The Cancer Imaging Program

The Quantitative Imaging Network

The National Cancer Institute

Then: 1939

And Now: 2016

Annual Report

January 2017

In Memoriam

It was a year ago that Dr. Larry Clarke passed away suddenly from acute myeloid leukemia. He was a pioneer in the field of quantitative imaging and its ability to measure or predict response to cancer therapies. We met at the 2016 Annual Meeting knowing that Larry was not well, but not knowing the extent of his illness. Now, we will be coming together for the 2017 meeting with a feeling of loss by his absence. It is a testimony to Larry's strength and dedication to the field of quantitative imaging that progress in the Network remains scientifically vibrant and active, moving toward the goas he set many years ago. Larry was not only a leader in this field, he was a good friend to everyone who participates in it.

Quantitative Imaging for Evaluation of Response to Cancer Therapies (U01)

The Quantitative Imaging Network

QIN

Fifth Annual Report

January 2017

FOREWORD

This is the fifth report of the Quantitative Imaging Network (QIN), covering the period from January 2016 to January 2017. The period has seen the stabilization of the network at a total of approximately 21 research teams with two of these teams funded by the Canadian Government (the Institute of Cancer Research of the Canadian Institutes of Health Research) The network of teams continues to create a number of collaborative projects among the members, and the Executive Committee continues to provide the governance for the Network, guiding it through external collaborations with organizations such as National Clinical Trials Network, (NCTN), the NCI Clinical Centers, QIBA, and other groups. This Executive Committee, comprised of the principal investigators from each of the QIN teams, will continue to grow as new teams enter the network.

Coordination of the internal activities of the network, including communication among the various working groups continues to be the responsibility of the Coordinating Committee. The chairs and co-chairs of the working groups comprise this committee. They meet every month by teleconference to discuss the activities of the working groups and to find avenues of collaboration whenever necessary. As the network grows and the domains of the working groups expand, this overall coordination is needed to ensure the most efficient use of time and resources within the network. A new responsibility of the Coordinating Committee during this past year has been the review of challenges. A challenge is an activity where researchers can test their algorithms against a common dataset to evaluate performance characteristics.

The activities of the working groups continue to evolve as the Network grows and matures. Although this is only the fifth written report from the QIN, the Network is in its ninth year of operation. Thus, several research teams have completed early-phase development and optimization, and have moved into the important work of clinical validation. This progress is presented in the individual reports from the more advanced teams. In addition, several teams have "graduated" from the Network, having completed a five-years research cycle. Despite this, they remain active in the consensus building and have contributed to this report.

Reporting individual team progress in a joint report is somewhat difficult because while some teams are just beginning their research in quantitative imaging in QIN, others are nearing the end of the five-year program period. Thus, there is a staggered level of accomplishment within the network. This staggering has led to considerable mentoring of the newer participants in the network by the more mature ones, and has created an environment where the new teams are contributing to overall progress more quickly than the older teams did when they were beginning in the QIN. This is because pathways to progress and solutions to organizational issues have been solved by the older teams, and the newer ones are inheriting the results. This is a distinct advantage for working within the organization structure of a network.

A meeting was held in December 2016 between members of QIN and the NCTN (National Clinical Trials Network) to discuss methods for placing quantitative tools in upcoming clinical trials. This was very helpful to the QIN members present, and future meeting on this subject are planned for the coming year. The overall goal for QIN remains

the translation of robust and well-tested tools into the clinical setting, and meeting such as the ones planned will be a big help in that direction.

This was the final year for operation under the funding opportunity announcement PAR-14-116. A new announcement PAR-17-128 and its companion announcement PAR-17-129 will be in force beginning May 2017. In the initial announcement, teams entering the QIN will propose a phased approach to development, optimization and validation of quantitative tools for measuring or predicting response to therapy. In the first phase of one or two years, the investigators will develop and optimize tools of their choosing. A review of progress will determine whether the team will proceed to the second phase, where tools are tested and validated in multisite clinical trials. The duration of the first phase is the choice of the investigator, and the total duration of the two phases cannot exceed five years.

For teams with well-developed and optimized tools already for clinical testing, the second PAR is available. This U01 mechanism is similar to the existing plan for QIN members. The NCI program staff is hopeful that this change in approach to quantitative imaging direction will accelerate translation to clinical adaption.

Robert J. Nordstrom
QIN Director

Table of Contents

Contributions to QIN from US Technical Teams

U01CA143062: Radiomics of NSCLC

H. Lee Moffitt Cancer Center

Robert Gillies, Ph.D.

INTRODUCTION

The first 5-year tenure of this award ended 02/29/2015, and was continued for an additional year under a no cost extension. The original grant was focused on developing the field of "radiomics"; that is, the conversion of images to structured, mineable data. At the end of this period, it was decided to change tack and use the radiomics approach to address a clinically relevant question and perhaps, change clinical practice. Hence, the renewed focus of the U01CA143062 program is to use radiomics to predict recurrence following surgical resection or early stage lung cancers, and thus inform the decision whether to treat with adjuvant chemotherapy. The projected accrual of this new project is 4558 domestic and 1675 foreign NSCLC patients; creating the largest such highly crated data set to date. A renewal with this focus was submitted 11/05/2014 and was reviewed on 1/30/2015, receiving an Impact Score of 24, which was not funded. An A1 application was submitted 07/02/2015 and was reviewed on 11/12/2015, receiving an Impact Score of 13, which was funded at the end of the cycle. The NGA was received on 8/12/2016 and it was finally activated by Moffitt 1/06/2017. Hence, the entirely of 2016 was unfunded by the QIN program. Nonetheless, we continued to be productive in a number of avenues, described below.

DISCUSSION OF PROGRESS

§ General Radiomics Opinion Pieces

As Sir Ernest Rutherford (1871-1937) said during a hiatus in his lab's funding funding: "Gentlemen, we have run out of money. It is time to start thinking". We continue to write reviews and contribute to opinion pieces on the current and future status of Radiomics. In particular, there were two high-impact opinion pieces, below, that emanated from QIN networks in the US (Yankeelov) and the UK (O'Connor) that are beginning to shape the way quantitative imaging is practiced in oncology:

1. Yankeelov TE, et al. Clin Cancer Res. 2016.
2. O'Connor JP, et al., Nature Reviews Clinical Oncology. 2016.

Gillies published a paper in Radiology with Hedy Hriackh and Paul Kinehan that described radiomics for a practicing radiologist, and took a look into the "radiology reading room of the future" with shared data, and "habitat imaging".

3. Gillies RJ, Kinahan PE, Hricak H. Radiology. 2016.

Finally, based on some comments made at an international meeting, Gillies and Tomas Beyer were invited to write an opinion piece on PET/MR in Cancer Research, which also discussed habitat imaging at length.

4. Gillies RJ, Beyer T. Cancer Research. 2016.

§ Radiogenomics of Lung Cancer

As a continuation of the original 5 year tenure of the grant, the U01 group finalized analyses comparing NSCLC CT features to gene expression. In NSCLC, the most important and well known driver mutations are k-ras, Alk fusion protein, and EGFR activating mutation; the latter two of which are actionable. The rationale behind a radiomic approach is that: 1) genomic information is often not available, while CT data are; 2) genomic information can have high false negatives either through sample preparation or sampling artifacts. Hence it was felt that an association of CT feature with gene expression, whether or not it was orthogonal information, could contribute to the clinical management of patients. The first of the papers looked at k-ras mutations in a large cohort with the interesting finding that spiculation was the only feature that was strongly associated with k-ras status and that k-ras status was not associated with survival. The only feature strongly predictive of non-survival was pleural attachment:

5. Wang H et al., Clinical lung cancer. 2016.

A further paper using this same cohort associated radiomic features with Alk and EGFR mutations:

6. Wang H et al. European journal of radiology. 2016.

Two papers focused in on EGFR mutation status in two large multi-institutional cohorts used both semantic (radiologist scored) and agnostic (computer derived) features. Both studies used large cohorts of 385 (manually scored by 2 radiologists) and 298 (computer extracted) patients. Significant associations with EGFR mutation status were found with a number of semantic features, which in turn were much more predictive of outcome than were the agnostic features:

7. Liu Y et al. Clinical Lung Cancer. 2016.
8. Liu Y et al. Radiology. 2016.

§ Nodule Diagnosis

We also initiated work under the U01 to begin investigating CT scans from then National Lung Screening Trial (NLST). A major effort was spent to curate and build appropriate cohorts for study, and this revealed otherwise unknown aspects wherein risk of death could be assessed based on screening history. Based primarily on his outstanding epidemiological work on this, Dr. Schabath was invited to be co-PI of the U01 renewal, which will require significant curation and cohort building.

9. Schabath MB et al. PLoS One. 2016.

Using these cohorts, we have begun our radiomic analyses and have shown, quite convincingly that quantitative computer derived features of CT screen-detected nodules at baseline can predict emergence of cancer 1 or 2 years hence. Notably, over half of the patients had radiomics scores in the lowest or highest quartiles, and these predicted good and bad outcomes, respectively, with >93% accuracy. This work was published in JTO, where it received the "Editor's choice" award in Dec., 2016 and was the subject of an editorial (http://www.jto.org/article/S1556-0864(16)31066-8/fulltext).

10. Hawkins S et al. Journal of Thoracic Oncology. 2016.

Further, we have performed a "semantic" analysis of incidentally detected indeterminate pulmonary nodules from 172 patients in collaboration with Pierre Massion's group at Vanderbilt (102 training, 70 test). 24 radiological traits were scored by 3 radiologists and observed that a parsimonious set of 4 features could predict the presence or absence of cancer with an AUROC of 0.88 (train) and 0.80 (test).

11. Liu Y et al. Clin Cancer Res. 2016.

§ Participation in QIN and QIBA activities

In addition to the above, we have also participated in a number of joint efforts between QIN members. Many of these efforts finally bore fruit in 2016, such as our participation in the QIBA lung nodule volume estimation challenge:

12. Athelogou M et al. Academic radiology. 2016.

We undertook a multi-institutional study in collaboration with Drs. Kalpathy-Cramer, Napel, and Zhou of the QIN to compare lung segmentation algorithms:

13. Kalpathy-Cramer et al. Journal of digital imaging. 2016;

This work was followed up by a multi-institutional study comparing the robustness of radiomic features extracted form segmented lung nodules:

14. J. Kalpathy-Cramer, et al. Tomography, 2016.

Finally, through his leadership of the PET-CT committee of the QIN, Dr. Goldgof has participated in a number of studies, one which was published in 2016 comparing 3-D PET segmentations across institutions:

15. R. Beichel et al. JMRI, 2016.

§ Future: Deep Learning, Habitats, and Rad-Path

Radiomics as it was first conceived involved the segmentation of tumors and extraction of 100's of quantitative features to describe them; and the subsequent mining of these data to generate decision support. While this continues to be a goal, new techniques and approaches are emerging that may supplant or complement this approach. One such approach is "habitat imaging", which combines orthogonal data sets to identify sub-regions within tumors with similar physiologies, and the subsequent extraction of radiomic features from these regions. This approach obviates the need for segmentation. Currently, Habitat imaging is best performed with MRI, so we have invested some effort into diseases, such as GBM (Zhou) and prostate (Stoyonava) for which MRI is routinely practiced clinically:

16. Zhou M et al. J Magn Reson Imaging. 2016.
17. Stoyanova R et al. Oncotarget. 2016.

Notably, in this latter paper, biopsy locations are marked by MR-US fusion and thus there is an effort to identify gene expression and histopathology associated with the Habitats observed in MR images. Habitat imaging lends itself to co-registration with pathology, and we have begun to explore quantitative histopathology measures of intra-tumoral heterogeneity. This work showed important distinctions between the edge and core of tumors, which is something that we observed radiomically in 2015 (Grove et al., PLoS One 2015; PMCID 4349806):

18. Lloyd MC et al. Cancer Research. 2016.

Another approach that is just emerging is the analyses of images with deep learning, or convolutional neural nets, CNNs. This holds promise as: 1) it is a mature technology in other classification paradigms (facial recognition, defense applications); and 2) it may prove to be immune to image acquisition heterogeneity, which is a challenge in radiomics. We have made an initial foray into this space, and will continue to pursue it in the future:

19. Paul R et al. Tomography. 2016.

PLANS FOR NEXT YEAR

Our plans for next year include:

1. Establish the data entry pipeline via Clinica and NCBI.
2. Populate the data base with > 100 patients from each site (Tianjin and Moffitt)
3. Disseminate Data Management tools and CDEs to EDRN, MCL and other groups who may have an interest.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

§ Publications

1. Yankeelov TE, Mankoff DA, Schwartz LH, Lieberman FS, Buatti JM, Mountz JM, Erickson BJ, Fennessy FM, Huang W, Kalpathy-Cramer J, Wahl RL, Linden HM, Kinahan PE, Zhao B, Hylton NM, Gillies RJ, Clarke L, Nordstrom R, Rubin DL. Quantitative Imaging in Cancer Clinical Trials. Clin Cancer Res. 2016;22(2):284-90. doi: 10.1158/1078-0432.CCR-14-3336. PubMed PMID: 26773162; PubMed Central PMCID: PMC4717912.
2. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G, Buckley DL, Chenevert TL, Clarke LP, Collette S, Cook GJ, deSouza NM, Dickson JC, Dive C, Evelhoch JL, Faivre-Finn C, Gallagher FA, Gilbert FJ, Gillies RJ, Goh V, Griffiths JR, Groves AM, Halligan S, Harris AL, Hawkes DJ, Hoekstra OS, Huang EP, Hutton BF, Jackson EF, Jayson GC, Jones A, Koh DM, Lacombe D, Lambin P, Lassau N, Leach MO, Lee TY, Leen EL, Lewis JS, Liu Y, Lythgoe MF, Manoharan P, Maxwell RJ, Miles KA, Morgan B, Morris S, Ng T, Padhani AR, Parker GJ, Partridge M, Pathak AP, Peet AC, Punwani S, Reynolds AR, Robinson SP, Shankar LK, Sharma RA, Soloviev D, Stroobants S, Sullivan DC, Taylor SA, Tofts PS, Tozer GM, van Herk M, Walker-Samuel S, Wason J, Williams KJ, Workman P, Yankeelov TE, Brindle KM, McShane LM, Jackson A, Waterton JC. Imaging biomarker roadmap for cancer studies. Nature reviews Clinical oncology. 2016. doi: 10.1038/nrclinonc.2016.162. PubMed PMID: 27725679.
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-77. doi: 10.1148/radiol.2015151169. PubMed PMID: 26579733; PubMed Central PMCID: PMC4734157.
4. Gillies RJ, Beyer T. PET and MRI: Is the Whole Greater than the Sum of Its Parts? Cancer research. 2016;76(21):6163-6. doi: 10.1158/0008-5472.CAN-16-2121. PubMed PMID: 27729326.
5. Wang H, Schabath MB, Liu Y, Stringfield O, Balagurunathan Y, Heine JJ, Eschrich SA, Ye Z, Gillies RJ. Association Between Computed Tomographic Features and Kirsten Rat Sarcoma Viral Oncogene Mutations in Patients With Stage I Lung Adenocarcinoma and Their Prognostic Value. Clinical lung cancer. 2016;17(4):271-8. doi: 10.1016/j.cllc.2015.11.002. PubMed PMID: 26712103; PubMed Central PMCID: PMC4887405.
6. Wang H, Schabath MB, Liu Y, Han Y, Li Q, Gillies RJ, Ye Z. Clinical and CT characteristics of surgically resected lung adenocarcinomas harboring ALK rearrangements or EGFR mutations. European journal of radiology. 2016;85(11):1934-40. doi: 10.1016/j.ejrad.2016.08.023. PubMed PMID: 27776643; PubMed Central PMCID: PMC5123695.
7. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, Ye Z, Gillies RJ. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. Clinical lung cancer. 2016;17(5):441-8 e6. doi: 10.1016/j.cllc.2016.02.001. PubMed PMID: 27017476.

8. Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, Ye Z, Gillies RJ. CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. Radiology. 2016;280(1):271-80. doi: 10.1148/radiol.2016151455. PubMed PMID: 26937803; PubMed Central PMCID: PMC4934516
9. Schabath MB, Massion PP, Thompson ZJ, Eschrich SA, Balagurunathan Y, Goldof D, Aberle DR, Gillies RJ. Differences in Patient Outcomes of Prevalence, Interval, and Screen-Detected Lung Cancers in the CT Arm of the National Lung Screening Trial. PloS one. 2016;11(8):e0159880. doi: 10.1371/journal.pone.0159880. PubMed PMID: 27509046; PubMed Central PMCID: PMC4980050.
10. Hawkins S, Wang H, Liu Y, Garcia A, Stringfield O, Krewer H, Li Q, Cherezov D, Gatenby RA, Balagurunathan Y, Goldgof D, Schabath MB, Hall L, Gillies RJ. Predicting Malignant Nodules from Screening CT Scans. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2016;11(12):2120-8. doi: 10.1016/j.jtho.2016.07.002. PubMed PMID: 27422797.
11. Liu Y, Balagurunathan Y, Atwater T, Antic S, Li Q, Walker RC, Smith G, Massion PP, Schabath MB, Gillies RJ. Radiological Image traits Predictive of Cancer Status in Pulmonary Nodules. Clin Cancer Res. 2016. doi: 10.1158/1078-0432.CCR-15-3102. PubMed PMID: 27663588.
12. Athelogou M, Kim HJ, Dima A, Obuchowski N, Peskin A, Gavrielides MA, Petrick N, Saiprasad G, Colditz Colditz D, Beaumont H, Oubel E, Tan Y, Zhao B, Kuhnigk JM, Moltz JH, Orieux G, Gillies RJ, Gu Y, Mantri N, Goldmacher G, Zhang L, Vega E, Bloom M, Jarecha R, Soza G, Tietjen C, Takeguchi T, Yamagata H, Peterson S, Masoud O, Buckler AJ. Algorithm Variability in the Estimation of Lung Nodule Volume From Phantom CT Scans: Results of the QIBA 3A Public Challenge. Academic radiology. 2016;23(8):940-52. doi: 10.1016/j.acra.2016.02.018. PubMed PMID: 27215408.
13. Kalpathy-Cramer J, Zhao B, Goldgof D, Gu Y, Wang X, Yang H, Tan Y, Gillies R, Napel S. A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. Journal of digital imaging. 2016;29(4):476-87. doi: 10.1007/s10278-016-9859-z. PubMed PMID: 26847203; PubMed Central PMCID: PMC4942386.
14. R. Beichel, B. Smith, J. Ulrich, C. Bauer, P. Ahmadvand, M. Budzevich, R. Gillies, D. Goldgof, M. Grkovski, G. Hamarneh, Q. Huang, P. Kinahan, C. Laymon, E. Moros, J. Mountz, J. Muzi, M. Muzi, S. Nehmeh, M. Oborski, Y. Tan, B. Zhao, J. Sunderland, J. Buatti, "Multi-site Quality and Variability Analysis of 3D FDG PET Segmentations based on Phantom and Clinical Image Data", Journal of Magnetic Resonance Imaging, 2016 (accepted for publication).
15. J. Kalpathy-Cramer, A. Mamomov, B. Zhao, L. Lu, D. Cherezov, S. Napel, S. Echegaray, M. McNitt-Gray, P. Lo, J.C. Sieren, J. Utho , S.K.N. Dilger, B. Driscoll, I. Yeung, L. Hadjiiski, K. Cha, Y. Balagurunathan, R. Gillies, D.Goldgof, "Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features", Tomography Journal, Special QIN Issue, V. 2(4), pp. 430-437, 2016.
16. Zhou M, Chaudhury B, Hall LO, Goldgof DB, Gillies RJ, Gatenby RA. Identifying spatial imaging biomarkers of glioblastoma multiforme for survival group prediction. J Magn Reson Imaging. 2016. doi: 10.1002/jmri.25497. PubMed PMID: 27678245.

17. Stoyanova R, Pollack A, Takhar M, Lynne C, Parra N, Lam LL, Alshalalfa M, Buerki C, Castillo R, Jorda M, Ashab HA, Kryvenko ON, Punnen S, Parekh DJ, Abramowitz MC, Gillies RJ, Davicioni E, Erho N, Ishkanian A. Association of multiparametric MRI quantitative imaging features with prostate cancer gene expression in MRI-targeted prostate biopsies. Oncotarget. 2016. doi: 10.18632/oncotarget.10523. PubMed PMID: 27438142.
18. Lloyd MC, Cunningham JJ, Bui MM, Gillies RJ, Brown JS, Gatenby RA. Darwinian Dynamics of Intratumoral Heterogeneity: Not Solely Random Mutations but Also Variable Environmental Selection Forces. Cancer research. 2016;76(11):3136-44. doi: 10.1158/0008-5472.CAN-15-2962. PubMed PMID: 27009166.
19. Paul R, Hawkins SH, Balagurunathan Y, Schabath MB, Gillies RJ, Hall LO, Goldgof DB. Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma. Tomography : a journal for imaging research. 2016;2(4):388-95. doi: 10.18383/j.tom.2016.00211. PubMed PMID: 28066809; PubMed Central PMCID: PMC5218828.

§ Presentations

01/2016 Schabath MB "Diagnostic and predictive quantitative-imaging features in lung cancer screening". Oral Presentation at AACR-IASLC International Joint Conference: Lung Cancer Translational Science from the Bench to the Clinic, San Diego, California

01/2016 Schabath MB, Gillies RJ*. "Radiomics of lung cancer". Oral Presentation at AACR-IASLC International Joint Conference: Lung Cancer Translational Science from the Bench to the Clinic, San Diego, California. *Presented by Dr. Gillies.

01/2016 Gillies, RJ. "Imaging Tumor Habitats" Danny Thomas Lecture, St. Jude's. Memphis TN.

02/2016 Gillies, RJ. "Imaging Habitats in Cancer" 5th Tübingen PET/MR Workshop, University of Tübingen, Germany

04/2016 Gillies, RJ. "Imaging Habitats of Cancer" AACR, New Orleans, LA

05/2016 LO Hall "Leveraging Big Data in Medical Image Analysis", Nanjing University of Science and Technology, Nanjing, China.

05/2016 LO Hall "Leveraging Big Data in Medical Image Analysis", Invited Talk, 2016 International Conference on Intelligence Science and Big Data Engineering, Guangzhou, China.

08/2016 Schabath MB "Diagnostic and predictive quantitative-imaging features in lung cancer screening". Oral Presentation at 15th Annual Guangdong Congress of Radiology, GuangZhou, China

07/2016 Schabath MB "Radiomics and Lung Cancer Screening". Oral Presentation at NCI Lung Cancer SPORE Workshop, Bethesda, Maryland

10/2016 Gillies, RJ. "Radiomics in Decision Support"; 3rd Personalized Medicine Conference, invited speaker, Orlando, FL

10/2016 Gillies, RJ. "Whither Radiomics?" 6th Annual Radiomics Workshop, meeting organizer, Clearwater Beach, FL

11/2016 LO Hall "Transfer Learning using Deep Features for Medical Image Analysis", University of Notre Dame,

12/2016 Gillies, RJ. “Radiomics and Tumor Habitats”; 28th EORTC-NCI-AACR symposium, invited speaker, Munich, Germany

§ Conference Proceedings

I. Tunali, J.E. Gray, J. Qi, M. Abdullah, Y. Balagurunathan, R.J. Gillies, M.B. Schabath. Quantitative Imaging Features Predict Response of Immunotherapy in Non-Small Cell Lung Cancer Patients. Int’l Assoc. Study of Lung Cancer; Vienna, AUT; 01/2017

D. Cherezov, S. Hawkins, D. Goldgof, L. Hall, Y. Balagurunathan, R.J. Gillies, M.B. Schabath. Quantitative Imaging Features Predict Incidence Lung Cancer in Low-Dose Computed Tomography (LDCT) Screening01/2017. Int’l Assoc. Study of Lung Cancer; Vienna, AUT; 01/2017

D. Cherezov, S. Hawkins,, D. Goldgof, L. Hall, Y. Balagurunathan, R. Gillies, M. Schabath, "Improving Prediction through Selecting Features Informed by Nodule Size Ranges in NLST", IEEE International Conference on Systems, Man and Cybernetics (SMC 2016), Budapest, Hungary, 10/2016.

R. Paul, S, Hawkins, L. Hall , D. Goldgof, R. Gillies, "Combining Deep Neural Network and Traditional Image Features to Improve Survival Prediction Accuracy for Lung Cancer Patients from Diagnostic CT", IEEE International Conference on Systems, Man and Cybernetics (SMC 2016), Budapest, Hungary, 10/2016.

H. Farhidzadeh, B. Chudhury, J. Scott, D. Goldgof, L. Hall, R. Gatenby, R. Gillies, M. Raghavan, "A Quantitative Histogram-based Approach to Predict Treatment Outcome for Soft Tissue Sarcoma Using Pre- and Post-treatment MRIs", IEEE International Conference on Systems, Man and Cybernetics (SMC 2016), Budapest, Hungary, 10/2016.

R. Liu, L. Hall, D. Goldgof, M. Zhou, R. Gatenby, K. Ahmed, "Exploring Deep Features from Brain Tumor Magnetic Resonance Images via Transfer Learning", 2016 International Joint Conference on Neural Networks, (IJCNN 2016), Vancouver, Canada, 7/2016.

B. Chaudhury, M. Zhou, D. Goldgof, L. Hall, R. Gatenby, R. Gillies, J. Drukteinis, "Predicting Ki67 expression from DCE-MR images of breast tumors using textural kinetic features in tumor habitats", SPIE Medical Imaging 2016, San Diego, CA, 2/2016.

B. Geiger, S. Hawkins, L. Hall, D. Goldgof, Y. Balagurunathan, R. Gatenby, R. Gillies, "Change Descriptors for Predicting Tumor Malignancy in NLST CT Screening Images", SPIE Medical Imaging 2016, San Diego, CA, 2/2016.

H. Farhidzadeh, J. Scott, D. Goldgof, L. Hall, R. Gatenby, R. Gillies, M. Raghavan, "Signal IntensityAnalysis of Ecological Defined Habitats in Soft Tissue Sarcomas to Predict Metastasis Development", SPIE Medical Imaging 2016, San Diego, CA, 2/2016.

§ Radiomics Retreat 2016

We again hosted the Radiomics Retreat in Clearwater Beach on Oct. 24-26. This was supported by generous gifts from both the Moffitt Cancer Center and the Department of Radiology at Stanford University. There were 102 attendees from 47 different institutions.

Attendees came from US, China, Germany, Denmark, and Canada. The summary agenda is attached. Once again, there was a sponsored Young Investigators dinner.

The major emerging theme at this meeting was the growing interest in Deep Learning and AI applied to medical images, and this will be a focus of this meeting going forward. The agenda for the meeting appears on the next page.

MON AM				Email
8:00	Continental Breakfast			
9:00	Gillies	Robert	Welcome, announcements, and: "Whither Radiomics?"	
	S1	DATA SETS (Bob Gillies moderator)		
9:45	Kalpathy-Cramer	Jayashree	Feature challenge	DATA SET EXTENDED
10:15	Bakr	Shaimae	NSCLC dataset	Available data sets
10:30	Fuller	Clifton	TBN	Available Data Sets
10:45	Kinahan	Paul	Image accrual in clinical trials	Available data sets
11:00	Kirby	Justin	"The Cancer Imaging Archive: Review of new data and functionality"	Available data sets
11:15	Schabath	Matthew	NLST	Data Sets
12:00	LUNCH			
MON PM				
13:00	Napel	Sandy	QIN Special Issue	
	S2	DATA ANALYTICS (John Hazle, moderator)		
13:45	Echegaray	Sebastian	title?	Feature robustness
14:00	Goetz	Michael	"Radiomics with MITK"	software Toolkit
14:15	Hall	Lawrence	Exploring the use of deep features in Radiomics	Predictive Models and Results
14:30	Hazle	John	Tentative - The need for Quality Control in Quantification of Imaging Biomarkers	robustness of Features
14:45	Mackin	Dennis	Results from CCR Phantom Studies	Robustness of features
15:00	BREAK			
15:45	Mihaylov	Ivaylo	CT texture features stability dependence on the image binning approach	Robustness of features
16:00	Shafiq ul Hassan	Muhammad	"Voxel volume dependent features in CT Radiomics"	Robustness of features
16:15	Yip	Stephen	PET radiomics	
16:30	Vallieres	Martin	The ONCORAY initiative: preliminary results	
17:00	Veeraraghavan	Harini	Correlating Inter-Site Tumor Heterogeneity to outcomes for recurrent ovarian cancers	
17:45	DINNER on Own			
	Young Investigators dinner (AAPM)			
Tue AM				
8:00	Continental Breakfast			
9:00	Aerts	Hugo	PLENARY	
	S3a.	PREDICTIVE MODELS and RESULTS (Jae Tian moderator)		
9:45	Dong	Di	Prediction of therapeutic resistance by a CT phenotypic signature in stage IV *EGFR*-mutant non-small cell lung cancer patients treated with tyrosine kinase inhibitors	
10:00	Fu	Srui	Clinical treatment of hepatocellular carcinoma	
10:15	Goldgof	Dmitry	Improving malignancy prediction through features selection informed by nodule size ranges in NLST	Predictive Models and Results
10:30	Hansen	Adam	'Combined Hyperpolarized Carbon MRS and PET in dog cancer patients'	'New imaging biomarkers'
10:45	Katiyar	Prateek	Multi-view learning on multiparametric PET MRI decodes intratumor heterogeneity	Predictive Models and Results
11:00	Li	Allen	"Radiation response assessment based on daily CT radiomics"	Response Monitoring (Delta-Radiomics)
11:15	Li	QIAN	Radiomics of RT response	PM&R
11:30	Liu	Zhenyu	Radiomics analysis allows for precise prediction of seizure occurrence in patients with low-grade gliomas	PM&R
11:45	Liu	Zaiyi	Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer	PM&R
12:00	Lunch			
	S3b	PREDICTIVE MODELS and RESULTS (Hugo Aerts moderator)		
13:00	Parekh	Vishwa	Multiparametric Radiological Imaging Radiomics & Geodesics for Cancer Detection & Characterization	Predictive Models and Results
13:15	Rusu	Mirabela	Predicting Diffuse Lung Disease on High Resolution CT	Predictive Models and results
13:30	Vallières	Martin	Assessing the risk of tumour recurrences and metastases in head and neck cancer by combining radiomics and clinical variables via imbalance-adjusted machine learning	Predictive Models and Results
13:45	Wei	Jingwei	"Therapy guiding system for stage-B HCC patients based on Radiomics"	
14:00	Wu	Jia	"Quantitative image features from DCE-MRI to predict treatment response in breast cancer"	Response Monitoring (Delta-Radiomics)
14:15	Yi	Darvin	Probably something on deep learning in GBM*	Predictive Models and Results
14:30	Break			
15:00	TBD		TBD	
	S4	HABITATS And RAD/PATH (Paul Kinehan, moderator)		
15:45	Balagurunathan	Yoga	Habitats in PrCa	habitts
16:00	Corollier	Thibaud	Genotype-Phenotype association in meningioma patients	RPC
16:15	Dimitrova	Nevenka	Genome Analytics framework for integrating expression and MRI data	Radiomic-Pathomic-Genomic Correlations
16:30	Erickson	Bradley	"MRI textures predict glioma genomics"	Radiomic-Pathomic-Genomic Correlations
16:45	ford	John	BOLD MRI as a Potential Imaging Biomarker for Prostate Cancer	Radiomic-Pathomic-Genomic Correlations
17:00	Raghunand	Natarajan	"Habitat imaging"?	habitat?
17:15	stoyanova	radka	Radiogenomics in PrCa habitats	habitats
17:30	Zhou	Mu	"Radiogenomics map of non-small cell lung cancer identifies relationships between molecular and imaging phenotypes with prognostic implications"	Radiomic-Pathomic-Genomic Correlations.
18:00	Dinner on Own			
WED AM				
7:30	Breakfast			
	S5	PREDICTIVE MODELS part deux (Sandy Napel moderator)		
8:30	Mitchell	Ross	Lessons learned from 20 years of Radiomics research -- & talk regarding difficulty in obtaining funding support??	
9:00	Chellis	Jack	Tensor Flow for Deep Learning on Images	
9:15	Mazurowski	Maciej	Computational imaging biomarkers and radiogenomics in breast and brain cancers	
9:30	Milchenko	Mikhail	Informatics and analysis tools for quantitative brain tumor research	
9:45	Simpson	Garrett	Texture changes during MR-guided radiotherapy of head and neck	
10:00	van Dijk	Lanne	PET and CT image biomarkers improve prediction of late radiation-induced xerostomia	
	S6	Network planning and recap		
			open session moderated by Gillies, Napel and Aerts	

U01 CA140206: Quantitative Imaging to Assess Response in Cancer Therapy Trials

University of Iowa

John M. Buatti, MD
Thomas L. Casavant, PhD
Michael M. Graham, PhD MD
Milan Sonka, PhD

INTRODUCTION

The University of Iowa QIN team has been consistently committed to improve and develop tools for quantitative image analysis both for assessment of response and for tumor targeting. The group remains committed to the QIN central mission of "improving the role of quantitative imaging for clinical decision making in oncology by the development and validation of data acquisition, analysis methods and tools to tailor treatment to individual patients and to predict or monitor response to drug or radiation therapy."

Our team completed its 6th year of participation and made significant progress building on both developed infrastructure and through multi-institutional working group teams as part of QIN. Several new publications highlight this progress and the interdisciplinary and inter-institutional efforts being led within QIN. Our group continues to move forward on 4 specific aims that build creatively from our previous work in a highly innovative fashion and help accelerate QIN progress and collaboration.

§ Specific Aim 1:

Develop a novel, robust imaging genomics-based decision support platform using a combination of our successful Phase-I developed and validated highly automated quantitative image analysis methods applied to linked and publicly-available well curated image (TCIA) and molecular (The Cancer Genome Atlas–TCGA) data warehouses along with an established outcomes database for H&N cancers. This will facilitate new methods and decision support tools necessary for future risk adaptive trials that will certainly include both genomic and quantitative image data.

§ Specific Aim 2:

Build and innovate based on Phase-I developed and validated image analysis tools: a) Apply highly and fully automated quantitative image analysis methods to a cooperative group data set of H&N cancers, b) Develop unique new tools through creative new image analysis methods for application to FLT/PET in H&N cancer, FLT/PET in pelvis and bone marrow, as well as DOTATOC for liver metastases in neuroendocrine cancers. These newly refined approaches will be made publicly available and will contribute to future clinical trials,

decision support, quantitative imaging response assessment and therapy targeting in a variety of cancer sites.

§ Specific Aim 3:

Create a novel link between our established work in PET quantification and calibration phantoms with our image analysis and decision support tools to create a clinically practical open source automated phantom analysis tool that can be applied to national efforts aimed to improve quantitative imaging quality assurance for clinical trials across multiple modalities including PET, CT, and MRI. This will provide a critical tool for improving the ease, accuracy and harmonization for clinical trials data acquisition.

§ Specific Aim 4:

Adapt, enhance and extend quantitative image-based response assessment in clinical trial decision-support through relevant active clinical trials. Several clinical trials are highlighted exploring: 1) FLT-PET as a predictor of bone marrow activity and toxicity in pelvic malignancies treated with chemoradiotherapy, 2) DOTATOC as an indicator of disease burden in neuroendocrine tumors and 3) quantitative MR imaging [T2, T1, T1ρ, quantitative susceptibility mapping (QSM) and MRSI] as effective predictors of response in malignant glial tumors treated with intravenous high dose vitamin C. These trials will facilitate quantitative image analysis tool development, decision support tools and risk adaptive approaches in future clinical trials.

DISCUSSION OF PROGRESS

During the previous period our efforts continue in several major integrated activities. Clinical data is provided for analysis including outcomes data using both an established head and neck cancer data base that was initiated as part of our phase I effort. In addition, we are increasingly reaching out to national data bases such as the TCGA and TCIA for other curated data sets. There is a team of clinicians and computer engineers and physicists that have worked closely to develop a group of computer algorithms applicable for quantitative analysis of PET images in FDG. This tool is not only available in 3D Slicer but now has been complemented by instructional videos on the Iowa QIN website (http://qin.iibi.uiowa.edu). Review of these videos and methods highlights the potential applicability of these tools for active clinical trials and makes practical application more feasible. The tools provide fully automated liver uptake measurements for normalization of PET/CT images and also provide "just enough interaction" methods for tumor segmentation in a series of head and neck squamous cancers on FDG PET-CT. Such tools further enable the calculation of a large number of radiomics features as well as automated and consistent response assessment or targeting routines for clinical decision making.

In the past year, a major activity included coordination of a challenge that evaluated the ability of 7 institutions to analyze a series of PET image data sets of both phantoms and clinical head and neck cancer cases [9]. This included analysis of a group of different image acquisition routines as well as different methods used as standard practice at the institutions.

Simultaneously the decision support group has worked on developing the pipeline for integrated analysis of both genomic and quantitative image analysis data through utilization of both TCIA and TCGA data. In the coming year additional clinical data and genome data will be added. Our group continues to tightly integrate efforts through biweekly meetings that discuss progress of our teams of bioinformaticists, computer engineers, statisticians, radiation physicists, nuclear medicine physicians, radiologists and radiation oncologists. We remain focused on advancing tools that can more effectively provide quantitative imaging based response assessment in cancer clinical trials.

§ Aim 1

During this past year, our efforts have been focused on refinement and enhancement of our genomic variant analysis pipeline necessary to identify highly-informative features for prediction and decision support. This novel informatics pipeline utilizes currently-available TCGA data for H&N cancers for which TCIA data is also available. Currently, the TCGA repository contains molecular data for more than 500 H&N squamous cell cancer patients. This molecular data consists of 5 modalities including: 499 with full or targeted exome (DNA) sequences corresponding to both tumor (T) and normal (N) samples, 526 with Copy Number Variation (CNV) data from either genome-wide single nucleotide polymorphisms (SNP) arrays or low-pass high-throughput DNA sequencing or both, 528 with epigenetic/methylome (HumanMethylation450) data, and 505 with high throughput RNAseq data for tumor samples. Figure 1 illustrates the basic analysis derived from alignments of this sequence data in three ways. Alignment of an N sample to the UCSC reference genome sequence will reveal the set of all germline (G) plus somatic (S) mutations present. Alignment of the T sample to the UCSC reference will reveal the additional set of tumor-specific T mutations present. One hundred forty eight H&N cases (of the 528 in TCGA) have some imaging data in TCIA. Within this H&N subset with imaging and TCGA data, 30 subjects have PET imaging data, 144 have CT, and 11 have MRI.

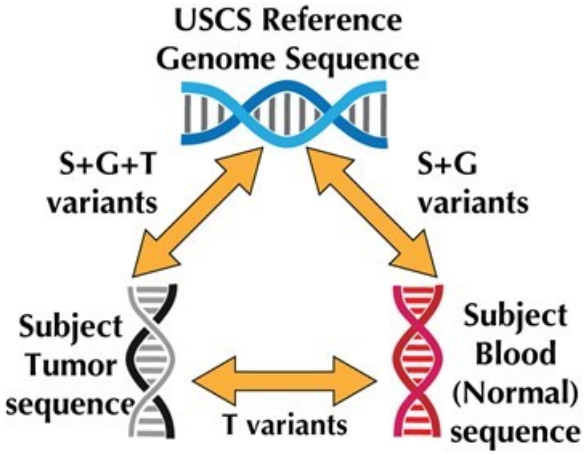


Figure 1: Simplified schematic of variant analysis. G-germline, S-somatic, and T-tumor-specific variants. Exome sequence available for blood (normal) DNA, and tumor DNA. The three comparisons represent possible choices of clustering features.

The dataset driving our pipeline development efforts consists of 512 TCGA subject exomes (>33Gb), and associated variant calls (regardless of imaging data availability). These files contain on average 44,507 T variants per subject (tumor *vs.* normal, Figure 2). Of the 22,787,886 (non-distinct) variants total, 1,683,888 *distinct variants* were reported across all 512 subjects. We first intersect this with a set of 70 genes most often associated with H&N cancer from the annotated genomic start of translation to genomic stop. The resulting set of 45,261 variants is then filtered by eliminating variants that appear in both the entire set of 512 subjects (not informative), and in fewer than 5% (25) of the 512 subjects. This resulting set of 1,836 variants is further filtered, so that only those with a minor allele frequency (MAF) of 0.5% or less from the 1000 genomes project were retained. This MAF filter reduces the number of variants per subject to 51 known SNPs, however in the T samples there were an additional 626 private mutations with no MAF values, bringing the total number of variants of interest to 677.

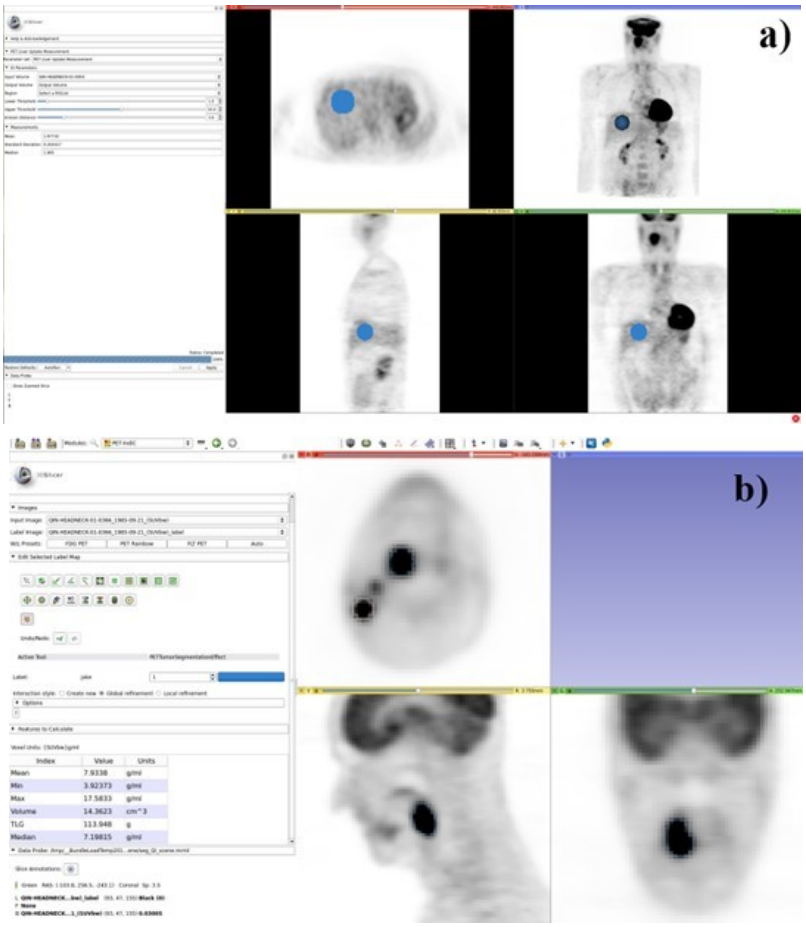


Figure 2: Examples for developed, freely available open-source software for quantitative PET image analysis in 3D Slicer. (a) 3D Slicer PET Liver Uptake Measurement tool. (b) 3D Slicer PET-IndiC tool for lesion segmentation and generation of quantitative uptake features for treatment outcome prediction.

We next construct the subset of 29 subjects with corresponding PET/CT data available, so that we could apply our quantitative image analysis methods to systematically extract image-based features within this subgroup. We continue the refinement of the set of 677 rare genomic variants by requiring that a variant must appear in at least 4, but in not more than 19 instances within the set of 29 cases analyzed. We thus arrived at a set of 395 distinct genomic loci within this subgroup of patients with TCGA and PET/CT data, and 16 QI metrics derived from our Phase I image analysis tool from which to build predictive models. As we continue with expansion of the set of patients, as well as our improved selection of features of interest for machine learning, we will be releasing our feature selection algorithms to the QIN community through shared BIDS working group tools and interfaces.

§ Aim 2 and 3

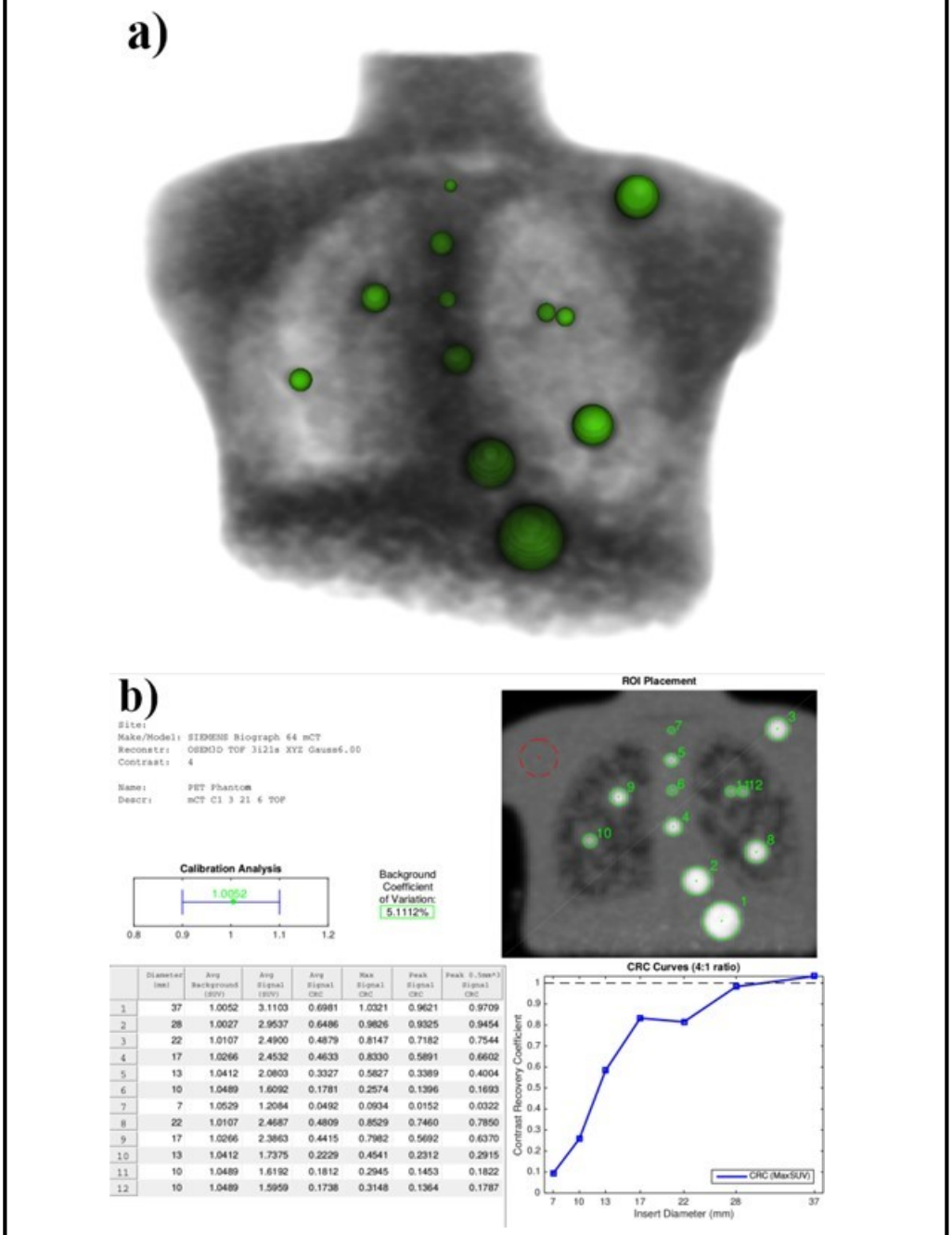
We have updated and improved our publicly released open-source software for quantitative PET image analysis, consisting of 3D Slicer PET Tumor Segmentation, 3D Slicer PET DICOM Extension, PET Liver Uptake Measurement (Fig. 2(a)), and 3D Slicer PET-IndiC (Fig. 2(b)) Extension as well as supporting libraries. To better document the released software and facilitate the broad dissemination, we have published a summary paper, which describes details of the implemented lesion segmentation algorithm as well as its validation [3]. We have also established a website: http://qin.iibi.uiowa.edu providing instructional videos and demonstrations that are fully narrated for public use. We are hopeful that this will facilitate a broad utilization of our developed tools.

We have developed a fully-automated quantitative PET phantom analysis algorithm, which allows the user to segment ACR/ACRIN-ECOG, SNMMI/CTN, and NEMA NU-2 image quality phantoms and will help to simplify the process of PET scan image quality assessment (Figure 3).

To augment the Iowa H&N PET/CT image collection (already available on TCIA collection: "QIN-HeadNeck"), we have encoded a) segmentations and quantitative measurements of lesions derived from Iowa H&N PET/CT image data and b) clinical data related to the Iowa H&N PET/CT image data in standard conform DICOM format and published it on TCIA, resulting in one of the most complex DICOM data collections currently available on TCIA. The data is accompanied by a recently published paper, which describes the data collection as well as underlying design decisions regarding the selected data representation in DICOM format [3].

We have finished the UI-led QIN PET Phantom and Clinical Head and Neck Segmentation challenge and have written a summary paper [9], which provides valuable insight on how to improve (multi-site) quantitative PET image analysis performance.

Development activities for an FLT based tool for head and neck cancer as well as for DOTATOC for tumor burden in liver are also under development but are not yet mature.



	Diameter (mm)	Avg Background (SUV)	Avg Signal (SUV)	Avg Signal CRC	Max Signal CRC	Peak Signal CRC	Peak (1.5mm^3) Signal CRC
1	37	1.0052	3.1103	0.6981	1.0321	0.9621	0.9709
2	28	1.0027	2.9537	0.6486	0.9826	0.9325	0.9454
3	22	1.0107	2.4900	0.4879	0.8147	0.7182	0.7544
4	17	1.0266	2.4532	0.4633	0.8330	0.5891	0.6602
5	13	1.0412	2.0803	0.3327	0.5827	0.3389	0.4004
6	10	1.0489	1.6092	0.1781	0.2574	0.1396	0.1693
7	7	1.0529	1.2084	0.0492	0.0934	0.0152	0.0322
8	22	1.0107	2.4687	0.4809	0.8529	0.7460	0.7850
9	17	1.0266	2.3863	0.4415	0.7982	0.5692	0.6370
10	13	1.0412	1.7375	0.2229	0.4541	0.2312	0.2915
11	10	1.0489	1.6192	0.1812	0.2945	0.1453	0.1822
12	10	1.0489	1.5959	0.1738	0.3148	0.1364	0.1787

Figure 3: Example of the fully-automated analysis of a PET Torso phantom. (a) Volume rendering of automatically identified inserts, and (b) corresponding quantitative analysis report.

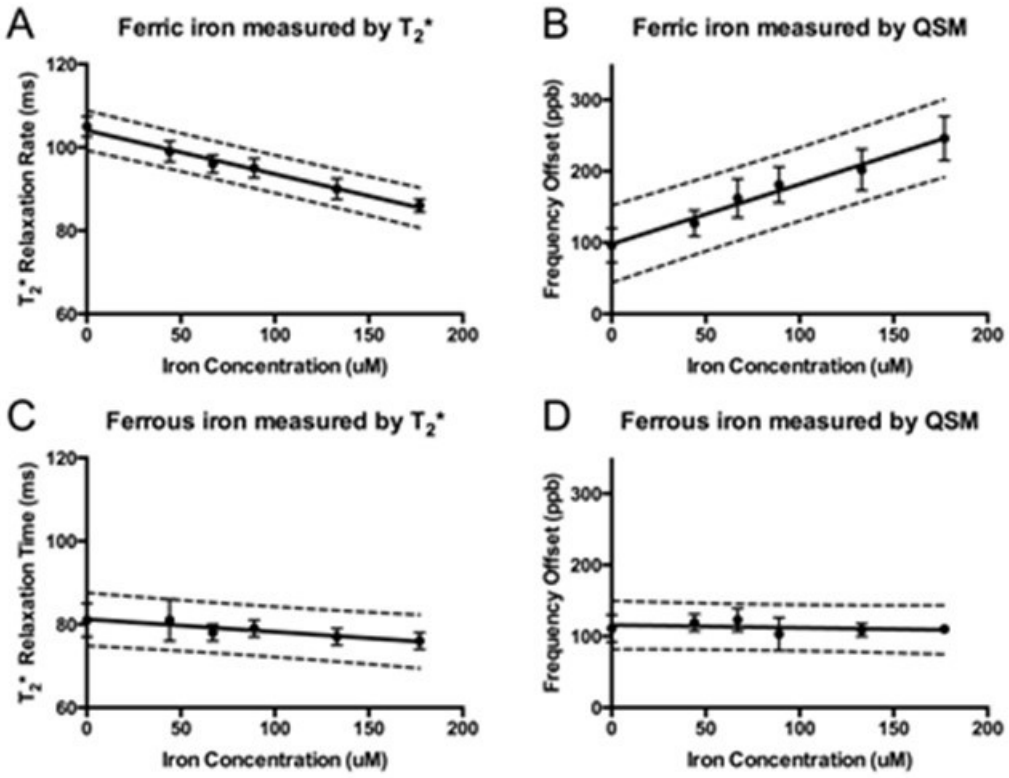


Figure 4: Iron (III) is primarily responsible for the alterations seen in T_2* relaxation time and tissue susceptibility in phantoms**.** T_2* relaxation times (Panels A and C) were calculated by fitting a mono-exponential decay curve to the data on a voxel-by-voxel basis. Quantitative susceptibility maps (Panels B and D) were generated using the MEDI algorithm. Mean and standard deviation within the phantom are plotted relative to the iron concentration.

§ Aim 4

We are pursuing imaging methods to assess tumor response to pharmacological ascorbate as an adjuvant to standard of care therapy. Peak plasma concentrations of ascorbate are currently measured as part of the trial but do not directly report the concentrations within the tumor. Therefore, we have pursued the development of methods capable of directly quantitating ascorbate within the tumor and measuring markers attributable to the reduction of labile iron by ascorbate. Imaging presents a unique opportunity for directly quantitating ascorbate *in vivo* and assessing an indirect marker of ascorbate efficacy. Since high concentrations of ascorbate have been shown *in vitro* to reduce labile Iron(III) to Iron(II), we have pursued the evaluation and reliability of T_2* relaxation times and tissue susceptibility to detect subtle changes in the net iron oxidation state that may occur after ascorbate infusion. Phantoms containing physiological concentrations of Iron(II) and Iron(III) were evaluated

using the quantitative imaging metrics to determine the sensitivity of each imaging metric to sensitivity to varying concentrations of each iron oxidation state. We have demonstrated that Iron(III) has a greater influence on T_2* relaxation times and tissue susceptibility than does an equal amount of Iron(II). Work is on-going in phantoms and pre-clinical models to further evaluate this relationship.

SIGNIFICANT RESULTS, INCLUDING MAJOR FINDINGS, DEVELOPMENTS, OR CONCLUSIONS (BOTH POSITIVE AND NEGATIVE)

The significant results from our research have been published or are being published as noted in the text and below. We have successfully developed a robust segmentation for commonly used phantoms and also evaluated the current methods for PET segmentation through a QIN based challenge, which resulted in an accepted paper in Medical Physics [9]. We believe these methods will improve harmonization and enable more facile analysis needed for clinical trials image acquisition consistency. We are making good progress according to our overall project timeline although proceeding with internal DNA sequencing and obtaining National studies has been slower than we had hoped. We believe continued progress in the coming year will enable completion of all elements of the proposed research.

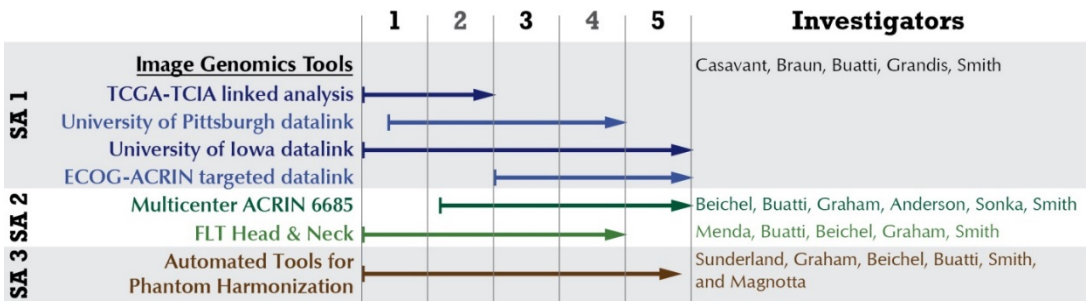


Table 1: Gantt Chart for mapping progress.

PLANS FOR NEXT YEAR

In the coming year we will begin sequencing head and neck cancers from our own University of Iowa database. We will also continue to work with national databases and begin image analysis on cooperative group data either from ECOG/ACRIN or the NRG or both. These will be analyzed using the automated methods and compared to traditional analysis. We will work to continue to define the platform for genomic-radiomic analyis in head and neck cancer. We also plan to make progress on our FLT based tools and DOTATOC tools. Publication of our phantom tool should also be accomplished and integrated into our website. Initial work on defining an MR based tool will also be pursued. Further progress on MR

imaging will be determined by progress on our phase II trial using ascorbate with standard therapy.

LIST OF QIN REFERENCES, PUBLICATIONS AND PRESENTATIONS

§ References and Publications

1. Graham MM, Wahl RL, Hoffman JM, Yap JT, Sunderland JJ, Boellaard R, Perlman ES, Kinahan PE, Christian PE, Hoekstra OS, Dorfman GS. Summary of the UPICT Protocol for 18F-FDG PET/CT Imaging in Oncology Clinical Trials. *J Nucl Med.* Jun 2015;56(6):955-961. PMID: 25883122. PMCID: 4587663.

2. Beichel RR, Van Tol M, Ulrich EJ, Bauer C, Chang T, Plichta KA, Smith BJ, Sunderland JJ, Graham MM, Sonka M, Buatti JM. Semiautomated segmentation of head and neck cancers in 18F-FDG PET scans: A just-enough-interaction approach. *Med Phys.* Jun 2016;43(6):2948. PMID: 27277044. PMCID: 4874930.

3. Fedorov A, Clunie D, Ulrich E, Bauer C, Wahle A, Brown B, Onken M, Riesmeier J, Pieper S, Kikinis R, Buatti J, Beichel RR. DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research. *PeerJ.* 2016;4:e2057. PMID: 27257542. PMCID: 4888317.

4. Kurland BF, Aggarwal S, Yankeelov TE, Gerstner ER, Mountz JM, Linden HM, Jones EF, Bodeker KL, Buatti JM. Accrual Patterns for Clinical Studies Involving Quantitative Imaging: Results of an NCI Quantitative Imaging Network (QIN) Survey. *Tomography.* Dec 2016;2(4):276-282. PMID: 28127586. PMCID: 5260812.

5. McGuire SM, Bhatia SK, Sun W, Jacobson GM, Menda Y, Ponto LL, Smith BJ, Gross BA, Bayouth JE, Sunderland JJ, Graham MM, Buatti JM. Using [(18)F]Fluorothymidine Imaged With Positron Emission Tomography to Quantify and Reduce Hematologic Toxicity Due to Chemoradiation Therapy for Pelvic Cancer Patients. *Int J Radiat Oncol Biol Phys.* Sep 01 2016;96(1):228-239. PMID: 27319286. PMCID: 4982822.

6. Pierce Ii LA, Byrd DW, Elston BF, Karp JS, Sunderland JJ, Kinahan PE. An algorithm for automated ROI definition in water or epoxy-filled NEMA NU-2 image quality phantoms. *J Appl Clin Med Phys.* Jan 08 2016;17(1):5842. PMID: 26894356. PMCID: 4874494.

7. Yankeelov TE, Mankoff DA, Schwartz LH, Lieberman FS, Buatti JM, Mountz JM, Erickson BJ, Fennessy FM, Huang W, Kalpathy-Cramer J, Wahl RL, Linden HM, Kinahan PE, Zhao B, Hylton NM, Gillies RJ, Clarke L, Nordstrom R, Rubin DL. Quantitative Imaging in Cancer Clinical Trials. *Clin Cancer Res.* Jan 15 2016;22(2):284-290. PMID: 26773162. PMCID: 4717912.

8. Farahani K, Kalpathy-Cramer J, Chenevert TL, Rubin DL, Sunderland JJ, Nordstrom RJ, Buatti J, Hylton N. Computational Challenges and Collaborative Projects in the NCI Quantitative Imaging Network. 2016.

9. R.R. Beichel, B.J. Smith, C. Bauer, E.J. Ulrich, P. Ahmadvand, M.M. Budzevich, R.J. Gillies, D. Goldgof, M. Grkovski, G. Hamarneh, Q. Huang, P.E. Kinahan, C.M.

Laymon, J.M. Mountz, J.P. Muzi, M. Muzi, S. Nehmeh, M.J. Oborski, Y. Tan, B. Zhao, J.J. Sunderland, and J.M. Buatti, "Multi-site Quality and Variability Analysis of 3D FDG PET Segmentations based on Phantom and Clinical Image Data," *Med Phys*, 2016, in print.

§ Presentations

1. Yusung Kim PhD, Kaustubh Anil Patwardhan MS, Reinhard R Beichel PhD., Brian J Smith, PhD, Ethan J.Ulrich BS, Christopher Mart MS, Tangel Chang, MD, Kristin A Plichta MD, John M Buatti MD. *Impact of Contouring Accuracy on Expected Tumor Control Probability for Head and Neck Cancer: Semi-Automated Segmentation versus Manual Contouring.* ASTRO Annual meeting 2016, Boston Convention Center, September 25-28, 2016

2. Rendleman MCR. Poster presentation: "*Developing Machine Learning Tools for Cancer Prediction from Genomic Data*". University of Iowa College of Engineering Research Open House, Iowa City, IA. April 7, 2016.

3. John Buatti, MD. *An Introduction to the Quantitative Imaging Network*. ASTRO Annual meeting 2016, Boston Convention Center, September 25-28, 2016

4. John Buatti, MD. *Quantitative Imaging Based Tools for Radiation Therapy Targeting and Response Assessment in Head and Neck Cancer*. ASTRO Annual meeting 2016, Boston Convention Center, September 25-28, 2016

U01CA148131: Advanced PET/CT Imaging for Improving Clinical Trials

University of Washington and the Seattle Cancer Care Alliance

Paul Kinahan, Ph.D. and Hannah Linden, M.D.

INTRODUCTION

The goal of this project is to improve cancer clinical trials by enhancing the effectiveness of quantitative PET/CT imaging of tumor response. This has three distinct and linked components:

1. Develop and implement a unified database and imaging platform for our phantoms and software tools.
2. Extend our biologically principled imaging tools developed for FDG to FLT (proliferation) and FES (receptor status) in multicenter studies.
3. Prospectively test the integration of the above tools and methods in a newly approved ECOG-ACRIN clinical trial that uses FES PET imaging to evaluate new breast cancer therapies.

DISCUSSION OF PROGRESS

We list methods, results, and conclusions for each Aim in order below.

§Aim 1

Develop and implement a unified database and imaging platform for our phantoms and software tools.

We have completed an evaluation of our PET/CT cross-calibration kit, which were designed in collaboration with RadQual, and are now available as a commercial product called the PET F-18 X-Cal System (Figure 1). The X-cal is designed to allow the monitoring of biases in SUV values by enabling the monitoring of biases. The kits, which contain sealed, long-lived 68Ge/68Ga sources in an epoxy matrix, were subjected to tests to evaluate the repeatability and reproducibility of their measurements, including tests on scanners and dose calibrators from multiple manufacturers across a network of local PET imaging centers.

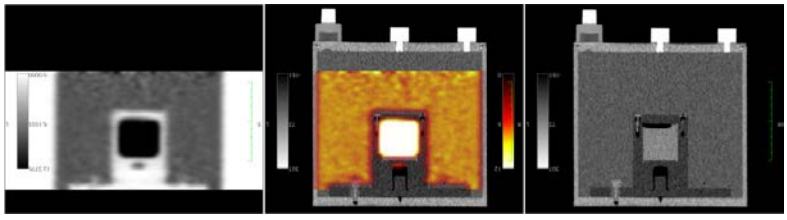


Figure 1: PET, PET+CT fused, and CT image of the X-Cal PET phantom inside a 20 cm diameter phantom filled with water. In this example a small amount of 18F-FDG was added to the background water.

Each X-Cal kit contains three sealed 68Ge/68Ga sources in an epoxy matrix for use in a PET/CT scanner, dose calibrator, and well counter, respectively. Each source's activity is known to ±2.5% with a 95% confidence level. The dose calibrator reference sources are directly traceable to NIST (National Institute of Standards and Technology) standards. The scanner and well counter sources are implicitly NIST-traceable, i.e. they are made following the same procedures but are not certified by NIST.

In testing at multiple sites, per-site average recovery coefficients ranged from 0.907 to 0.983, with per-site standard deviations between 0.019 and 0.034. The 24 measurements overall had a mean of 0.944 ± 0.038. Dose calibrator recovery coefficients were 0.964 ± 0.033. For a single site, Figure 2 (right) shows the estimated SUV bias calculated from the recovery coefficients as described by $b = \left(R_P / R_D\right) - 1$, here b is the estimated SUV bias, R_P and R_D are the recovery coefficients for the PET phantom and the dose calibrator sources. A comparison of the pre-test PET scanner and dose calibrator biases did not show any correlations in the biases.

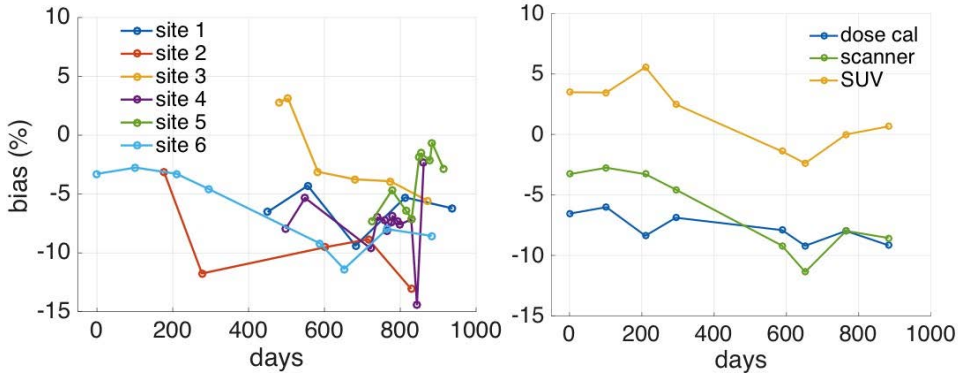


Figure 2: (left) PET scanner recovery values versus time from 6 local network hospitals. 24 scans were collected over 2.5 years. Overall the ensemble recovery coefficient had a mean of 0.944 ± 0.038. (right) Scanner, dose calibrator, and resultant SUV bias versus time for a single site showing variability of SUV bias in time.

These results were published in the QIN special issue of Tomography [1] showing longitudinal variations in bias at single-center and multi-center studies. The X-cal phantom kit was deployed in a QIN multi-center study that has completed analysis and is being submitted for publication.

Aim 2

Extend our biologically principled imaging tools developed for FDG to FLT (proliferation) and FES (receptor status) in multicenter studies.

We developed a method called 'virtual clinical trials to evaluate variations in the PET imaging process to characterize the ability of static and dynamic metrics to measure breast cancer response to therapy in a clinical trial setting. We have competed and published three studies: Estimating the effect of FDG uptake time on lesion detectability in PET imaging of early stage breast cancer showing that delayed imaging improves detection [2], estimating the

effects of uptake time variability on required sample size showing that variability in uptake time can double the needed number of patient studies in clinical trials [3], and comparing static versus dynamic PET imaging in measuring response to breast cancer therapy showing that as expected, dynamic imaging improves the correct discrimination of response [4].

In this last study we generated 540 i.i.d. PET study realizations for each of 22 18F-FDG breast cancer patient studies pre- and post-therapy. Each noise realization accounted for known sources of variability in the imaging process. We then performed a ROC analysis on the resulting SUV and kinetic parameter uncertainty distributions to assess the impact of the variability on the measurement capabilities of each metric. Analysis showed that the kinetic macro parameter, Ki, shows more variability than SUVmax (CV of 16.6% compared to 13.5%). However, for the patients who did not show perfect separation between the pre- and post-therapy parameter uncertainty distributions (AUROC<1), dynamic imaging outperformed SUVmax in distinguishing metabolic change in response to therapy (14/16 patients, $p<0.05$).

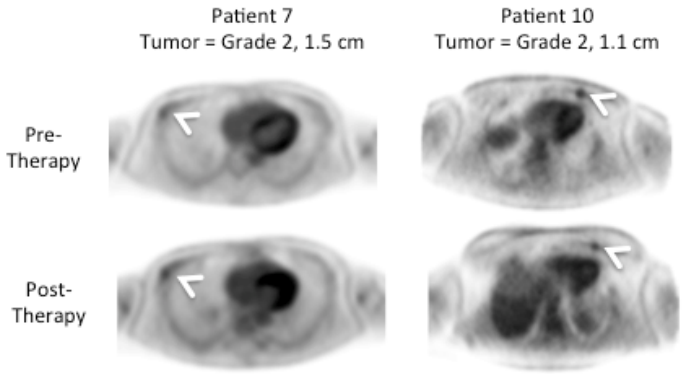


Figure 3: Two of the 22 patient studies used.

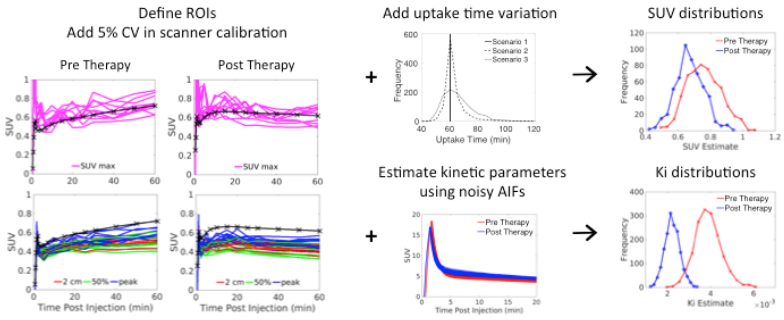


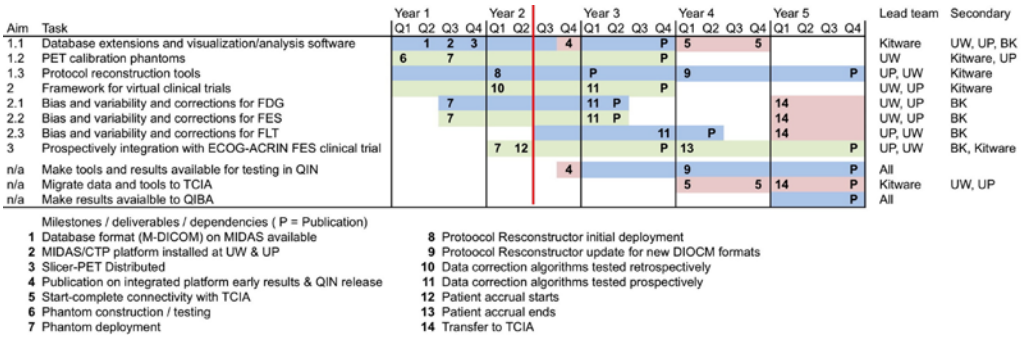
Figure 4: Virtual clinical trial simulation of the imaging process and results for patient 11. SUVmax at 60 minutes post-injection and dynamic ROIs were measured from the tumors in the reconstructed images. Uptake time uncertainty was added to SUVmax to generate the final SUVmax uncertainty distributions pre- and post-therapy. The dynamic ROIs were re-input into the kinetic model to generate the Ki uncertainty distributions.

Aim 3

Prospectively test the integration of the above tools and methods in a newly approved ECOG-ACRIN clinical trial that uses FES PET imaging to evaluate new breast cancer therapies.

We have constructed a new set of X-cal phantom kits for deployment in the ECOG-ACRIN trial I142 '[18F] Fluoroestradiol (FES) as a Predictive Measure for Endocrine Therapy in Women with Newly Diagnosed Metastatic Breast Cancer'. It is a multi-center trial for which Dr Linden is the co-PI.

PROGRESS AND PLANS FOR NEXT YEAR



LIST OF RECENT PUBLICATIONS

[1] Byrd DW et al. Tomography 2(4):353-360, 2016. PMID: 28066807. PMCID: PMC5214172.

[2] Wangerin KA, et al. Tomography 1:53-60, 2015. PMID: 26807443. PMCID: PMC4721230.

[3] Kurland BF et al. J Nucl Med 57:226-230, 2016. PMID: 26493206. PMCID: PMC4749350.

[3] Wangerin KA, et al. Phys Med Biol. (accepted) 2017.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, O'Sullivan F, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Besa C, Kalpathy-Cramer J, Mountz JM, Layman CM, Muzi M, Kinahan PE, Schmainda K, Cao Y, Chenevert T, Taoluli B, Yankeelov TE, Fennessy FM, Li X. The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. Tomography 2:56-66, 2016. PMID: 27200418. PMCID: PMC4869732

Rosen M, Kinahan PE, Gimpel J, Opanowski A, Hill GC, Weiss L, Shankar L. Performance Observations of Scanner Qualification of NCI-designated Cancer Centers: Results from the Centers of Quantitative Imaging Excellence (CQIE) Program. Academic Radiology. vol.24(2):232–245, 2017.

Scheuermann JS, Reddin JS, Opanowski A, Kinahan PE, Siegel BA, Shankar LK, Karp JS. Qualification of NCI-Designated Cancer Centers for Quantitative PET/CT Imaging in Clinical Trials. Journal of Nuclear Medicine (accepted) 2017.

U01CA190214: Qualification and Deployment of Imaging Biomarkers of Cancer Treatment Response

Stanford University Department of Radiology (1)

Daniel L. Rubin, M.D., M.S.

INTRODUCTION

Response to cancer therapy in clinical trials has traditionally been assessed via simple linear tumor size measurement on images. However, linear measurement may be less effective for newer targeted agents that can arrest tumor growth without causing shrinkage. While novel imaging biomarkers, such as those being developed in the NCI Quantitative Imaging Network (QIN), may be more appropriate for detecting and predicting treatment response to these agents, few as of yet have been used in clinical trials, primarily due to three major challenges: (1) it is difficult to introduce new imaging biomarkers into the workflow of clinical trials, since current image viewing tools are generally closed systems and limited to linear assessment of target lesions, and time does not allow for more complex human-guided measurements; (2) there are no decision support tools that can employ new quantitative imaging biomarkers to assess treatment response in individual patients or drug effectiveness in clinical trial cohorts; and (3) it is difficult to repurpose existing clinical trial imaging data to accrue aggregate evidence needed to show that new imaging biomarkers predict survival, thereby qualifying them as surrogate endponts in clinical trials

.

We recently developed the electronic Physician Annotation Device (ePAD) to facilitate collecting annotations and measurements on target lesions in compliance with standards in the cancer imaging community. In this proposal, we will leverage our prior work, our active collaborations with current QIN researchers, and our engagement with the ECOG-ACRIN national cooperative group to develop and evaluate a software platform, algorithms, and tools that meet all of these challenges.

Our project will tackle the foregoing challenges by developing a software platform that incorporates ePAD for image viewing, enhancing it with a plugin architecture to deploy novel quantitative imaging biomarkers developed by QIN and other researchers, and by providing tools that facilitate translating and evaluating novel imaging biomarkers in clinical trials. Our infrastructure will contain a workflow engine that computes these novel imaging biomarkers during image interpretation, and tools for decision making about treatment response and drug effectiveness based on them. It will also enable repurposing imaging data from previous clinical trials to assess the benefit of these imaging biomarkers for predicting treatment response.

Our flexible platform and tools will have substantial impact in cancer research and ultimately in clinical care, specifically by (1) advancing cancer research and accelerating clinical trials by enabling novel quantitative imaging biomarkers being developed by QIN researchers and others, which may be more appropriate for newer, targeted anti-cancer agents,

to be introduced into the clinical trial workflow, (2) improving both clinical trials and clinical practice by providing decision support about cancer treatment response based on these biomarkers, and (3) accelerating the acquisition of sufficient data needed to qualify new and potentially better imaging biomarkers of cancer treatment response and survival.

DISCUSSION OF PROGRESS

§ Specific Aims

Specific Aim 1: We will develop a platform and tools to facilitate deploying new imaging biomarkers in clinical trials and using them for decision support. We will create a plug-in mechanism to our ePAD platform that allows novel quantitative imaging algorithms developed by us or by others to be incorporated into the clinical trial workflow with minimal impact on the time required for image interpretation. To assess individual and cohort response based on new imaging biomarkers, we will develop decision support tools that summarize their output in relation to clinical outcome. We will also develop tools that compare the assessments of novel and conventional (e.g., linear dimension) imaging biomarkers of cancer treatment response.

Specific Aim 2: We will develop methods to repurpose existing imaging data from clinical trials to study new imaging biomarkers. We will develop automated image segmentation methods that use seed points from conventional clinical trial lesion measurements to derive volumetric lesion outlines, from which novel quantitative imaging biomarkers of treatment response can be computed efficiently in the workflow of clinical trials. With the ultimate goal of generalizability, we will develop and deploy two exemplar quantitative image biomarkers: (1) target lesion volume in carcinoid tumors imaged by CT and (2) functional quantitative image parameters in hepatocellular carcinoma (HCC) imaged by MRI. We will deploy these as plugins to our ePAD platform so that they can be used for repurposing existing imaging data, and can be incorporated into the clinical trial workflow.

Specific Aim 3: We will deploy and evaluate our platform and tools in the core imaging laboratories of two cancer centers and the ECOG-ACRIN national cooperative group. We will apply our tools retrospectively to a recently-completed ECOG-ACRIN cooperative group trial (carcinoid tumors imaged by CT, linear measure vs. volumetric image biomarkers to assess treatment response) and a prospective investigator-initiated trial (HCC imaged by MRI, linear measure vs. novel functional quantitative MRI biomarkers to assess treatment response), with image assessments performed at two cancer centers (Stanford and Vanderbilt University). For both studies, we will compare the efficiency of the analysis done with and without our platform. Finally, we will use aggregate image biomarker data we acquire in conjunction with survival data from these clinical trials to study the important hypothesis that radiological response based on quantitative image biomarkers can predict overall survival.

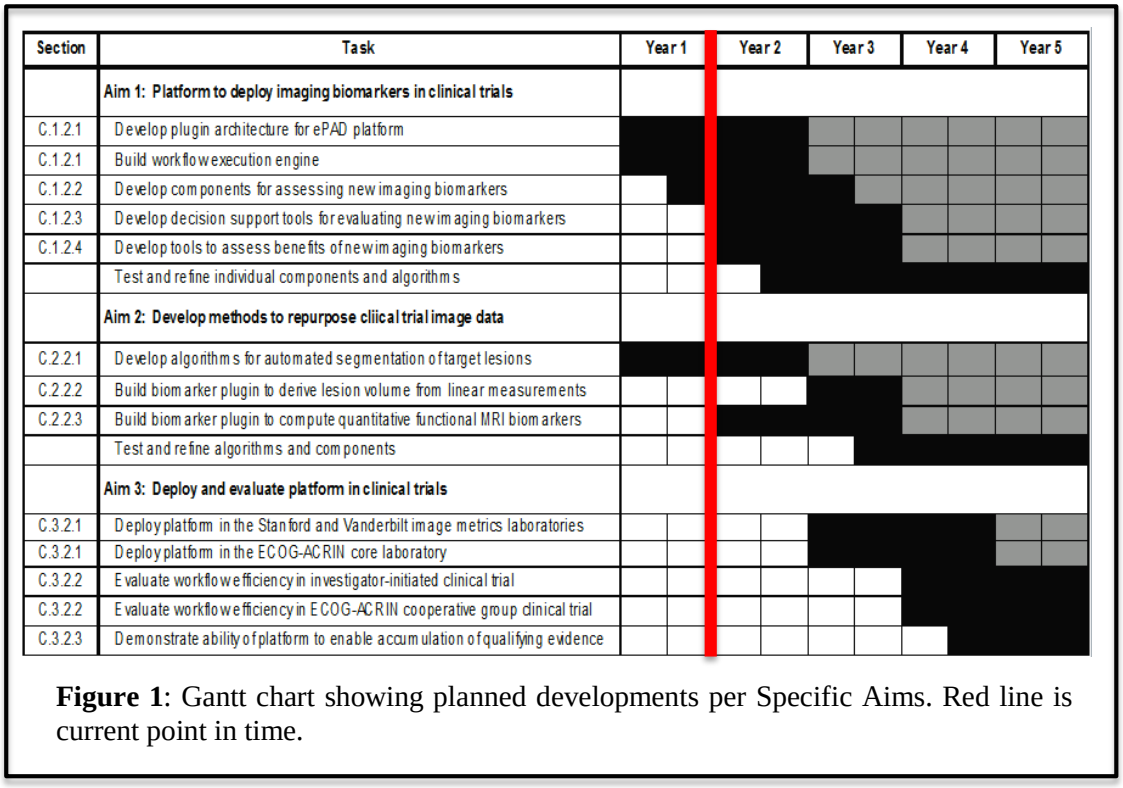


Figure 1: Gantt chart showing planned developments per Specific Aims. Red line is current point in time.

§ Progress on the Specific Aims

Our specific objectives and progress against these Aims for Years 1 & 2 were to: (labels C.n.m refer to our grant proposal and the Gantt Chart (Figure 1)):

<u>AIM 1: Develop a suite of configurable image feature characterization algorithms:</u>

C.1.2.1 **Plugin architecture and workflow execution engine for deploying new imaging biomarkers**: We completed a prototype of a plugin mechanism to our ePAD platform that allows novel quantitative imaging algorithms to be incorporated into the clinical trial workflow. There are three components of this architecture:

- **Biomarker plugins** are code modules that can be added to the ePAD virtual machine to execute the algorithms that QIN or others develop to compute novel imaging biomarkers, or for producing automated segmentation of lesions during image viewing.
- **Application modules** are software applications that leverage data in the ePAD platform, typically implemented as web-based applications that access data in ePAD via a RESTful application interface. We created an application to show images that have similar imaging features, called BIMM (Fig. 2) and an application that tracks lesions and produces a Word file summary of treatment

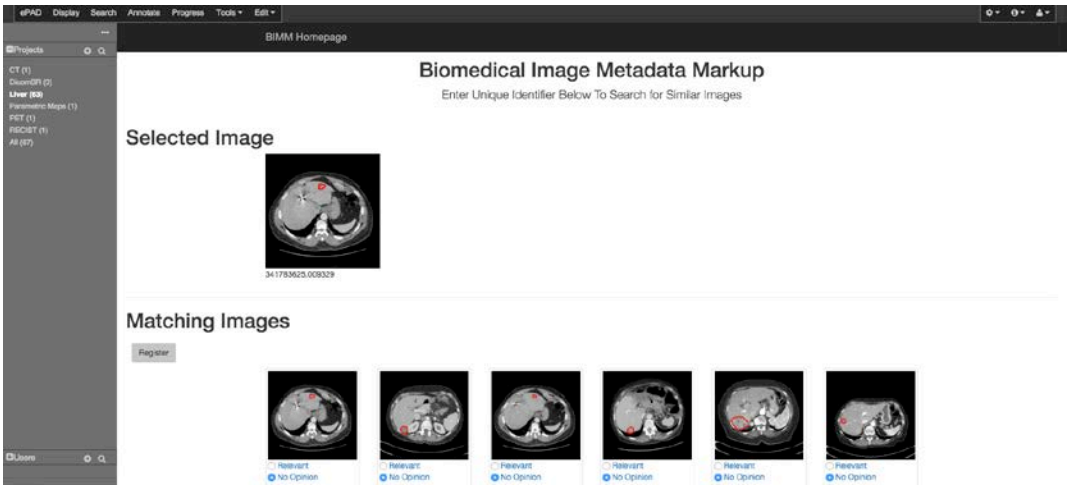


Figure 2: ePAD application: BIMM. This application, a "biomedical image metadata markup manager" searches the database of image features collected by ePAD to find similar images. The figure shows a query image ("selected image") and the results of searching for images that contain lesions having similar features to that in the query image ("matching images").

response that can be entered into the record of a clinical trial based on our recent work [1].

- **Workflow execution engine**: We have made plans to implement this via an interface to the Quantitative Imaging Feature Pipeline (QIFP), a separate QIN project undertaking developing a workflow engine.

We already are seeing third party developers beginning to develop new imaging biomarkers into ePAD [2]. In addition, one of the QIN sites is going to deploy ePAD for collecting quantitative imaging data in an upcoming multi-site trial.

We also made enhancements to the core ePAD functionality, including more robust DICOM segmentation object (DSO) support, interoperability with 3D Slicer, and better AIM template support. We also began supporting AIM interoperability with DICOM-SR.

C.1.2.2 **Image viewing to facilitate assessment of quantitative imaging biomarkers**: This includes the several tasks (see Plans), and we made progress on one of the tasks:

- **Managing projects, users, and clinical trial information**: ePAD now groups images into projects, managing users, and securing access to the data. It also associates radiologists with the images they interpret, so that ePAD viewer can

produce summaries of image interpretations that need to be performed in clinical trials. This functionality has been helpful in a new project that adopted ePAD: the MGH/HST Martinos Center for Biomedical Imaging) used this for MEDICI project.

In addition, we engaged actively in community outreach and dissemination. We have set up a public website for ePAD, http://epad.stanford.edu/ that contains introductory material, a demo move, documentation, a detailed description of the developer interface, and download information. ePAD is open source, and the license is posted as well. We have regular releases, at least 6 times per year, and release notes are at https://epad.stanford.edu/documentation/release-notes.

AIM 2: Develop methods to repurpose existing imaging data from clinical trials to study new imaging biomarkers:

C.2.2.1 **Exemplar #1a—Automated segmentation of cancer lesions**: Cancer lesions are challenging to segment since they vary in appearance in different organs. We created a novel method for adaptive estimation of active contour parameters for lesion segmentation (Fig. 3) [3]. The method is fully automatic once the lesion has been detected. The location of the level set contour relative to the lesion is estimated using a convolutional neural network (CNN). The output CNN probabilities are then used to adaptively calculate the parameters of the active contour functional during the segmentation process. Finally, the adaptive window size surrounding each contour point is re-estimated by an iterative process that considers lesion size and spatial texture. We evaluated the method in a diverse dataset of 164 MRI and 112 CT images of liver lesions that includes low contrast and heterogeneous lesions as well as noisy images. Our method, as assessed by Dice similarity coefficients, performed significantly better than currently available methods. An average Dice improvement of 0.27 was found across the entire dataset over all comparisons. We also analyzed two challenging subsets of lesions and obtained a significant Dice improvement of **0.24** with our method ($p < 0.001$, Wilcoxon) [3].

PLANS FOR NEXT YEAR

We will continue our software developments as follows (labels C.n.m refer to our grant proposal and the Gantt Chart (Fig. 1)):

C.1.2.1 **Plugin architecture and workflow execution engine for deploying new imaging biomarkers**:

- **Biomarker plugins** We will continue developing new plugins and incorporating those submitted by the community. In particular we will work with Dr. Abramson to create a plugin to compute a biomarker of response based on pixel histogram characteristics along the long axis (“ADLA plugin”).
- **Application modules**: We will develop applications to enable tracking lesions and producing waterfall plots (see C.1.2.3).

- **Workflow execution engine**: We will link ePAD to QIFP, a separate QIN project undertaking developing a workflow engine, to provide this functionality.

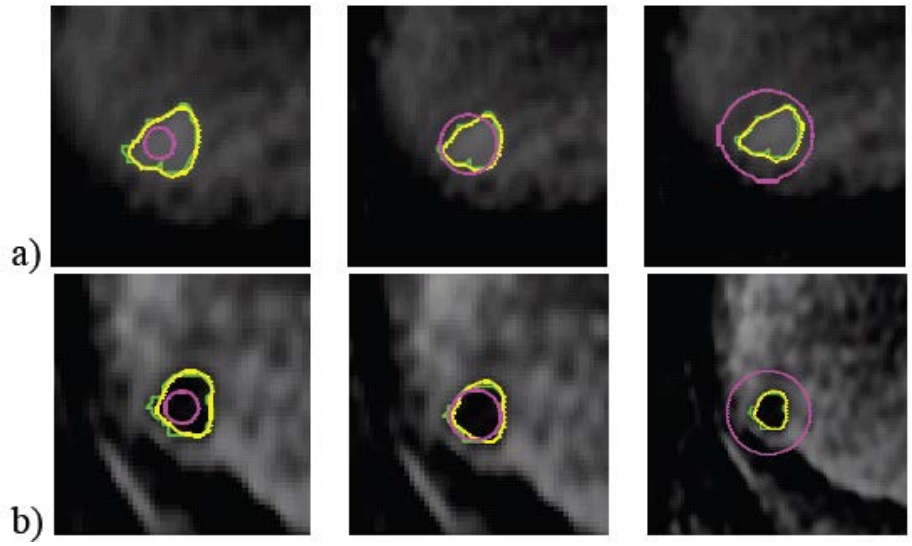


Figure 3: Lesion segmentation using the proposed method with different initializations. Left column - small initialization (3-pixels radius), middle column – more accurate initialization (5-pixels radius), right column – large initialization (9-pixels radius). a) low-contrast lesion, b) noisy and heterogeneous tissue surrounding the lesion. For both cases, lesion is located close to the liver boundary. Magenta – initial contour, yellow – our final segmentation, green – manual radiologists' annotation.

C.1.2.2 **Image viewing to facilitate assessment of quantitative imaging biomarkers**: We will pursue several tasks:

- **Facilitating image biomarker assessments by clinical centers**, such as incorporating several tools into ePAD viewer that streamline the assessment of quantitative imaging biomarkers.
- **Facilitating oversight of image readings by clinical trial researchers and sponsors**: We will develop a study monitoring application module that permits ePAD to monitor the status of image interpretations made in multiple clinical trials and summarized as a table in ePAD viewer.

C.1.2.3 **Decision support in assessing treatment response**: We will develop tools to assist decision making based on image biomarker assessments in two major clinical trial tasks: (1) determine treatment response in patients, and (2)

evaluate treatment effectiveness by determining the cohort-based treatment response.

C.1.2.4 **Tools to assess the benefits of new imaging biomarkers**: We will develop the biomarker comparison module, an application module in ePAD viewer that compares the cohort treatment response results obtained when using novel vs. conventional (e.g., linear dimension) imaging biomarkers. This module will summarize the treatment response in patient cohorts based on the new imaging biomarker (using linear measurement for comparison) using several methods: waterfall plots to show the best overall response rates in the cohort, progression-free survival (PFS), MRR, and MTP.

C.2.2.2 **Develop algorithms for automated segmentation of target lesions**: (Exemplar #1a—Automated segmentation of cancer lesions; we will pursue additional exemplars as outlined in our timeline in future years). We will also engage with the QIN community to test these modules.

In addition:

1. We will make a working prototype available to interested QIN participants.
2. We will make regular public releases of ePAD and will submit an educational exhibit to RSNA 2017 that will allow us to begin to train the broader community regarding the use of the QIFP.

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

§ Published papers:

1. Abajian AC, Levy M, Rubin DL. Informatics in Radiology Improving Clinical Work Flow through an AIM Database: A Sample Web-based Lesion Tracking Application. Radiographics 2012; 32:1543-1552. PMID:WOS:000308632900027.
2. Hoogi A, Subramaniam A, Veerapaneni R, Rubin D. Adaptive Estimation of Active Contour Parameters Using Convolutional Neural Networks and Texture Analysis. IEEE Trans Med Imaging 2016. PMID:27849525.
3. Hwang KH, Lee H, Koh G, Willrett D, Rubin DL. Building and Querying RDF/OWL Database of Semantically Annotated Nuclear Medicine Images. J Digit Imaging 2016. PMID:27785632.
4. Barker J, Hoogi A, Depeursinge A, Rubin DL. Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles. Med Image Anal 2016; 30:60-71. PMID:26854941.
5. Yuan Y, Hoogi A, Beaulieu CF, Meng MQ, Rubin DL. Weighted locality-constrained linear coding for lesion classification in CT images. Conf Proc IEEE Eng Med Biol Soc 2015; 2015:6362-6365. PMID:26737748.
6. Diamant I, Hoogi A, Beaulieu C, Safdari M, Klang E, Amitai M, Greenspan H, Rubin D. Improved patch based automated liver lesion classification by separate analysis of

the interior and boundary regions. IEEE J Biomed Health Inform 2015. PMID:26372661. PMCID:PMC5164871

§ Accepted Manuscripts:

1. J. Wu, M. F. Gensheimer, X. Dong, D. L. Rubin, S. Napel, M. Diehn, B. W. Loo, R. Li, "Robust Intra-tumor Partitioning to Identify High-risk Subregions in Lung Cancer: a Pilot Study," International Journal of Radiation Oncology, Biology, Physics 95(5):1504-12, 2016. PMID: 27212196. PMCID: PMC4969127.
2. K. Lekadir, A. Galimzianova, À. Betriu, L. Igual, D. L. Rubin, E. Fernández, P. Radeva, and S. Napel, "A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound*," in press, IEEE Journal on Biomedical and Health Informatics, October 2016 NIHMSID 832055.*
3. Hoogi A, Lambert JW, Zheng Y, Comaniciu D, and Rubin DL, A Fully Automated Pipeline for Detection and Segmentation of Liver Lesions and Pathological Lymph Nodes, *in press, Thirtieth Annual Conference on Neural Information Processing Systems (NIPS 2016).*
4. Fedorov A, Rubin DL, Clunie D, Pieper S, Kikinis R, Standardized communications of quantitative image analysis results using DICOM: Establishing interoperability through outreach and community engagement, *in press, AMIA Joint Summits on Translational Science,* 2017.

§ Submitted Manuscripts:

1. M. Zhou, A. N. C. Leung, S. Echegaray, J. B. Shrager, K. C. Jensen, G. G. Berry MD, S. K. Plevritis, D. L. Rubin, S. Napel, O. Gevaert, "Radiogenomics mapping of non-small cell lung cancer identifies prognostic relationships between semantic image features and metagenes captured using RNA sequencing, *revised and re-submitted to Radiology August 2016.*
2. R. Minamimoto, M. Jamali, O. Gevaert, S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, S. K. Plevritis, D. L. Rubin, A. N. C. Leung, S. Napel, A. Quon, "Prediction of EGFR and KRAS Mutation in non-small cell lung cancer using advanced quantitative 18F FDG-PET/CT metrics," *submitted to Oncotarget, September 2016.*
3. O. Gevaert, S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, K. C. Jensen, G. J. Berry, S. K. Plevritis, D. L. Rubin, S. Napel, A. N. Leung, "Predictive radiogenomics modeling of EGFR mutation status in lung cancer," *revised and re-submitted to Nature Scientific Reports, October 2016.*
4. A. Hoogi, C. F. Beaulieu, G. M. Cunha, E. Heba, C. B. Sirlin, S. Napel and D. L. Rubin, "Adaptive Local Window for Level Set Segmentation of CT and MRI Liver Lesions," *revised and resubmitted, Medical Image Analysis, November 2016.*
5. S. Echegaray, S. Bakr, D. L. Rubin, S. Napel, "Quantitative Image Feature Engine (QIFE): An open-source, modular engine for 3D quantitative feature extraction from volumetric medical images," *submitted to Medical Physics, Nov. 2016.*

6. Farahani K, Kalpathy-Cramer J, Chenevert TL, Rubin DL, Sunderland JJ, Nordstrom RJ, Buatti J, Hylton N, Computational Challenges and Collaborative Projects in the NCI Quantitative Imaging Network, *Tomography* 2 (4): 242-249, 2016.
7. Echegaray S, Nair V, Kadoch M, Leung A, Rubin DL, Gevaert O, Napel S, A rapid segmentation-insensitive "digital biopsy" method for radiomic feature extraction; method and pilot study using CT images of non-small cell lung cancer, *Tomography*, 2(4): 283-294, 2016.
8. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin DL, McNitt-Gray M, Lo P, Sieren JC, Uthoff J, Dilger SKN, Driscoll B, Yeung I, Hadjiiski L, Cha K, Balagurunathan Y, Gillies R, and Goldgof D, Radiomics of Lung Nodules: A Multi- Institutional Study of Robustness and Agreement of Quantitative Imaging Features, *Tomography* 2(4): 430-437, 2016.

§ Presentations:

1. Rubin DL, Altindag C, Alkim E, New developments in the ePAD platform to support quantitative imaging assessment in the research workflow, Scientific Exhibit in the Quantitative Imaging Reading Room of the Future (QIRR), One hundred and second annual scientific meeting of the RSNA, Chicago, IL, 2016
2. Rubin DL, Altindag C, Alkim E, New developments in the ePAD platform to support quantitative imaging assessment in the research workflow, Scientific Exhibit in the Quantitative Imaging Reading Room of the Future (QIRR), One hundred and second annual scientific meeting of the RSNA, Chicago, IL, 2016.
3. Fedorov A; Rubin DL, Clunie DA, Flade D, Nolden M, Hafey C, Rannou N, Baumhauer M, Meine H, Mongkolwat P, Kalpathy-Cramer J, Kirby J, Onken M, Riesmeier J, Pienaar R, Marcus DS, Harris GJ, Pieper SD, Kikinis R, Baumann M, Interoperable Communication of Quantitative Image Analysis Results Using the DICOM Standard, Scientific Exhibit in the Quantitative Imaging Reading Room of the Future (QIRR), One hundred and second annual scientific meeting of the RSNA, Chicago, IL, 2016.

U01CA142565: Quantitative MRI for Predicting Response of Breast Cancer to Neoadjuvant Therapy

Vanderbilt University and University of Texas at Austin

Thomas Yankeelov, Ph.D.

INTRODUCTION

The long-term vision of this program is to significantly improve patient care by optimizing, validating, and then extending quantitative MRI methods for the early prediction of breast cancer response to neoadjuvant therapy (NAT). During the first period of support we incorporated quantitative dynamic contrast enhanced MRI (DCE-MRI) and diffusion weighted MRI (DW-MRI) into a predictive statistical model to achieve an area under of the receiver operator characteristic curve of 0.87 for predicting the eventual response of breast tumors after the first cycle of neoadjuvant therapy (NAT). We now seek to extend these results in multi-site clinical trials.

The ability to predict—early in the course of therapy—patients who will eventually achieve a pathological complete response remains a highly relevant clinical objective. Accurate and early response assessment would provide the opportunity to replace an ineffective treatment with an alternative regimen, and in so doing potentially avoid or curtail debilitating side effects or toxicities. With the numerous options for NAT that have become available, development of a method to predict response early in the course of therapy is especially needed.

We have developed several experimental and computational tools for improving DCE-MRI and DW-MRI of the breast, and we have successfully applied these tools in clinical trials at Vanderbilt University. We are now applying these techniques in multi-site clinical trials at Vanderbilt University and The University of Chicago. Furthermore, we have an exploratory component in the community setting in place at The University of Texas at Austin.

The knowledge acquired through this study will provide direction on developing personalized treatment strategies for breast cancer patients undergoing NAT and may motivate a fundamental shift in existing paradigms of therapy monitoring and selection in breast cancer. Furthermore, MRI assessment of early response could be more broadly applicable to other solid malignancies where NAT is appropriate (e.g., pancreas, osteosarcoma, rectal, ovarian); thus, the results of this study could potentially have a significant impact beyond breast cancer.

DISCUSSION OF PROGRESS

§ Our primary work

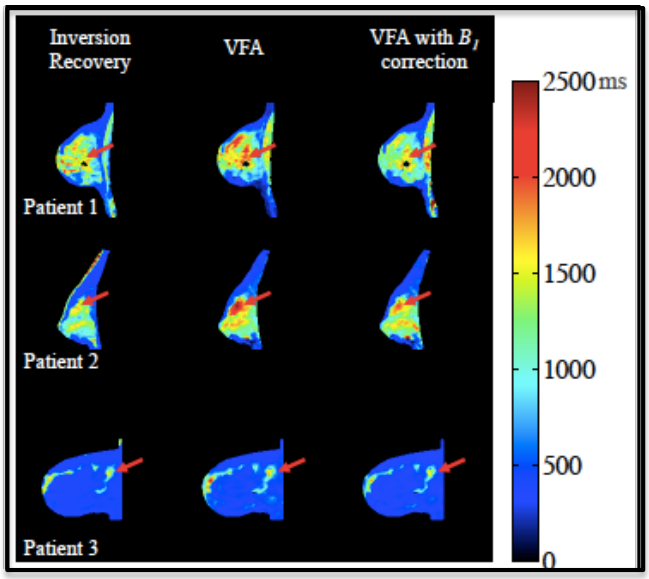


Figure 1: T_1 parametric maps are shown for inversion recovery (left column), uncorrected variable flip angle (VFA, center) and B_1-correctedVFA (right) data collected from each of three patients (show in rows).

We have recently completed a set of repeatability/reproducibility studies in phantoms, healthy volunteers, and a limited number of patients; these data consist of quantitative, Bloch-Siegert corrected T_1-mapping, apparent diffusion coefficient maps, and quantitative magnetization transfer (qMT) maps. Figure 1 displays inversion recovery, variable flip angle (VFA) and B_1-corrected VFA T_1 maps on three patients. Compared to the VFA data, the B_1-corrected VFA T_1 values of the fibroglandular tissue (FGT), adipose tissue (AT), and tumor in all three patients are more similar to the IR T_1 values. After B_1 correction, %error significantly ($p < 0.001$) decreased from 17% to 8.6% and the concordance correlation coefficient increased from 0.55 to 0.83 in the FGT. The 95% CI of the mean difference decreased from ±94 ms to ±38 ms after B_1 correction. Similar accuracy and reproducibility results were observed in the AT and tumor tissues. These data show that Bloch-Siegert B_1 mapping significantly improves accuracy and precision of VFA-derived T_1 measurements.

Quantitative MT imaging potentially provides more specific information on tissue composition, including the ratio of macromolecular protons to the protons in the free water pool, or pool size ratio (PSR). We have assessed the reproducibility of PSR measurements of FGT in healthy controls. Figure 2 displays PSR maps from the central slice of three subjects

for scan 1 (top row) and scan 2 (bottom row). The mean difference for all subjects (-0.06) was not significantly different from zero, and the 95% confidence interval limits were ±0.64 ($\alpha = 0.05$) and the repeatability measure (2.77 × wSD) was 1.87. The B_1-corrected T_1, ADC, PSR measurements are implemented in identical protocols on nearly identical 3T Philips scanners at Chicago and Vanderbilt. The clinical trials that were selected to deploy these measurements have also been successfully opened at both data acquisition sites and we have begun to acquire patient data. This essentially completes the majority of Aim 1 and has us well-positioned to address Aims 2 and 3.

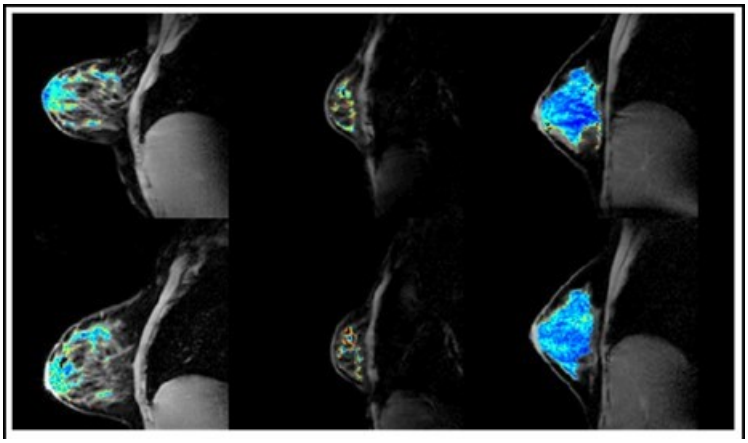


Figure 2: PSR maps from the central slices of three subjects for scan 1 (top row) and scan 2 (bottom row). Note the range of adipose tissue to fibraglandular tissue ratios observed in these healthy volunteers.

In addition to collecting the above data in our ongoing clinical trials, we are exploring the utility of ultrafast imaging of the breast during the first minute after the administration of contrast media. Using standard Fourier techniques, we achieved temporal resolutions of 2 to 9 seconds by reducing spatial resolution, and increasing parallel imaging and partial Fourier factors. While several techniques have been developed to increase temporal resolution without sacrificing spatial resolution (e.g., view-sharing, sliding window), they rely on under sampling the edges of *k*-space and mixing data acquired at different times. In a situation when the signal is rapidly changing, and much of the enhancement is occurring in small areas, undersampling and view sharing at the periphery of *k*-space could lead to artifacts and errors in parameters descriptive of lesion kinetics. Our initial experience with ultrafast breast imaging showed that malignant lesions, on average, enhanced earlier and faster than benign lesions and normal uninvolved parenchyma. Because of this, lesion conspicuity is increased in ultrafast images. Ultrafast imaging also allows for more accurate estimations of parameters descriptive of contrast media uptake, and for the measurement of these parameters relative to the time at which the contrast bolus arrived in the breast (see Figures 3 and 4), rather than the time of injection, removing the dependence on global variables such as cardiac output. Imaging the early phases of contrast uptake allows for the use of simplified pharmacokinetic models to estimate parameters such as the volume transfer constant, K^{trans}.

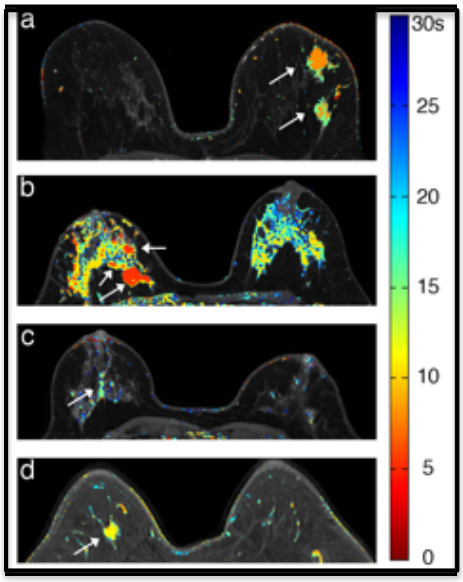


Figure 3: Time of arrival (TOA) is defined as the time the lesion began to significantly enhance relative to when the contrast agent firs arrived in the arteries of the breast. The TOA is show as a color overlay in these images. Lesions are marked with arrows: a) and b) invasive ductal carcinoma, c) complex sclerosing lesion, d) fibroadenoma. The results show her are typical of our initial results, with malignant lesions having a shorter average time of arrival than benign lesions (6.9 s +/- 4.6 s and 15.5 s+/- 13.6 s, respectively).

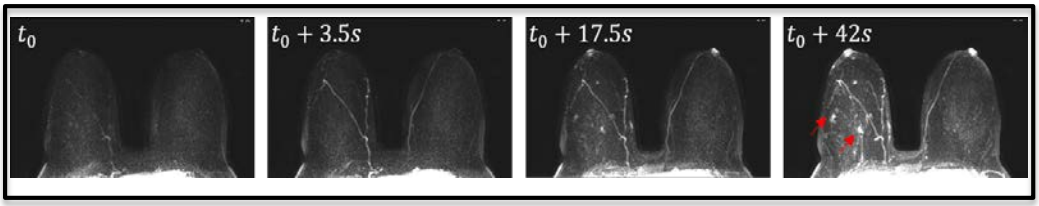


Figure 4: Maximum intensity projections (MIPs) for post-contrast minus pre-contrast difference images acquired at a temporal resolution of 3.5s to is defined as the time that arterial enhancement is first measured in the breast. Two likely benign lesions are marked with arrows in the latest image shown.

We have also used the high-temporal resolution data to automatically detect, segment, and track tumor associated vasculature within the breast. Representative results are shown in Figure 5. In Figure 5a, two tracked paths based on data obtained with 2 sec resolution data are indicated and overlain on the MIP of a post-contrast image. Observe how the vessels appear to originate from the most lateral lesion and extend to the internal thoracic veins. Figure 5b shows the normalized signal intensity time series associated with four different locations within the most anterior vessel (i.e., P1, P2, P3, P4 as labeled in Figure 5a). The order of

enhancement is what allows for the color coding of the vessel displayed in Figure 5a. The high temporal resolution of the acquisition thus allows for not only vessel tracking, but also determining the direction of flow within the breast tissue. Figure 5c prevents a 3D rendering of the tracked vessels within the breast volume to better visualize their trajectory. An abstract summarizing this work is to be presented at the upcoming ISMRM meeting. Longer term, we aim to include these measurements in our ongoing trials.

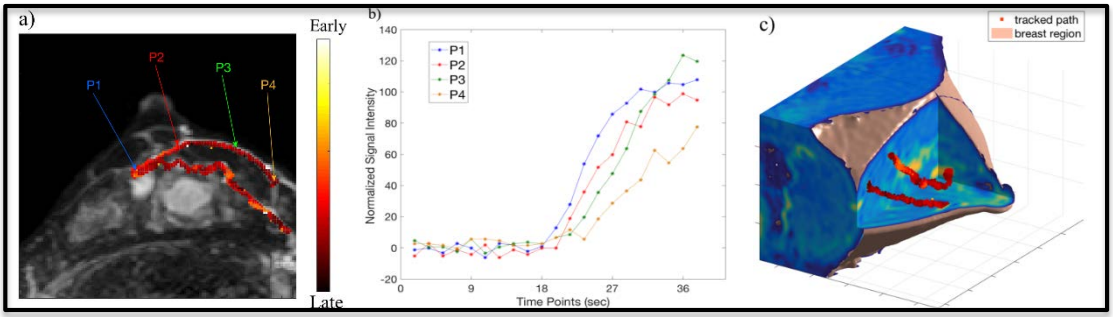


Figure 5: Panel a displays two tracked vessels overlain on a maximum intensity projection. The (normalized) signal intensity time courses associated with locations P1-P4 are indicated on Panel b for one of the tracked vessels. Note how the high temporal resolution acquisition is critical to determining the direction of flow within the tracked vessel. Panel c shows a 3D rendering of the same two vessels within the breast volume.

We are thrilled to report that we have established a public-private collaboration between The University of Texas at Austin, Austin Radiological Associates, Texas Oncology, and Seton Healthcare. This is truly noteworthy as our team only arrived on campus in February of 2016 and we were able to establish this formal collaboration (complete with contracts, etc.) and open a clinical trial by September. Our first patient was enrolled in this study in October of 2016. More specifically, we are testing the hypothesis that quantitative dynamic contrast enhanced MRI (DCE-MRI), diffusion-weighted MRI (DW-MRI), and magnetization transfer MRI (MT-MRI) can predict, early in the course of NAT, the eventual response (i.e., pathological complete response vs. residual disease) of the individual patient, and that this can be achieved in the community setting. A real, practical, advance of these studies is that we have implemented these advanced MRI measures in the community setting; that is, in the locations around Austin where patients go for their standard-of-care imaging session. The patients do not have to come to an academic setting, they can simply go to where they usually receive their care. If our methods prove successful, then the barrier between the bench and the bedside is dramatically lowered with this approach. Figure 6 presents an illustrative data set from a patient enrolled in our study. This constitutes a “new aim” in the Gantt plot below.

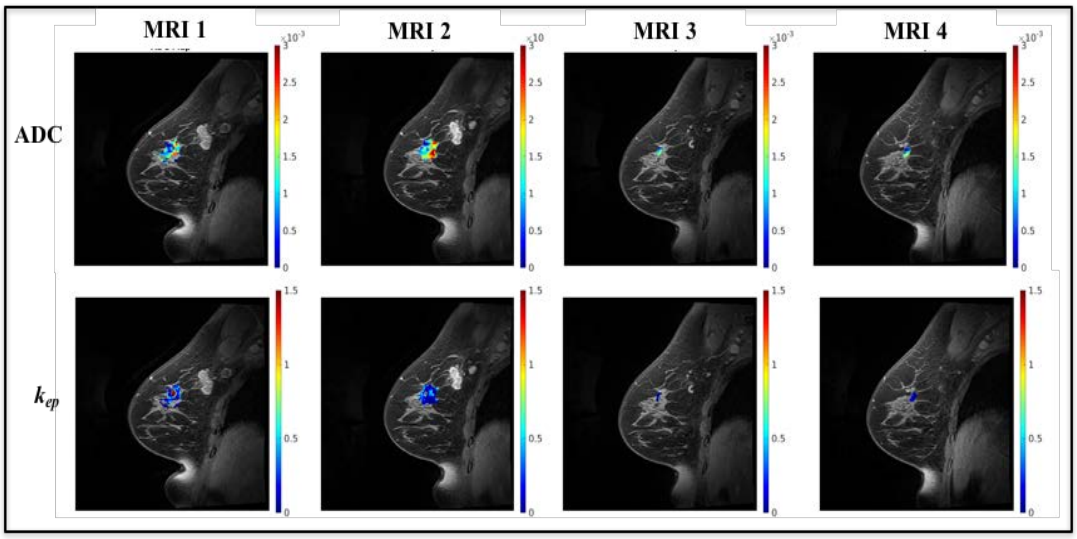


Figure 6: The figure displays two of the quantitative parametric maps we acquire in patients undergoing neoadjuvant chemotherapy; ADC = apparent diffusion coefficient which is a surrogate for cellularity, and kep = efflux transfer constant which is a surrogate for vascularity. We perform various computations on these parametric maps to predict who will achieve pCR and who will have residual disease at the time of surgery. Note: although only a single slice is show here, we collect full 3D data sets at each time point.

		Year 1				Year 2				Year 3				Year 4				year 5			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Aim 1	Planned																				
	Actual																				
Aim 2	Planned																				
	Actual																				
Aim 3	Planned																				
	Actual																				
Aim 4	Planned	*New aim, not initially planned.*																			
	Actual																				

In addition to the above progress, we have also participated in a large number of intra-QIN projects which we now briefly summarize.

§ IAPM MRI Subgroup: ADC DICOM Challenge

Drs. Dariya Malyarenko and Tom Chenevert from the University of Michigan have distributed DICOM files of diffusion weighted MRI data of their ice water phantom from three obtained on GE, Siemens, and Philips scanners. We were asked to generate ADC maps and save the results in DICOM format. We have processed the data and sent back to Dr. Malyarekno for processing.

§ Data Acquisition Working Group: T_1-mapping Challenge

Our team has scanned the phantom that was sent to us as part of the challenge and submitted the data to Dr. Octavia Bane. We used one of head and neck protocols with a 32-channel head coil to perform the generic T_1 mapping sequence as all other participants. Additionally, we performed our multi-flip angle T_1 mapping sequence as well. We are co-authors on the abstract accepted to ISMRM.

§ Data Acquisition Working Group: Diffusion, Phase II

We were one of the sites that scanned the fBIRN phantom for the second phase of the diffusion gradient nonlinearity challenge led by Drs. Malyarenko and Chenevert from the University of Michigan group. We scanned the phantom, deliveted the data to Dr. Malyarkenko and co-authored manuscript submitted to and accepted by *Tomography*.

§ DCE-MRI Data Challenge: Effects of AIF Quantification in Soft Tissue Sarcoma

This effort was led by Dr. Wei Huang at the Oregon Health Sciences University. to investigate the effects of AIF variations on DCE-MRI prediction of soft tissue sarcoma response to preoperative therapy. As did all the participating centers, we determined individual AIFs for each patient in the cohort from the femoral artery using the DCE-MRI data with our site-specific method and submitted them to the OHSU for pharmacokinetic modeling of the tumor voxel DCE-MRI time-course data using the standard Tofts model. This effort will be presented as an abstract at the upcoming ISMRM meeting.

§ DCE-MRI Data Challenge: Effects of AIF Quantification on Shutter-Speed Analysis

This effort was led by Dr. Wei Huang at the Oregon Health Sciences University. The goal was to determine the effects of AIF characterization on the ability to perform a robust Shutter-Speed analysis at multiple sites. The study design was similar to that described in the previous paragraph and this effort will also be presented as an abstract at the upcoming ISMRM meeting.

§ Participation in National Clinical Trials

We have participated in two previous ECOG-ACRIN trials investigating advanced quantitative MRI techniques in breast cancer, and are currently ramping up to participate in another consortium trial that is comparing an abbreviated breast MRI exam to digital breast

tomosynthesis in breast cancer screening of women with dense breasts. The primary objective of this study is to compare the detection rates of invasive cancers between the imaging technologies. This additional scan sequence was developed by Drs. Karczmar and Pineda at The University of Chicago, and both are current investigators in our QIN multisite breast MRI program.

PLANS FOR NEXT YEAR

The main goal during the next reporting period is to accrue a significant number of patients at the two institutions. Now that the protocols are up and running at both institutions, we are well-positioned to attack these Aims. In addition to accruing patients in support of the ultimate goals of the application, we will also pursue a number of technical—but practical—issues. In particular, as described in the initial application, we will assess the repeatability and reproducibility of data acquired in the same healthy volunteers at both VU and UC. We will also perform a cross-validation of independent site analyses. The goals here are to determine if our acquisition and analysis toolbox can provide statistically identical answers when used at different institutions.

We will continue to participate in inter-QIN collaborative projects and assist in the writing and publication of these efforts.

At The University of Texas at Austin we will continue to expand our footprint into the community setting by completing repeatability and reproducibility studies in health subjects at multiple private practice settings.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

[1] F. Pineda, M. Medved, S. Wang, X. Fan, D. Schacht, C. Sennett, A. Oto, G. Newstead, H. Abe, and G. Karczmar. Ultrafast bilateral DCE-MRI of the breast with conventional Fourier sampling: preliminary evaluation of semi-quantitative analysis. Acad Radiol 2016; 23:1137-1144.

[2] Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, Muzi M, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Abramson RG, Besa C, Kalpathy-Cramer J, Laymon CM, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy F, Li X. The Impact of Arterial Input Function Determination Variation on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. Tomography, 2016;2:56-66.

[3] Chan E, Arlinghaus LR, Cardin DB, Goff L, Berlin JD, Parikh A, Abramson RG, Yankeelov TE, Hiebert S, Merchant N, Bhaskara S, Chakravarthy AB. Phase I trial of vorinostat added to chemoradiation with capecitabine in pancreatic cancer. Radiotherapy and Oncology 2016;119:312-8.

[4] Whisenant JG, Dortch RD, Grissom W, Kang H, Arlinghaus LR, Yankeelov TE. Bloch-Siegert B1-mapping Improves Accuracy and Precision of Longitudinal Relaxation Measurements in the Breast at 3T. Tomography, 2016;2:250-9.

[5] O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brown G, Brady M, Buckley DL, Chenevert TL, Clarke LP, Collette S, Cook GJ, deSouza NM, Dickson JC, Dive C, Evelhoch JL, Faivre-Finn C, Gallagher FA, Gilbert FJ, Gillies RJ, Goh V, Griffiths JR, Groves AM, Halligan S, Harris AL, Hawkes DJ, Hoekstra OS, Huang EP, Hutton BF, Jackson EF, Jayson GC, Jones A, Koh D-M, Lacombe D, Lambin P, Lassau N, Leach MO, Lee T-Y, Leen EL, Lewis JS, Liu Y, Lythgoe MF, Manoharan P, Maxwell RJ, Miles KA, Morgan B, Morris S, Ng T, Padhani AR, Parker GJM, Partridge M, Pathak AP, Peet AC, Punwani S, Reynolds AR, Robinson SP, Shankar LK, Sharma RA, Soloviev D, Stroobants S, Sullivan DC, Taylor SA, Tofts PS, Tozer GM, van Herk M, Walker-Samuel S, Wason J, Williams KJ, Workman P, Yankeelov TE, Brindle KM, McShane LM, Jackson A, Waterton JC. Imaging Biomarker Roadmap for Cancer Studies. Nature Reviews Clinical Oncology, 2017;14:169-86.

[6] Arlinghaus LA, Dortch RD, Whisenant JG, Kang H, AbramsonRG, Yankeelov TE. Quantitative Magnetization Transfer Imaging of the Breast at 3.0 T: Reproducibility in Healthy Volunteers. Tomography, 2016;2:260-6.

[7] Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, Muzi M, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Abramson RG, Besa C, Kalpathy-Cramer J, Laymon CM, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy F, Li X. The Impact of Arterial Input Function Determination Variation on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. Tomography, 2016;2:56-66.

[8] Chan E, Arlinghaus LR, Cardin DB, Goff L, Berlin JD, Parikh A, Abramson RG, Yankeelov TE, Hiebert S, Merchant N, Bhaskara S, Chakravarthy AB. Phase I trial of vorinostat added to chemoradiation with capecitabine in pancreatic cancer. Radiotherapy and Oncology 2016;119:312-8.

[9] Abramson RG, Arlinghaus L, Dula A, Quarles CC, Stokes A, Weis JA, Whisenant J, Williams J, Chekmenev EY, Zhukov I, Yankeelov TE. MRI Biomarkers in Oncology Clinical Trials. MRI Clinics of North America, 2016;24:11-29.

[10] Yankeelov TE, Mankoff DA, Schwartz L, Lieberman F, Buatti J, Mountz JM, Erickson BJ, Fennessy FM, Huang W, Kalpathy-Cramer J, Wahl RL, Linden H, Kinahan P, Zhao B, Hylton N, Gillies RJ, Clarke L, Nordstrom R, Rubin D. Quantitative Imaging in Cancer Clinical Trials. Clinical Cancer Research, 2016; 22:284-90.

[11] Weis JA, Miga MI, Yankeelov TE. Three-dimensional Image-based Mechanical Modeling for Predicting the Response of Breast Cancer to Neoadjuvant Therapy. Computer Methods in Applied Mechanics and Engineering, 2017;314:494-512.

[12] Kurland BF, Aggarwal S, Yankeelov TE, Gerstner ER, Mountz JM, Linden HM, Jones EF, Bodeker KL, Buatti JM. Accrual patterns for clinical studies involving quantitative imaging: Results of an NCI Quantitative Imaging Network (QIN) survey. Tomography, 2016;2:276-282.

[13] Williams JM, Arlinghaus LR, Rani SD, Shone MD, Abramson VG, Pendyala P, Chakravarthy AB, Gorge WJ, Knowland JG, Lattanze RK, Perrin SR, Scarantino CW, Townsend DW, Abramson RG, Yankeelov TE. Towards real-time topical detection and characterization of FDG dose infiltration prior to PET imaging. European Journal of Nuclear Medicine and Molecular Imaging, 2016:43;2374-2380.

[14] Stokes AM, Skinner JT, Yankeelov TE, Quarles CQ. Assessment of a Simplified Spin and Gradient Echo (sSAGE) Approach for Human Brain Tumor Perfusion Imaging. Magnetic Resonance Imaging, 2016;34:1248-55.

U01CA151261: Quantitative MRI of prostate cancer as a biomarker and guide for treatment

Brigham and Women's Hospital

Fiona M. Fennessy M.D. Ph.D.

INTRODUCTION

Prostate cancer (PCa) continues to be the most common malignancy and third leading cause of cancer-related mortality in American men. PCa is the most common cancer in men in North America and Europe, with over 180,000 new cases to be diagnosed in 2016, when 26,120 men will die from their disease (1). The natural history of PCa is remarkably heterogeneous and still not completely understood. Autopsy and early observational studies have shown that approximately one in three men ≥ 50 years old has histologic evidence of prostate cancer; a significant portion of these tumors are small and possibly clinically insignificant, although others are extremely aggressive and lethal(2). As the number of men with localized prostate cancer increases, the need for an accurate non-invasive imaging tool increases. MR imaging has been shown to contribute significant incremental value to both digital rectal exam and TRUS-guided biopsy in cancer detection and localization within the prostate.

REPORT OF PROGRESS

The specific aims (SA's) and summary of the important findings of our first cycle are as follows:

§ SA 1: To optimize prostate MR Image analysis tools

We optimized DCE-MRI modeling tools through investigation of T1 mapping effects (3), assessment of optimal Arterial Input Function (AIF) (4,5), and assessment of the effects of the bolus arrival time (BAT) measurement(6). We demonstrated the sensitivity of pharmacokinetic (PK) parameters to tissue T1 values, and found that using either a flip angle corrected VFA method, or a VTR FSE method with judiciously chosen TR values, increased the accuracy of T_1 values (3). We demonstrated that the method for automated determination of AIF can lead to variability in DCE-MRI parameters (5). Therefore, PK values obtained using different AIF methods may not be comparable. We found that inaccuracies in BAT, another choice in DCE MRI analysis, leads to variability among DCE-MRI PK model parameters, diminishes the quality of model fit, and produces fewer voxels suitable for modeling (6).

We subsequently shared, through the TCIA, a subset of our prostate imaging data with 9 QIN centers to evaluate variations in PK parameters in PCa due to differences in AIF determination, and showed that AIF variations significantly affect PK parameter values for prostate DCE-MRI data (7). We also contributed to a multi-center study investigating the

role of platform-specific data encoding on the accuracy of quantitative analysis (8), and to a study investigating the role of the analysis platform on the PK analysis of DCE-MRI (9).

Our first cycle also supported a repeatability study of treatment-naive men undergoing ecoil prostate mpMRI, within a 2-week period. Knowledge of measurement test-retest repeatability is critical in longitudinal studies to enable differentiation of true change vs. measurement noise. Our results indicate that PI-RADS v2 suspicion scores are highly repeatable, while tumor volume change in response to therapy may not be considered significant unless it exceeded 70% on T2-WI, or 120% on DCE-MRI and ADC imaging. A change in mean ADC may not be significant unless it exceeds 40%. We are in the final stages of preparing this manuscript and, importantly, we plan to accompany the manuscript with the dataset which will be publicly shared on TCIA.

§ SA 2: To clinically validate prostate MR quantitative analytic tools

We clinically validated our MR quantitative analytic tools from SA 1. Using whole mount pathology (WMP) validation, we both automatic and model-based AIF methods for DCE-MRI to be excellent in discriminating PCa from normal tissue, but the same method should be used throughout a biomarker study(74). We validated PK maps used for guidance in our MRgBX program(10), which demonstrated higher PCa detection rates, when compared to TRUS biopsy samples. We also compared 2 approaches for correlating pathology to mpMRI (11), and found that WMP correlation is superior to standard path report for accurate localization of all index lesions, but is not required to distinguish differences of quantitative MRI parameter values within tumor. We also determined if tumor cell density and % GP within an index tumor on WMP correlated with ADC values on mpMRI. We found tumor cell density and ADC to be significantly negatively correlated (ρ=-0.61, p=0.005) (12).

§ SA 3: To determine the clinical use of the analysis tools as a biomarker guide for targeted therapy and as a surrogate for disease recurrence in low-risk PCa patients

The last enrolled patients have completed their follow up MRIs within the last 6 months, and we are in the final stages of data analysis of this project, which explores the feasibility of mpMR as an imaging biomarker to assess response of bulky localized prostate cancer to combined ADT/EBRT.

§ SA 4: To determine the clinical use of the analysis tools in evaluating tumor response to treatment with neoadjuvant second-generation androgen receptor inhibitor enzalutamide in patients with high-risk PCa

This study is in the final stages of manuscript preparation, and demonstrated no significant difference between mpMRI-based residual tumor burden (RTB) and RTB at RP. In addition, there is a strong positive correlation between DCE-MRI and RTB (ρ=0.79, p=0.03), and a strong negative correlation (ρ=-0.91, p=0.005) between ADC and RTB, indicating a very promising role for mpMRI as a biomarker for treatment of localized PCa with neoadjuvant therapy.

In summary: 1) We optimized/validated our prostate mpMRI protocol (3,4,6,13,5)**, 2)** we established a MRI-pathology validation workflow (4,10,11,12)**: 3),** we established open source tools within 3D Slicer for annotation/quantitative analysis of mpMRI(14): **4),** we determined the clinical use of the analysis tools as a biomarker guide for targeted therapy and as a surrogate for disease recurrence in high and low-risk PCa populations, and in the final stages of manuscript preparation for these 2 trials; **5)** we contributed to multi-site QIN manuscripts(5,7-9,15–18). Finally, the data we shared in Cycle 1 was invaluable to several QIN community projects/challenges (3,7).

LIST OF REFERENCES

1. Miller KD, Siegel RL, Lin CC, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin. 2016;66(4):271–289.

2. Dall'Era MA, Albertsen PC, Bangma C, et al. Active surveillance for prostate cancer: a systematic review of the literature. Eur Urol. 2012;62(6):976–983.

3. Fennessy FM, Fedorov A, Gupta SN, Schmidt EJ, Tempany CM, Mulkern R V. Practical considerations in T1 mapping of prostate for dynamic contrast enhancement pharmacokinetic analyses. Magn Reson Imaging. 2012;30(9):1224–1233.

4. Fennessy FM, Fedorov A, Penzkofer T, et al. Quantitative pharmacokinetic analysis of prostate cancer DCE-MRI at 3T: comparison of two arterial input functions on cancer detection with digitized whole mount histopathological validation. Magn Reson Imaging. 2015;33(7):886–894.

5. Fedorov A, Fluckiger J, Ayers GD, et al. A comparison of two methods for estimating DCE-MRI parameters via individual and cohort based AIFs in prostate cancer: A step towards practical implementation. Magn Reson Imaging. Elsevier; 2014

6. Mehrtash A, Gupta SN, Shanbhag D, et al. Bolus arrival time and its effect on tissue characterization with dynamic contrast-enhanced magnetic resonance imaging. J Med Imaging Radiat Oncol. 2016;3(1):014503

7. Huang W, Chen Y, Fedorov A, et al. The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. Tomography. ncbi.nlm.nih.gov; 2016;2(1):56–66

8. Chenevert TL, Malyarenko DI, Newitt D, et al. Errors in Quantitative Image Analysis due to Platform-Dependent Image Scaling. Transl Oncol. 2014;7(1):65–71

9. Huang W, Li X, Chen Y, et al. Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. Transl Oncol. 2014;7(1):153–166.

10. Penzkofer T, Tuncali K, Fedorov A, et al. Transperineal in-bore 3-T MR imaging-guided prostate biopsy: a prospective clinical observational study. Radiology. 2015;274(1):170–180.

11. Fedorov A, Penzkofer T, Hirsch MS, et al. The role of pathology correlation approach in prostate cancer index lesion detection and quantitative analysis with multiparametric MRI. Acad Radiol. 2015;22(5):548–555.

12. Glazer DI, Hassanzadeh E, Fedorov A, et al. Diffusion-weighted endorectal MR imaging at 3T for prostate cancer: correlation with tumor cell density and percentage Gleason pattern on whole mount pathology. Abdom Radiol (NY). 2016;http://dx.doi.org/10.1007/s00261-016-0942-1

13. Hegde J V, Mulkern R V, Panych LP, et al. Multiparametric MRI of prostate cancer: An update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. J Magn Reson Imaging. 2013;37(5):1035–1054.

14. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. Elsevier Inc.; 2012;30(9):1323–1341.

15. Levy MA, Freymann JB, Kirby JS, et al. Informatics methods to enable sharing of quantitative imaging research data. Magn Reson Imaging. 2012;30(9):1249–1256.

16. Kurland BF, Gerstner ER, Mountz JM, et al. Promise and pitfalls of quantitative imaging in oncology clinical trials. Magn Reson Imaging. Elsevier Inc.; 2012;30(9):1301–1312.

17. Yankeelov TE, Mankoff DA, Schwartz LH, et al. Quantitative Imaging in Cancer Clinical Trials. Clin Cancer Res. 2016;22(2):284–290.

18. Mountz JM, Yankeelov TE, Rubin DL, et al. Letter to cancer center directors: Progress in quantitative imaging as a means to predict and/or measure tumor response in cancer therapy trials. J Clin Oncol. 2014;32(19):2115–2116.

U01 CA15460: Quantitative MRI of Glioblastoma Response

Massachusetts General Hospital

Jayashree Kalpathy-Cramer, PhD, MS

INTRODUCTION

Patients with glioblastoma (GBM), a deadly form of brain cancer, have extremely poor prognosis and few treatment options [1-3]. Assessment of response to therapies is critically needed to aid in clinical decision-making [2,4,5]. The QIN team from MGH continues to make good progress in developing and validating novel biomarkers for measuring the response to therapy in GBM. Our techniques for the analysis of diffusion weighted imaging, dynamic contrast enhanced MRI (DCE-MRI) and dynamic susceptibility contrast MRI (DSC-MRI) are being applied in a number of prospective and retrospective clinical trials [6]. The software tools used for these analyses are publicly available as open-source packages, both standalone as modules for 3D Slicer [7]. These techniques have been developed using phantom studies, simulations, retrospective analysis, and prospective analysis in patients undergoing treatment with anti-angiogenic therapies. Having worked with other teams in the QIN as well as other groups in the community in establishing common, standardized approaches to image analysis and acquisition for patients with GBM [8,9,10,11], we are now implementing these protocols in clinical trials.

Advanced MRI methods may improve our ability to provide an accurate prognosis and potentially guide treatment choices for glioblastoma patients [2,4,5]. We continue to be fortunate to work closely with our clinical colleagues who provide us with great access to high quality imaging data acquired both as part of clinical trials and as part of routine clinical care. Our "double baseline" studies have established the variability in the DCE-MRI, DSC-MRI and diffusion MRI based parameters (Ktrans, rCBV, rCBF, MTT, ADC, FA) in patients scanned 2-5 days apart. These studies also establish best practices for image analysis to achieve maximal robustness [8,11]. Our novel image acquisition methods include a double-echo DSC and DCE sequences and a multi-shell, multi-directional diffusion sequence that help us better elucidate the tumor microenvironment. In addition to progress in the image acquisition arena, we have also made significant strides in image analysis and informatics. We have also developed a number of open-source image analysis tools for tumor segmentation and registration, multimodal atlases, personalized tumor growth models and hardware and software approaches to improve image resolution. We continue to develop open-source software tools for the analysis of DCE data, especially T1 mapping, and the "double-echo" sequence that has been shown to be able to quantify R2* effects. Some of these have already been disseminated as Slicer modules, in conjunction with the QIICR ITCR project.

The team from MGH continues to actively participate in the Quantitative Imaging Network (QIN) and has made significant contributions as part of the various working group. We have been very actively engaged in the "challenges" being conducted as part of the QIN, both as organizers and participants. We have also developed a number of close and fruitful

collaborations with other members of the QIN, resulting in joint publications, planned multi-institutional studies and successful grant applications.

In addition to our primary work in the area of GBM, the MGH team has been participating actively in the PET/CT working groups and has developed radiomics pipelines and statistical analysis tools. These are also publicly available [https://github.com/QTIM-Lab].

DISCUSSION OF PROGRESS

§ Quantitative imaging pipeline for GBM

We followed up on our previous work creating a within-patient and longitudinal registration pipeline. In particular, we developed a pipeline to register and upsample T1, T2, DTI, DSC, MPRAGE, FLAIR, and DCE maps into 1x1x1mm isotropic space. This involved ordering images in a chain of sequential registrations, such that only visually-similar images are registered in any given step. This allowed us to coregister and upsample every modality in every patient visit without error. Our registration scheme is provided below. This pipeline is available as a Docker container and is being shared with our collaborators at Tata Memorial Hospital in India.

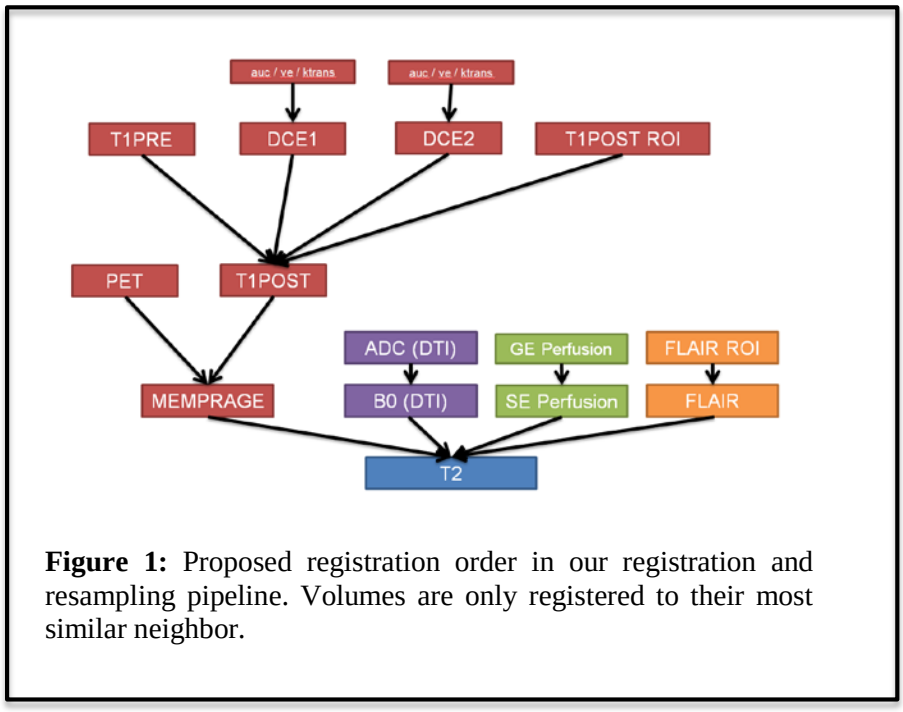


Figure 1: Proposed registration order in our registration and resampling pipeline. Volumes are only registered to their most similar neighbor.

§ Tumor growth modeling and personalization of radiotherapy:

We continue to develop our tumor growth models and their use in personalization of radiotherapy. Using a radiation therapy plan based on the expected growth patterns of tumors can results in improved tumor control and better sparing of normal tissues compared

to a uniform expansion of the visible tumor. As shown in two recent publications [13,14], this is a promising approach and had the potential for use in adaptive radiotherapy.

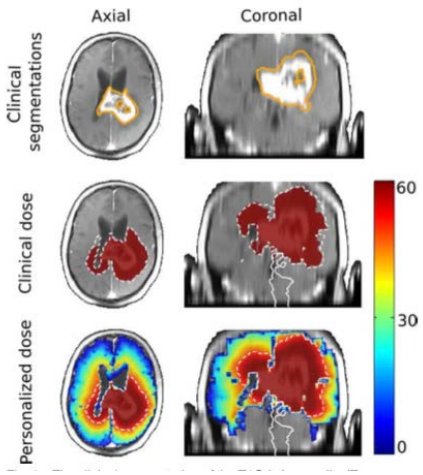


Figure 2: Clinical segmentation is used to define the clinical target volume (CTV, white dashed line) as a 2 cm expansion of the segmentation. In clinical settings, 60 Gy is prescribed to the CTV. We propose to personalize the prescription dose (Bottom) to account for tumor infiltration and segmentation uncertainty

§ Tumor segmentation and Normative atlases

We have created tools for automatic tumor segmentation as well as normative atlases that are useful in the semi-automated segmentation process. In diagnosing brain tumors, gadolinium contrast agent is usually injected to patients to highlight enhancing tumor in T1-weighted magnetic resonance imaging (MRI). Gadolinium, or other contrast agents, help localize the tumors and are used in surgery planning, treatment design and prognosis prediction. Automatic segmentation tools, such as those based on "subtraction images" (comparing post contrast with pre-contrast) can tools incorrectly identify normal enhancing areas (such as large vessels) are being tumor tissue. In order to quantify"normal" enhancement to correctly distinguish tumors, we took the approach of constructing normative atlases, from patient images as normal patients are typically not given contrast. These normative enhancement maps are publicly available [https://www.nitrc.org/projects/stamp_atlases]and have been used by a number of groups (including ours) in conjunction with automatic segmentation algorithms to remove false positive and clean up segmentations.

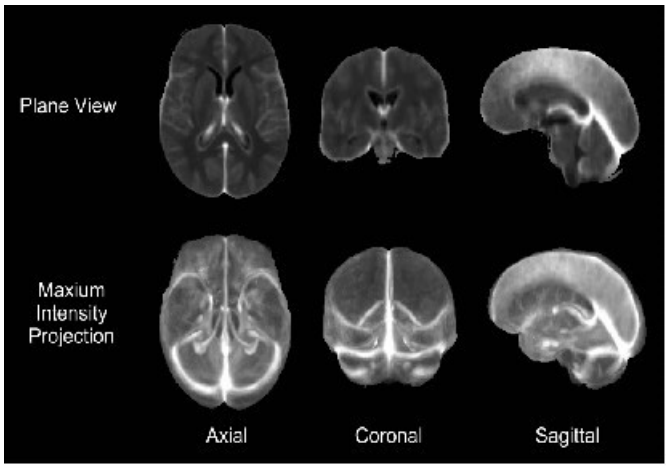


Figure 3: The STAMP T1d atlas (top row) and its maximum intensity projection (MIP) rendering quantifying locations and extent of enhancement of a (virtual) average normal.

§ QIN Challenges and collaborative projects

The MGH group participated in several QIN challenges and project last year, including DSC challenge, ADC challenge, T1 mapping challenge and the breast response challenge.

QIN DSC Challenge: Though DSC-MRI perfusion is of well-known benefit for the evaluation of brain tumors, clinical translation has been hampered by a lack of confidence in the consistency of the derived relative cerebral blood volume (RCBV) and cerebral blood flow (CBF) values across sites and platforms. This multi-site and multi-platform study, for which the same patient data set was analyzed, demonstrated substantial consistency in RCBF across software sites and platforms and the ability of each to distinguish low-grade from high-grade tumor. In addition, a single RCBV threshold was identified for which all platforms maintained good accuracy. This study was summarized in an abstract and accepted as an oral presentation at ISMRM. A full manuscript is under preparation. The MGH software package, used to produce the results of the challenge is available as an open-source module for 3D Slicer [7]. The results of our module showed excellent concordance with other commercial packages and performed similarly in distinguishing high grade tumors from low grades.

QIN ADC Challenge: Reproducibility of diffusion metrics is essential given the increasing role quantitative diffusion weighted imaging plays in diagnosis and treatment monitoring. Here we examined the variability in apparent diffusion coefficient (ADC) measures resulting from different post-processing software implementations utilized by researchers across the NCI Quantitative Imaging Network. Agreement between the majority of implementations was good; typical biases for in vivo ADC measures of 2-3%, and lower

biases in phantom scans. Higher deviations (above 5%) detected among individual implementations and scanner-generated parametric maps highlighted inadequacies in meta-data and post-processing parameters that need to be addressed in multi-site study settings. This study was summarized in an abstract and submitted to the ISMRM. A full manuscript is under preparation. Again, our open source implementation provided excellent results and showed good concordance with other software packages.

QIN T1 Mapping Challenge: This multicenter study examined variability in T1 quantification by testing common inversion-recovery spin echo and variable flip angle (VFA) protocols, as well as T1 mapping methods used by participating sites, using a phantom with known T1 values. We found field strength dependence of the accuracy, and platform dependence of the repeatability of T1 measurements with the common VFA protocol. Accuracy for site-specific protocols was influenced by site, while repeatability, by type of protocol. Our findings suggest modified IR methods and VFA protocols with multiple flip angles and B1 correction as good methods for repeatable T1 measurement.

Lung volume interval challenge: As members of in the PET/CT working group, MGH was an active participant for the lung volume interval challenge and developed a data visualization platform to analyze the effectiveness of five institutions' automatic lung nodule segmentation algorithms.

Our visualization platform [http://cbibop.cloudapp.net:3838/Interval_Lung_Challenge_ShinyApp/] has been used effectively during the group meetings for the statistical data analysis. Examples of the analyses techniques supported by our system are found below. As seen in Figure 4, there were nodules with substantial disagreement between segmentation algorithms while in other cases, there was good agreement. As seen in Figure 5, there was a range of Dice agreements reflecting the range of agreement. The AUCs using percent change as a measure of malignancies were somewhat consistent between the groups. Our visualization also demonstrated the difficulty of recognizing segmentation failures without a human observer. A manuscript is forthcoming on the results derived from this visualization platform.

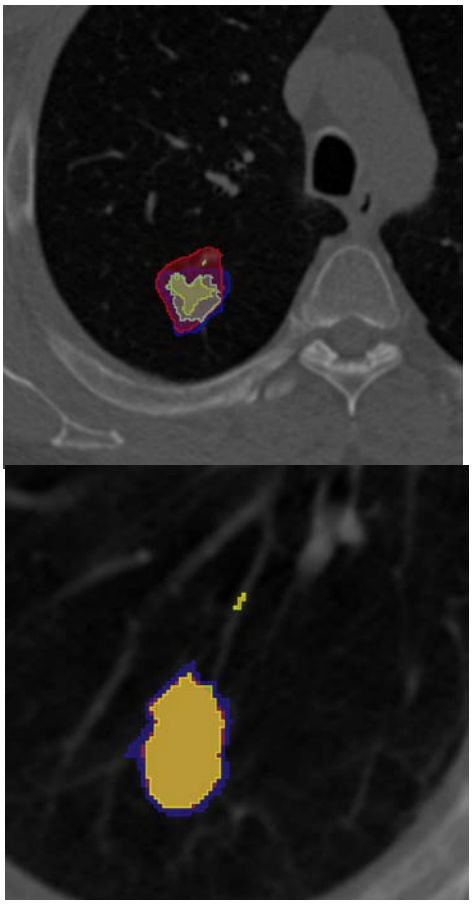


Figure 4: Low agreement (left) and high agreement (right) segmentations of lung nodules from the Lung Interval Volume Challenge.

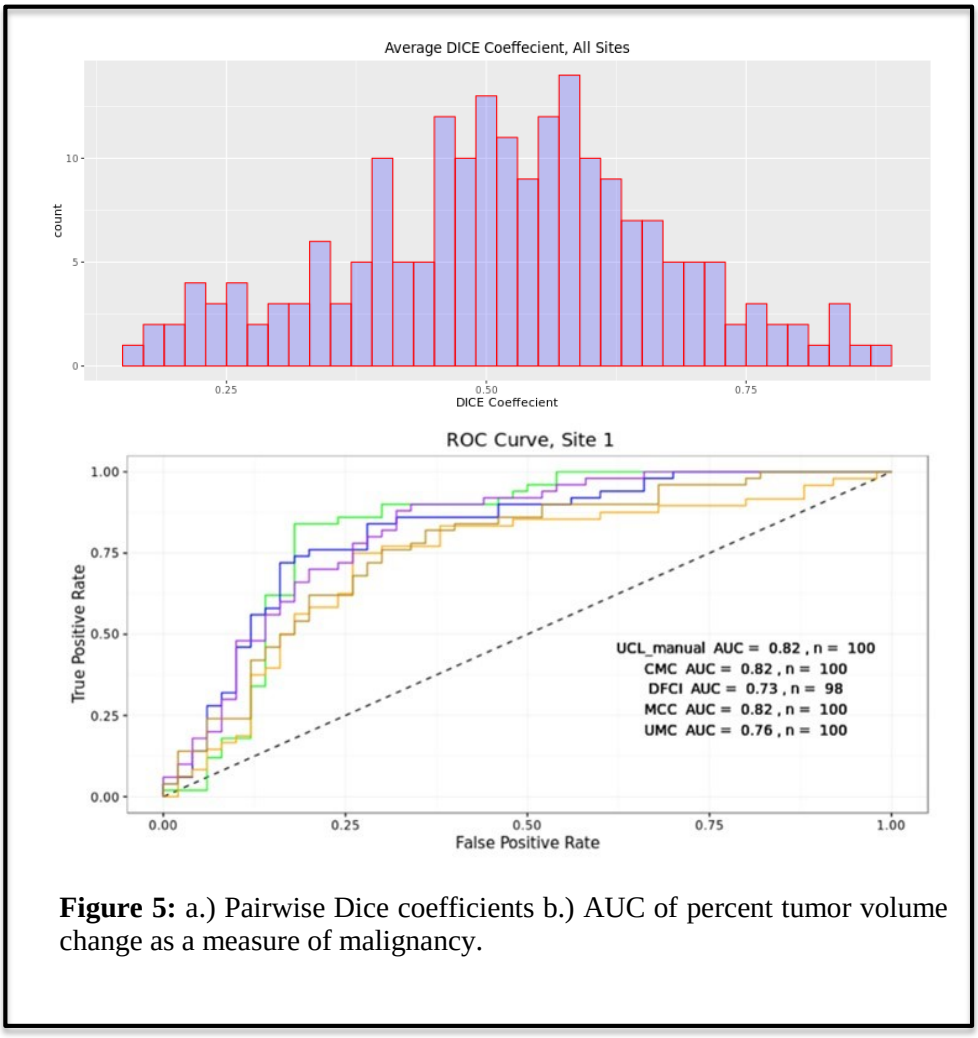


Figure 5: a.) Pairwise Dice coefficients b.) AUC of percent tumor volume change as a measure of malignancy.

Lung feature challenge: Working with the data from the lung feature challenge, we analyzed the results of 800+ unique features generated from lung nodule segmentations by seven different institutions' feature extraction software. The group developed a lexicon of radiomics feature and categorized the features are being related to shape, size, texture, margins and local and global shape descriptors. We found that many texture features were highly and unexpectedly correlated with segmentation volume, either because of errors in coding implementation or because of errors in methodology. Additionally, we found that the choice to include highly-textured border regions between tumor tissue and normal tissue can significantly change the values of certain texture features. As such, any texture features that purport to measure texture across an entire tumor area may only reflect this highly-textured area on the tumor border. The development of ground-truth "phantoms" for certain texture features will help achieve standardization across the field of texture features. We used a graphical model approach to visualize the correlation between features from different sites as well as features from different classes, as seen in Figure 6.

This work resulted in a recent publication in the Tomography QIN special issue [15]

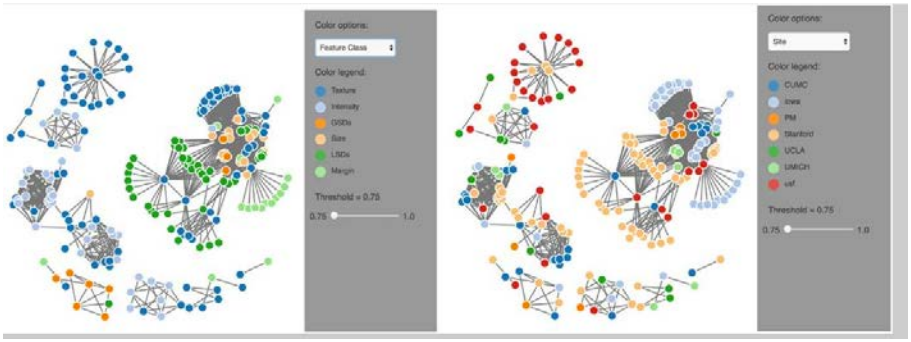


Figure 6: Graphical model showing correlation between features by feature type and site.

§ DCE repeatability study

We used two separate double-baseline studies of patients with glioblastoma to evaluate the repeatability of pharmacokinetic variables (k_{trans}, Ve) derived from DCE-MRI images. Patients (n = 45) were scanned twice without any intervening treatment between 3-4 days apart, and then pharmacokinetic parameter maps were generated using in-house software. We found that using individually-calibrated AIFs and variable flip-angle T1 mapping was less repeatable than using population AIFs and static T1 values, despite the theoretical accuracy benefits these patient-specific methods could achieve. Furthermore, we found that values obtained from methods using individual AIFs did not correlate well with methods using population-based AIFs, and that individual AIF methods generated higher median k_{trans} values than population-AIF methods. This suggests that one's chosen method for deriving pharmacokinetic parameters has a significant effect on those parameters' accuracy.

We have also found that many publically-available software packages for DCE-MRI parameter mapping do not perform well on the publically available QIBA digital reference object (a software phantom). We have identified several ways in which flawed implementations of parameter-mapping software can lead to systematically biased values for k_{trans} and Ve. These biases are compounded in noisy, real-world data, likely leading to inaccurate results in practical and clinical settings. We have developed a Python package and a C++ module in the open-source program 3D Slicer that addresses these implementation errors. It performs better than existing proprietary software packages on both noisy and non-noisy ground truth data. We also developed open source software that produces perfect results on the QIBA digital reference object (a software phantom), and the best results in a multi-institutional study (to be published shortly). A manuscript is forthcoming detailing the specific changes to our software's parameter-optimization methods.

§ Applications of imaging tools and pipeline to clinical trials and retrospective data analysis.

Finally, our suite of tools was applied to a phase II study of Tivozanib in recurrent glioblastoma. In this study, we a recurrent glioblastoma population (N=10, median age 62 (51-72)) receiving tivozanib who underwent baseline and follow-up MRIs (once every 4-week cycle). We reported that tivozanib was well tolerated but most patients progressed rapidly, and the majority of patients had little changes in tumor enhancement and perfusion imaging suggesting that his anti-angiogenic agent had limited impact on brain tumor vasculature. This paper is in press in the Journal of Neuro-Oncology [6].

PLANS FOR NEXT YEAR

We are in the final year of our current participation in the QIN. For the next year, we will continue to apply our tools to ongoing clinical studies of GBMs. We will disseminate our research through publication and presentations and making available our tools as open-source packages to the community. We hope to continue to be part of the QIN in the future.

LIST OF REFERENCES

1. Kalpathy-Cramer J, Gerstner ER, Emblem KE, Andronesi OC, Rosen B. Advanced magnetic resonance imaging of the physical processes in human glioblastoma. Cancer Res. 2014;74(17):4622-37. doi: 10.1158/0008-5472.CAN-14-0383. PubMed PMID: 25183787; PubMed Central PMCID: PMC4155518.

2. Batchelor TT, Gerstner ER, Emblem KE, Duda DG, Kalpathy-Cramer J, Snuderl M, Ancukiewicz M, Polaskova P, Pinho MC, Jennings D, Plotkin SR, Chi AS, Eichler AF, Dietrich J, Hochberg FH, Lu-Emerson C, Iafrate AJ, Ivy SP, Rosen BR, Loeffler JS, Wen PY, Sorensen AG, Jain RK. Improved tumor oxygenation and survival in glioblastoma patients who show increased blood perfusion after cediranib and chemoradiation. Proceedings of the …. 2013. PubMed PMID: 24190997.

3. Batchelor TT, Duda DG, di Tomaso E, Ancukiewicz M, Plotkin SR, Gerstner E, Eichler AF, Drappatz J, Hochberg FH, Benner T, Louis DN, Cohen KS, Chea H, Exarhopoulos A, Loeffler JS, Moses MA, Ivy P, Sorensen AG, Wen PY, Jain RK. Phase II study of cediranib, an oral pan-vascular endothelial growth factor receptor tyrosine kinase inhibitor, in patients with recurrent glioblastoma.2010;28(17):2817-23. doi: 10.1200/JCO.2009.26.3988. PubMed PMID: 20458050; PubMed Central PMCID: PMCPMC2903316.

4. Emblem KE, Farrar CT, Gerstner ER, Batchelor TT, Borra RJ, Rosen BR, Sorensen AG, Jain RK. Vessel caliber--a potential MRI biomarker of tumour response in clinical trials. Nature reviews Clinical oncology. 2014;11(10):566-84. doi: 10.1038/nrclinonc.2014.126. PubMed PMID: 25113840.

5. Emblem KE, Mouridsen K, Bjornerud A, Farrar CT, Jennings D, Borra RJ, Wen PY, Ivy P, Batchelor TT, Rosen BR, Jain RK, Sorensen AG. Vessel architectural imaging

identifies cancer patient responders to anti-angiogenic therapy. Nat Med. 2013;19(9):1178-83. doi: 10.1038/nm.3289. PubMed PMID: 23955713; PubMed Central PMCID: PMC3769525.

6. Kalpathy-Cramer J, Chandra V, Da X, Ou Y, Emblem KE, Muzikansky A, Cai X, Douw L, Evans JG, Dietrich J, Chi AS, Wen PY, Stufflebeam S, Rosen B, Duda DG, Jain RK, Batchelor TT, Gerstner ER. Phase II study of tivozanib, an oral VEGFR inhibitor, in patients with recurrent glioblastoma. Journal of neuro-oncology. 2016. doi: 10.1007/s11060-016-2332-5. PubMed PMID: 27853960.

7. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic Resonance Imaging. 2012;30(9):1323-41. doi: 10.1016/j.mri.2012.05.001. PubMed PMID: 22770690; PMCID: PMC3466397.

8. Prah MA, Stufflebeam SM, Paulson ES, Kalpathy-Cramer J, Gerstner ER, Batchelor TT, Barboriak DP, Rosen BR, Schmainda KM. Repeatability of Standardized and Normalized Relative CBV in Patients with Newly Diagnosed Glioblastoma. AJNR Am J Neuroradiol. 2015;36(9):1654-61. doi: 10.3174/ajnr.A4374. PubMed PMID: 26066626; PubMed Central PMCID: PMC4567906.

9 Prust MJ, Jafari-Khouzani K, Kalpathy-Cramer J, Polaskova P, Batchelor TT, Gerstner ER, Dietrich J. Standard chemoradiation for glioblastoma results in progressive brain volume loss. Neurology. 2015;85(8):683-91. doi: 10.1212/WNL.0000000000001861. PubMed PMID: 26208964; PubMed Central PMCID: PMC4553035.

10. Ellingson BM, Bendszus M, Boxerman J, Barboriak D, Erickson BJ, Smits M, Nelson SJ, Gerstner E, Alexander B, Goldmacher G, Wick W, Vogelbaum M, Weller M, Galanis E, Kalpathy-Cramer J, Shankar L, Jacobs P, Pope WB, Yang D, Chung C, Knopp MV, Cha S, van den Bent MJ, Chang S, Yung WK, Cloughesy TF, Wen PY, Gilbert MR, Jumpstarting Brain Tumor Drug Development Coalition Imaging Standardization Steering C. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. Neuro Oncol. 2015;17(9):1188-98. doi: 10.1093/neuonc/nov095. PubMed PMID: 26250565; PubMed Central PMCID: PMC4588759.

11. Jafari-Khouzani K, Emblem KE, Kalpathy-Cramer J, Bjornerud A, Vangel MG, Gerstner ER, Schmainda KM, Paynabar K, Wu O, Wen PY, Batchelor T, Rosen B, Stufflebeam SM. Repeatability of Cerebral Perfusion Using Dynamic Susceptibility Contrast MRI in Glioblastoma Patients. Transl Oncol. 2015;8(3):137-46. doi: 10.1016/j.tranon.2015.03.002. PubMed PMID: 26055170; PubMed Central PMCID: PMC4486737.

12. Jafari-Khouzani K. MRI upsampling using feature-based nonlocal means approach. IEEE Trans Med Imaging. 2014;33(10):1969-85. doi: 10.1109/TMI.2014.2329271. PubMed PMID: 24951680.

13. Le M, Delingette H, Kalpathy-Cramer J, Gerstner E, Batchelor T, Unkelbach J, Ayache N., MRI Based Bayesian Personalization of a Tumor Growth Model., IEEE Trans Med Imaging. 2016 Apr 29. [Epub ahead of print], PMID: 28113549

14. Le M, Delingette H, Kalpathy-Cramer J, Gerstner ER, Batchelor T, Unkelbach J, Ayache N., Personalized Radiotherapy Planning Based on a Computational Tumor Growth Model. IEEE Trans Med Imaging. 2016 Nov 8. [Epub ahead of print], PMID: 27845656

15. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren JC, Uthoff J, Dilger SK, Driscoll B, Yeung I, Hadjiiski L, Cha K, Balagurunathan Y, Gillies R, Goldgof D., Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features., Tomography. 2016 Dec;2(4):430-437. PMID: 28149958

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

1. Gutman DA, Dunn WD Jr, Cobb J, Stoner RM, Kalpathy-Cramer J, Erickson B., Web based tools for visualizing imaging data and development of XNATView, a zero footprint image viewer. Front Neuroinform. 2014 May 27;8:53, eCollection 2014. (PMID: 24904399)

2. Reuter M, Gerstner ER, Rapalino O, Batchelor TT, Rosen B, Fischl B., Impact of MRI head placement on glioma response assessment. J Neurooncol. 2014 May;118(1):123-9. Epub 2014 Feb 25. PMID: 24566765

3. Obuchowski NA, Reeves AP, Huang EP, Wang XF, Buckler AJ, Kim HJ, Barnhart HX, Jackson EF, Giger ML, Pennello G, Toledano AY, Kalpathy-Cramer J, Apanasovich TV, Kinahan PE, Myers KJ, Goldgof DB, Barboriak DP, Gillies RJ, Schwartz LH, Sullivan AD; (for the Algorithm Comparison Working Group). Quantitative imaging biomarkers: A review of statistical methods for computer algorithm comparisons, Stat Methods Med Res. 2014 Jun 11.[Epub ahead of print] (PMID: 24919829)

4. Obuchowski NA, Barnhart HX, Buckler AJ, Pennello G, Wang XF, Kalpathy-Cramer J, Kim HJ, Reeves AP; for the Case Example Working Group., Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example, Stat Methods Med Res. 2014 Jun 11.[Epub ahead of print](PMID: 24919828)

5. Kalpathy-Cramer J, Gerstner ER, Emblem KE, Andronesi OC, Rosen B., Advanced magnetic resonance imaging of the physical processes in human glioblastoma. Cancer Res. 2014 Sep 1;74(17):4622-37. PMID: 25183787

6. Jafari-Khouzani K., MRI upsampling using feature-based nonlocal means approach. IEEE Trans Med Imaging. 2014 Oct;33(10):1969-85. Epub 2014 Jun 12, PMID: 24951680

7. Eichner C, Jafari-Khouzani K, Cauley S, Bhat H, Polaskova P, Andronesi OC, O. Rapalino, R. Turner, L. L. Wald, S. Stufflebeam, and K. Setsompop, “Slice Accelerated Gradient-Echo Spin-Echo Dynamic Susceptibility Contrast Imaging With Blipped CAIPI for Increased Slice Coverage”, Magnetic Resonance in Medicine, vol. 72, no. 3, pp. 770-778, 2014.

8. Mohamed AS, Ruangskul MN, Awan MJ, Baron CA, Kalpathy-Cramer J, Castillo R, Castillo E, Guerrero TM, Kocak-Uzel E, Yang J, Court LE, Kantor ME, Gunn GB, Colen RR, Frank SJ, Garden AS, Rosenthal DI, Fuller CD., Quality Assurance Assessment of Diagnostic and Radiation Therapy-Simulation CT Image Registration for Head and Neck Radiation Therapy: Anatomic Region of Interest-based Comparison of Rigid and Deformable Algorithms, Radiology. 2014 Nov 7:132871. [Epub ahead of print]

9. Prah MA, Stufflebeam SM, Paulson ES, Kalpathy-Cramer J, Gerstner ER, Batchelor TT, Barboriak DP, Rosen BR, Schmainda KM., Repeatability of Standardized and Normalized Relative CBV in Patients with Newly Diagnosed Glioblastoma., AJNR Am J Neuroradiol. 2015 Jun 11. [Epub ahead of print] PMID: 26066626

10. Jafari-Khouzani K, Emblem KE, Kalpathy-Cramer J, Bjørnerud A, Vangel MG, Gerstner ER, Schmainda KM, Paynabar K, Wu O, Wen PY, Batchelor T, Rosen B, Stufflebeam SM., Repeatability of Cerebral Perfusion Using Dynamic Susceptibility Contrast MRI in Glioblastoma Patients., Transl Oncol. 2015 Jun;8(3):137-46. PMID: 26055170

11. Ellingson BM, Bendszus M, Boxerman J, Barboriak D, Erickson BJ, Smits M, Nelson SJ, Gerstner E, Alexander B, Goldmacher G, Wick W, Vogelbaum M, Weller M, Galanis E, Kalpathy-Cramer J, Shankar L, Jacobs P, Pope WB, Yang D, Chung C, Knopp MV, Cha S, van den Bent MJ, Chang S, Al Yung WK, Cloughesy TF, Wen PY, Gilbert MR; Jumpstarting Brain Tumor Drug Development Coalition Imaging Standardization Steering Committee, Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials., Neuro Oncol. 2015 Sep;17(9):1188-98. Epub 2015 Aug 5., Review., PMID: 2625056

12. Prust MJ, Jafari-Khouzani K, Kalpathy-Cramer J, Polaskova P, Batchelor TT, Gerstner ER, Dietrich J., Standard chemoradiation for glioblastoma results in progressive brain volume loss. Neurology. 2015 Aug 25;85(8):683-91. Epub 2015 Jul 24, PMID: 26208964

13. Ding Y, Hazle JD, Mohamed ASR,Frank SJ, Hobbs BP, Colen RR, Gunn BG, Wang J, Kalpathy-Cramer J, Garden AJ, Lai SY, Rosenthal DI, Fuller CD, Intravoxel Incoherent Motion Imaging Kinetics during Chemoradiotherapy for Human Papillomavirus-Associated Squamous Cell Carcinoma of the Oropharynx: Preliminary Results from a Prospective Pilot Study" accepted for publication in NMR in Biomedicine

14. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, et al, The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," in Medical Imaging, IEEE Transactions on, vol.34, no.10, pp.1993-2024, Oct. 2015

15. Lê M, Delingette H, Kalpathy-Cramer J, Gerstner ER, Batchelor T, Unkelbach J, Ayache N. Bayesian Personalization of Brain Tumor Growth Model. In Frangi, Hornegger, Navab, and Wells, editors, MICCAI - Medical Image Computing and Computer Assisted Intervention - 2015, Lecture Notes in Computer Science - LNCS, Munich, Germany, October 2015. Springer

16. Kalpathy-Cramer J, Napel S, Goldgof D, Zhao B, QIN multi-site collection of Lung CT data with Nodule Segmentations, QIN multi-site collection of Lung CT data with Nodule Segmentations http://dx.doi.org/10.7937/K9/TCIA.2015.1BUVFJR7

17. Ou Y, Zollei L, Da X, Retzepi K, Grant PE, Murphy SN, Kalpathy-Cramer J, Gollub RL, Field of View Normalization for Neuroimaging Analysis: Necessity, Approach and Effects. NeuroImage. (Under Review)

18. Kalpathy-Cramer J, Zhao B, Goldgof D, Gu Y, Wang X, Yang H, Tan Y, Gillies R, Napel S, A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. (in revision at J. Digital. Imaging)

19. Awan M, Kalpathy-Cramer J, Gunn GB, Beadle BM, Garden AS, Phan J, Holliday E, Jones WE, Maani E, Patel A, Choi J, Clyburn V, Tantiwongkosi B, Rosenthal DI, Fuller CD, Prospective assessment of an atlas-based intervention combined with real-time software feedback in contouring lymph node levels and organs-at-risk in the head and neck: Quantitative assessment of conformance to expert delineation, Practical Radiation Oncology, 20 December 2012 (PMID: 23853674)

20. Pinho MC, Polaskova P, Kalpathy-Cramer J, Jennings D, Emblem KE, Jain RK, Rosen BR, Wen PY, Sorensen AG, Batchelor TT, Gerstner ER., Low Incidence of Pseudoprogression by Imaging in Newly Diagnosed Glioblastoma Patients Treated With Cediranib in Combination With Chemoradiation. Oncologist. 2013 Dec 5. [Epub ahead of print], (PMID: 24309981)

21. Batchelor TT, Gerstner ER, Emblem KE, Duda DG, Kalpathy-Cramer J, Snuderl M, Ancukiewicz M, Polaskova P, Pinho MC, Jennings D, Plotkin SR, Chi AS, Eichler AF, Dietrich J, Hochberg FH, Lu-Emerson C, Iafrate AJ, Ivy SP, Rosen BR, Loeffler JS, Wen PY, Sorensen AG, Jain RK. Improved tumor oxygenation and survival in glioblastoma patients who show increased blood perfusion after cediranib and chemoradiation. Proc Natl Acad Sci., 2013 Nov 19; 110(47):19059-64. (PMID: 24190997)

22. Kalpathy-Cramer J, Freymann JB, Kirby JS, Kinahan PE, Prior FW, Quantitative Imaging Network Data Sharing and Competitive Algorithm Validation Leveraging The Cancer Imaging Archive, Translational Oncology, Transl Oncol. 2014 Feb 1;7(1):147-52. eCollection 2014 Feb. (PMID: 24772218)

23. Huang W, Li X, Chen Y, Li X, Chang MC, Oborski MJ, Malyarenko DI, Muzi M, Jajamovich GH, Fedorov A, Tudorica A, Gupta SN, Laymon CM, Marro KI, Dyvorne HA, Miller JV, Barbodiak DP, Chenevert TL, Yankeelov TE, Mountz JM, Kinahan PE, Kikinis R, Taouli B, Fennessy F, Kalpathy-Cramer J., Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. Transl Oncol. 2014 Feb 1;7(1):153-66. eCollection 2014 Feb. (PMID: 24772219)

24. Kalpathy-Cramer J, Gerstner ER, Emblem KE, Andronesi OC, Rosen B., Advanced magnetic resonance imaging of the physical processes in human glioblastoma. Cancer Res. 2014 Sep 1;74(17):4622-37. PMID: 25183787

25. Awan M, Dyer BA, Kalpathy-Cramer J, Bongers E, Dahele M, Yang J, Walker GV, Thaker NG, Holliday E, Bishop AJ, Thomas CR Jr, Rosenthal DI, Fuller CD., Auto-segmentation of the brachial plexus assessed with TaCTICS - A software platform for rapid multiple-metric quantitative evaluation of contours., Acta Oncol. 2014 Oct 3:1-5. [Epub ahead of print]

26. Mohamed AS, Ruangskul MN, Awan MJ, Baron CA, Kalpathy-Cramer J, Castillo R, Castillo E, Guerrero TM, Kocak-Uzel E, Yang J, Court LE, Kantor ME, Gunn GB, Colen RR, Frank SJ, Garden AS, Rosenthal DI, Fuller CD., Quality Assurance Assessment of Diagnostic and Radiation Therapy-Simulation CT Image Registration for Head and Neck Radiation Therapy: Anatomic Region of Interest-based Comparison of Rigid and Deformable Algorithms, Radiology. 2014 Nov 7:132871. [Epub ahead of print]

27. Shoultz-Henley S, Garden AS, Mohamed AS, Sheu T, Kroll MH, Rosenthal DI, Brandon Gunn G, Hayes AJ, French C, Eichelberger H, Kalpathy-Cramer J, Smith BD, Phan J, Ayoub Z, Lai SY, Pham B, Kies M, Gold KA, Sturgis E, Fuller CD.Prognostic value of pre-therapy platelet elevation in oropharyngeal cancer patients treated with chemoradiation. Int J Cancer. 2015 Sep 28.[Epub ahead of print] PMID: 26414107

28. Ding Y, Hazle JD, Mohamed ASR,Frank SJ, Hobbs BP, Colen RR, Gunn BG, Wang J, Kalpathy-Cramer J, Garden AJ, Lai SY, Rosenthal DI, Fuller CD, Intravoxel Incoherent Motion Imaging Kinetics during Chemoradiotherapy for Human Papillomavirus-Associated Squamous Cell Carcinoma of the Oropharynx: Preliminary Results from a Prospective Pilot Study" NMR Biomed. 2015 Oct 9. [Epub ahead of print] PMID: 26451969

29. Yankeelov TW, Mankoff DA, Schwartz LH, Lieberman FS, Buatti JM, Mountz JM, Erickson BJ, Fennessy FMM, Huang W, Kalpathy-Cramer J, Wahl RL, Linden HM, Kinahan P, Zhao B, Hylton NM, Gillies RJ, Clarke L, Nordstrom R, Rubin DL, Quantitative Imaging in Cancer Clinical Trials, Clinical Cancer Research, Clin Cancer Res. 2016 Jan 15;22(2):284-90, PMID: 26773162

30. Messer JA, Mohamed AS, Hutcheson KA, Ding Y, Lewin JS, Wang J, Lai SY, Frank SJ, Garden AS, Sandulache V, Eichelberger H, French CC, Colen RR, Phan J, Kalpathy-Cramer J, Hazle JD, Rosenthal DI, Gunn GB, Fuller CD., Magnetic resonance imaging of

swallowing-related structures in nasopharyngeal carcinoma patients receiving IMRT: Longitudinal dose-response characterization of quantitative signal kinetics., Radiother Oncol. 2016 Feb;118(2):315-22. Epub 2016 Jan 28. PMID: 26830697

31. MD Anderson Head and Neck Cancer Symptom Working Group., Beyond mean pharyngeal constrictor dose for beam path toxicity in non-target swallowing muscles: Dose-volume correlates of chronic radiation-associated dysphagia (RAD) after oropharyngeal intensity modulated radiotherapy. Radiother Oncol. 2016 Feb;118(2):304-14, Epub 2016 Feb 17. PMID: 26897515

32. Holliday E, Fuller CD, Kalpathy-Cramer J, Gomez D, Rimner A, Li Y, Senan S, Wilson LD, Choi J, Komaki R, Thomas CR Jr., Quantitative assessment of target delineation variability for thymic cancers: Agreement evaluation of a prospective segmentation challenge. J Radiat Oncol. 2016 Mar;5(1):55-61. Epub 2015 Nov 3., PMID: 27570583

33. Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, O'Sullivan F, Abramson RG, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Besa C, Kalpathy-Cramer J, Mountz JM, Laymon CM, Muzi M, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy F, Li X, The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge, Tomography. 2016 Mar;2(1):56-66, PMID: 27200418

34. Joint Head and Neck Radiotherapy-MRI Development Cooperative, Dynamic contrast-enhanced MRI detects acute radiotherapy-induced alterations in mandibular microvasculature: prospective assessment of imaging biomarkers of normal tissue injury, Sci Rep. 2016 Aug 8;6:29864. PMID: 27499209

35. Kalpathy-Cramer J, Zhao B, Goldgof D, Gu Y, Wang X, Yang H, Tan Y, Gillies R, Napel S., A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. J Digit Imaging. 2016 Aug;29(4):476-87, PMID: 26847203

36. Hansen MB, Tietze A, Kalpathy-Cramer J, Gerstner ER, Batchelor TT, Østergaard L, Mouridsen K., Reliable estimation of microvascular flow patterns in patients with disrupted blood-brain barrier using dynamic susceptibility contrast MRI, J Magn Reson Imaging. 2016 Nov 30. [Epub ahead of print] PMID: 27902858

37. Kalpathy-Cramer J, Chandra V, Da X, Ou Y, Emblem KE, Muzikansky A, Cai X, Douw L, Evans JG, Dietrich J, Chi AS, Wen PY, Stufflebeam S, Rosen B, Duda DG, Jain RK, Batchelor TT, Gerstner ER., Phase II study of tivozanib, an oral VEGFR inhibitor, in patients with recurrent glioblastoma, J Neurooncol. 2016 Nov 16. [Epub ahead of print], PMID: 27853960

38. Le M, Delingette H, Kalpathy-Cramer J, Gerstner E, Batchelor T, Unkelbach J, Ayache N., MRI Based Bayesian Personalization of a Tumor Growth Model., IEEE Trans Med Imaging. 2016 Apr 29. [Epub ahead of print], PMID: 28113549

39. Le M, Delingette H, Kalpathy-Cramer J, Gerstner ER, Batchelor T, Unkelbach J, Ayache N., Personalized Radiotherapy Planning Based on a Computational Tumor Growth Model. IEEE Trans Med Imaging. 2016 Nov 8. [Epub ahead of print], PMID: 27845656

40. Farahani K, Kalpathy-Cramer J, Chenevert TL, Rubin DL, Sunderland JJ, Nordstrom RJ, Buatti J, Hylton N, Computational challenges and Collaborative projects in the NCI Quantitative Imaging Network, Tomography, Tomography. 2016 Dec;2(4):242-249.

41. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren JC, Uthoff J, Dilger SK, Driscoll B, Yeung I, Hadjiiski L, Cha K, Balagurunathan Y, Gillies R, Goldgof D., Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features., Tomography. 2016 Dec;2(4):430-437. PMID: 28149958

U01 CA140207: Quantitative Volume and Density Response Assessment: Sarcoma and HCC as a Model

Columbia University Medical Center, New York

Lawrence H Schwartz, M.D. and Binsheng Zhao, D.Sc.

INTRODUCTION

When new cancer treatments are being tested, longitudinal images of patients' tumors are used to determine whether the treatment is working. The current uni-dimensional RECIST method and standard cut-off values for response assessment are outdated. The goal of our research is to develop new response parameters and assessment criteria for cancer treatment based on CT imaging of changes in tumor volume and density (e.g., necrosis fraction). This study will seek a proof of concept using two types of tumors (HCC and sarcoma) in which RECIST is known to correlate poorly with tumor response to treatment and clinical outcome. HCC is one of the most common malignancies worldwide, and sarcomas, though rare, carry the same molecular alterations as many other heterogeneous cancers; they are the classic cancer studied in drug discovery.

Our specific aims are therefore,

Aim 1a. To establish the reproducibility of volumetric and unidimensional measurements obtained with our advanced segmentation algorithms, using images from the SARC 011 multicenter clinical trial.

Aim 1b. To continue the development of the different algorithms over the time of the grant to reduce the fraction of lesion measurements that must be corrected by a radiologist.

Aim 2. To validate new imaging response parameters and criteria based on tumor volume, necrosis volume, and their combination using data from SARC 011 (sarcoma) and CALGB 80802 (HCC) trials.

Aim 3. To explore the correlation between the new imaging biomarker with biochemical biomarkers and the added value of the combination of both in the prediction of patient survival.

The proposed research will first develop criteria based on quantitative imaging biomarkers (tumor volume and necrosis fraction) and then compare the predictive value of these criteria to the current clinical standard. Gaining evidence that volume and necrosis are early biomarkers of response or progression would aid clinical trials in the development of cancer drugs/treatments and help match patients to the treatments that work best for them. The new criteria will be widely applicable to clinical practice because CT is the most common imaging modality for cancer, the new quantification algorithms run on popular imaging platforms, and this method will enhance radiologist productivity.

By the time we complete this project, we will deliver the followings:

- new response metrics and criteria, based on CT imaging of changes in tumor volume and necrosis fraction, for better assessing sarcoma and HCC treatments,
- robust computer algorithms for segmentation of solid tumors including tumors in the lungs, liver and lymph nodes,
- insight into the variability that exists in measuring these new response parameters using the computer-aided methods, and
- a CT image dataset containing radiologists' mark-up of tumors made from a subset of the studying data.

Success of our study will help resolve the urgent, unmet need for early and more accurate response assessment methods in the study of targeted therapies in drug discovery by rapidly translating the new imaging biomarkers into clinical trials.

DISCUSSION OF PROGRESS

There has been no modification to the Specific Aims stated in the original application. For this reporting period, from October 2015 to today, we continued moving forward our project smoothly. The following subsections will briefly address our accomplishments in the previous year.

§ Segmentation algorithms

We completed the development and validation of the segmentation algorithm for lymph nodes, the last algorithm proposed to be developed and optimized in our grant. We are now writing up a manuscript to report this technique. To date, we have successfully developed all of the three proposed algorithms; one for lung lesions, one for liver metastases, and one for lymph node metastases.

Clinical correlative studies: We proposed two clinical trial studies to validate our new volumetric and density-based response assessment method.

Clinical Trial Study #1: SARC 011, a Phase II study of patients with recurrent or refractory Ewing's sarcoma treated with IGF1R antibody (R1507)

This study is completed and now published in JCO (1).

Clinical Trial Study #2: CLAGB 80802, a Phase III randomized study of sorafenib plus doxorubicin versus sorafenib in patients with advanced HCC

To date, we completed the collection of HCC CLAGB 80802 patient data and tumor measurements as well. In total, we measured 207 patients (681 scan time points). We are in the process of final data analysis.

§ Image-based response assessment platform

We developed and published a prototype imaging platform for efficient assessment of tumor response to therapies using uni-, bi-dimensional and volumetric techniques (2) (Figure 1). We integrated our segmentation algorithms developed for lung, liver and lymph node lesions into this platform to validate the volumetric response assessment technique in clinical trials and clinical practice. Our current imaging platform offers standard functions to view, manipulate and process CT and MR images. We have used this platform for measuring tumor volumes in various clinical studies including SARC 011 and CALGB 80802.

§

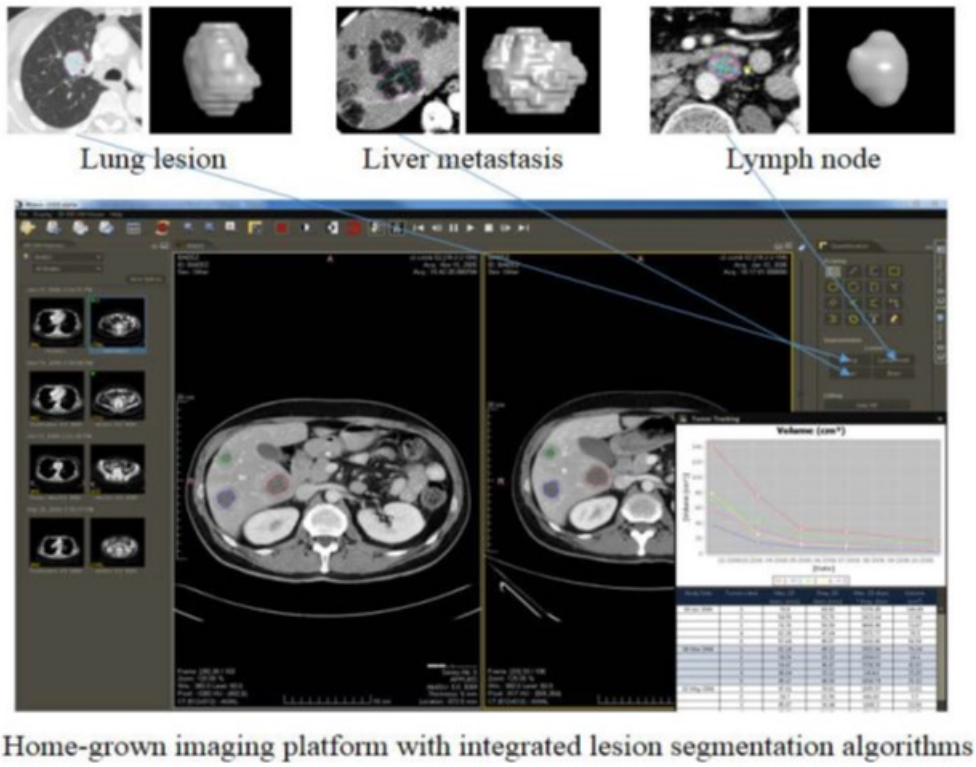


Figure 1: Home-grown imaging platform for efficient assessment of tumor response to therapies

Collaborations within other teams at QIN

Dr. Lawrence Schwartz, the contact PI of this grant, served as the Chair for QIN Executive Committee (EC) and organized monthly t-cons for the EC since May 2016.

Challenges: After successful collaboration with the other QIN teams on the Lung Segmentation Challenge (3), we continue actively participating in all CT- and PET-related challenge projects within the Image Analysis & Performance Metrics Working Group (IAPMWG) – PET/CT subgroup of QIN.

PET tumor segmentation challenge: We participated in the PET Tumor Segmentation Challenge and completed the three phases' experiments, DRO report (measuring max SUV, SUV_std, etc), and H&N lesion segmentation, using our home-developed PET lesion segmentation algorithm. This work is now accepted for publication in Medical Physics (4).

NIST Lung Nodule Change Challenge: We participated in the Lung Nodule Change Challenge to study variations in measuring tumor and tumor change over time. There are six QIN team participants (Moffitt/USF, MGH, Columbia, Stanford, UCLA, and U Michigan). We completed nodule segmentations for 100 NIST patients (50 cancer subjects and 50 non-cancer subjects) on diagnostic and 1-year follow up CT scans and provided both computer-generated and radiologist's edited tumor volumes. This project is currently under intensive data reviewing and analysis. We have participated in weekly or bi-weekly t-con discussions.

CT Feature Comparison Study: This project was among eight QIN sites (Columbia, Stanford, MGH, Moffitt/USF, UCLA, U of Iowa, Princess Margaret Cancer Centre, U Michigan). Using the Moist Run Lung Segmentation Challenge project's results (i.e., 52 segmented lung lesions x 3 algorithms x 3 repeats/algorithm), we computed the quantitative image features implemented at each site from each of these segmentations to explore features' definitions and repeatability between repeated runs of each algorithm, and reproducibility across segmentation algorithms. The comparison was performed through the C-BIBOP, the informatics platform being developed by the joint U24 grant from the four QIN sites, MGH, Columbia, Stanford andMoffitt/USF. Our preliminary result is now published (5).

§ Joint grants

After receiving a joint U24 grant entitled "Informatics Tools for Optimized Imaging Biomarkers for Cancer Research & Discovery" by the four QIN sites, as a result of the continued collaboration on the radiomic feature development and comparison study mentioned above, the 8 QIN sites jointly submitted an R24 grant application entitled "Community based terminology standards for quantitative imaging (radiomics) metadata to advance precision medicine" in middle November 2016. The purpose of this R24 grant is to develop terminology standards for radiomic features to reduce the chaos, enable identification of best features for particular uses and ultimately improve the repeatability and reliability of these features and ensure better predictive models that use them. Our portal will provide access to datasets of images, features, terminology and tools for comparison and visualization of feature. We will also encourage community-based participation in this effort by providing a means for participants to suggest new features, relationships to known semantics and clinical terminology.

In the past year, in addition to participating in the U24 regular bi-week t-cons and worked together with the other three sites to build this informatics system, we also participated in the regular t-cons to develop and submit the new R24 grant.

PLANS FOR NEXT YEAR

During the 1st No-Cost Extension year (next year),

1. We plan to publish our lymph node segmentation algorithm. The manuscript is under development.
2. We will integrate the necrosis (in HCC) segmentation algorithm into our Weasis-based response assessment imaging platform.
3. We will complete the analysis of CALGB 80802 HCC data and publish this study.
4. We will submit 100 de-identified cases (CT images) collected from the CALGB 80802 HCC clinical study to the NCI public database.
5. We will continue actively participating in the existing and new QIN challenge initiatives to which our expertise, algorithms and datasets can contribute.

LIST OF REFERENCES

.

1. Koshkin VS, Bolejack V, Schwartz LH, Schuetze S, Wahl RL, Chugh R, Reinke DL, Zhao B, Joo HO, Patel S, Schuetze SM and Baker LH. Assessment of Imaging Modalities and Response Metrics in Sarcoma - Correlation with Survival. JCO 2016; 34(30): 3680-5.

2. Yang H, Schwartz LH, and Zhao B. A Response Assessment Platform for Development and Validation of Imaging Biomarkers in Oncology. Tomography. 2016; 2(4):406-410.

3. Kalpathy-Cramer J, Zhao B, Goldgof D, Gu Y, Wang X, Yang H, Tan Y, Gillies R, Napel S, A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. J Digit Imaging 2016; 29(4):476-87.

4. Beichel R, Smith B, Bauer C, Ulrich E, Ahmadvand P, Budzevich M, Gillies R, Goldgof D, Grkovski M, Hamarneh G, Huang Q, Kinahan P, Laymon C, Mountz J, Muzi J, Muzi M, Sadek N, Oborski M, Tan Y, Zhao B, Sunderland J, and Buatti J. Multi-site Quality and Variability Analysis of 3D FDG PET Segmentations based on Phantom and Clinical Image Data. Med Phys. (in press)

5. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren JC, Uthoff J, Dilger SKN, Driscoll B, Yeung I, Hadjiiski L, Cha K, Balagurunathan Y, Robert Gillies R, and Goldgof D. Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. Tomography. 2016; 2(4):430-437.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

1. Beichel R, Smith B, Bauer C, Ulrich E, Ahmadvand P, Budzevich M, Gillies R, Goldgof D, Grkovski M, Hamarneh G, Huang Q, Kinahan P, Laymon C, Mountz J, Muzi J, Muzi M, Sadek N, Oborski M, Tan Y, Zhao B, Sunderland J, and Buatti J. Multi-site Quality and Variability Analysis of 3D FDG PET Segmentations based on Phantom and Clinical Image Data. Med Phys. (in press)

2. Yang H, Schwartz LH, and Zhao B. A Response Assessment Platform for Development and Validation of Imaging Biomarkers in Oncology. Tomography. 2016; 2(4):406-410.

3. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren JC, Uthoff J, Dilger SKN, Driscoll B, Yeung I, Hadjiiski L, Cha K, Balagurunathan Y, Robert Gillies R, and Goldgof D. Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. Tomography. 2016; 2(4):430-437.

4. Zhao B, Tan, Y, Qi J, Xie C, Tsai W-Y, Schwartz LH. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Nat Sci Rep 6; 23428, 2016. doi:10.1038/srep23428.

5. Lu L, Ehmke R, Schwartz LH, and Zhao B. Assessing Agreement between Radiomic Features Computed for Multiple Imaging Settings. PLoS One. 2016 Dec 29;11(12):e0166550. doi: 10.1371/journal.pone.0166550.

6. Hugo A, Grossmann P, Tan Y, Oxnard G, Schwartz LH, and Zhao B. Defining a Radiomic Response Phenotype: A Pilot Study using TKI therapy in NSCLC. Nat Sci Rep 6; 33860, 2016. doi:10.1038/srep33860.

7. Koshkin VS, Bolejack V, Schwartz LH, Schuetze S, Wahl RL, Chugh R, Reinke DL, Zhao B, Joo HO, Patel S, Schuetze SM and Baker LH. Assessment of Imaging Modalities and Response Metrics in Sarcoma - Correlation with Survival. JCO 2016; 34(30): 3680-5

8. Yankeelov TE, Mankoff DA, Schwartz LH, Lieberman F, Buatti J, Mountz JM, Erickson B, Fennessy F, Huang W, Kalpathy-Cramer J, Wahl R, Linden H, Zhao B, Rubin D, Quantitative Imaging in Cancer Clinical Trials. Clin Cancer Res 2016 Jan 15;22(2):284-90.

9. Kalpathy-Cramer J, Zhao B, Goldgof D, Gu Y, Wang X, Yang H, Tan Y, Gillies R, Napel S, A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. J Digit Imaging 2016; 29(4):476-87.

10. Yan J, Schwartz LH, and Zhao B. Semi-automatic segmentation of liver metastases on volumetric CT images. Med Phys. 2015 Nov;42(11):6283. doi: 10.1118/1.4932365.

11. Mountz JM, Yankeelov TE, Rubin DL, Buatti JM, Erikson BJ, Fennessy FM, Gillies RJ, Huang W, Jacobs MA, Kinahan PE, Laymon CM, Linden HM, Mankoff DA, Schwartz LH, Shim H, Wahl RL. Letter to cancer center directors: Progress in quantitative imaging as a means to predict and/or measure tumor response in cancer therapy trials. J Clin Oncol. 2014 Jul 1;32(19):2115-6.

12. Zhao B, Lee S, Lee HJ, Tan Y, Qi J, Persigehl T, Mozley PD and Schwartz LH. Variability in assessing treatment response: metastatic colorectal cancer as a paradigm. Clin Cancer Res. 2014; 20(13):3560-8.

13. Zhao B, Tan Y, Tsai WY, Schwartz HL, Lu L, Exploring variability in CT characterization of tumors: a preliminary phantom study. Transl Oncol. 2014 Feb 1;7(1):88-93.

14. Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, Goldgof DB, Hall LO, Korn R, Zhao B, Schwartz LH, Basu S, Eschrich S, Gatenby RA, Gillies RJ, Test ReTest Reproducibility Analysis of Lung CT image Features. J Digit Imaging. 2014; 27(6):805-23.

15. Tan Y, Schwartz LH and Zhao B, Segmentation of lung tumors on CT Scans using Watershed and Active Contours. Med Phys. 2013; 40(4):043502.

16. Zhao B, Tan Y, Bell DJ, Mann H, Guo P, Guo X, Marley S, Scott M , Schwartz LH, Ghiorghiu D. Exploring manual and computer-aided intra- and inter-reader variability in tumor uni-dimensional (1D), bi-dimensional (2D) and volumetric measurements. European Journal of Radiology 2013; 82:959-968.

17. Tan Y, Mann H, Guo P, Marley S, Scott M, Montalv C, Ghiorghiu D, Schwartz LS, and Zhao B. Assessing the effect of computed tomographic (CT) slice thickness on unidimensional (1D), bidimensional (2D) and volumetric measurements of solid tumors. Cancer Imaging 2012; 12: 497-505.

18. Kurland BF, Gerstner ER, Mountz JM, Schwartz LH, Ryan CW, Graham MM, et al. Promise and pitfalls of quantitative imaging in oncology. Magn Reson Imaging. 2012 Nov; 30 (9): 1301-12.

U01 CA-140204: Multi-Modality Quantitative Imaging for Evaluation of Response to Cancer Therapy

Johns Hopkins University and Washington University in Saint Louis

Eric C. Frey, Ph.D.
Michael A. Jacobs, Ph.D.
Richard L. Wahl, M.D.

INTRODUCTION

The underlying hypothesis of this project is that combining multiple quantitative image-derived parameters, whether different quantities from the same modality, multiple modalities or multiple tracers, can provide a more robust prediction and assessment of treatment response than a single imaging metric. Substantial effort has been focused on developing single-modality metrics and assessing and reducing their variability. Modalities investigated include DWI and ADC MRI, FDG and FLT PET/CT, In-111 octreotide and Y-90 and SPECT/CT. In MRI we have studied the stability and reproducibility of DWI and ADC mapping. In PET/CT we have investigated the variability of SUVmax and SUVpeak and proposed an index of defect heterogeneity. In SPECT/CT we have studied the stability of SPECT systems and developed protocols for calibrating quantitative SPECT imaging that reduce instrumentation-related sources of variability. We have investigated the variability of In-111 octreotide uptake in normal organs and investigated the reliability of simplified methods for determining normal organ VOIs based on simplified ROIs. We have validated Quantitative Y-90 bremsstrahlung SPECT reconstruction methods by phantom studies and in vivo comparison to Y-90 PET. Integrating multiparametric and multimodality images requires registration of images obtained with a variety of technical parameters including field-of-view, matrix size, and scan planes. These differences result in a challenging registration problem. We have developed a registration method based on a 3D wavelet transformation and nonlinear affine transformation that performs 3D resampling and interpolations of the reference and target radiological images without loss of information. The registration method was validated using synthetic and multiparametric MRI and PET/CT images applied to breast and prostate cancer data.

PROGRESS OVER THE PREVIOUS YEAR

§ Specific Aim 1: Optimize and characterize individual methods

We have initial results on our newly developed Radiomic-Informatic modeling of radiological imaging. Multiparametric radiological imaging is a very effective technique for diagnosis of breast cancer in patients. Conventionally, radiologists produce diagnosis using a set of carefully designed features defined by BI-RADS. However, the process of manually engineering features is difficult, time consuming, requires expert knowledge and leads to a limited set of features. Moreover, “hidden” features such as complex interaction between

different MRI images are not visually perceivable and hence are not extracted by radiological experts. In order to automatically extract useful features or representations directly from the raw multiparametric radiological imaging datasets, we developed an advanced unsupervised machine learning algorithm called the multidimensional imaging radiomics-geodesics (MIRaGe) and shown in Figure 1.

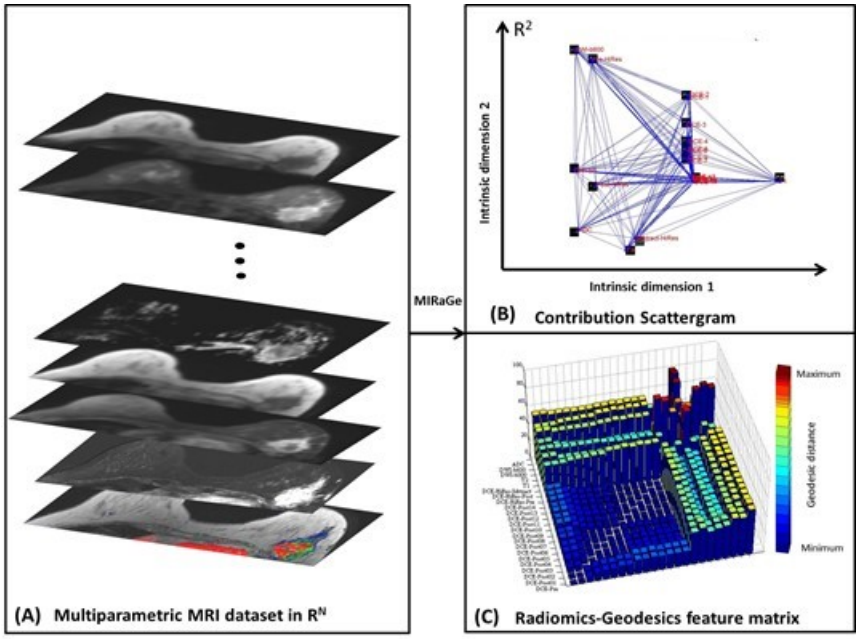


Figure 1: Demonstration of the MIRaGe algorithm in a breast cancer patient. A. Multiparametric radiological parameters from the patient. B. The resulting Contribution scattergram and informatics structure in feature space from the MIRaGe algorithm. C. The multidimensional histogram from the radiomic-geodesic distance metrics between each radiological parameter. The geodesic distances are represented using heights of different bars as well as color-coded from blue (low) to red (high).

We investigated seventy-six breast tumor patients (mean age = 52, age range = 24-80) who underwent 3T MRI breast imaging were used to test the ability of the MIRaGe algorithm to extract feature representations relevant to the task of classification of breast tumors as benign or malignant. The MRI parameters used were T1-weighted imaging, T2-weighted imaging, dynamic contrast enhanced MR imaging (DCE-MRI) and diffusion weighted imaging (DWI). The MIRaGe algorithm extracted the radiomics-geodesics features (RGF) from multiparametric breast MRI datasets of all the patients by learning their intrinsic manifold representations. The radiomics-geodesics features (RGF) represented as RGF(Ia,Ib) characterize the complex interactions between all possible image pairs (Ia,Ib) in the multiparametric MRI. The feature selection and classification model (tIso-SVM) was implemented using a combination of Student's ttest, Isometric feature mapping (Isomap) and

support vector machine (SVM) algorithms. The tIso-SVM model first filtered the set of top N significant features using Student's ttest and transformed them into a two-dimensional feature space using the Isomap algorithm and then trained the support vector machine classifier on the two-dimensional feature space to classify patients as benign or malignant. The tIso-SVM model outputs the most informative RGFs as well as the trained model for the given task, validated using k-fold and leave one out cross validation. We found that new graph theory metrics resulted in the average path length (mean RGF) and the graph diameter (maximum RGF) for the contribution scattergrams were obtained at 24.9±7.7 and 72.6±20.6 for benign patients and 25.3±5.5 and 69±15.1 for malignant patients in image distance units respectively. The tIso-SVM model was built using all the RGFs extracted by the MIRaGe algorithm. The tIso-SVM model successfully classified malignant lesions from benign lesions with a sensitivity of 93% and a specificity of 91%. The tIso-SVM model identified a total of 50 RGFs as the most informative features for classification of malignant from benign breast lesions. The top 50 RGFs primarily involved the contribution scattergram edges or paths between different dynamic contrast enhanced images. Therefore, we developed the novel MIRaGe feature extraction algorithm for automatic feature extraction from multiparametric radiological imaging and demonstrated the power of the MIRaGe algorithm at automatically discovering useful feature representations directly from the raw multiparametric MRI data. In conclusion, the MIRaGe informatics model provides a powerful tool with applicability in cancer diagnosis and a possibility of extension to other kinds of pathologies.

We are performing a comprehensive repeatability test of FDG PET/MR and PET/CT. We have accrued 8 patients to date. The patient type and data tables are shown on the next page. We have whole tumor data, but not voxel by voxel data, or subtumor region analyses. Also are working on PEAK analyses. The test and re test are highly correlated for both PET and MRI. The ADC means are reasonably reproducible. We have not yet reached our target accrual to perform Bland Altmann analyses or tests of significance. We note one case in which the ADC values were not measurable due to image artifacts. These assessments continue.

We have developed and evaluated methods for comparing quantitative methods using patient that do not require a gold standard. A paper on these methods and validation with simulated data was published in Physics in Medicine and Biology (publication 6, below).

As described in previous years, we have investigated simplified methods for defining VOIs to estimate activity concentration in normal organs. These methods are important in the context of quantitative response metrics as normal organ activities play a role as threshold or image QC metrics for metrics such as PERCIST 1.0. A paper on this was just accepted for publication in Medical Physics.

We have developed resampling methods for estimating the precisions of quantities estimated from images such as SUVmean, SUVmax, etc. A series of three papers on this work have been written. One was submitted to IEEE Transactions on Medical Imaging. Revisions and resubmission were required, and these are in process. Drafts of two other manuscripts are written and undergoing coauthor review.

	PETCT		*PETMR*				
SUBJECT	**SULmax-V1**	**SULmax-V2**	**SULmax-V1**	**SULmax-V2**	**ADCmean-V1**	**ADCmean-V2**	**HISTOLOGY**
1	10.02	9.88	12.59	11.72	-	-	rectal adenocarcinoma
2	1.97	1.4	2.31	2.41	1382.1	816.8	rectal carcinoma
3	16.24	22.36	20.78	19.77	1227.2	1221.7	squamous cell carcinoma of the cervix
4	9.63	8.21	8.76	7.83	1160.2	1002.4	endometrioid adenocarcinoma
5	13.24	12.3	14.36	13.42	1184.1	1146.6	squamous cell carcinoma of the cervix
6	13.67	15.95	14.18	14.98	943.32	944.36	squamous cell carcinoma of rectovaginal septum
7	14.73	14.9	15.11	14.17	890.57	925.8	squamous cell carcinoma of the cervix
8	11.99	14.63	15.48	16.19	1063.2	1082.2	carcinoma of the cervix

Table 1: Results from an initial set of 8 patients.

§ Specific Aim 2: Develop methods to optimally combine methods

A key premise of this grant is that the ability to accurately characterize different tissue types and response to therapy in cancer requires information from multiple radiological modalities. For example, multiparametric and multimodality radiological imaging methods, such as, magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET), provide multiple types of tissue contrast and anatomical information for tissue and response characterization. We have developed and applied multi-parameter MRI to breast cancer used a novel machine-learning model based nonlinear dimension reduction methods to integrate multiparametric MRI data (T_1WI, T_2WI, ADC, pre- and post-dynamic contrast enhanced) for improved tissue characterization of breast tissue with demonstrated excellent diagnostic ability. Moreover, we have moved these type of tissue characterization to include our advanced Radiomic discussed above. Discussion: A patent for this method has been granted.US Patent 9,256,966; Inventors: Jacobs MA, Akhbardeh A. Multiparametric Non-linear dimension reduction methods for segmentation and classification of radiological images.

In addition, we investigated the integration of Mammography, ultrasound(US) and MRI modalities used for breast cancer detection. The BIRADS lexicon provides a set of descriptors that facilitates consistent structure for assessment and reporting of breast lesions. To predict recurrence, oncologists use OncotypeDX, which stratifies patients into three risk groups: low, medium, and high]. We hypothesize that there is a relationship between imaging features defined by BIRADS and the genetic profile of cancers. To test this, we developed a machine-learning non-linear dimension reduction(NLDR) algorithm with embedded informatics. Using these techniques, we compare BIRADS descriptors to the OncotypeDX for recurrence prediction.

Methods: Patients(n=48) who underwent diagnostic breast imaging, were ER+, with available OncotypeDX were tested with the algorithm. The clinical and BIRADS parameters for mammography included breast density, asymmetry, microcalcifications(morphology, distribution), mass(size, shape, margins, density) and architectural distortion. Ultrasound parameters included mass presence, size, echogenicity, shape, margins, vascularity, and orientation. These parameters were assigned numerical values to reflect relative suspicion of each descriptor. RESULTS: There were 24 patients with low(0-17), 13 with intermediate(18-31), and seven with high risk(>31) scores from OncotypeDX. The top predictors were mammographic beast density, and mass margins and US directional size. These predictors resulted in a significant AUC(0.86±0.07). The mammographic tumor sizes in high risk groups were larger(1.9±0.58cm) compared to the low-risk group(1.38±0.58cm) with similar results for US measurements in the radial(2.7±1.2cm vs.1.2±0.8cm), AP (1.8±0.76cm vs. 0.98±0.61cm) and antiradial (2.1±1.3cm vs.1.0±0.58cm) dimensions. We created a visualization informatics heat map detailing the contribution of each parameter. The resulting risk map is shown in Figure 2.

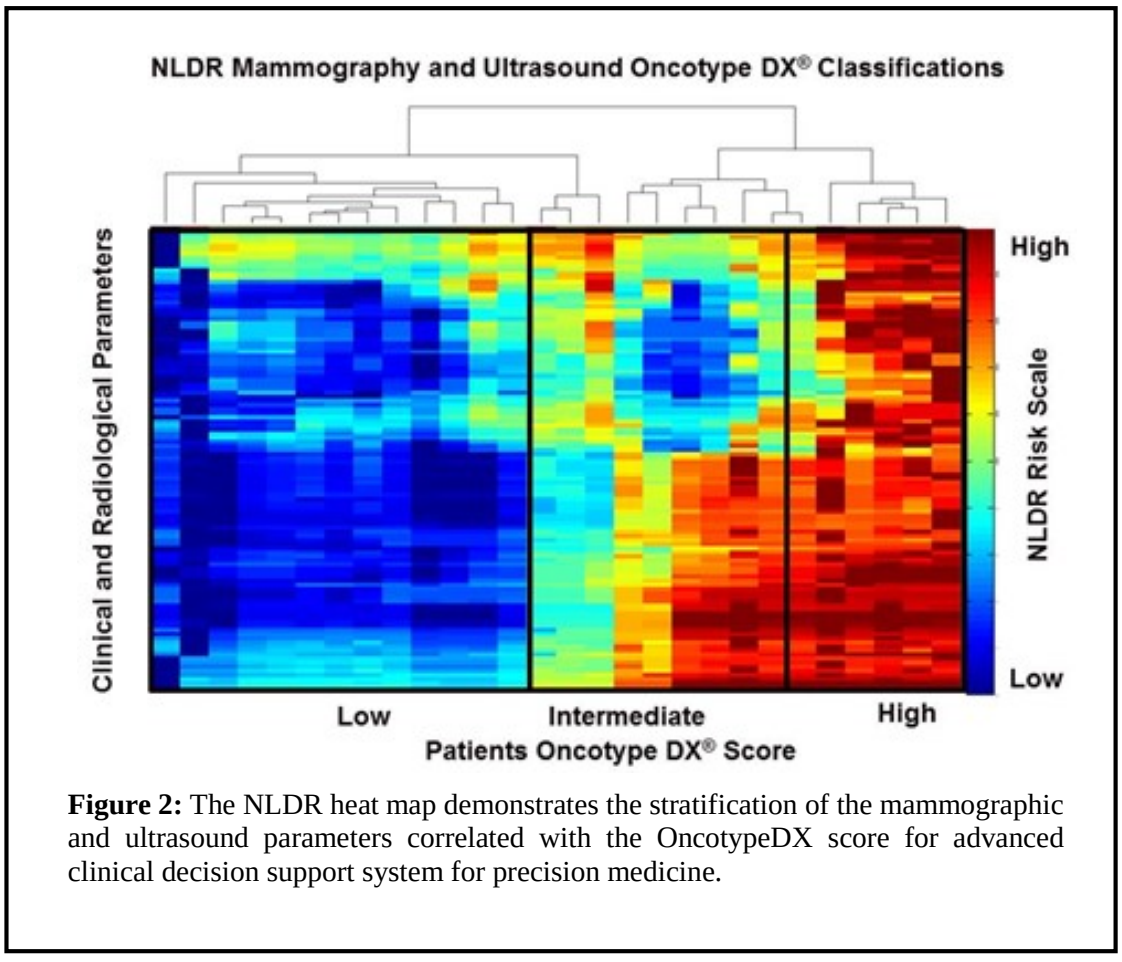


Figure 2: The NLDR heat map demonstrates the stratification of the mammographic and ultrasound parameters correlated with the OncotypeDX score for advanced clinical decision support system for precision medicine.

§ Specific Aim 3. Apply methods to data from clinical trials

As described in previous reports, we have been studying repeatability of quantitative 18F-3'-fluoro-3'-deoxy-L-thymidine (18F-FLT) positron emission tomography (PET). A paper on this has been recently accepted in has potential as a non-invasive tumor biomarker for the objective assessment of response to treatment. To guide interpretation of these quantitative data, we evaluated the repeatability of 18F-FLT PET as part of a multi-center trial involving patients with high grade glioma. A paper on this has been accepted for publication in the Journal of Nuclear Medicine (see item 4). 18F-FLT PET was performed on 10 patients with recurrent high grade glioma at 5 different institutions within the Adult Brain Tumor Consortium trial ABTC1101. Data were acquired according to a double baseline protocol in which PET was repeated within 2 days of each other with no intervening treatment. On each of the 2 imaging days, dedicated brain PET was performed at 2 time-points, 1 and 3 hours after 18F-FLT administration. Tumor standardized uptake values (SUVs) and related parameters were measured at a central lab using various volumes-of-interest: isocontour at 30% of the maximum pixel (SUVmean_30%); gradient-based segmentation (SUVmean_gradient); the maximum pixel (SUVmax); and a 1 mL sphere at the region of highest uptake (SUVpeak). Repeatability coefficients (RCs) were calculated from the relative differences between corresponding SUV measurements obtained on the 2 days. RCs for tumor SUVs were: 22.5 % (SUVmean_30%), 23.8 % (SUVmean_gradient), 23.2 % (SUVmax) and 18.5 % (SUVpeak) at 1 hour post injection. Corresponding data at 3 hours were: 22.4, 25.0, 27.3 and 23.6 %. Normalizing the tumor SUV data with reference to a background region improved repeatability and the most stable parameter was the tumor-to-background (T-to-B) ratio derived using SUVpeak (RC 16.5 %). SUV quantification of 18F-FLT uptake in glioma has an RC in the range of 18-24 % when imaging began 1 hour after 18F-FLT administration. The volume-of-interest methodology had a small but not negligible influence on repeatability, with the best performance obtained using SUVpeak. Although changes in 18F-FLT SUV following treatment cannot be directly interpreted as a change in tumor proliferation, we have established ranges beyond which SUV differences are likely due to legitimate biological effects.

We have applied the no-gold-standard methods described in publication 6 to patient data in the context of comparing methods for evaluating segmentation methods for estimating metabolic tumor volume in FDG PET/CT. A paper on this was submitted to the Journal of Medical Imaging. The initial reviews were positive, and we have addressed the concerns and resubmitted the paper.

We have developed methods for automatic segmentation of bone lesions that uses clustering methods and joint information from bone SPECT and CT scans. An abstract on this was submitted to the SNMMI annual meeting.

COLLABORATIONS WITHIN THE NETWORK

Two publications (8 and 9) have resulted from previous QIN collaborations.

PLANS FOR NEXT YEAR

This project is halfway through a no-cost extension. A renewal has been submitted that builds on the methods developed her and applies them to the problem of monitoring metastasis of prostate cancer to bones. We will continue to adapt methods to this problem and are planning a resubmission for March.

We will complete the PET-MRI clinical repeatability at Washington University. We will use these data to evaluate the repeatability of various single and multi-modality metrics for tumor response.

We have two drafts of papers on the use of bootstrap resampling methods to evaluate the precision and accuracy of VOI definition methods. We plan to submit this paper in the next two months.

We will continue studies on quantitative bone imaging in the context of response to therapy of metastatic prostate cancer.

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

1. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. *Expert Review of Precision Medicine and Drug Development.* 2016;1(2):207-226
2. Parekh V, Jacobs MA. Multidimensional Imaging Radiomics-Geodesics: A Novel Manifold Learning Based Automatic Feature Extraction Method for Diagnostic Prediction in Multiparametric Imaging. Med Phys. 2016 Jun;43(6):3373-3374
3. Ahlawat S, Baig A, Blakeley JO, Jacobs MA, Fayad LM. Multiparametric whole-body anatomic, functional, and metabolic imaging characteristics of peripheral lesions in patients with schwannomatosis. J Magn Reson Imaging. 2016 Oct;44(4):794-803
4. Blakeley JO, Xiaobu Ye, Duda DG., Halpin C, Bergner AL., Muzikansky A, Merker VL., Gerstner ER. Fayad LM; Ahlawat, S; Jacobs MA, Jain RK, Zalewski C, Dombi E, Widemann B, Plotkin S. Efficacy and biomarker study of bevacizumab for hearing loss due to neurofibromatosis type 2 associated vestibular schwannomas. Journal of Clinical Oncology. 2016;34(14):1669-75.
5. Lodge MA, Holdhoff M, Leal JP, Bag AK, Nabors LB, Mintz A, Lesser GJ, Mankoff DA, Desai AS, Mountz JM, Lieberman FS. Repeatability of 18F-FLT PET in a Multi-Center Study of Patients with High Grade Glioma. Journal of Nuclear Medicine, accepted 2016 Sep 29, J Nucl Med.
6. Jha AK, Caffo B, Frey EC. A no-gold-standard technique for objective assessment of quantitative nuclear-medicine imaging methods. Physics in medicine and biology. 2016 Mar 15;61(7):2780.
7. Malyarenko DI, Newitt D, J Wilmes L, Tudorica A, Helmer KG, Arlinghaus LR,Jacobs MA, Jajamovich G, Taouli B, Yankeelov TE, Huang W, Chenevert TL. Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials. Magn Reson Med. 2016 Mar;75(3):1312-23

8. Malyarenko DI, Wilmes L, Arlinghaus LR Jacobs MA, Huang W, Helmer KG, Taouli B, Yankeelov TE, Newitt D, Chenevert TL. QIN DAWG Validation of Gradient Nonlinearity Bias Correction in Quantitative Diffusion Weighted Imaging. Tomography, 2016 2(4) 396-405

PATENTS

US Patent Application 20160132754: Inventors: Akhbardeh A., **Jacobs MA.** *An integrated real-time tracking system for normal and anomaly tracking and the methods therefore* Date Awarded: May 12, 2016.

US Patent Application 20160171695: Inventors **Jacobs MA,** Akhbardeh, A. *Advanced Treatment Response Prediction Using Clinical Parameters and Advanced Unsupervised Machine Learning: The Contribution Scattergram*, filed on July 31, 2014, Date Awarded: June 23, 2016.

Patent Cooperation Treaty Applications

PCT: Inventors: **Jacobs MA**, Parekh V. *(IRIS):Informatics Radiomics Integration System: A novel informatics radiomics method for the integration of many types of data for classification into different groups.*

U01 CA154602: Shutter-Speed Model DCE-MRI for Assessment of Response to Cancer Therapy

Oregon Health & Science University

Wei Huang, PhD
Christopher Ryan, MD

INTRODUCTION

§ Dynamic contrast-enhanced MRI (DCE-MRI)

DCE-MRI is one of the mostly used functional imaging modalities for evaluation of cancer response to treatment. It involves collection of serial T_1-weighted images before, during, and after the IV injection of a contrast reagent (CR). It mostly measures tissue perfusion and permeability. DCE-MRI evaluations of cancer therapy response frequently use empirically quantitative (semi-quantitative) approaches to analyze DCE-MRI signal intensity time-course data, such as initial area under the curve (iAUC) [1-3], wash-in rate [4-6], time-to-peak [6], and enhancement ratio [7]. However, the results obtained are often dependent on the MRI scanner [vendor, magnetic field strength (B_0)], data acquisition details (pulse sequence and parameters), CR dose and/or injection rate, personnel skills, *etc.*, which often vary from one institution to another. This leads to high variability and low reproducibility for DCE-MRI monitoring of tumor response to therapy. Fitting a pharmacokinetic model to signal intensity time-course data to extract tissue parameters, quantitative DCE-MRI [8], is a more desirable approach. These "imaging biomarkers" are physiological quantities, in principle independent of all of the factors listed above. The parameters are usually variants of: K^{trans}, a rate constant for passive CR plasma/interstitium transfer, and v_e, the interstitial space (extracellular, extravascular) volume fraction (the putative CR distribution volume). The K^{trans} value is directly related to tumor vessel wall permeability and/or blood flow, while the v_e parameter may be a complementary measure of tumor cellularity. However, the commonly used Standard Model (SM) (or Tofts model (TM) [8]) for DCE-MRI data analysis incorrectly assumes that equilibrium inter-compartmental water exchange kinetics is infinitely fast. This is physically unrealistic, and contrary to more than 40 years of experimental results [9,10]. In a DCE-MRI study of 92 suspicious breast lesions [11], we found this erroneous assumption (for transcytolemmal water exchange) causes pharmacokinetic parameter *underestimation*. Remarkably, for K^{trans} this is significant for *only* malignant lesions. This is the major source of the limited TM DCE-MRI specificity for breast cancer detection [9-11] and of pharmacokinetic parameter dependence on CR dose and dose delivery rate [12]. Such dependencies violate the definitions of the K^{trans} and v_e parameters.

Shutter-Speed Model for Pharmacokinetic DCE-MRI Data Analysis: We have recently developed the "Shutter-Speed" Model (SSM) for DCE-MRI data analysis to account for finite inter-compartmental water exchange kinetics [13,14]. It removes the CR dose delivery rate- and/or dose-dependence mentioned above [12]. With the 92 breast lesion cohort [11], we have shown that at 100% sensitivity, tumor region-of-interest (ROI) SSM K^{trans} has

significantly (p = 0.02) higher specificity than TM K^{trans} in breast cancer diagnosis. This success is because the *TM K^{trans} underestimation is uniquely amplified for malignant tumors*. In a meta-population analysis of137 breast lesions [15], we found that the excellent SSM K^{trans} discriminative ability is *independent of MRI instrument vendor (platform/software), B_0, pulse sequence and parameters, and the CR used – the essence of a quantitative imaging biomarker.* Similar success for prostate cancer diagnosis has been found using the SSM DCE-MRI method [16].

The finite water exchange effects in DCE-MRI pharmacokinetic modeling, at least on K^{trans} estimation, become more prominent (and thus the greater extent of SM or TM K^{trans} underestimation) with increased CR extravasation [9,10]. Once the vascular shutdown begins to occur with successful cancer therapy, it can be expected that DCE-MRI shutter-speed effects will be significantly diminished. *The potential major impact of the SSM DCE-MRI method for assessment of response to cancer therapy is embedded in two rational hypotheses: (a) by correcting SM or TM DCE-MRI pharmacokinetic parameter underestimation, the SSM-derived parameters have greater dynamic ranges and thus, will be more sensitive to therapy-induced changes; and (b) vascular changes as a result of treatment will lead to amplified decreases in shutter-speed effects, which can be measured with a novel imaging biomarker, such as $\Delta K^{trans}[= K^{trans}(SSM) – K^{trans}(TM)]$*. In Specific Aim 1, SSM DCE-MRI will be compared with TM DCE-MRI, and tumor size measurement for early prediction of treatment response and assessment of residual cancer following therapy completion. Breast cancer and soft tissue sarcoma will be studied for this aim.

Currently, there is no widely adopted standard DCE-MRI protocol in data acquisition and processing for assessment of therapy response. As in the case of TM DCE-MRI, accuracy and reproducibility of parameters derived from SSM DCE-MRI may be influenced by choices of data acquisition and processing schemes, such as arterial input function (AIF) quantification [17,18]. In Specific Aim 2, the effects of DCE-MRI acquisition duration, temporal resolution (tRes), AIF quantification, and MR system platform on imaging biomarkers will be evaluated within the context of monitoring therapy response. These are necessary steps in validating SSM DCE-MRI as a reliable and reproducible tool for assessing therapeutic response before it can be standardized across multiple sites.

Informatics Approach to Software Development: For quantitative imaging biomarkers to be used in clinical practice for assessment of cancer response to therapy, software framework is needed to integrate imaging biomarkers with other patient-specific information including clinical data and molecular biomarkers. This will enable the translation of novel imaging techniques into clinical practice. Different interfaces should be presented to clinicians and imaging scientists, fitting into their respective workflows. The software framework proposed for our Specific Aim 3 will leverage the cancer Biomedical Informatics Grid (caBIG®), an information network created by NCI that enables researchers, clinicians and patients in the cancer community to share data and knowledge. A caBIG compliant approach to this aim will enable us to more readily disseminate the advances made during the course of this project to the larger cancer community.

DISCUSSION OF PROGRESS

§ Specific Aim 1: Compare SSM DCE-MRI with TM DCE-MRI, and tumor size for early prediction and evaluation of cancer therapy response

DCE-MRI Evaluation of Soft Tissue Sarcoma Response to Preoperative Chemoradiotherapy: Twenty patients (15 male, 5 female; mean age: 49 years; age range: 25 – 69 years) with histologically confirmed, ≥ 5 cm, intermediate to high grade extremity soft-tissue sarcomas, who were planned for preoperative systemic therapy and surgical resection, provided written informed consent to participate in a longitudinal research MRI study that included DCE-MRI. The tumors were located in the thigh (n = 13), knee (n = 3), and calf (n = 4).

Twelve patients were treated with our institutional standard chemoradiotherapy regimen consisting of ifosfamide and epirubicin (IE) combined with preoperative hypofractionated radiation. Each 21-day chemotherapy cycle included epirubicin 30 mg/m^2/day I.V. infusion over 3-5 minutes on days 1-4 (epirubicin was omitted during cycle 2) and ifosfamide 2.5 g/m^2/day I.V. infusion over 90 minutes on days 1-4 along with I.V. hydration, mesna, anti-emetics, and filgrastim or pegfilgrastim. Chemotherapy was planned for 3 preoperative and 3 postoperative cycles. Surgery was planned for week 9 and chemotherapy was resumed approximately 4 weeks after surgery. External beam radiation therapy was initiated concomitantly at the start of cycle 2 of chemotherapy and consisted of 28 Gy administered as 8 fractions of 3.5 Gy each over 10 days. The other eight patients were treated on a phase I clinical trial that included the addition of sorafenib (200 mg daily, 400 mg daily, or 400 mg twice daily), a vascular endothelial growth factor receptor (VEGFR) tyrosine kinase inhibitor, to the same chemoradiotharapy regimen, except that 3 rather than 4 days of chemotherapy was administered. Sorafenib administration began 2 weeks before the first cycle of chemotherapy [19]. The clinicopathologic characteristics of the patients are presented in Table 1.

The research MRI exams were performed before treatment (Visit 1, V1), after two weeks of sorafenib-only treatment in the phase I trial or after the first cycle of IE treatment in the standard regimen (V2), and after completion of preoperative therapy but prior to surgery (V3). Several patients dropped out of the MRI study at V2 and V3 due to various personal or medical reasons, resulting in completed MRI exams in 16 subjects (9 on the standard regimen and 7 on the sorafenib trial) at V2 and 12 subjects (7 on the standard regimen and 5 on the sorafenib trial) at V3.

Table 1. Clinicopathologic Characteristics of Patients

Patient Number	Age (yr)	Gender	Histologic Tumor Subtype	Tumor Grade	Pre-Therapy Size (cm)	Chemotherapy Regimen	NP (%)	Pathologic Response
1	55	Male	myxoid liposarcoma	Inter	13.5	IE+S	95	Optimal
2	60	Female	myxoid liposarcoma	Inter	13.1	IE+S	85	Sub-optimal
3	62	Female	myxofibrosarcoma	Inter	20.6	IE	50	Sub-optimal
4	38	Male	pleomorphic/undifferentiated/spindle	Inter	22.5	IE+S	95	Optimal
5	58	Male	myxoid liposarcoma	Inter	24.6	IE+S	95	Optimal
6	43	Male	spindle cell sarcoma	Inter	6.4	IE+S	30	Sub-optimal
7	58	Male	pleomorphic/undifferentiated/spindle	High	7.3	IE+S	99	Optimal
8	53	Male	synovial sarcoma	Inter	12.7	IE+S	60	Sub-optimal
9	25	Male	synovial sarcoma	Inter	10.9	IE+S	80	Sub-optimal
10	40	Female	pleomorphic liposarcoma	High	15.9	IE	80	Sub-optimal
11	53	Male	pleomorphic/undifferentiated/spindle	High	5.0	IE	99	Optimal
12	26	Male	myxofibrosarcoma	Inter	10.4	IE	99	Optimal
13	64	Male	pleomorphic/undifferentiated/spindle	High	8.6	IE	98	Optimal
14	33	Male	synovial sarcoma	High	8.0	IE	70	Sub-optimal
15	57	Male	pleomorphic/undifferentiated/spindle	Inter	9.0	IE	99	Optimal
16	34	Male	myxoid liposarcoma	Inter	5.6	IE	90	Sub-optimal
17	64	Female	pleomorphic/undifferentiated/spindle	High	5.7	IE	98	Optimal
18	69	Male	pleomorphic/undifferentiated/spindle	High	18.8	IE	90	Sub-optimal
19	40	Female	myxofibrosarcoma	Inter	6.6	IE	5	Sub-optimal
20	46	Male	synovial sarcoma	Inter	12.8	IE	30	Sub-optimal

Inter: intermediate; Pre-therapy tumor size was the longest diameter (LD) measured from post-contrast DCE-MRI images;

IE: Ifosfamide + Epirubicin; S: Sorafenib; NP: necrosis percentage

Table 1: Clinicopathologic characteristics of study patients.

DCE-MRI Data Acquisition: All the research MRI studies were performed using a 3T Siemens Tim Trio system with the body coil as the radio frequency (RF) transmitter and a phased-array body matrix coil (combined with a phased-arrayed spine matrix coil) as the RF receiver. Following scout and multi-slice axial T2-weighted MRI with fat-suppression to locate the tumor, 3D sagittal DCE-MRI data acquisition with fat-suppression was conducted using a RF-spoiled gradient-echo sequence, covering the spatial extent of the tumor. The acquisition parameters included 10^{o} flip angle, 1.5/6.0 ms TE/TR, a parallel imaging acceleration factor of two, 24 - 36 cm FOV, 448 x 224 in-plane matrix size, and 5.0 mm slice thickness. The total acquisition time for a DCE-MRI series was ~10 minutes for 36 - 80 frames of image volume of 12 - 30 slices each with 6.8 – 16.0 s temporal resolution. The variations in number of frames, number of slices per volume, and temporal resolution were due to differences in tumor size. The I.V. injection of the contrast agent (CA), Gd(HP-DO3A) [ProHance (Bracco Diagnostic Inc.)] (0.1 mmol/kg at 2 mL/s), by a programmable power injector was timed to commence after acquisition of five frames of baseline image volumes, followed by a 20-mL saline flush.

For quantification of the pre-CA T_1 value, T_{10}, proton density-weighted images were acquired immediately before and spatially co-registered with the DCE-MRI scan [20,21]. The data acquisition pulse sequence and parameters were the same as for the DCE-MRI scan except for 5° flip angle and 50 ms TR.

Pharmacokinetic Analysis of DCE-MRI Data: The soft-tissue sarcoma region of interest (ROI) was manually drawn by an experienced musculoskeletal radiologist on contiguous post-CA (approximately 120 – 180 s after CA injection) DCE-MRI image slices that cover the entire spatial extent of the CA-enhanced tumor. The radiologist also measured the longest diameter (LD) of the tumor from these images based on the RECIST guidelines [22]. Table 1 lists the tumor LD values prior to treatment (V1).

For each DCE-MRI data set, the voxel signal intensity time-courses within the multi-slice tumor ROIs were subjected to pharmacokinetic analysis using a two-compartment-three-parameter fast-exchange-regime (FXR)-allowed version of the Shutter-Speed model (SSM). The three fitting parameters of the FXR-SSM are K^{trans} (rate constant for plasma/interstitium CA transfer), v_e (volume fraction of extravascular and extracellular space), and τ_i (mean intracellular water molecule lifetime). The τ_i parameter is used to account for the finite cross-cell membrane water exchange kinetics. The CA intravasation rate constant, k_{ep}, was calculated as $k_{ep} = K^{trans}/v_e$.

Used for pharmacokinetic data analysis, the voxel T_{10} values were determined by comparing signal intensities between the spatially registered proton density-weighted images and the averaged baseline images from the DCE series [20,21]. The arterial input function (AIF), the plasma CA concentration time-course, was determined for each individual DCE-MRI data set through direct measurement. An ellipsoidal ROI was placed within the clearly visible femoral artery on a post-CA DCE image slice that was approximately through the center of the artery. The ROI signal intensity time-course was recorded and then converted to blood R_1 ($\equiv 1/T_1$) time-course using the steady-state signal intensity equation for RF-spoiled gradient-echo sequence, which was further converted to plasma CA concentration time-course using a linear relationship between R_1 and CA concentration with an CA relaxivity of 3.8 $mM^{-1}s^{-1}$ at 3T, a fixed pre-CA blood R_1 of 0.61 s^{-1} [23], and a hematocrit value set at 0.45 [21,24].

Following the FXR-SSM fittings of the DCE-MRI data, voxel-based multi-slice parametric maps of the derived pharmacokinetic parameters were generated. The mean pharmacokinetic parameter value of the whole tumor was calculated by averaging the returned voxel parameter values. For each imaging metric, including pharmacokinetic parameters and RECIST LD, the percent changes for later MRI visits relative to V1, V21% (V2 relative to V1) and V31%, were calculated.

Pathological Analysis: Pathological analysis of the post-therapy resection specimens of each soft-tissue sarcoma was performed under light microscopy using standard pathologic procedures. The pathologist estimated the amount of viable tumor and the percentage of necrosis. Pathologic response to preoperative chemoradiotherapy was classified as either optimal ($\geq$ 95% necrosis) or sub-optimal ($<$ 95% necrosis).

Statistical Analysis: Descriptive statistical analysis was conducted to summarize the pharmacokinetic parameter and RECIST LD values at each MRI visit, as well as the percent changes of these imaging metrics relative to baseline (V1). In assessing the abilities of MRI metrics (absolute values and percent changes) for evaluation of therapy response, the univariate logistic regression (ULR) analysis was used to correlate V1, V2, V3 MRI metrics, and the corresponding V21% and V31% changes, with dichotomous pathologic response endpoints, optimal *vs.* sub-optimal. A ULR C statistics value, equivalent to the area under the Receiver Operating Characteristic curve (ROC AUC), in the range of 0.9 – 1.0 indicates an excellent marker; 0.8 – 0.9, a good marker; 0.7 – 0.8, a fair marker; < 0.7, a poor marker. Two sample t test was used to evaluate the differences in imaging metrics and the corresponding percent changes between the two response groups, as well as between the two cohorts that received standard chemoradiotherapy and sorafenib plus standard chemoradiotherapy, respectively. Fisher's exact test was used to determine if there was association between therapy regimen (with and without sorafenib) and response status (optimal *vs.* sub-optimal). Pearson's correlation analysis was used to examine relationships between MRI metrics and necrosis percentage (NP) of the resection specimens.

Results and Discussion: As shown in Table 1, pathological analyses of the surgical specimens revealed that 9 (45%) patients (5 on the standard regimen and 4 on the sorafenib trial) achieved optimal response to preoperative chemoradiotherapy, while the other 11 patients (7 on the standard regimen and 4 on the sorafenib trial) had sub-optimal response. There was no statistically significant (Fisher's exact test, $P = 1.0$) association between the use of sorafenib and pathologic response status, nor any significant (two sample t test, $P > 0.2$) differences in any MRI metric (RECIST LD and pharmacokinetic parameters) at any visit and the corresponding percent changes between the two cohorts on different therapy regimens. Therefore, we combined the two patient cohorts in assessing the utility of quantitative DCE-MRI for evaluation of response to preoperative therapy.

Table 2 lists the mean ± SD whole tumor MRI metric values of the optimal and sub-optimal response groups and the corresponding ULR C statistics values for discrimination of the two response groups. Only the absolute pharmacokinetic parameters and the V21% and V31% changes with $C \geq 0.7$, representing fair or better imaging biomarkers, are listed. The C value (0.69) for V31% RECIST LD change is presented for the purpose of comparison. V1, V2, and V21% metrics were obtained before and 2-3 weeks after the start of therapy, and thus, are potential early predictors of therapy response. The V2 K^{trans} parameter was an excellent ($C = 0.9$) early discriminator of optimal *vs.* sub-optimal pathologic response, while V1 and V2 k_{ep}, V1 and V21% K^{trans}, V21% v_e, and V21% τ_i were fair to good ($0.7 \leq C \leq 0.8$) markers for early prediction of response. Compared with good to excellent predictive abilities of the K^{trans} and k_{ep} metrics, the V21% change in RECIST LD was just a fair early predictor of response. Several pharmacokinetic metrics obtained after the completion of chemoradiotherapy, including V3 k_{ep}, K^{trans} and v_e, and V31% v_e, were good to excellent ($0.8 < C < 1.0$) discriminators of optimal *vs.* sub-optimal response, whereas the V31% change in RECIST LD was a poor ($C < 0.7$) marker of response. For the imaging metrics listed in Table 2 with ULR C values ≤ 0.77, including V21% RECIST LD, the differences between the two response groups were not statistically significant ($P > 0.05$).

Table 2. Evaluation of Pathologic Response (Optimal *vs.* Sub-Optimal Response)

MRI Metric	Optimal	Sub-Optimal		ULR C value
	Mean ± SD	Mean ± SD	P value	
V3 k_{ep} (min^{-1})	0.11 ± 0.03	0.45 ± 0.40	0.0024	0.97
V3 K^{trans} (min^{-1})	0.02 ± 0.01	0.21 ± 0.26	0.0012	0.94
V3 v_e	0.15 ± 0.08	0.36 ± 0.24	0.021	0.91
V2 K^{trans} (min^{-1})	**0.05 ± 0.03**	**0.20 ± 0.13**	**0.0020**	**0.90**
V31% v_e	-52% ± 28%	53% ± 92%	0.021	0.84
V21% K^{trans}	**-38% ± 25%**	**-9% ± 33%**	**0.038**	**0.80**
V1 k_{ep} (min^{-1})	**0.32 ± 0.24**	**0.75 ± 0.43**	**0.010**	**0.80**
V2 k_{ep} (min^{-1})	**0.29 ± 0.26**	**0.68 ± 0.48**	**0.045**	**0.78**
V31% K^{trans}	-68% ± 21%	2% ± 75%	0.043	0.78
V3 τ_i (s)	1.42 ± 0.83	0.85 ± 0.84	0.25	0.77
V1 K^{trans}(min^{-1})	**0.10 ± 0.09**	**0.21 ± 0.16**	**0.055**	**0.72**
V21% RECIST LD	**7% ± 10%**	**-3% ± 8%**	**0.13**	**0.72**
V21% τ_i	**38% ± 66%**	**13% ± 43%**	**0.13**	**0.71**
V21% v_e	**-18% ± 42%**	**15% ± 50%**	**0.16**	**0.70**
V31% RECIST LD	-11% ± 22%	-7% ± 8%	0.66	0.69

ULR: univariate logistic regression; SD: standard deviation; P value: two sample t test; V1, V2, and V21% metrics are bolded as early predictors of therapy response.

Table 2: Evaluation of pathologic response (optimal vs. sub-optimal response).

Figure 1 shows examples of V1 – V3 colored tumor K^{trans} maps from two soft-tissue sarcoma patients who had optimal (1A, left column; Patient 13 in Table 1) and sub-optimal (1B, right column; Patient 6 in Table 1) responses, respectively. The K^{trans} color scales are different for the two patients, but kept the same throughout the three visits for each patient to demonstrate changes in the longitudinal study. The six panels in Fig. 1 are cropped images (without zooming) of K^{trans} maps overlaid on post-CA DCE-MRI image slices that were approximately through the center of the tumor. The FOV of DCE-MRI acquisition was kept the same for all three visits for each patient. Thus, it is rather apparent in Fig. 1 that there was not only minimal change in the imaging tumor size for each patient but also little difference in tumor size change between the optimal and sub-optimal responders in the longitudinal study. However, substantial decrease in tumor K^{trans} was observed at V2 compared to V1, and continued to V3 for the optimal responder, while there were no noticeable K^{trans} changes from V1 to V2, and to V3 for the sub-optimal responder.

The Pearson's correlation coefficient, R, and the P value for statistical significance are summarized in Table 3 for correlations between the absolute MRI metric values (and percent changes) and the pathologically measured NP values of the resection specimens. Only the imaging metrics with statistically significant ($P < 0.05$) correlations with NP are listed, except for V1, V2, V3, V21%, and V31% RECIST LD metrics which are listed for comparison. Figures 2 and 3 show examples of linear regressions between NP and MRI metrics pre-therapy (Figure 2), at the early stage of therapy (Figure 2), and post-therapy (Figure 3). While the negative correlations of V1 K^{trans} (Fig. 2A) and k_{ep} (Figure 2B), V2 K^{trans} (Figure 2C), and V3 K^{trans} (Figure 3A), v_e (Figure 3B), and k_{ep} (Figure 3C) with NP were statistically significant ($P < 0.05$), there were no significant ($P > 0.2$) associations between any RECIST LD measures and NP.

Table 3. Pearson's Correlation of MRI Metric with NP

MRI Metric	R	P
V3 K^{trans}	-0.93	< 0.0001
V3 k_{ep}	-0.92	< 0.0001
V31% K^{trans}	-0.89	0.0001
V3 v_e	-0.75	0.005
V2 K^{trans}	-0.62	0.010
V1 k_{ep}	-0.55	0.012
V31% v_e	-0.63	0.028
V1 K^{trans}	-0.45	0.047
V21% RECIST LD	0.31	0.25
V31% RECIST LD	-0.20	0.52
V3 RECIST LD	0.19	0.56
V1 RECIST LD	0.071	0.76
V2 RECIST LD	0.078	0.77

NP: necrosis percentage of the resection specimen; R: Pearson's correlation coefficient; $P < 0.05$ indicates statistically significant correlation.

Table 3: Pearson's Correlation of MRI metric with NP

This study shows that changes in tumor functions as measured by quantitative DCE-MRI are superior to changes in RECIST-based imaging tumor size measurement for early prediction of soft-tissue sarcoma pathologic response to preoperative therapy, suggesting that therapy-induced tumor functional changes precede changes in tumor size. These results suggest that soft-tissue sarcomas with low perfusion and permeability at baseline (pre-therapy) and/or after one cycle of chemotherapy may have less angiogenesis-induced abnormal vasculature, and therefore, better drug delivery and response. The potential of noninvasive functional imaging methods, such as DCE-MRI, for accurate early prediction of therapy response may have profound importance in the emerging era of precision and personalized medicine. Early identification of poor responders to a therapy regimen may allow rapid adjustment in treatment planning and spare these patients from ineffective therapies and the associated toxicities.

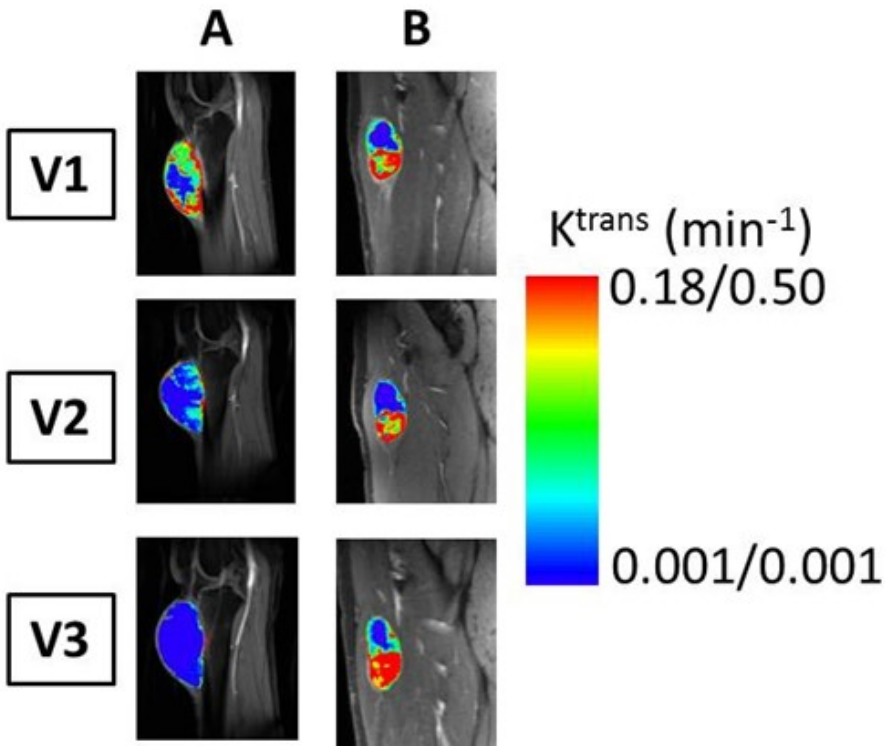


Figure 1: Visit 1 (V1, pre-therapy), V2 (after two weeks of sorafenib or one cycle of chemotherapy), and V3 (after completion of chemoradiotherapy) color parametric K^{trans} maps of two soft-tissue sarcomas: an optimal (A, left column, 98% necrosis in resection specimen) and a sub-optimal (B, right column, 30% necrosis in resection specimen) responder to preoperative therapy. The maps were generated for tumor ROIs defined on multiple contiguous image slices, and the ones on the image slices through the central portion of the tumors are displayed here. For each tumor, the K^{trans} color scale is kept the same for all three visits for easy visualization of therapy-induced changes. The left and right color scales correspond to K^{trans} maps in A and B, respectively.

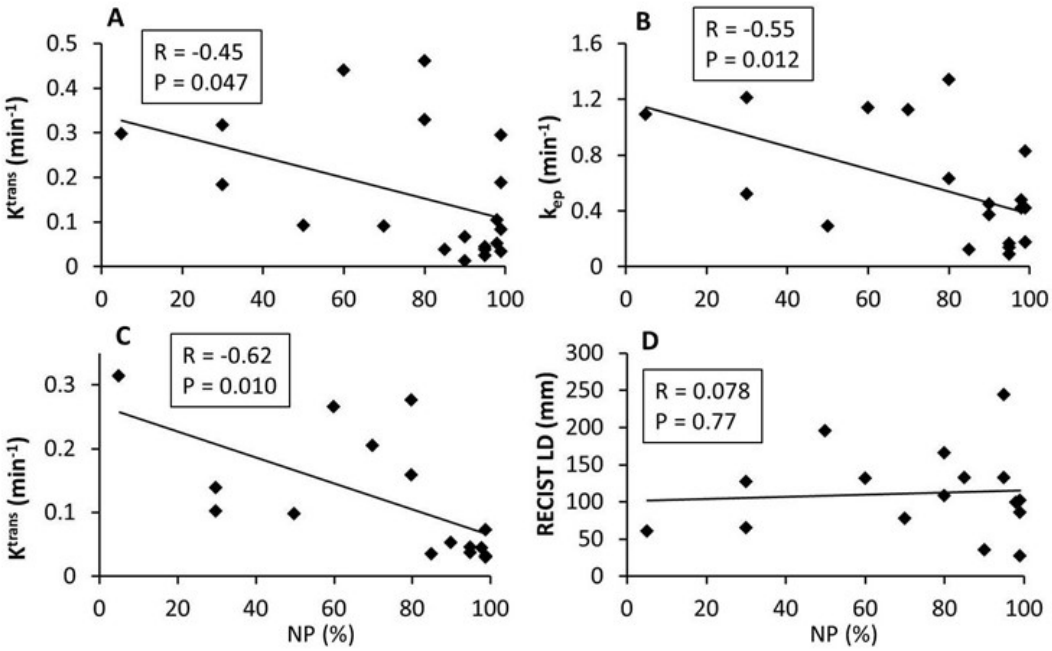


Figure 2: Scatter plots of pathologically measured necrosis percentage (NP) of the resection specimen against K^{trans} (A) and k_{ep} (B) pre-therapy (V1), and K^{trans} (C) and RECIST LD (D) after two weeks of sorafenib or one cycle of chemotherapy (V2). The straight line in each panel represents a linear regression. The Pearson's correlation coefficient R and P values for the four imaging metrics are listed in Table 3 and shown in each panel. The data points are from the initial cohort of 20 patients for the V1 metrics (A and B) and the 16 patients who continued to have the V2 MRI studies (C and D).

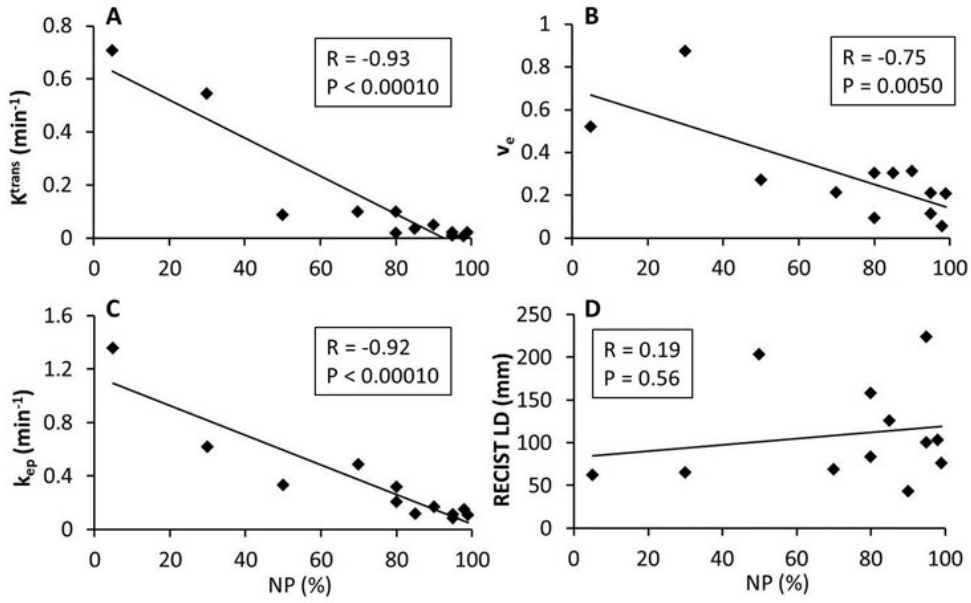


Figure 3: Scatter plots of pathologically measured necrosis percentage (NP) of the resection specimen against post-therapy (V3) MRI metrics: (A) K^{trans}, (B) v_e, (C) k_{ep}, and (D) RECIST LD. The straight line in each panel represents a linear regression. The Pearson's correlation coefficient R and P values for the four imaging metrics are listed in Table 3 and shown in each panel. The data points are from 12 patients who completed the V3 MRI studies among the initial cohort of 20 patients.

The post-therapy (V3) K^{trans}, k_{ep}, and v_e parameters all showed strong negative correlations with NP of the resection specimens and were excellent markers (ULR C value > 0.9) for discrimination of optimal and sub-optimal responders. However, there was no significant correlation between V3 RECIST LD and NP. This suggests that a functional imaging study such as DCE-MRI following preoperative therapy may yield additional information potentially useful for surgical planning and subsequent management. The negative correlations of post-therapy K^{trans} and k_{ep} with NP are expected, as increased tumor necrosis is usually associated with decreased perfusion, and thus the DCE-MRI measures of microvascular properties. The similar relationship observed between post-therapy v_e and NP is, however, intriguing. With cancer cell death and increased necrosis after the preoperative chemoradiotherapy, the v_e value is generally expected to increase with increased necrosis. The opposite was seen in this study and the probable reason for this is that, though defined as extravascular and extracellular volume fraction, v_e as measured by DCE-MRI is in principle the putative CA distribution volume fraction. With increased necrosis and decreased viable perfused tumor area, the CA distribution volume fraction, which was reported as an averaged value over the whole tumor volume, was presumably decreased as well. It is possible that the estimated v_e value may actually increase with increased necrosis if the DCE-MRI acquisition time is long enough to allow substantial diffusion of CA molecules into the necrotic area.

In conclusion, we have demonstrated the utility of quantitative DCE-MRI for early prediction and evaluation of soft-tissue sarcoma response to preoperative chemoradiotherapy in 20 patients with lower extremity tumors. Tumor functional changes as measured by quantitative DCE-MRI parameters such as K^{trans} and k_{ep} provided better early prediction of pathologic response outcome than the conventional approach of measuring changes in imaging tumor size. Post-therapy DCE-MRI parameters, not the RECIST LD metric, were found to significantly correlate with percent necrosis of the resection specimens. The SSM-unique τ_i parameter could be a useful imaging biomarker of metabolic activity that can be used to evaluate tumor response to therapy.

DCE-MRI Prediction and Assessment of Breast Cancer Response to Neoadjuvant Chemotherapy

Neoadjuvant chemotherapy (NACT) is increasingly used before surgery to treat locally advanced breast cancer. Though pathological response is a good indicator of survival, it can be determined only after surgery. Thus, there is genuine need of noninvasive imaging method to monitor and provide early prediction of therapeutic response. This allows swift introduction of alternative treatment for non-responding patients. In addition, accurate assessment of residual disease following NACT completion improves surgery decision making such as lumpectomy *vs.* mastectomy. Conventionally, tumor size measurement is used to evaluate response. However, changes in tumor size often occur late during treatment and may over- or under-estimate residual disease.

Quantitative DCE-MRI has been shown effective for early prediction of breast cancer response NACT [21,25]. However, few have investigated the utility of DCE-MRI for evaluation of post-NACT residual disease, which can have important implications for surgical decision making of mastectomy *vs.* lumpectomy. In this project we compared quantitative imaging biomarkers estimated from pharmacokinetic (PK) analysis of DCE-MRI data with imaging tumor size measurement for early prediction of breast cancer NACT response and evaluation of residual disease, and the standard Tofts model (TM) with the Shutter-Speed model (SSM) PK analysis within the context of response assessment. Here we report our results in the first five years of this U01 project using SSM and TM DCE-MRI for assessment of breast cancer response to NACT.

A total of 6 breast cancer patients have been recruited for the MRI studies in the past year, making a total accrual of 65 subjects thus far. They all underwent six-eight cycles of NACT before surgery. As shown in the study schema, the research DCE-MRI studies were performed at Visit 1 (V1) - before NACT, at Visit 2 (V2) - after first NACT cycle, at Visit 3 (V3) - at NACT midpoint or before change of drugs, and at Visit 4 (V4) - after NACT completion, but before surgery. At the time of this report, 47 subjects with 49 independent primary tumors have completed the longitudinal MRI studies and undergone surgeries, and their MRI data have been analyzed and correlated with pathological endpoints.

DCE-MRI Data Acquisition and Analysis: Axial bilateral DCE-MRI images with fat-saturation and full breast coverage were acquired with a 3D gradient-echo TWIST (Time-resolved angiography WIth Stochastic Trajectories) sequence using a 3T Siemens scanner

[21]. The TWIST sequence is a k-space undersampling and data sharing gradient-echo sequence delivering both high spatial and temporal resolution for breast DCE-MRI. Other details of DCE-MRI acquisition included 10^0 flip angle, 2.9/6.2 ms TE/TR, a parallel imaging acceleration factor of two, 30-34 cm FOV, 320x320 matrix size, and 1.4 mm slice thickness. The total acquisition time was ~ 10 min with 16-20 s temporal resolution. Gd contrast agent (Prohance®) IV injection (0.1 mmol/kg at 2 mL/s) was carried out following acquisition of two baseline image volumes. Tumor ROIs were drawn by experienced radiologists who also measured tumor size according to well-established (one dimensional) RECIST guidelines. The ROI and pixel-by-pixel (within ROI) DCE time-course data were subjected to both the TM and the SSM pharmacokinetic analyses to extract K^{trans}, v_e, k_{ep} (= K^{trans}/v_e), and τ_i (from SSM only) parameters. The whole tumor ROI DCE-MRI parameter values were calculated by averaging the ROI values from each of the image slices covering the entire tumor, weighted by the pixel numbers within the ROI in each image slice.

Pathology and Statistical Analyses: The pre-therapy biopsy specimens along with the post-therapy surgical specimens and lymph nodes were analyzed to evaluate pathological responses. Two pathological metrics [26], RCTD (relative changes in tumor density) and RCB (residual cancer burden), were computed. Pathologic complete response (pCR) is defined as RCTD = -1.0 and RCB = 0; non-response (pNR) as RCTD $\geq$ 0; and partial response (pPR) as -1.0 < RCTD < 0. Non-pCR includes both pPR and pNR and can be further stratified into RCB classes based on RCB index values [26]: RCB-I: 0 < RCB $\leq$ 1.36; RCB-II: 1.36 < RCB $\leq$ 3.28; RCB-III: RCB > 3.28. Since the MRI metrics were measured from the primary breast tumor only, the in-breast component of RCB was also computed for correlation with the MRI results.

The pathologic endpoints were correlated with the MRI metrics using the univariate logistic regression (ULR) analysis to identify imaging biomarkers for early prediction of response. A ULR C statistics value, a measure equivalent to the area under the Receiver Operating Characteristic curve (ROC AUC), in the range of 0.9 – 1.0 indicates an excellent predictor; 0.8 – 0.9, a good predictor; 0.7 – 0.8, a fair predictor; < 0.7, a poor predictor. ULR analysis and the Spearman's correlation (SC) were used to correlate MRI metrics with RCB ranks and numerical values, respectively.

Results and Discussion: 12 patients achieved pathologic complete response (pCR) (RCB = 0) while the other 35 (37 tumors) were non-pCRs. Table 4 shows the mean ± SD values of the PK parameters and the percent changes (e.g., V21%: percent change of V2 relative to V1) for the two groups and P values for comparison, as well as the ULR C statistics values (equivalent to AUC of ROC analysis) for early prediction of pCR *vs.* non-pCR. Only the metrics at V3 or earlier and with C $\geq$ 0.7 (indicating fair or better early predictor of response) are listed. RECIST LD and its percent changes are listed for comparison. V21% values of several PK parameters were good (C > 0.8) early predictors of response, with parameters of both PK models performing equally well. However, even at NACT midpoint (V3), RECIST LD and its percent changes remained poor (C < 0.7) predictors of response. Figure 4 shows a column graph of the mean±SD V21% changes of some of these MRI metrics for the pCR (black column) and non-pCR (gray column) groups. Figure 5 shows representative K^{trans}(SSM) and τ_i color maps of a pCR (1A) and a non-pCR (1B) at V1 and

V2. Compared to the non-pCR, the pCR tumor had considerable decrease in K^{trans} and increase in τ_i after only one NACT cycle.

Table 4. Early Prediction of Pathologic Response (pCR *vs.* non-pCR)

MRI Metric	pCR	non-pCR		ULR C value
	Mean ± SD	Mean ± SD	P value	
V21% K^{trans}(TM)	-58% ± 17%	-13% ± 36%	0.006	0.89
V21% K^{trans}(SSM)	-65% ± 17%	-16% ± 43%	0.006	0.89
V21% τ_i	33% ± 38%	-5% ± 36%	0.011	0.82
V21% k_{ep} (TM)	-60% ± 41%	-20% ± 43%	0.014	0.82
V3 k_{ep} (SSM)	0.04 ± 0.02	0.20 ± 0.31	0.072	0.77
V2 K^{trans} (SSM)	0.05 ± 0.03	0.10 ± 0.08	0.040	0.75
V31% K^{trans} (SSM)	-75% ± 12%	-47% ± 36%	0.035	0.75
V2 k_{ep} (TM)	0.16 ± 0.14	0.26 ± 0.14	0.035	0.75
V21% k_{ep} (SSM)	-51% ± 82%	-19% ± 87%	0.270	0.74
V31% K^{trans} (TM)	-66% ± 17%	-38% ± 37%	0.044	0.74
V2 K^{trans} (TM)	0.03 ± 0.02	0.06 ± 0.04	0.038	0.74
V31% k_{ep} (TM)	-77% ± 18%	-40% ± 60%	0.054	0.74
V31% k_{ep} (SSM)	-81% ± 20%	-15% ± 77%	0.086	0.73
V2 k_{ep} (SSM)	0.10 ± 0.09	0.20 ± 0.15	0.060	0.73
V3 k_{ep} (TM)	0.07 ± 0.05	0.20 ± 0.20	0.089	0.72
V3 v_e (SSM)	0.71 ± 0.09	0.63 ± 0.16	0.156	0.71
V31% τ_i	61% ± 91%	9% ± 69%	0.089	0.70
V31% RECIST LD	-46% ± 29%	-29% ± 26%	0.101	0.68
V3 RECIST LD	16.1 ± 10.7	26.4 ± 18.9	0.116	0.65
V1 RECIST LD	33.3 ± 20.0	41.3 ± 23.8	0.295	0.62
V2 RECIST LD	28.4 ± 18.9	36.9 ± 21.9	0.240	0.62
V21% RECIST LD	-17% ± 32%	-10% ± 14%	0.293	0.51

TM metrics are listed in black; SSM metrics in red; and RECIST LD in purple; K^{trans} and k_{ep} have the unit of min^{-1}; RECIST LD has the unit of mm.

Table 4: Early prediction of pathologic response (pCR vs. non-pCR)

Table 5 lists coefficient R and P values for significant ($P < 0.05$) SC between V4 imaging metrics and RCB index value, while Table 6 is the Table 2 equivalent for in-breast RCB. After NACT completion, K^{trans} and k_{ep} of both models and RECIST LD were positively correlated with RCB, while τ_i was negatively associated with in-breast RCB. The correlation was generally strengthened when in-breast RCB was used, as the imaging metrics were from the primary tumor only.

Our results thus far suggest that changes in tumor vasculature precede size changes in response to NACT. After only one cycle of NACT, the % changes (relative to baseline) or actual values of quantitative DCE-MRI biomarkers can provide excellent early prediction of eventual pathologic response to the entire course of NACT, while the RECIST measure of tumor size is not a good predictor of response at early time point or even the midpoint of NACT (results not shown here). V21 percent changes of both TM and SSM K^{trans} parameters

are excellent early predictors, suggesting the systematic differences between the two models are of less concern when % change is used for response evaluation. However, *the absolute SSM parameter values generally offer larger separations of the two response groups (see Fig. 4) than their TM counterparts and thus are more sensitive measures of therapeutic response*. This is most likely due to the incorporation of the exchange effects in the SSM analysis. Furthermore, SSM analysis allows quantification of τ_i, a potential imaging biomarker of metabolic activity [21, 27, 28]. The utility of τ_i is clearly demonstrated in early prediction of response (Table 4) and assessment of RCB (Table 6). The potential of τ_i as a robust early predictor of breast cancer therapy response is further supported by our observation that it is the only pre-NACT imaging metric that correlates with RCB with near statistical significance (P = 0.053).

Table 5. Spearman's Correlation with RCB Index Value

MRI Metric	R	P
V4 RECIST LD	0.46	0.0043
V4 K^{trans}(SSM)	0.41	0.013
V4 K^{trans}(TM)	0.39	0.020
V4 k_{ep}(TM)	0.38	0.023
V4 k_{ep}(SSM)	0.38	0.023

R: correlation coefficient. TM metrics are listed in black; SSM metrics in red; RECIST LD in purple.

Table 5: Spearman's Correlation with RCB index value

Table 6. Spearman's Correlation with In-Breast RCB Index Value

MRI Metric	R	P
V4 K^{trans}(SSM)	0.60	0.00010
V4 K^{trans}(TM)	0.59	0.00020
V4 RECIST LD	0.52	0.0010
V4 τ_i	-0.39	0.020
V4 k_{ep}(SSM)	0.38	0.023
V4 k_{ep}(TM)	0.37	0.026

R: correlation coefficient. TM metrics are listed in black; SSM metrics in red; RECIST LD in purple.

Table 6: Spearman's Correlation with Breast RCB index value.

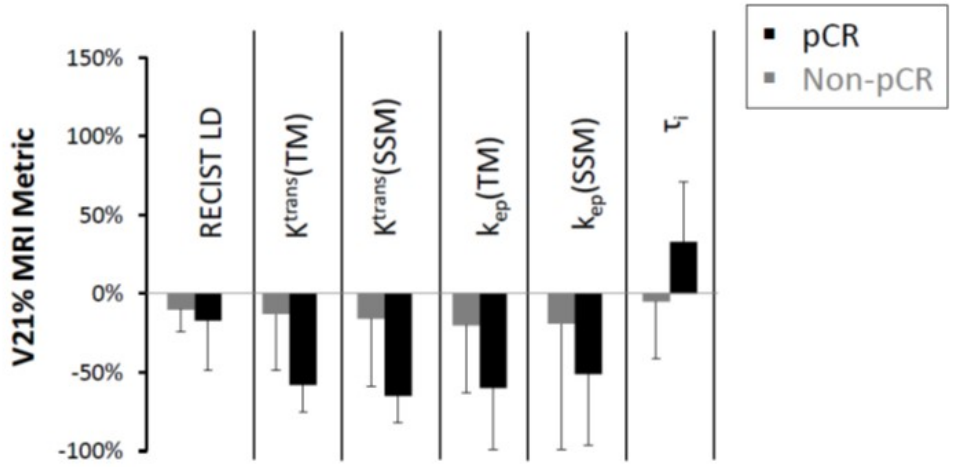


Figure 4: Column graphs of the mean V21% change values of RECIST LD and several DCE-MRI metrics (K^{trans}, k_{ep}, and τ_i, estimated from the TM and SSM pharmacokinetic analyses) for the pCR (black column) and non-pCR (gray column) patient groups. The error bar represents the standard deviation (SD). V21%: percent change of MRI metric at visit 2 (V2, after one NACT cycle) relative to visit 1 (V1, pre-NACT).

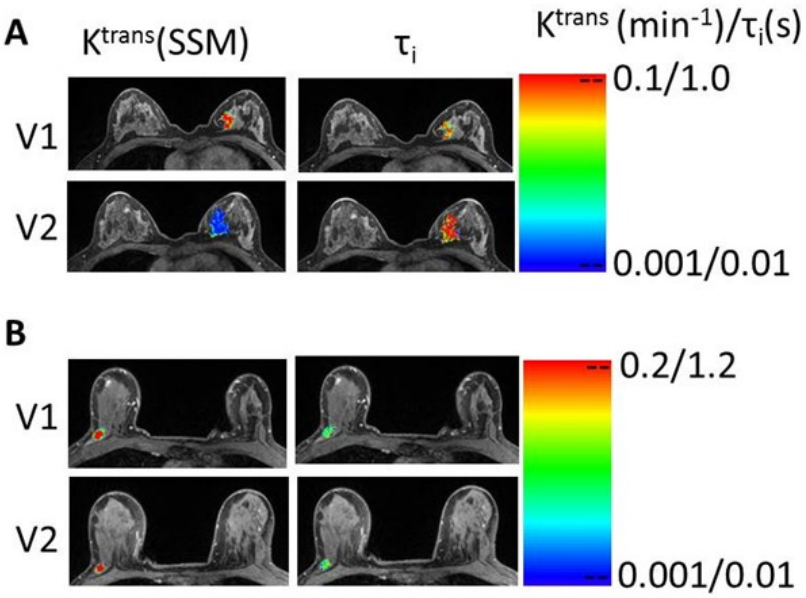


Figure 5: V1 (pre-NACT) and V2 (after one NACT cycle) color parametric maps of K^{trans}(SSM) and τ_i from a pCR (A) and a non-pCR (B) breast tumor. For each tumor, the color scale of each DCE-MRI metric is kept the same between the two visits for easy visualization of NACT-induced changes. There are no noticeable changes in the parametric maps from V1 to V2 for the non-pCR, while the substantial decrease in K^{trans}(SSM) and increase in τ_i are clearly visible for the pCR.

§ Specific Aim 2: Investigate the effects of data acquisition and processing schemes on DCE-MRI biomarkers within the context of assessing therapy response.

Effects of Temporal Resolution on DCE-MRI Prediction of Breast Cancer Response to Therapy

Pharmacokinetic (PK) analysis of high temporal resolution (tRes) DCE-MRI data has been shown effective for early prediction of breast cancer response to NACT. However, high tRes breast DCE-MRI studies are currently limited to research and early phase clinical trial settings. Due to the trade-off of tRes and spatial resolution (sRes) in data acquisition and clinical needs for bilateral full breast coverage and high sRes, low tRes (60-120 s) breast DCE-MRI protocols are commonly used in large-scale clinical trials and clinical practice. Consequently, because of inaccuracies in PK parameter estimation from low tRes data [29, 30], semi-quantitative analysis (such as uptake slope, etc.) is often employed for low tRes data. Unlike quantitative PK parameters (such as K^{trans}) which are direct measures of biological properties, semi-quantitative metrics are directly related to MR signal change, not tissue biology, and the values are often dependent on data acquisition protocols and scanner platforms and settings, making it difficult to compare studies across institutions. There has been no literature evidence on whether PK analysis of low tRes data can still provide useful early prediction of breast cancer therapy response despite expected PK parameter errors. Here

we report our initial results comparing PK analyses of low and high tRes breast DCE-MRI data for early prediction of NACT response, using data sets from the same patient cohort.

Methods: 15 breast cancer patients enrolled in a multicenter ISPY-2 NACT trial consented to high tRes (14-18 s) research DCE-MRI (2) at visit 1 (V1, before NACT), V2 (after 1 NACT cycle), V3 (at NACT midpoint), and V4 (after NACT). They also underwent a low tRes (80-100 s) ISPY-2 DCE-MRI protocol at the same four time points. PK analyses of the low and high tRes DCE-MRI data were performed using the Shutter-Speed model (SSM) which takes into account transcytolemmal water exchange kinetics. Tumor mean PK parameter values were calculated by averaging tumor voxel parameter values from all slices covering the tumor, which included K^{trans}, v_e, k_{ep} ($=K^{trans}/v_e$), and the SSM-unique τ_i parameter, mean intracellular water lifetime.

Estimated PK parameters from the low and high tRes data at V1 and V2, and the percent changes (V21%, V2 relative to V1) were compared, and correlated with pathologic response status (determined from resection specimens after NACT) to assess abilities for early prediction of response through ROC analysis. A nonparametric method was used to compare ROC AUC between results from the two tRes data sets.

Results and Discussion: Following NACT, 4 patients had pathologic complete response (pCR) while the other 11 had non-pCR. Table 7 lists tumor mean ± SD values of V1, V2, and V21% PK parameters estimated from the high and low tRes data, showing statistically significant underestimations of K^{trans}, k_{ep}, and τ_i and overestimation of v_e from the low tRes data compared to the high tRes data. However, there were no significant differences in V21% values of these parameters. For example, Figure 6 shows scatter plots of V2 and V21% K^{trans} from the high and low tRes data. Table 8 lists the ROC AUC values of several DCE-MRI metrics for early discrimination of pCR vs. non-pCR. For each metric there was no statistically significant difference in ROC AUC between the two tRes data sets.

The findings of K^{trans} underestimation and v_e overestimation from the low tRes data are consistent with a previous study [29]. The errors in PK analysis of low tRes data are largely systematic with PK parameter values changing in the same direction going from low to high tRes. This is why there are no significant differences in V21% values, and the likely reason that DCE-MRI metrics that are good early predictors of NACT response when obtained from the high tRes data perform comparably well in early prediction when obtained from the low tRes data (Table 8). This preliminary study suggests that despite expected errors in estimated PK parameters, PK analysis of low tRes DCE-MRI data could be useful for assessment of breast cancer therapy response. Since low tRes data is usually collected in large-scale breast cancer clinical trials, the utility of PK analysis of low tRes data for therapy response evaluation may have significant impact on future imaging biomarker development, taking advantage of the large, retrospective database from the past and current trials that include breast DCE-MRI.

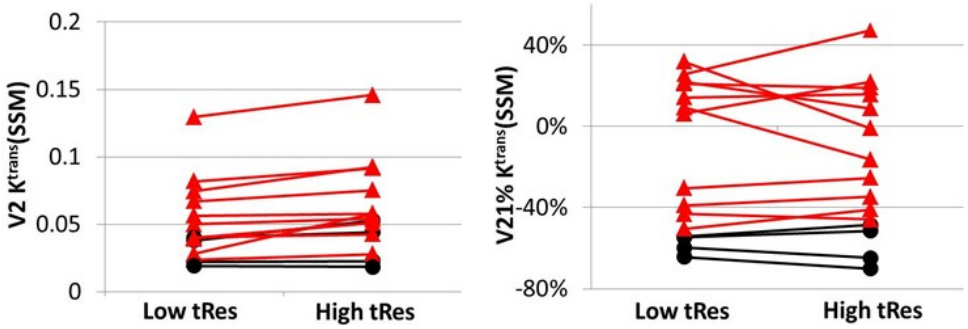


Figure 6: Scatter plots of V2 K^{trans} (left) and V21% K^{trans} (right) estimated from the low and high tRes data. The straight line connects data points from the same subject. pCRs are represented by black circles while non-pCRs by red triangles.

Table 7. Effects of tRes on Estimated DCE-MRI PK Parameters

tRes	V1				V2				V21 (%)			
	K^{trans}	v_e	k_{ep}	τ_i	K^{trans}	v_e	k_{ep}	τ_i	K^{trans}	v_e	k_{ep}	τ_i
Low	0.066 ± 0.036*	0.34 ± 0.12*	0.25 ± 0.11*	0.59 ± 0.26#	0.050 ± 0.028*	0.52 ± 0.24*	0.16 ± 0.11*	0.50 ± 0.21*	-14 ± 43$	69 ± 85$	-30 ± 53&	4 ± 73&
High	0.076 ± 0.038	0.23 ± 0.07	0.35 ± 0.11	0.73 ± 0.27	0.059 ± 0.032	0.38 ± 0.21	0.25 ± 0.16	0.67 ± 0.24	-15 ± 43	68 ± 99	-25 ± 54	-4 ± 30

t test for low *vs.* high tRes: *, $P < 0.001$; #, $P < 0.01$; $, $P > 0.9$; &, $P > 0.5$.

Table 7: Effects of tRes on estimated DCE-MRI PK parameters.

Table 8. Early Prediction of Response		
MRI Metrics	**ROC AUC**	
	Low tRes	**High tRes**
V21% K^{trans}	1.00	1.00
V21% τ_i	1.00	1.00
V21% k_{ep}	0.93	0.97
V21% v_e	0.93	0.93
V2 k_{ep}	0.93	0.97
V2 K^{trans}	0.91	0.93
V1 τ_i	0.91	0.88

Table 8: Early prediction of response.

§ Specific Aim 3: Develop software tools that can provide clinicians with imaging metrics together with clinical and molecular biomarkers to aid clinical decision-making in evaluation of therapy response

The OHSU Informatics group (Shannon McWeeney, PhD, Jayashree Kalpathy-Cramer, PhD, Fred Loney, MS, Lara Fournier, MS and Erik Segerdell, BS) continued work on several tasks pertaining to Aim 3. The activities completed in Year 5 include the following:

- Add web application database update to the image data analysis pipeline
- Improve imaging pipeline features and scalability
- Add ROI and modeling overlays to the Quantitative Imaging Profile (QuIP) image display (Fig.ure 7)
- Investigate alternative image registration methods
- Deploy a production QuIP web server instance
- Review the QuIP application with imaging scientists and clinicians
- Added 16 DICOM studies to TCIA

Open Source Tool Utilization

OHSU is utilizing the following open source tools in the Quantitative Imagine Pipeline (Figure 7). XNAT imaging repository platform developed at Washington University; Nipype, a Python workflow integration framework; ANTS, a diffeomorphic registration and image mapping toolkit; CTP, the Washington University TCIA image uptake utility; XTK, an image visualization module; and NVD3, a charting utility.

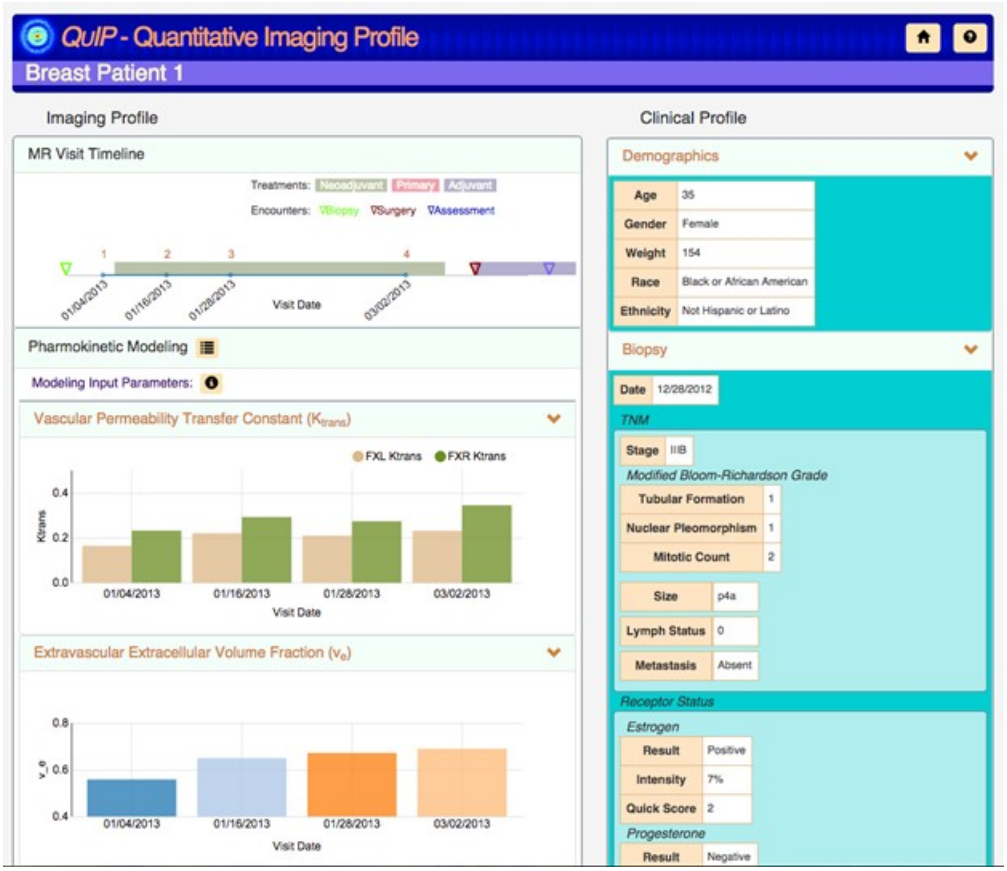


Figure 7: Quantitative Imaging Profile (QuIP) web application.

Collaborations within the QIN Network

1. The OHSU team participated in a multi-center project investigating gradient nonlinearity bias in measurement of apparent diffusion coefficient across MRI scanner platforms. As a collaborative study of the QIN Data Acquisition Working Group (DAWG), this project is headed by Dr. Thomas Chenevert of University of Michigan. The results of this study have been published in *Magnetic Resonance in Medicine* in 2016 and *Tomography* in 2016.

2. The OHSU PI (Dr. Huang) has initiated a multicenter AIF challenge project within the MRI WG to evaluate the effects of variations in AIF determination on estimated DCE-MRI pharmacokinetic parameters, as well as on therapy response assessment. A total of 9 QIN centers participated in this challenge and quantified AIFs with site-specific methods from shared pre-therapy prostate and pre- and post-therapy soft-tissue sarcoma DCE-MRI data. The AIFs were submitted by each center to the managing center OHSU. Dr. Huang and his team then analyzed the shared data with submitted AIFs (from multi-centers) and fixed pharmacokinetic model, pre-contrast T1, and tumor ROI definition to assess the effects of AIF variations only. The analysis of the shared eleven prostate data sets using the standard Tofts model has been completed and the manuscript has been published in *Tomography* in 2016. The same data was also analyzed by OHSU using the Shutter-Speed model and the results were submitted as an abstract to the 2017 annual ISMRM meeting. The analysis of AIF variation on a longitudinal sarcoma therapy response study has been completed and the results were submitted as an abstract to the 2017 annual ISMRM meeting. We expect to submit the manuscript on the effects of AIF variation on DCE-MRI evaluation of soft tissue sarcoma response to preoperative chemoradiotherapy in mid-2017.

3. The OHSU team participated in the T1 measurement collaborative project in the DAWG, led by Dr. Bachir of Mount Sinai. The goal of the challenge is to assess reproducibility across scanner platform in T1 measurement using commonly used pre-contrast T1 determination methods for DCE-MRI, such as the multi-flip angle method and the inversion-recovery method. A NIST phantom is used for challenge with known T1 values for the solutions included in the phantom. Data analysis has been completed and a 2017 ISMRM abstract has been submitted. The manuscript is expected to be submitted in mid-2017.

4. The OHSU team participated in the ADC mapping challenge of the MRI WG organized by Dr. Newitt of UCSF. The goal is to evaluate the concordance of ADC maps generated with scanner manufacturers' software and tools used by QIN centers. The initial data analysis has been completed by Dr. Newitt and an 2017 ISMRM abstract was submitted.

PLANS FOR NEXT YEAR

For Specific Aim 1, we will continue breast cancer and soft-tissue sarcoma patient accrual, with no modification in the research protocols. The primary focus will be on MRI data collection, analysis, correlation with pathologic endpoints, and statistical analysis to look for the best imaging biomarker or combination of biomarkers for cancer therapy response. For Specific Aim 2, we have evaluated the effect of DCE-MRI data acquisition duration, AIF quantification, as well as temporal resolution on evaluation of breast cancer response to NACT. We plan to submit a manuscript on these results in 2017. Additionally, we will perform similar analyses using the soft tissue sarcoma data and determine if we can draw similar conclusions with regard to the effects of variations in DCE-MRI data acquisition and analysis within the context of therapeutic monitoring. For the bioinformatics aim, the Specific

Aim 3, we will continue working on Aim 3.2 to integrate imaging and non-imaging biomarkers in the web-based informatics tool, QuIP, for cancer therapy response evaluation, and test the tool with clinicians (radiologists and surgical and medical oncologists) for its utility for clinical decision making.

Collaborations within QIN Network

The OHSU team will be participating in or initiating the following multi-center challenge projects:

1. DCE-MRI AIF challenge phase II (MRI WG, Leader: Wei Huang)
2. DSC-MRI DRO (digital reference object) challenge (MRI WG, Leader: Chad Quarles)
3. Effects of k-space under-sampling on quantitative DCE-MRI analysis (DAWG and MRI WG, Leader: Wei Huang)

LIST OF REFERENCES

1. Johansen R, Jensen LR, Rydland J, Goa PE, Kvistad KA, Bathen TF, Axelson DE, Lundgren S, Gribbestad IS. Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using DCE-MRI. J Magn Reson Imaging 2009;29:1300-1307.
2. Viglianti BL, Lora-Michiels M, Poulson JM, Lan L, Yu D, Sanders L, Craciunescu O, Vujaskovic Z, Thrall DE, Macfall J, Charles CH, Wong T, Dewhirst MW. Dynamic contrast-enhanced magnetic resonance imaging as a predictor of clinical outcome in canine spontaneous soft tissue sarcomas treated with thermoradiotherapy. Clin Cancer Res 2009;15:4993-5001. [PMCID: PMC2763531]
3. Walker-Samuel S, Leach MO, Collins DJ. Evaluation of response to treatment using DCE-MRI: the relationship between initial area under the gadolinium curve (IAUGC) and quantitative pharmacokinetic analysis. Phys Med Bio 2006;51:3593-3602.
4. Thoeny HC, De Keyzer F, Vandecaveye V, Chen F, Sun X, Bosmans H, Hermans R, Verbeken EK, Boesch C, Marchal G, Landuyt W, Ni Y. Effect of vascular targeting agent in rat tumor model: dynamic contrast-enhanced versus diffusion-weighted MR imaging. Radiology 2005;237:492-499.
5. van Rijswijk CS, Geirnaerdt MJ, Hogendoorn PC, Peterse JL, van Coevorden F, Taminiau AH, Tollenaar RA, Kroon BB, Bloem JL. Dynamic contrast-enhanced MR imaging in monitoring response to isolated limb perfusion in high-grade soft tissue sarcoma: initial results. Eur Radiol 2003;13:1849-1858.
6. Egmont-Petersen M, Hogendoorn PC, van der Geest RJ, Vrooman HA, van der Woude H, Janssen JP, Bloem JL, Reiber JH. Detection of areas with viable remnant tumor in postchemotherapy patients with Ewing's sarcoma by dynamic contrast-enhanced MRI using pharmacokinetic modeling. Magn Reson Imaging 2000;18:525-535.
7. Martincich L, Montemurro F, De Rosa G, Marra V, Ponzone R, Cirillo S, Gatti M, Biglia N, Sarotto I, Sismondi P, Regge D, Aglietta M. Monitoring response to primary chemotherapy in breast cancer using dynamic contrast-enhanced magnetic resonance imaging. Breast Cancer Res Treat 2004;83:67-76.

8. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp, MV, Larsson HB, Lee, TY, Mayr NA, Parker GJ, Port RE, Taylor J, Weisskoff RM. Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging 1999; 10: 223-232.
9. Huang W, Li X, Morris EA, Tudorica LA, Seshan VE, Rooney WD, Tagge I, Wang Y, Xu J, Springer CS. The magnetic resonance shutter speed discriminates vascular properties of malignant and benign breast tumors *in vivo.* Proc Natl Acad Sci 2008; 105: 17943-17948. [PMCID: PMC2582583]
10. Li X, Huang W, Morris EA, Tudorica LA, Seshan VE, Rooney WD, Tagge I, Wang Y, Xu J, Springer CS. Dynamic NMR effects in breast cancer dynamic-contrast-enhanced MRI. Proc Natl Acad Sci 2008; 105: 17937-17942. [PMCID: PMC2582941]
11. Huang W, Tudorica LA, Li X, Thakur SB, Chen Y, Morris EA, Tagge IJ, Korenblit M, Rooney WD, Koutcher JA, Springer CS. Discrimination of benign and malignant breast lesions by using shutter-speed dynamic contrast-enhanced MR imaging. Radiology 2011;261:394-403.
12. Zhou R, Pickup S, Yankeelov TE, Springer CS, Glickson JD. Simultaneous measurement of arterial input function and tumor pharmacokinetics in mice by dynamic contrast enhanced imaging: effects of transcytolemmal water exchange. Mag Reson Med 2004; 52:258-257.
13. Yankeelov TE, Rooney WD, Li X, Springer CS. Variation of the relaxographic "Shutter-Speed" for transcytolemmal water exchange affects the CR bolus-tracking curve shape. Magn Reson Med 2003; 50: 1151-1169.
14. Li X, Rooney WD, Springer CS. A unified pharmacokinetic theory for intravascular and extracellular contrast agents. Magn Reson Med 2005; 54: 1351-1359. [Erratum. Magn Reson Med 2006; 55:1217.]
15. Springer CS, Tudorica LA, Li X, Thakur S, Morris EA, Oh KY, Kettler MD, Chen Y, Tagge IJ, Hemmingson SL, Korenblit M, Grinstead JW, Laub G, Koutcher JA, Huang W. Meta-population breast cancer screening with the ΔK^{trans} DCE-MRI parameter. Proc Intl Soc Magn Reson Med 2011;19: 3097.
16. Li X, Priest RA, Woodward WJ, Tagge IJ, Siddiqui F, Huang W, Rooney WD, Beer TM, Garzotto MG, Springer CS. Feasibility of Shutter-Speed DCE-MRI for improved prostate cancer detection. Magn Reson Med 2012. doi: 10.1002/mrm.24211. [Epub ahead of print]
17. Yang C, Karczmar GS, Medved M, Oto A, Zamora M, Stadler WM. Reproducibility assessment of a multiple reference tissue method for quantitative dynamic contrast enhanced-MRI analysis. Magn Reson Med 2009;61:851-859. [PMCID: PMC2758034]
18. Yankeelov TE, Gore JC. Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples. Curr Med Imaging Rev 2009;3:91-107. [PMCID: PMC2760951]
19. Meyer JM, Perlewitz KS, Hayden JB, Doung YC, Hung AY, Vetto JT, Pommier RF, Mansoor A, Beckett BR, Tudorica A, Mori M, Holtorf ML, Afzal A, Woodward WJ, Rodler ET, Jones RL, Huang W, Ryan CW. Phase I trial of preoperative chemoradiation plus Sorafenib for high risk extremity soft tissue sarcoma with dynamic contrast-enhanced magnetic resonance imaging correlates. Clin Cancer Res 2013;19:6902-6911.
20. Huang W, Wang Y, Panicek DM, Schwartz LH, Koutcher JA. Feasibility of using limited-population-based average R_{10} for pharmacokinetic modeling of osteosarcoma dynamic contrast-enhanced MRI data. Magn Reson Imaging 2009;27:852-858.

21. Tudorica A, Oh KY, Chui SYC, Roy N, Troxell ML, Naik A, Kemmer K, Chen Y, Holtorf ML, Afzal A, Springer CS, Li X, Huang W. Early Prediction and Evaluation of Breast Cancer Response to Neoadjuvant Chemotherapy Using Quantitative DCE-MRI. Transl Oncol 2016;9:8-17.
22. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, Gwyther SG. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, Nation Cancer Institute of Canada. J Natl Cancer Inst 2000;92:205-216.
23. Lu H, Clingman C, Golay X, van Zijl PC. Determining the longitudinal relaxation time (T1) of blood at 3.0 Tesla. Magn Reson Med 2004;52:679-682.
24. Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, O'Sullivan F, Abramson RG, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Besa C, Kalpathy-Cramer J, Mountz JM, Laymon CM, Muzi M, Kinahan PE, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy FMM, Li X. The impact of arterial input function determination variations on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: a multicenter data analysis challenge. Tomography 2016;2:56-66.
25. Li X, Kang H, Arlinghaus LR, Abramson RG, Chakravarthy AB, Abramson VG, et al. Analyzing spatial heterogeneity in DCE- and DW-MRI parametric maps to optimize prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Transl Oncol 2014;7:14-22.
26. Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol 2007;25:4414-4422.
27. Zhang Y, Poirier-Quinot M, Springer CS, Balschi JA. Active trans-plasma membrane water cycling in yeast is revealed by NMR. Biophys J 2011;101:2833-2842.
28. Springer CS, Li X, Tudorica LA, Oh KY, Roy N, Chui SYC, et al. Intratumor mapping of intracellular water lifetime: metabolic images of breast cancer? NMR Biomed 2014;27:760-773.
29. Heisen M, Fan X, Buurman J, van Riel NAW, Karczmar GS, ter Haar Romeny BM. The influence of temporal resolution in determining pharmacokinetic parameters from DCE-MRI data. Magn Reson Med 2010;63:811-816.
30. Di Giovanni P, Azlan CA, Ahearn TS, Semple SI, Gilbert FJ, Redpath TW. The accuracy of pharmacokinetic parameter measurement in DCE-MRI of the breast at 3T. Phys Med Biol 2010;55:121-132.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

§ Peer-Reviewed Papers

1. Thibault G, Tudorica A, Afzal A, Chui SYC, Naik A, Troxell ML, Kemmer KA, Oh KY, Roy N, Holtorf ML, Huang W, Song X. Early prediction of breast cancer therapy response to neoadjuvant chemotherapy through texture analysis of DCE-MRI. Proc MICCAI-BIA 2015; pp 145-152.
2. Yankeelov TE, Mankoff DA, Schwartz LH, Lieberman FS, Buatti JM, Mountz JM, Erickson BJ, Fennessy FMM, Huang W, Kalpathy-Cramer J, Wahl RL, Linden HM,

Kinahan PE, Zhao B, Hylton NM, Gillies RJ, Clarke L, Nordstrom R, Rubin DL. Quantitative imaging in cancer clinical trials. Clin Cancer Res 2016;22:284-290. PMID: 26773162, PMCID: PMC4717912

3. Malyarenko DI, Newitt D, Wilmes LJ, Tudorica A, Helmer KG, Arlinghaus LR, Jacobs MA, Jajamovich G, Taouli B, Yankeelov TE, Huang W, Chenevert TL. Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials. Magn Reson Med 2016;75:1312-1323. PMID: 25940607, PMCID: PMC4630210
4. Jajamovich GH, Huang W, Besa C, Li X, Afzal A, Dyvorne HA, Taouli B. DCE-MRI of hepatocellular carcinoma: perfusion quantification with Tofts model versus shutter-speed model - initial experience. Magn Reson Mater Phy Biol Med 2016;29:49-58. PMID: 26646522
5. Tudorica A, Oh KY, Chui SYC, Roy N, Troxell ML, Naik A, Kemmer K, Chen Y, Holtorf ML, Afzal A, Springer CS, Li X, Huang W. Early Prediction and Evaluation of Breast Cancer Response to Neoadjuvant Chemotherapy Using Quantitative DCE-MRI. Transl Oncol 2016;9:8-17. PMID: 26947876, PMCID: PMC 4800060
6. Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, O'Sullivan F, Abramson RG, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Besa C, Kalpathy-Cramer J, Mountz JM, Laymon CM, Muzi M, Kinahan PE, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy FMM, Li X. The impact of arterial input function determination variations on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: a multicenter data analysis challenge. Tomography 2016;2:56-66. PMID: 27200418, PMCID: PMC4869732
7. Li X, Cai Y, Moloney B, Chen Y, Huang W, Woods M, Coakley FV, Rooney WD, Garzotto MG, Springer CS. Relative sensitivities of DCE-MRI pharmacokinetic parameters to arterial input function (AIF) scaling. J Magn Reson 2016;269:104-112. doi: 10.1016/j.jmr.2016.05.018
8. Huang W, Beckett BR, Tudorica A, Meyer JM, Afzal A, Chen Y, Mansoor A, Hayden JB, Doung YC, Hung AY, Holtorf ML, Aston TJ, Ryan CW. Evaluation of soft tissue sarcoma response to preoperative chemoradiotherapy using dynamic contrast-enhanced magnetic resonance imaging. Tomography 2016;2:308-316 DOI: 10.18383/j.tom.2016.00202
9. Malyarenko DI, Wilmes LJ, Arlinghaus LR, Jacobs MA, Huang W, Helmer KG, Taouli B, Yankeelov TE, Newitt D, Chenevert TL. QIN DAWG validation of gradient nonlinearity bias correction workflow for quantitative diffusion-weighted imaging in multicenter trials. Tomography 2016;2:396-405.
10. Hectors SJ, Besa C, Wagner M, Jajamovich GH, Haines GK, Tewari A, Huang W, Taouli B. DCE-MRI of the prostate using shutter-speed vs. Tofts model for tumor characterization and assessment of aggressiveness. J Magn Reson Imaging, in press.
11. Thibault G, Tudorica A, Afzal A, Chui SYC, Naik A, Troxell ML, Kemmer KA, Oh KY, Roy N, Holtorf ML, Huang W, Song X. DCE-MRI Texture Features for Early Prediction of Breast Cancer Response. Tomography, in press.

§ Peer-Reviewed Abstracts and Meeting Presentations

1. Thibault G, Tudorica A, Afzal A, Chui SYC, Naik A, Troxell ML, Kemmer KA, Oh KY, Roy N, Holtorf ML, Huang W, Song X. Texture feature analysis of quantitative and semi-quantitative DCE-MRI metrics for early prediction of breast cancer therapy response. Proc Intl Soc Magn Reson Med 2016;24:127.
2. Le Y, Afzal A, Chen X, Spottiswoode B, Huang W, Lin C. A prototype image quality assurance system for accelerated quantitative breast DCE-MRI. Proc Intl Soc Magn Reson Med 2016;24:189.
3. Huang W, Chen Y, Fedorov A, Li X, Jajamovich G, Malyarenko DI, Aryal M, LaViolette PS, Oborski MJ, O'Sullivan F, Abramson RG, Muzi M, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Besa C, Kalpathy-Cramer J, Mountz JM, Laymon CM, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy F, Li X. The impact of arterial input function determination variation on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: a multicenter data analysis challenge. Proc Intl Soc Magn Reson Med 2016;24:580.
4. Huang W, Tudorica A, Oh KY, Chui SYC, Roy N, Troxell ML, Naik A, Kemmer KA, Chen Y, Holtorf ML, Afzal A, Li X, Springer CS. Metabolic imaging biomarker k_{io} discriminates breast tumor therapy time-courses. Proc Intl Soc Magn Reson Med 2016;24:2653.
5. Huang W, Afzal A, Tudorica A, Chen Y, Chui SYC, Naik A, Troxell M, Kemmer K, Oh KY, Roy N, Holtorf ML, Li X. Effects of temporal resolution on quantitative DCE-MRI prediction of breast cancer therapy response. Proc Intl Soc Magn Reson Med 2016;24:2792.
6. Tudorica A, Oh KY, Chui SYC, Roy N, Troxell ML, Naik A, Kemmer K, Chen Y, Holtorf ML, Afzal A, Springer CS, Li X, Huang W. Early prediction and evaluation of breast cancer response to neoadjuvant chemotherapy using quantitative DCE-MRI. Proc Intl Soc Magn Reson Med 2016;24:2638.
7. Li X, Garzotto MG, Coakley FV, Moloney B, Woodward WJ, Chen Y, Huang W, Rooney WD, Springer CS. DCE-MRI high-resolution metabolic prostate imaging is insensitive to AIF uncertainty. Proc Intl Soc Magn Reson Med 2016;24:447.
8. Li K, Mohamed ASR, Ding Y, Awan MJ, Frank SJ, Wang J, Hazle JD, Hutcheson K, Lai SY, Kalpathy-Cramer J, Li X, Fuller CD, Huang W. DCE-MRI evidence of biological changes in irradiated healthy muscle following chemoradiotherapy treatment of head and neck cancer. Proc Intl Soc Magn Reson Med 2016;24:2775.
9. Springer CS, Li X, Tudorica A, Oh KY, Chui SYC, Roy N, Troxell ML, Naik A, Kemmer KA, Chen Y, Holtorf ML, Afzal A, Huang W. Breast tumor metabolic imaging biomarker k_{io} correlates with histopathology residual cancer measures. Proc Intl Soc Magn Reson Med 2016;24:2722.
10. Huang W. Characterization of tumors with DCE-MRI. Multiparametric Assessment of Cancer; Sunrise Lecture. Proc Intl Soc Magn Reson Med 2016;24.

U01 CA151235: Quantitative Imaging for Assessing Breast Cancer Response to Treatment

University of California, San Francisco

Nola Hylton, Ph.D.;

David Newitt, Ph.D.; Ella Jones, Ph.D.; Lisa Wilmes, Ph.D.;

Wen Li, Ph.D.; Laura Esserman, MD

INTRODUCTION

Our research program continues to focus on the development and clinical implementation of quantitative breast MRI for assessing response to treatment. Neoadjuvant chemotherapy (NAC) for breast cancer, in which systemic therapy is administered prior to surgery, has important benefits for patients, including down-staging inoperable cancers and improving breast conservation rates [1, 2]. It is now well established that women whose tumor is completely eradicated at the time of surgery (pathologic complete response, or pCR) have excellent survival rates [3-5]. Conversely, women with substantial residual disease at surgery, have much poorer outcomes, with recurrence rates of over 50% at 5 years[6]. The problem of identifying ineffective treatment remains one of the most critical unmet needs in neoadjuvant chemotherapy for locally advanced breast cancer. We are developing quantitative imaging methods to assess response to treatment in the I-SPY 2 TRIAL, a multi-center phase II treatment trial using response-adaptive randomization within breast cancer subtypes to evaluate investigational agents for women with high-risk stage II/III breast cancer [7-9]. The goal of our QIN research is to develop accurate and reliable breast imaging techniques that can be applied as diagnostic tools for individualizing patient treatment. Over the past year our QIN efforts have focused on 1) developing advanced DWI approaches for breast cancer evaluation for introduction into I-SPY 2, 2) beginning dedicated breast PET studies in I-SPY 2, 3) providing curated image data and outcomes from I-SPY 1 for public data-sharing and 4) leading a QIN Grand Challenge to identify high performing imaging biomarkers of response.

PROGRESS OVER THE PRIOR YEAR

§ Advanced diffusion-weighted MRI (DWI) methods for assessing breast tumor response to neoadjuvant chemotherapy

High spatial resolution breast DWI

We previously developed a high-spatial resolution reduced field of view (rFOV) DWI sequence to overcome limitations of standard DWI methods for evaluating the breast. We continue to evaluate this method in comparison to standard DWI for measuring response to treatment. The rFOV sequence utilizes a 2D spatially-selective echo-planar RF excitation

pulse and a 180-degree refocusing pulse to reduce the FOV in the phase-encode (PE) direction [10], resulting in improved spatial resolution and reduced off-resonance effects. We optimized the rFOV HR-DWI sequence for breast imaging to acquire data with voxel sizes 5-6 times smaller than standard commercially available single shot echo-planar imaging (ssEPI) DWI sequences (STD-DWI) while retaining sufficient SNR for accurate calculation of ADCs. We showed that the sequence improved image quality compared to standard ssEPI DWI in breast, as seen in Figure 1 showing a comparison of breast images acquired with standard (STD) and high resolution (HR) DWI in a patient with invasive breast carcinoma [11]. Studies comparing STD-DWI and HR-DWI in breast cancer patients undergoing NAC demonstrated that the lower tumor ADC percentile from HR-DWI (15^{th} and 25^{th} percentile) derived tumor ADC histograms have stronger association with final MRI-measured tumor volume change than that from STD-DWI [12]. We also found an association between early changes in tumor ADC metrics measured by HR-DWI and pathologic complete response (pCR). An increasing trend in the area under the receiver operating curve (AUC) for predicting pCR was found with decreasing ADC percentile. Additionally, AUCs for the lower percentile tumor ADC were higher than those for early functional tumor volume (FTV) change (Table 1). Our findings here are consistent with a previous study showing the sensitivity of lower percentile tumor ADC values to early treatment changes [12]. The higher AUCs found for ADC metrics versus FTV suggest that HR-DWI may be of value in evaluating early breast tumor response to neoadjuvant chemotherapy and support the investigation of this technique in a larger cohort.

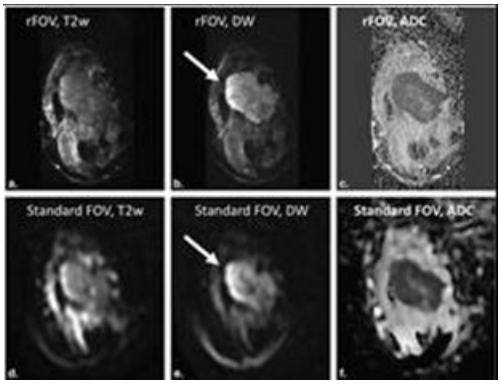


Figure 1: Representative DWI (b=0 and b=600) and ADC maps of an invasive breast carcinoma acquired with HR-DWI (top row) with voxel size of 4.8 mm^3 compared to STD-DWI (bottom row) with voxel size of 29 mm^3. It shows improved image quality and reduced distortion. The tumor is visible as a hypointense region in the center of the breast on the ADC maps.

Early Treatment Change	full cohort (n=20)
ADC Predictors	**AUC**
% Change Mean ADC	0.61
% Change 5th percentile	0.67
% Change 15th percentile	0.67
% Change 25th percentile	0.65
% Change 50th percentile	0.65
% Change 75th percentile	0.59
% Change 95th percentile	0.57
MRI Tumor Volume Predictor	**AUC**
% Chng Tumor Volume	0.59

Table 1: AUCs for the early percent change in HR-DWI ADC and tumor volume as predictors of pCR showing an increasing trend with decreasing ADC percentile.

Diffusion-tensor imaging (DTI) of the breast

We are also evaluating diffusion-tensor imaging (DTI) as a further refinement of DWI methods that may carry additional prognostic information for breast cancer response. DTI and contrast-enhanced MRI were acquired at 1.5 T in 34 patients before treatment and after 3 cycles of taxane-based therapy (early treatment). Tumor fractional anisotropy (FA), principal eigenvalues ($\lambda 1$, $\lambda 2$, and $\lambda 3$), and apparent diffusion coefficient (ADC) were estimated for tumor regions of interest drawn on DTI data. The association between DTI metrics and final tumor volume change was evaluated with Spearman rank correlation. DTI metrics were investigated as predictors of pathological complete response (pCR) by calculating the AUC. Early changes in tumor FA and ADC significantly correlated with final tumor volume change post therapy (ρ = -0.38, p = 0.03 and ρ = -0.71, $p < .001$, respectively). Pretreatment tumor ADC was significantly lower in the pCR than in the non-pCR group (p = 0.04). At early treatment, patients with pCR had significantly higher percent changes of tumor eigenvalues $\lambda 1$, $\lambda 2$, $\lambda 3$, and ADC than those without pCR. The AUCs for early percent changes in tumor FA and ADC were 0.60 and 0.83, respectively. The early percent changes in tumor eigenvalues and ADC were the strongest DTI-derived predictors of pCR. Although early percent change in tumor FA had a weak association with pCR, the significant correlation with final tumor volume change suggests that this metric changes with therapy and may merit further evaluation. These results were recently published in the QIN Special Issue of the journal Tomography [13].

§ Breast MRI predictors of NAC response

Optimization of breast MRI biomarkers by cancer subtype

Under its primary aim, ACRIN 6657, the imaging component of I-SPY 1, prospectively tested the functional tumor volume (FTV) biomarker and found it to be highly predictive of both pathologic response and recurrence-free survival following NAC [14, 15]. FTV is

defined as the image volume with enhancement kinetics exceeding both an early percentage enhancement threshold (PEt) and a signal enhancement ratio threshold (SERt). Primary study analysis used empirically determined values for these thresholds. In subsequent studies, we examined the effect of varying PEt and SERt on prediction of pCR, to determine if optimization of these parameter thresholds can improve predictive performance [16]. We also hypothesized that predictive performance varied by cancer subtype and therefore independent optimization within subtype groups would result in the greatest improvement. The ACRIN 6657 cohort included women with locally advanced breast cancer (tumor size ≥ 3cm) having up to four DCE-MRI examinations: before NAC (MR1), after one cycle of NAC (MR2), between the anthracycline-based regimen and taxane (MR3), and after NAC and prior to surgery (MR4). Patients were stratified into 3 groups by cancer subtypes defined by hormone receptor (HR), and human epidermal growth factor receptor 2 (HER2) status: HR+/HER2-, HER2+, and triple negative (TN, HR-/HER2-). MRI-measured FTV and change in FTV (ΔFTV) were investigated as predictors of the outcome pCR.

For our optimization study PEt was varied from 30% to 200% in 10% intervals, and SERt was varied from 0.0 to 2.0 in 0.2 unit intervals. FTV was measured at each examination (FTV_1, FTV_2, FTV_3, FTV_4) and ΔFTV was measured relative to the first examination (ΔFTV_2, ΔFTV_3, ΔFTV_4). For each pair of thresholds (PEt, SERt), the absolute FTV and ΔFTVs were calculated and analyzed for prediction of pCR using AUCs. 116 patients from the ACRIN 6657 / I-SPY 1 TRIAL with complete data on all four MRI visits, HR/HER2 status, and pCR outcome were included. Mean age was 48 (range 29-69). The 116 patient cohort was divided into subgroups: 45 (39%) HR+/HER2-; 39 (34%) HER2+; and 30 (26%) TN. Lower AUCs with less variation were observed in patients in the HER2+ subgroup than patients with HR+/HER2- and TN breast cancer. When examining prediction by visit, maximum AUCs were found at later time points in all patient cohorts. Specifically, maximum AUC was observed for: the full cohort at ΔFTV_3 with AUC of 0.78 (CI: 0.69 – 0.87) at (PEt, SERt) = (130%, 0.0); the HR+/HER2- subtype at ΔFTV_3 with AUC of 0.9 (CI: 0.84 – 0.97) also at (PEt, SERt) = (130%, 0.0); the HER2+ subtype at FTV_3 with AUC of 0.77 (CI: 0.62 – 0.92) when (PEt, SERt) = (70%, 1.4); and the TN at FTV_4 with AUC of 0.89 (CI: 0.76 – 1) with (PEt, SERt) = (40%, 2.0). The analysis suggests that MRI thresholds need to be adjusted by breast cancer subtype to improve the predictive performance. The PEt may need to be set higher in HR+/HER2- than other subtypes, which may be due to higher background parenchymal enhancement among HR+ patients, and SERt may need to be set at higher level for TN subtype. These data were recently published in Tomography [17] and a validation study is currently underway in I-SPY 2, with a larger patient population.

§ Dedicated breast PET (dbPET) in I-SPY2

Breast cancer is increasingly recognized to represent a heterogeneous group of diseases that vary in their treatment response, recurrence risk and overall prognosis [18]. Ever since the first description of breast cancer subtypes based on gene expression profiles [19], there has been growing emphasis on the molecular characteristics of breast cancer. While contrast-enhanced MRI depicts breast tumor morphology and vascularity [20], positron emission tomography (PET) with tumor specific tracers can provide

complementary molecular information that elucidates the underlying biology of the disease. Recent advances of organ specific PET scanner have allowed us to incorporate PET into the clinical workflow of breast imaging. In our ongoing effort to expand our breast imaging capability to the realm of molecular imaging, we evaluated the use of a dedicated breast PET (dbPET) to characterize breast tumor behavior and its response to treatment.

A 32 year-old female *BRCA1* gene mutation carrier with bilateral synchronous breast cancers was imaged with breast MRI (1.5 T) and a new FDA-approved dedicated breast PET scanner (MAMMI dbPET, OncoVision, Spain) before and after three weeks of neoadjuvant chemotherapy. The patient had two biopsy-proven invasive ductal carcinomas in the right breast, one of which was estrogen and progesterone receptor positive, HER2-negative (ER/PR+, HER2-) and the other triple receptor negative (TN), as well as a TN invasive ductal carcinoma in the left breast. Standard DCE-MRI was obtained using a dedicated breast coil. The patient also underwent MAMMI dbPET imaging with a low dose of F-18 FDG (5 mCi) at 45 min post-injection. The same imaging protocol was repeated after three weeks of chemotherapy.

Prior to treatment, breast MRI showed two malignant masses in the right breast measuring 4.0 cm (ER+) and 5.3 cm (TN), respectively, in longest diameter. Overall functional tumor volume (FTV) of both masses, defined as the volume of enhancing tumor exceeding an early enhancement threshold of 70% above baseline [20], was 73.2 cm^3 (Figure 2A). DbPET showed two FDG avid lesions with the maximum standard uptake value (SUV_{max}) of 19.1 for the ER + tumor and 19.5 for the TN tumor (Figure 2B).

After 3 weeks of paclitaxel treatment, MRI showed a decrease in size of the ER+ tumor to 3.2 cm, but there was slight enlargement of the TN tumor to 5.8 cm. Overall FTV of both masses also increased to 89.5 cm^3 (Figure 2C). As MRI appeared to show disease progression, carboplatin was added to the regimen and dbPET was obtained 1 week later. DbPET showed a complete resolution of FDG uptake in the ER + tumor and a 22% reduction of SUV_{max} in the TN tumor (SUV_{max} at 15.3) (Figure 2D). Repeat MRI obtained one week later showed minimal decrease in size of the right breast TN tumor to 5.2 cm and further decrease in the right ER+ tumor to 2.3 cm.

Within the left breast, baseline MRI showed a 1.2 cm malignant mass with overall FTV of 0.67 cm^3 and MAMMI dbPET showed an FDG avid mass with SUV_{max} of 6.7. After 3 weeks of chemotherapy, MRI showed residual disease (measuring 0.7 cm with FTV at 0.12 cm^3, whereas dbPET showed no FDG uptake in the left breast mass after 4 cycles of treatment.

After 12 weeks of paclitaxel chemotherapy, MRI demonstrated marked improvement of all 3 lesions with a residual ill-defined 3.8 cm TN mass and a 2.2 cm ER+ mass in the right breast with combined FTV at 1.82 cm^3. The left breast mass had resolved completely on MRI. The patient subsequently completed 4 cycles of doxorubicin and cyclophosphamide (AC). The final MRI prior to surgery showed a residual 0.8 cm TN mass with surrounding faint non-mass enhancement and faint non-mass enhancement at the site of the ER+ cancer (overall FTV at 0.22 cm^3).

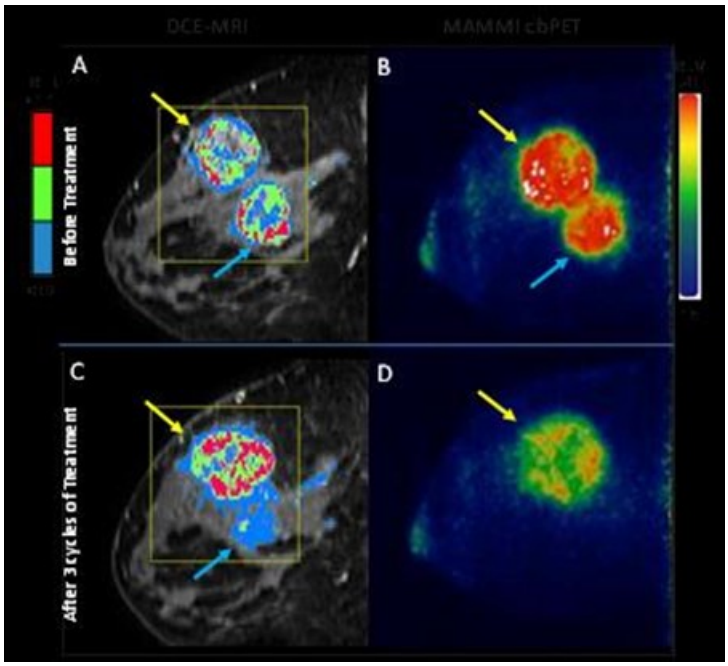


Figure 2: Breast imaging of a 32-year-old female patient with biopsy confirmed ER+/PR-/HER2- and TN invasive carcinomas in the right breast. **A**: Before treatment DCE-MRI showing the malignant lesions with the mapping of contrast signal enhancement ratio (SER) and overall FTV at 73.2 cm^3. **B**: Before treatment MAMMI dbPET imaging with FDG confirmed MRI findings, showing high FDG avidity in ER+ (blue arrow, SUVmax = 19.2) and TN (yellow arrow, SUVmax = 19.5) tumors. **C**: After 3 cycles of treatment, DCE-MRI showed residual disease in the ER+ tumor and progression of the TN tumor with the FTV at 89.5 cm^3, whereas **D:** 1 week after, MAMMI dbPET showed a complete resolution of FDG uptake in the ER+ tumor and reduction of SUVmax by 22% in the TN tumor.

Pathology from the subsequent right mastectomy revealed two residual foci of weakly ER+, HER2-negative, high-grade invasive ductal carcinoma measuring 1.5 cm and 0.7 cm. There was also residual high-grade ductal carcinoma in situ, which was present as scattered microscopic foci less than 1 mm each. Left mastectomy showed no evidence of residual disease.

This pilot study demonstrates that dbPET may be more sensitive than dynamic contrast enhanced MRI for evaluating early treatment response, revealing functional changes that precede anatomic changes at MRI. Further studies involving larger numbers of patients are underway to validate our initial observations. This work was recently published in the journal Clinical Breast Cancer [21].

§ Data-sharing efforts and TCIA Collections

Over the past year we worked with The Cancer Imaging Archive (TCIA) to provide three imaging collections for public access. In collaboration with TCIA, ACRIN, I-SPY and QIN, we developed and implemented MRI data sharing procedures for clinical studies focused on DCE and DWI of breast cancer. DICOM private attributes were defined and documented for embedding quality assessment and SER FTV results within the shared datasets, and a de-identification scheme suitable for our treatment response studies was developed. The primary effort was to provide highly curated and quality-assessed image data from the ACRIN 6657/I-SPY 1 trial for public data-sharing. This data was made available with limited access in August 2015 and became fully public September 1, 2016 [22]. The collection includes images and clinical data on 222 patients with 847 MRI studies. Protocol compliance and quality assessment enabled curating into multiple, easily accessible, collection subsets for different levels of analysis including basic radiologic evaluation, e.g. tumor size, and full SER FTV. Using the developed methods additional collections were established on TCIA including a 64-patient pilot study for SER FTV evaluation of treatment response in NAC, and a 13-subject collection for use in the QIN ADC Mapping Challenge led by Dr. Newitt.

DISCUSSION OF COLLABORATIONS

This U01 is being conducted in the context of the ongoing multi-center I-SPY 2/ACRIN 6698 Trial integrating molecular biomarkers and imaging to maximize the effectiveness of neoadjuvant treatment for patients with locally-advanced breast cancer. We leverage our existing partnerships with the American College of Radiology Imaging Network (ACRIN) Imaging Core, the National Institute of Standards and Technology (NIST) and Quantitative Imaging Network (QIN) sites to develop and evaluate a robust image quality assurance (QA) process for our ongoing and future clinical trials, and to optimize quantitative image classifiers for prediction of treatment response. In addition to the face-to-face meeting in April 2016 and regular teleconference with working groups (WG) and sub-groups, we have been interacting with other QIN sites on a regular basis. In particular, we have been working closely with Dr. Thomas Chenevert at the University of Michigan to develop the image quality ranking system and gradient non-linearity correction (GNC) in breast DWI.

§ Participation in QIN Network Committees and Working Groups

- Dr. Hylton served as Chair of the Executive Committee for the term April 2015-March 2016. During this time, she worked with QIN program leaders to establish a Working Group to develop guidelines for managing Challenges and Collaborative Projects (CCPs). Dr. Hylton continued to lead the Executive Committee Working Group on Breast MR Metrics of Response (BMMR).

- Dr. Newitt continued to lead the ADC Mapping Collaborative Project under the QIN MRI Subgroup of the Image Analysis and Performance Metrics Working Group

(IAPMWG). Dr. Newitt completed his service as Co-Chair of the MRI Subgroup of the IAPMWG in April 2016 and continues to participate as a member.

- Dr. Wilmes was a central participant in the Data Acquisition Working Group (DAWG). She led UCSF's participation and site data acquisition for the multi-center challenges for 1) Diffusion weighted imaging: Characterization (Phase I) [23] and correction (Phase II) of gradient non-linearity [24] and 2) Assessment of inter-platform variability of T_1 quantification methods used for DCE-MRI in a multicenter phantom study [25].

- Dr. Ella Jones was a regular participant in the Clinical Trial Design and Development Working Group (CTDD WG). In 2016, she served as the Vice Chair of the CTDD WG and co-authored a paper surveying the accrual pattern in clinical studies involving quantitative imaging [26].

§ Participation in QIN Challenges

UCSF investigators led two QIN Challenges/Collaborative Projects (CCPs) and were also major participants in two additional CCPs, as described below.

§ The Breast MRI Metrics of Response (BMMR) Challenge

A QIN-sponsored challenge for evaluation of Breast MRI Metrics of Response (BMMR) was designed in 2016 by a QIN Executive Committee working group led by Dr. Hylton, implementing the challenge and collaborative project procedures developed by the QIN in 2015. The objective of the challenge was the prediction of recurrence free survival time (RFS) for patients with invasive breast cancer undergoing neoadjuvant chemotherapy, utilizing serial DCE-MRI studies taken over the course of therapy. The BMMR challenge opened in May 2016 and ran through October 2016 and was the 1st QIN challenge to be performed under the new QIN guidelines for CCPs. The BMMR Challenge used MRI data from 162 ACRIN 6657/I-SPY 1 patients, annotated with RFS outcome and breast cancer subtype defined by hormone receptor (HR) and HER2 receptor status. Separate data on 64 patients with RFS outcomes from a UCSF pilot neoadjuvant breast cancer study was provided as a training data set. Both training and test data sets were made available to Challenge participants on TCIA. The challenge was managed in collaboration with Dr. Jayashree Kalpathy-Cramer through the QINLABS website. Three QIN groups (U. Chicago (M. Giger), Moffitt Hospital (J. Drukteinis) and MGH (J. Kalpathy-Cramer)), and one non-QIN group (U. Pennsylvania (D. Kontos)) submitted results for evaluation. Three other groups (Stanford, U. Washington, and OHSU) participated in development of the challenge but did not complete analysis of the test phase data. Statistical analysis of the challenge results was performed by members of the ACRIN Biostatistical Center (Zheng Zhang and Helga Marques of the Brown U. Center for Statistical Sciences). The BMMR Challenge results are currently being prepared for publication and will be presented at the 2017 QIN face-to-face meeting.

§ QIN ADC Mapping Collaborative Project: Multi-site Concordance of DWI Metrics

The ADC Mapping CCP, led by Dr. David Newitt of UCSF was undertaken to examine the variability in apparent diffusion coefficient (ADC) measures resulting from different post-processing software implementations utilized by researchers across the NCI Quantitative Imaging Network. Participating QIN sites included UCSF, University of Michigan-1, BNI, BWH, JHU, Mount Sinai, MCW, MGH, OHSU, University of Michigan-3, University of Washington and Vanderbilt. A secondary aim of the ADC Mapping Challenge was to evaluate the feasibility and practical challenges involved in centralized analysis of multi-center ADC data. MRI data from both phantom [27] (Ph) and in vivo breast (Br) DWI was analyzed, including data from three major MRI scanner manufacturers: Siemens, Philips and GE Medical Systems (Table 2). The breast MRI studies [28] were curated and de-identified at UCSF and shared via TCIA [29] in a private collection for this CCP. Phantom data was provided by U. Michigan via the NCIPHub [30]. Eleven QIN sites calculated parametric maps using 12 DWI analysis platforms, with analysis implementations using IDL, Matlab, 3D Slicer, OsiriX, AFNI, C++ and QIBAPhan1.3, and file formats DICOM, NIFTI, NRRD and Matlab. Manufacturers' software (scanner-generated) DICOM ADC maps were also evaluated where available. All comparative and statistical analyses were done by D. Newitt and J. Gibbs at UCSF. For comparisons, all maps were converted to a modified-DICOM format and scaling factors were set in the meta-data to produce ADC maps in common units of 10^{-6} mm^2/sec. ROIs were defined as shown in Figure 3 and applied to the parametric maps yielding mean values of the diffusion metrics. Concordance was evaluated from the percent difference of each measurement from the median value for all QIN sites. Pairwise within-subject coefficient of variation (wCV) was calculated for all site pairs and metrics to establish groupings of similar (wCV<0.1%) results.

Group label	DWI Scan Description	N studies *ID values*	b values (s/mm^2)	Scanner manufacturers *	Parameters @	Analysis ROIs
Ph4b	4 b-value, 3 direction QIBA diffusion phantom	3 *401, 402, 403*	0, 500, 900, 2000	GEMS, SM, PM	ADC4 $ADC_{hi\text{-}low}$	1cm circles 13 vials
Br2b	2 b-value, 3 direction bilateral axial breast	3 *101, 102, 103*	0, 800	GEMS, SM, PM	ADC2	multi-slice tumor
Br4b	4 b-value, 3 direction bilateral axial breast	8 *201-208*	0, 100, 600, 800	GEMS, SM, PM	ADC4 $ADC3_{slow}$ PerfFrac	multi-slice tumor

* Manufacturers: General Electric Medical Systems (GEMS); Siemens Medical (SM); Philips Medical (PM)

@ DWI Parameter Definitions:
ADC<n> (n=# b-values) mono-exponential ADC
$ADC_{hi\text{-}low}$ mono-exponential ADC using only highest and lowest b-values
$ADC3_{slow}$ mono-exponential ADC using 3 highest b-values
PerfFrac fraction of b=0 signal attributed to fast-decaying perfusion component

Table 2: Protocol and analysis metric descriptions for data sets included in ADC Mapping Challenge.

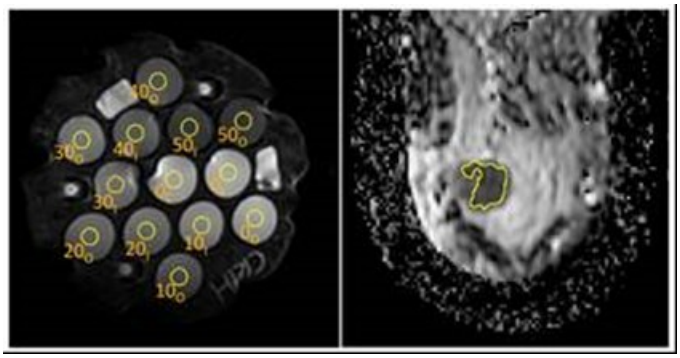


Figure 3: ROI definitions for PVP phantom scans (left) and in vivo breast scans (right). The breast image shows a single representative slice of the multi-slice tumor ROI.

Results

All 12 platforms were able to produce mono-exponential ADC maps for the Br2b and Br4b groups, and perfusion-suppressed $ADC3_{slow}$ values for Br4b. 8 platforms provided perfusion-fraction maps for the Br4b studies. All sites were able to handle all multi-vendor DICOM image sets, but interpretation of the full directional data from the GEMS scanners (Br4b, IDs 203, 204) was challenging for several sites due to unfamiliarity with this format, requiring assistance from UCSF. All maps were centrally analyzable, but required a variety of manipulations including scaling, slice order reversal, and masking of NaN values, illustrating the necessity of adoption of a uniform DICOM standard for parametric maps [31]. Preliminary analysis was completed in Fall 2016, and submitted as an abstract for the 2017 ISMRM meeting. Sample results for the 4 b-value breast ADC are shown in Figure 4. Inter-site wCV tables revealed eight of the sites were grouped into 2 separate groups: sites [1, 4, 13] with wCV<0.01% and sites [3, 5, 6, 8, 9] wCV<0.1%, while the other 4 sites and the scanner-generated maps showed more individualistic behavior. ADC values differed 2.8±0.2% between the two groups and up to 5% for non-grouped sites. The Philips scanner map had a 28% error due to inaccurate scaling information in the DICOM. Phantom results showed similar groupings amongst analysis implementations, though with smaller differences between the groups: RMS percent difference in ADC values for all phantom ROI of 0.29%, 0.30%, 0.62% for GEMS, Siemens, and Philips scans respectively. Full results will be submitted for publication in 2017.

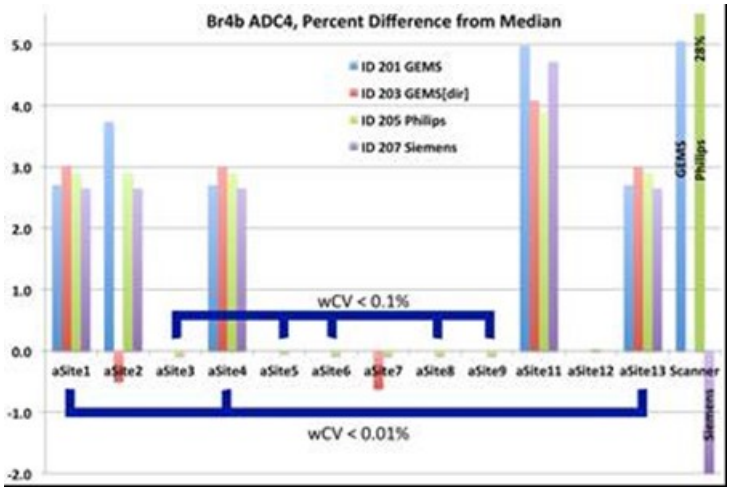


Figure 4: Sample results: Percent difference from median values for all sites for the mono-exponential ADC from the 4 b-value breast scans. Horizontal bars indicate the 2 groups of sites with close to identical results as measured by the wCV. Scanner-generated results are shown at the far right.

§ AutoPERCIST Challenge

In our effort to incorporate PET into the workflow of breast imaging, we recognized the need for accurate objective measurements of standard uptake values (SUVs), required by PERCIST 1.0, to evaluate FDG uptake and tumor response to treatment. We sought to collaborate with Drs. Richard Wahl (Washington University) and Jeff Leal (John Hopkins University) to assess the AutoPERCIST software to semi-automatically identify and measure reference tissue (liver), set disease threshold values and calculate SUVs (peak, max, mean, volume and total lesion glycolysis). Using the latest version of AutoPERCIST, we were able to accurately identify and measure breast cancer patient's primary breast tumor and axillar lymph node (Figure 5). We also participated in a multicenter reader variability study of AutoPERCIST through the CTDD WG. Sixteen sites including six international institutions participated in this study. Thirty paired sets of anonymized FDG PET-CT images were downloaded for evaluation and up to 5 tumor lesions from each PET image will be selected. All selections will be recorded and sent to the central database at Johns Hopkins Image Response Assessment Team for quality control.

§ DWI Gradient non-linearity correction Challenge: Phase I and Phase II

Our group participated in a multi-site MRI data challenge led by Tom Chenevert's group at University of Michigan to improve the accuracy of diffusion weighted imaging

(DWI). Gradient non-linearity in DWI introduces a significant spatial bias in apparent diffusion coefficient (ADC) values. In the DAWG Phase I project, our site and others acquired DWI data from an imaging phantom with known ADC value for multiple locations within the MRI bore. From these data, GNL bias was characterized and "corrector functions" were generated [23]. In the Phase II project, our group was one of a subset of initial participants that acquired DWI data in a different phantom to validate the corrector functions derived in Phase I. This work demonstrated that the GNL correction resulted in increased quantitative accuracy in measured ADCs across multiple sites and MRI scanner vendors [24].

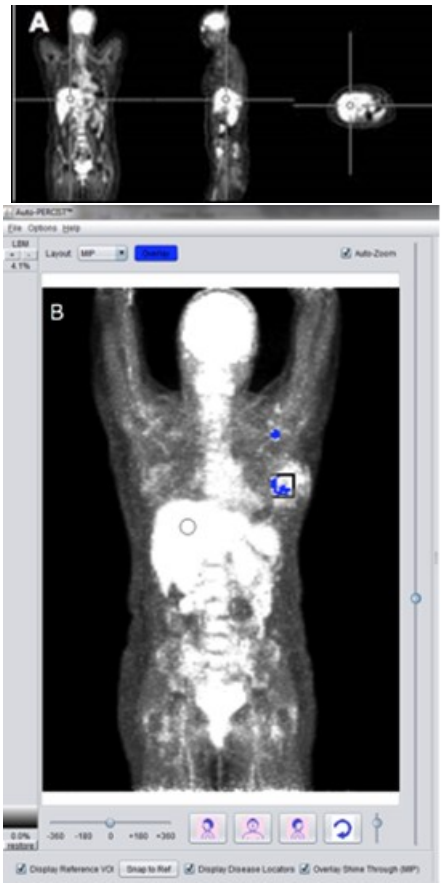


Figure 5: Screenshots from AutoPERCIST. A: Automatic identification of the liver as a reference organ. B: Measurement of the primary breast tumor and the axillar lymph node.

§ T1 mapping Challenge: Assessment of inter-platform variability of T1 quantification methods used for DCE-MRI in a multicenter phantom study

Our group participated in a multi-site MRI data challenge led by Bachir Taouli's group at Mt. Sinai hospital to evaluate different MRI T1 mapping methods for accuracy and

variability. To this end we implemented standard T1 mapping protocols provided by the challenge and acquired data in a phantom of known T1 values using the prescribed sequences as well as our local site protocol. These data were provided to the challenge and in combination with data from other imaging sites, were used to determine the variability of different T1 mapping sequences and which sequence is the most accurate for calculating T1 values [25].

ACTIVITIES OUTSIDE THE QIN NETWORK

§ The Cancer Imaging Archive (TCIA)

We worked with investigators at TCIA and the ACRIN Imaging Core Lab to enable the transfer of ACRIN 6657 image data and to ensure that appropriate patient de-identification and DICOM standards for QI are incorporated. We implemented specific capabilities to support the archival of derived images and metadata associated with the tumor volumetric analysis used to generate the primary imaging endpoint, functional tumor volume (FTV) for ACRIN 6657, as well as for storing results from QA/QC evaluations.

§ American College of Radiology Imaging Network (ACRIN)

To improve image quality assurance, we collaborated with the ACRIN Imaging Core to develop the image QA/QC program for the I-SPY2/6698 trial. The resulting DWI quality ranking system has been implemented as part of the image review and analysis process used to generate the primary study endpoint for the trial.

§ NIST and Industry

We are collaborating with the National Institute of Standards and Technology (NIST) and two industrial partners, High Precision Devices, Inc. (HPD) and The Phantom Laboratory through the SBIR Phase I award mechanism. We previously worked with NIST to design and prototype a universal breast MRI phantom that could be used in multi-center clinical trials for standardization and quality control of breast DCE and DWI data [32, 33]. Based on specifications provided by UCSF, the next generation of prototype phantom is being designed by both industrial partners. These two prototypes are compatible with most major breast MRI coils and magnet systems, and contain compartments with materials mimicking the MRI properties of normal fibroglandular, adipose breast tissue and breast tumor, with a representative range of T1, T2 and ADC values. In addition, geometrical objects for evaluation of image distortion and resolution are in place. In conjunction with the final physical design and single-site testing at UCSF, we will develop comprehensive image acquisition protocols and measurement methods to efficiently monitor breast-imaging critical parameters including T1 and ADC measurements, image distortion, fat suppression and SNR. The phantom and associated protocols will then be evaluated in a pilot multicenter study at 3-5 I-SPY 2 clinical centers. Additionally, in collaboration with the QIN Data Acquisition Working Group (DAWG) and Image Analysis and Performance Metrics Working Group (IAPMWG), we plan to design and execute a phantom-based challenge focused on breast-specific imaging using the finalized phantom.

PLANS FOR NEXT YEAR

Our renewal application received an Impact Score of 20 and is pending. If funded, our continuing U01 program will focus on the clinical evaluation of advanced QI methods in I-SPY 2. Through our experience with ACRIN 6698, we encountered a number of issues that pose limitations to the quantitative use of DWI in the NAC setting, including systematic errors in acquired data due to gradient non-linearity and B_0 inhomogeneity, and variability in image quality and consistency. We propose to address these in the ongoing U01 project and to perform more in-depth and robust evaluation of breast DWI for assessing NAC response in I-SPY 2. We plan to introduce improvements to both the DWI data acquisition methods and diffusion quantification approaches. Through continued collaboration with the National Institute of Standards and Technology, we will utilize the universal breast MRI phantom to implement a phantom-based quality control (QC) process at the participating I-SPY 2 clinical centers. We will also implement vendor-specific GN correction for all breast DWI data in I-SPY 2. We will continue collaborative efforts begun under ACRIN 6698 to improve diffusion quantification approaches and develop DWI-based metrics that can be used in combination with DCE metrics to improve predictive performance of imaging. The project will be conducted in the clinical context of the ongoing I-SPY 2 trial, allowing us to measure the impact of each proposed refinement, as well as the overall effectiveness of DWI for predicting response and survival.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS - 2016

§ Publications

1. Hylton NM, Gatsonis CA, Rosen MA, et al: Neoadjuvant Chemotherapy for Breast Cancer: Functional Tumor Volume by MR Imaging Predicts Recurrence-free Survival-Results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. Radiology 279:44-55, 2016
2. Rugo HS, Olopade OI, DeMichele A, et al: Adaptive Randomization of Veliparib-Carboplatin Treatment in Breast Cancer. N Engl J Med 375:23-34, 2016
3. Park JW, Liu MC, Yee D, et al: Adaptive Randomization of Neratinib in Early Breast Cancer. N Engl J Med 375:11-22, 2016
4. Bolan PJ, Kim E, Herman BA, et al: MR spectroscopy of breast cancer for assessing early treatment response: Results from the ACRIN 6657 MRS trial. J Magn Reson Imaging, 2016
5. Lo WC, Li W, Jones EF, et al: Effect of Imaging Parameter Thresholds on MRI Prediction of Neoadjuvant Chemotherapy Response in Breast Cancer Subtypes. PLoS One 11:e0142047, 2016
6. Li W, Arasu V, Newitt DC, et al: Effect of MR Imaging Contrast Thresholds on Prediction of Neoadjuvant Chemotherapy Response in Breast Cancer Subtypes: A

Subgroup Analysis of the ACRIN 6657/I-SPY 1 TRIAL. Tomography 2:378-387, 2016

7. Jones EF, Ray KM, Li W, et al: Dedicated Breast Positron Emission Tomography for the Evaluation of Early Response to Neoadjuvant Chemotherapy in Breast Cancer. Clinical Breast Cancer Online release, 2016
8. Kurland BF, Aggarwal S, Yankeelov TE, et al: Accrual Patterns for Clinical Studies Involving Quantitative Imaging: Results of an NCI Quantitative Imaging Network (QIN) Survey. Tomography 2:276-282, 201
9. Malyarenko DI, Wilmes LJ, Arlinghaus LR, et al: QIN DAWG Validation of Gradient Nonlinearity Bias Correction Workflow for Quantitative Diffusion-Weighted Imaging in Multicenter Trials. Tomography 2:396-405, 2016
10. Wilmes L, Li W, Shin HJ, et al: Diffusion Tensor Imaging for Assessment of Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer. Tomography 2:438-447, 2016
11. Malyarenko DI, Newitt D, L JW, et al: Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials. Magn Reson Med 75:1312-23, 2016
12. Keenan KE, Peskin AP, Wilmes LJ, et al: Variability and bias assessment in breast ADC measurement across multiple systems. J Magn Reson Imaging 44:846-55, 2016
13. Keenan KE, Wilmes LJ, Aliu SO, et al: Design of a breast phantom for quantitative MRI. J Magn Reson Imaging 44:610-9, 2016

§ Abstracts

1. Arasu V, Harnish R, McHarque C, Li W, Wilmes L, Newitt D, Jones E, Esserman L, Joe B, Hylton N. Automated Breast MRI Segmentation Method for Background Parenchymal Enhancement. 2016 ISMRM Annual Meeting, Singapore.

2. Li W, Arasu V, Jones E, Newitt D, Wilmes L, Kornak J, Esserman L, Hylton N. Breast MRI for Early Prediction of Residual Disease Following Neoadjuvant Chemotherapy: Optimization of Response Cut-Point by Tumor Subtype. 2016 ISMRM Annual Meeting, Singapore.

3. Li W, Arasu V, Jones EF, Newitt DC, Wilmes LJ, Kornak J, Esserman LJ, Hylton NM. Effect of MR imaging contrast kinetic thresholds for prediction of neoadjuvant chemotherapy response in breast cancer subtypes - Results from ACRIN 6657 / I-SPY 1 trial. 2016 San Antonio Breast Cancer Symposium.

4. Nayak L, Ray KM, Woodard GA, Joe BN, Hylton NM, Price ER, Gibbs J, Lobach I, Newitt D, Baehner R. Quantitative Radiogenomics: Association between Breast MRI Functional Tumor Volume and Oncotype DX Recurrence Score. 2016 RSNA Annual Meeting.

5. Bane O, Hectors S, Wagner M, Arlinghaus L, Aryal M, Cao Y, Chenevert T,

Fennessy F, Huang W, Hylton N, Kalpathy-Cramer J, Keenan K, Malyarenko D, Mulkern R, Newitt D, Russek SE, Stupic KF, Wilmes L, Yankeelov T, Yen Y F, Boss M, Taouli B, Assessment of interplatform reproducibility of T_1 quantification methods used for DCE-MRI: results from a multicenter phantom study. Accepted International Society of Magnetic Resonance In Medicine, Annual Meeting, Honolulu, HI, April 2017.

6. Dariya I. Malyarenko, Yuxi Pang, Lisa J. Wilmes, Ek T. Tan, John Kirsch, Julien Senegas, Michael A. Jacobs, David Newitt and Thomas L. Chenevert. Vendor-provided gradient characteristics correct nonuniform diffusion weighting. Accepted International Society of Magnetic Resonance In Medicine, Annual Meeting, Honolulu, HI, April 2017.
7. Yuxi Pang, Dariya I. Malyarenko, Michael Schär, Lisa J. Wilmes, David C. Newitt, Michael A. Jacobs, and Thomas L. Chenevert. Characterization of B0 Shim-Induced Bias in Diffusion Weighting Gradients Accepted International Society of Magnetic Resonance In Medicine, Annual Meeting, Honolulu, HI, April 2017.
8. David C. Newitt, Dariya Malyarenko, Thomas L. Chenevert et al. Multi-site Concordance of DWI Metrics: Results of the NCI Quantitative Imaging Network ADC Mapping Collaborative Project. Accepted International Society of Magnetic Resonance In Medicine, Annual Meeting, Honolulu, HI, April 2017.
9. David C. Newitt, Zheng Zhang, Jessica Gibbs et al. Reproducibility of ADC measures by Breast DWI: Results of the ACRIN 6698 Trial. Accepted International Society of Magnetic Resonance In Medicine, Annual Meeting, Honolulu, HI, April 2017.
10. Lisa J. Wilmes, Ek T. Tan, Evelyn Proctor, Jessica Gibbs, Nola M. Hylton, and David C. Newitt. Correction of image distortion and gradient nonlinearlity in DTI of breast cancer. Accepted International Society of Magnetic Resonance In Medicine, Annual Meeting, Honolulu, HI, April 2017.
11. Jessica Gibbs, Ella Jones, Lisa Wilmes, David C. Newitt, John Kornak, Melanie Regan, and Nola Hylton. Comparison of whole tumor and single slice ROIs for measuring ADC in breast cancer. Accepted International Society of Magnetic Resonance In Medicine, Annual Meeting, Honolulu, HI, April 2017.

LIST OF REFERENCES

1. Fisher B, Bryant J, Wolmark N, et al: Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 16:2672-85, 1998
2. Bear HD, Anderson S, Smith RE, et al: Sequential preoperative or postoperative docetaxel added to preoperative doxorubicin plus cyclophosphamide for operable breast cancer:National Surgical Adjuvant Breast and Bowel Project Protocol B-27. J Clin Oncol 24:2019-27, 2006
3. Peintinger F, Sinn B, Hatzis C, et al: Reproducibility of residual cancer burden for prognostic assessment of breast cancer after neoadjuvant chemotherapy. Mod Pathol 28:913-20, 2015
4. Esserman LJ, Berry DA, DeMichele A, et al: Pathologic complete response predicts recurrence-free survival more effectively by cancer subset: results from the I-SPY 1 TRIAL--CALGB 150007/150012, ACRIN 6657. J Clin Oncol 30:3242-9, 2012
5. Sheri A, Smith IE, Johnston SR, et al: Residual proliferative cancer burden to predict long-term outcome following neoadjuvant chemotherapy. Ann Oncol 26:75-80, 2015
6. Symmans WF, Peintinger F, Hatzis C, et al: Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol 25:4414-22, 2007
7. Barker AD, Sigman CC, Kelloff GJ, et al: I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clin Pharmacol Ther 86:97-100, 2009
8. Rugo HS, Olopade OI, DeMichele A, et al: Adaptive Randomization of Veliparib-Carboplatin Treatment in Breast Cancer. N Engl J Med 375:23-34, 2016
9. Park JW, Liu MC, Yee D, et al: Adaptive Randomization of Neratinib in Early Breast Cancer. N Engl J Med 375:11-22, 2016
10. Saritas EU, Cunningham CH, Lee JH, et al: DWI of the spinal cord with reduced FOV single-shot EPI. Magn Reson Med 60:468-73, 2008
11. Singer L, Wilmes LJ, Saritas EU, et al: High-resolution diffusion-weighted magnetic resonance imaging in patients with locally advanced breast cancer. Acad Radiol 19:526-34, 2012
12. Wilmes LJ, McLaughlin RL, Newitt DC, et al: High-resolution diffusion-weighted imaging for monitoring breast cancer treatment response. Acad Radiol 20:581-9, 2013
13. Wilmes L, Li W, Shin HJ, et al: Diffusion Tensor Imaging for Assessment of Response to Neoadjuvant Chemotherapy in Patients With Breast Cancer. Tomography 2:438-447, 2016
14. Hylton NM, Blume JD, Bernreuter WK, et al: Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy--results from ACRIN 6657/I-SPY TRIAL. Radiology 263:663-72, 2012
15. Hylton NM, Gatsonis CA, Rosen MA, et al: Neoadjuvant Chemotherapy for Breast Cancer: Functional Tumor Volume by MR Imaging Predicts Recurrence-free Survival-Results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. Radiology 279:44-55, 2016

16. Lo WC, Li W, Jones EF, et al: Effect of Imaging Parameter Thresholds on MRI Prediction of Neoadjuvant Chemotherapy Response in Breast Cancer Subtypes. PLoS One 11:e0142047, 2016
17. Li W, Arasu V, Newitt DC, et al: Effect of MR Imaging Contrast Thresholds on Prediction of Neoadjuvant Chemotherapy Response in Breast Cancer Subtypes: A Subgroup Analysis of the ACRIN 6657/I-SPY 1 TRIAL. Tomography 2:378-387, 2016
18. Polyak K: Heterogeneity in breast cancer. J Clin Invest 121:3786-8, 2011
19. Perou CM, Sorlie T, Eisen MB, et al: Molecular portraits of human breast tumours. Nature 406:747-52, 2000
20. Hylton NM, Blume JD, Bernreuter WK, et al: Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy--results from ACRIN 6657/I-SPY TRIAL. Radiology 263:663-72, 2012
21. Jones EF, Ray KM, Li W, et al: Dedicated Breast Positron Emission Tomography for the Evaluation of Early Response to Neoadjuvant Chemotherapy in Breast Cancer. Clinical Breast Cancer Online release, 2016
22. I-SPY 1 / ACRIN 6657 TCIA Collection, 2015
23. Malyarenko DI, Newitt D, L JW, et al: Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials. Magn Reson Med 75:1312-23, 2016
24. Malyarenko DI, Wilmes LJ, Arlinghaus LR, et al: QIN DAWG Validation of Gradient Nonlinearity Bias Correction Workflow for Quantitative Diffusion-Weighted Imaging in Multicenter Trials. Tomography 2:396-405, 2016
25. Bane O, Hectors S, Wagner M, et al: Assessment of interplatform reproducibility of T1 quantification methods used for DCE-MRI: results from a multicenter phantom study. , International Society of Magnetic Resonance In Medicine, Annual Meeting, April 2017, pp 2571
26. Kurland BF, Aggarwal S, Yankeelov TE, et al: Accrual Patterns for Clinical Studies Involving Quantitative Imaging: Results of an NCI Quantitative Imaging Network (QIN) Survey. Tomography 2:276-282, 2016
27. Boss M CT, et al.: Multicenter study of reproducibility of a wide range of ADC at 0C. , Proceedings RSNA. Chicago, IL, 2016
28. American College of Radiology Imaging Network (ACRIN) 6698 Trial: Diffusion-weighted MR Imaging Biomarkers for Assessment of Breast Cancer Response to Neoadjuvant Treatment, https://http://www.acrin.org/6698_protocol.aspx,
29. Clark K, Vendt B, Smith K, et al: The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 26:1045-57, 2013
30. NCIPHub, https://nciphub.org,
31. Parametric Map DICOM storage, dicom.nema.org/Dicom/News/june2014/docs/sup172.pdf
32. Keenan KE, Peskin AP, Wilmes LJ, et al: Variability and bias assessment in breast ADC measurement across multiple systems. J Magn Reson Imaging 44:846-55, 2016

33. Keenan KE, Wilmes LJ, Aliu SO, et al: Design of a breast phantom for quantitative MRI. J Magn Reson Imaging 44:610-9, 2016

U01CA166104: Advancing Quantification of Diffusion MRI for Oncologic Imaging

University of Michigan (1)

Thomas L. Chenevert, Ph.D., Brian D. Ross, Ph.D., Craig J. Galban, Ph.D. and Dariya Malyarenko, Ph.D.

INTRODUCTION

The overarching goal of this project is to provide for standardized implementation and clinical validation of advanced quantitative diffusion-weighted MRI (qDWI) analytical techniques for quantification of tumor diffusion values across multiple MRI systems in order to improve use in multi-site cancer imaging trials [1]. This research effort is focused on identification and mitigation of significant technical impediments to DWI as a quantitative imaging (QI) metric for cancer patients. Activities are aligned with three specific objectives involving strategic collaborations within the NCI Quantitative Imaging Network (QIN), Imbio, LLC (industrial partner), the National Institute of Standards and Technology (NIST), Eastern Cooperative Oncology Group and the American College of Radiology Imaging Network (ECOG-ACRIN), and the Imaging and Radiation Oncology Core (IROC). The major goals of this U01 research effort are focused on an integrated series of three Specific Aims involving image data transfer and analysis with quantitative software testing (Aim 1), data acquisition quality assurance and system characterization (Aim 2), removal of technical bias (Aim 3) and statistical evaluation (Aim 1). Scans will be obtained for site/system certification and quality control as described in Aim 2. Diffusion MR clinical data will undergo bias removal as described in Aim 3 followed by generation of histogram and voxel-wise parametric response map (PRM) metrics using the proposed advanced software algorithms described in Aim 1. The image data source for ACRIN 6702 & 6698 trials is the TRIADv4 application. Deliverables by Aim are as follows:

§AIM 1

To develop and evaluate a reproducible and robust computational environment for quantification of diffusion-weighted MR images using data collected from the ISPY-2 breast cancer trial.

For this QIN effort, diffusion-PRM will be evaluated using a novel computing platform allowing DICOM data to be securely uploaded and processed through a web browser. This will enable easy collaboration between Imbio and UM QIN investigators. All software under this grant proposal will be developed and tested in accordance to Imbio's quality system that ensures standardization of quantitative measures. This approach will be undertaken using a semi-automated spatial alignment of serial data. Image registration will be performed using a multi-layered approach. Rigid body registration is first performed followed by a geometric warping interpolant, i.e. B-spline, algorithm used for mapping the tumor volumes from interval examinations onto the tumor volume from the pre-therapy

anatomical image, which is defined by the user. Registration will be based on the optimization of mutual information between two image data sets, allowing for multimodal registration, and implemented through the use of automatically determined control points.

§ AIM 2

To devise the next generation DWI phantom for absolute quantitation that spans the tissue ADC range and incorporates internal MR-thermometry; and to extend QA/QC metrics to include characterization of systematic bias for ongoing multi-center breast cancer trials.

Through QIN and ACRIN collaborations, we provide QA/QC services and facilitate incorporation of developed quality assurance methods utilizing an ice water diffusion phantom in ongoing multi-center ACRIN 6698 and 6702 clinical trials. Acceptable performance of each newly added MR system will be confirmed upon entry to the study, and reaffirmed bi-annually or after significant hardware/software upgrades to each system. Our developed uniform data structure format (regardless of vendor-specific DICOM) will be used for data screening for protocol compliance testing and data reduction. Each QA/QC DWI scan is performed in four passes in rapid succession such that system noise & short-term instability artifact level are measurable for each pixel by variance over these passes. QA procedures will be amended to include the long-tube ice-water phantom distribution among participating sites, as well as implementation of uniform quality assurance protocol for routine assessment of systematic spatial GNL bias on relevant scanner platforms. Analogous to the multi-center DWI phantom study, a detailed phantom preparation and QA/QC scan protocol will be provided with each phantom set. Reported performance metrics will include measures of random noise and bias over FOV, as well as DWI directional spatial uniformity coefficients, and scanning protocol compliance.

§ AIM 3

Enhancement of predictive power for quantitative diffusion metrics by retrospective correction of DW-MRI gradient nonlinearity (GNL) errors in multi-center therapy-response trials.

Our proposed GNL bias correction approach will follow the recently described algorithm, based on system characterization from regular QC measurements on the ice water diffusion phantom (Aim 2). This correction can be implemented independent of proprietary information on gradient design. This aim will include (1) system GNL and SNR evaluation (data from Aim 2); (2) modeling of system-specific nonlinearity tensors; (3) construction of digital 3D maps for DW bias; (4) application of corrector maps to patient DWI-ADC data from ongoing clinical trials (data from Aim 1); (5) performance evaluation for quantitative population statistics (e.g., fDM and histogram metrics) with and without bias correction.

DISCUSSION OF PROGRESS

§ Progress Toward Next Generation DWI Phantom and QA/QC

A formal collaboration has been established with NIST scientists to develop an ambient temperature diffusion phantom based on polyvinylpyrrolidone (PVP) that offers absolute quantitative diffusion coefficients tunable by PVP concentration, long-term stability, no toxicity, and spans the tissue ADC range 0.4→2.5(x10-3mm2/s). This builds on an existing temperature-controlled PVP-based DWI phantom design (http://www.nist.gov/pml/div686/grp08/biomagnetics.cfm#) that has become "a standard" since endorsed by QIBA (http://www.rsna.org/qiba/) and is commercially available (http://hpd-online.com/MRI-phantoms/php). UM investigators were deeply involved in its design and understand its two key limitations: (a) it requires ice-water temperature control to achieve absolute quantification, thus phantom preparation is relatively tedious with limited duration for use (<1-2hr) , and (b) 0^{o}C PVP solutions only span half the tissue ADC range. Our proposition is to use PVP solutions at ambient temperature thereby eliminating phantom preparation while achieving the full relevant tissue ADC range. However, this approach now requires determination of absolute temperature internal to the phantom. To achieve this, a temperature-sensitive chemical probe insert, combined a single-shot low bandwidth EPI (LB-EPI) sequence will provide an estimate of temperature based on spatial separation of chemical moieties visible as ghosts on the image. When deployed in the field, we believe an image-based read-out of internal temperature using a standard sequence with parameters set to maximize spatial separation of chemical shift ghosts (i.e. low bandwidth) would be less dependent on shim quality and operator skill than single-voxel spectroscopy. NIST has been contracted to provide essential calibration of chemical shift vs temperature of candidate probes chemically designed for: (a) long T2; (b) temperature sensitivity; and (c) comparable signal between chemical moieties. The calibration curve (Figure 2A) performed at 500MHz (11.7T) of a pH-adjusted t-butanol with dilute deuterium provides excellent NMR properties and reasonable temperature-sensitivity. Figure 2B illustrates temperature read-out via LB-EPI performed on a clinical 1.5T (64MHz) system. A software routine was developed to automatically analyze LB-EPI to convert spatial shift (via cross-correlation) to chemical-shift, thus temperature. Magnetic field-dependent bias discovered in our fiber-optic temperature probe (used for independent confirmation) was recently rectified and new batches of chemically-designed probes are being evaluated.

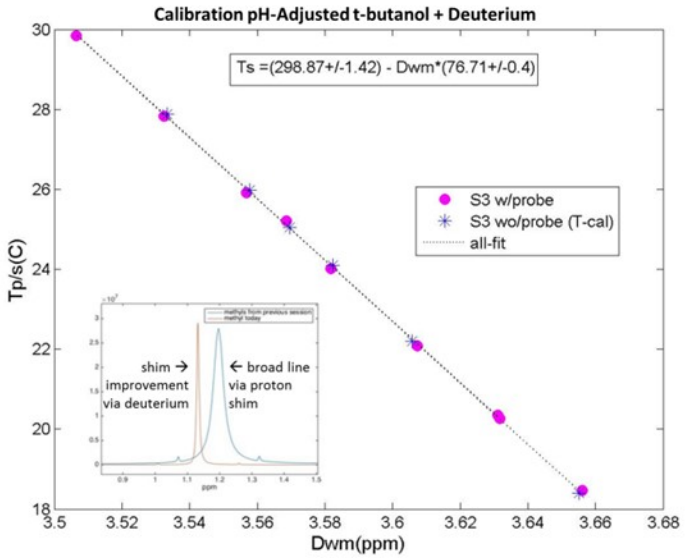


Figure 2A: Chemical shift vs temperature calibration performed at 500MHz on pH-adjusted t-butanol + deuterium sample. The deuterium improves shimming to narrow line width (inset), thus improves calibration precision. Once calibration is available, fit coefficients are used in conversion of low-bandwidth EPI chemical shift "distance" into internal phantom temperature when scanned on standard clinical system.

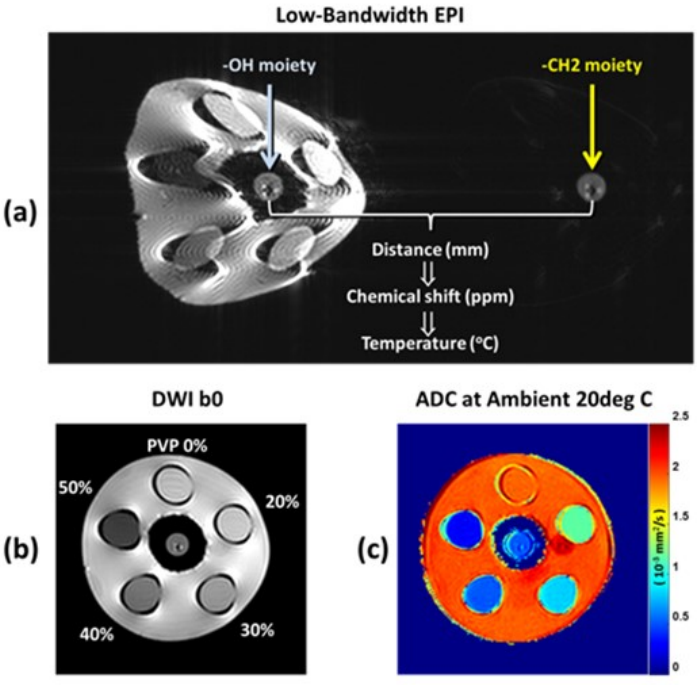


Figure 2B: (a) Low-bandwidth EPI, to increase sensitivity to chemical shift difference between –OH and –CH2, hence improve temperature sensitivity; (b) DWI of PVP phantom containing chemical temperature probe; and (c) corresponding ADC map spanning (0.3 to 2.0)x10^{-3}mm^2/s.

In parallel to work on this next generation DWI phantom, QA/QC software tools, fabrication and delivery of DWI ice-water phantoms, and Site Certification services have been established with aim to improve uniformity and quality of DWI in clinical trials. To date, these services have been utilized in four Oncology trials: ACRIN6698 (Breast); ACRIN6701 (Prostate); ACRIN6702 (Breast); NRG-BN001 (Brain). Over 200 DWI phantom datasets have been analyzed. Approximately 15% failed certification tests due to: significant protocol violations; high ADC non uniformity due to gradient non linearity; and low SNR. Figure 2C illustrates the automated DWI QA/QC analysis output (Fig 2C (a)) used for system certification report (Fig 2C (b)) for one NRG-BN001 trial site.

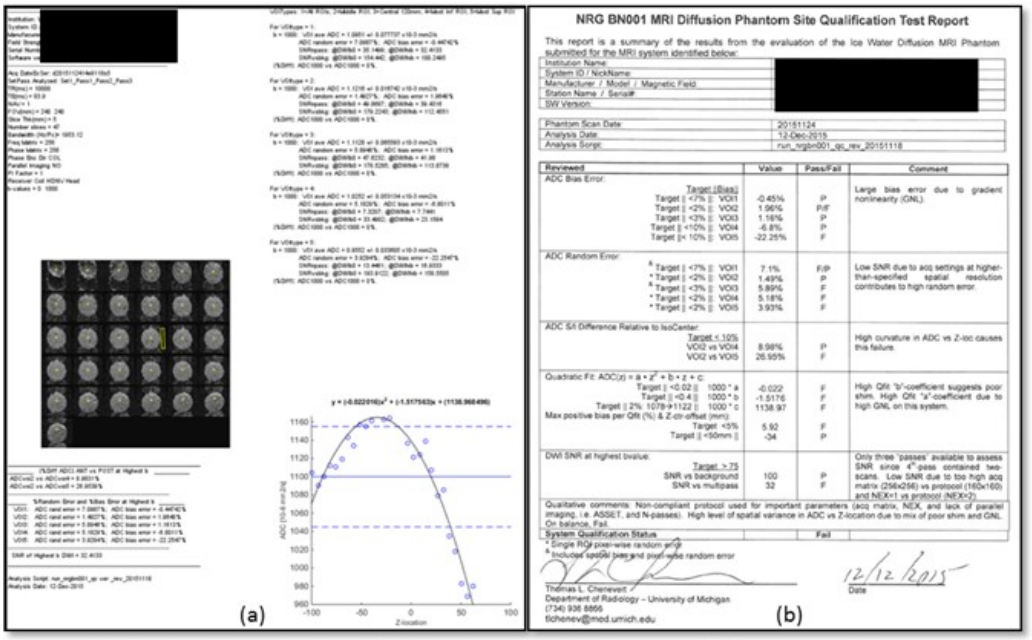


Figure 2C: (a) Sample output of DWI phantom analysis to assess ADC bias, random error, gradient non linearity and DWI SNR which form the basis for (b) the site/system certification report to qualify for the advanced imaging arm of the NRG-BN001 trial.

The NRG-BN001 trial also involves dynamic susceptibility contrast (DSC) perfusion MRI for which we lack a dynamic phantom, therefore human subject DSC series are assessed for quality using: whole-brain SNR estimated via voxel-wise signal stability from pre-injection dynamics; peak change in signal upon bolus passage; and inspection of whole-brain leakage-corrected rCBV histogram (Figure 2D).

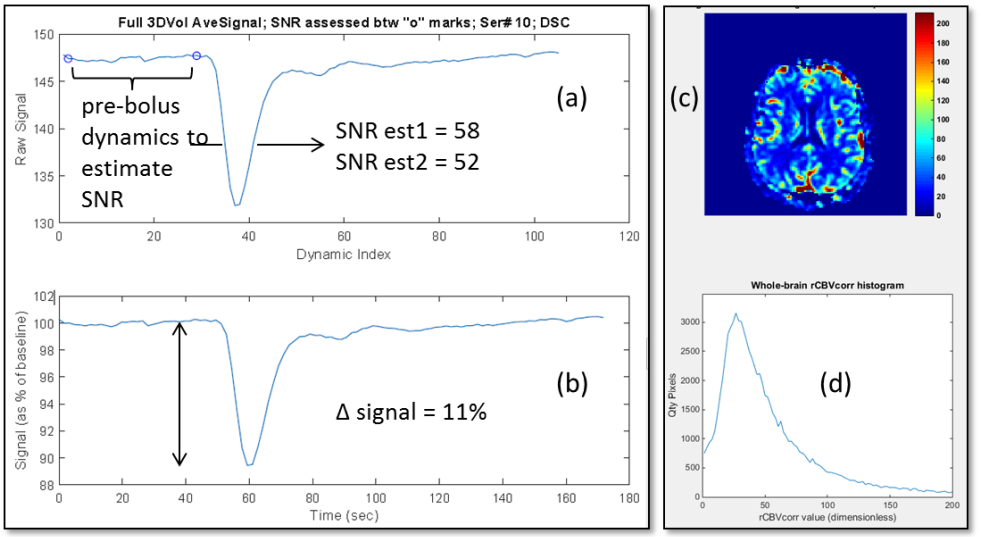


Figure 2D: Quality is assessed on each DSC series via (a) a SNR estimate using variance over pre-bolus timepoints measured for each voxel, averaged over the brain; (b) peak DSC contrast due to bolus passage; (c) a representative leakage-corrected rCBV map and (d) the whole-brain histogram of rCBVcorr.

§ Progress Toward Gradient Nonlinearity Correction

During the previous U01 cycle, our team has designed and launched a QIN-wide multi-center "phase 1" collaborative project within Data Acquisition Work Group (DAWG) to obtain empiric descriptors of gradient nonlinearity (GNL) along primary magnet directions on representative MRI systems utilized in clinical trials. As a result of "phase 1" project, channel-specific GNL was characterized for ten distinct gradient systems by three vendors [2]. In Year 1 of U01 renewal project, based on "GNL phase 1" results, the system-specific nonlinearity tensors were modeled using previously developed empiric approach [3], and the corresponding empiric 3D diffusion weighting bias maps were derived for six representative systems (e.g., Figure 3A), two from each vendor (GE, Siemens and Philips). The follow-up GNL correction validation "phase 2" project was launched within QIN DAWG to validate the empirically derived GNL correctors using independent phantom (FBIRN, 1.5% agar) scans outside of the (spatially limited) "phase 1" measurements performed with different DWI phantom (ice-water tube). In addition to test scans using arbitrary (orthogonal) DWI directions at two arbitrary off-center locations within the bore with substantial (predicted) GNL bias (> 10 %), the participating QIN sites obtained a "reference" measurement at bore isocenter representing the true diffusion coefficient free of GNL bias. The retrospective empirical correction was then applied to "test DWI" DICOM by our team, blinded to "reference DWI". The empirical GNL correctors were also compared to vendor design GNL characteristics provided for several systems. A convenient procedural simplification for isotropic phantom medium was that direction-average DW bias correctors are independent of DWI schema (LAB or non-LAB) and could be applied directly to ADC maps. Degree of similarity of ADC histograms from reference and corrected off-

center locations were used as a figure of merit for the retrospective correction. The performed correction improved uniformity of diffusion weighting for all systems at least three-fold, which lead to seven-fold gain in ADC precision and two-fold reduction of cross-site variability (Figures 3B and 3C). These results confirmed feasibility of centralized GNL-bias correction in multi-site trial setting warranted by general stability of system-specific GNL. The results of this project were published in QIN special issue of Tomography [4].

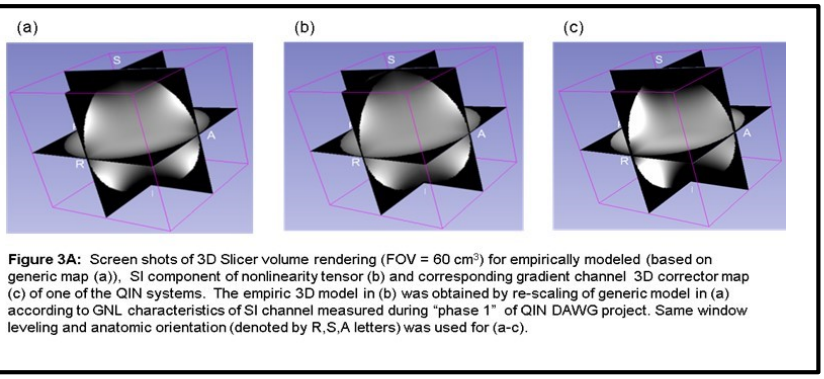


Figure 3A: Screen shots of 3D Slicer volume rendering (FOV = 60 cm^3) for empirically modeled (based on generic map (a)), SI component of nonlinearity tensor (b) and corresponding gradient channel 3D corrector map (c) of one of the QIN systems. The empiric 3D model in (b) was obtained by re-scaling of generic model in (a) according to GNL characteristics of SI channel measured during "phase 1" of QIN DAWG project. Same window leveling and anatomic orientation (denoted by R,S,A letters) was used for (a-c).

The automated retrospective correction tools were developed to recast the static (empiric) 3D correctors for arbitrary scan geometry as recorded in DICOM. These tools would help streamline analysis of multiple data sets from clinical trials. Although, the developed empiric correctors appear sufficient to remove the bulk of observed GNL bias, these correctors are only approximations of actual bias, best predicted based on system design coefficients known to vendors ([5], Figure 3B). Along with the finite accuracy of empiric GNL scaling, finite contribution from local shim (revealed during "phase 1" project [2]) and EPI distortion errors were identified as the main sources of residual ADC error. Another limitation of the proposed "retrospective" GNL correction approach is related to the type of DICOM data available from ongoing clinical trials. The clinical trials typically store only trace-DWI DICOM, which would preclude channel-specific GNL corrector application, desired especially for anisotropic tissue [2]. Nevertheless, for majority of breast DWI data targeted by this U01, the assumption of nearly isotropic tissue diffusion is valid, and direction-average corrector approach should be viable for ADC maps [5].

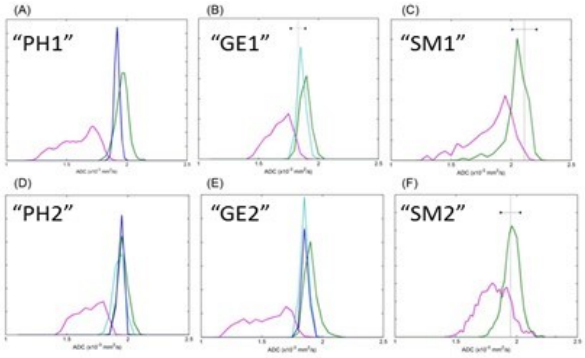


Figure 3B: ADC ROI histograms for all systems for a single coronal slice close to AP=0 and SI offset > 10 cm: "PH1", "GE1", "SM1" (in A-C), and "PH2", "GE2", "SM2" (in D-F). Magenta histograms indicate original ADC measurements (before correction), and green denotes corrected ADC maps, using generic (scaled) GNL models (as described in Methodology). Results of correction based on vendor-provided system GNL models are over-plotted (where available) as dashed cyan traces in (B, D, E). Reference ADC histograms, exported off-line by several sites, are shown in blue in (A, D, E), and reference values with corresponding 95% CI, independently measured on scanner console by other sites, are shown as vertical and horizontal black dotted lines, respectively.

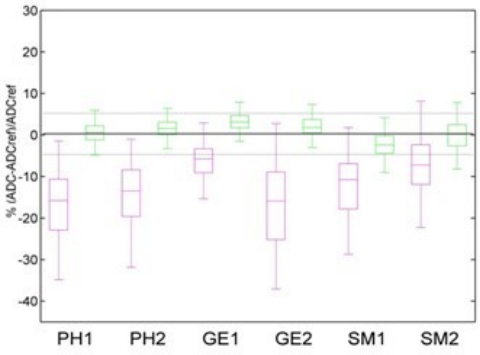


Figure 3C: Percent-bias box-plot summary for ADC ROI histograms in Figure 3 measured in respect to reference ADC for all systems before (magenta) and after (green) "scaled" GNL bias correction. Median ADC bias is marked with the central line inside the box. The edges of the box correspond to 25th and 75th percentiles, while whiskers encompass the 5th to 95th percentile data points. The dashed horizontal lines delineate ±5% error ranges.

PLANS FOR NEXT YEAR

§ Aim 2

We believe significant absolute temperature error stems from uncertainly in the calibration data, therefore we have altered our apparatus to allow simultaneous direct (i.e. non-MR) and NMR/MRI temperature measurement. The goal of this component of the proposal is to identify, fabricate, calibrate, and implement chemical systems that can be used as internal thermometers to measure absolute temperature in the MRI environment, using image features. To date, we have designed and implemented several systems having water, and methyl or methylene resonances with long T2 times. The dual frequency mixtures allow internal calibration and absolute temperature accuracy to approximately 1-2 °C in clinical MRI systems at 1.5 at 3.0 T. To improve this accuracy, we will explore the ability of chelates of the lanthanide metals europium, praseodymium, and thulium to create enhanced temperature sensitivity due to paramagnetic induced chemical shifts. These metals have short electron T1 times and have been used previously as chemical shift reagents to with minimal line broadening.

§ Aim 3

In Year 2 of the project, toward Aim3, we plan to identify a representative DWI data set from ACRIN 6698 breast clinical trial, acquired on (multiple) characterized systems with

validated GNL correction. The longitudinal phantom QC data for the corresponding systems will be analyzed to ensure consistency with the empirically modeled GNL bias and to establish SNR thresholds for retrospective GNL correction. The appropriate permissions will be obtained to request the de-identified DWI DICOM from the data managing center (UCSF) for retrospective correction for trial subjects scanned on the corresponding systems.

LIST OF QIN PUBLICATIONS, PRESENTATIONS AND COLLABORATIVE PROJECTS

[1] Galban, CJ, Ma B, Malyarenko D, Pickles MD, Heist K, Henry, NL, Schott AF, et al. Multi-Site Clinical Evaluation of DW-MRI as a Treatment Response Metric for Breast Cancer Patients Undergoing Neoadjuvant Chemotherapy. *PLOS one*. 2015; 10: e0122151. PMCID: PMC4376686

[2] Malyarenko DI, Newitt D, Wilmes LJ, Tudorica A, Helmer KG, Arlinghaus LR, Jacobs MA, Jajamovich G, Taouli B, Yankeelov TE, Huang W, Chenevert TL. Demonstration of Nonlinearity Bias in the Measurement of the Apparent Diffusion Coefficient in Multicenter Trials. Magn. Reson. Med. 2016; 75: 1312 (MRM March 2016 Editor's pick). PMCID: PMC4630210.

[3] Malyarenko DI and Chenevert TL. Practical Estimate of Gradient Nonlinearity for Implementation of ADC Bias Correction. J Magn. Reson. Imaging, 2014; 40: 487. PMCID: PMC4324556

[4] Malyarenko DI, Wilmes LJ, Arlinghaus LR, Jacobs MA, Huang W, Helmer KG, Taouli B, Yankeelov TE, Newitt D, Chenevert TL. QIN DAWG Validation of Gradient Nonlinearity Bias Correction Workflow for Quantitative Diffusion-Weighted Imaging in Multicenter Trials. Tomography 2016; 2:396

[5] Malyarenko DI, Pang Y, Senegas J, Ivancevic MK, Ross BD, Chenevert TL. Correction of Gradient Nonlinearity Bias in Quantitative Diffusion Parameters of Renal Tissue with Intra Voxel Incoherent Motion. Tomography, 2015; 1:145. PMCID: PMC 4724207

[6] Farahani K, Kalpathy-Cramer J, Chenevert TL, Rubin DL, Sunderland JJ, Nordstrom RJ, Buatti J, and Hylton N. Computational Challenges and Collaborative Projects in the NCI Quantitative Imaging Network. TOMOGRAPHY, December 2016, Volume 2, Issue 4: 242-249. DOI: 10.18383/j.tom.2016.00265

[7] Mulkern, RV, Ricci KI, Vajapeyam S, Chenvert TL, Malyarenko DI, Kocak M, Poussaint TY. Pediatric Brain Tumor Consortium multisite assessment of apparent diffusion coefficient z-axis variation assessed with an ice–water phantom. *Acad. Rad*, 2015; 22:363. PMCID: PMC4323850

[8 – QIN CCP] Huang W, Chen Y, Fedorov A, Li X, Jajamovich JH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, et. al. The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance

Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. Tomography, 2016; 2: 56.

[9] Galbán CJ, Lemasson B, Hoff BA, Johnson TD, Sundgren PC, Tsien C, Chenevert TL, Ross BD. Development of a Multiparametric Voxel-Based Magnetic Resonance Imaging Biomarker for Early Cancer Therapeutic Response Assessment. Tomography. 2015 Sep;1(1):44-52. PubMed PMID: 26568982; PubMed Central PMCID: PMC4643274.

[10] Galbán CJ, Hoff BA, Chenevert TL, Ross BD. Diffusion MRI in early cancer therapeutic response assessment. NMR Biomed. 2016 Jan 15;PubMed PMID: 26773848; NIHMSID: 746325.

Presentations:

Swanson SD, Malyarenko DI, and Chenevert TL. Quantitative temperature imaging in chemically designed phantoms. ISMRM'24 Annual Conference, Singapore, May 2016.

Malyarenko DI, Boss MA, Obuchowski NA and Chenevert TL. "QibaPhan" Software for Quality Assurance of Quantitative Diffusion Weighted Imaging in Multi-Center Trials. QIRR educational exhibit, RSNA '2016 Annual Meeting, Chicago, IL, Dec. 2016.

In addition, University of Michigan – Team 1 has been active member of the QIN MRI Subgroup and participant in ongoing QIN-wide Challenges and Collaborative Projects (CCP):

1) "ADC Challenge" – Led by David Newitt, UCSF
2) "ADC DICOM Challenge" – Led by Dariya Malyarenko, UMich – Team1
3) "DCE AIF Challenge" – Led by Wei Huang, OHSU
4) "DSC Challenge" – Led by Kathleen Schmainda, UWisc
5) "T1-Mapping Challenge" Led by Octavia Bane, Mount Sinai
6) "DSC DRO" – Led by Chad Quarles, Barrows Neurological Institute

CA 160045: Objective Decision Support Environment for Clinical Trials

Mayo Clinic

Bradley Erickson, M.D., Ph.D.

INTRODUCTION

The goal of our grant is to apply machine learning techniques to objectively identify those imaging features that best characterize tumors--either tumor biology or therapy response. In particularly, we aimed to develop virtual biopsy technology that will enhance the clinical decision making process in brain cancer by providing tools for investigation of image-based therapy response assessment. We anticipate this project will impact clinical trials by enabling identification of superior outcome measures using objective algorithmic selection methods.

There is significant potential for Machine Learning (ML) to improve how we use imaging in clinical trials to assess therapy. Multi-spectral MRI processing is not new, but using ML methods to identify more complex relationships than simple linear ones could be important, and is novel in the therapy assessment and Radiogenomics space. There has been relatively little work focusing on the estimation of the information content of features in medical images

The specific deliverables for the grant include:

1. A library of easily applied tools for computing both widely used standard features and biologically relevant features from DICOM images. We expect this library should be usable by QIN community.
2. An interactive tool for applying several FS methods to identify the most informative features and best performing machine learning methods for the selected feature set. We expect this tool should be usable by QIN community.
3. A family of decision support applications for three clinical situations.

DISCUSSION OF PROGRESS

§ Automated Segmentation of hyperintense regions in FLAIR MRI

Brain tumor segmentation is a challenging task with many researchers and competitions focusing on creating and evaluating newly developed algorithms. In 2012, the brain tumor image segmentation benchmark (BRATS) competition (1,2) was established as part of the MICCAI conference and since then has been the “gold standard” for brain segmentation algorithm testing.

The brain tumor segmentation algorithms commonly described in the literature usually exploit classical image analysis techniques or pattern recognition techniques (3–5) with the more recent approaches utilizing deep convolutional neural networks (6–13).

Each MRI series (image type) reveals different information about the tumor. For instance, T1-weighted (T1w) post contrast acquisitions reveal information regarding the enhancing part of the tumor, while fluid-attenuated inversion recovery (FLAIR) acquisitions capture the edema part of the tumor. Lesion size in FLAIR images is an important clinical parameter for patient assessment and follow-up. Manual estimation of the volume of the lesions in FLAIR images is time-consuming and highly user-dependent.

Autoencoders have recently been gaining attention for their ability to perform segmentation tasks in medical images (14–16). One advantage of autoencoders against other deep learning approaches is the use of decoders that enables estimation of features suitable for pixel-wise classification (16).

Over the past year, our team focused on accurate quantification of the abnormal signal areas in the FLAIR acquisitions in glioma patients. For the purpose of this study, we utilized convolutional autoencoders trained on the publically available BRATS dataset and evaluate the accuracy on a dataset where three expert segmentations were available. Figure 1 captures the main idea of an autoencoder and its application to image segmentation. The primary concept is that the autoencoder learns how to reconstruct the segmented desired output (namely the segmentation mask). The encoder layer consists of 7 convolutional layers. The convolutions are used to produce the feature maps. In addition, a rectified-linear non-linearity (ReLU) is applied followed by maxpooling with a 2 × 2 window. The resulting output is sub-sampled by a factor of 2. Max-pooling achieves translation invariance, accounting for small spatial shifts. The decoder component consists of a hierarchy of decoders, one corresponding to each encoder. Of these, the appropriate decoders use the max-pooling indices received from the corresponding encoder to perform non-linear up-sampling of their input feature maps. This allows for improved boundary delineation (16). The high decoder output is forwarded to a trainable soft-max classifier which classifies each pixel independently. The number of input channels is the number of classes (in our case, tumor or not tumor) and the output of the sigmoid classifier is a 2 channel image of probabilities. The predicted segmentation corresponds to the class with maximum probability at each pixel.

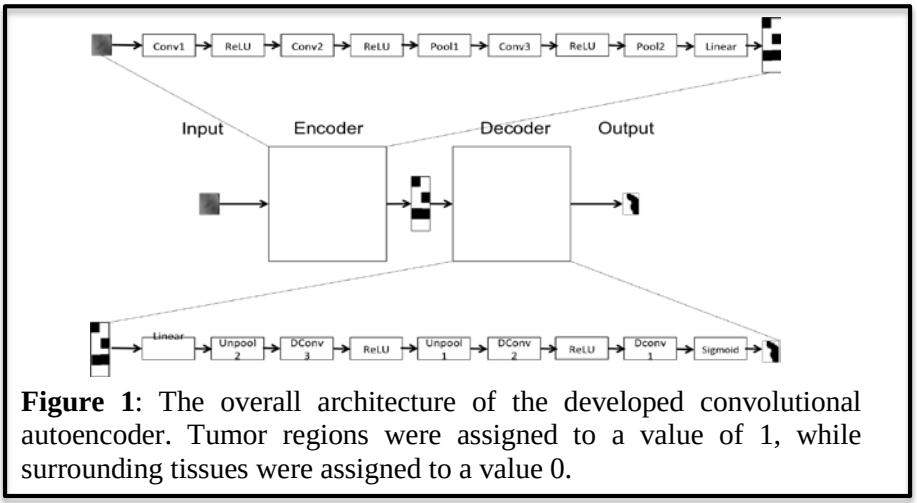


Figure 1: The overall architecture of the developed convolutional autoencoder. Tumor regions were assigned to a value of 1, while surrounding tissues were assigned to a value 0.

Comparison of the proposed method and the three manual segmentations available against the STAPLE algorithm is shown in Table 1.

Measure	Statistic	User 1	User 2	User 3	Proposed
Jaccard	*Mean*	0.923	0.840	0.758	0.785
	SD	0.051	0.077	0.057	0.095
	Max	1.000	1.000	0.865	0.917
	Min	0.760	0.550	0.649	0.458
	Median	0.931	0.856	0.747	0.821
	Q1	0.901	0.815	0.711	0.729
	Q3	0.957	0.879	0.809	0.849
Dice	*Mean*	0.959	0.911	0.861	0.876
	SD	0.029	0.048	0.037	0.066
	Max	1.000	1.000	0.928	0.957
	Min	0.864	0.710	0.787	0.629
	Median	0.964	0.922	0.855	0.901
	Q1	0.948	0.898	0.831	0.843
	Q3	0.978	0.935	0.895	0.919
FPF	*Mean*	0.079	0.198	0.190	0.291
	SD	0.055	0.135	0.111	0.210
	Max	0.253	0.819	0.460	1.181
	Min	0.000	0.000	0.020	0.090
	Median	0.070	0.164	0.169	0.219
	Q1	0.044	0.136	0.100	0.172
	Q3	0.101	0.227	0.275	0.370
TPF	*Mean*	0.993	0.996	0.899	0.995
	SD	0.032	0.015	0.062	0.016
	Max	1.000	1.000	0.994	1.000
	Min	0.793	0.923	0.720	0.931
	Median	1.000	1.000	0.895	1.000
	Q1	1.000	1.000	0.860	1.000
	Q3	1.000	1.000	0.956	1.000

Table 1: Comparison of the proposed method and the three manual segmentations available against the STAPLE algorithm.

Figure 2 captures the probabilistic output generated from the autoencoder for an input image.

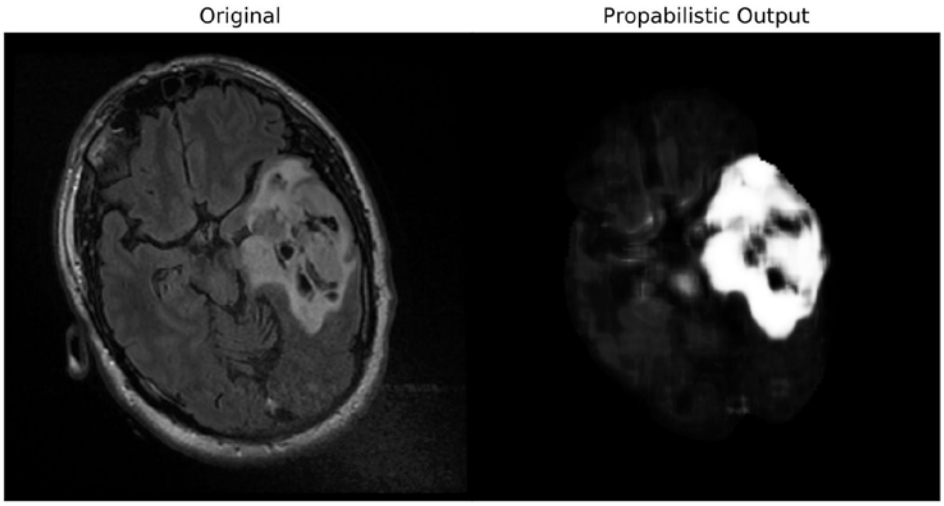


Figure 2. Probabilistic output of the proposed system for a case in our validation dataset. As depicted (right panel) tumor areas appear to be brighter than surrounding tissue.

The proposed automated system is indistinguishable from expert derived segmentations in its ability to perform glioma segmentation. This approach will be useful for alleviating the inherent variability of human derived tumor delineation thereby improving the reproducibility of image-derived biomarkers.

§ Grunt--a Flexible Pipeline Technology

Docker (https://www.docker.com/) is an open source technology that allows one to capture a complete execution environment as a file that can then be executed on any Docker host platform (which can be LINUX, MacOS or Windows Server). This is much like virtual machine technology, but has much lower computing requirements.

In research, we frequently think of 'pipelines' where a series of tools are applied to a dataset, producing a final output at the end of the pipe. With Docker technology, it is feasible to connect a number of tools (Dockers) together that might otherwise not be compatible. One minor challenge to this approach is providing access to the image files to process, and the result, in a secure and controlled fashion. Security is increasingly recognized as an important part of proper research computing, and others working on pipelines have largely ignored security, and we believe that will become a critical error.

To leverage Dockers while addressing the security issue, we propose deploying the image analysis algorithms as web application and interact with them though a RESTful Application Programming Interface (API). We have extended an open source software tool called Grunt (https://githuib.com/Mayo-QIN/grunt). The aim of Grunt is to simplify the creation and deployment of web apps utilizing Dockers with an easy and well-documented connection method.

The only documentation needed is a description of the endpoints of the RESTful api basically the description of the required algorithm inputs. Grunt can be configured based on configuration files (yml or cli- - enables compatibility with existing tools like Slicer)). The configuration files consist of multiple services described as endpoints of the RESTful api.

By deploying the service through a RESTful api also enables the researched to leverage cloud architectures. Since one institution can create the grunt based app publish it in a private or public cloud and subsequently potential collaborators would be able to stream the data to the service and retrieve the results. A very crucial requirement for creating automated pipelines. Grunt also contains a web interface where users can check the jobs running. Furthermore a job scheduler is provided for long running jobs with functionality for monitoring and notification when the jobs are finished.

§ Predicting MGMT methylation status utilizing machine learning

Glioblastoma multiforme tumors (GBMs) with Methylguanine methyltransferase (MGMT) promoter methylation can be expected to respond better to an alkylating agent like temozolomide(17). In addition, MGMT methylation may be considered as a predictive biomarker for a patient's desirable response to radiation therapy. Several reports in the literature indicate that MGMT promoter methylation is associated with longer survival(18). However, while determination of MGMT methylation status has been standard of care for some time, an accurate result is not always obtained due to the requirement of large tissue specimens.

A retrospective study of 155 GBM patients with known MGMT methylation status was conducted. Co-occurrence and run length texture features were calculated and both support vector machines (SVMs) and random forest classifiers (RFCs) were used to predict MGMT methylation status.

The best classification system (an SVM-based classifier) had a maximum area under the ROC curve (AUC) of 0.85 (95% CI: 0.78 to 0.91) using four texture features (correlation, energy, entropy, and local intensity) originating from the T2-weighted images, yielding at the optimal threshold of the ROC curve a sensitivity of 0.803 and a specificity of 0.813 (Tables 2, 3).

Results show that supervised machine learning of MRI texture features can predict MGMT methylation status in preoperative GBM tumors, thus providing a new noninvasive imaging biomarker.

# Selected Features	Classifier parameter (estimator)	Window filter size	Gray Level (GL)	A_z
10	50	3	16	0.82 95% CI: 0.662 to 0.849
7	10	3	64	0.7 95% CI: 0.521 to 0.732
13	10	3	128	0.756 95% CI: 0.432 to 0.798
7	**100**	**5**	**32**	**0.84** 95% CI: 0.757 to 0.892

Table 2: Results from random forest classifier (RFC) with feature extracted from T2 images (best performing system is in bold).

# Selected Features	Classifier parameter (estimator)	Window filter size	Gray Level (GL)	A_z
4	C: 10.0, σ: 0.1	3	16	0.83 95% CI: 0.637 to 0.867
4	**C: 10.0,** **σ: 0.01**	**3**	**32**	**0.85** 95% CI: 0.782 to 0.913
4	C: 1.0, σ: 1e-03	3	64	0.78 95% CI: 0.594 to 0.804
4	C: 1.0, σ: 1.1e-03	5	16	0.78 95% CI: 0.633 to 0.821
4	C: 1.0, σ: 1.0	5	16	0.8 95% CI: 0.512 to 0.822
8	C: 10.0, σ: 1e-04	5	16	0.76 95% CI: 0.422 to 0.824
4	C: 100.0, σ: 1e-02	5	64	0.75 95% CI: 0.410 to 0.816

Table 3: Results from support vector machine (SVM) with feature extracted from T2 images (best performing system is in bold).

IMAGE ANALYSIS TOOLS FOR LOW GRADE GLIOMAS

We have developed a complete software package which includes several brain image analysis tools such as rigid, nonrigid, and atlas image registration, bias field correction, skull striping, and semi-automated segmentation algorithm for low grade glioma (LGGs) (19) and shared it publicly for other QIN members (https://github.com/aqqush/LGG_Software). We have also developed a deep learning based classification approach for predicting chromosomal arms 1p/19q deletion from MRI images (Akkus et al. 2016). Furthermore, our work on fully-automated segmentation of LGGs in pre- and post- operative images using deep learning and assessment of LGGs progression are still ongoing.

§ Semi-automated Segmentation of Preoperative Low Grade Gliomas

Segmentation of pre-operative LGGs from magnetic resonance imaging is a crucial step for studying imaging biomarkers. However, segmentation of LGGs is particularly challenging because they rarely enhance after gadolinium administration. Like other gliomas, they have irregular tumor shape, heterogeneous composition, ill-defined tumor boundaries, and limited number of image types. To overcome these challenges we propose a semi-automated segmentation method that relies only on T2-weighted (T2W) and optionally post-contrast T1-weighted (T1W) images. First, the user draws a region-of-interest (ROI) that completely encloses the tumor and some normal tissue. Second, a normal brain atlas and post-contrast T1W images are registered to T2W images. Third, the posterior probability of each pixel/voxel belonging to normal and abnormal tissues is calculated based on information derived from the atlas and ROI. Finally, geodesic active contours use the probability map of the tumor to shrink the ROI until optimal tumor boundaries are found. This method was validated against the true segmentation (TS) of 30 LGG patients for both 2D (1 slice) and 3D. The TS was obtained from manual segmentations of three experts using the Simultaneous Truth and Performance Level Estimation (STAPLE) software. Dice and Jaccard indices and other descriptive statistics were computed for the proposed method, as well as the experts' segmentation versus the TS. We also tested the method with the BraTS datasets, which supply expert segmentations. For 2D segmentation vs. TS, the mean Dice index was 0.90 ± 0.06 (standard deviation), sensitivity was 0.92, and specificity was 0.99. For 3D segmentation vs. TS, the mean Dice index was 0.89 ± 0.06, sensitivity was 0.91, and specificity was 0.99. The automated results are comparable with the experts' manual segmentation results. We present an accurate, robust, efficient, and reproducible segmentation method for preoperative LGGs.

§ Predicting Chromosomal Arms 1p19q Codeletion from MRI images

In this study, we predict the 1p/19q chromosomal arm deletion from MR images using convolutional neural networks (CNN), which could be a noninvasive alternative to surgical biopsy and histopathological analysis. Our method consists of three main steps: image registration, tumor segmentation, and classification of 1p/19q status using CNN. We included a total of 159 LGG subjects (57 nondeleted and 102 codeleted) and preoperative postcontrast-T1 (T1C) and T2 images. The T1-weighted images were rigidly registered to the T2 images. For all images, the image where the tumor had the largest cross-sectional area as well as the slice immediately above and below were segmented using the semi-automated tool that we developed above. We divided our data

into training, validation, and test sets. The training data was balanced for equal class probability. We used data augmentation, including random translational shift, rotation, and horizontal and vertical flips to increase the size of the training set at each epoch. Finally, we evaluated several configurations of a multi-scale CNN architecture until training and validation accuracies became consistent. We also compared the performance of our method to the performance of a classical machine-learning algorithm using support vector machine (SVM) classifier with greedy feature selection. Using seven selected features (from intensity-based features, local binary patterns, Gabor filters, Laplacian of Gaussian, gray-level co-occurrence matrix, and boundary sharpness) the SVM classifier was trained and tested on the same data. The multiscale CNN overfits the original (limited size) data when data augmentation is not used. The training accuracy was 100% for both the training and validation sets, but remained below 80% for the test data. The results of the best performing configuration on the unseen test set were 96% (sensitivity), 82% (specificity), and 89% (accuracy). The results of the SVM on the test set were 80% (sensitivity), 82% (specificity), and 81% (accuracy). Multi-scale CNN, which learns a hierarchy of complex features directly from raw image data with their self-learning capability, provides promising results for predicting 1p/19q status noninvasively based on T1C and T2 images.

PLANS FOR NEXT YEAR

Our aim for the last year of our grant is to fully focus on deep learning and more specifically in its application in segmentation and prediction of genomics utilizing MRI data. We have obtained a large collection of LGGs from UCSF used in a recent paper that has genomic data. Furthermore we are planning to compare the performance of the deep learning architectures with traditional machine learning approaches.

We are also planning to extend the functionality of the Grunt pipeline work with BIDS group to ensure that our tool is compatible with the pipeline tools available in QIN.

LIST OF REFERENCES

1. Ellingson BM, Wen PY, van den Bent MJ, Cloughesy TF. Pros and cons of current brain tumor imaging. Neuro Oncol. 2014 Oct;16 Suppl 7:vii2–11.
2. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging. 2015 Oct;34(10):1993–2024.
3. Xiao Y, Zhou LH, Zhen X. FISER: Feature Image Space Enhanced Random Walker Algorithm for Brain Tumor Segmentation in Multimodal MR Images. Journal of Medical Imaging and Health Informatics. 2015;5(8):1977–81.
4. Demirhan A, Ayse D, Mustafa T, Inan G. Segmentation of Tumor and Edema Along With Healthy Tissues of Brain Using Wavelets and Neural Networks. IEEE Journal of Biomedical and Health Informatics. 2015;19(4):1451–8.
5. Menze BH, Van Leemput K, Danial L, Tammy R-R, Ezequiel G, Esther A, et al. A Generative Probabilistic Model and Discriminative Extensions for Brain Lesion Segmentation— With Application to Tumor and Stroke. IEEE Trans Med Imaging. 2016;35(4):933–46.

6. Havaei M, Mohammad H, Francis D, Chris P, Hugo L, Pierre-Marc J. A Convolutional Neural Network Approach to Brain Tumor Segmentation. In: Lecture Notes in Computer Science. 2016. p. 195–208.
7. Pereira S, Sergio P, Adriano P, Victor A, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Trans Med Imaging. 2016;35(5):1240–51.
8. Havaei M, Mohammad H, Axel D, David W-F, Antoine B, Aaron C, et al. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017;35:18–31.
9. Bricq S, Ch. C, Armspach JP. Unifying framework for multimodal brain MRI segmentation based on Hidden Markov Chains. Med Image Anal. 2008;12(6):639–52.
10. Li Y, Yuhong L, Qi D, Jinze Y, Fucang J, Jing Q, et al. Automatic Brain Tumor Segmentation from MR Images via a Multimodal Sparse Coding Based Probabilistic Model. In: 2015 International Workshop on Pattern Recognition in NeuroImaging [Internet]. 2015. Available from: http://dx.doi.org/10.1109/prni.2015.18
11.Maier O, Oskar M, Matthias W, Heinz H. Image Features for Brain Lesion Segmentation Using Random Forests. In: Lecture Notes in Computer Science. 2016. p. 119–30.
12. Islam A, Reza SMS, Iftekharuddin KM. Multifractal texture estimation for detection and segmentation of brain tumors. IEEE Trans Biomed Eng. 2013 Nov;60(11):3204–15.
13. Kamnitsas K, Ledig C, Newcombe V, Simpson JP, Kane AD, Menon DK, et al. Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation [Internet]. 2016. Available from: http://hdl.handle.net/10044/1/30897
14. Zhao G, Wang X, Niu Y, Tan L, Zhang S-X. Segmenting Brain Tissues from Chinese Visible Human Dataset by Deep-Learned Features with Stacked Autoencoder. Biomed Res Int. 2016 Jan 26;2016:5284586.
15. Shin H-C, Orton MR, Collins DJ, Doran SJ, Leach MO. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans Pattern Anal Mach Intell. 2013 Aug;35(8):1930–43.
16. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [Internet]. arXiv [cs.CV]. 2015. Available from: http://arxiv.org/abs/1511.00561
17. Rivera AL, Pelloski CE, Gilbert MR, Colman H, De La Cruz C, Sulman EP, et al. MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma. Neuro Oncol. 2010 Feb;12(2):116–21.
18. Zhang K, Wang X-Q, Zhou B, Zhang L. The prognostic value of MGMT promoter methylation in Glioblastoma multiforme: a meta-analysis. Fam Cancer. 2013 Sep;12(3):449–58.
19. Akkus Z, Sedlar J, Coufalova L, Korfiatis P, Kline TL, Warner JD, et al. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging. Cancer Imaging. 2015 Aug 14;15:12.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

1. Kelm ZS, Korfiatis P, Lingineni RK, Daniels JR, Buckner JC, Lachance DH, Parney IF, Carter RE, Erickson BJ. Variability and accuracy of different software packages for dynamic susceptibility contrast magnetic resonance imaging for distinguishing glioblastoma progression from pseudoprogression. J Med Imaging. 2015 Apr; 2(2): 026001 doi: 10.1117/1.JMI.2.2.026001. PMID 26158114, PMCID:PMC4478857

2. Akkus Z, Sedlar J, Coufalova L, Korfiatis P, Kline TL, Warner JD, Agrawal J, Erickson BJ. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging. Cancer Imaging. 2015 Aug 14;15(1):12. doi: 10.1186/s40644-015-0047-z. PMID: 26268363, PMCID:PMC4535671
3. Korfiatis P, Kline TL, Coufalova L, Lachance DH, Parney IF, Carter RE, Buckner JC, Erickson BJ. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. Med Phys. 2016 Jun;43(6):2835. doi: 10.1118/1.4948668. PMID: 27277032, PMCID:PMC4866963
4. Korfiatis P, Kline TL, Blezek DJ, Langer SG, Ryan WJ, Erickson BJ. PESSCARA:: A Content Management System for medical image analysis research. Radiographics. 2015 Sep-Oct; 35 (5): 1461-8. PMCID:PMC26284301
5. Korfiatis P, Kline TL, Erickson BJ. Automated Segmentation of Hyperintense Regions in FLAIR MRI Using Deep Learning. Tomography. PMC Journal - In Process.
6. Korfiatis P, Kline TL, Kelm ZS, Hu LS, Erickson BJ. Dynamic Susceptibility Contrast MRI Quantification Software Tool: Development and Evaluation. Tomography. PMC Journal - In Press.
7. Peter R, Korfiatis P, Blezek DJ, Beitia O, Stepan-Buksakowska I, Horinek D, Flemming K, Erickson BJ. Quantitative Symmetry-based Analysis of Hyperacute Ischemic Stroke Lesions in Non-Contrast Computed Tomography. Medical Physics. PMC Journal - In Press.
8.Erickson BJ, Korfiatis P, Akkus Z, Kline T. Machine Learning For Medical Imaging. Radiographics, March 2017. PMC Journal - In Process.
9. Korfiatis P, Erickson B. The basics of diffusion and perfusion imaging in brain tumors. Appl Radiol 2014 Jul; 43 (7):22-29 Epub 2014 July 04 PMID: 26456989 PMCID: 4599787
10. Korfiatis P, Erickson BJ. PESSCARA: An Example Infrastructure for Big Data Research In: Big Data on Real-World Applications. (Book chapter) DOI: 10.5772/63815
11.Kline TL, Korfiatis P, Edwards ME, Warner JD, Irazabal MV, King BF, Torres VE, Erickson BJ. Automatic total kidney volume measurement on follow-up magnetic resonance images to facilitate monitoring of autosomal dominant polycystic kidney disease progression. Nephrol Dial Transplant 2016 Feb; 31 (2):241-8 Epub 2015 Aug 31 PMID: 26330562 PMCID: 4725388 DOI: 10.1093/ndt/gfv314
12.Kline TL, Edwards ME, Korfiatis P, Akkus Z, Torres VE, Erickson BJ. Semiautomated Segmentation of Polycystic Kidneys in T2-Weighted MR Images. AJR Am J Roentgenol 2016 Sep; 207 (3):605-13 Epub 2016 June 24 PMID: 27341140 PMC Journals - In Process.

CONFERENCE ABSTRACTS

1. Korfiatis, P., Hu, L., Kelm, Z., Erickson, B., A DSC Digital Brain Phantom for Assessment of Leakage Correction Methods. Radiological Society of North America 2013 Scientific Assembly and Annual Meeting, December 1 - December 6, 2013 ,Chicago IL
2. Korfiatis, P, Hu, L, Kelm, Z, Erickson, B, Perfusion Analysis Software Accuracy Evaluation: A Digital Phantom Based Study. Radiological Society of North America 2013 Scientific Assembly and Annual Meeting, December 1 - December 6, 2013 ,Chicago IL.
3. Kelm, Z., Hu, L., Korfiatis, P., Lingineni, R., Carter, R., Daniels, J., Erickson, B., Variability of rCBV Measurements of Glioblastoma between Three FDA-approved

Software Packages. Radiological Society of North America 2013 Scientific Assembly and Annual Meeting, December 1 - December 6, 2013 ,Chicago IL.
4. Kline L.T., Korfiatis P, Warner DJ., Torres EV., King FB., Erickson BJ., Making Medical Image Analysis Research Easier, Faster, and More Reproducible with IPython. Scientific Computing with Python conference 2014, July 6th-12th 2014, Austin, Texas
5. Korfiatis P., Kline LT., Blezek JD., Warner DJ., Langer GS., Erickson BJ., Platform to Enable Shared Scientific Computing Advances Research Assets (PESSCARA). Scientific Computing with Python conference 2014, July 6th-12th 2014, Austin, Texas
6. Kaufmann TJ, Erickson BJ, Agrawal J, Korfiatis P , Carter RE. NIMG-26comparison of post-gadolinium 3D fast spin echo images versus registered and subtracted mprage images in glioma patients. Society for Neuro-Oncology 20th Annual Scientific Meeting (2015).
7. Korfiatis P. Kline TL. Erickson BJ. Residual Deep Convolutional Neural Network Predicts Management Methylation Status. Machine Intelligence in Medical Imaging Alexandria, Virginia (2016).
8. Akkus Z, Sedlar J., Ali I., Giannini G., Parney I, Erickson B.J.. Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images using Machine Intelligence. Conference on Medical Intelligence in Medical Imaging. Sep 2016
9. Akkus Z, Ali I, Sedlar J., Kline TL., Agrawal J.P, B. Erickson J.. Predicting Response of Low Grade Gliomas to Therapy from MRI Images using Convolutional Neural Networks (CNNs). RSNA 2016
10. Erickson BJ, Korfiatis P, Akkus Z, Kline T. Man vs Machine: Hands-on tutorial on Machine Learning. RSNA 2016

1 U01 CA172320-04: Evaluation of HCC Response to Systemic Therapy with Quantitative MRI

Icahn School of Medicine at Mount Sinai

Bachir Taouli, MD

INTRODUCTION

§ Specific Aim 1

Develop a framework for quality control (QC) in functional MRI of the liver in patients with hepatocellular carcinoma (HCC).

§ Specific Aim 2

Validate a quantitative multiparametric scoring system combining measurements of MR diffusion, perfusion and hypoxia against histopathologic measures of tumor grade/cellularity, aggressiveness, angiogenesis and hypoxia in human HCC.

Specific aims 1 and 2 are as initially stated, and not modified.

§ Specific Aim 3

Validate new imaging response parameters based on multiparametric quantitative MRI in patients with advanced HCC treated with sorafenib in an independent study. We have slightly modified specific aim 3 by changing the treatment from sorafenib to Yttrium 90 radioembolization.

Updated specific aim 3 reads as follows: Validate new imaging response parameters based on multiparametric quantitative MRI in patients with advanced HCC treated with Y90 radioembolization in an independent study. Rationale for the modification: we have decided to switch to radioembolization, as it is a more effective treatment than sorafenib.

DISCUSSION OF PROGRESS

§ Specific Aim 1

Data on repeatability has been reported last year and published (1-3).

§ Specific Aim 2

R1 and R_2* measurement under oxygen and carbogen challenge: data already reported last year and published in (1).

IVIM DWI results and correlation with DCE-MRI: data already reported last year and published in (2).

DCE-MRI quantification with Tofts model vs. shutter-speed model: data already reported last year and published in (3).

§ New Data

Quantification of HCC tumor heterogeneity using mpMRI [manuscript submitted]

Introduction: Many studies that employ mpMRI to assess/predict tumor response use central tendency parameters, such as mean or median, over entire regions of interest (ROIs) to determine longitudinal changes in tumor tissue after treatment (5). However, such analysis may not represent the exact tumor status, given the intrinsic heterogeneous tumor composition (5). Heterogeneity analysis of tumor MRI measurements may provide accurate markers of tumor heterogeneity at the genetic, cellular and molecular levels (5) and thereby allow for a better understanding of tumor characteristics that affect treatment. HCC lesions are known to exhibit substantial intra- and inter-tumor heterogeneity, due to a large variety in etiological and genetic backgrounds and the longtime development of the disease (6). Tumor heterogeneity poses a significant challenge for treatment stratification. While morphological and genetic heterogeneity in HCC lesions has been assessed previously (7), imaging reports on HCC heterogeneity are extremely limited (8). Tumor imaging phenotypes, including histogram features, potentially correlate with the underlying genotype and subsequently noninvasive imaging, including MRI, can be used as a surrogate for genomics and transcriptomics (radiogenomics) (9). Recently, there has been considerable interest in immunotherapy of a wide variety of cancers, including HCC (10). The success of such treatment heavily depends on tumor expression of immunotherapy targets, such as immune checkpoints. Identification of imaging features that correlate with gene expression of immunotherapy targets potentially allows for noninvasive prediction of immunotherapy outcome at baseline.

Purpose: To quantify tumor heterogeneity in HCC using mpMRI, and to report preliminary data correlating quantitative MRI parameters with histopathology and gene expression in a subset of patients.

Materials and Methods: We included 32 HCC patients (M/F 26/6, mean age 59y) who underwent mpMRI including DWI, BOLD, TOLD and DCE-MRI. Histogram characteristics [central tendency (mean, median) and heterogeneity (standard deviation, kurtosis, skewness) MR imaging parameters] in HCC and liver parenchyma were compared using Wilcoxon signed-rank tests. Inter-tumor heterogeneity was assessed using the

coefficient of variation between histogram features across tumors. Histogram data was correlated between MRI methods in all patients and with histopathology and gene expression in 14 patients.

Results: 39 HCC lesions were assessed (mean size 4.4±3.3 cm). HCCs exhibited significantly higher intra-tissue heterogeneity vs. liver with all MRI methods (P<0.042). Inter-tumor heterogeneity was significantly higher for kurtosis and skewness vs. mean parameters (P<0.001). While there were significant correlations for central tendency parameters between MRI methods and with each of histopathology and gene expression, heterogeneity parameters exhibited additional complementary correlations between BOLD and DCE-MRI and with histopathologic hypoxia marker HIF1α and gene expression of Wnt target *GLUL*, pharmacological target *FGFR4*, stemness markers *EPCAM* and *KRT19* and immune checkpoint *PDCD1*.

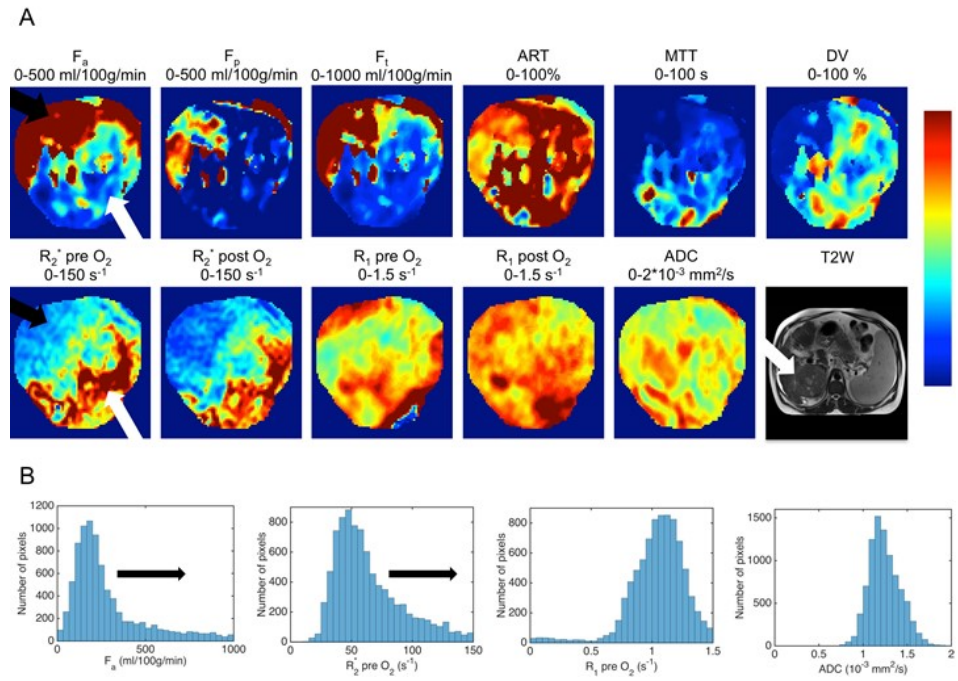


Figure 1: 54 year-old male patient with HBV cirrhosis and HCC. A) Representative magnified parametric maps (DCE-MR imaging, BOLD, TOLD and ADC) of a large (8.3 cm) HCC. Location of the tumor within the liver is indicated by the white arrow on the T_2-weighted image (bottom row, right). A distinct region in the anterior portion of the tumor of high arterial flow (F_a) and low R_2* was observed, reflective of high tumor perfusion and normoxia (black arrow in F_a and R_2* pre O_2 maps). The posterior portion of the tumor displays low F_a and high R_2*, suggestive of poor perfusion and hypoxia (white arrow in F_a and R_2* pre O_2 maps). B) Histograms of F_a, R_2* pre O_2, R_1 pre O_2 and ADC in the same lesion. The extensive heterogeneity observed in the parameter maps of F_a and R_2* pre O_2 is also reflected in the histograms, as illustrated by the fat tails and pronounced skewness, indicated by the black arrows. ADC = apparent diffusion coefficient, ART = arterial fraction, DV = distribution volume, F_a, arterial flow, F_p = portal flow, F_t = total flow, MTT = mean transit time, R_1 = longitudinal relaxation rate, R_2* = transverse relaxation rate.

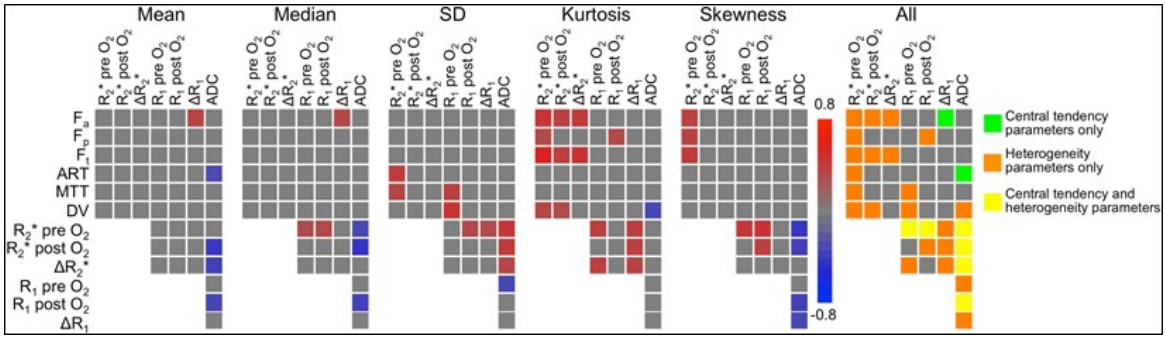


Figure 2: Heatmaps of correlations between mean, median, SD, kurtosis and skewness of MR imaging parameters in 39 HCC lesions (in 32 patients). Significant correlations (P<0.05) are colored according to the scale bar. A combined heatmap of all significant correlations between MR imaging features is shown on the right, illustrating additional information provided by heterogeneity parameters (SD, kurtosis and skewness). Significant correlations between DCE-MR imaging and BOLD were for example only seen for heterogeneity parameters and not for central tendency parameters (mean and median). ADC = apparent diffusion coefficient, ART = arterial fraction, DV = distribution volume, F_a, arterial flow, F_p = portal flow, F_t = total flow, MTT = mean transit time, R_1 = longitudinal relaxation rate, R_2* = transverse relaxation rate.

Conclusions: Histogram analysis combining central tendency and heterogeneity mpMRI features is promising for noninvasive HCC characterization on the functional, histologic and genomics level.

	Mean			Median			SD			Kurtosis			Skewness		
	Liver	**HCC**	***P****	**Liver**	**HCC**	***P****	**Liver**	**HCC**	***P****	**Liver**	**HCC**	***P****	**Liver**	**HCC**	***P****
F_a	49±45	247±240	**<0.001**	47±45	241±250	**<0.001**	16.3±13.0	83.0±66.7	**<0.001**	4.3±1.8	5.5±6.0	0.754	0.76±0.66	0.44±1.34	0.445
F_p	145±123	112±156	**0.048**	146±126	103±166	**0.01**	39.9±32.5	78.6±78.6	**0.02**	6.3±14.0	64.1±183.6	**0.005**	0.54±1.57	4.00±6.41	**<0.001**
F_t	193±122	358±337	**0.023**	193±125	342±350	0.092	43.7±36.2	144.3±126.3	**<0.001**	4.5±3.1	17.1±54.9	0.969	0.62±0.76	1.14±3.11	0.378
ART	33.9±27.9	78.9±21.5	**<0.001**	33.0±29.1	80.3±24.2	**<0.001**	9.7±5.1	13.1±9.0	0.055	7.6±10.5	30.1±61.8	0.256	0.68±1.71	-2.49±4.37	0.085
MTT	20.3±12.3	21.2±15.8	0.724	19.7±12.1	18.8±16.3	0.667	4.2±2.9	10.9±7.3	**<0.001**	5.2±4.8	9.1±17.4	0.938	0.90±0.87	1.54±2.06	0.273
DV	37.4±22.7	34.6±24.7	0.695	36.8±22.8	34.0±26.0	0.71	4.9±3.4	9.9±5.7	**<0.001**	3.7±1.7	5.5±10.8	0.24	0.58±0.61	0.70±1.59	0.799
R_2* pre O_2	55.3±34.4	34.5±15.1	**<0.001**	54.7±34.8	33.1±15.0	**<0.001**	7.6±7.5	7.6±5.2	0.681	4.0±2.9	5.2±4.3	0.433	0.35±0.89	0.92±1.09	**0.042**
R_2* post O_2	51.8±30.8	33.7±14.5	**<0.001**	51.5±31.4	32.7±14.6	**<0.001**	8.3±10.7	8.0±4.7	0.347	3.5±1.8	4.0±2.7	0.984	0.15±0.64	0.72±0.81	0.075
ΔR_2*	-3.47±10.0	-0.78±8.74	0.281	-3.12±8.8	-0.86±7.48	0.264	8.8±9.5	8.6±5.2	0.638	3.9±2.0	4.7±3.6	0.83	-0.18±0.72	-0.16±0.94	0.085
R_1 pre O_2	1.76±0.47	1.67±0.94	0.052	1.76±0.46	1.58±0.73	**0.033**	0.12±0.14	0.75±3.31	**0.008**	3.7±2.2	8.1±17.7	0.531	-0.12±0.76	0.72±2.07	0.224
R_1 post O_2	1.87±0.73	1.90±1.79	**0.027**	1.77±0.63	1.54±0.61	**0.026**	2.26±7.73	1.44±5.48	0.422	71.4±257.6	16.5±34.5	0.327	2.13±8.07	1.41±2.76	0.176
ΔR_1	0.11±0.45	0.23±1.15	0.176	0.01±0.24	-0.05±0.25	0.829	2.26±7.73	1.06±3.46	0.196	72.2±257.4	17.6±44.6	0.814	2.14±8.08	1.50±3.03	0.318
ADC	1.31±0.58	1.43±0.68	**<0.001**	1.31±0.58	1.41±0.70	**<0.001**	0.2±0.1	0.3±0.2	0.142	3.8±2.2	4.0±6.4	**0.03**	0.10±0.79	0.33±0.90	0.652

Table 1

§ Specific Aim 3

We have enrolled so far 5 patients with HCC treated with radioembolization. Patients underwent mpMRI before and 6 weeks after treatment. We will look at the perceptive value of baseline, early follow-up and changes in parameters as possible markers of response at 6-12 months. Results will be analyzed after at least 10 patients are enrolled.

DISCUSSION OF COLLABORATIONS

Within the Network: 1) we are in the process of submitting a manuscript assessing involving a QIN challenge involving multicenter quantification of T1 mapping in vitro, used for DCE-MRI purposes. 2) We have collaborated in several challenges: the DWI linearity challenge, the ADC challenge, and the prostate AIF challenge.

Other Institutions Outside the Network: we have an ongoing collaborating with NYU (Daniel Sodickson's group) on the use of radial GRASP sequence for perfusion acquisition. Data analysis showed major truncation artifacts in the arterial input function. We are looking for solutions to this.

Industrial: we are collaborating with Siemens to test new sequences including DWI, DCE-MRI (using k-space sharing) and 3D T1 mapping.

PLANS FOR NEXT YEAR

Continue patient recruitment for specific aim 3 in patients treated with radioembolization.

LIST OF QIN PUBLICATIONS

1. Bane O, Besa C, Wagner M, Oesingmann N, Zhu H, Fiel MI, Taouli B. Feasibility and reproducibility of BOLD and TOLD measurements in the liver with oxygen and carbogen gas challenge in healthy volunteers and patients with hepatocellular carcinoma. J Magn Reson Imaging. 2016;43(4):866-76. doi: 10.1002/jmri.25051. PubMed PMID: 26417669; PubMed Central PMCID: PMC4803537.
2. Hectors SJ, Wagner M, Besa C, Bane O, Dyvorne HA, Fiel MI, Zhu H, Donovan M, Taouli B. Intravoxel incoherent motion diffusion-weighted imaging of hepatocellular carcinoma: Is there a correlation with flow and perfusion metrics obtained with dynamic contrast-enhanced MRI? J Magn Reson Imaging. 2016. doi: 10.1002/jmri.25194. PubMed PMID: 26919327.
3. Jajamovich GH, Huang W, Besa C, Li X, Afzal A, Dyvorne HA, Taouli B. DCE-MRI of hepatocellular carcinoma: perfusion quantification with Tofts model versus shutter-speed model--initial experience. MAGMA. 2016;29(1):49-58. doi: 10.1007/s10334-015-0513-4. PubMed PMID: 26646522.
4. Huang W, Li X, Li X, Chang M, Oborski MJ, Malyarenko DI, Muzi M, Jajamovich GH, Fedorov A, Chen Y, Tudorica A, Gupta SN, Laymon CM, Marro KI, Dyvorne HA,

Miller JV, Chenevert TL, Yankeelov TE, Mountz J, Kinahan PE, Kikinis R, Taouli B, Fennessy F, Kalpathy-Cramer J. Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. Translational Oncology 2014 Feb 1;7(1):153-166

5. Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, O'Sullivan F, Abramson RG, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Besa C, Kalpathy-Cramer J, Mountz JM, Laymon CM, Muzi M, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy F, Li X. The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. Tomography: a journal for imaging research. 2016;2(1):56-66. doi: 10.18383/j.tom.2015.00184. PubMed PMID: 27200418; PubMed Central PMCID: PMC4869732.
6. Malyarenko DI, Newitt D, L JW, Tudorica A, Helmer KG, Arlinghaus LR, Jacobs MA, Jajamovich G, Taouli B, Yankeelov TE, Huang W, Chenevert TL. Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials. Magn Reson Med. 2016;75(3):1312-23. doi: 10.1002/mrm.25754. PubMed PMID: 25940607; PubMed Central PMCID: PMC4630210.

§ Submitted

Hectors SJ, Wagner M, Bane O, Besa C, Lewis S, Remark R, Chen N, Fiel MI, Zhu H, Merad M, Hoshida Y, Taouli B. Characterization of hepatocellular carcinoma heterogeneity with multiparametric MRI. Radiology (submitted)

REFERENCES CITED

1. Bane O, Besa C, Wagner M, Oesingmann N, Zhu H, Fiel MI, Taouli B. Feasibility and reproducibility of BOLD and TOLD measurements in the liver with oxygen and carbogen gas challenge in healthy volunteers and patients with hepatocellular carcinoma. J Magn Reson Imaging. 2016;43(4):866-76. doi: 10.1002/jmri.25051. PubMed PMID: 26417669; PubMed Central PMCID: PMC4803537.
2. Hectors SJ, Wagner M, Besa C, Bane O, Dyvorne HA, Fiel MI, Zhu H, Donovan M, Taouli B. Intravoxel incoherent motion diffusion-weighted imaging of hepatocellular carcinoma: Is there a correlation with flow and perfusion metrics obtained with dynamic contrast-enhanced MRI? J Magn Reson Imaging. 2016. doi: 10.1002/jmri.25194. PubMed PMID: 26919327.
3. Jajamovich GH, Huang W, Besa C, Li X, Afzal A, Dyvorne HA, Taouli B. DCE-MRI of hepatocellular carcinoma: perfusion quantification with Tofts model versus shutter-speed model--initial experience. MAGMA. 2016;29(1):49-58. doi: 10.1007/s10334-015-0513-4. PubMed PMID: 26646522.
4. Hectors SJ, Wagner M, Bane O, Besa C, Lewis S, Remark R, Chen N, Fiel MI, Zhu H, Merad M, Hoshida Y, Taouli B. Characterization of hepatocellular carcinoma heterogeneity with multiparametric MRI. Radiology (submitted).

5. Just N. Improving tumour heterogeneity MRI assessment with histograms. Br J Cancer. 2014;111(12):2205-13. doi: 10.1038/bjc.2014.512. PubMed PMID: 25268373; PubMed Central PMCID: PMC4264439.
6. Anfuso B, El-Khobar KE, Sukowati CH, Tiribelli C. The multiple origin of cancer stem cells in hepatocellular carcinoma. Clin Res Hepatol Gastroenterol. 2015;39 Suppl 1:S92-7. doi: 10.1016/j.clinre.2015.05.011. PubMed PMID: 26186879.
7. Friemel J, Rechsteiner M, Frick L, Bohm F, Struckmann K, Egger M, Moch H, Heikenwalder M, Weber A. Intratumor heterogeneity in hepatocellular carcinoma. Clin Cancer Res. 2015;21(8):1951-61. doi: 10.1158/1078-0432.CCR-14-0122. PubMed PMID: 25248380.
8. Fujita N, Nishie A, Kubo Y, Asayama Y, Ushijima Y, Takayama Y, Moirta K, Shirabe K, Aishima S, Honda H. Hepatocellular carcinoma: clinical significance of signal heterogeneity in the hepatobiliary phase of gadoxetic acid-enhanced MR imaging. Eur Radiol. 2015;25(1):211-20. doi: 10.1007/s00330-014-3349-9. PubMed PMID: 25063395.
9. Rutman AM, Kuo MD. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. Eur J Radiol. 2009;70(2):232-41. doi: 10.1016/j.ejrad.2009.01.050. PubMed PMID: 19303233.
10. Patel P, Schutzer SE, Pyrsopoulos N. Immunobiology of hepatocarcinogenesis: Ways to go or almost there? World J Gastrointest Pathophysiol. 2016;7(3):242-55. doi: 10.4291/wjgp.v7.i3.242. PubMed PMID: 27574562; PubMed Central PMCID: PMC4981764.

U01 CA172027: Quantitative MRSI to predict early response to HDAC inhibitor therapy in new GBM management

Emory University/Johns Hopkins University

Hyunsuk Shim, Ph.D.
Hui-Kuo Shu, M.D., Ph.D.
Jeffrey Olson, M.D.
Peter Barker, Ph.D.

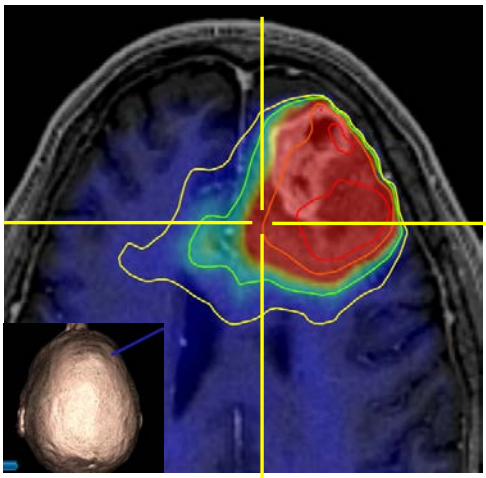
INTRODUCTION

A major impediment to the development of new therapies for glioblastoma (GBM) is a lack of biomarkers to quantitatively monitor response. Standard of care diagnostic images (contrast-enhanced T1 weighted MRI and T2-weighted/FLAIR) are used to guide surgical resection and radiation therapy planning, While these images are excellent images to differentiate higher grade gliomas from the lower grade gliomas, they do not show entirety of the infiltration of GBMs. Proton magnetic resonance spectroscopic imaging (MRSI), which can characterize regions of brain based on levels of various metabolites and other substances, is a candidate imaging modality for defining high risk regions that are not identified by standard MRI. Metabolites that can be evaluated include: choline (Cho), a peak reflecting cell membrane synthesis that is elevated in highly proliferating, non-necrotic gliomas; creatine (Cr), an energy metabolite; and N-acetyl aspartate (NAA), a healthy neuronal biomarker that is decreased as healthy tissue is displaced. Early studies established that the MR spectra of GBMs differ significantly from normal brain, with increased levels of Cho, and decreased levels of NAA.

DISCUSSION OF PROGRESS

§ Correlation Between sMRI Cho/NAA with Histology

We have been developing an advanced spectroscopic technique we have termed spectroscopic MRI (sMRI), which combines advanced imaging technologies such as 3D whole brain echo-planar spectroscopic imaging (EPSI) and acceleration acquisition technologies with a new clinical platform for registering spectroscopy data with standard of care images, enabling easy visualization and an efficient clinical workflow. We have been using sMRI in various clinical studies at Emory to demonstrate it's superiority compared with standard imaging alone to identify the entirety of GBMs, including non-enhancing infiltrative tumor. Based on a recently completed clinical study to correlate sMRI and histology in tumor samples collected via stereotactic biopsy-manner (Emory IRB00051663), we learned that the Cho/NAA ratio showed significant correlations with tumor cell density as determined via histological analysis ($\rho = 0.82$, $p < 0.001$). (Cordova et al. published in Neuro-Oncology 2016)



sMRI Biomarkers vs SOX2 (glioma marker) Density

Biomarker	ρ	p-value
NAA	-0.50	0.01*
Cho	0.63	5E-4*
Cho/NAA	0.82	<1E-4*
DWI-ADC	0.17	0.40

Figure 1. A normalized metric of tumor infiltration, SOX2 (glioma marker) density, identifies tumor outside of conventional imaging and exhibits striking correlations with sMRI biomarkers. Though no obvious abnormality can be found on preoperative T1w-CE or T2w imaging in this patient, a striking elevation in Cho/NAA on sMRI suggests substantial tumor infiltration. Statistically significant correlations were seen between various normalized metabolic markers and SOX2 density with Cho/NAA exhibiting the strongest association

§ Correlation Between sMRI Cho/NAA with Recurrence Pattern

Based on data from the control arm of our current U01 study (Emory IRB00055973), sMRI Cho/NAA ratio map before RT treatment initiation matched well with contrast-enhancement at sites of tumor recurrence and exhibited an inverse relationship with progression-free survival (Cordova et al. published in Neuro-Oncology 2016). Figure 2 shows the zoom-in view of recurrence case that Cordova et al. published in the QIN special issue of Tomography 2016)

The 40-year-old male shown in Figure 2 exhibited a striking anterior tail of Cho/NAA elevation outside of the T_{1w}-CE lesion that passed along the posterior horn of the left lateral ventricle before surgery. This metabolic abnormality continued to grow through the duration of RT, ultimately resulting in subependymal invasion along the trajectory of the Cho/NAA elevation. This patient underwent salvage surgery, which histologically confirmed GBM tumor at recurrence site. This case also exhibited T1w-CE lesion morphological changes that approached a Dice coefficient of unity, or perfect agreement, when compared to pre-RT Cho/NAA 2X NORM volume (ΔDice: +19.0%, ΔMED: -35.6.0%). This increase in agreement accounts for an addition of 12.9 cm^3 of tumor in the pre-RT Cho/NAA 2X NORM volume from preRT to recurrence. These exciting results will be reported in the future once confirmed with other similar cases.

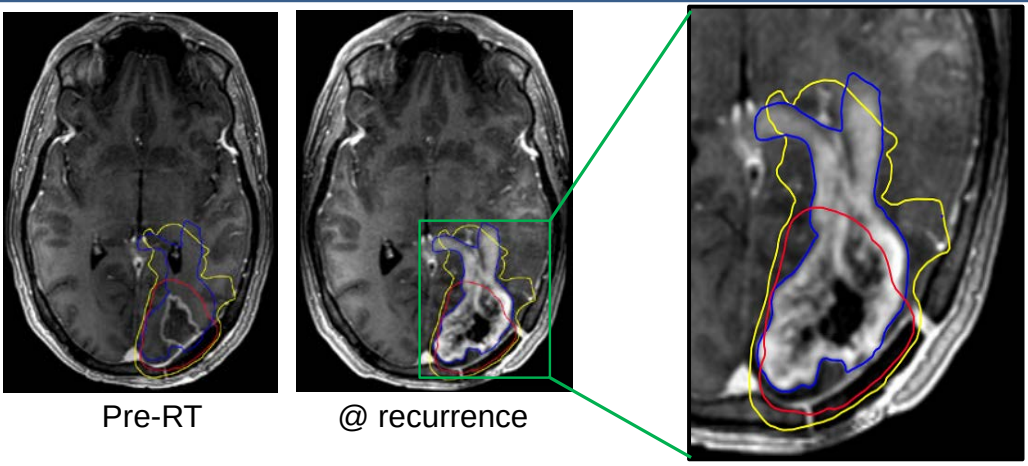


Figure 2. Abnormalities in Cho/NAA describe regions at high risk for recurrence before RT in GBM. The **yellow** contour illustrates the regions that exhibit a Cho/NAA abnormality ≥ 2-fold higher than normal contralateral brain co-registered with T1w-CE taken before RT (left) and at the first recurrence after RT (regions of tumor recurrence - **blue** contour). Though no T1w-CE abnormality was found before RT, preRT Cho/NAA abnormality (**yellow**) clearly shows infiltration of subependymal space that becomes contrast-enhancing 4 months later. **Red** contour was where 60 Gy (CTV60) was applied for RT. The fusion of 2.0-fold Cho/NAA abnormalities to the CTV60 resulted in a target covering a significantly larger proportion of the recurrence.

§ Development of Web-based spectroscopic MRI Clinical Interface

There is widespread agreement that MR spectroscopy can provide valuable information without the need for exogenous contrast agents, however the infrastructure needed to incorporate sMRI into the clinical workflow is lacking. We have been developing a web-based sMRI clinical interface for analysis, visualization and integration of sMRI data into patient management. This "scanner-to-clinician" platform is designed to provide quantitative, expedient, and objective analysis to integrate sMRI into routine clinical usage, including diagnosis and therapy planning (radiation or surgery). In addition, this user-friendly tool can be highly valuable in the sMRI-based diagnosis and evaluation of numerous other neuropathologies aside from cancer, including hypoxic-ischemic injury, multiple sclerosis (and other demyelinating diseases), inborn errors of metabolism, and neurodegenerative diseases, such as Alzheimer's.

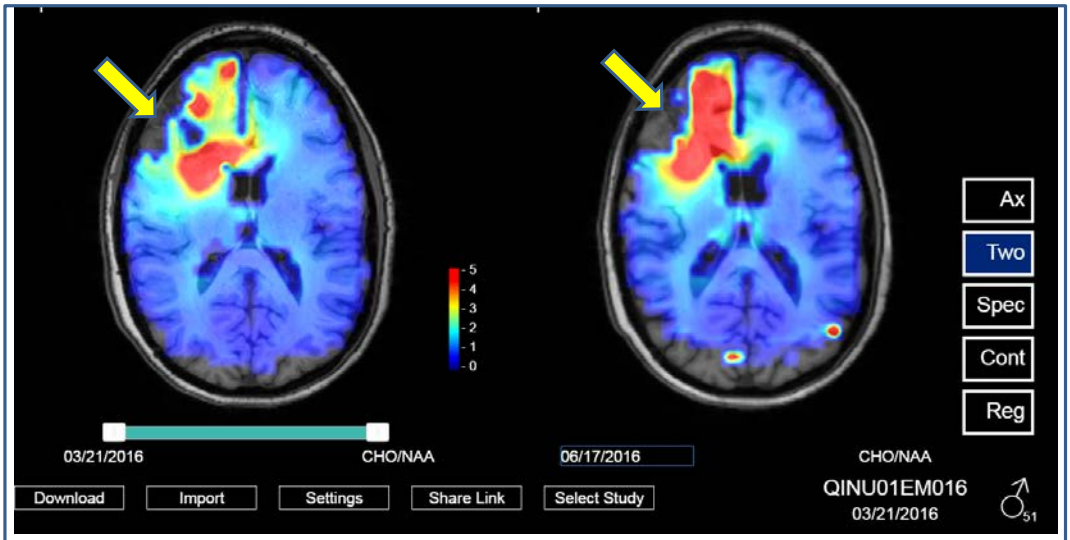


Figure 3. We have developed an imaging technology known as "spectroscopic MRI," which can detect changes in tissue metabolism. This helps us monitor the response to therapy in patients by tracking the changes in the metabolism of tumor cells without the need for any injected contrast agents. sMRI clinical interface is an easy-to-use web application for visualization and collaborative treatment planning using sMRI. The left image shows a 51 year old female diagnosed with glioblastoma before standard care treatment, and the right image shows the same patient after. Unfortunately, she did not respond to chemo & radiation, and the sMRI highlights the corresponding metabolic response (red represents tumor infiltrative activity, blue represents healthy tissue).

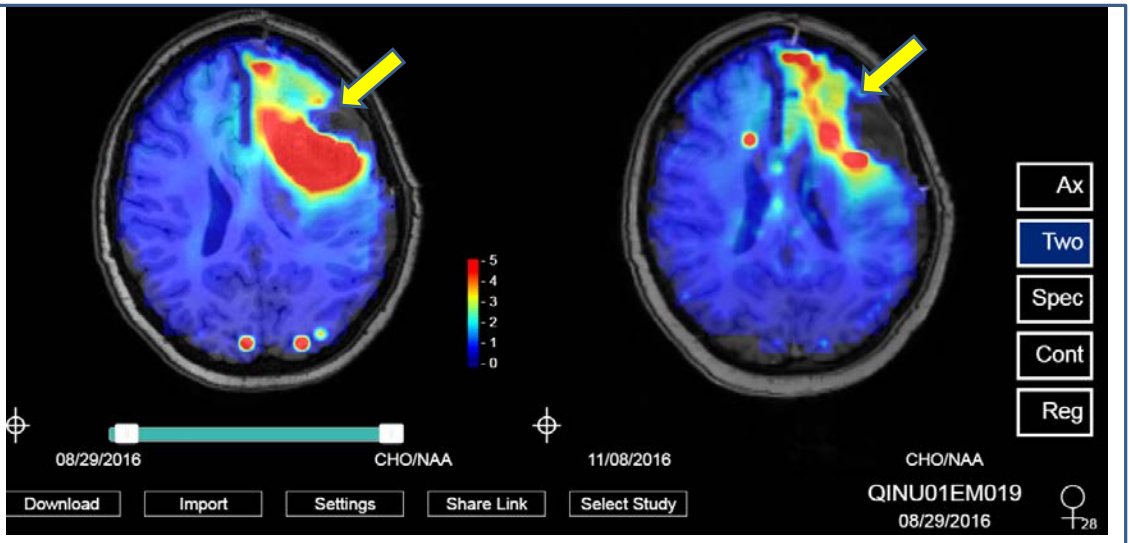


Figure 4: The image shows a 28 year old female diagnosed with GBM who was treated with HDAC inhibitor, an investigational drug being tested in the current U01 study, in addition to standard chemo and radiation therapy. The left image is before treatment and the right image is 4 weeks after treatment completion. As the sMRI highlights, the metabolic changes show that she positively responded to the therapy (red represents tumor infiltrative activity, blue represents healthy tissue, and lack of metabolite signals represents necrosis).

COLLABORATIONS WITHIN THE NETWORK

Our current project is a two-site clinical study, with the Emory team collaborating with Johns Hopkins. The sMRI clinical interface has been shared between these sites, and is in the process of expanding testing to the University of Miami and New York University. We hope to deploy this technology with several QIN sites later in 2017.

PLANS FOR NEXT YEAR

We plan to continue with patient enrollment for our clinical study at two sites. We will continue to develop easier (automated) quality control components to display and report the reliable sMRI results: we are now focusing on developing spectral quality filter to eliminate the voxels with poor quality spectra or poor fitting.

Our sMRI resolution is 108 microliters and the scan time for 3D whole brain sMRI for 6 different metabolite maps takes 15 mins. We have purchased a new Siemens Prisma 3T scanner with 32 channel head coil array that will be available for use in February 2017. We are working together with Dr. Maudsley at University of Miami (consultant) and Siemens to implement the same advanced sMRI sequence on Prisma. We anticipate a 40% signal-to-noise ratio improvement while maintaining the same spatial coverage as our current systems.

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

§ Manuscripts published directly as a result of this grant

Cordova, J.S., Shu, H.G., Liang, Z., Gurbani, S. S., Cooper, L.A.D., Holder, C.A. Olson, J.J., Kairdolf, B., Schreibmann, E., Neill, S., Hadjipanayis, C.G., Shim, H. (2016) Whole-brain, spectroscopic MRI biomarkers identify infiltrating margins in glioblastoma patients. Neuro-Oncology, Neuro-Oncology, 18(8): 1180-9. PMC4933486

Cordova, J.S., Gurbani, S. S., Olson, J.J., Liang, Z., Cooper, L.A.D., Shu, H.G., Schreibmann, E., Neill, S., Hadjipanayis, C.G., Holder, C.A., Shim, H. (2016) A systemic pipeline for the objective comparison of whole-brain spectroscopic MRI with histology in biopsy specimens from grade III glioma. Tomography, 2(2): 106-116. PMC4968944.

Cordova, J.S., Kandula, S., Gurbani, S. S., Zhong, J., Tejani, M., Kayode, O., Patel, K., Prabhu, R., Schreibmann, E., Crocker, I., Holder, C.A., Shim, H., Shu, H.G. (2016) The impact of integrating volumetric whole-brain spectroscopic MRI into radiation treatment planning for glioblastoma. QIN special issue, Tomography 2(4): 366-73.

Schreibmann, E., Cordova, J.S., Gurbani, S., Holder, C.A., Cooper, L.A., Shu, H.G., Shim, H. (2016) Automated segmentation of high resolution 3D wholebrain spectroscopic MRI for glioblastoma treatment planning. Medical Physics 43(6) 3428.

§ National meeting education session presentations directly as a result of this grant

We organized several CME sessions during SNMMI mid-Winter meeting in Orlando. One of them was First-in-Human MR Molecular Imaging that included our research talks. In addition, we organized several sessions during SNMMI 2016 Annual Meeting in San Diego including CME Categorical Whole day Session and a NCI Cancer Imaging Program (CIP) Quantitative Imaging Network (QIN) session.

§ National Presentations directly as a result of this grant

Invited Lectures at the National Meetings:

Shim H. et al "IDH mutation detection in gliomas" – Society of Nuclear medicine mid-Winter meeting, a continuing education session, Orlando, January 2016

Shu H. et al. "Spectroscopic MRI identifies infiltrating margins in glioblastoma for 5-ALA fluorescence-guided surgery" - Society of Nuclear medicine mid-Winter meeting, a continuing education session, Orlando, January 2016

Barker, P. et al. "Tumor tutorial" – International Society of Magnetic Resonance in Medicine, a continuing education session, Singapore, May 2016

Barker, P. et al. "Brain Tumor Spectroscopy" – American Society of Neuro-Radiology, SAM session, Washington DC, May 2016

Shim H. et al "Overview of cancer metabolism: glucose and amino acids" – Society of Nuclear medicine Annual meeting, a continuing education session, San Diego, June 2016

Shu H et al. "1. Critical unmet needs for treatment planning imaging in brain tumor patients; 2. Spectroscopic MRI for brain tumor patients" - Society of Nuclear medicine Annual meeting, a continuing education session, San Diego, June 2016

Shu H. et al. "Feasibility of whole brain, high resolution spectroscopic MRI for glioblastoma tumor imaging" – American Society for Radiation Oncology, Boston, September 2016

Shim H. et al. "Molecular Imaging mini-course: Clinical application of molecular imaging – Neuro" RSNA Refresher Course, Chicago, December 2016

Invited Lectures at the Academic Centers:

Shim H. et al. "The use of high resolution 3D whole brain MR spectroscopic imaging in the management of brain tumor patients", Cedars Sinai Hospital, February 2016

Shim H. et al. "Critical unmet needs for treatment planning imaging in GBM patients & spectroscopic MRI", Cedars Sinai Hospital, June 2016

Shim H. et al. “Critical unmet needs for treatment planning imaging in GBM patients & spectroscopic MRI”, Mount Sinai, July 2016

Shim H. et al. “Critical unmet needs for treatment planning imaging in GBM patients & spectroscopic MRI”, New York University, July 2016

Barker, P. et al. “Brain Tumor Spectroscopy”, German Cancer Research Center in Heidelberg, Germany, August 2016.

Shim H. et al. “The use of high resolution 3D whole brain MR spectroscopic imaging in the management of brain tumor patients”, Georgia State University, October 2016

Shim H. et al. “Improved whole brain spectroscopic MRI to guide radiation dose escalation for glioblastomas”, University of Pennsylvania, Radiology Grand Rounds, December 2016

Shim H. et al. “Improving cancer patient management through drug discovery and whole brain spectroscopic MRI” Seoul National University Hospital, Nuclear Medicine Grand Rounds, December 2016

U01 CA176110: Quantitative Perfusion and Diffusion MRI Biomarkers to Measure Glioma Response

Medical College of Wisconsin

Kathleen M. Schmainda, Ph.D.

INTRODUCTION

The overall goal of this project is to develop and validate both standard and novel perfusion-weighted MRI (PWI) and diffusion-weighted MRI (DWI) biomarkers to monitor treatment response for both therapeutic clinical trials and standard of care treatment patients with brain tumors. This goal addresses an urgent need for better ways to monitor targeted therapies, for which standard measures of enhancing tumor volumes are no longer sufficient. Two PWI methods will be characterized for clinical trials. The first more wide-spread DSC (dynamic susceptibility contrast) approach provides tumor rCBV (relative cerebral blood volume) measurements obtained after a pre-load of contrast agent and corrected for confounding contrast agent leakage effects. The second approach, while less-proven has high-potential to become the most comprehensive perfusion solution. It consists of using a dual-echo gradient-echo spiral method, which enables the simultaneous collection of both DSC (dynamic susceptibility contrast) and DCE (dynamic contrast enhanced) perfusion data using only a single dose of contrast agent and incorporates comprehensive correction for leakage effects [1-3]. In addition, we will continue to explore the potential of DWI methods for the evaluation of treatment response, specifically by computing changes in the apparent diffusion coefficient (ADC) across time and creating functional diffusion maps (fDM) within non-contrast-agent-enhancing regions.

While both PWI and DWI have demonstrated great promise for treatment monitoring, studies defining their test-retest repeatability, necessary for use of these techniques in clinical trials, are lacking, and thus represent the focus of Aim 1. In addition, early results suggest that hybrid PWI/DWI maps will likely provide the most complete assessment of treatment response, a hypothesis that will be tested in Aim 2. Finally, in order to make the optimized PWI/DWI technology and workflow available in a robust and cost-effective manner for clinical trials and standard practice, Aim 3 involves the development of a commercial integrated image analysis platform for use in large-scale multi-center clinical trials.

DISCUSSION OF PROGRESS

§ Specific Aim1

Manuscripts: Published and in Progress

Perfusion Repeatability: In collaboration with Massachusetts General Hospital, another QIN member, we *published* two papers describing the repeatability of DSC-pMRI methods [4, 5], and their dependence on post-processing methods, as well as the minimum

number of patients need to power a clinical trial [4]. The results show that when ordered by the RC, methods utilizing post-processing leakage-correction and ΔR2*(t) techniques largely offered superior repeatability. Across processing techniques the standardized RCBV[6] estimates were less variable (13-20%) than normalized rCBV (nRCBV) (24-67%) estimates. It was also found that normalization of rCBV rather than AIF deconvolution (to estimate an 'absolute' value of CBV) resulted in a more repeatable measurement.[5]

> **Significance:** Knowledge of the repeatability of DSC-MRI perfusion methods has been lacking. These first reports providing this information are important for clinical translation and use in clinical trials.

SPICE: We recently published a paper describing the theory and initial feasibility of the dual-echo sequence, which has been renamed SPICE (spiral perfusion imaging with consecutive echoes) [7]. This paper appeared in the December 2016 issue of the journal, *Tomography*, includes a detailed mathematical description of the novel SPICE perfusion imaging acquisition and post-processing method. This method can be used to simultaneously acquire DSC- and DCE-MRI data with only a single dose of gadolinium contrast agent. It also does not require the collection of a precontrast T1 map for DCE-MRI processing and eliminates confounding contrast agent effects due to contrast extravasation.

> **Significance:** We are hopeful that the publication of the theory underlying SPICE will motivate more groups to adopt dual-echo sequences for perfusion.

Prostate AIF Challenge: Dr. Peter LaViolette, Co-Investigator on this grant, has participated in the arterial input (AIF) challenge headed by Wei Huang from Oregon Health Science University (OHSU) by applying independent component analysis to extract the AIFs automatically. He processed both prostate cancer and sarcoma datasets and submitted them to the host institution. This study was recently published [8].

> **Significance:** There was good consistency across DCE parameter estimations using a variety of AIF tools. Only the ICA tool, contributed by Dr LaViolette, showed the greatest discrepancy.

Ongoing Experimental Studies

SPICE vs DSC-MRI study: We completed the study to compare SPICE-derived rCBV maps to DSC-derived rCBV maps. The results demonstrated that the rCBV values are comparable in both low-grade and high-grade tumors. This is further proof that the SPICE method may provide similar information, yet be superior to standard methods since it requires only a single dose of Gd contrast agent while also providing additional (ie DCE) perfusion metrics. An initial submission of the manuscript was not accepted. A primary concern of the reviewers was the extra leakage-correction analysis and comparisons that were included. (We had data to show the importance of preload to standard DSC-MRI.) However, this secondary comparison is not necessary for the validation of SPICE and served only to confuse the readers. It will therefore be removed and the manuscript submitted to

another journal and will focus more on the rCBV comparison and applications. In the interim we decided it was important to get a description of the basic theory and feasibility published, which resulted in the SPICE manuscript described above. A revised paper of the comparison will be submitted in 2017 Q1.

> **Significance**: The validation of SPICE is quite timely given the many discussions regarding the optimal dose of contrast for DSC studies in the context of using the least amount possible. With SPICE this discussion will eventually become a non-issue since all data can be acquired with a single dose of contrast agent. Publication of this work will also move us closer towards a QIN network goal of distributing the acquisition and post-processing software to all interested QIN sites for further evaluation.

Collection of double-baseline SPICE and diffusion data: Double-baseline data will be collected in patients with high-grade brain tumors within a short time-interval during which no change in tumor status is presumed to occur. As described previously, the prospective collection of the SPICE repeatability data has been delayed due to an upgrade of our GE MRI system, which made a current version unworkable. The sequence and image reconstruction software have now been revised and recompiled for the GE 3T clinical platform as well as the GE 3T research MRI. This software upgrade together with the purchase of a power injector and the hiring of a certified radiology technologist for the research GE 3T MRI system enables us to perform many more add-on research studies. The number of SPICE datasets collected has increased tremendously over the past two months such that we should be able to easily complete this study during this next funding cycle.

> **Significance:** Previous studies to determine the repeatability of diffusion have not been performed and thus is the focus of the planned studies. At the same time, the repeatability of SPICE, the new perfusion imaging method will be undertaken.

§ Specific Aim 2

Manuscripts: In progress

Using Perfusion and Diffusion MRI to Distinguish Tumor from Treatment Effect: We demonstrated that normalized and standardized rCBV values could be used to distinguish tumor from treatment effect (TE). Forty-eight tissue samples from sixteen brain tumor subjects were spatially correlated with pre-surgical MRI, which included DSC-MRI and DWI [9]. Biopsy locations were determined via a StealthStation® S7™ surgical navigation unit (Medtronic, Minneapolis, MN). Pathologic diagnosis confirmed 11 samples with pure treatment effect and 37 samples with pure GBM. All perfusion metrics distinguished treatment effect from GBM while ADC did not (Table 1). Of particular note, the normalized rCBV threshhold determined by us is comparable to the threshold determined by Dr Leland Hu (QIN associate member) at Barrow Neurological Institute in Phoenix Arizona. A manuscript describing this work was submitted to the journal, *Radiology*, but not accepted. It is being revised for submission in another journal.

Table 1: Image parameter results for distinguishing tumor from treatment

Parameter	P-Value	Threshold	Specificity	Sensitivity	AUC
ADC	0.066	-	-	-	-
sRCBV	0.001	≥ 3575 a.u.	90.9	78.4	0.870
nRCBV	0.0002	≥ 1.13 a.u.	90.9	81.1	0.902
nRCBF	0.0005	≥ 1.03 a.u.	81.8	75.7	0.853

Table 1: Image parameter results for distinguishing tumor from treatment.

Significance: This study is a first step towards addressing the longstanding need for a method to accurately distinguish true treatment response from pseudoprogression or pseudoresponse. Given the consistency of results between two institutions, it appears that this is a good approach to develop test further and possibly become the basis of a new QIN network challenge.

Fractional tumor burden (FTB) maps to predict treatment response: Using the thresholds obtained for nRCBV (or sRCBV), fractional tumor burden (FTB) maps can be created to spatially visualize the portion of enhancing tumor that is treatment effect or GBM (Figure 1). Preliminary results, shown in Figure 2, demonstrate that FTB may serve as an important marker useful for treatment management decisions. In this group of patients with newly diagnosed GBM, and after undergoing chemo-radiation therapy, only FTB was useful for distinguishing both PFS and OS. A similar result was found for patients treated with bevacizumab (not shown). This work has been submitted as two separate abstracts for the 2017 International Society of Magnetic Resonance in Medicine meeting. In addition, a manuscript describing these FTB results is being prepared for journal submission in early 2017.

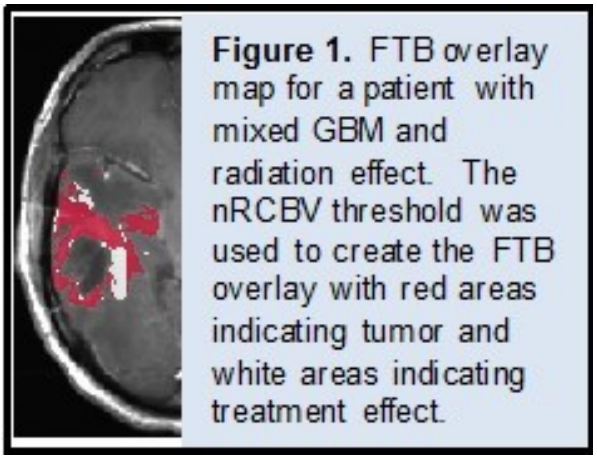


Figure 1. FTB overlay map for a patient with mixed GBM and radiation effect. The nRCBV threshold was used to create the FTB overlay with red areas indicating tumor and white areas indicating treatment effect.

Significance: These results demonstrate that rCBV metrics provide information relevant to treatment evaluation, and can be used to create fractional tumor burden (FTB) maps, which are demonstrating promise as a new biomarker for evaluating treatment response.

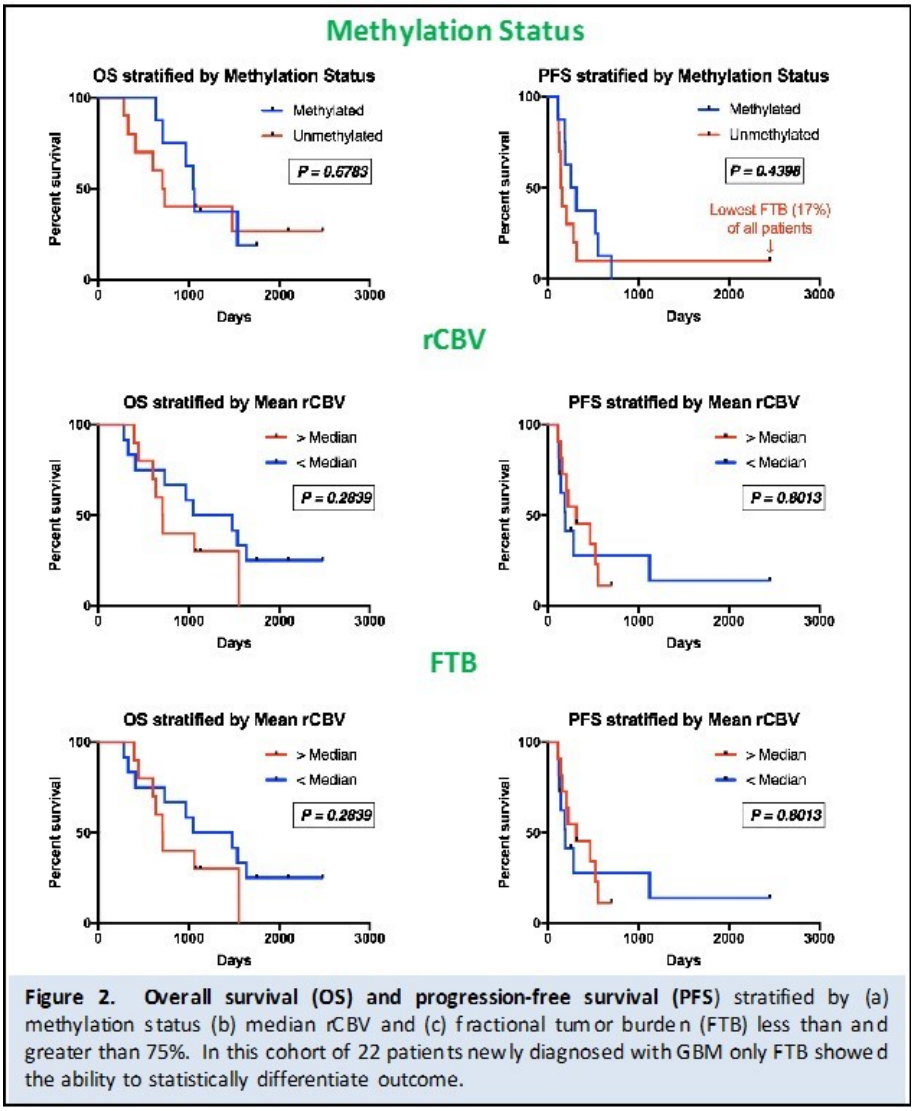


Figure 2. Overall survival (OS) and progression-free survival (PFS) stratified by (a) methylation status (b) median rCBV and (c) fractional tumor burden (FTB) less than and greater than 75%. In this cohort of 22 patients newly diagnosed with GBM only FTB showed the ability to statistically differentiate outcome.

Radiomic Profiling: Additional studies using diffusion MRI and radiomic profiling demonstrate promise to distinguish tumor from TE and predict prognosis. This work, led by co-investigator Dr Peter LaViolette is described in detail in two recently published journal articles [10, 11].

Significance: These results demonstrate that diffusion MRI continues to play in role in understanding treatment response and together with other parameters, via radiomic profiling, can predict prognosis.

DSC-MRI Challenge: A DSC Challenge was undertaken to boost confidence in DSC-MRI post-processing across sites and platforms. The purpose of this challenge was to reach consensus regarding the post-processing of DSC-MRI data through a comparison of multi-site/multi-platform analyses of a shared brain tumor patient data set. A total of 49 low-grade (n=13) and high-grade (n=36) glioma DSC-MRI datasets were uploaded to the cancer imaging archive (TCIA). All glioma grades were confirmed by histopathology within 41 days following the DSC-MRI study. The datasets were co-registered with T1w images and included a predetermined AIF, necessary for the determination of CBF, ROIs of whole brain for efficient DSC processing, normal appearing white matter (NAWM), for the creation of normalized parameter maps, normal appearing cerebral cortex (NACC), as well as enhancing tumor ROIs. Seven sites using seven different software (SW) platforms provided median ROI values for 18 different normalized rCBV (nRCBV), 2 standardized rCBV (sRCBV) and 12 normalized CBF (nCBF) metrics. As listed in Table 2, there was excellent concordance across sites and platforms and each could statistically distinguish low-grade from high-grade glioma. However, the thresholds that gave the best sensitivity and sensitivity varied from 1.3 to 1.7. But with a nRCBV of 1.45 all platforms had a sensitivity and specificity of at least 80%. This work was submitted as an abstract for the 2017 International Society of Magnetic Resonance in Medicine meeting. In addition, a manuscript describing theses results is being prepared for journal submission in early 2017.

Significance: These results demonstrate that DSC-MRI methods can be used more routinely, with confidence, for the evaluation of adult primary brain tumor.

Table 2: Linn's Concordance Correlation Coefficient

TEAM		1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19
	Leakage	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No
1	Yes	LESION NACC	1.00	1.00	0.99	0.99	0.99	0.99	0.97	0.88	0.93	0.96	0.91	0.98	0.96	0.96	0.92	0.79	0.88
2	Yes	1.00	LESION NACC	1.00	0.99	0.99	0.99	0.99	0.97	0.88	0.93	0.96	0.91	0.98	0.96	0.96	0.92	0.79	0.88
3	Yes	1.00	1.00	LESION NACC	0.99	0.99	0.99	0.99	0.97	0.88	0.93	0.96	0.91	0.98	0.94	0.96	0.92	0.79	0.88
4	Yes	0.82	0.82	0.82	LESION NACC	1.00	1.00	1.00	0.98	0.90	0.94	0.98	0.88	0.99	0.98	0.98	0.94	0.81	0.90
5	Yes	0.82	0.82	0.82	1.00	LESION NACC	0.99	0.99	0.98	0.89	0.94	0.98	0.88	0.99	0.98	0.98	0.94	0.81	0.89
6	Yes	0.70	0.70	0.70	0.95	0.95	LESION NACC	1.00	0.98	0.88	0.94	0.97	0.91	0.98	0.97	0.97	0.92	0.79	0.88
7	Yes	0.70	0.70	0.70	0.95	0.95	1.00	LESION NACC	0.98	0.88	0.94	0.97	0.91	0.98	0.97	0.97	0.92	0.79	0.88
8	Yes	0.98	0.98	0.98	0.77	0.77	0.63	0.63	LESION NACC	0.92	0.92	0.98	0.83	0.98	1.00	0.99	0.98	0.84	0.92
9	Yes	0.92	0.92	0.91	0.63	0.63	0.48	0.48	0.95	LESION NACC	0.92	0.92	0.77	0.92	0.93	0.93	0.95	0.93	1.00
10	Yes	0.60	0.60	0.60	0.74	0.74	0.72	0.72	0.57	0.49	LESION NACC	0.94	0.87	0.93	0.92	0.92	0.89	0.86	0.92
11	Yes	0.82	0.82	0.82	0.99	0.98	0.95	0.95	0.77	0.64	0.64	LESION NACC	0.84	0.98	0.99	0.99	0.96	0.85	0.92
12	Yes	0.40	0.40	0.40	0.73	0.73	0.86	0.86	0.33	0.21	0.59	0.74	LESION NACC	0.85	0.82	0.82	0.76	0.68	0.77
13	Yes	0.93	0.93	0.93	0.94	0.94	0.82	0.82	0.91	0.82	0.69	0.93	0.52	LESION NACC	0.98	0.98	0.96	0.83	0.92
15	No	0.96	0.96	0.95	0.71	0.70	0.56	0.56	0.98	0.98	0.52	0.71	0.27	0.87	LESION NACC	1.00	0.98	0.86	0.93
16	No	0.96	0.96	0.95	0.71	0.70	0.56	0.56	0.98	0.98	0.52	0.71	0.27	0.87	1.00	LESION NACC	0.98	0.86	0.93
17	No	0.91	0.91	0.91	0.62	0.62	0.50	0.50	0.94	0.97	0.46	0.63	0.23	0.79	0.98	0.98	LESION NACC	0.91	0.95
18	No	0.87	0.87	0.87	0.83	0.83	0.70	0.70	0.88	0.85	0.66	0.84	0.42	0.92	0.86	0.86	0.81	LESION NACC	0.97
19	No	0.92	0.92	0.92	0.64	0.64	0.46	0.49	0.95	1.00	0.49	0.64	0.21	0.82	0.98	0.98	0.97	0.86	LESION NACC

LCCC ≥ 0.90; 0.80 ≤ LCCC ≥ 0.89; LCCC < 0.80

§ Specific Aim 3

IB Rad Tech

IB Rad Tech is a customizable workflow wizard developed by our industrial collaborator, Imaging Biometrics (IB), to service many different "work in progress" workflows such as those being developed by QIN members. (See: http://69.162.134.80/~imagingb/files/9914/3050/6176/IB_Clinic.pdf) It was recently enhanced with longitudinal processing capability, allowing the generation of comparison data for multiple patient time points.

The core processing library used by IB Rad Tech was enhanced by the additional of several DWI-based outputs, including alpha-diffusion and IVIM parameter maps. This technology was tested as part of one of the diffusion MRI challenges led by Dr David Newitt. It is being further revised in response to feedback resulting from the participation. This functionality has not yet been exposed in the IB Rad Tech workflows, but is planned to be implemented during year four. Meanwhile, IB Rad Tech 2.0 is expected to be released as an FDA-cleared and CE-Marked product late in 2016 or early in 2017.

In addition, a workflow for the creation of FTB (fractional tumor burden maps) as described under Specific Aim 2, has been developed. It was used entirely for the processing described and will thus make it seamless to distribute the workflow to others for testing and use.

> Significance: Having parallel development of an industrial platform ensures efficient and timely translation of the most proven technologies for widespread use in both the research and clinical communities.

PLANS FOR NEXT YEAR

§Specific Aim1

Characterize the repeatability of DSC and DEGES PWI and DWI (fDM) parameters in primary brain tumors.

- Revise and submit SPICE comparison paper with a working title of "*Spiral Perfusion Imaging with Consecutive Echoes (SPICE) for the Simultaneous Mapping of DSC- and DCE-MRI Parameters in Brain Tumor Patients using a Single Dose of Gadolinium Contrast*."
- Increase prospective data collection using SPICE sequence on both the clinical and research 3T MRI systems.
- Complete diffusion and perfusion MRI repeatability studies.
- Initiate new perfusion QIN network challenges, which may include the determination of AIF for DSC-MRI studies and/or the ability of each site to distinguish tumor from

treatment effect with the upload of a new patient dataset with the appropriate diagnostic information.

§ Specific Aim 2

To prospectively determine the ability of pMRI and DWI to predict treatment response in glioblastoma patients.

- Revise and submit the "Tumor versus Treatment Effect" paper.
- Submit manuscript describing fractional tumor burden (FTB) results in patients treated with chemo-radiation therapy and bevacizumab therapy. This may be one or two separate papers.
- Complete and submit the "DSC-MRI Platform Challenge" paper.
- Begin to evaluate the role of FTB in distinguishing pseudoprogression and psuedoresponse from true response.
- Continue to evaluate the role of diffusion, and in particular the different diffusion approaches (IVIM, flow-compensated diffusion, and RSI (restricted spectrum imaging)) for treatment evaluation.
- Continue to evaluate radiomic profiling for the detection of response, prediction of outcomes and, of great interest, the ability to detect tumor that is "invisible" with standard imaging.

Specific Aim 3

To develop a commercial integrated PWI/DWI image analysis platform for use in large-scale multi-center multi-platform clinical trials.

- Re-analyze the data from the (Dr David Newitt) DWI challenge. In particular, the IVIM metrics need to be recalculated using the newly incorporated/modified Imaging Biometrics plugins.
- Finish the dual-echo (ie SPICE) post-processing plugin tools. Test these tools on both SPICE and EPI-based dual-echo datasets. Accomplishing this task will position us well for a possible dual-echo DSC challenge.
- MRI processing and workflows. Test standard methods and models against the "fixed T1" model, which has been recently incorporated. Initial results show that this may provide more useable DCE parameter maps.
- Finalize testing on the deltaT1 standardization workflow. Upon completion Imaging Biometrics plans to initiate another QIN challenge regarding tumor image segmentation.

LIST OF REFERENCES

1. Paulson, E.S., D.E. Prah, and K.M. Schmainda. *Correction of Contrast Agent Extravasation Effects in DSC-MRI using Dual-Echo SPIRAL Provides Better Reference for Evaluating PASL CBF Estimates in Brain Tumors*. in *Proceedings International Society of Magnetic Resonance in Medicine, 15th Annual Meeting*. 2007. Berlin.
2. Paulson, E.S., D.E. Prah, and K.M. Schmainda. *Compensation of confounding T1 and T2 dipolar and residual susceptibility effects in DSC-MRI using dual-echo SPIRAL*. in *Proc. Intl. Soc. Mag. Reson. Med.* 2007.
3. Paulson, E.S. and K.M. Schmainda, *Comparison of dynamic susceptibility-weighted contrast-enhanced MR methods: recommendations for measuring relative cerebral blood volume in brain tumors.* Radiology, 2008. **249**(2): p. 601-13.
4. Prah, M.A., et al., *Repeatability of Standardized and Normalized Relative CBV in Patients with Newly Diagnosed Glioblastoma.* AJNR Am J Neuroradiol, 2015. **36**(9): p. 1654-61.
5. Jafari-Khouzani, K., et al., *Repeatability of Cerebral Perfusion Using Dynamic Susceptibility Contrast MRI in Glioblastoma Patients.* Transl Oncol, 2015. **8**(3): p. 137-46.
6. Bedekar, D., T.R. Jensen, and K.M. Schmainda, *Standardization of relative cerebral blood volume (rCBV) image maps for ease of both inter and intra-patient comparisons.* Magn Reson Med, 2010. **64**(3): p. 907-913.
7. Paulson, E.S., D.E. Prah, and K.M. Schmainda, *Spiral Perfusion Imaging with Consecutive Echoes (SPICETM) for the Simultaneous Mapping of DSC- and DCE-MRI Parameters in Brain Tumor Patients: Theory and Initial Feasibility.* Tomography, 2016: p. (In Press).
8. Huang, W., et al., *The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge.* Tomography, 2016. **2**(1): p. 56-66.
9. Prah, M.A., et al. *Comparison of diffusion and perfusion parameters in distinguishing radiation effect and necrosis from GBM*. in *Proceedings International Society of Magnetic Resonance in Medicine*. 2015. Toronto, Ontario, Candada.
10. Nguyen, H.S., et al., *Progressing Bevacizumab-Induced Diffusion Restriction Is Associated with Coagulative Necrosis Surrounded by Viable Tumor and Decreased Overall Survival in Patients with Recurrent Glioblastoma.* AJNR Am J Neuroradiol, 2016.
11. McGarry, S.D., et al., *Magnetic Resonance Imaging-Based Radiomic Profiles Predict Patient Prognosis in Newly Diagnosed Glioblastoma Before Therapy.* Tomography, 2016. **2**(3): p. 223-228.

QIN PUBLICATIONS AND PRESENTATIONS FOR 2016

§ Manuscripts

1. Paulson, E.S., D.E. Prah, and **K.M. Schmainda**, Spiral Perfusion Imaging with Consecutive Echoes (SPICETM) for the Simultaneous Mapping of DSC- and DCE-MRI Parameters in Brain Tumor Patients: Theory and Initial Feasibility. Tomography, 2016: p. (In Press).
2. Nguyen, H.S., et al., Progressing Bevacizumab-Induced Diffusion Restriction Is Associated with Coagulative Necrosis Surrounded by Viable Tumor and Decreased Overall Survival in Patients with Recurrent Glioblastoma. AJNR Am J Neuroradiol, 2016.
3. McGarry, S.D., et al., Magnetic Resonance Imaging-Based Radiomic Profiles Predict Patient Prognosis in Newly Diagnosed Glioblastoma Before Therapy. Tomography, 2016. **2**(3): p. 223-228.

§ Abstracts

1. Anderson C, **Schmainda K**, Muftuler LT, Nencka A, Koch K. "Volume-Parcellated Quantitative Susceptibility Mapping." 24th Annual Meeting of the International Society of Magnetic Resonance in Medicine (ISMRM), Singapore (2016). ***Summa Cum Laude Award**
2. Mickevicius NJ, Cochran EJ, Rand SD, Connelly J, **Schmainda KM**, LaViolette PS. "Brain tumor imaging based, histology trained maps (IBHTMs) of cellularity predict tumor presence in pathologically confirmed regions sampled ex-vivo." 23rd Annual Meeting of the International Society for Magnetic Resonance in Medicine, Toronto, Canada. May 2015.
3. Salmon AE, Pellatt BJ, Mickevicius NJ, Cochran EJ, **LaViolette PS**. "In vivo MRI-based 3-D Printed Molds and Individualized Tissue Sectioning Apparatuses Improve MRI-Histopathologic Co-Registration in Brain Cancer Patients." 23rd Annual Meeting of the International Society for Magnetic Resonance in Medicine, Toronto, Canada. May 2015.
4. LaViolette PS, Mickevicius NJ, Cochran EJ, Connelly J, **Schmainda KM**, Rand SD. Brain tumor cellularity maps trained with co-registered histology predict tumor presence in pathologically confirmed regions sampled ex-vivo. World Molecular Imaging Congress Annual Meeting, Honolulu, HI. Sept. 2015.
5. Al-Gizawiy MM, Doan N, **Schmainda KM**, Mueller WM, Connelly JM, Mirza S. "Glioma characterization using the novel biomarker acid ceramidase." 21st Annual Meeting of the Society for Neuro-Oncology, Scottsdale, AZ (2016).
6. Al-Gizawiy MM, Doan N, **Schmainda KM**, Palomares J, Mueller WM, Connelly JM, Mirza S. "Acid ceramidase is a supplementary biomarker of survival in glioma." 21st Annual Meeting of the Society for Neuro-Oncology, Scottsdale, AZ (2016).

U01 CA179106: Biomarkers for Staging and Treatment Response Monitoring of Bladder Cancer

University of Michigan (2)

Lubomir Hadjiyski, PhD

INTRODUCTION

Bladder cancer is a common type of cancer that can cause substantial morbidity and mortality among both men and women. Bladder cancer causes over 16,870 deaths per year in the United States [1]. It is estimated that 79,030 new bladder cancer cases will be diagnosed in 2017. Early diagnosis and treatment of these lesions is important to reduce the morbidity, mortality and their attendant costs compared to diagnosis at a later, more symptomatic stage that might involve deep invasion and/or metastasis.

Correct staging of the bladder cancer is crucial for the decision of neoadjuvant chemotherapy treatment and minimizing the risk of under-treatment or over-treatment [2-8]. Only patients with stage T2 to T4 of muscle-invasive operable urothelial carcinoma of the bladder are recommended for treatment with neoadjuvant chemotherapy. If the response to chemotherapy can be estimated with sufficient accuracy and precision, it is possible to identify those patients that do not respond, stop the treatment early, and seek alternative treatment [8]. CT is an effective non-invasive modality for measuring primary site gross tumor volume (GTV) and the addition of MRI is on the rise. GTV has been used as a biomarker for predicting treatment outcome of bladder tumors [9]. Other pathological information and diagnostic test (bimanual evaluation, cystoscopy) results and immunohistochemical biomarkers are also useful for staging and treatment response monitoring. Although CT and MRI are promising methods for evaluation of a variety of bladder cancers, the time and costs required for the clinicians to outline cancer margins on a large number of CT and MRI slices for each case makes it difficult to advocate the use of this method for GTV estimation of every patient and of every pre- and post-treatment tumor evaluation.

The goal of this project is to develop a novel multimodality quantitative image analysis tool for bladder cancer (QIBC) to assist radiologists in estimation of GTV and analysis of image characteristics, thereby improving the efficacy of image biomarkers. The QIBC will be designed to use either one or more than one modalities from CT and MRI.

Another goal of this project is to develop novel decision support systems CDSS-S and CDSS-T for bladder cancer staging and for monitoring of bladder cancer treatment response based on multi-modality image-based, pathology-based and immunohistochemical biomarkers. The proposed QIBC, CDSS-S and CDSS-T have the potential to provide non-invasive, objective, and reproducible decision support, thereby reducing the subjectivity and variability in these processes. In order to achieve these goals we are performing the following specific tasks: (1) to collect a database of multi-modality MR, CT exams of

bladder cancers for development, training and testing of the QIBC and CDSS algorithms; (2) to develop advanced computer vision techniques to quantitatively estimate bladder GTV and image characteristics; and (3) to develop predictive models using machine learning techniques to combine multimodality image based, pathological and immunohistochemical biomarkers for cancer staging and determination of non-responders.

In addition, although we will focus on the specific application to the bladder tumors in this project, we plan to design the image analysis and decision support tools in a modular, expandable, and re-trainable framework. The software packages will be versatile and can be adapted to other tumor types or imaging modalities in the future by proper retraining with case samples of the tumor type of interest and expansion of the decision support tools as needed. Therefore, the development of the QIBC, CDSS-S and CDSS-T will potentially benefit not only the bladder tumor patients but also patients with other types of tumors that require staging and monitoring of treatment response.

DISCUSSION OF PROGRESS

§ Specific Aim 1: Data Collection

During the current time period of the project we have collected additionally 102 CTU bladder cancer cases from CTU examinations performed at University of Michigan. This includes 81 pre- and post- neoadjuvant chemotherapy treatment cases with clinical stage larger than T1, and 21 cases of which the clinical stage were called T1 and did not underwent neoadjuvant chemotherapy treatment. As a result we have collected in total 226 CTU bladder cancer cases: 132 pre- and post- neoadjuvant chemotherapy treatment cases and 94 cases of clinical stage T1. For each patient, the images are downloaded from the PACS system. The treatment records, pathology reports, and the clinically estimated treatment outcome after completion of the chemotherapy, are collected from patient files. All collected images and clinical information are stored into our CAD Lab information infrastructure (CADii). At present all patients undergo 3 cycles of chemotherapy. After completion of the chemotherapy treatment, the patients undergo radical cystectomy. The gold reference standard for the chemotherapy treatment outcome is determined by histopathology findings after radical cystectomy. Our clinical co-investigators marked each lesion and provided descriptors seen on the images. Two radiologists have manually drawn 3D outlines as gold standard for 171 cases.

In addition, we are part of a team which has started prospective collection of pathological information, diagnostic test results, immunohistochemical biomarkers, and CT scans from bladder cancer patients after the first cycle of chemotherapy. The protocol for data collection is approved by IRB. We have started the data collection. We have full access to the collected data. So far we have collected 13 cases. This will be very valuable dataset allowing us to develop tools for very early prediction of response to treatment.

§ Specific Aim 2: Design of quantitative image analysis tool (QIBC) for evaluation of bladder GTV and image characteristics

For both decision support systems CDSS-S and CDSS-T, an important component is to quantify the bladder gross tumor volume (GTV) and image characteristics. During the current time period of the project we have continued the development of a quantitative image analysis tool for bladder cancer, QIBC, specifically designed for these applications. We have been exploring further the use of a deep learning convolution neural network (DL-CNN) in QIBC. The details of the QIBC design and evaluation of the segmentation of the bladder lesions and bladder wall thickenings are presented in the following:

Segmentation of Inner and Outer Bladder Wall using Deep-Learning Convolutional Neural Network in CT Urography

We have explored the use of a deep-learning convolutional neural network (DL-CNN) to segment the bladder wall. This task is challenging due to differences in the wall between the contrast and non-contrast-filled regions, significant variations in appearance, size, and shape of the wall among cases, overlap of the prostate with the bladder wall, and the wall being extremely thin and occasionally invisible compared to the overall size of the bladder.

Methods: We trained a DL-CNN to distinguish the bladder wall from the inside of the bladder and the outside of the bladder using neighborhood information. A training set of about 240,000 regions of interest (ROIs) (Figure 3) were extracted from training cases for which the boundaries of the inner and outer wall of the bladder had been manually drawn. Half of the 16x16-pixel ROIs were determined to include the bladder wall and the other half were selected to exclude the bladder wall with some being inside the bladder wall and the rest outside the bladder entirely. The DL-CNN trained on these ROIs was applied to the test cases slice by slice to generate a bladder wall likelihood map where the gray level of a given pixel represents the likelihood that a given pixel would belong to the bladder wall. In addition, we used the DL-CNN likelihood map as an energy term in the energy equation of a cascaded level sets method to segment the inner and outer bladder wall (Figure 1 and Figure 2). A data set of 173 cases collected as described in Specific Aim 1 was used in this study. The data set was randomly split into two independent sets of training (81 cases) and testing (92 cases). Of this data set, 79 of the training cases and 37 of the test cases were hand outlined for both the inner and outer wall and used in this study. The DL-CNN segmentation with level sets was compared to these 3D hand-segmented contours as a reference standard. The accuracy of the segmentation was evaluated with four performance metrics: average volume intersection %, average % volume error, average absolute % volume error, and average distance. These performance measures were applied to the inner and outer wall contours independently. We compared the outer wall contours to the bladder segmentation contours based on our previous method. We also evaluated the accuracy of the bladder wall segmentation by the average bladder wall volume intersection %, average % bladder wall volume error, and average absolute % bladder wall volume error.

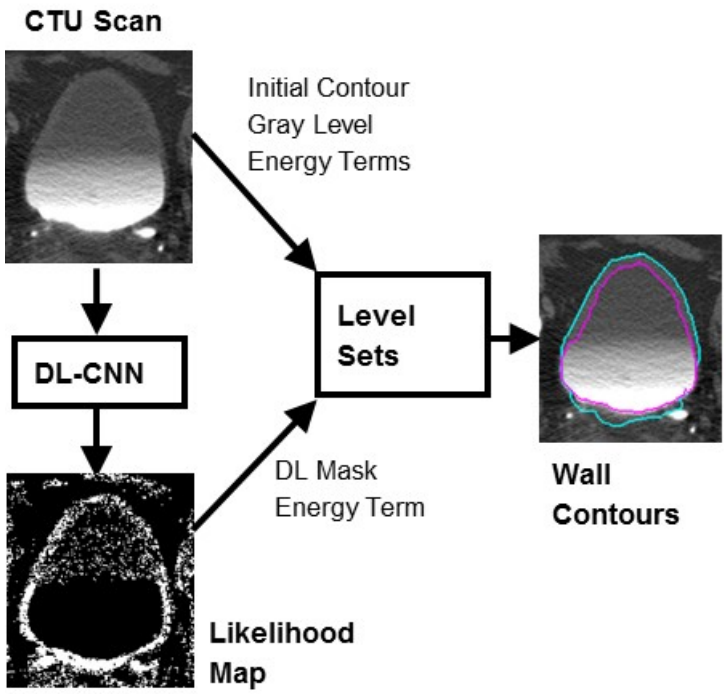


Figure 1: Flowchart of DL-CNN segmentation that shows generation of a bladder wall likelihood map and use of level sets to obtain inner and outer wall contours. The DL-CNN likelihood map was used as an energy term in the energy equation of a cascaded level sets method.

(a)

(b)

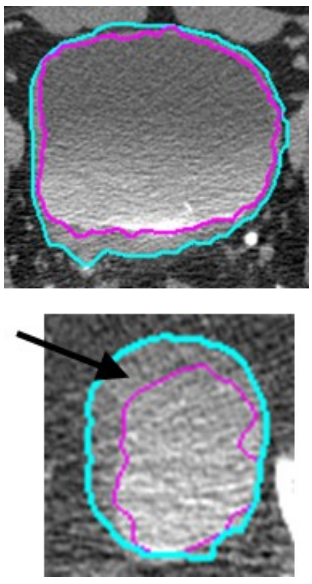


Figure 2: Example of bladder segmentation. The blue contour is for the outer wall and the pink contour is for the inner wall. (a) The DL-CNN effectively follows the inner and outer wall with slight over-segmentation of both. (b) The DL-CNN under-segments the inner contour, but successfully segments the thickened wall pointed to by the arrow.

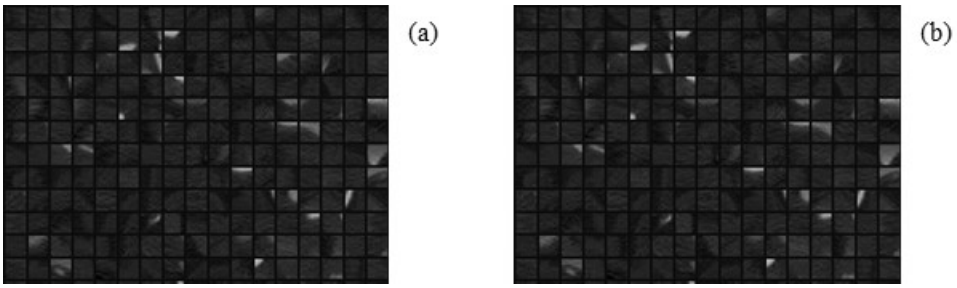


Figure 3: Subset of 240,000 ROIs used to train the DL-CNN. Each ROI is 16x16 pixels. (a) ROIs that are labeled as within the bladder wall. (b) ROIs that are labeled as outside the bladder wall.

Results: For the training set, the inner wall contour achieved the average volume intersection %, average % volume error, average absolute % error, and average distance of 90.0±8.7%, -4.2± 18.4%, 12.9±13.9%, and 3.0±1.6mm (Table 1). For the test set, the inner wall achieved values of 86.9±9.6%, -8.3±37.7%, 18.4±33.8%, and 3.4±1.8mm respectively. For the training set, the outer wall contour achieved the values of 93.7±3.9%, -7.8±11.4%, 10.3±9.3%, and 3.0±1.2mm respectively. For the test set, the outer wall contour achieved values of 87.5±9.9%, -1.2±20.8%, 11.9±17.0%, and 3.5±2.3mm respectively (Figure 4). The outer bladder wall segmentation was compared to the bladder segmentation based on our previous method in Table 2. If the segmented bladder wall was evaluated with the average bladder wall volume intersection %, average % bladder wall volume error, and average absolute % bladder wall volume error, the values were 61.0±11.4%, -13.7±49.1%, and 34.5±37.3%, respectively, for the training set, and 54.6±10.4%, 10.7±28.0%, and 25.1±15.8%, respectively, for the test set. The direct measurement of the bladder wall obtained less accurate results because slight deviations of the wall contour would lead to a much larger % error due to the much smaller wall volume compared to the inner and outer bladder volume.

Conclusions: A DL-CNN with level sets can effectively segment bladder walls from the inner bladder and outer structures despite a lack of consistent distinctions along the inner wall. The outer wall segmentation was improved compared to our previous method and the DL-CNN was also able to segment the inner bladder wall with similar results.
(QIN Publications and Presentations: #10).

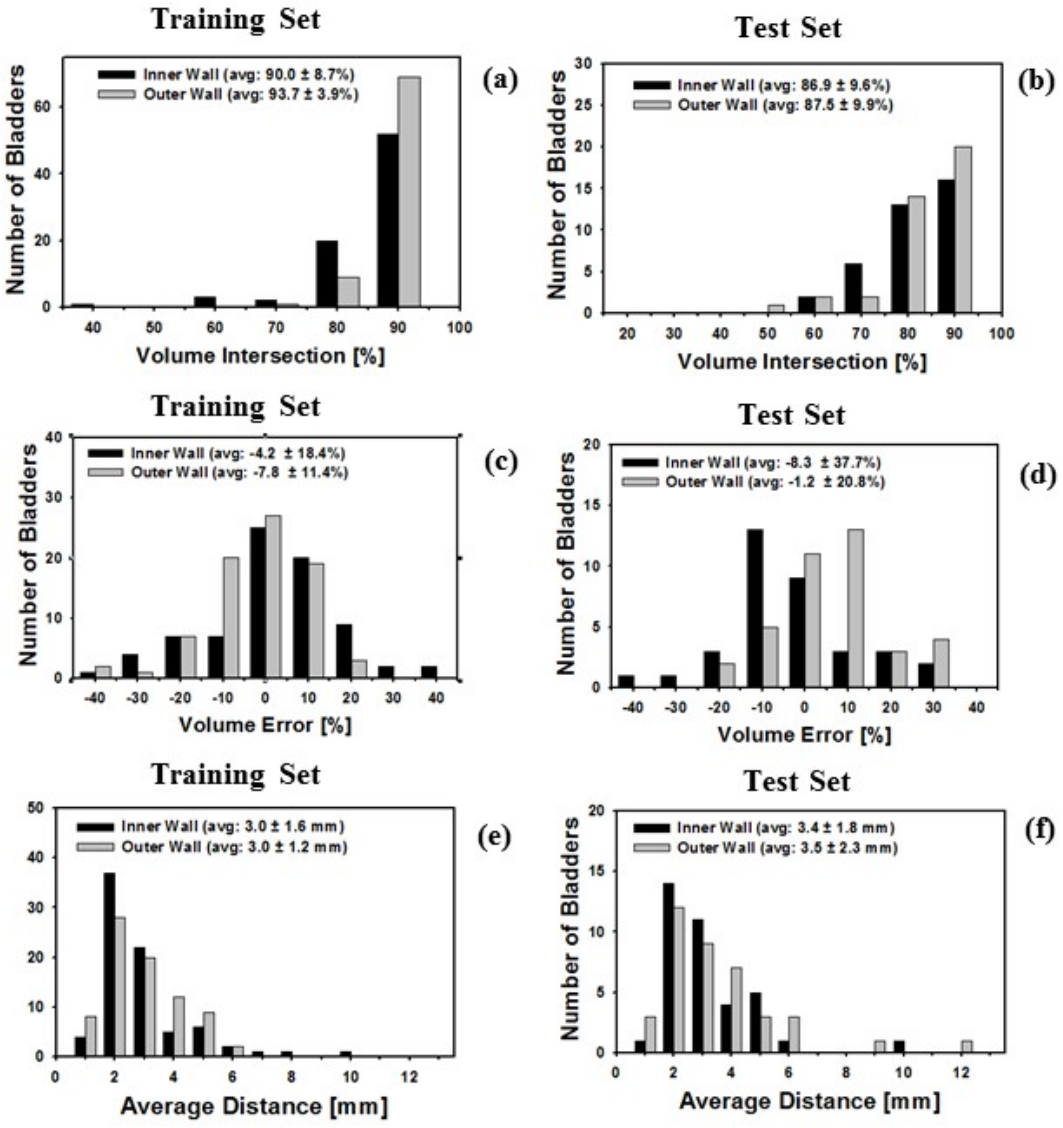


Figure 4: Histograms of the (a,b) volume intersection %, (c,d) volume % error, and (e,f) average distance for the training and test sets, respectively for the inner and outer bladder wall segmentations

	Inner Training	Inner Testing	Outer Training	Outer Testing
Volume Intersect %	90.0 ± 8.7	86.9 ± 9.6	93.7 ± 3.9	87.5 ± 9.9
Volume % Error	-4.2 ± 18.4	-8.3 ± 37.7	-7.8 ± 11.4	-1.2 ± 20.8
Absolute Volume % Error	12.9 ± 13.9	18.4 ± 33.8	10.3 ± 9.3	11.9 ± 17.0
Average Distance(mm)	3.0 ± 1.6	3.4 ± 1.8	3.0 ± 1.2	3.5 ± 2.3

Table 1: Performance metrics for inner and outer bladder wall.

Method	Training Volume Intersection %	Testing Volume Intersection %	Training Volume % Error	Testing Volume % Error
DL-CNN LS	93.7 ± 3.9	87.5 ± 9.9	-7.8 ± 11.4	-1.2 ± 20.8
Previous	86.5 ± 6.4	77.6 ± 12.0	7.3 ± 9.1	18.0 ± 12.5

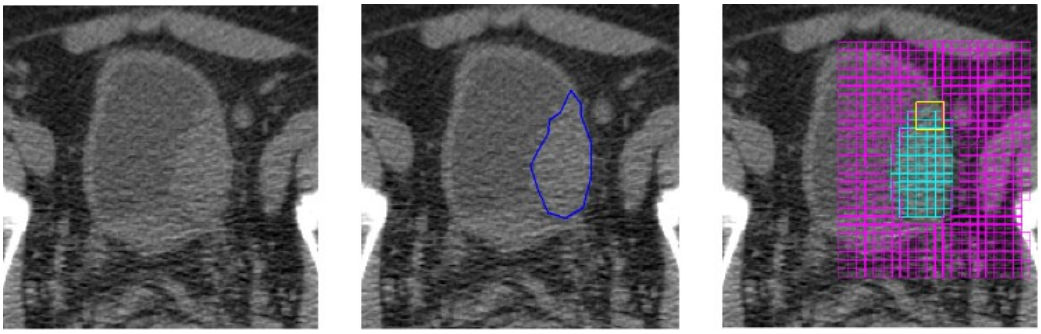
Table 2: Comparison between current and previous DL-CNN outer wall contours for volume intersection % and volume percent error.

Segmentation of Bladder Cancer for Treatment Response Assessment using Deep-Learning Convolution Neural Network

In this study, we applied DL-CNN to bladder lesion segmentation. The DL-CNN was trained to recognize the patterns in the regions that were inside and outside of the bladder lesion and generate a lesion likelihood map. Minor refinement on the likelihood map was performed by level sets to obtain the segmented boundaries of the bladder cancer.

Methods: A data set of 62 cases was collected as described in Specific Aim 1. All of the patients in the data set had undergone CT examination before and after chemotherapy. The data set contained 64 tumors. A reference standard for the computerized segmentation was obtained via 3D hand-segmented contours of the bladder tumors in the pre- and post-treatment CTs by two radiologists (reference standard 1 and reference standard 2,

respectively). The DL-CNN by Krizhevsky et al. called cuda-convnet [10, 11] was used. The neural network was trained to classify regions of interests (ROIs) on 2D slices as being inside or outside of the bladder cancer. Details on the DL-CNN can be found in the literature [12]. The DL-CNN was trained with the pre-treatment scans of the cases. For each axial slice of the cases, a large number of overlapping 16 x 16-pixel ROIs were extracted from the region including the cancer marked by the radiologist. If more than 80% of an ROI was within the hand-outlined bladder cancer, the ROI was labeled as being inside of the cancer, whereas the ROI had to be completely outside of the cancer in order for it to be classified as being outside the cancer. ROIs not labeled as either inside or outside of the cancer were excluded. Figure 5 shows an example of ROIs obtained from a CT slice. The number of ROIs within the two classes was balanced, resulting in approximately 65,000 ROIs. The output of the DL-CNN can be interpreted as the likelihood of an input ROI being classified into one of the two categories. Leave-one-case-out cross-validation was employed for this study. In each of the leave-one-case-out partitions, all ROIs associated with a case were removed and the DL-CNN was trained using the remaining ROIs. For each leave-one-case-out partition, the trained DL-CNN network was applied to the removed case to generate the bladder cancer segmentation likelihood map. Figure 6 shows the bladder cancer likelihood map for the CT slice shown in Figure 5. The DL-CNN was applied to the CT scan for both the pre- and post-treatment scans for each bladder cancer case.



(a) **(b)** **(c)**

Figure 5: An axial slice of a pre-treatment CT scan from a training case. (a) Cropped CT slice centered at the bladder. (b) Radiologist's hand-outline of the cancer overlaid on the CT slice. (c) ROIs extracted from this slice. The yellow ROI shows the size of a 16 x 16-pixel ROI. The ROIs are partially overlapping. The blue ROIs are labeled as inside the bladder cancer. The pink ROIs are labeled as outside the bladder cancer for training the DL-CNN.

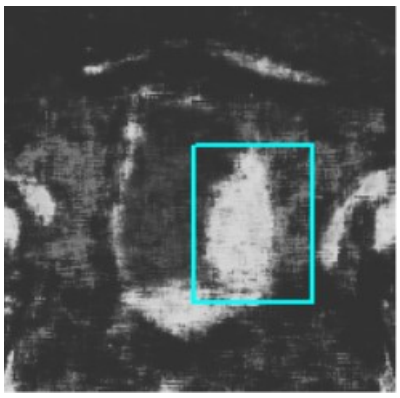


Figure 6: Bladder Cancer likelihood map of the CT slice shown in Figure 5. Regions that are highly likely to be bladder cancer have higher intensity values. The VOI that was used for this lesion is shown in blue. For demonstration purposes, the bladder cancer likelihood map was generated in the region around the entire bladder.

As seen in the example of Figure 6, the likelihood map identifies the bladder tumor region very well but the tumor boundary is not sharply demarcated. 3D and 2D level sets, are used to perform minor refinements to the contour. A 3D level set is applied to the initial segmentation surface, and the segmentation on each slice is further refined by a 2D level set. Details on the level sets used can be found in the literature [13]. Figure 7 shows the final contour of the bladder cancer on the CT slice from Figure 5 using the likelihood map shown in Figure 6. Segmentation performance was evaluated by comparing quantitatively the automatic segmentation results to the 3D hand-segmented contours. The average minimum distance, and the Jaccard index [14] between the hand-segmented contours and computer segmented contours were calculated.

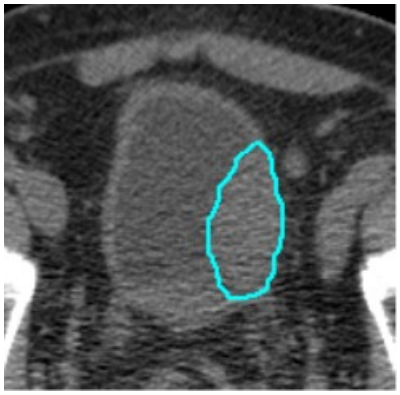


Figure 7: Bladder cancer segmentation on the CT slice shown in Figure 1 using the bladder likelihood map shown in Figure 6.

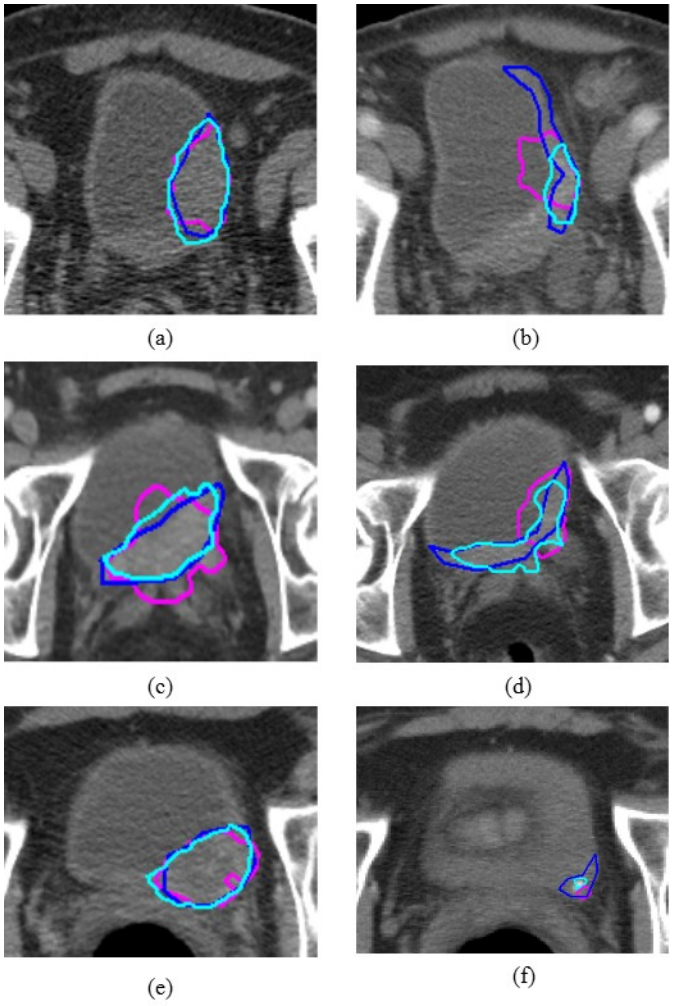


Figure 8: Examples of segmentations of bladder tumors in pre-treatment (a, c, e) and post-treatment (b, d, f) CT scans. The DL-CNN segmentation is shown in light blue. The AI-CALS segmentation is shown in pink. The hand outline is shown in dark blue. (a) DL-CNN segmentation with AI-CALS segmentation and hand outline for the cancer shown in Figure 5. Both computer methods segmented the lesion reasonably. (b) The cancer shrunk due to treatment, and became a part of the bladder wall. The DL-CNN under-segmented the cancer, not extending enough into the bladder wall. AI-CALS over-segmented the lesion, leaking into the bladder. (c) The DL-CNN segmentation outlined the cancer relatively accurately, while the AI-CALS segmentation leaked. (d) In this post-treatment scan, the cancer along the bladder wall was reasonably segmented by DL-CNN, while the AI-CALS was unable to follow the shape and leaked into the bladder. (e) Both DL-CNN and AI-CALS segmented the bladder cancer reasonably well, but the AI-CALS slightly under-segmented the cancer. (f) The bladder cancer responded to treatment, thus had shrunk considerably, making the segmentation difficult. Both the DL-CNN and the AI-CALS under-segmented the lesion.

Results: Examples of DL-CNN segmented bladder cancer on pre- and post-treatment CT scans, along with the AI-CALS segmentation, are shown in Figure 8. The segmentation performance measures of both the DL-CNN and AI-CALS methods compared with reference standard 1 are presented in Table 3. For all lesions, the difference in the average minimum distance was statistically significant with a p-value of 0.001, while the difference in the Jaccard index approached significance with a p-value of 0.058. The differences in the pre-treatment lesion segmentation performances were statistically significant with p-values of less than 0.001 and 0.015 for the average minimum distance and the average Jaccard Index, respectively. The differences in the post-treatment lesion segmentation performances did not reach statistical significance. The segmentation performance measures of the DL-CNN and AI-CALS methods compared with the two reference standards averaged over the pre-treatment lesions, post-treatment lesions, and both pre- and post-treatment lesions for a subset of 29 cases are presented in Table 4. None of the differences reached statistical significance for this subset of cases.

Conclusions: Our results demonstrate that DL-CNN is useful for 3D segmentation of bladder cancers for a variety of bladder cancer shapes and sizes. The DL-CNN and the AI-CALS methods were able to automatically segment the cancers, with results similar to those of the radiologists. This study suggests that computerized segmentation of bladder cancers using DL-CNN has the potential to assist in the assessment of tumor volume of bladder cancer by providing the more accurate 3D information without the extensive effort of manual segmentation. (QIN Publications and Presentations: #3)

		DL-CNN vs RS1	AI-CALS vs RS1	p-value
Average minimum distance	**Pre-treatment**	4.8 ± 2.3 mm	6.1 ± 3.6 mm	0.001*
AVDIST	**Post-treatment**	4.6 ± 1.8 mm	4.9 ± 2.6 mm	0.389
	Both	4.7 ± 2.1 mm	5.5 ± 3.2 mm	0.001*
Jaccard index	**Pre-treatment**	39.5 ± 17.1%	34.7 ± 15.8%	0.015*
JACCARD ^{3D}	**Post-treatment**	32.6 ± 17.8%	32.7 ± 14.4%	0.936
	Both	36.3 ± 17.7%	33.8 ± 15.1%	0.058
* Statistically significant at $p < 0.05$				

Table 3: Lesion segmentation evaluation using reference standard 1 (RS1). The results are shown in groups of pre-treatment, post-treatment, and both pre- and post-treatment lesions (126 lesions). The p-values from Student's two-tailed paired t-test for the differences between the DL-CNN and the AI-CALS segmentation methods are also shown. Some post-treatment lesions were determined to have shrunk completely by radiologist, thus no segmentation was performed

		DL-CNN vs RS1	AI-CALS vs RS1	DL-CNN vs RS2	AI-CALS vs RS2
Average minimum distance	Pre-treatment	4.8 ± 1.8 mm	5.3 ± 2.7 mm	4.9 ± 3.4 mm	4.5 ± 1.9 mm
AVDIST	Post-treatment	4.3 ± 1.7 mm	4.4 ± 1.8 mm	4.7 ± 3.1 mm	4.9 ± 3.7 mm
	Both	4.6 ± 1.8 mm	4.8 ± 2.3 mm	4.8 ± 3.2 mm	4.7 ± 2.9 mm
Jaccard index	Pre-treatment	45.3 ± 8.5%	42.5 ± 14.1%	46.8 ± 9.3%	42.8 ± 12.5%
JACCARD ^{3D}	Post-treatment	29.8 ± 17.7%	32.9 ± 14.8%	28.8 ± 19.7%	28.6 ± 18.2%
	Both	37.5 ± 15.8%	37.7 ± 15.2%	37.8 ± 17.8%	35.7 ± 17.1%

Table 4: Lesion segmentation evaluation results for a subset of 29 cases divided into pre-treatment, post-treatment, and both pre- and post-treatment lesions (58 lesions) between hand-segmented reference standards (RS1, RS2) by two different readers for DL-CNN and the AI-CALS segmentation methods. None of the paired differences between the two methods reached statistical significance for this subset, probably due to the small sample size.

§ Specific Aim 3: Design of CDSS-S and CDSS-T decision support systems to assist clinicians in staging and monitoring of treatment response of bladder cancer.

During the current time period of the project we have continued the development of the decision support systems for bladder cancer staging and treatment response monitoring.

Specific Aim 3.1: Design of computer decision support system (CDSS-S) for bladder cancer staging.

Correct staging of bladder cancer is crucial for the decision of neoadjuvant chemotherapy treatment and minimizing the risk of under-treatment or over-treatment. At clinical staging, approximately 30% of patients are under-staged or over-staged. Subjectivity and variability of clinicians in utilizing various diagnostic information may lead to inaccuracy in staging bladder cancer. An objective decision support system that merges the information in a predictive model based on statistical outcomes of previous cases and machine learning may assist clinicians in making more accurate and consistent staging assessments.

We have continued the design of CDSS-S. During the current time period of the project we have developed a CDSS-S to stage bladder cancer based on different machine learning techniques. The details of the CDSS-S design and evaluation are presented in the following:

Methods: A data set consisting of 84 bladder cancer lesions from 76 CTU cases collected as described in Specific Aim 1, was used to train and test the classifier. The cases were grouped into two classes based on pathological stage ≥T2 or below T2, which is the decision threshold for neoadjuvant chemotherapy treatment clinically. There were 43

cancers below stage T2 and 41 cancers at stage T2 or above. All 84 lesions were automatically segmented using our previously developed auto-initialized cascaded level sets (AI-CALS) method. Each lesion was marked with a bounding box by a radiologist. This box served as the input to our 3D AI-CALS automated segmentation system. The segmentation of bladder lesions can be challenging as some lesions are very small, subtle in contrast, or have irregular boundaries. Additionally, lesions are sometimes located in the non-contrast enhanced region of the bladder and the contrast between the lesion boundary and the surrounding background is very low.

Morphological [15] and texture features [16, 17] were extracted. The morphological features included gray level features, contrast features, and the lesion volume. The texture features included filtered Disarthy East-West and Horizontal direction features, and the gray level radial gradient direction features. The features were divided into subspaces of 26 morphological features only, 65 texture features only, and a combined set of 91 morphological and texture features.

The data set was split into Set 1 and Set 2 for two-fold cross validation. The cancers were evenly and randomly split into two sets with 42 cancers each by balancing the number of cancers of each class. Set 1 consisted of 22 cancers below stage T2 and 20 cancers stage T2 or above. Set 2 consisted of 21 cancers below stage T2 and 21 cancers stage T2 or above. The average size for cancers of stage <T2 and ≥T2 in Set 1 were 26.4±17.3 mm and 45.6±19.1 mm, respectively (Figure 9). The average size for cancers of stage <T2 and ≥T2 in Set 2 were 27.3±10.8 mm and 40.6±17.3 mm, respectively (Figure 9). Stepwise feature selection was used to select the most effective features. A linear discriminant analysis (LDA), a neural network (NN), a support vector machine (SVM), and a random forest (RAF) classifier were used to combine the features into a single score. In the first fold, Set 1 was used for feature selection and for training of the classifiers. The trained classifiers were then tested on Set 2. In the second fold, feature selection and classifiers' training were performed on Set 2 and then tested on Set 1. The classification accuracy was quantified using the area under the ROC curve (A_z) for both the training and test sets.

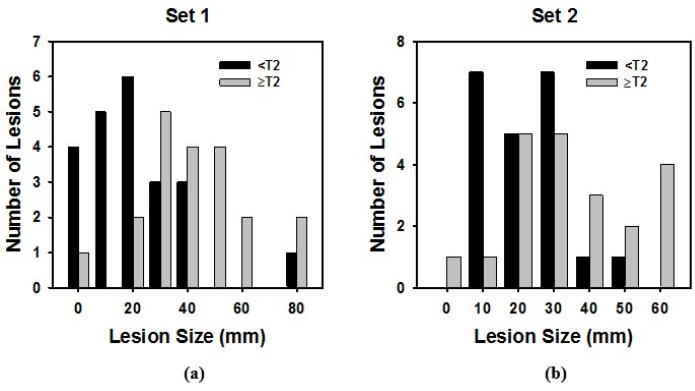


Figure 9: Distribution of tumor sizes for Set 1 and Set 2. (a) Set 1: The average tumor sizes of stage < T2 and ≥ T2 were 26.4±17.3 mm and 45.6±19.1 mm respectively. (b) Set 2: The average tumor sizes of stage < T2 and ≥ T2 were 27.3±10.8 mm and 40.6 ±17.3 mm respectively.

Results: The performance of the classifiers based on different machine learning techniques, the LDA, NN, SVM, and RAF is summarized in Table 5. Different feature spaces containing the morphological features, the texture features, and the combined set of both morphological and texture features were used for classification. The features selected with LDA were used in the SVM and NN classifiers. The LDA classifier with morphological features achieved a training A_z of 0.91 on Set 1 and a test A_z of 0.81 on Set 2. For training on Set 2 it achieved a A_z of 0.97 and a test A_z of 0.90 on Set 1. The selected features on the training sets included volume, a contrast feature, and gray level features. The test A_z of the NN for Set 1 and Set 2 was 0.88 and 0.91 respectively. The SVM achieved test A_z of 0.88 on Set 1 and test A_z of 0.90 on Set 2. The test A_z of the RAF for Set 1 and Set 2 was 0.83 and 0.88 respectively. The distribution of the discriminant scores from the four classifiers for testing on Set 1 and Set 2 in two fold cross-validation in the morphological feature space are presented in Figure 10. It can be observed that most of the classifiers were able to provide a relatively good separation between the two classes.

		LDA		NN		SVM		RAF	
Feature Type	***Number of Features***	***Training***	***Testing***	***Training***	***Testing***	***Training***	***Testing***	***Training***	***Testing***
Morphological Features									
Training (Set 1)	4	0.91	0.81	0.96	0.91	0.95	0.9	1	0.88
Testing (Set 2)									
Training (Set 2)	4	0.97	0.9	0.98	0.88	0.97	0.88	1	0.83
Testing (Set 1)									
Texture Features									
Training (Set 1)	2	0.91	0.88	0.95	0.92	0.92	0.89	1	0.97
Testing (Set 2)									
Training (Set 2)	7	1	0.91	1	0.89	1	0.91	1	0.89
Testing (Set 1)									
Combined Features									
Training (Set 1)	3	0.92	0.9	0.97	0.95	0.92	0.89	1	0.96
Testing (Set 2)									
Training (Set 2)	7	1	0.89	1	0.91	1	0.92	1	0.86
Testing (Set 1)									

Table 5: Summary results for LDA, NN, SVM and RAF classifiers in morphological, texture, and combined feature spaces.

By using the texture features the LDA classifier achieved a test A_z of 0.91 on Set 1 and a test A_z of 0.88 on Set 2. When trained on Set 1 and Set 2 the LDA classifier selected subsets of the filtered Disarthy East-West direction features, the filtered Disarthy Horizontal direction features and the gray level radial gradient direction features. The test A_z of the NN classifier for Set 1 and Set 2 was 0.89 and 0.92, respectively. The SVM classifier achieved test A_z of 0.91 on Set 1 and test A_z of 0.89 on Set 2. The test A_z of the RAF classifier for Set 1 and Set 2 was 0.89 and 0.97, respectively.

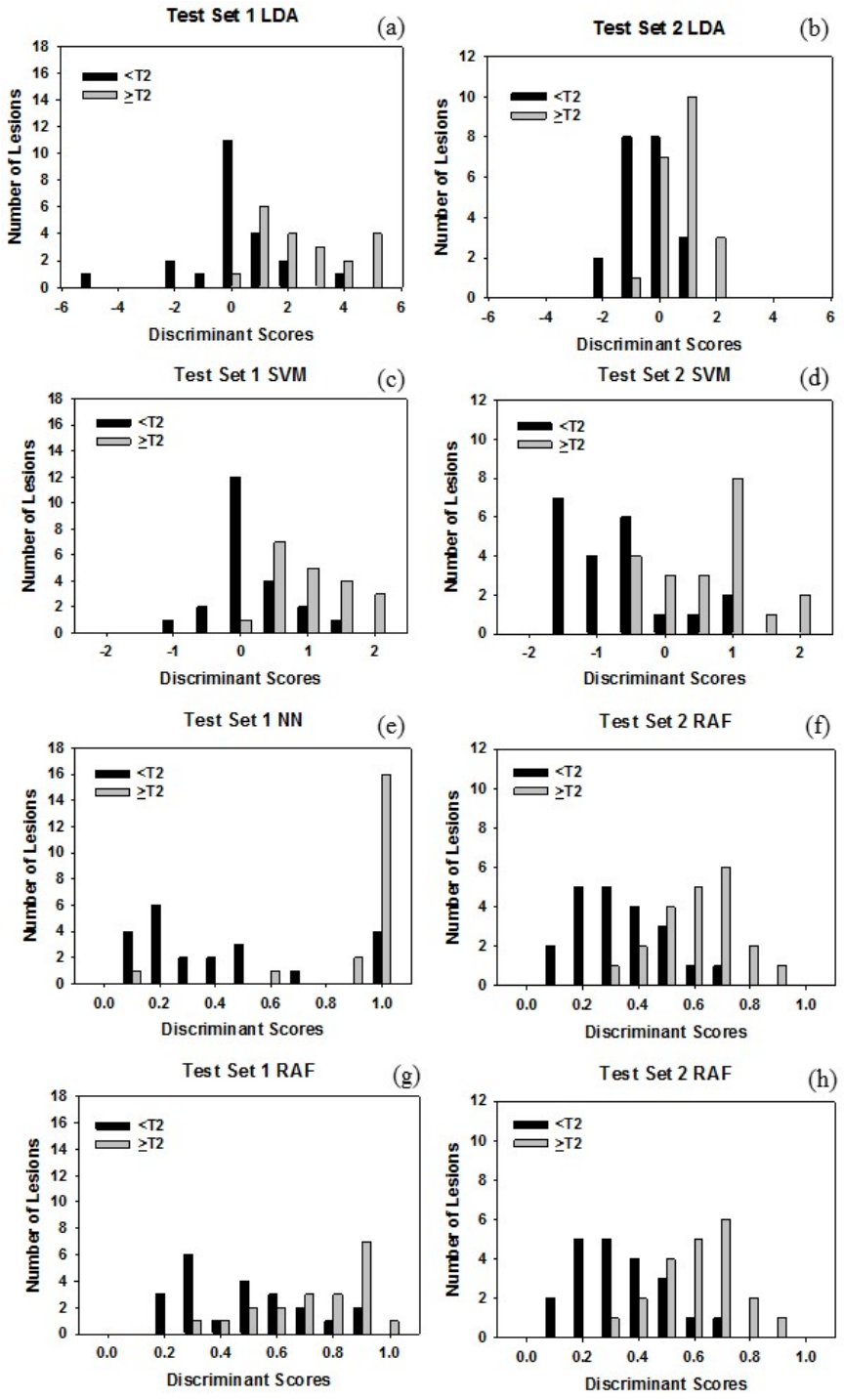


Figure 10: Distribution of the classifiers discriminant scores for testing on Set 1 and Set 2 in two-fold cross validation using the morphological features. (a) LDA (Set 1) $A_z = 0.90$, (b) LDA (Set 2) $A_z = 0.81$, (c) SVM (Set 1) $A_z = 0.88$, (d) SVM (Set 2) $A_z = 0.90$, (e) NN (Set 1) $A_z = 0.88$, (f) NN (Set 2) $A_z = 0.91$, (g) RAF (Set 1) $A_z = 0.83$, (h) RAF (Set 2) $A_z = 0.88$.

When the morphological and the texture features were combined, the LDA classifier achieved a test A_z of 0.89 on Set 1 and a test A_z of 0.90 on Set 2. When trained on Set 1 and Set 2 the LDA classifier selected a contrast feature, subsets of the filtered Disarthy Horizontal direction features, and subsets of the gray level radial gradient direction features. The test A_z of the NN classifier for Set 1 and Set 2 was 0.91 and 0.95, respectively. The SVM classifier achieved test A_z of 0.92 on Set 1 and test A_z of 0.89 on Set 2. The test A_z of the RAF classifier for Set 1 and Set 2 was 0.86 and 0.96, respectively. The test ROC curves for all of the classifiers when tested on Set 1 and Set 2 in the two fold cross-validation in the different feature spaces are shown in Figure 11.

The classifiers achieved slightly higher A_z values in the texture feature space than in the morphological and combined feature spaces; however, the differences did not achieve statistical significance. Examples of bladder cancers with stages $\geq$ T2 or $<$ T2 with the corresponding computer outlines and classifier scores are presented in Figure 12.

Conclusion: Staging of bladder cancer is crucial in minimizing the risk of under-treatment or over-treatment. The performance of the LDA classifier in staging different bladder cancer lesions shows promise in assessing bladder cancer stage using quantitative image analysis from CTU. Our preliminary results demonstrate the feasibility of an image-based predictive model that can assist with bladder cancer staging. (QIN Publications and Presentations: #7)

Specific Aim 3.2: Design of computer decision support system (CDSS-T) for bladder cancer treatment response monitoring.

Early assessment of therapeutic efficacy and prediction of treatment failure would help clinicians decide whether to discontinue chemotherapy at an early phase before additional toxicity develops, and thus improve the quality of life of a patient and reduce unnecessary morbidity and cost. The ultimate goal is to improve survival for those with a high risk of recurrence while minimizing toxicity to those who will have minimal benefit. Therefore, development of an accurate and early predictive model of the effectiveness of neoadjuvant chemotherapy is important for patients with bladder cancer.

We have continued the design of CDSS-T by merging (1) image biomarkers obtained by QIBC, and (2) changes in descriptors of local tumor tissue characteristics. We designed predictive models using the image biomarkers and local tumor descriptors to distinguish between bladder cancers that have fully responded to chemotherapy and those that have not, based upon analysis of pre- and post-treatment CT images. We evaluated three unique predictive models, which employ different fundamental design principles: 1) a pattern recognition method (DL-CNN), 2) a more deterministic radiomics feature based approach (F-SL), and 3) a bridging method between the two, which extracts features from image patterns (F-ROI). We studied both the properties of the different predictive models and the relationship between these different radiomics approaches. We also compared the performance of the models in predicting a complete response of bladder cancer to neoadjuvant chemotherapy with that of expert physicians. The details of the study are presented below.

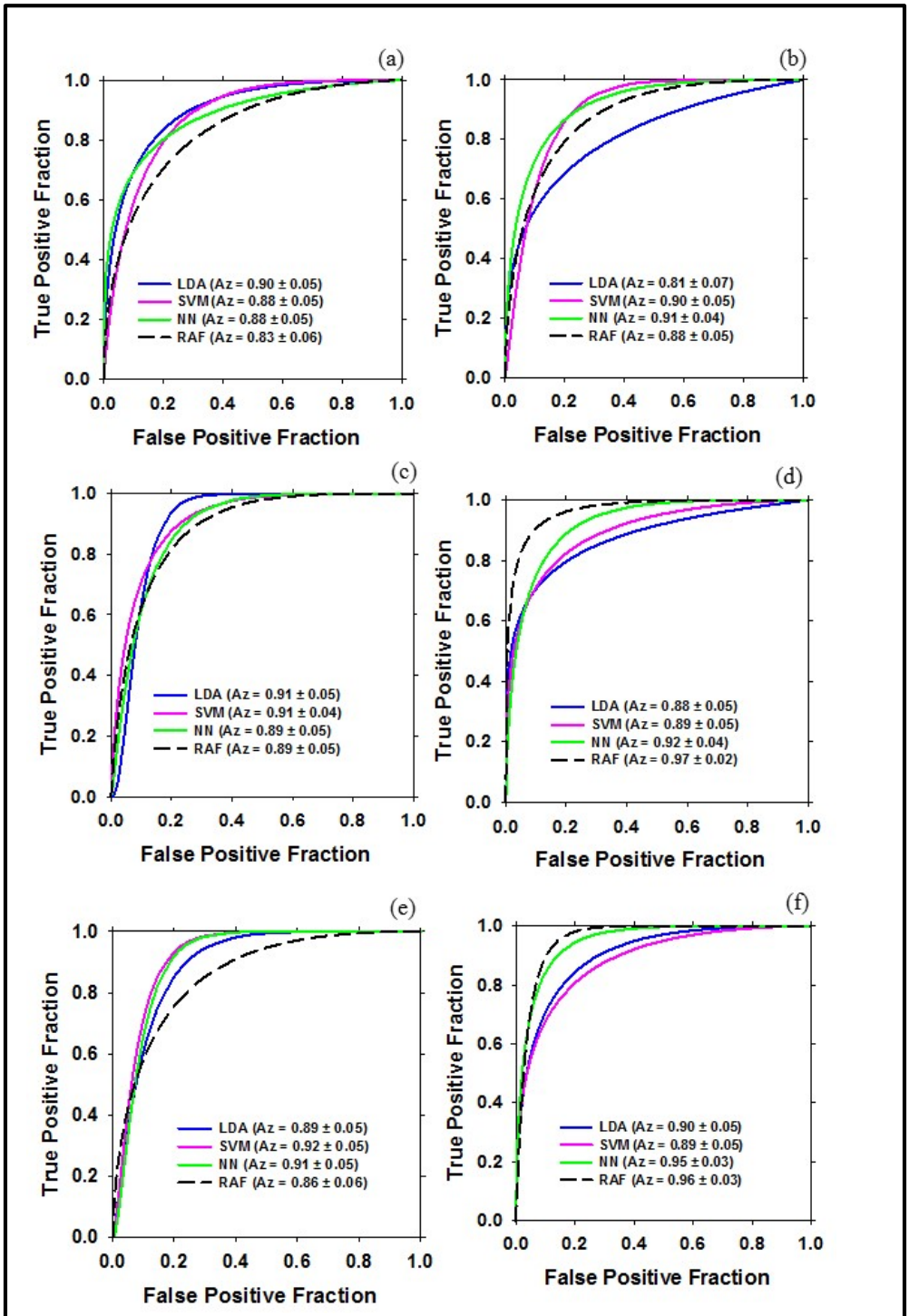


Figure 11: ROC curves for testing on Set 1 and Set 2 in two-fold cross validation for LDA, SVM, NN, and RAF classifiers: Left column: testing on Set 1, right column: testing on Set 2. (a) and (b) morphological features; (c) and (d) texture features; (e) and (f) combined features.

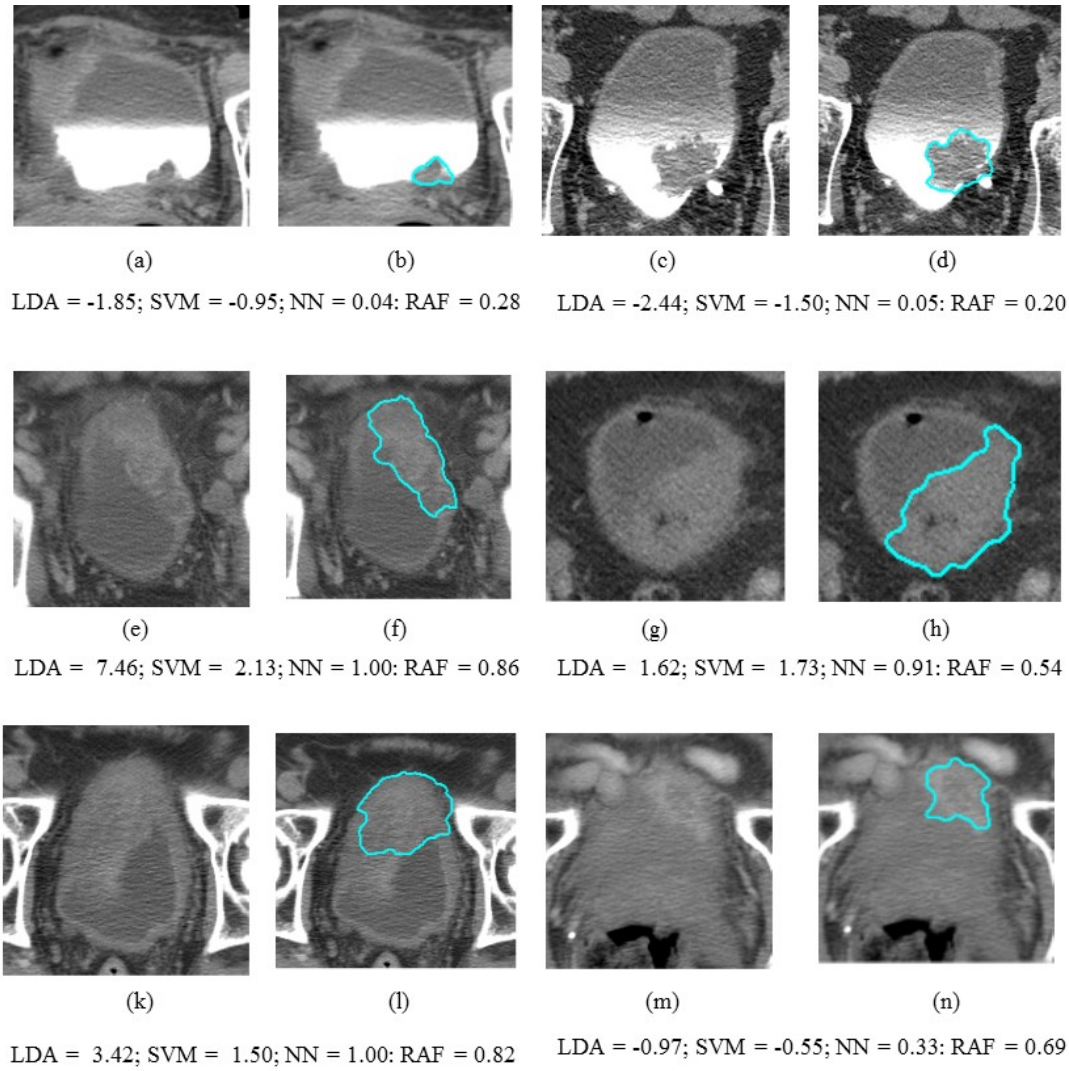


Figure 12: Examples of bladder cancers with stages ≥ T2 or < T2. The blue outlines represent the AI-CALS segmentation. The reported scores are test scores for the LDA, SVM, NN, and RAF classifiers based on the morphological features. The two cases in (a)(b) and (c)(d) both contained is a T1 stage cancer that was properly classified with low scores from all classifiers. (e)(f) is a T3 stage case that was properly classified with high scores from all classifiers. (g)(h) is a T2 stage case that was properly classified with high scores from all classifiers. (k)(l) is a case that was clinically identified as T1 pre-surgery but was identified as a T2 stage cancer post-surgery. The classifiers classified the cancer as ≥T2 with high scores. (m)(n) is T2 stage cancer that was incorrectly identified by the LDA, SVM, and NN classifiers with low scores and correctly identified by the RAF with a high score.

Methods: A training data set of 82 patients with 87 lesions who underwent pre- and post-neoadjuvant chemotherapy CTU scans was collected as described in Specific Aim 1. Using the 87 lesions, 104 pre- and post-treatment lesion pairs were generated, and 27% of the training set patients had T0 cancer stage after neoadjuvant chemotherapy. T0 stage corresponds to a complete response to treatment. An additional 41 patients with 43 lesions were collected as a test set. Fifty-four pre- and post-treatment pairs were generated from the 42 lesions, and 22% of the test set lesion pairs had T0 cancer after neoadjuvant chemotherapy. Cystectomy was performed at the end of treatment, and the cancer stage after treatment was used as the reference standard to determine if a patient responded to treatment. Bladder lesions in the CTU scans were segmented using our Auto-Initialized Cascaded Level Sets (AI-CALS) system.

Regions of interests (ROIs) were extracted from within the segmented lesions from corresponding pre- and post-treatment scans of a patient and were paired together in multiple combinations to generate pre-post-treatment paired ROIs (Figure 13). We trained a DL-CNN to distinguish between bladder lesions that were diagnosed as stage T0 post-treatment and those that were greater than stage T0. The "per-lesion" score was obtained by using the average value among the ROI scores associated with the lesion.

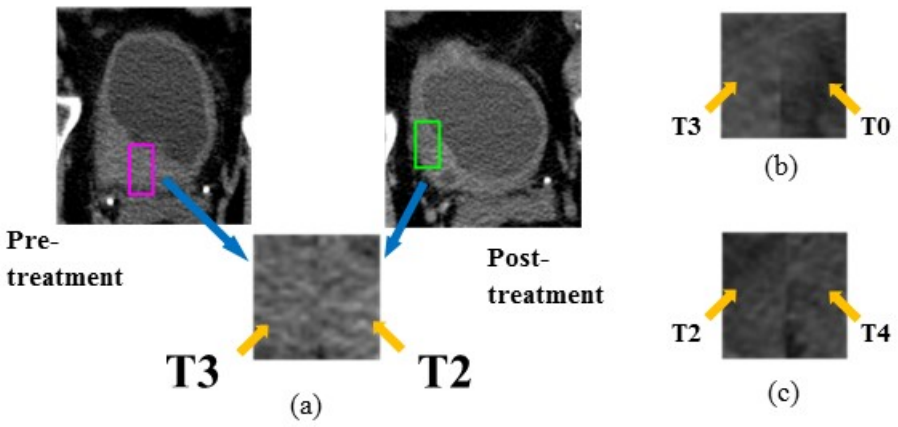


Figure 13: Creating ROIs to train the DL-CNN. (a) ROIs were generated by combining regions from the pre- and post-treatment scan lesions. In this example, the pre-treatment stage was T3, and the post-treatment stage was T2. Therefore, the ROI was labeled as being greater than stage T0 after treatment. (b) ROI of a case that was stage T3 pre-treatment and stage T0 after treatment. (c) ROI of a case that was stage T2 pre-treatment and stage T4 post-treatment. Therefore, the ROI was labeled as greater than stage T0 after treatment.

A radiomics-feature-based analysis was applied to the segmented lesions (RF-SL) to build a classifier for the prediction of complete responders to chemotherapy. Ninety-one features were extracted from every segmented lesion, which included morphological features, gray level features, texture features, and gradient field features. For every temporal lesion pair, the percent change between each radiomics feature extracted from the pre- and

post-treatment lesion was calculated. The percent change of each of the feature values before and after the treatment was calculated. Feature selection was performed and a random forest classifier (RAF) was trained to use the selected radiomics features to predict the likelihood of the post-treatment lesion being T0 stage.

Radiomic features from paired ROIs (RF-ROI) were also used to build a classifier for the prediction of complete responders to chemotherapy. Gray-level and texture features were extracted from the paired ROIs used for the DL-CNN. Thirty-eight features, including gray-level histogram statistics, and run length statistics features, were calculated for every ROI. The "per-lesion" features were generated by averaging the feature values among the ROIs associated with the lesion. Similar to the RF-SL model, feature selection was performed and a RAF classifier was trained to use the selected radiomics features to predict the likelihood of the post-treatment lesion being T0 stage.

An observer performance study with two experienced radiologists was also performed independently, in which the radiologist estimated the likelihood of stage T0 after viewing each pre-post-treatment CTU pair. ROC analysis was performed and the A_z was calculated for the DL-CNN and radiologists' estimates.

Results: Table 6 shows the performances for the DL-CNN, RF-SL, and RF-ROI methods, along with the radiologists' results for the test set. Figure 14 shows the ROC curves for the DL-CNN, F-SL, and F-ROI methods, and the radiologists for the test set. The test A_z values for prediction of T0 disease after treatment were 0.73 ± 0.08, 0.77 ± 0.08, 0.67 ± 0.08 for the DL-CNN, F-SL, and F-ROI methods, respectively. The two radiologists had A_z values of 0.76 ± 0.08 and 0.77 ± 0.07 on the test set. None of the pairwise differences in the methods reached statistical significance.

	DL-CNN	**RF-SL**	**RF-ROI**	**Radiologist 1**	**Radiologist 2**
AUC	0.73 ± 0.08	0.77 ± 0.08	0.69 ± 0.08	0.76 ± 0.08	0.77 ± 0.07

DL-CNN: Deep-learning convolution neural network
RF-SL: Features extracted from segmented lesions
RF-ROI: Features extracted from pre- and post-treatment paired ROIs
The area under the curve (A_z) is shown with the standard deviations

Table 6: Performances of bladder cancer treatment response assessment on the test set.

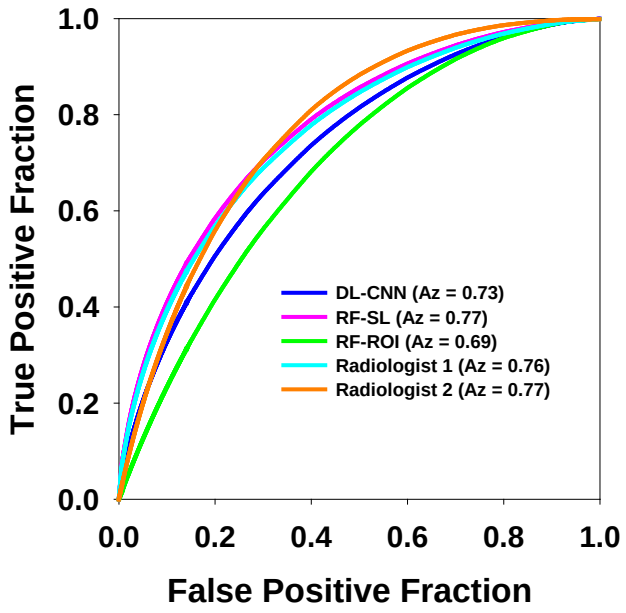


Figure 14: Test set ROC curves for the three models and two expert radiologists. The results from the test set for prediction of T0 stage after neoadjuvant chemotherapy for the three models. The differences between any pairs of A_z values did not reach statistical significance.

Examples of the treatment response prediction of pre- and post-treatment case pairs are shown in Figure 15. Given the fact that in some instances the computer models were correct about complete tumor responses and the radiologists were incorrect, we speculate that use of one or more of these models alongside a radiologist might improve the radiologist's ability to identify patients who responds fully to chemotherapy. In cases like that in Figure 15(d), radiologists will generally decide that the case is a non-responder because they see residual bladder wall thickening, which is an indicator of cancer. If the computer models suggested that there was a high likelihood of T0 after treatment in this case, it might lead the radiologists to re-evaluate their decision, and, possibly come to a different (and correct) conclusion. Of course, it is also possible for the computer models to sway a radiologist's decision in the wrong direction. Further study of the accuracy of the computer models in tandem with radiologist assessment is needed to determine whether or not such decision support systems will improve radiologist performance in treatment response assessments for bladder cancers.

Conclusion: This study indicates the potential of using DL-CNN and image features obtained by QIBC, as well as the changes in the descriptors of local tumor tissue characteristics from the pre- and post-treatment CT of patients who have undergone

neoadjuvant chemotherapy for bladder cancer has the potential to assist in assessment of treatment response. (QIN Publications and Presentations: #4, #8, and #9)

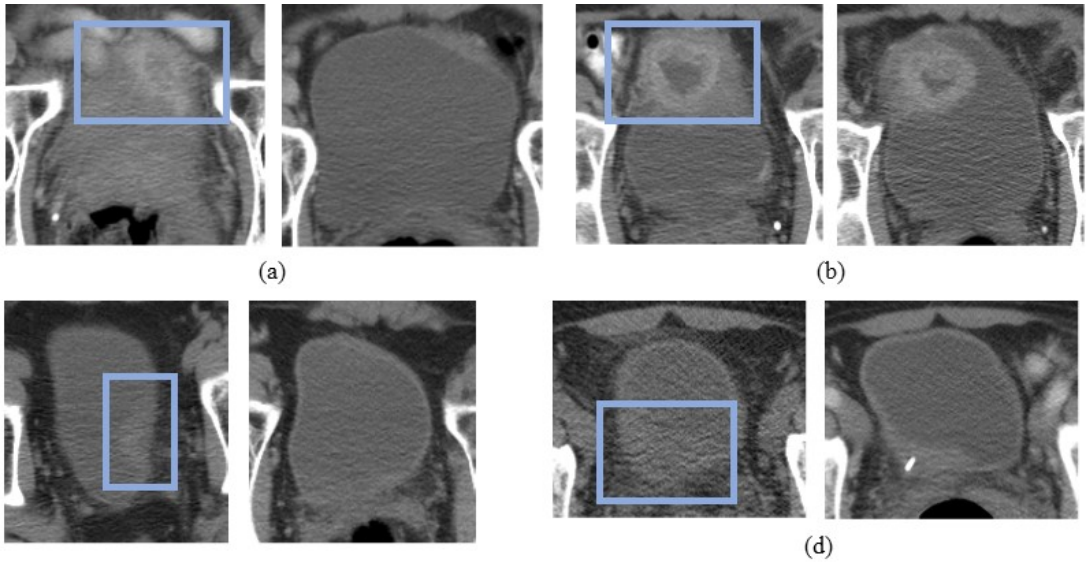


Figure 15: Examples of pre- and post-treatment bladders and their predictions. (a) The computer methods and the radiologists correctly predicted the treatment outcome for this case, which was a non-responding, progressive disease that went from stage T2 before treatment to T3a after treatment. (b) In this stable disease case (stage T3), the computer methods and the radiologists correctly identified the case as non-responding. (c) This case fully responded, going from stage T2 to T0, and the computer methods and the radiologists correctly predicted the treatment response. (d) A full-responding case, going from stage T3 to T0. The computers correctly predicted the response, while the radiologists did not. The region around the right ureterovesicular junction was asymmetrically thickened, which might have misled the radiologist to assess that cancer was present. The pre-treatment scan is on the left and the post-treatment scan is located on the right of each pair. The box on the pre-treatment scan represents the location of the lesion as marked by one of the radiologists.

COLLABORATIONS WITHIN THE NETWORK

We are actively involved in the collaboration activities within the QIN.

§ QIN committees and working groups

We participate in the QIN committees (the Executive Committee and the Coordinating Committee) and in the QIN working groups (PET-CT working subgroup, Image Analysis Performance Metrics working group, Bioinformatics/IT & Data Sharing working group, and Clinical trial Design & Development working group). Dr. Hadjiyski serves as a chair of PET-CT working subgroup.

§ QIN Grand Challenges

We also participate in two grand challenges organized within the PET-CT working subgroup: (1) Use of NLST as a dataset for assessing lung nodule interval change, and (2) CT Feature Comparison Study. We are very enthusiastic about this QIN opportunity, because this allows to test our tools on a different modality and different type of lesions (lung nodules) as well as to compare the tools to the systems of the other QIN participants in the challenges. We actively participate in the data analysis and the publications preparation related to the challenges, which resulted in a joint publication in the QIN Special issue of Tomography. (QIN Publications and Presentations: #11)

§ Computer demonstrations at the QIN face to face meeting

We also have participated in the live computer demonstrations at the Face to Face meeting in April 2016 and have demonstrated our GUI and QIBC segmentation tool. We are also very enthusiastic about this QIN opportunity, because (1) it was possible to present our tool to the other members of QIN, (2) we got very useful feedback from the experts in the field and (3) it allowed discussions for potential collaboration for integration of our GUI in Slicer.

PLANS FOR NEXT YEAR

In the next year we will continue to collect CTU pre- and post- neoadjuvant chemotherapy treatment cases. We also will continue the prospective collection of pathological information, diagnostic test results, immunohistochemical biomarkers, and CT scans from bladder cancer patients after the first cycle of chemotherapy. Our clinical collaborators will continue to annotate and outline the bladder lesions. We will concentrate our efforts to continue the development of our segmentation bladder lesion system (QIBC) and the decision support systems for bladder cancer staging (CDSS-S) and treatment response monitoring (CDSS-T) with a larger data set. We also will continue to extract additional 3D morphological and texture radiomic descriptors, define new descriptors, and use machine learning methods for the design of the predictive model to predict the cancer stage and to combine the descriptors in a “combined response index” as a predictor of the treatment response.

LIST OF REFERENCES

1. American Cancer Society, www.cancer.org 2017, "Cancer Facts & Figures 2017" (2017).
2. C.N. Sternberg, "The treatment of advanced bladder cancer," Annals of Oncology, 6, 113-126 (1995).
3. S.L. Fagg, P. Dawsonedwards, M.A. Hughes, T.N. Latief, E.B. Rolfe, and J.W.L. Fielding, "CIS-Diamminedichloroplatinum (DDP) as initial treatment of invasive bladder cancer," British Journal of Urology, 56, 296-300 (1984).

4. D. Raghavan, B. Pearson, G. Coorey, W. Woods, D. Arnold, J. Smith, J. Donovan, and P. Langdon, "Intravenous CIS-platinum for invasive bladder cancer – safety and feasibility of a new approach," Medical Journal of Australia, 140, 276-278 (1984).

5. J.J. Meeks, J. Bellmunt, B.H. Bochner, N.W. Clarke, S. Daneshmand, M.D. Galsky, N.M. Hahn, S.P. Lerner, M. Mason, T. Powles, C.N. Sternberg, and G. Sonpavde, "A Systematic Review of Neoadjuvant and Adjuvant Chemotherapy for Muscle-invasive Bladder Cancer," European Urology, 62, 523-533 (2012).

6. H. Abol-Enein, A.V. Bono, M. Boyer, N.W. Clarke, C.M.L. Coppin, E. Cortesi, P.J. Goebell, S. Groshen, R.R. Hall, A. Horwich, P.U. Malmstrom, J.A. Martinez-Pineiro, M.K.B. Parmar, D. Raghavan, J.T.G. Roberts, L. Sengelov, A. Sherif, L.A. Stewart, M. Stockle, R. Sylvester, J.F. Tierney, F.M. Torti, C.L. Vale, D.M.A. Wallace, and A.B.C.M.-A. Collaboration, "Neoadjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis," Lancet, 361, 1927-1934 (2003).

7. H.B. Grossman, R.B. Natale, C.M. Tangen, V.O. Speights, N.J. Vogelzang, D.L. Trump, R.W.D. White, M.F. Sarosdy, D.P. Wood, D. Raghavan, and E.D. Crawford, "Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer," New England Journal of Medicine, 349, 859-866 (2003).

8. J.A. Witjes, M. Wullink, G.O.N. Oosterhof, and P. deMulder, "Toxicity and results of MVAC (methotrexate, vinblastine, adriamycin and cisplatin) chemotherapy in advanced urothelial carcinoma," European Urology, 31, 414-419 (1997).

9. E.M. Bessell, H.M. Price, and P.J. McMillan, "The measurement of the regression of carcinoma of the bladder using ultrasonography and CT scanning during and after radical radiotherapy," Radiotherapy and Oncology, 19, 145-157 (1990).

10. A. Krizhevsky, "cuda-convnet," (2012).

11. A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in NIPS 2012: Neural Information Processing Systems (Lake Tahoe, Nevada 2012).

12. K. H. Cha, L. Hadjiiski, R. K. Samala, H.-P. Chan, E. M. Caoili, and R. H. Cohan, "Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets," Medical Physics 43, 1882-1896 (2016).

13. E. Street, L. Hadjiiski, B. Sahiner, S. Gujar, M. Ibrahim, S.K. Mukherji, H.P. Chan, "Automated Volume Analysis of Head and Neck Lesions on CT Scans Using 3D Level Set Segmentation," Medical Physics **34**, 4399-4408 (2007).

14. P. Jaccard, "The distribution of the flora in the alpine zone," New phytologist 11, 37-50 (1912).

15. B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, L.M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," Medical Physics **28**, 1455-1465 (2001).

16. B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie, M.M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," Medical Physics **25**, 516-526 (1998).

17. T.W. Way, L.M. Hadjiiski, B. Sahiner, H.-P. Chan, P.N. Cascade, E.A. Kazerooni, N. Bogot, C. Zhou, "Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours," Medical Physics **33**, 2323-2337 (2006).

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

1. Hadjiiski L, Weizer AZ, Alva A, Caoili EM, Cohan RH, Cha K, Chan HP, "Bladder cancers on CT: preliminary study of treatment response assessment based on computerized volume analysis, WHO and RECIST Criteria." *American Journal of Roentgenology* 2015; 205(2): 348-352.

2. Cha KH, Hadjiiski L, Samala RK, Chan H-P, Caoili EM, Cohan RH, "Urinary Bladder Segmentation in CT Urography using Deep-Learning Convolutional Neural Network and Level Sets." *Medical Physics* 2016; 43: 1882-96.

3. Cha KH, Hadjiiski L, Samala RK, Chan H-P, Cohan RH, Caoili EM, Paramagul C, Alva A, Weizer AZ, "Bladder Cancer Segmentation in CT for Treatment Response Assessment: Application of Deep-Learning Convolution Neural Network – A Pilot Study." *Tomography* 2016; 2: 421-8.

4. Cha KH, Hadjiiski L, Chan H-P, Weizer AZ, Alva A, Cohan RH, Caoili EM, Paramagul C, Samala RK, "Bladder Cancer Treatment Response Assessment in CT using Radiomics with Deep-Learning." *Scientific Reports* 2017 (under revision)

5. Hadjiiski L, Chan HP, Cha KH, Cohan RH, Caoili EM, Weizer AZ, Alva A, Wei J, Zhou C., "Analysis of Treatment Response of Bladder Cancers on CT Scans: Comparison of Assessments by Radiomic Features, WHO and RECIST Criteria." Poster presentation at the *101st Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA),* Chicago, IL. Nov 29-Dec 4, 2015.

6. Cha KH, Hadjiiski L, Samala RK, Chan H-P, Cohan RH, Caoili EM, "Comparison of bladder segmentation using deep-learning convolutional neural network with and without level sets." *Progress in Biomedical Optics and Imaging – Proceedings of SPIE* 2016; 9785

7. Garapati, S, Hadjiiski L, Cha KH, Chan H-P, Caoili E M, Cohan RH, Weizer A, Alva A, Paramagul C, Wei J, Zhou C, "Automatic staging of bladder cancer on CT urography." *Progress in Biomedical Optics and Imaging – Proceedings of SPIE* 2016; 9785

8. Cha KH, Hadjiiski L, Chan H-P, Samala RM, Cohan RH, Caoili E M, Weizer AZ, Alva A, "Deep-Learning Bladder Cancer Treatment Response Assessment in CT Urography." Oral presentation at the *102nd Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA),* Chicago, IL. Nov 27-Dec 2, 2016.

9. Cha KH, Hadjiiski L, Chan H-P, Samala RK, Cohan RH, Caoili EM, Paramagul C, Alva

A, Weizer AZ, “Bladder Cancer Treatment Response Assessment using Deep Learning in CT with Transfer Learning.” Oral presentation at the SPIE Medical Imaging, Orlando, FL, Feb 11-Feb 16, 2017.

10. Gordon M, Hadjiiski L, Cha KH, Chan H-P, Samala RK, Cohan RH, Caoili EM “Segmentation of Inner and Outer Bladder Wall using Deep-Learning Convolutional Neural Network in CT Urography.” Oral presentation at the SPIE Medical Imaging, Orlando, FL, Feb 11-Feb 16, 2017.

11. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren J C, Uthoff J, Dilger S K N, Driscoll B, Yeung I, Hadjiiski L, Cha K, Balagurunathan Y, Gillies R, Goldgof D 2016 Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features Tomography 2 430-7.

U01 CA181156: Quantitative CT Imaging for Response Assessment when Using Dose Reduction Methods

David Geffen School of Medicine at UCLA

Michael McNitt-Gray, PhD
Matthew Brown, PhD, Jonathan Goldin, MD, PhD, Grace Kim, PhD.

INTRODUCTION

Despite concerns over radiation dose, CT continues to be widely used for assessing response to therapy in many clinical trials settings. There have been significant developments which allow the reduction of radiation dose from CT, including advances in iterative reconstruction techniques, detector technologies and others that promise significant dose reductions (50-60%) to patients, while maintaining clinical image quality. While these technologies should be investigated wherever possible in a clinical environment, their effects on quantitative measures extracted from CT images are unclear and need to be investigated before they are deployed in clinical trials. Simply reducing tube current time product (mAs) will increase image noise, which may increase variability in quantitative measures. Size measures may be affected differently depending on the anatomic region; lung lesions (typically high contrast objects) may be affected differently from liver lesions (typically lower contrast). Peak values measured when contrast enhanced studies are used may also respond to dose reductions differently. In addition, because new iterative reconstruction methods reduce noise, they often also smooth the image somewhat, which may affect size and density (e.g. average HU) measures. Therefore, this application proposes to systematically investigate the effects of radiation dose reduction methods on quantitative metrics used in clinical trials. The goal is to determine how far we can decrease dose under different conditions before we increase variance to unacceptable levels in the context of using quantitative measures to assess response to therapy.

We have proposed two specific aims to carry out this research. In the first aim, we proposed to create a collection of cases that represent a range of low dose acquisition and reconstruction scenarios in specific quantitative imaging tasks. This is being accomplished using a calibrated dose reduction simulation method (noise insertion tool) and then reconstructing images under a variety of dose reduction levels and reconstruction methods. In the second specific aim, we are extracting quantitative Imaging measures from these reconstructed image data sets and analyzing the variance of quantitative measures across dose levels and reconstruction methods. The overall goal is to provide guidance to the QIN, and clinical trials in general, regarding the use of both standardized protocols and the use of dose reduction methods, with the ultimate goal of determining the levels of dose reduction that yield acceptable levels of measurement variance in several assessment tasks/environments.

DISCUSSION OF PROGRESS

Since the beginning of the project period, we have made progress on a number of projects. These are reported below.

§ The Effects of Radiation Dose and Reconstruction Method on Tumor Volumetrics

A study into the effects of radiation dose level and reconstruction method on estimating the volume of lesions observed on CT was published in Medical Physics (Young et al, Medical Physics, May 2015). In this study, we analyzed the effects of radiation dose level and reconstruction method on measured lesion volumes of lung lesions in cancer patients. We used the original dose level (approximately 20 mGy) and then simulated reduced dose levels of 25% 10%, and 3% of the dose of our clinical protocol. Simulated reduced-dose data were reconstructed with both conventional filtered backprojection (B45 kernel) and iterative-reconstruction methods (SAFIRE: I44 strength 3 and I50 strength 3). Three lab technologist readers contoured "measurable" nodules in 33 patients under each of the different acquisition/reconstruction conditions in a blinded study design. Of the 33 measurable nodules, 17 were used to estimate repeatability with our clinical reference protocol, as well as interdose and inter-reconstruction-method reproducibility.

The clinical-dose repeatability experiment yielded a mean proportional difference of 1.1% and SD of 5.5%. The inter-dose reproducibility experiments gave mean differences ranging from -5.6% to -1.7% and SDs ranging from 6.3% to 9.9%. The inter-reconstruction-method reproducibility experiments gave mean differences of 2.0% (I44 strength 3) and -0.3% (I50 strength 3), and SDs were identical at 7.3%. For the subset of repeatability cases, inter-reconstruction-method mean/SD pairs were (1.4%, 6.3%) and (-0.7%, 7.2%) for I44 strength 3 and I50 strength 3, respectively. Analysis of representative nodules confirmed that reader variability appeared unaffected by dose or reconstruction method.

Lung-nodule volumetry was shown to be extremely robust to the radiation-dose level, down to the minimum scanner supported dose settings. In addition, volumetry was robust to the reconstruction methods used in this study, which included both conventional filtered back projection and iterative methods.

§ The Effects of Radiation Dose and Reconstruction Methods on Lung Lesion Density and Texture Based Features

Following the above effort, we investigated the effects of radiation dose level and reconstruction method on other features of interest to the Quantitative Imaging community, namely those based on density and texture (local variations). So using the lesions identified and analyzed in the Young study described above (and previously contoured as well), we extended the analysis to features extracted from the nodule contours. Our study had two major components. In the first component, a uniform water phantom was scanned at 3 dose levels and images were reconstructed using both conventional filtered back-projection (FBP) and iterative reconstruction (IR) methods with four kernels for each method for a total of 24 different combinations of acquisition and reconstruction conditions (4 FBP reconstructions

and 4 IR reconstructions at each of 3 dose levels). Example water phantom images are shown in Figure 1. In the second component, raw projection data (sinogram data) was obtained for 33 lung nodules from patients scanned as part of our clinical practice. For the nodule cases, low dose acquisitions were simulated by adding noise to sinograms acquired at clinical dose level and then reconstructed using one FPB kernel and 2 IR kernels for a total of 12 conditions (4 dose levels and 3 reconstructions at each dose level). Examples of these are shown in Figure 2.

For the water phantom, spherical regions of interest (ROI) were created at multiple locations within the water phantom on one reference image obtained at a reference condition. For the lung nodule cases, the ROI of each nodule (represented as a three dimensional boundary) was obtained using semi-automated contouring methods with manual editing allowed from images obtained at a reference condition. All ROIs were then applied to their corresponding images constructed at different conditions. For 17 of the nodule cases, repeat contours were performed to assess repeatability. For all ROIs, both histogram (8 features) and gray level co-occurrence matrix based texture features (34 features) were computed. For the lung nodule cases, the reference condition was selected to be 100% of clinical dose with FBP reconstruction using the B45f kernel; feature values calculated under all other acquisition/reconstruction conditions were compared to this reference condition. In order to measure the stability of features across different combinations of acquisition and reconstruction parameters, a *Q* measure was introduced, which is defined as the ratio of reproducibility (across acquisition/reconstruction conditions) to repeatability (across repeat contours) of each feature.

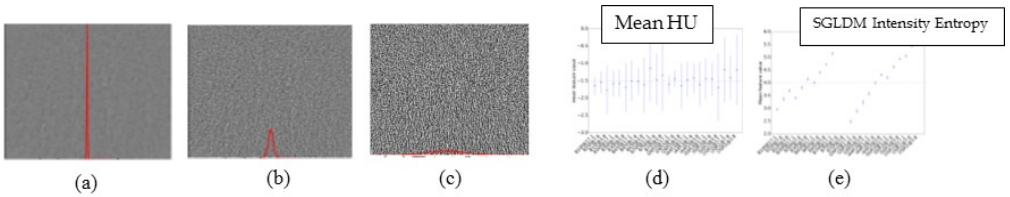


Figure 1: (Images from Figures 1 and 2 of Lo et al, Med. Phys. 2016). CT images of a **water phantom** illustrating differences in appearance and HU value distribution across different dose levels and reconstructions for: (a) Original clinical dose with iterative recon (I26 Str 5); (b) Reference condition – Original dose, FBP recon (B45 kernel) and (c) Simulated reduced dose with sharp FBP recon (B70). The next plots show feature valuess from these water phantom ROIs across different dose/reconstruction conditions, where the points and whiskers indicate the mean and the standard deviation of the feature value, respectively. The y-axes are the mean feature value at each condition and the x-axes are the various dose/reconstruction conditions. These are shown for: (d) **mean intensity value** and (e) **Spatial Gray Level Dependence Matrix texture value Intensity Entropy**. These plots show that the mean intensity value is stable across conditions, while this texture value varies substantially across conditions.

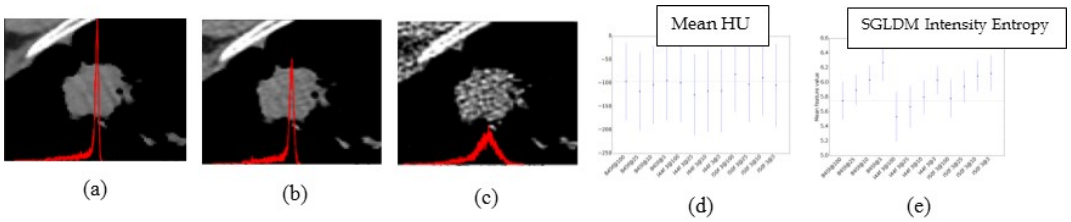


Figure 2 (Images from Figures 4 and 5 of Lo et al, Med. Phys 2016) describing nodules used in [26] and specifically describing CT images of a **lung nodule** illustrating differences in appearance and HU value distribution across different dose levels and reconstructions for: (a) Original clinical dose with iterative recon (I44 Str 3); (b) Reference condition – Original dose, FBP recon (B45 kernel) and (c) Simulated reduced dose FBP recon (B45). The overlaid red lines are the histogram of HU values within the nodule. Similar to **Figure 1** above, the next plots show feature valuess from these lung nodules across different dose/reconstruction conditions. These are shown for: (d) **mean intensity value** and (e) **SGLDM texture value Intensity Entropy**. These plots also show that the mean value is stable across conditions, while this texture value again varies substantially across conditions.

The water phantom results demonstrated substantial variability among feature values calculated across acquisition and reconstruction conditions, with the exception of the mean value of the density (mean HU of the region) which was robust across all conditions. Features calculated from lung nodules demonstrated similar results with histogram mean as the most robust feature ($Q <= 1$), having a mean and standard deviation Q of 0.37 and 0.22 respectively. Surprisingly, the other two histogram features that are also quite robust were standard deviation and variance. Some of the GLCM measures were also quite robust across different conditions, namely diff. variance, 35 sum variance, sum average, variance and mean. As expected, the histogram mean is the most robust feature in our study. The effects of acquisition and reconstruction conditions on GLCM texture features vary widely, though there was a trend toward features calculated based on the sum of the product of intensities and probability being more robust in general, with a few exceptions.

The conclusion of this work was that care should be taken to account for variation in density and texture features if a variety of dose and reconstruction conditions are used for the quantification of lung nodules in CT, otherwise a change in quantification results may be more reflective of acquisition and reconstruction conditions than the nodule itself. Preliminary results of this work were presented at AAPM in July 2015 and a peer-reviewed manuscript was published this past year (Lo et al, Medical Physics, 2016).

§ Extensions to Previous Software and Data Collection Efforts

During the current project period, we have: (a) extended the capabilities of our software that reads sinogram data from Siemens Scanners to read several different formats (.IMA, .CTD and .PTR) as well as reading the files from the newest CT scanner from Siemens – the Dual Source Definition Force (adding to our capabilities to read data from the Sensation 64, Definition AS); (b) extended the software that adds noise to sinogram data and simulates

specific amounts of radiation dose reduction. The new capabilities reflect an improved capability to both characterize and model the effects of reduced electronic noise in newer scanners with advanced detector technologies. This will be critical as we move to lower and lower doses in more modern scanners where the electronic noise might provide a substantial limit to dose reduction; (c) Extended our collection of anonymized image data. These datasets represent a wide range of reduced dose acquisition levels as well as reconstruction methods (both conventional filtered back projection –FBP – as well as scanner provided iterative reconstruction methods (Siemens Safire or Admire).

Our current inventory of raw data (sinogram) from different clinical protocols includes:

Protocol Name	Total # of Cases
NLST Low Dose Lung Cancer Screening	**481**
Low Dose Lung Cancer Screening (Current)	**(to date) 583**
Routine Chest (Diagnostic) Cases	**99**
Renal Cell Carcinoma (3 phase)	**(to date) 27**

Table 1: Description of raw data inventory for different types of scans. Note: "NLST" cases were acquired with fixed tube current protocols (see Cagnon et al, Academic Radiology, 2006). "Current" cases were acquired with tube current modulation (CareDose4D, Siemens Healthcare).

§ Lung Nodule Detection for Reduced Dose CT Scanning

For the low dose Lung Cancer Screening Cases obtained during the NLST, we have: (a) an inventory of 481 cases; (b) using our noise addition software, we have simulated reduced dose acquisitions for all cases at both 50% of the original dose (~1 mGy) and 25% of the original dose (~0.5 mGy); (c) and all 3 dose levels (original and 2 reduced dose levels) have been reconstructed using conventional Filtered Back Projection (FBP) on the Siemens scanner. Of these 481 cases, 82 had at least one nodule (prevalence of 17%) and 399 did not (83%). A total of 118 nodules were identified: 27 nodules (23%) corresponded to LungRADS category 4 based on size and composition, while 18 (15%) corresponded to LungRADS category 3 and 73 (61%) corresponded to LungRADS category 2. The lungs were segmented semi-automatically, and all images and segmentations were input to an in-house CAD algorithm trained on higher-dose scans (75-300 mAs). CAD findings were compared to a reference standard generated by an experienced reader. Nodule- and patient-level sensitivities

were calculated along with false positives per scan, all of which were evaluated in terms of the relative change with respect to dose. Nodules were subdivided based on size and solidity into categories analogous to the LungRADS assessment categories, and sub-analyses were performed.

For solid nodules ≥ 8 mm, patient-level median sensitivities were 100% at all three dose levels, and mean sensitivities were 72%, 63%, and 63% at original, 50%, and 25% dose respectively. Overall mean patient-level sensitivities were 38%, 37%, and 38% at original, 50%, and 25% dose. These low sensitivities were primarily due to the prevalence of smaller nodules and non-solid nodules in our reference standard. The mean false-positive rates were 3, 5, and 13/case.

This work showed that CAD sensitivity decreased very slightly for larger nodules as dose was reduced, indicating that reducing the dose to 50% of original levels may be investigated further for use in CT screening. However, the effect of dose was small relative to the effect of the nodule size and solidity characteristics. The number of false positives per scan increased substantially at 25% dose, illustrating the importance of tuning CAD algorithms to very challenging, high-noise screening exams. This work has been presented at both AAPM and RSNA conferences and was just accepted for publication and is in press at Medical Physics with an expected publication date of Feb. 2017.

§ Open Source Image Reconstruction Software (wFBP) and Creation of an Image Acquisition/Reconstruction Pipeline

One of the issues that we have been running into in our research is just the size and scale of the problems we are trying to address in terms of the numbers of cases (quite large for the NLST cases) and the number of dose levels and reconstruction kernel settings we wish to analyze. Though we have been successful in reconstructing raw projection data at the scanner on which the data was originally acquired, clinical CT scanners are not designed for: (a) high throughput of raw projection data files and subsequent reconstructions; (b) multiple versions of the same patient dataset (at different reduced dose levels); (c) batch mode processing for a large variety of reconstruction conditions such as different reconstruction kernels and slice thicknesses. So, we developed an open source implementation of a commonly used reconstruction method referred to as weighted Filtered Backprojection (wFPB). Our implementation was based on the original article published by Stierstorfer et al (Physics and Biology, 2004). Our project successfully implemented the wFBP algorithm on a medium cost GPU and showed excellent image quality and computational performance. This approach will help us overcome many of the limitations described above in that this can be done in batch mode with multiple prospective reconstructions performed over a relatively short period of time. This work is critical in the development of our image acquisition/reconstruction pipeline (described below). This work was published in Medical Physics as a technical note and available online at
http://scitation.aip.org/content/aapm/journal/medphys/43/3/10.1118/1.4941953

With the availability of an offline, GPU implementation of image reconstruction, we were able to create a pipeline that is capable of high throughput processes for our research.

Specifically, the raw data reading modules, the noise addition (simulating reduced dose acquisitions) and wFBP reconstruction engine have been organized into a fully automated pipeline (Figure 3) that take the raw data files and creates the desired set of image datasets that represent a range of dose levels, slice thicknesses and reconstruction kernels for wFBP. These datasets will be used (see future plans) in large scale investigations into the effects of acquisition and reconstruction parameters on quantitative imaging tasks.

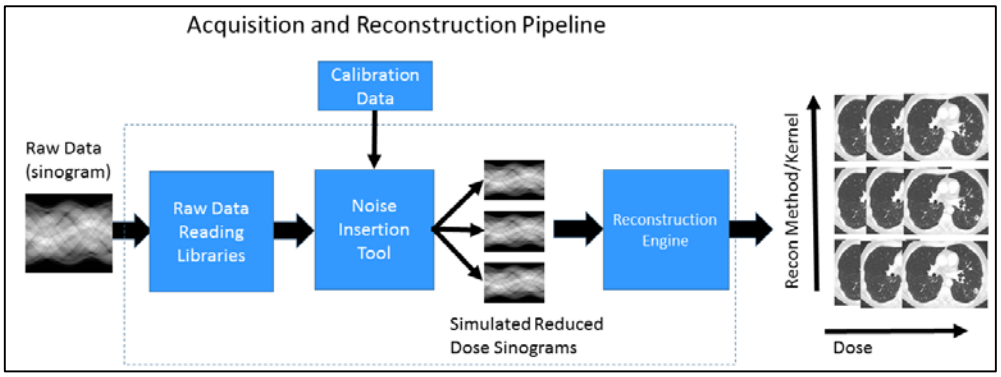


Figure 3: Diagram illustrating the components and data flow for the image acquisition/reconstruction pipeline, which takes raw projection (sinogram) data and creates image datasets that represent a range of image acquisition and reconstruction parameter settings.

The operation and control of the Pipeline was designed to be fully automatic and provide a high-throughput system for the creation of a large number of image datasets representing a wide range of acquisition and reconstruction conditions. To accomplish this, the initial system uses an HT condor computation environment which allows the queuing of jobs (using HTCondor queuing) with python control script to initiate each job and execute each step. We are currently developing the ability to execute all steps in one system (linux based machine with GPU capabilities). Initial performance benchmarks indicate that on a system with 4 GPUs (e.g. a "Deep Learning" system from NVIDIA), a performance of 1.25 minutes per case/condition. Table 2 illustrates the expected benefits from the pipeline implemented on a 4 GPU system using an example comparing our previous experience with 481 NLST cases reconstructed at 3 dose levels (1 thickness, 1 kernel) (Young et al, Med. Phys. 2017) to the performance we expect to get with the described pipeline with 500+ UCLA lung cancer screening cases we have collected to create datasets that represent 3 dose levels (original plus two simulated reduced doses), 3 slice thicknesses (0.6, 1 and 2mm) and 3 reconstruction kernels (smooth, standard and sharp). Thus, the high throughput, batch mode processing used here will allow us more than 2 orders of magnitude increase in throughput, which provide a much broader exploration of the acquisition and reconstruction parameter space than is currently achievable.

Cases	Dose Levels	Thicknesses	Kernels	# Datasets	Time to Create
481 NLST (scanner recon)	3	1	1	1,443	~ 6 months
500 UCLA screening cases (Acq/Recon Pipeline)	3	3	3	13,500	11.25 days

Table 2. Illustration of increased throughput from Acquisition\Recon Pipeline described in Figure 3 when implemented on a 4 GPU system compared to throughput when using current conventional approach**.**

§ Automating Phantom Assessment for Clinical Trials Using Quantitative Imaging

One of this issues for clinical trials that seek to use quantitative imaging methods is to assure that the acquisition and reconstruction parameters specified are indeed being used at each participating site. Preferably this should be done prospectively to avoid having to exclude a case because of technical differences. In addition, trials may also desire to have some assurance that the CT scanner is performing well (e.g. is calibrated) prior to scanning subjects. In some trials, the use of a phantom (test object) is used to evaluate both of these objects (protocol adherence and system performance under the desired protocol). However, the use of phantoms can be a burden as each site has to scan the phantom according to the protocol and then someone (e.g. a central site) has to read and evaluate the phantom scan. Through our extensive experience with clinical trials in CT, we have developed standardized processes for evaluation phantoms scanned on CT scanners at participating sites.

Recently, we have developed methods to perform these assessments automatically. While the assessment seems reasonably straightforward, there are several issues that needed to be addressed including: (a) the heterogeneity of available CT phantoms at sites (each manufacturer supplies a QC phantom, but they are quite different between manufacturers), so identification of the phantom is a first step; (b) phantom scanning may or may not include the entire phantom or just the water portion of the phantom, so identification of the water region needs to be done. (c) the manufacturers report some (but not all) technical parameters in the DICOM headers and there is heterogeneity in how these values are reported, especially in the context of modern scanners using Automatic Exposure Control (e.g. Tube Current Modulation) systems and iterative reconstruction methods.

This approach that we developed was based on several computer vision techniques as well as registration methods (e.g. automatically matching the submitted phantom to one of the known types of phantoms). Further analysis was needed to correctly identify the image on which the desired analysis was to be performed. Then the desired analyses (e.g. water calibration, scan field homogeneity) were designed to be performed automatically. Finally, an analysis of the DICOM headers was designed with manufacturer-specific analyses to account

for different reporting schemes and then compared to the scanner specific properties spelled out in the trial's protocol documents. We have presented this work at the AAPM conference in 2016 and are preparing a peer-reviewed publication (target submission date is 1st quarter 2017).

§ Participation in QIN PET-CT Group "Feature Challenge"

The UCLA QIN team participated in the CT Image Feature Challenge (coordinated by Moffit QIN). We submitted a limited set of feature data (15 features, one from several different categories) to participate in this challenge. The purpose of this study was to investigate the sensitivity of quantitative descriptors of pulmonary nodules to segmentations and to illustrate comparisons across different feature types and features computed by different implementations of feature extraction algorithms. The concordance correlation coefficients of the features were calculated as a measure of their stability with the underlying segmentation.

This study showed that 68% of the 830 features in this study had a concordance CC of 0.75. Pairwise correlation coefficients between pairs of features were used to uncover associations between features, particularly as measured by different participants. A graphical model approach was used to enumerate the number of uncorrelated feature groups at given thresholds of correlation. At a threshold of 0.75 and 0.95, there were 75 and 246 subgroups, respectively, providing a measure for the features' redundancy. This work resulted in a peer-reviewed publication in the special issue of the journal Tomography (DOI:10.18383/j.tom.2016.00235 .

PLANS FOR NEXT YEAR

During the next year we will extend our work in several different ways as described below.

§ Construction of a Pipeline for Open Source Image Reconstruction Software (wFBP)

The first activity will be to extend the pipeline described above (illustrated in Figure 3) by bringing it together with two other components to create a tightly integrated, high throughput system (illustrated in Figure 4). The image acquisition/reconstruction pipeline will create inputs to the segmentation/CAD/Quantitative Imaging feature pipeline that will identify anatomic and pathologic structures and extract features of interest. A performance evaluation pipeline will compare the extracted results (e.g. detections or feature values) to reference condition results (e.g. radiologist markings, feature values obtained under a reference condition) and evaluate performance metrics across acquisition and reconstruction conditions (e.g. changes in feature values across dose levels and/or reconstruction conditions). This is all being designed for high throughput performance for large numbers of cases and with a wide variety of acquisition and reconstruction parameters.

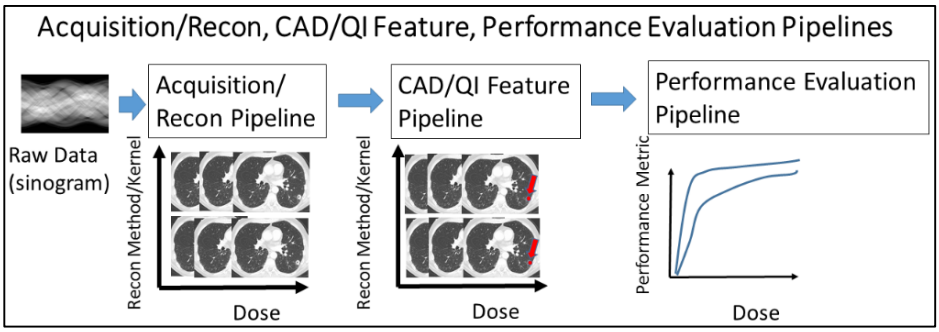


Figure 4. Diagram illustrating the relationship and workflow between the different components of the overall pipeline. This will allow evaluation of large numbers of cases with a range of acquisition and reconstruction parameter settings.

§ Extension of Image Reconstruction Capabilities

We also plan to extend our image reconstruction capabilities beyond wFBP to an iterative reconstruction algorithm while allowing the reconstructions to be performed in batch mode using a computationally efficient approach. This will allow us to create a wide range of simulated reduced dose scans reconstructed under a wide variety of approaches (conventional wFPB and iterative), which will allow us to assess the robustness of quantitative features being extracted from image data.

§ Perform Analysis on 3-Phase Kidney CT scans to evaluate Renal Cell Carcinoma (RCC)

We have been collecting raw projection data from patients undergoing our RCC protocol and we have been reviewing medical records to establish diagnoses on these scans. In coming year we plan to evaluate the effects of dose reduction and reconstruction method on the ability to distinguish cell types in RCC using methods already published by our investigators This will further extend our work to go beyond just volume and texture to quantitative features that are derived from contrast enhancement (functional features). We hope to be able to publish our results on iterative reconstruction methods, automated phantom QA as well as contrast uptake information in RCC during the coming reporting period.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

§ Published in Peer-Reviewed Journals

1. J. Kalpathy-Cramer, B. Zhao, L. Lu, D. Goldgof, D. Cherezov, S. Napel, S. Echegaray, M. McNitt-Gray, J.C. Sieren, J. Uthoff, B. Driscoll, I. Yeung, L. Hadjiiski, Y. Balagurunathan, R. Gillies, D. Goldgof. Radiomics in of lung nodules: a multi-institutional study of robustness and agreement of imaging features. TOMOGRAPHY, December 2016, Volume 2, Issue 4: 430-437 DOI:10.18383/j.tom.2016.00235
2. P. Lo, S. Young, H. J. Kim, M. S. Brown, & M. F. McNitt-Gray, "Variability in CT lung-nodule quantification : Effects of dose reduction and reconstruction methods on density and texture based features", Medical Physics, 4854(43). http://doi.org/10.1118/1.4954845
   (NOTE: Selected to be highlighted under the Editors' Choice column for the Medical Physics Scitation and medphys.org websites for the August issue.
3. J. Hoffman, S. Young, F. Noo, & M. F. McNitt-Gray, "Technical Note: FreeCT_wFBP: A robust, efficient, open-source implementation of weighted filtered backprojection for helical, fan-beam CT", Medical Physics, 43(3), 1411. http://doi.org/10.1118/1.4941953.
4. S. Young, H. J. G. Kim, M. M. Ko, W. W. Ko, C. Flores, M. F. McNitt-Gray, "Variability in CT lung-nodule volumetry: Effects of dose reduction and reconstruction methods", Medical Physics, 42(5), 2679. doi:10.1118/1.4918919.
5. D. Chong, H. J. Kim, P. Lo, S. Young, M. F. McNitt-Gray, F. Abtin, J. G. Goldin, M. S. Brown, "Robustness-driven feature selection in classification of fibrotic interstitial lung disease patterns in computed tomography using 3D texture features", IEEE Transactions on Medical Imaging, 35(1), 144–57. http://doi.org/10.1109/TMI.2015.2459064.

§ In press

1. S. Young, P. Lo, H. J. Kim, M. S. Brown, J. Hoffman, W. Hsu, W. Wahi-Anwar, C. Flores, G. Lee, F. Noo, J. Goldin, and M. F. McNitt-Gray, "The Effect of Radiation Dose Reduction on Computer-Aided Detection (CAD) Performance in a Low-Dose Lung Cancer Screening Population", in press, Medical Physics; to be published Feb. 2017.

§ In preparation (submission 1st or 2nd quarter 2017)

1. Wahi-Anwar M, Lo P , Kim HG , Brown MS, Goldin, JG, McNitt-Gray MF. A Fully Automated CT Tool to Facilitate Phantom Image QA for Quantitative Imaging in Clinical Trials. To be submitted to Journal of Digital Imaging, 1st quarter 2017.
2. Emaminejad N, Lo P, Ghahremani S, Kim HG, Brown MS, McNitt-Gray MF, The effects of slice thickness and radiation dose level variations on computer-aided diagnosis (CAD) nodule detection performance in pediatric chest CT scans. To be submitted to Journal of Medical Imaging, 1st quarter 2017.

3. Wahi-Anwar M, Young S, Lo P, Coy H, Ashen-Garry D, Pace-Soler E, Raman S, Kim H, Brown MS, McNitt-Gray MF. Effects of Radiation Dose Reduction On Renal Cell Carcinoma Discrimination Using Multi-Phasic CT Imaging. To be submitted to Medical Physics, 1st quarter 2017.

§ Conference Abstracts, Presentations and Posters

1. Emaminejad N, Lo P, Ghahremani S, Kim HG, Brown MS, McNitt-Gray MF, The effects of slice thickness and radiation dose level variations on computer-aided diagnosis (CAD) nodule detection performance in pediatric chest CT scans. Paper 10134-10; oral presentation at . SPIE Medical Imaging Conference (CAD Conference). Orlando, FL, February 13, 2017. http://spie.org/MI/conferencedetails/computer-aided-diagnosis
2. Young,S, Lo,P, Hoffman,J, Kim,H, Hsu,W, Flores,C, Lee,G, Brown,M, McNitt-Gray,M, CAD Performance on a Large Cohort of National Lung Screening Trial Patients at Screening and Sub-screening Doses. Radiological Society of North America 2016 Scientific Assembly and Annual Meeting, November 27 - December 2, 2016, Chicago IL. archive.rsna.org/2016/16016041.html
3. S. Young, P. Lo, J. Hoffman, W. Wahi-Anwar, F. Noo, M. Brown, M. McNitt-Gray, "A fully-automated pipeline for generating CT images across a range of doses and reconstruction methods", AAPM Annual Meeting; Washington, DC (Aug 4, 2016). http://www.aapm.org/meetings/2016AM/PRAbs.asp?mid=115&aid=34124
4. S. Young, P. Lo, J. Hoffman, G. Kim, W. Hsu, C. Flores, G. Lee, M. Brown, M. McNitt-Gray, "CT Lung Cancer Screening and the Effects of Further Dose Reduction On CAD Performance", AAPM Annual Meeting; Washington, DC (Aug 4, 2016). http://www.aapm.org/meetings/2016AM/PRAbs.asp?mid=115&aid=34420
5. M Wahi-Anwar, P Lo , H Kim , M Brown , M McNitt-Gray. A Fully Automated CT Tool to Facilitate Phantom Image QA for Quantitative Imaging in Clinical Trials. Snap Oral presentation at AAPM Annual Meeting Washington, DC (July 31, 2016) http://www.aapm.org/meetings/2016AM/PRAbs.asp?mid=115&aid=32708
6. M Wahi-Anwar, S Young , P Lo , S Raman , H Kim , M Brown , M McNitt-Gray , H Coy , D Ashen-Garry , E Pace-Soler. Effects of Radiation Dose Reduction On Renal Cell Carcinoma Discrimination Using Multi-Phasic CT Imaging. Poster presented at AAPM Annual meeting, Washington DC (Jul 31 2016). http://www.aapm.org/meetings/2016AM/PRAbs.asp?mid=115&aid=33811
7. S. Young, J. Hoffman, F. Noo, and M. F. McNitt-Gray, "Vendor-independent, model-based iterative reconstruction on a rotating grid with coordinate-descent optimization for CT imaging investigations", Poster presented at AAPM Annual Meeting; Washington, DC (July 31, 2016). http://www.aapm.org/meetings/2016AM/PRAbs.asp?mid=115&aid=33355
8. M Wahi-Anwar, S Young , P Lo , S Raman , H Kim , M Brown , M McNitt-Gray , H Coy , D Ashen-Garry , E Pace-Soler. Effects of Radiation Dose Reduction On Renal Cell Carcinoma Discrimination Using Multi-Phasic CT Imaging. Poster presentation at AAPM annual scientific meeting 2016. http://www.aapm.org/meetings/2016AM/PRAbs.asp?mid=115&aid=33811

9. J Hoffman, F Noo , S Young , M McNitt-Gray ,Tailoring TCM Schemes to a Task: Evaluating the Impact of Customized TCM Profiles On Detection of Lung Nodules in Simulated CT Lung Cancer Screening - Oral presentation at AAPM annual scientific meeting 2016. http://www.aapm.org/meetings/2016AM/PRAbs.asp?mid=115&aid=33818
10. Young S, Lo P, Hoffman JM, Kim GHJ, Brown MB, McNitt-Gray M. (2014). Effects of CT dose and nodule characteristics on CAD performance in a cohort of 90 National Lung Screening Trial patients. Proc. SPIE 9785, Medical Imaging 2016: Computer-Aided Diagnosis, 97850G; doi:10.1117/12.2217405
11. Hoffman JM; Noo F; McMillan K; Young S; McNitt-Gray MF. Assessing nodule detection on lung cancer screening CT: the effects of tube current modulation and model observer selection on detectability maps. Proc. SPIE 9787, Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment, 97870Q (24 March 2016); doi: 10.1117/12.2217348
12. P. Lo, S. Young, G. Kim, J. Hoffman, M. Brown, M. McNitt-Gray, "The Effects of CT Acquisition and Reconstruction Conditions On Computed Texture Feature Values of Lung Lesions", AAPM Annual Meeting; Anaheim, CA (July 14, 2015). http://www.aapm.org/meetings/2015AM/PRAbs.asp?mid=99&aid=29699
13. S. Young, P. Lo, G. Kim, J. Hoffman, M. Brown, M. McNitt-Gray, "A Research Pipeline to Simulate a Wide Range of CT Image Acquisition and Reconstruction Parameters and Assess the Performance of Quantitative Imaging and CAD Systems", AAPM Annual Meeting; Anaheim, CA (July 14, 2015). http://www.aapm.org/meetings/2015AM/PRAbs.asp?mid=99&aid=29110
14. S. Young, P. Lo, H. J. Kim, W. Hsu, J. Hoffman, M. S. Brown, M. F. McNitt-Gray, "The effects of Reduced-Dose Lung Cancer Screening CT on Lung Nodule Detection Using a CAD Algorithm", AAPM Annual Meeting; Anaheim, CA (July 14, 2015). http://www.aapm.org/meetings/2015AM/PRAbs.asp?mid=99&aid=29580
15. S. Young, J. Hoffman, F. Noo, and M. F. McNitt-Gray, "Development of Stand-Alone Filtered Backprojection and Iterative Reconstruction Methods Using the Raw CT Data Exported From Clinical Lung Screening Scans", Poster presented at AAPM Annual Meeting; Anaheim, CA (July 14, 2015). http://www.aapm.org/meetings/2015AM/PRAbs.asp?mid=99&aid=29425
16. S. Young, Michael F. McNitt-Gray, "Evaluating the consistency of lesion volumetry in chest CT after substantial dose reduction", AAPM Annual Meeting; Austin, TX (July 22, 2014). http://www.aapm.org/meetings/2014AM/PRAbs.asp?mid=90&aid=25133
17. Young S, McNitt-Gray M. Lung lesion volume measurements in simulated reduced-dose CT scans. Society of Thoracic Radiology annual meeting, 2014.
18. S. Young, MF McNitt-Gray, "Solid lung nodule volumetry: effects of dose reduction and reconstruction algorithms", 3rd International CT Meeting; Salt Lake City, UT (June 22-25, 2014).
19. Young S, McNitt-Gray M. (2014). Estimating lesion volume in low-dose chest CT: How low can we go? SPIE Medical Imaging, 9033, 1–13. doi:10.1117/12.2043730.

U01CA183848: Quantitative MRI Models of HN Cancers for Physiological Adaption of RT

University of Michigan (3)

Yue Cao, Ph.D. and Avraham Eisbruch, MD

INTRODUCTION

§ The Clinical Problem

Current state-of-art therapy of high-risk, advanced head-and-neck cancers (HNC) (e.g., HPV-), concurrent radiation therapy with chemotherapy and followed by adjuvant chemotherapy, still leads to 30-50% of local and regional failure. Physiological imaging based adaptive radiation boosting of the resistant subvolume of the tumor has the potential to improve outcomes. However, clinical utilization of metabolic and physiological imaging is challenging due to issues such as reproducibility of physiological images, tumor heterogeneity, and lack of tools to support therapy adaptation.

§ Quantitative Image Features

We have been developing and investigated quantitative image tools using pattern recognition techniques to identify the subvolumes of HNC with low blood volume (LBV) derived from DCE MRI and low ADC quantified from diffusion-weighted MRI (Figure 1). Currently, these tools are used to support a randomized phase II clinical trial for boosting the potential "risk for failure" subvolumes of the tumor in the advanced HNC. This trial involves two sites, University of Michigan Hospital and VA hospital at Ann Arbor. This phase II clinical trial allows us to test feasibility of using our QI tools in the clinical environment. Also, the clinical trial allows us to further identify issues and barriers that need to be overcome before deploying them in a multi-center clinical trial.

PROGRESS

§ Standardization of Delineation of the LBV Subvolumes in HNC

Evaluation of Basic Methodology on Different Scanners: The basics methodology was developed using the DCE scans acquired on one vendor scanner. After quantifying physiological parameters e.g., blood volume, from DCE data using the modified Toft model, we applied our methodology to delineate the low blood volume (LBV) component of the tumor using a pattern recognition technique. For clinical usage, a threshold was established from a probability map of LBV of the tumor to define the subvolume with LBV. If the DCE data acquired on a different vendor scanner using a different pulse sequence with different acquisition parameters would lead to different BV values even though quantification is done by using the same software, this threshold simply cannot be applied. We further realized

that “standardizing” or “harmonizing” acquisition is not always possible due to differences in underlying technologies of each vendor. Then, we attempt to “standardize” the “content”, e.g., BV derived from DCE MRI acquired from different scanners.

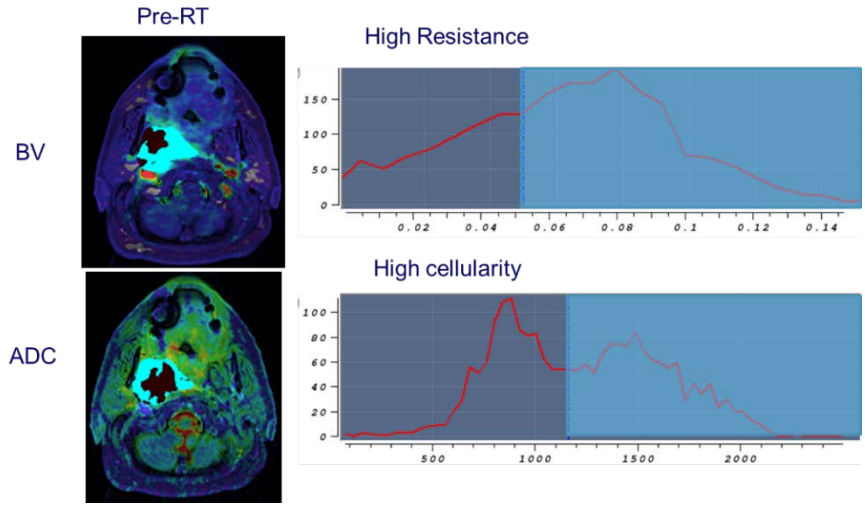


Figure 1: The subvolume of the tumor with low blood volume (dark region) indicates tumor resistance to treatment (up), while the subvolume with low ADC (dark region) suggests high cellularity of the tumor. Cyan color depicts the gross tumor volume (GTV).

Standardization of Blood Volumes (“content”): *We hypothesize that measured BV in a certain type of tissue should follow a same distribution in the population even the data acquired on different scanners and using different pulse sequences.* We selected cerebellum as the tissue of interest for ‘standardization” since cerebellum is always in the FOV when imaging HNC (that requires a large FOV to cover the extended primary and nodal diseases. We found that the mean and standard deviation of BV in the cerebellum VOIs of a group of patients scanned on one vendor scanner differed from those on another vendor scanner. If the measured BV values are normally distributed, based upon our hypothesis, we can have:

$$\frac{BV_1 - mean_1}{SD_1} = \frac{BV_2 - mean_2}{SD_2}$$

where index 1 or 2 indicates scanner 1 or 2. The equation suggests that the distribution of BV in cerebellum should be the same regardless how the measurement is done. Using this equation, we related the BV values measured from scanner 1 to the BV values from scanner 2 to “standardize” the “BV” values. Using this approach, we overcome the “non-standardized” acquisition-caused discrepancies in the parametric maps and subsequent threshold values for the LBV subvolume of the tumor. This concept can be generalized to

other body sites, for which a standardized tissue of region is selected for "standardization" of the "content" of interest. Using this approach, we standardize our threshold to define the subvolume of the tumor with LBV for different scanners.

Individual Patient QA: Although the system QA is performed and quality of images is controlled, sometimes, the quantitative parametric maps of an individual patient scan still can be off from the distribution of the group or population. Then, we used the same concept described above to re-normalize the BV values of individual patients to the group mean if the BV in the cerebellum VOI of the individual patient is 2SD above or below the group mean.

We have been using this approach to "standardize" our results from two different vendor scanners, and to control the unexpected variations in individual patients in supporting the phase II clinical trial that has enrolled approximately 40 patients.

We have presented this concept in the Quantitative Imaging Track in the annual meeting of AAPM 2014, Quantitative Imaging Series in the annual meeting of RSNA 2015 and the QIN panel at 2016 ASTRO annual meeting. [1-3].

§ Automation of Delineation of LBV Subvolumes in HNC

Our current workflow for delineation of the LBV subvolume of the tumor in the clinical trial involves a two-step image process: 1) quantification of the parametric maps from the DCE MRI using a pharmacokinetic model and 2) delineation of the subvolume of the tumor with LBV. The question is whether we can fully automate this process and reduce the process to a single step. A fully automated process will be better for supporting the trials in the clinical environment.

We applied the basic principles in the radiomics and machine learning to the temporal-domain analysis of DCE MRI. Development of this approach involves training and testing. After the algorithm is trained and tested, we were able to delineate directly the subvolume of the tumor with LBV from extracted DCE features, which is a rapid process. Compared to the conventional two-step approach, we were able to achieve the similar accuracy (Figure 2).

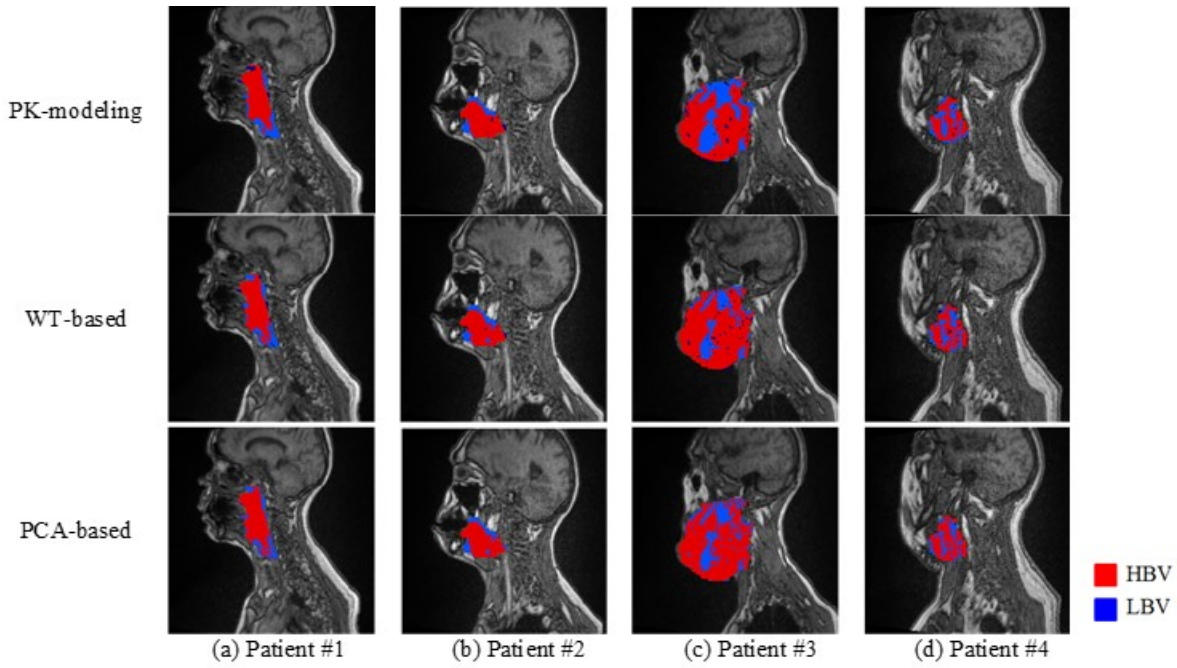


Figure 2: Subvolumes extracted by the two-step method via PK modeling (top row), the Wavelet-based method (middle row), and the PCA-based method (bottom row). Note the similarity of voxels identified to be LBV among the three methods.

We further tested whether the algorithm trained on one dataset could be applied to a dataset acquired with different parameters using different pulse sequence on different scanners. The results from our preliminary test indicated that the accuracy for the data acquired differently was not lower than the data acquired using the exactly same parameters. This algorithm can tolerance to differences in data acquisitions than the pharmacokinetic models. This type of the algorithms has the potential to support variations in DCE acquisition in the clinical environment with further development and validation. We believe this is unique in the field.

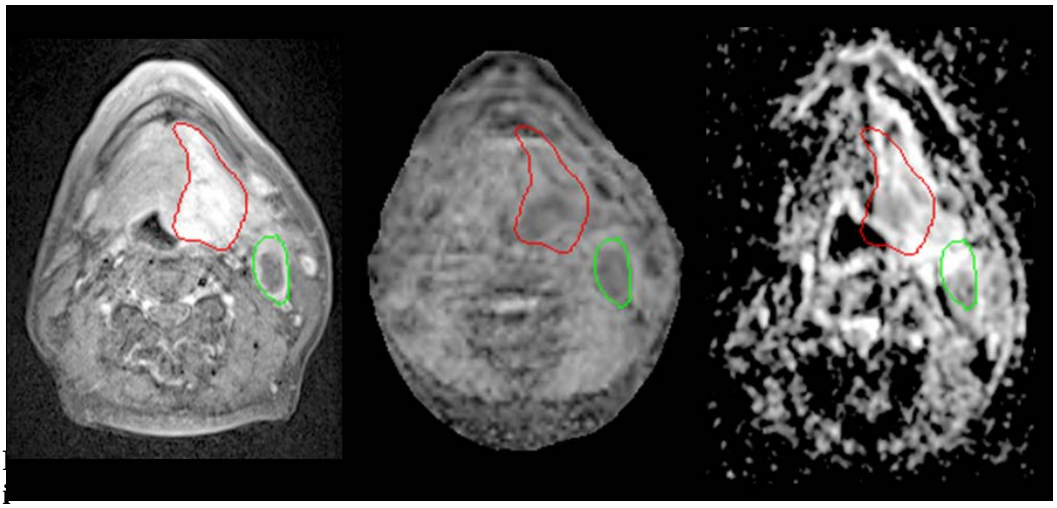
We have presented this work in the annual meeting of AAPM 2015 and in the 2016 MRI in RT workshop and a paper is published in Tomography of the QIN special issue in 2016. [4-6]

§ **Reduction of Susceptibility Effects on Diffusion Weighted Images in HN**

Anatomy in the neck and the base of skull produces large variations in magnetic susceptibility, and results in signal loss and geometric distortion in diffusion weighted images (Figure 3). Also, metals in dental works cause signal loss and geometric distortion. In order to support precision radiation therapy, e.g., boosting the high cellular subvolume (low ADC) of the tumor, and quantitative analysis of diffusion weighted images in HN, it requires to have a pulse sequence that reduces the susceptibility effect in the diffusion weighted images.

One solution is to adopt the RESOLVE pulse sequence that can reduce susceptibility effects on diffusion weighted images dramatically. A trade-off of the RESOLVE pulse sequence is the longer scanning time and sensitive to motion. We have optimized the parameters to balance the acquisition time and quality of diffusion weighted images, including geometric and signal quality. Examples of the slices acquired by the RESOLVE and single shot EPI sequences are shown in Figure 3.

We have been evaluating both geometric distortion and signal quality of diffusion weighted images acquired by the RESOLVE sequence compared to the convention sequence in HN. We have presented the preliminary results in the MRI in RT workshop in 2016.[7] We will continue this evaluation in the next year.



gure 3: Post-Gd T1 weighted image (left), and ADC maps acquired using the RESOLVE pulse sequence and single shot EPI pulse sequence (middle and right, respectively). Red and green contours are gross tumor volumes of primary cancer and affected nodes. Note that the geometric distortion and signal loss in the ADC map acquired by the single shot EPI.

Figure 4: The RT HN protocol including both conventional and advanced imaging sequences.

§ **Optimization of the HN MRI Protocol for the RT Workflow**

Using the quantitative MR images for RT planning requires to position a patient in the RT treatment configuration during the MRI scan. This requires to use the RT immobilization devices, such as five-point mask and bit bar, for the HN MRI scan, which limits the duration of the scan due to patient tolerance. Also, RT planning has unique requirement on the FOV, slice thickness, 2D or 3D, orientation, with or without fat saturation and so on. We have optimized the RT MRI scanning protocol by working with a group of users including radiation oncologists, radiologists, and physicists (Figure 4). We have shared our protocol at national and international conferences in last two years.

§ **Analysis of Heterogeneity in the advanced HN Cancer**

Heterogeneity in the cancers represents a challenge on treatment and assessment of response. Tumor heterogeneity leads to that a single imaging modality often is insufficient to guide for precision treatment and inadequate or even mis-led on response. We leveraged on the image data collected in the randomized phase II clinical trial, including pre-RT FDG PET, and DCE and diffusion MRI pre-RT and during RT, to analyze the image-phenotype features in the advanced HNC and early changes during the course of RT. These image-phenotype features as well as early changes during RT will be correlation with local and regional outcomes, which will tell us which image-phenotype features could be the best radiation boosting target.

Spatial Overlap Between Low Blood Volume and Low ADC in HNC: We have described that the subvolume of the tumor with LBV is potentially radiation resistant while the subvolume with low ADC had high cellularity. The question is whether these two subvolumes of the tumor in HNC have any spatial overlap. Our preliminary analysis of 28 patients showed that 26% and 14% of primary gross tumor volumes (GTVs) had LBV pre-RT and after receiving two weeks of RT, respectively; while 35% and 19% of primary GTVs had low ADC ($<1.2x10^{-3}$ mm^2/s). However, only 9% of the GTVs had both LBV and low ADC before RT and two weeks radiation reduced it to 4%, suggesting the two image-phenotype features represent the two different aspects of the HNC (Figure 5). This work was presented in the annual meeting of ASTRO in 2016 as a research paper as well as in a QIN panel [8-9]. This work will be extended to the 40 patients who have been enrolled in the clinical trial.

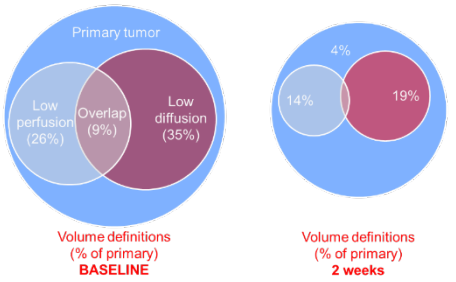


Figure 5: Spatial relationships between the subvolumes of the tumor with LBV and low ADC pre-RT and after receiving 2 weeks RT.

Spatial Relationship Between FDG PET, LBV and Low ADC in the HNC: FDG PET plays an important role in HNC management, including RT target and response

assessment. However, how does the metabolic tumor volume (MTV) defined based FDG uptake have LBV or low ADC is largely unknown. We investigated this question. Table 1 show the volumes defined based upon these metabolic images. Note that the MTV based upon a threshold of 50% of SUV max ($MTV_{50\%}$) was only approximately ¼ of the primary GTV defined based upon the post-Gd T1 weighted images; while the $MTV_{50\%}$ was approximately ½ of the nodal GTV. The median subvolume of the tumor with LBV of the primary tumor was 11 cc, which was approximately 2/3 of $MTV_{50\%}$. The subvolume of the tumor with low ADC was 26 cc.

	Median volume (cc)	Range (cc)
GTVp	69.1	10.2-595.2
MTV50%p	15.5	2.2-259.7
MTV30%p	33.7	3.9-362.0
LBVp	10.9	0.2-158.8
LADCVp	25.9	1.1-180.6
GTVn	11.2	1.3-172.5
MTV50%n	5.2	0.7-61.5
MTV30%n	12.7	1.9-126.4
LBVn	4.6	0-114.2

Table 1: Tumor Volumes Based upon FDG, LBV and Low ADC. p indicates primary tumor; n notes nodal tumor.

We further investigated the spatial overlaps between these subvolumes defined based upon FDG, LBV and low ADC. We found that 98% and 86% of $MTV_{50\%}$ within the primary and nodal GTVs, respectively. However, only 10%-12% of $MTV_{50\%}$ had LBV and 13-15% of $MTV_{50\%}$ had low ADC. Table 2 shows these spatial relationships in detail.

$\%MTV_{50\%p}$ in GTV	$\%MTV_{30\%p}$ In GTV	$\%MTV_{50\%p}/MTV_{30\%p}$	$\%MTV_{50\%p}$ with LBV	$\%MTV_{30\%p}$ with LBV	$\%MTV_{50\%p}$ with LADC	$\%MTV_{30\%p}$ with LADC
97.8% (53.2%-100%)	89.2% (1.9%-100%)	45.9% (10.9%-71.7%)	10.3% (0-77.7%)	13.1% (0-73.0%)	71.3% (3.8%-93.0%)	53.3% (2.0%-82.6%)
$\%MTV_{50\%n}$ in GTV	$\%MTV_{30\%n}$ In GTV	$\%MTV_{50\%n}/MTV_{30\%n}$	$\%MTV_{50\%n}$ with LBV	$\%MTV_{30\%n}$ with LBV		
86.2% (14.1%-100%)	73.1% (12.4%-99.8%)	38.6% (11.9%-71.0%)	12.2% (0-89.8%)	14.6% (0-80.0%)		

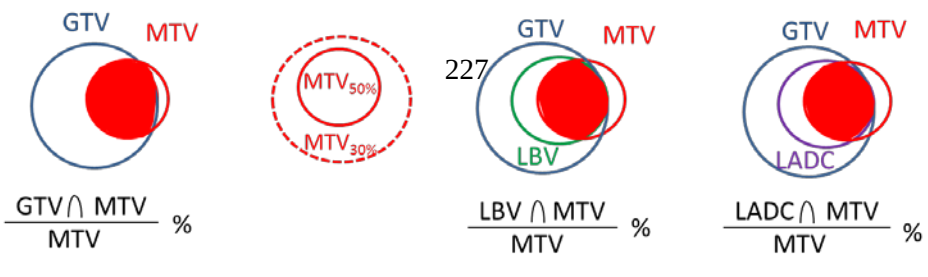


Table 2: Spatial Relationship Between of GTV, MTV, LBV and low ADC

In the 40 HN tumors, the voxel-level correlations between SUV FDG and BV values had the correlation coefficients varied from 0.55 to -0.12, and between SUV FDG and ADC had the correlation coefficients from 0.15 to -0.6.

These data suggest that advanced HN cancers exhibit a large extent of heterogeneity. However, which of these image-phenotype features represent the most aggressive or radiation resistant subvolume of the tumor is to be determined when we correlate the features with outcomes.

§ Participation in QIN Challenges

We participated in several QIN challenges: 1) arterial input function for DCE analysis led by Wei Huang, 2) diffusion quantification challenge led by David Newitt; 3) T1 measurement challenge led by Octavia Bane, and 4) DSC challenge led by Kathleen Schmainda. All challenges led to submitted abstracts for the 2017 ISMRM annual meeting[10-12]. The third challenge has a RNSA abstract. The first part of the first challenge had led to one publication in Tomography [13].

§ Development and Evaluation of Other QI Tools

We have developed other QI tools that are not directly related to our QIN HNC project.

Hypercellularity Volume Delineation for GB*M*: GBM represents many challenges, including target definition for surgery and radiation and response assessment due to diffuse disease and edema. We have developed a QI tool to delineate the hypercellular tumor volume (HCV) by suppressing edema using high b-value diffusion weighted imaging to reveal the solid tumor (Figure 6). We have found that 40% of the HCV were non-enhanced. The large HCV was associated with short progression-free survival, suggesting that HCV represents one of aggressive components of GBM. We have published a paper on this research. [14]

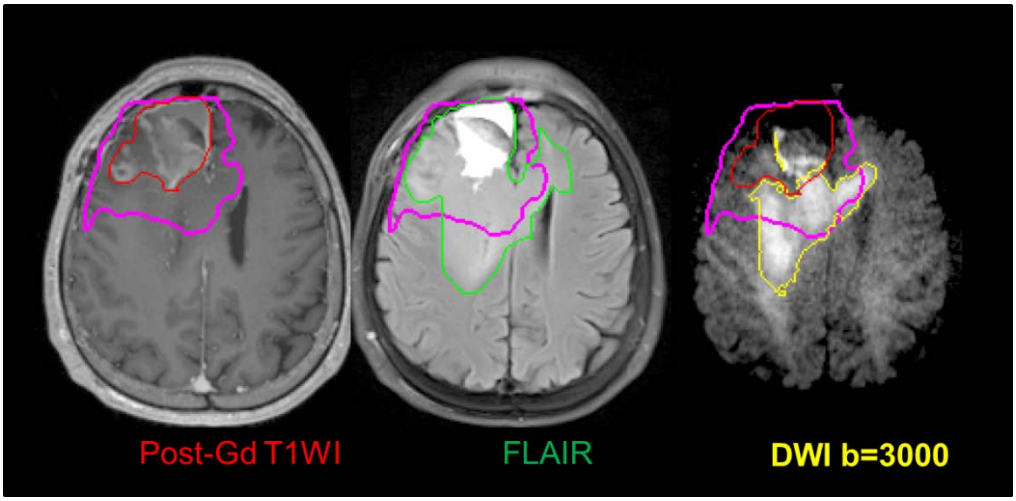


Figure 6: Post-Gd T1 weighted, T2 FLAIR and diffusion weighted images with b=3000 s/mm2. Red, green and yellow contours represent enhanced GTV, FLAIR abnormality volume and hypercellularity volume. Dark pink contour depicts the 95% prescribed dose volume which missed a portion of HCV based upon the conventional treatment planning.

Recently, we have tested whether we can bring this QI tool to a site where there is no MRI expert to conduct a multi-center clinical trial. First, we tested diffusion imaging on the ice water phantoms at both sites and led to 1.3% discrepancies between two sites. Secondly, we implemented a fully automated version of software to delineate the HCV for defining the radiation boost target. Thirdly, we tested the geometric accuracy of diffusion weighted images using a RESULVE sequence. Except the first 1-2 mm around the brain surface, the diffusion images have the accuracy at the level of the spatial resolution uncertainty. This technique does not require a long time for image acquisition. This technique is ready to be deployed in different sites for support clinical trials. Also, this technique could be a useful

tool to differentiate true progression from pseudo-progression, particularly when the GBM vasculature is altered by anti-angiogenesis drugs.

T1 Repeatability Test: We have performed T1 repeatability tests on brain since T1 is important in DCE quantification. In this work, we compared two methods for T1 quantification. Our results are published in 2015 [15-16].

PLANS FOR NEXT YEAR

1. We will continue to support the workflow for image acquisition and analysis for the randomized phase II clinical trial. Right now, we have enrolled approximately 40 patients from two sites.
2. We will complete analysis of the image-phenotype features in advanced HN cancers, and write two papers.
3. We will develop a method to automate our current workflow to reduce expert efforts required to support the clinical trial.
4. We will participate in other QIN challenges.
5. We will further improve and develop the one-step method to delineate the subvolume of the tumor with LBV.
6. We will collaborate with other sites to explore the radiomics analysis.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

1. E. Jackson, M. McNitt-Gray, R. Jeraj, Y. Cao. Quantitative Imaging: Techniques, Applications, and Challenges. *Med Phys.* MO-E-12A-1, 2014. Presented in the QI track symposium at the annual meeting of AAPM 2014.

2. Y. Cao, "Quantitative Imaging for DCE-MRI: Applications and Future Directions", in the series course of RC225 of Quantitative Imaging Mini-Course: Image Modality Specific Issues, 101th Scientific Assembly and Annual Meeting of RSNA, Nov 29-Dce 4, 2015, Chicago, IL

3. Y. Cao, "DCE-Perfusion and Diffusion-Weighted MR Imaging for Clinical Decision Support in Head and Neck cancer", on Panel 3 of Advanced Quantitative Imaging for the Radiation Oncologist: Response Assessment and Targeting for Clinical Trials and Practice, A View from the NCI's Quantitative Imaging Network in the 58th Annual meeting of ASTRO, Sept 25-28, 2016, Boston, MA

4. D. You, M. Aryal, S. Samuels, A. Eisbruch, and Y. Cao. Wavelet-Based Temporal Feature Extraction from DCE-MRI to Identify Sub-Volumes of Low Blood Volume in Head-And-Neck Cancer. *Med Phys*. SU-E-J-241. AAPM 2015.

5. Daekeun You, Madhava Aryal, Yue Cao. TEMPORAL FEATURE EXTRACTION FROM DCE-MRI TO IDENTIFY POORLY PERFUSED TUMOR SUB-VOLUMES

IN HEAD-AND-NECK CANCER. Abstract, The 4th MRI in RT workshop, June 18-19, 2016, Ann Arbor, MI

6. Daekeun You, Madhava Aryal, Stuart Samuels, Avraham Eisbruch, and Yue Cao. Wavelet-based Temporal Feature Extraction from DCE-MRI to Identify Significant Subvolumes of Tumors Related to Outcomes of Radiation Therapy in Head-and-Neck Cancer. *Tomography*, Vol 2(4): 341-352, NCI QIN issue, 2016.

7. Madhava P Aryal, Timothy Ritter, Michelle Mierzwa, Avraham Eisbruch, and Yue Cao Readout Segmented EPI in Diffusion Weighted Imaging of Head and Neck Cancer: Comparison with Single-shot EPI. Abstract, The 4th MRI in RT workshop, June 18-19, 2016, Ann Arbor, MI

8. J.Y. Lee, S. Samuels, M.P. Aryal, C. Lee, Y. Cao, A. Eisbruch,. Characterizing Regions of Hypoperfusion and Restricted Diffusion in Head and Neck Cancer Patients Enrolled on a Prospective Phase 2 Randomized Trial. Int J Rad Onc Biol Phys, 96(2S), S71, 2016.

9. Y. Cao, "DCE-Perfusion and Diffusion-Weighted MR Imaging for Clinical Decision Support in Head and Neck cancer", on Panel 3 of Advanced Quantitative Imaging for the Radiation Oncologist: Response Assessment and Targeting for Clinical Trials and Practice, A View from the NCI's Quantitative Imaging Network in the 58th Annual meeting of ASTRO, Sept 25-28, 2016, Boston, MA

10. Kimberly Li, Yiyi Chen, Yun Yu, Xia Li, Andriy Fedorov, Guido H. Jajamovich, Dariya I. Malyarenko, Madhava P. Aryal, Peter S. LaViolette, Matthew J. Oborski, Finbarr O'Sullivan, Richard G. Abramson, Kourosh Jafari-Khouzani, Aneela Afzal, Alina Tudorica, Brendan Moloney, Sandeep N. Gupta, Cecilia Besa, Jayashree Kalpathy-Cramer, James M. Mountz, Charles M. Laymon, Mark Muzi, Paul E. Kinahan, Kathleen Schmainda, Yue Cao, Thomas L. Chenevert, Bachir Taouli, Fiona Fennessy, Thomas E. Yankeelov, Xin Li, Christopher Ryan, Wei Huang. The Effects of AIF Quantification Variations on DCE-MRI Prediction of Soft Tissue Sarcoma Response to Preoperative Therapy: A Preliminary Multicenter Study. 2017 ISMRM abstract (submitted).
11. O. Bane, S. Hectors, M. Wagner, L. R. Arlinghaus, M. Aryal, M. Boss, Y. Cao, T. L. Chenevert, F. Fennessy, W. Huang, N. Hylton, J. Kalpathy-Cramer, K. E. Keenan, D. Malyarenko, R. Mulkern, D. Newitt, K. F. Stupic, L. Wilmes, T. Yankeelow, Y. Yen, S. E. Russek, and B. Taouli. Assessment of interplatform reproducibility of T1 quantification methods used for DCE-MRI: results from a multicenter phantom study. 2017 ISMRM abstract (submitted).

12. D. C. Newitt, D. Malyarenko, T. L. Chenevert, C. C. Quarles, L. Bell, A. Fedorov, F. Fennessy, M. A. Jacobs, M. Solaiyappan, S. Hectors, B. Taouli, K. M. Schmainda, M. A. Prah, Y. Yen, J. Kalpathy-Cramer, E. Taber, C. Kroenke, Y. Cao, M. Aryal, M. Muzi, P. Kinahan, T. Yankeelow, L. R. arlinghaus, M. A. Boss, A. Shukla-Dave, and N. Hylton.

Multi-site Concordance of DWI Metrics: Results of the NCI Quantitative Imaging Network ADC Mapping Collaborative Project. 2017 ISMRM abstract (submitted).

13. W. Huang, Y. Chen, A. Fedorov, X. Li, G. H. Jajamovich, D. I. Malyarenko, M. P. Aryal, P. S. LaViolette, M. J. Oborski, F.O'Sullivan, R. G. Abramson, K. Jafari-Khouzani, A. Afzal, A.Tudorica, B. Moloney, S. N. Gupta, C. Besa, J. Kalpathy-Cramer, J. M. Mountz, C. M. Laymon, M. Muzi, K. Schmainda, Y. Cao, T. L. Chenevert, B. Taouli, T. E. Yankeelov, F. Fennessy, and X. Li. The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. *Tomography*, Vol 2(1):56-66 (2016). DOI: 10.18383/j.tom.2015.00184.

14. P. P. Pramanik, H. A. Parmar, A. G. Mammoser, L. R. Junck, M. M. Kim, C. I. Tsien, T. S. Lawrence, and Y. Cao. Hypercellularity Components of Glioblastoma Identified by High b-value Diffusion-Weighted Imaging. *Int J Rad Onbc Biol Phys*, 92(4):811-819, 2015. PMID: 26104935. PMCID: PMC4507284

15. M. Aryal and Y. Cao. Impact of Uncertainty in T1 Measurements on Quantification of Dynamic Contrast Enhanced MRI. Med Phys, SU-D-303-3. AAPM 2015.

16. Madhava P Aryal, Thomas L Chenevert, Yue Cao. Impact of Uncertainty in Longitudinal T1 Measurements on Quantification of Dynamic Contrast Enhanced MRI. NMR in Biomedicine, 29(4):411-9, 2016.

U01CA187013: Resources for development and validation of Radiomic analyses & Adaptive Therapy

University of Arkansas for Medical Sciences and Emory University

Fred Prior, PhD
Ashish Sharma, PhD

INTRODUCTION

Since it's inception in 2011 the Cancer Imaging Archive (TCIA) has been NCI's primary resource for acquiring, curating, managing and distributing images and related data to support Cancer Research. TCIA is visited each month by more than 4000 users from around the world, actively supports over 8000 registered users (and a larger community who access data anonymously) and has provided data used in over 400 peer reviewed publications and graduate theses. Figure 1 summarizes some of the key TCIA utilization metrics and of June 2016.

TCIA is the primary image repository for the Quantitative Imaging Network (1), the Center for Multiple Myeloma Nanotherapy (2), the National Lung Screening Trial (3), and a number of NCI internal and sponsored research programs including the exceptional responder's initiative and the Data Science Bowl 2017 Challenge. TCIA also supports challenge competitions organized and/or managed by QIN and ITCR funded researchers and NCI (4-6). With NCI's approval, TCIA is currently supporting a PCORI funded prospective clinical trial as the Radiorepository for all image and radiation oncology planning data (7).

TCIA is a Recommended Data Repository for Nature Scientific data (8), one of the repositories recommended by PLOS One as part of its Open Data Policy(9), an Elsevier Supported Data Repository (10) and a repository recommended by F1000Research (11). BioSharing.org maintains a complete list of publishers that consider TCIA as an approved repository.

Careful curation and strict quality control processes are two key activities that have led to the success of TCIA. The TCIA service includes expert curation and quality control of incoming data sets, extension of the software/technology used to meet production standards of service and availability, and dissemination of knowledge to the wider research community in areas of DICOM de-identification and open data. Significant advances have been made in TCIA curation automation and accuracy.

TCIA is a data platform to support radiomics research. We continue to expand this capability by using the TCIA API to stream data into radiomics pipelines under the control of the QIN Portal. The Galaxy Project (https://galaxyproject.org/) provides a web-based interface for biomedical research. There are a wide range of Galaxy plugins that support such tasks as accessing data from existing repositories, processing data, converting between data formats, statistical analysis and data visualization. The vast majority of these tools are

based on Next Generation Sequencing. The Galaxy plugin interface has been identified as the basic mechanism for integrating radiomic analysis pipelines into the Galaxy framework to create a QIN Portal.

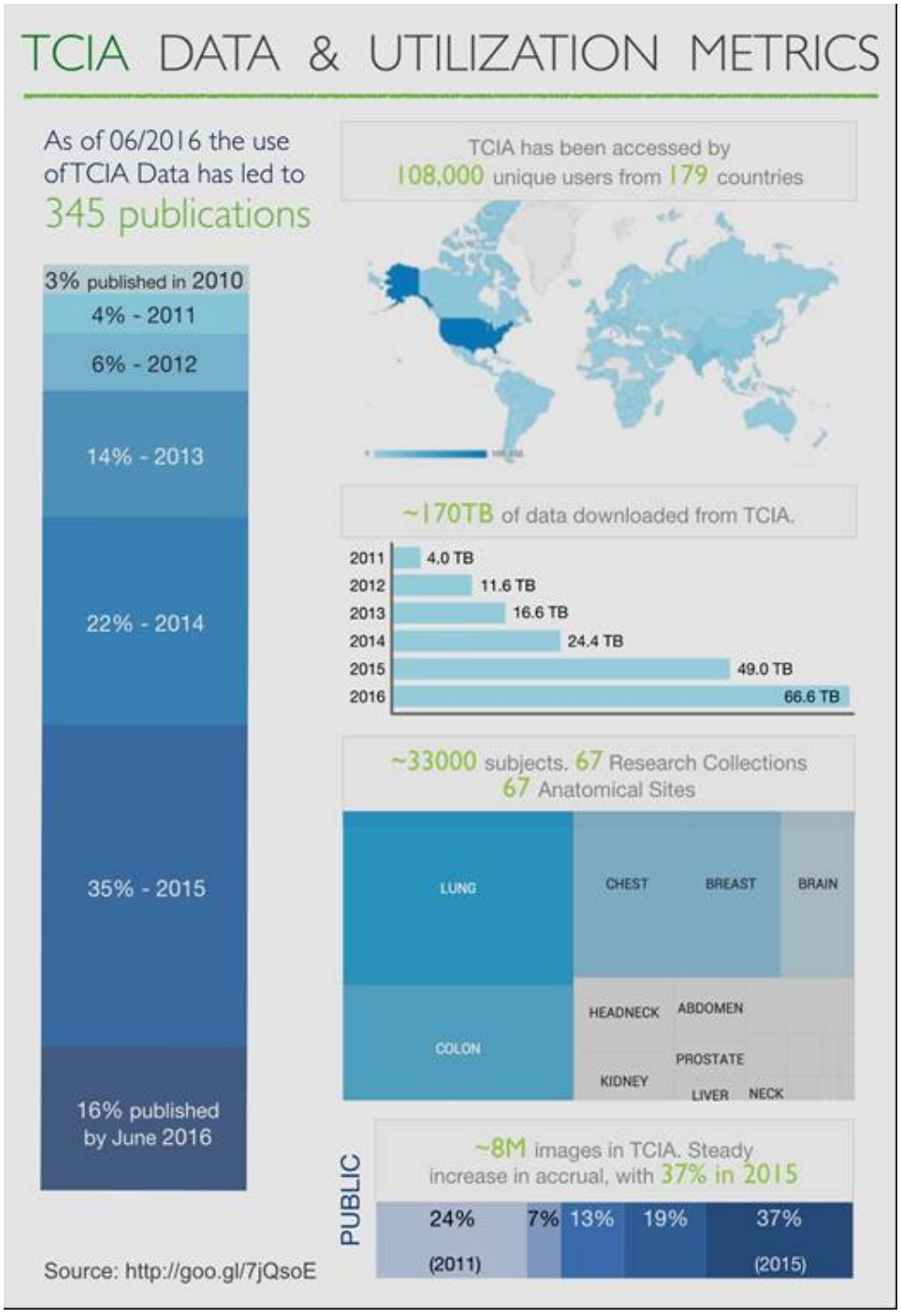


Figure 1: TCIA Utilization Metrics

The analysis of radiomics feature sets combined with clinical data to produce imaging phenotypes remains a key focus of our team's research. We have made progress in our radiomic pipeline development and in the use of Eureka to identify cancer phenotypes.

PROGRESS OVER THE PREVIOUS YEAR

Curation workflows for Radiology data are now implemented in TCIA using the Posda open source toolkit (12, 13). Posda is a custom set of curation workflow tools developed by our team and currently in daily use by TCIA staff. Posda relies heavily on our in-depth knowledge of the DICOM standard and employs a DICOM validation rule set. Posda supports inspection and editing of PHI, and allows editing of the data to correct inconsistencies found particularly in complex DICOM data objects such as those associated with Radiation Therapy. The current Posda release supports: DICOM send and receive, summary spreadsheets of image counts by subject, study, series and modality for verification that the data was received or sent; automatic detection of subject, study, and series level DICOM inconsistencies, semi-automatic correction of inconsistencies; detection and correction tools for RTSTRUCT linkage errors; checks for duplicate SOP Instance UID's; semi-automated detection and correction of PHI. A publication detailing the rapid evolution of Posda is ready for submission.

RT collections have been acquired from a number of sites, stored in TCIA but not yet linked to treatment records. Because of the use of TCIA for the PCORI RT trial referenced above we have moved beyond the proposed verification study for this aim and we are demonstrating the ability to download longitudinal RT data sets and generate a composite dose for each subject as part of the PCORI trial. All calculated dose results for the trial will be added back to the TCIA collection once the trial is fully underway.

The TCIA infrastructure has been enhanced by extending the application programming interface (API). The API's security infrastructure has been upgraded and now includes security fixes to avoid SQL injection attacks. Progress has been made on support of secure programmatic access to QIN private collections. The overall workflow is as follows: TCIA users are mapped to groups, with one group per TCIA collection. This group information is maintained in a central authorization database. We have developed plugins that inspect incoming API requests and can make authorization decisions (allow/deny). These authorization plugins have been developed to work with LDAP as the central authorization database where user-group mapping is maintained. NBIA currently uses CSM (common security module) as its authorization database. However, CSM is being retired by the NBIA development team. We are currently collaborating with the NBIA team and evaluating authorization databases such as LDAP.

§ Data Integration and Mashups

During the past year we began work on a data integration layer, called Data Café, which facilitates the creation of domain/problem specific data cohorts called biomedical data

lakes. These lakes can contain imaging metadata, radiomic features, clinical data and other structured data sets. They are stored in HDFS (Hadoop File System) and accessed via Apache Drill. This was presented at the VLDB workshop on Data Management and Analytics for Medicine. Additionally, we have also used the TCIA API to develop a data sharing middleware, called MeDIATOR that extends the TCIA shared list capability. TCIA shared lists are a popular way for creating a sharable reference. MeDIATOR expands upon the shared list paradigm, and allows investigators to create references of images as well as non-imaging data. With an increase in the number of TCIA collections that include clinical data as well as derived imaging feature data, MeDIATOR is expected to play a major role in encouraging and facilitating reproducible research with TCIA. MeDIATOR works by providing an API that can be integrated within the TCIA API ecosystem.

§ QIN Portal

We proposed extending the Galaxy Project web-based interface for biomedical research to include custom processing tools for radiomic analysis creating a QIN portal. One of the first tasks was extending Galaxy to recognize the DICOM and NIfTI imaging formats. We initially did this by triggering on the image suffix, which is similar to the method used by many imaging libraries such as the ITK. This proved impractical, as Galaxy renames all files internally using a generic name scheme dataset_XXXX.dat. This required us to further extend Galaxy by providing a binary 'sniffer' function to read and recognize the magic number of the imaging formats (Figure 2). We next added the capability to view DICOM/NIfTI files utilizing Papaya, a Javascript-based DICOM/NIfTI viewer. We were able to use the Galaxy visualization plugin tools to associate DICOM/NIfTI files with the Papaya viewer (Figure 3).

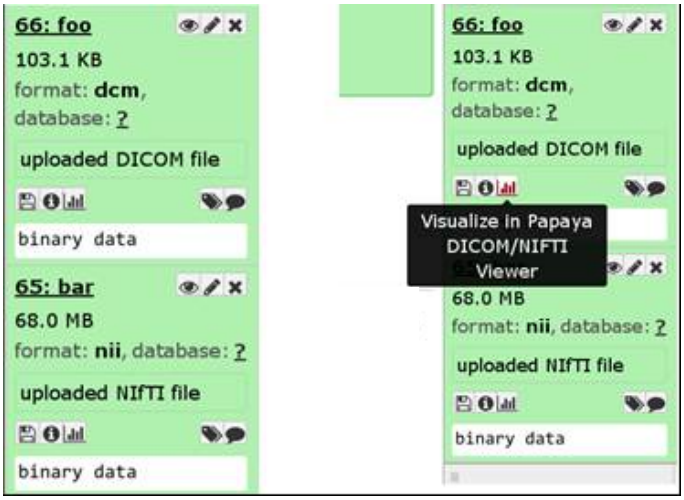


Figure 2: QIN Portal Control panel. On the image on the left, we see two files that have been uploaded into Galaxy named foo and bar with no file suffixes. In both cases Galaxy was able to identify the file type by reading the image header. In the image on the right, if we hover over the visualization button for the data set, we see the option to visualize the data set using Papaya.

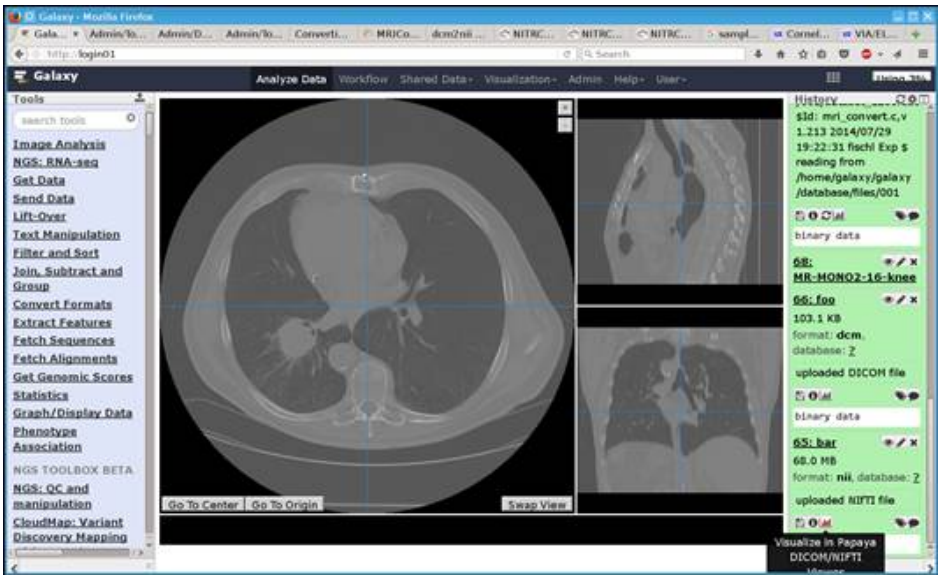


Figure 3: QIN Portal Image Viewer. Papaya provides a Javascript-based DICOM/NIfTI viewer, which has been integrated into our instance of Galaxy to provide visualization capabilities within our QIN portal.

We have an operational lung segmentation pipeline that includes extraction of a feature set defined by Gierada et al.(14). This pipeline has been implemented as a Galaxy plugin and is currently running on the high performance computing system at Washington University (CHPC).

We have also been exploring the use of Docker containers as a means to encapsulate Radiomics and Radiogenomics applications and deploy imaging pipelines on the cloud. We have developed a tool called YunPipe (https://github.com/sharmalab/yunpipe) that allows researchers to execute imaging pipelines on the cloud. Imaging pipelines are authored using the Common Workflow Language (CWL). YunPipe is capable of running pipeline on Amazon Cloud. In the coming year we will be extending YunPipe to work with Google Cloud, and integrating it with TCIA, via the TCIA API. We will also be containerizing a lung segmentation pipeline that was developed at Washington University, St. Louis and UAMS (lead by Dr. Prior). A paper describing YunPipe is under preparation. Dr. Sharma is leading this work.

This complements and enhances our parallel efforts with Galaxy. Galaxy supports the dockerized deployment of tools as well as execution of these tools on public cloud environments. The eventual goal remains one where QIN researchers will be provided a Radiomics portal that would allow them to upload Radiomics tools and execute them on TCIA image data without an explicit download of images

§ Lung Cancer Radiomics Pipeline

We are continuing our work exploring radiomic analyses of lung cancer using TCIA data and further developing pipelines for deployment in QIN Portal. We previously reported preliminary work using NLST data(15) in which we first segmented the lung and then all closed objects in the lung that are larger than a voxel. We used hierarchical clustering to produce object classes then attempted to determine which feature classes best correlated with clinical outcomes (cancer diagnosis in this instance). Unfortunately the results thus far have not been promising.

We shifted focus to improving feature extraction using radiologist-identified objects of interest. We chose the LIDC/IDRI dataset to train and test two kinds of models to identify image features for prediction of lung cancer. One model is based on the set of image features defined by Gierada et al.(14), the other is based on deep convolutional neural networks (CNN)(16), which learn a feature set. The LIDC/IDRI datasets contains 1018 CT cases with "truth" established by four experienced thoracic radiologists, which makes it optimal for testing and validation of the models. Both models were given segmented nodules and randomly selected regions of healthy lung as illustrated in Figure 4. All analyses were performed over the population of nodules. We used 1000 non-nodules and 1065 nodules with 80% of the data used for training the models and 20% for validation.

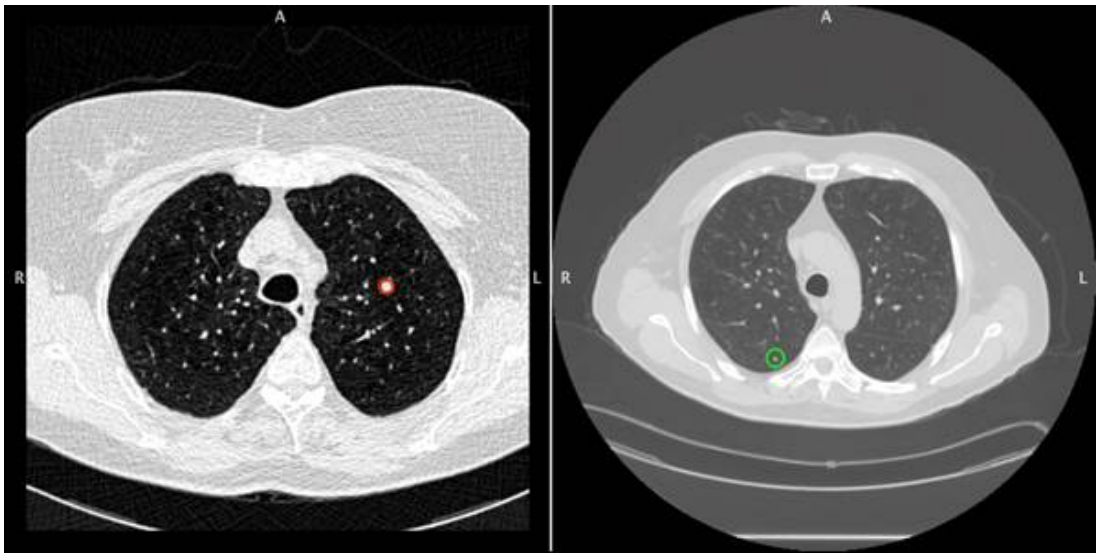


Figure 4: Nodule vs. non-nodule comparison. The left image is a nodule (rated malignancy = 3 from LIDC-IDRI) with consensus radiologist segmentation (red). The image on the right is a "non-nodule", with computer segmentation (green).

The Gierada feature set was extracted from segmented regions. A principal components analysis (PCA) was performed for dimensionality reduction. As shown in Figure 5A, the first 3 components (representing 50 features) explained most of the data. A support vector machine (SVM) with a radial basis function was trained to partition the test

set. The resulting ROC analysis is illustrated in Figure 5B.

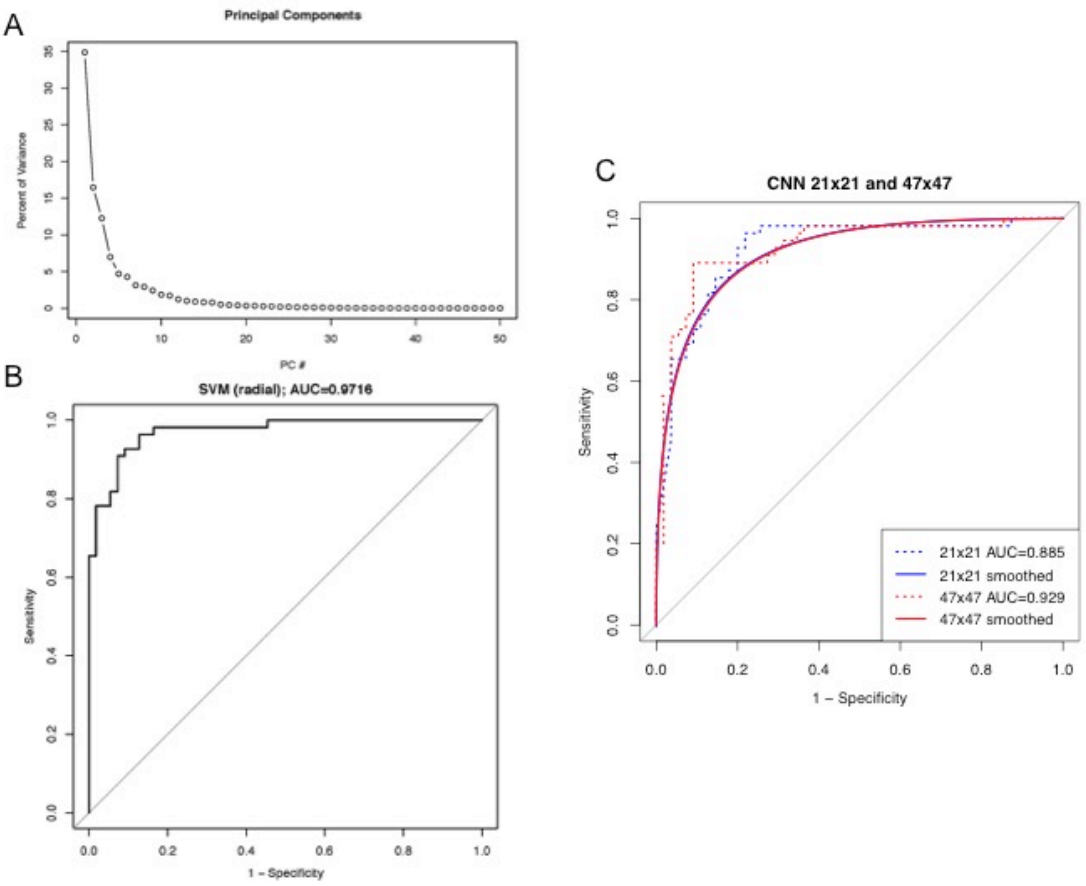


Figure 5: Preliminary Radiomics results derived from LIDC data. A) PCA analysis of feature space, B) ROC analysis of Gierada et al. feature set and Support Vector Machine classifier; C) ROC analysis using 2 different sized voxel cubes as input to the CNN.

Recently we started working on a CNN, which used cubes of voxels centered in the region of interest (nodule, non-nodule region) and learns a 200 dimensional feature representation that partitions the training set. Preliminary results from the CNN are presented in Figure 5C. Relative to the literature these preliminary results are too precise suggesting the models are over fitting the data. Independent test sets are currently being acquired and experiments with differing percentages of the data allocated to the training and test sets are underway.

§ Eureka

In support of evaluating and validating clinical-imaging phenotypes, Eureka implements processes for parsing diverse data and exporting it into diverse tools for clinical research. Eureka provides APIs for creating data adapters that can be configured to parse a variety of clinical datasets in flat file and database formats. A data adapter was implemented

for QIN for accessing NLST data. Similar adapters have been implemented as part of other projects for accessing data from Emory's clinical systems and the publicly available MIMIC II dataset. Data thus transformed can be output by Eureka into a graph database or into the widely deployed i2b2 data warehousing system. I2b2 supports rapid interactive data query and export from a web browser into formats expected by standard statistical analysis tools.

§ Outreach

End user training and easily accessible documentation and tutorials are an essential component of community engagement. Each year we present hands-on TCIA training courses at professional society meetings including the Radiological Society of North America annual meeting. A short course on the advanced features added to TCIA was presented at RSNA 2016: Course number RCA55 (Accessing and using 'Big Data' Diagnostic Image Archives for the Study of Cancer Proteomics, Genetics and Pathology). In the past year we also presented a demonstration of Eureka at the QIN annual meeting. The complete list of presentations and training courses is included in other sections of this report.

The imphub open source community support and software development environment was ported to the University of Arkansas for Medical Sciences and is in general use to support TCIA operations, Posda and API development, and other aspects of this project. This site also supports interactions with Google Summer of Code students.

§ Personnel Changes

In year 3, two new faculty joined the QIN team — Dr. Suprateek Kundu and Dr. Yasir Rahmatallah. Dr. Kundu is on the faculty of the Biostatistics and Bioinformatics department at Emory. Under his direction, we will continue work on developing graph-based statistics, that use phenotypes that are developed with Eureka, and helps researchers evaluate and validate these integrated clinical-imaging phenotypes. This work was carried out using the NLST datasets, as well as other, publicly available, TCIA datasets. We will also explore integration with genomic data, via the Genomic Data Commons. Dr. Rahmatallah is an assistant professor of biomedical informatics at UAMS. He will work with Dr. Prior to explore new approaches to graph based analysis of radiomics features.

COLLABORATIONS WITHIN THE NETWORK

Collaboration with the Mayo-QIN: We are evaluating Grunt — a Docker middleware developed by the Mayo-QIN that supports the creation of REST APIs for Docker containers. We have dockerized our lung segmentation algorithm, and integrated it with Grunt. We have since deployed it on AWS. In the coming months, as part of a BIDS WG project, we will be evaluating various workflow techniques such as the Common Workflow Language.

Collaboration with members of the BIDS and IAMS WG: We are conducting a study that is evaluating the integration of imaging data with non-imaging and clinical data. The objective of this study is to determine best practices for data curation and management, as well as outlining the value proposition of such integration. The goal is to publish a white

paper and begin QIN projects involving curation and sharing of radiomics feature data.

Collaboration with prospective members: Our QIN team has partnered with Stony Brook University on a new QIN proposal and an ITCR U24 proposal. These collaborations establish cross network projects at the outset.

RELATED COLLABORATIONS OUTSIDE THE NETWORK

Dr. Sharma is a member of the BIDS working group and is participating in two WG projects. Dr. Sharma is working with the MGH team and Dr. Jayashree Kalpathy-Cramer (ITCR U24) on a project where the Emory team will deploy lung segmentation pipelines onto AWS using YunPipe. Dr. Sharma has also started collaboration with the ISB Cancer Genomics Cloud Pilot, to explore an integration of TCIA data with the cloud pilots. These collaborations will continue in the upcoming year.

Dr. Bosch is an active member of DICOM WG-7 (RT Information Objects) where he is currently involved in the development of second-generation DICOM RT Dose objects and ROI templates. Dr. Bosch serves as Connectathon Test Manager for the IHE Radiation Oncology (IHE-RO) Domain, which seeks to improve interoperable exchange of DICOM RT objects among commercial software systems. He also participates in AAPM TG-263, which seeks to standardize nomenclature for radiotherapy treatment planning.

Dr. Bosch (lead PI) and Dr. Prior direct the Radiorepository core of the PCORI funded Pragmatic Randomized Trial of Proton vs. Photon Therapy for Patients with Stage II or III Breast Cancer (PI: Bekelman). The Radiotherapy Comparative Effectiveness (RADCOMP) Consortium was given permission by NCI to use TCIA in this prospective trial. The ability of TCIA to support this trial was a direct result of our efforts under AIM 1 to expand the capabilities of TCIA to collect and curate RT objects.

The UAMS-Emory team joined Dr. Saltz and his Stony Brook university team to submit a QIN application focusing on radiomic-pathomic correlations and management of feature space representations. This application featured cross-network projects.

§ Summary of Progress

Table 1 lists the publications and presentations produced by the UAMS-Emory team during the past year. Figure 6 summarizes the tools our team has produced thus far and how they are deployed in trials or other applications.

Tool Name	Tool Type	Tool Description	Created as part of QIN?	Clinical trials & other uses
TCIA	Open Access Information Repository	An information resource that provides open access Radiology and Pathology images, clinical trial and other patient related data to support cancer research.	Partially	PCORI Pragmatic Randomized Trial of Proton versus Photon Therapy
POSDA	Curation and De-identification Software for DICOM images and objects	A Database for storing relationships among DICOM images and a set of tools and user interfaces for exploring relationships among DICOM Images, making bulk changes to DICOM images, and removing PHI.	Partially	Deployed in TCIA
Project Bindaas	Data Management and Sharing	A middleware tool to simplifies the development and deployment of programmatic interfaces for databases.	No	Deployed in TCIA
YunPipe	Cloud Based Data Processing	A tool that allows you to run imaging worksflows on Amazon Cloud Services. The tool requires that all components of the workflow be containerized as Docker images. Workflows are authored using CWL	YES	
DataScope	Data Exploration and Visualization	A platform for creation of scientific mashups and visualizing multi-dimensional datasets.	NO	
Data Café	Data Integration	Big Data integration platform that can combine mutliple types of data such as clinical data and Radiomic/Pathomic features	YES	
caMicroscope	Digital Pathology	A platform for digital pathology data management, visualization and analysis.	NO	Deployed in TCIA
Eureka! Clinical Analytics	Clinical Phenotyping	Eureka! Clinical Analytics enables biomedical researchers to create databases containing clinical data of interest for cohort discovery and analysis. It allows one to define clinical or imaging phenotypes and represent these as graphs in a graph database.	Partially	

Figure 6: Tools developed by the UAMS-Emory QIN

Rosenstein B, Capala J, Efstathiou J, Hammerbacher J, Kerns S, Kong F-M, Ostrer H, Prior F, Vikram B, Wong J, Xia Y, How Will Big Data Improve Clinical and Basic Research in Radiation Therapy? International Journal of Radiation Oncology, 2015; doi:10.1016/j.ijrobp.2015.11.009
Benedict SH, Hoffman K, Martel MK, Abernethy AP, Asher AL, Capala J, Chen RC, Chera B, Couch J, Deye J, Efstathiou JA, Ford E, Fraass B, Gabriel PE, Huser V, Kavanagh BD, Khuntia D, Marks LB, Mayo C, McNutt T, Miller RS, Moore KL, Prior F, Roelofs E, Rosenstein BS, Sloan J, Theriault A, Vikram B. Overview of the American Society for Radiation Oncology–National Institutes of Health–American Association of Physicists in Medicine Workshop 2015: Exploring Opportunities for Radiation Oncology in the Era of Big Data. International Journal of Radiation Oncology• Biology• Physics. 2016; 95(3):873-9.
Kathiravelu, P. & Sharma, A. (2015). MEDIator: A Data Sharing Synchronization Platform for Heterogeneous Medical Image Archives. In Workshop on Connected Health at Big Data Era (BigCHat'15), co-located with 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2015). Aug. 2015. http://doi.org/10.13140/RG.2.1.3709.4248
Post AR, Pai AK, Willard R, May BJ, West AC, Agravat S, et al. Metadata-driven Clinical Data Loading into i2b2 for Clinical and Translational Science Institutes. AMIA Summits Transl Sci Proc. 2016:184-93.
Wu, M. (2016). Association between Clinic Characteristics and Positive Low-dose CT outcome among current and Former Smokers (Master's Thesis). Retrieved from http://pid.emory.edu/ark:/25593/rhxbz
Tobias M, Porter J, Sharma A, Politite D, Prior F, "Utilizing the Galaxy web-based platform to simplify access to imaging data and processing pipelines," Abstract and poster presentation, Quantitative Imaging Network (QIN) Face-to-Face meeting in support of 1U01CA187013-01, April 11, 2016.
Bennett W, Bosch W, Smith K, Prior F, "Analysis and Remediation of DICOM consistency issues using the Posda open source DICOM toolkit," Abstract and poster presentation, Quantitative Imaging Network (QIN) Face-to-Face meeting in support of 1U01CA187013-01, April 11, 2016.
Post A, Wu M, Sharma A, "Eureka! Clinical Analytics," Abstract and Hands on Demonstration, Quantitative Imaging Network (QIN) Face-to-Face meeting in support of 1U01CA187013-01, April 11, 2016.
Sharma, A, Prior F. "Large Scale Data Management, Computation, and Analysis for Quantitative Imaging Research", presented by A. Sharma, Quantitative Imaging Network (QIN) Face-to-Face meeting in support of 1U01CA187013-01, April 11, 2016
Kathiravelu, P. & Sharma, A. "SPREAD - System for Sharing and Publishing Research Data". In Society for Imaging Informatics in Medicine Annual Meeting (SIIM 2016). June 2016.
"**Informatics resources for cancer radiomics**" Presented by F.W. Prior, NCI-Sponsored Session, AACR, New Orleans, LA, April 19, 2016.

Table 1: Publications and presentations from QIN involvement

PLANS FOR NEXT YEAR

During the coming year we plan to complete the following work on each of our aims.

§ AIM 1

Create a mashup between TCGA clinical data (available in GDC) and TCGA images available in TCIA. This will allow researchers to create cohorts and access images using clinical and imaging data. As described earlier, we are working with members of the BIDS WG to develop an information model for clinical data that accompanies all non-TCGA images in TCIA. We will begin work on making this accessible via an API as well as link the clinical data to the imaging data. This work is in preliminary stages and will be limited to a prototype in the coming year. We will continue to upgrade the API and add new APIs to meet user needs. A publication summarizing our work in APIs and data mashups is under preparation and will be submitted in early 2017.

§ AIM 2

We will continue work on yunpipe (a cloud based imaging pipeline system). In the coming year, we will add support to retrieve data from TCIA. We will also test yunpipe using existing image pipelines. Some of this work will be done as part of an ongoing BIDS WG cooperative project. A publication describing yunpipe and cloud based image pipelines is under preparation and will be submitted in late spring.

During the coming year we will work to refine our automated lung segmentation tool and its integration with the other Galaxy tools that we are developing. We will also compare our automated routines to either manually segmented routines or manually seeded routines. We will also integrate tools for doing automated lung nodule characterization based on our preliminary experience with CNNs.

We are currently summarizing our testing on image features and CNNs and preparing a manuscript for publication: Image features for prediction of lung cancer malignancy with low dose CT scan (LDCT). This manuscript will include the results of our over fitting analysis and utilization of independent test sets from LungX and Dr. Gierada.

Based on the preliminary testing with the LIDC/IDRI dataset of the models, we plan to return to the NLST problem where lesions are not identified and explore the use of Convolutional Neural Networks to learn how to identify patients who developed lung cancer from those who did not using the entire lung as input to the analysis.

§ AIM 3

We will continue our work on statistical methods to develop graph based imaging phenotypes that combine clinical data and imaging features. The first case study on this topic is looking into LungRADS and how it’s precision can be improved. The existing LungRADS relies on clinical features to determine risk and then relies exclusively on

imaging features to recommend a screening protocol. We are exploring alternate scoring criteria, in a retrospective study of NLST subjects that uses imaging features as well as clinical and demographic features, when recommending a screening protocol. The goal is a publication outlining this work, as well as a specific plan to proceed with this work, assuming a successful outcome, in an observational study with patients undergoing lung screening using LDCT.

Over the coming year, we will implement data adapters for additional TCIA datasets. As part of this work, we expect to enhance Eureka to process non-clinical data in order to incorporate imaging features into graphs and support a wider range of studies that compare populations by imaging features in addition to clinical data. This will support our goal of making TCIA datasets broadly accessible to QIN investigators in graph form. We will also advance making TCIA datasets broadly available for secure interactive query in i2b2, which will provide data export that is suitable for a broader range of statistical analysis methods.

§ AIM 4

A short course on the advanced features added to TCIA has been accepted for presented at RSNA 2017. Presentations are planned for the 2017 QIN annual meeting. Several new cross-network collaborations are in the planning stage**.**

LIST OF REFERENCES

1. Clarke LP, Nordstrom RJ, Zhang H, Tandon P, Zhang Y, Redmond G, Farahani K, Kelloff G, Henderson L, Shankar L. The Quantitative Imaging Network: NCI's historical perspective and planned goals. Translational oncology. 2014;7(1):1-4.
2. NCI. Center for Multiple Myeloma Nanotherapy 2016 [cited 2016 5/26/2016]. Available from: http://nano.cancer.gov/action/programs/washup3/.
3. Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, Clingan KL, Duan F, Fagerstrom RM, Gareen IF, Gatsonis CA, Gierada DS, Jain A, Jones GC, Mahon I, Marcus PM, Rathmell JM, Sicks J. Results of the two incidence screenings in the National Lung Screening Trial. N Engl J Med. 2013;369(10):920-31. Epub 2013/09/06. doi: 10.1056/NEJMoa1208962. PubMed PMID: 24004119; PubMed Central PMCID: PMCPMC4307922.
4. Kalpathy-Cramer J, Freymann JB, Kirby JS, Kinahan PE, Prior FW. Quantitative Imaging Network: Data Sharing and Competitive AlgorithmValidation Leveraging The Cancer Imaging Archive. Translational oncology. 2014;7(1):147-52.
5. Armato IIISG, Hadjiiski L, Tourassi GD, Drukker K, Giger ML, Li F, Redmond G, Farahani K, Kirby JS, Clarke LP. Guest Editorial: LUNGx Challenge for computerized lung nodule classification: reflections and lessons learned. JMIOBU. 2015;2(2):020103-. doi: 10.1117/1.JMI.2.2.020103.
6. Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, O'Sullivan F. The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. Tomography: a journal for imaging research. 2016;2(1):56.
7. PCORI. Pragmatic Randomized Trial of Proton vs. Photon Therapy for Patients with Non-Metastatic Breast Cancer Receiving Comprehensive Nodal Radiation: A Radiotherapy Comparative Effectiveness (RADCOMP) Trial 2015 [cited 2016 5/23/2016]. Available from: http://www.pcori.org/research-results/2015/pragmatic-randomized-trial-proton-vs-photon-therapy-patients-non-metastatic.
8. Nature Scientific Data Recommended Data Repositories Nature Publishing Group; 2016 [cited 2016 June 2, 2016]. Available from: http://www.nature.com/sdata/policies/repositories.
9. PLOS ONE Data Availability: PLOS; 2016 [cited 2016 6/2/2016]. Available from: http://journals.plos.org/plosone/s/data-availability - loc-recommended-repositories.
10. Elsevier. Supported Data Repositories 2017 [cited 2017 1/30/17]. Available from: https://http://www.elsevier.com/books-and-journals/enrichments/data-base-linking/supported-data-repositories - TCIA.
11. F1000Research Data Guidelines: F1000 Research Ltd. ; 2016. Available from: http://f1000research.com/for-authors/data-guidelines.
12. Rosenstein BS, Capala J, Efstathiou JA, Hammerbacher J, Kerns S, Ostrer H, Prior FW, Vikram B, Wong J, Xiao Y. How Will Big Data Improve Clinical and Basic Research in Radiation Therapy? International Journal of Radiation Oncology* Biology* Physics. 2015.
13. Bennett W, Matthews J, Bosch W. SU-GG-T-262: Open-Source Tool for Assessing Variability in DICOM Data. Medical Physics. 2010;37(6):3245-.

14. Gierada DS, Politte DG, Zheng J, Schechtman KB, Whiting BR, Smith KE, Crabtree T, Kreisel D, Krupnick AS, Patterson GA. Quantitative Computed Tomography Classification of Lung Nodules: Initial Comparison of 2-and 3-Dimensional Analysis. Journal of computer assisted tomography. 2016;40(4):589-95.
15. Pinsky PF, Gierada DS, Nath PH, Kazerooni E, Amorosa J. National Lung Screening Trial: Variability in Nodule Detection Rates in Chest CT Studies. Radiology. 2013.
16. Paul R, Hawkins SH, Balagurunathan Y, Schabath MB, Gillies RJ, Hall LO, Goldgof DB. Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma. Tomography: a journal for imaging research. 2016;2(4):388.

U01CA190234: Tumor Genotype and Radiomic Phenotype of Lung Cancer

Harvard University, Dan a Farber Cancer Institute, and -Brigham & Womens Hospital

Stephen Yip, Ph.D.; Andriy Fedorov, Ph.D.; Joost van Griethuysen, M.D.; Thibaud Coroller, M.Sc.; Fiona Fennessy, Ph.D.: Ahmed Hosny, M.Sc.; Roman Zeleznik, M.Sc.; Steven Pieper,Ph.D.; John Quackenbush, Ph.D.; Hugo Aerts, Ph.D.

INTRODUCTION

.

Despite the advances in treatment strategies, lung cancer remains the leading cause of cancer related death globally. Early prediction of treatment response and clinical outcomes may help tailor therapy for individual patients. Cancer genomics has demonstrated that the initiation and progression of lung cancer are caused by specific genetic abnormalities, such as mutations in EGFR, KRAS, and ALK and therapies that target these genetic abnormalities have shown great promise. Tumor tissues acquired from biopsies and surgical resection are used for genotyping but these procedures are invasive and are not generally repeated during treatment.

Genetic mutations drive various biological processes in malignant tumors that may manifest in macroscopic phenotypic. Medical imaging can non-invasively and quantitatively assess lung cancer phenotype, and, possibly, the tumor genotype. Radiomics is capable of extracting a large number of quantitative imaging features that can be used to describe tumor phenotypes. Studies by our group and others have suggested that radiomic features are significantly associated with clinical outcomes, treatment response, and genomic mutations. The features have great potential for noninvasive and quantitative tumor characterization, and through integration with other sources of information, personalized therapy.

The central hypothesis of our ongoing research is that radiomic features combined with genetic mutation profiles can improve tumor characterization and predict therapeutic response and clinical outcome. In addition, we are exploring the relationship between radiomic features and genomic abnormalities

DISCUSSION AND RESULTS OF PROGRESS MADE OVER THE PREVIOUS YEAR

In the past year, our research has been divided along the four main aims outlined in our funded project. We are actively building a large database consist of imaging and

mutational data (PROFILE) at DFCI/BWH. Currently, our database has the majority of PROFILE lung patients (~1000) were have been treated at DFCI/BWH since 2012. Investigators are actively engaged in this project and communicate through weekly meetings during which progress is assessed. Diagnostic CT and PET images have been curated by radiologists to define accurate tumor segmentations. Genomic and clinical data were collected and curated by medical and radiation oncologists. Moreover, we are developing a 3D-Slicer-based platform for automatic radiomic feature extraction and analysis. 3D tumor masks have been segmented on all of our curated images for radiomic features extraction.

As mentioned, tumor initiation and progression can be driven by somatic mutations, such as EGFR and KRAS, in NSCLC. These somatic mutations drive various biological processes that are ultimately reflected in tumor phenotypes. Quantitative radiomics non-invasively characterizes tumor phenotypes by using a automatic image characterization algorithms to extract a large number of quantitative features from medical images. An analysis of a first batch of the PROFILE patients has been completed, allowing us to explore correlations between genomic and radiomic features and investigate the value of radiomic features in predicting tumor mutational status.

^{18}F-FDG-PET imaging is used to assess tumor glucose metabolism (Figure 1). Because mutations alter processes that control the cell cycle, one could hypothesize that mutations may each alter metabolism in distinct ways. However, it is unknown whether somatic mutations can be predicted by PET-based radiomic features that describe the tumor metabolic phenotypes. For the first time, we assessed the power of ^{18}F-FDG-PET-based radiomic features to predict somatic mutations in non-small cell lung cancer (NSCLC) patients [1]. We identified 348 NSCLC patients who were profiled for somatic mutations and underwent diagnostic ^{18}F-FDG-PET/CT scans. Of thesem 13% (44/348) and 28% (96/348) were found to harbor an EGFR (EGFR+) or KRAS (KRAS+) mutation, respectively. We used AUD to assess the ability of 21 PET-based radiomic features to predict mutation status. The significance of the AUC was compared to a random guess (AUC=0.5) using the Noether's test. All p-values were corrected for multiple hypothesis testing by controlling the false discovery rate ($FDR_{Wilcoxon}$, $FDR_{Noether}$) with a significance threshold of 10%.

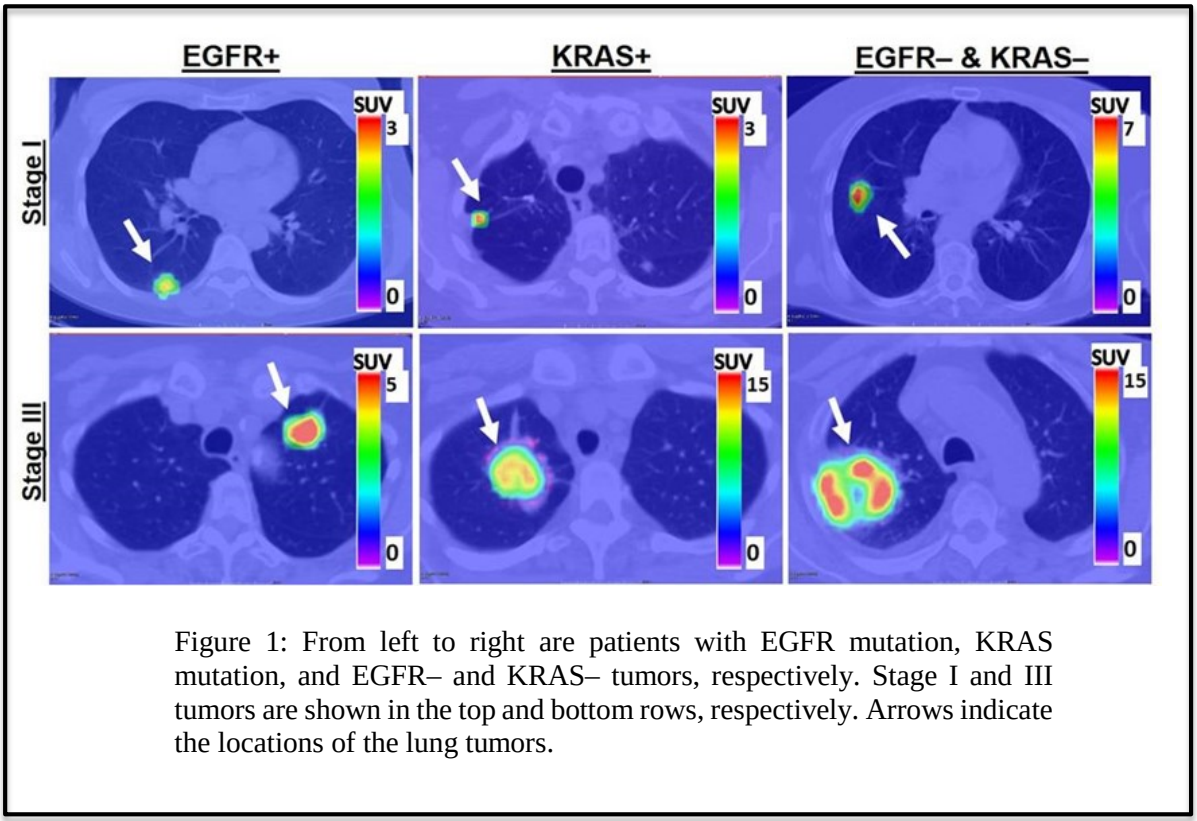


Figure 1: From left to right are patients with EGFR mutation, KRAS mutation, and EGFR– and KRAS– tumors, respectively. Stage I and III tumors are shown in the top and bottom rows, respectively. Arrows indicate the locations of the lung tumors.

Ten radiomic features were significantly predictive of EGFR mutation status (Figure 2). One radiomic feature (normalized inverse difference moment) outperformed all other features in predicting EGFR status (EGFR+ vs EGFR-, AUC=0.67, $FDR_{Noether}$=0.0032), as well as differentiating between KRAS+ and EGFR+ (AUC=0.65, $FDR_{noether}$=0.05). None of the features were associated with or predictive of KRAS mutation status (KRAS+ vs. KRAS-, AUC=0.50-0.54). Our results indicate that EGFR mutations may drive different metabolic tumor phenotypes that are captured in PET images, whereas KRAS mutated tumors do not. This study sheds light on genotype-phenotype interactions, using radiomics to capture and describe the phenotype, and may have potential for developing non-invasive imaging biomarkers for somatic mutations [1].

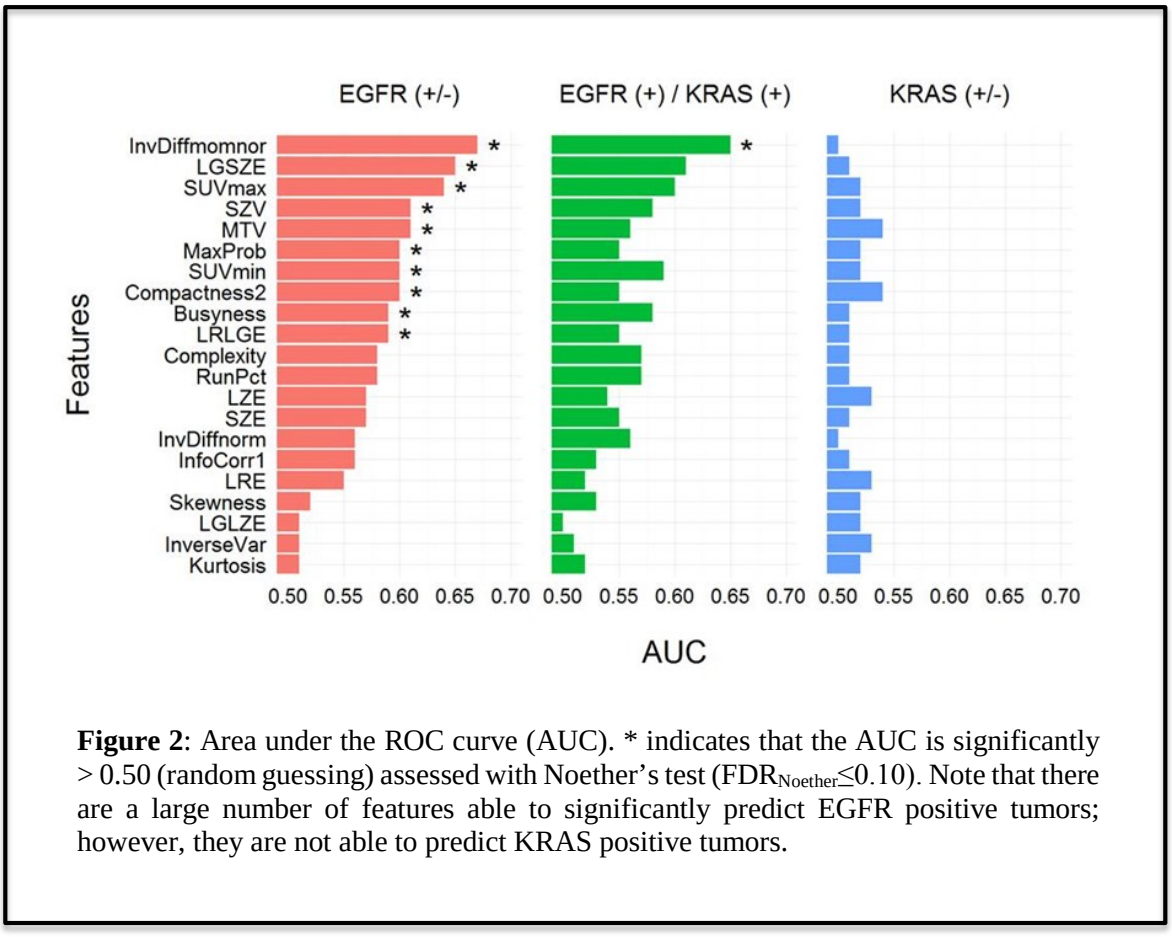


Figure 2: Area under the ROC curve (AUC). * indicates that the AUC is significantly > 0.50 (random guessing) assessed with Noether's test ($FDR_{Noether} \leq 0.10$). Note that there are a large number of features able to significantly predict EGFR positive tumors; however, they are not able to predict KRAS positive tumors.

We recently published two papers focusing on lung cancer pathological response that have potential clinical implications [2, 3]. For patients with non-small cell lung cancer (NSCLC) treated with trimodality therapy (chemoradiation followed by surgery), pathological response is a direct measure of therapeutic response that is assessed at the time of surgery. Pathological response can be used as a surrogate marker to aid clinical decision-making. However, the benefits of adding surgery to chemoradiation for stage IIIA NSCLC remain unclear. Therefore, noninvasive early predictors of pathological response are needed to identify patients most likely to benefit from continuing chemoradiation versus proceeding to surgery. We investigated the value of radiomic data extracted from pretreatment CT images of the primary tumor and lymph nodes in predicting pathological response after neoadjuvant chemoradiation before surgery (**Figure 3**).

After selecting features based on stability (test/retest robustness) and variance (Principal component analysis and correlation) we identified 10 features from the primary tumor and 10 from the lymph nodes. All results were corrected for multiple testing using false discovery rate. We found that tumors that did not respond well to neoadjuvant chemoradiation were more likely to present a rounder shape (spherical disproportionality, AUC = 0.63, p-value = 0.009) and heterogeneous texture (entropy, AUC = 0.61, p-value = 0.03). Additionally, two features extracted from lymph nodes (quantifying homogeneity) were

predictive of non-response (AUC range 0.72–0.75, $p < 0.05$) and performed significantly better than information from the primary tumor site (AUC = 0.62). Multivariate analysis showed (Figure 4) the radiomic features set alone had the best-performing classification (median AUC = 0.68) for identifying complete responders. For non-responder classification, the combination of radiomic and clinical data significantly outperformed all other predictors (median AUC = 0.73). This project has potential to develop noninvasive biomarkers that could capture the total tumor burden and provide important complementary information to aid clinical decision-making.

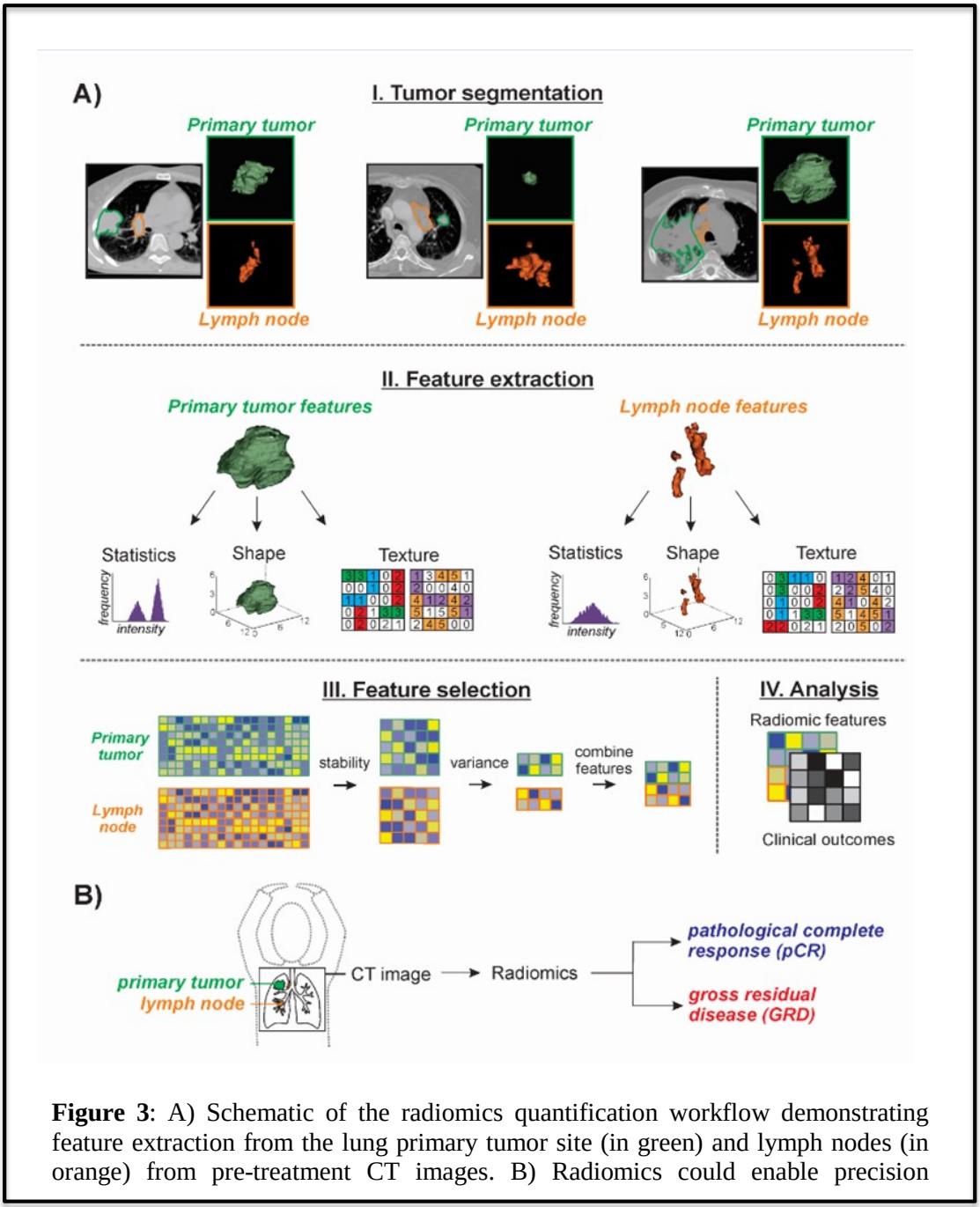


Figure 3: A) Schematic of the radiomics quantification workflow demonstrating feature extraction from the lung primary tumor site (in green) and lymph nodes (in orange) from pre-treatment CT images. B) Radiomics could enable precision

medicine by classifying patients prior to therapy based on how they will respond to chemoradiation.

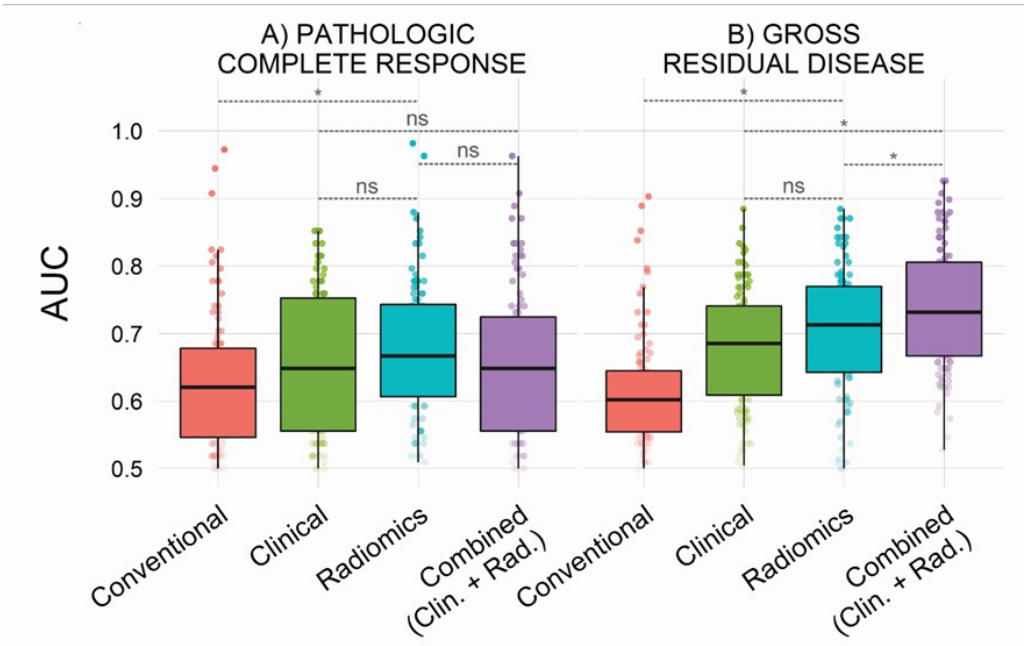


Figure 4: The performances of random forest classification of models for A) pathological complete response (pCR) and B) gross residual disease (GRD). AUCs are reported from nested cross validation analysis. "*" indicates significant difference between the feature models (permutation test, p-value 0.05)

Finally, because batch effects may be important in radiomic studies, we tested the effects of different image acquisition protocols. The majority of radiomics studies have used features extracted from a single image type. However, different imaging methods may be used to assess the tumor phenotype and different aspects of its behavior may be uniquely captured in different types of images, However, differences in collection protocol can introduce differences in radiomic features even within the same imaging modality. For example, in radiation therapy treatment planning, computed tomography (CT) is the primary imaging modality used, but different types of CT images are acquired to provide additional information for the treatment plan. Commonly, treatment plans are designed on static free breathing (FB) helical CT images, however, in cases where organ motion is a concern, such as with lung tumors, four-dimensional (4D) CT image datasets are also acquired and converted to 3D volumes using average intensity projection (AIP). This is the case for early stage non-small cell lung cancer (NSCLC) patients that are treated with stereotactic body radiation therapy (SBRT) (Figure 5). The use of both types of CT scans has contributed to the excellent survival and local control of NSCLC patients treated with SBRT.

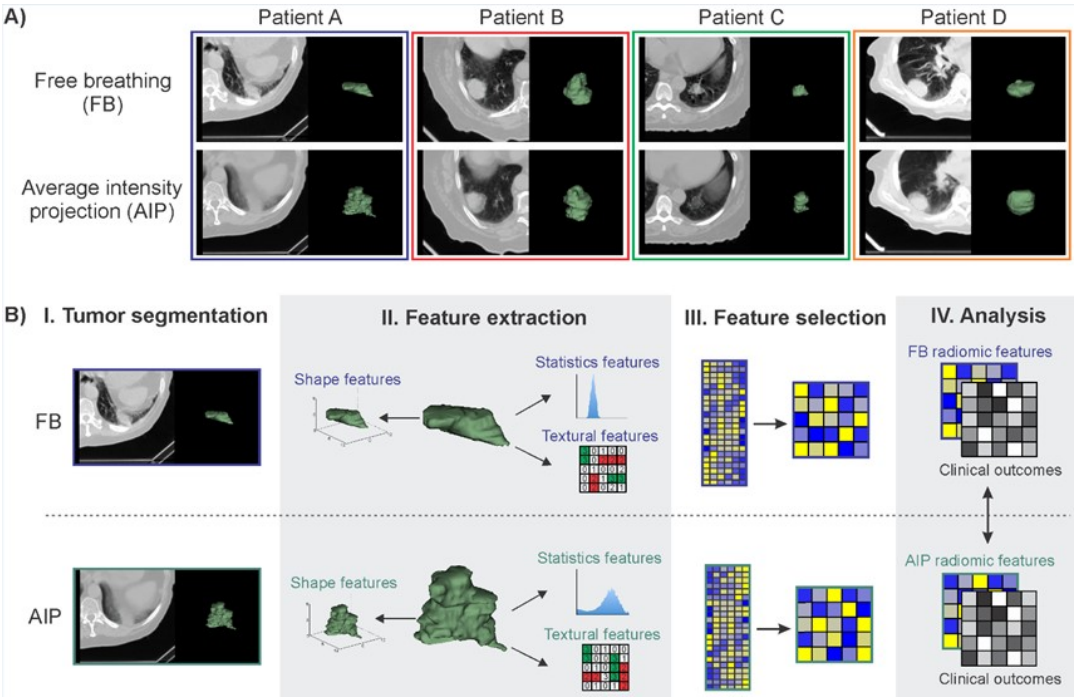


Figure 5.: A) Examples of free breathing (FB) and average intensity projection (AIP) images, demonstrating the observable differences in tumor phenotype between each image type. AIP images were reconstructed from 4D computed tomography (CT) scans. B) Schematic representation of the radiomics workflow for FB and AIP images. I. CT images of the patient are acquired and the tumor is segmented. II. Imaging features (radiomic and conventional features) are extracted from the tumor volume. III. Radiomic features undergo a feature dimension reduction process to generate a low-dimensional feature set based on feature stability and variance. IV. Imaging features are then analyzed with clinical outcomes to evaluate their prognostic power. FB and AIP radiomics features are compared.

Nineteen radiomic features were selected for analysis from FB and AIP images. Notably, thirteen of the ninteen features were different between FB and AIP images. These features were selected based on maintaining the variance in the dataset, and therefore, the difference in feature sets indicates that the images contain different radiomics information and the image type influences the feature values. Only one AIP radiomics feature was associated with loco-regional recurrence (LRR), however, none of the conventional volumetric or radiomics features from FB or AIP images were prognostic for LRR. Furthermore, none of the imaging features were associated with distant metastases (DM), although several features were prognostic for DM.

This highlights the important notion that although a feature may be associated with a clinical outcome, it may not necessarily be prognostic since the properties of the feature

distribution that qualify a feature as associative are not the same as the properties that qualify a feature as prognostic**.** The selected FB and AIP radiomic feature sets had six common radiomic features between both image types and thirteen funique features. None of the FB radiomic features were prognostic of DM, however, seven AIP radiomic features, that described tumor shape and heterogeneity, were prognostic (CI range: 0.638-0.676). Conventional volumetric features from FB images were not prognostic of DM, however, AIP conventional features were (CI range: 0.643-0.658). AIP radiomic multivariate models (median CI = 0.667) outperformed all other models (median CI range: 0.601-0.630) in predicting DM. None of the imaging features were prognostic of LRR (Figure 6).

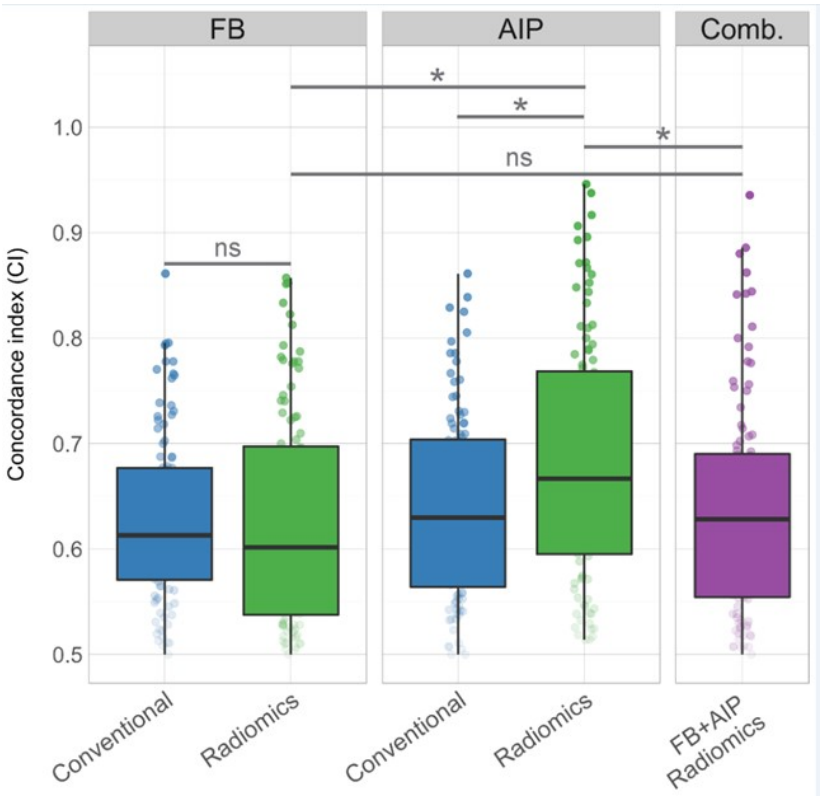


Figure 6: Performance of each multivariate model in predicting distant metastasis. Concordance indices are reported for the FB and AIP conventional and radiomic models, and a combined FB+AIP radiomics model, comparing the performance of each of model and image type. Cross validation was performed (80% training, 20% validation) to generate 100 models for each model type. Comb. Indicates the combined FB and AIP radiomics model. *p-value < 0.05; “ns” indicates not significant (p-value > 0.05).

COLLABORATIONS WITHIN THE QIN AND ITCR NETWORKS

§Lung Nodule Segmentation algorithms.

Our group also participated the regular teleconference of the PET/CT subgroup and in the nodule segmentation challenge 2016. In this challenge, we used a publically available semiautomatic segmentation algorithm that is implemented in the 3D Slicer 4.5 Chest Imaging Platform (CIP). A seed point needs to be placed within the nodule region to initialize segmentation. The nodule segmentation is then generated based on a level set formulation to propagate according to a Geodesic Active Contour functional. The robustness of the CIP-based and radiologist manually-defined segmentation were assessed using the region of uncertainty (δ) and the Dice similarity index (dsi). The median computational time of the CIP segmentation on a personal computer was only 10s. CIP segmentations were significantly more robust than manual segmentations (Figure 7). We demonstrated that CIP segmentation can potentially reduce the physician workload and inter-observer variability owing to its computational efficiency and superior stability compared to manual segmentation (Yip et al submitted 2016b).

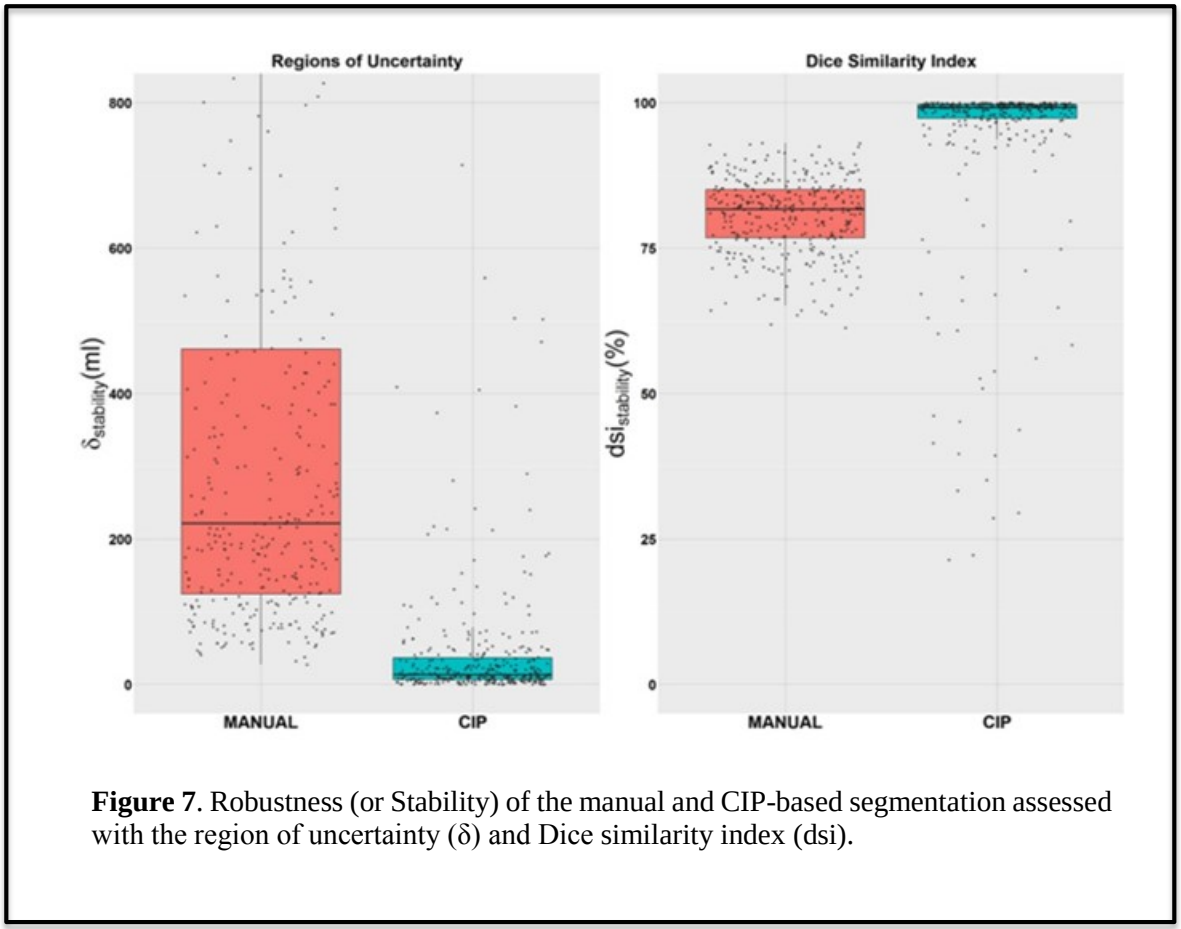


Figure 7. Robustness (or Stability) of the manual and CIP-based segmentation assessed with the region of uncertainty (δ) and Dice similarity index (dsi).

§ **Python Radiomics Platform:**

A large part of radiomics analyses rely on engineered hard-coded features. However, there is a lack of standardization both of the feature definitions and the preprocessing of the image, which has been shown to have a substantial impact on the performance of the extracted data. Many studies use in-house developed software, which is not always shared, making reproduction of results more difficult. By an ITCR funded effort, we aimed to develop a supported open source comprehensive radiomics platform to simplify and clarify feature extraction from image to output. This resulted in the development of *PyRadiomics*, written in easy-to-read Python code with additional C extensions for performance, which is freely available on github. It is freely distributed and requires no expensive licenses to run the code, as the Python language itself is freely distributed open source.The extraction of radiomics features by *PyRadiomics* comprises of four main steps: 1) Loading and preprocessing of the image and segmentation maps, 2) Application of enabled filters, 3) Calculation of features using the different feature classes and 4) Returning results. The dynamic and modular design simplifies the addition or removal of features. In interactive mode, *PyRadiomics* can be incorporated in larger image analysis pipelines, while the CLI scripts combined with the parameter file facilitate feature extraction without requiring extensive programming skills of the user. Furthermore, generated results can be stored in a CSV-format, enabling easy import directly into many statistical packages for analysis, including R and SPSS. For more information see: www.radiomics.io and www.github.com/radiomics

PLANS FOR NEXT YEAR

During the next year, we plan to continue our successful research program, pursuing both the refinement of existing methods, the development of new analytical approaches, and the expansion our collection of integrated data sets.

First, we will curate a second batch of Profile datasets by collecting and analyzing, imaging, genomic, and clinical outcome data. The analysis of the first batch of patients is finalized, and we aim to publish this analysis this year.

Second, we will incorporate additional radiomic classification methods, including deep learning, into our machine-learning framework. We will refine these methods over time by integrating additional sources of data as they become available, benchmarking their performance against current methods.

Third, we will develop further validate developed radiomic, genomic, and integrated biomarkers. For this purpose we have access to multiple independent datasets that we can use to develop and test signatures. Several machine-learning techniques have been evaluated and will be applied to build predictors for mutational status as well as clinically relevant outcomes, such as overall survival, local control, and distant metastasis.

Fourth, we will make all computational resources that we develop freely available as open source tools. We will share the radiomic system implemented in the open source software suite 3D-Slicer and instantiate our data analysis methods in freely-available Bioconductor

packages. We will educate and help other investigators with applying the radiomic system to their own data using the "project week" of 3D-Slicer, which is an open forum held twice each year.

While the focus of this project is on NSCLC, we are using this disease as a model. Our ultimate goal is to develop computational methods that can be more broadly applied in cancer research and clinical applications. Therefore, we aim to make all software as independent as possible for disease site, imaging modality, and genomic data type. Lastly, we are planning to design and participate in more QIN challenges for the coming year.

QIN PUBLICATIONS PUBLISHED AND SUBMITTED

1. Yip SS, Kim J, Coroller T, Parmar C, Rios Velazquez E, Huynh E, et al. Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer. J Nucl Med. 2016; doi:10.2967/jnumed.116.181826
2. Coroller TP, Agrawal V, Huynh E, Narayan V, Lee SW, Mak RH, et al. Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC. J Thorac Oncol. 2016; doi:10.1016/j.jtho.2016.11.2226
3. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. Radiother Oncol. 2016;119: 480–486.
4. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2016; doi:10.1038/nrclinonc.2016.162
5. Aerts HJWL, Grossmann P, Tan Y, Oxnard GG, Rizvi N, Schwartz LH, et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. Sci Rep. 2016;6: 33860.
6. Aerts HJWL. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. JAMA Oncol. 2016;2: 1636–1642.
7. Grossmann P, Gutman DA, Dunn WD Jr, Holder CA, Aerts HJWL. Imaging-genomics reveals driving pathways of MRI derived volumetric tumor phenotype features in Glioblastoma. BMC Cancer. 2016;16: 611.
8. Huynh E, Coroller TP, Narayan V, Agrawal V, Hou Y, Romano J, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. Radiother Oncol. 2016;120: 258–266.
9. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. Phys Med Biol. 2016;61: R150–66.
10. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol. 2015;114: 345–350.
11. Huynh E, Coroller TP, Narayan V, Agrawal V, Romano J, Franco I, et al. Associations of Radiomic Data Extracted from Static and Respiratory-Gated CT Scans with Disease

Recurrence in Lung Cancer Patients Treated with SBRT. PLoS One. 2017;12: e0169172.

12. Huynh E, Coroller T, Narayan V, Agrawal V, Romano J, Franco I, et al. SU-F-R-52: A Comparison of the Performance of Radiomic Features From Free Breathing and 4DCT Scans in Predicting Disease Recurrence in Lung Cancer SBRT Patients. Med Phys. 2016;43: 3385–3385.

13. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep. 2015;5: 13087.

14. Gutman DA, Dunn WD Jr, Grossmann P, Cooper LAD, Holder CA, Ligon KL, et al. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. Neuroradiology. 2015;57: 1227–1237.

15. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. Front Oncol. 2016;6: 71.

16 Patrick Grossmann, Olya Grove, Nehme El-Hachem, Emmanuel Rios-Velazquez, Chintan Parmar, Ralph T.H. Leijenaar, Benjamin Haibe-Kains, Philippe Lambin, Robert J. Gillies, Hugo J.W.L. Aerts, “Identification of Molecular Phenotypes in Lung Cancer by Integrating Radiomics and Genomics”, submitted

17. Stephen SF Yip, Chintan Parmar, Daniel Blezek, Raul San Jose Estepar, Steve Pieper, John Kim and Hugo J.W.L. Aerts, Robust Lung Nodule Segmentation Using the 3D Slicer Chest Imaging Platform (submitted)

18. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, Ye Z, Makrigiorgos M, Fennessy F, Mak RH, Gillies R, Quackenbush J, Aerts HJ, Somatic mutations drive distinct imaging phenotypes in lung cancer, submited

1U01CA190254-01 ECOG-ACRIN-Based QIN Resource for Advancing Quantitative Cancer Imaging in Clinical Trials

ECOG-ACRIN

MITCHELL D. SCHNALL, MD, PHD
DAVID MANKOFF, MD, PHD
PAUL E. KINAHAN, PHD
MARK ROSEN, MD, PHD

INTRODUCTION

The goal of this project is to accelerate the development and deployment of quantitative imaging methods that improve the effectiveness and efficiency of clinical trials by using the combined resources of the NCI-sponsored cooperative group ECOG-ACRIN and the Quantitative Imaging Network (QIN). To achieve this goal, and in accord with NOT-CA-13-011 (PAR-11-150), this project will create QIN-wide research resources.

Aim 1: Optimize the efficiency of the qualification and QA/QC processes to reduce costs and improve the quantitative accuracy of multi-center trials using advanced imaging.

Aim 2: Develop the ECOG-ACRIN QIN Resource to support retrospective testing for single- or multi-site QIN projects that seek to develop effective and efficient metrics and analysis methods for trials using advanced imaging.

Aim 3: Develop the ECOG-ACRIN QIN Resource to support prospective testing of novel quantitative imaging methods developed in the QIN.

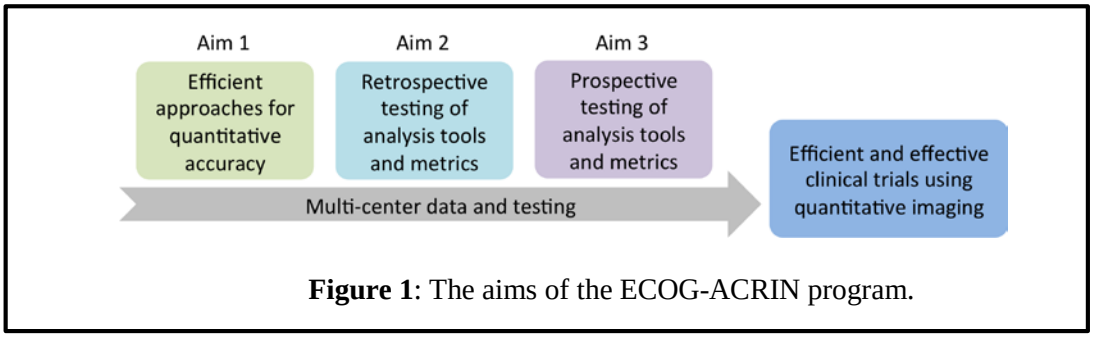


Figure 1: The aims of the ECOG-ACRIN program.

As part of the QIN, ECOG-ACRIN QIN Resource will act as a scientific site for evaluating methodologies and metrics for quality assurance of imaging and associated data, focusing on understanding the costs of efficient and effective site qualifications that result in high-quality imaging studies and the metrics required to appropriately define the number of participants required for adequate analysis.

This project will evaluate quality control at participating QIN laboratories, comparing practices currently applied by the NCI (e.g., CQIE) and ACR Imaging Core Laboratory (Aim 1) at each participating QIN site. The ECOG-ACRIN QIN Resource will further act as a resource development platform (Aims 2 and 3). ECOG-ACRIN, in league with the Brown Statistical Center, proposes to develop datasets for method testing and validation using completed ACRIN research for assessment of QIN metrics and validation purposes (Aim 2). In the Resource, outcomes and progression data will be made available for correlation with computational findings.

Finally, the ECOG-ACRIN QIN Resource will bring expertise across QIN Working Group platforms—in PET, MRI, CT, imaging statistical design, and informatics—to clinical trials by integrating quality assurance and QIN quantitative tools into prospective National Clinical Trial Network research (Aim 3). The ECOG-ACRIN QIN Resource PIs stand at the front lines within the ECOG-ACRIN clinical trials development structure as leaders of the Experimental Imaging Science Committee (EISC) and Biomarker Group and Imaging Science Advisory Committees (ISAC), which review imaging studies prior to submission to NCI for consideration, and thus open the door to identifying appropriate opportunities for prospective evaluation of QIN laboratory projects (Aim 3).

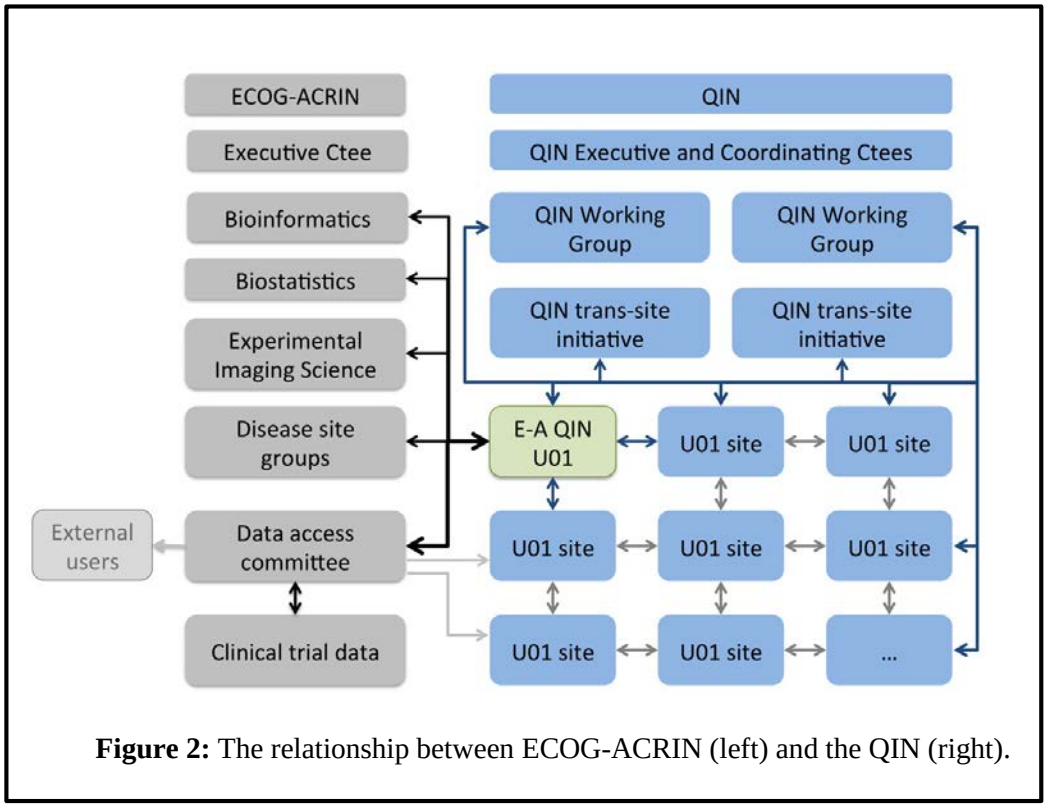


Figure 2: The relationship between ECOG-ACRIN (left) and the QIN (right).

PROGRESS OVER THE PREVIOUS YEAR

§ Major Activities:

The scientific team has implemented monthly conference calls of the project team and weekly calls of the project team leadership in order to manage the activities funded through the grant and to ensure consistent progress with respect to all of the goals. This mechanism has proven to be an effective way to aggregate the unique expertise of the PIs and stakeholders who are associated with institutions across the country. Details associated with progress achieved within each of the 3 aims follows.

Aim 1: Manuscript entitled Performance Observations of Scanner Qualification of NCI-designated Cancer Centers: Results from the Centers of Quantitative Imaging Excellence (CQIE) Program has been published by Academic Radiology, disseminating findings and standards for quantitative imaging in clinical research trials. The CQIE database has been used to identify QIN funded sites who have successfully met CQIE qualification standards in the past and this data is being used to create a site profile in the Qualification Utility for Imaging Clinical Trials (QUIC) of qualified QIN sites. This is intended to be dynamic as QIN sites change we will update the QUIC dataset.

Aim 2: Established a prioritized list of completed trials with datasets that QIN sites felt were best positioned to support QIN development needs. Collaborated with NCI to develop standard guidelines and workflows for transferring the datasets to TCIA, resulting in the transfer of 4 trial datasets (with others in progress). In parallel, EA QIN leveraged the ACR's commitment to development of the Data Access and Retrieval Tool (DART) and developed a workflow involving anonymization methods which ensure compliance with safe harbor regulations and HIPAA standards. In addition to the datasets transferred to TCIA, fully anonymized datasets for these trials and for other ACRIN legacy trials will be made accessible to QIN and other researchers. The DART environment will also offer the capacity to analyze datasets in the cloud and to process data through select applications hosted in DART, thereby enabling image processing to happen at the host and reducing the burden associated with transfers of large datasets.

Aim 3: Working with NCI QIN leadership, the scientific team successfully carried out a 1-day planning meeting on 12/13/16 that brought together thought leaders from the NCTN together and leaders from QIN and other interested stakeholders. Leading oncologists from different NCTN groups shared their perspectives on the value of quantitative imaging to therapy, while QIN representatives share their vision and perspective on how quantitative imaging may be able to benefit oncologists and provided a list of QI tools currently offered by QIN cites. Breakout sessions focused on 4 areas (response assessment, quantitative imaging biomarkers, informatics and precision medicine, and image/data curation and archiving) provided further discussion and yield of a number of ideas for the use of QIN tools in NCTN trials, and expanded use of NCTN image and clinical outcome datasets. We identified several immediate opportunities to incorporate prospective testing of quantitative imaging tools in national level clinical trials that will be further developed.

COLLABORATIONS WITHIN THE NETWORK

We successfully navigated numerous logistical issues related to data transfer and prepared 4 datasets to be transferred to the TCIA. We completed building the infrastructure for sharing imaging datasets via the DART, including implementing safe-harbor method of anonymization and de-identification to ensure compliance with HIPAA standards. A total of 4 datasets have either been transferred to TCIA or are in the process of being transferred and available to the QIN membership.

The scientific team hosted a 1-day planning meeting that brought together thought leaders from the NCTN together with leaders from QIN and related groups for some roundtable discussions on what oncologists need for quantitative imaging with their trials and what imagers have to offer. The outcome of the meeting is to generate 4 – 6 ideas on how to develop prospective testing of quantitative imaging tools in national level clinical trials.

PLANS FOR NEXT YEAR

Our plans for the coming year are as follows:

Aim 1: We will leverage the information gained from Aim 1 to inform Aim 3 and provide centers with the capability to participate in a quantitative study. This step will include gathering information from previous ECOG-ACRIN studies with phantom use and development (Ex: ACRIN 6701).

Aim 2: The EA QIN Resource Center will continue to interact with QIN leadership to establish prioritized datasets to be made available to QIN researchers, exploit the two alternative methods for making these datasets accessible (TCIA and DART), and develop optimal processes for accessing the data which reflect the intended use (multi-institutional trial process, single center test, or challenge grant process). Based on the needs of the QIN sites, EA QIN Resource Center will research additional datasets which would meet the needs of QIN sites to determine if there are opportunities to create new datasets that will better serve current QIN development needs.

Aim 3: The EA QIN Resource Center will produce a report summarizing the results and future directions generated by the QIN-NCTN Planning Meeting held December 13, 2016. Trials already approved or activated which are identified for hosting a QIN tool or method will be pursued through the clinical trial leadership of the LPO with oversight of that trial. We also anticipate that we will develop improved processes to engage QIN researchers in the LPO meetings, with the goal of engagement in early clinical trial development activities. We also anticipate that the Planning Meeting will identify a need to consider the possible development of a clinical consulting board to support direction and development.

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

The QIN team submitted two manuscripts for publication; one of the manuscripts has been published by Academic Radiology, disseminating findings and standards for quantitative imaging in clinical research trials (Performance Observations of Scanner Qualification of NCI-designated Cancer Centers: Results from the Centers of Quantitative Imaging Excellence (CQIE) Program.

LIST OF REFERRENCES

1. Atuegwu NC, Gore JC, Yankeelov TE. The integration of quantitative multimodality imaging data into mathematical models of tumors. *Phys Med Biol*. 2010;55:2429–2449. PMID: 20371913
2. Meyer CR, Armato SG, Fenimore CP, et al. Quantitative imaging to assess tumor response to therapy: common themes of measurement, truth data, and error sources. *Transl Oncol*. 2009;2:198–210. PMID:19956379
3. *Quantitative Imaging For Evaluation of Response to Cancer Therapies. The Quantitative Imaging Network (QIN).* US DHHS / NIH National Cancer Institute, Cancer Imaging Program. Second Annual Report, January 2013. (326 pages).
4. Maitland ML, Schilsky RL. Clinical trials in the era of personalized oncology. *CA Cancer J Clin*. 2011;61:365–381. doi: 10.3322/caac.20135. Epub 2011 Oct 27. PMID: 22034206
5. O'Dwyer PJ, Eckhardt SG, Haller DG, et al. Priorities in colorectal cancer research: recommendations from the Gastrointestinal Scientific Leadership Council of the Coalition of Cancer Cooperative Groups. *J Clin Oncol*. 2007;25:2313–2321. PMID: 17538178
6. McShane LM, Hayes DF. Publication of tumor marker research results: the necessity for complete and transparent reporting. *J Clin Oncol.* 2012;30(34):4223–4232. PMID: 23071235
7. Henry NL, Hayes DF. Cancer biomarkers. *Mol Oncol.* 2012;6(2):140–146. doi: 10.1016/j.molonc.2012.01.010. Epub 2012 Feb 6. PMID: 22356776
8. Buckler AJ, Aerts HJWL, Bendriem B, et al. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. *Radiology*. 2011;259(3):875–884. PMID: 21325035
9. Kumar V, Nath K, Berman CG, et al. Variance of SUVs for FDG-PET/CT is greater in clinical practice than under ideal study settings. *Clin Nucl Med*. 2013;38(3):175–182. PMID: 23354032
10. Velasquez LM, Boellaard R, Kollia G, et al. Repeatability of ^{18}F-FDG PET in a multicenter phase i study of patients with advanced gastrointestinal malignancies. *J Nucl Med*. 2009;50(10):1646–1654. doi: 10.2967/jnumed.109.063347. Epub 2009 Sep 16. PMID: 19759105
11. Ng CS, Raunig DL, Jackson EF, et al. Reproducibility of perfusion parameters in dynamic contrast-enhanced MRI of lung and liver tumors: effect on estimates of patient sample size in clinical trials and on individual patient responses. *AJR.* 2010;194:W134–W140. doi: 10.2214/AJR.09.3116. PMID: 20093564

12. Doot RK, Kurland BF, Kinahan PE, Mankoff DA. Design considerations for using PET as a response measure in single site and multicenter clinical trials. *Acad Radiol*. 2012;19(2):184–190. PMID: 22104290
13. Kurland BF, Doot RK, Linden HM, Mankoff DA, Kinahan PE. Multicenter trials using FDG PET to predict chemotherapy response: Effects of differential measurement error and bias on power calculations for unselected and enrichment designs. Clinical Trials. 2013;10:886–895. PMID: 24169628
14. Jackson EF, Barboriak DP, Rosen MA, et al. QIBA Perfusion, Diffusion, and Flow MRI Technical Committee: *Current Status, Special Poster Exhibit - RSNA 2012 - Quantitative Imaging Biomarkers Alliance (QIBA)*. 2012.
15. *Site Imaging Manual ACRIN 6701: Repeatability Assessment of Quantitative DCE-MRI and DWI: A Multicenter Study of Functional Imaging Standardization in the Prostate*. Available online at: www.acrin.org/Portals/0/Protocols/6701/ImagingMaterials/ACRIN%206701%20SIM%2010092012%20Final.pdf. Accessed February 3, 2014.
16. *NCI-CQIE Qualification Materials*. Available online at: http://www.acrin.org/CORELABS/NCICQIEQUALIFICATIONPROGRAM/SITEQUALIFICATIONMATERIALS.aspx. Accessed February 3, 2014.
17. Mukhtar RA, Yau C, Rosen M, Tandon VJ; The I-SPY 1 TRIAL and ACRIN 6657 Investigators, Hylton N, Esserman LJ. Clinically meaningful tumor reduction rates vary by prechemotherapy MRI phenotype and tumor subtype in the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Ann Clin Oncol*. 2013;20(12):3823–3830. doi: 10.1245/s10434-013-3038-y. Epub 2013 Jun 19. PMID: 23780381
18. Machtay M, Duan F, Siegel BA, et al. Prediction of survival by [^{18}F]fluorodeoxyglucose positron emission tomography in patients with locally advanced non-small-cell lung cancer undergoing definitive chemoradiation therapy: results of the ACRIN 6668/RTOG 0235 trial. *J Clin Oncol*. 2013;31(30):3823–3830. doi: 10.1200/JCO.2012.47.5947. Epub 2013 Sep 16. PMID: 24043740
19. Boxerman JL, Zhang Z, Safriel Y, et al. Early post-bevacizumab progression on contrast-enhanced MRI as a prognostic marker for overall survival in recurrent glioblastoma: results from the ACRIN 6677/RTOG 0625 Central Reader Study. *Neuro Oncol*. 2013 Jul;15(7):945–954. doi: 10.1093/neuonc/not049. Epub 2013 Jun 19. PMID: 23788270
20. Targeted for publication in 2014.
21. Yu EY, Duan F, Muzi M, Gorelick J, et al. ^{18}F-fluoride PET response to dasatinib in castration-resistant prostate cancer bone metastases correlates with progression-free survival: Preliminary results from ACRIN 6687. American Society of Clinical Oncology (ASCO). Poster/Abstract. 2013 Annual Meeting.
22. Yu EY, Duan F, Muzi M, et al. Effects of dasatinib on prostate cancer bone metastases and normal bone measured by ^{18}F-fluoride PET: Preliminary results from ACRIN 6687. Society of Nuclear Medicine and Molecular Imaging (SNMMI). Poster/Abstract. 2013 Annual Meeting.
23. Still accruing in 2014.
24. Clark KW, Gierada DS, Marquez G. Collecting 48,000 CT exams for the Lung Screening Study of the National Lung Screening Trial. *J Digit Imaging*. 2009;22(6):667–680. PMID: 18777192
25. Pepe, MS. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, NY: Oxford University Press.

26. Zhou Z, Obuchowski N, McClish D. (2011). *Statistical Methods in Diagnostic Medicine*. 2nd Ed. New York: Wiley.
27. Chakraborty DP and Berbaum KS. Observer studies involving detection and localization: Modeling, analysis and validation. *Medical Physics*. 2004;31(8): 2313–2330. PMID: 15377098
28. Edwards DC, Kupinski MA, Metz CE, Nishikawa RM. Maximum likelihood fitting of FROC curves under an initial-detection-and-candidatEAnalysis model. *Medical Physics*. *2002;*29(12):2861–2870. PMID: 12512721
29. Bandos A, Rockette H, Song T, Gur D. Area under the free-response ROC curve (FROC) and a related summary index. *Biometrics*. 2009;65:247–256. PMID:18479482
30. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387. PMID: 8668867
31. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105–1117. doi: 10.1002/sim.4154. Epub 2011 Jan 13.PMID: 21484848
32. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56(2):337–344. PMID: 10877287
33. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005;61(1):92–105. PMID: 15737082
34. Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics*. 2000;56(2):345–351. PMID: 10877288
35. Pencina M, D'Agostino Sr, R, D'Agostino Jr, R, Vasan R. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med*. 2008 ;27(2):157–-172; discussion 207–212. PMID: 17569110
36. Steyerberg E, Vickers A, Cook N, et al. Assessing the performance of prediction models a framework for traditional and novel measures. *Epidemiology*. 2010;21:128–138. doi: 10.1097/EDE.0b013e3181c30fb2. PMID: 20010215

QUALIFICATION-RELATED PAPERS FROM ACR IMAGING CORE LABORATORY

Scheuermann J, Opanowski A, Maffei J, et al. Qualification of NCI-Designated Comprehensive Cancer Centers for Quantitative PET/CT Imaging in Clinical Trials. Presented at SNMMI 2013.

Scheuermann JS, Saffer JR, Karp JS, et al. Qualification of PET scanners for use in multicenter cancer clinical trials: the American College of Radiology Imaging Network experience. *J Nucl Med*. 2009 Jul;50(7):1187–1193. doi: 10.2967/jnumed.108.057455. Epub 2009 Jun 12. PMID: 19525463

Madsen EL, Berg WA, Mendelson EB, Frank GR. Investigators for ACRIN Protocol 6666. Anthropomorphic breast phantoms for qualification of Investigators for ACRIN Protocol 6666. Radiology. 2006;239(3):869–874. Epub 2006 Apr 26. PMID: 16641345

BIBLIOGRAPHY OF INTEREST

1. Wu W, Shi Q, Sargent DJ. Statistical considerations for the next generation of clinical trials. *Semin Oncol*. 2011;38:598–604.
2. Sledge GW, Jr. The evolution of targeted biologic therapies. *Breast*. 2007;16 Suppl 2:S13.
1. Sledge GW, Jr, Jotwani AC, Mina L. Targeted therapies in early stage breast cancer: achievements and promises. *Surg Oncol Clin N Am*. 2010;19:669–679.
3. Stone AM, Bushnell W, Denne J, et al. Research outcomes and recommendations for the assessment of progression in cancer clinical trials from a PhRMA working group. *Eur J Cancer*. 2011;47:1763–1771.
4. Zia MI, Siu LL, Pond GR, et al. Comparison of outcomes of phase II studies and subsequent randomized control studies using identical chemotherapeutic regimens. *J Clin Oncol*. 2005;23:6982–6991.
5. Mullard A. 2010 FDA drug approvals. *Nat Rev Drug Discov*. 2011;10:825.
6. Arrowsmith J. Trial watch: Phase II failures: 2008–2010. *Nat Rev Drug Discov*. 2011;10:328–329.
7. DiMasi JA, Feldman L, Seckler A, et al. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther*. 2010;87:272–277.
8. Evelhoch J, Garwood M, Vigneron D, et al. Expanding the use of magnetic resonance in the assessment of tumor response to therapy: workshop report. *Cancer Res*. 2005;65:704–714.
9. Mankoff DA. Imaging studies in anticancer drug development, in Hidalo H, Eckhardt SG, Garrett-Meyer E, et al (eds): *Principles of anticancer drug development*. New York, Springer. 2011:275–304.
10. Mankoff DA, Krohn KA. PET imaging of response and resistance to cancer therapy, in Teicher B (ed): *Drug resistance in cancer*. Totowa, NJ, Humana Press. 2006:105–122.
11. Weber WA. Positron emission tomography as an imaging biomarker. *J Clin Oncol*. 2006;24:3282–3292.
12. Mankoff DA. Molecular imaging to select cancer therapy and evaluate treatment response. *Q J Nucl Med Mol Imaging*. 2009;53:181–192.
13. Meyer CR, Armato SG, Fenimore CP, et al. Quantitative imaging to assess tumor response to therapy: common themes of measurement, truth data, and error sources. *Transl Oncol*. 2009;2:198–210.
14. O'Connor JP, Jackson A, Parker GJ, et al. Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. *Nat Rev Clin Oncol*. 2012:167–177.

U01CA195564: Quantitative Image Analysis for Assessing Response to Breast Cancer Therapy

The University of Chicago

Maryellen L. Giger, Ph.D.

INTRODUCTION

In this second annual report on our research, we note that the goal of our research is to develop quantitative image-based surrogate markers of breast cancer tumors for use in predicting response to therapy and ultimately aiding in patient management. There is a large variation in the clinical presentation of breast cancer in women, and it has been shown that in many instances, biological characteristics, i.e., features, of the primary tumor correlate with outcome. Methods to assess such biological features for the prediction of outcome, however, may be invasive, expensive or not widely available. Our hypothesis is that MRI-based features obtained through quantitative image analysis will prove useful as non-invasive biomarkers for the assessment of, and prediction of, the response of breast cancer to neoadjuvant therapy. We propose to validate such image-based biomarkers using magnetic resonance (MR) images of breast tumors from the ACRIN 6657 clinical trial, which includes pathological response data. Specifically, (1) We will investigate the relationship of breast cancer therapy outcome and MR image-based tumor characteristics (features), and changes in these features over time, using a University of Chicago database and the ACRIN 6657 I-SPY clinical trial dataset of breast cancer tumors from patients who have undergone neoadjuvant treatment, (2) We will develop and evaluate the MRI-derived 'signatures' of breast cancer tumors for the prediction of, and assessment of, response to therapy using the ACRIN 6657 dataset, and (3) We will conduct preliminary, initial stratification and association of the MRI features with cancer subtype and other clinical/histopathological data from the ACRIN dataset. We will build on our 25-year history of taking innovation to the clinical setting by extending our prior development, validation, and translation of quantitative image analysis methods for computer-aided diagnosis to the post-diagnosis, predictive component in order to assess response to neoadjuvant therapy. Our research addresses the development and validation of algorithms using the existing ACRIN 6657 dataset with the goal of "improving the ability to measure the response of targeted tumors to therapy quantitatively". Our proposed research is aligned with the QIN U01 PAR-11-150 goals of including robustness investigations and multi-site trial data (UChicago and ACRIN). Through this QIN grant, our participation in the QIN community will yield deliverables including an open-platform system that will provide tools for linking segmentation/feature extraction/classification, for comparing performance metrics across acquisition and/or analysis systems, and for discovery through dimension reduction techniques. Our research will yield a set of validated lesion signatures that will serve as quantitative tools for use in clinical studies/trials to predict and/or assess tumor response. Given that other studies/trials may use different treatments, we will make available to the QIN community our tools for training, testing, and presenting the quantitative signatures so that predictive signatures for a range of treatments can be determined.

DISCUSSION OF PROGRESS DURING PAST YEAR

§ Relationship of breast cancer therapy outcome and MR image-based tumor characteristics (features), and *changes* in these features over time from patients who have undergone neoadjuvant treatment.

We modified our current quantitative MRI analysis software to automatically and objectively calculate pre-, during-, and post-treatment breast cancer tumor characteristics (features) including volumetric, morphological, textural, and kinetic features. These image-based features are based on our investigations over the past 30 years in developing and translating CAD/quantitative image analysis methods and we have published extensively on the use of multi-modality features for diagnosis and prognosis, but not yet treatment response or risk of recurrence (Refs. 14,16,17,20,21,23,24,35,41,42-58, 107, 128).

During the summer, we participated in the QIN BMMR Challenge, which related MRI-based features (phenotypes) with "risk of recurrence" using a UCSF dataset for training and the I-SPY 1 dataset of 162 cases for testing. We used our automatic, computerized lesion-segmentation algorithms and lesion feature-extraction algorithms on the breast MRIs. Here a c-statistics, which was used in Hylton et al. (REF) along with race and receptor status in the model, served as the index of performance. Of the four teams participating, we were the only one to use automatic lesion segmentation and also we did not use any of the given ACRIN data in our model in order to push the usefulness of quantitative radiomics in assessing risk of recurrence. Participating in the Challenge yielded many "lessons learned", which will be described in the group's future publication. One of our features – one that automatically assessed the tumor's most-enhancing volume did well in the I-SPY 1 prediction model. Because of the varied differences between the training and testing datasets, robust merged models were difficult to train.

§ Relationship of MRI phenotypes to genomics

Using the collected de-identified datasets of invasive breast carcinomas from The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA), cancer research resources supported by the National Cancer Institute (NCI) of the U. S. National Institutes of Health (Ref. 3, 4), the TCGA Breast Phenotype Group (Link in Ref. 5) investigated relationships between computer-extracted quantitative radiomic MRI lesion features and various clinical, molecular, and genomics markers of prognosis and risk of recurrence, including gene expression profiles. At the time of analysis, 91 biopsy-proven invasive breast cancers from the TCGA had DCE-MR images. On these cases, we assessed the predictive ability of the quantitative radiomic MRI features relative to four tasks: (i) pathologic stage, (ii) cancer subtypes, (iii) risk of cancer recurrence, and (iv) genomics.

Investigators have developed multi-gene assays with which to relate breast cancer expression profiles to risk of cancer recurrence, including the 21-gene Oncotype DC assay, the 50-gene PAM50 assay, and the 70-gene MammaPrint microarray assay. To investigate the relationships between quantitative MRI radiomic features and risk of breast cancer

recurrence, we conducted association studies with research versions of these multi-gene assays (Ref. 8). Multiple linear regression analyses demonstrated significant associations between the MRI radiomics signatures (incorporating tumor size and enhancement heterogeneity) and the multi-gene assay recurrence scores. Use of radiomics in the task of distinguishing between high and low likelihoods of cancer recurrence yielded AUC values of 0.88, 0.76, and 0.68 for MammaPrint, Oncotype DX, and PAM50 risk of relapse based on subtype, respectively, with all showing statistical difference from chance. Such computer-extracted MR imaging radiomics shows potential for image-based phenotyping in assessing the risk of cancer recurrence.

Through an extensive investigation, we identified statistically significant associations between quantitative MRI radiomic features and various clinical, molecular, and genomic features in breast invasive carcinoma (Refs. 6-10). Among the many novel findings, we discovered some highly specific imaging-genomic assocations, which may be potentially useful in (a) imaging-based diagnoses that can inform the genetic progress of tumor and (b) discovery of genetic mechanisms that regulate the development of tumor phenotypes.

§ Robustness of MRI phenotypes

We continue to investigate the robustness of our computer-extracted MRI lesion phenotypes. Our recent robustness study focused on the robustness of features across MR scanners of two different manufacturers, GE (N = 91 cases) and Philips (N = 332 cases), in the prognostic task of distinguishing positive and negative lymph node status and receptor statuses of breast cancers. Our results demonstrated that robustness in values and in performance across MR scanners varies for different features. Additionally, we demonstrated that a classification model trained on a dataset of one MR manufacturer did not always generalize to a dataset of another MR manufacturer, thus requiring further optimization and harmonization. We are now expanding this robustness study to include the UCSF and I-SPY 1 datasets.

§ Role of deep learning in assessing response to therapy

We investigated CNN features extracted with pre-trained AlexNet, (AlexNet is the CNN that won the ImageNet Large Scale Visual Recognition Challenge in 2012) in the task of breast malignancy assessment on DCE-MRIs. The CNN features merged with a support vector machine (SVM) classifier showed promising performance with an AUC = 0.81 (standard error (se) = 0.01) in the task of distinguishing between malignant and benign breast tumors on DCE-MRI. Based on our promising preliminary results, we will assess these pre-trained networks and optimize steps in the CNN feature extraction for use in assessing response to treatment, monitoring treatment, and assessing recurrence.

Gantt Chart of Progress: We are on schedule with our aims.	Year 1	Year 2	Year 3	Year 4	Year 5
Aim 1: We will investigate the relationship of breast cancer therapy outcome and MR image-based tumor characteristics (features), and *changes* in these features	We further developed our quantitative radiomics workstation for response to therapy and incorporated deep learning.	We participated in the QIN BMMR Challenge, using the datasets for training and testing for assessing recurrence.	We will assess the deep learning-based extracted features in predicting recurrence.		
Aim 2: We will develop and evaluate the MRI-derived 'signatures' of breast cancer tumors for the prediction of, and assessment of, response to therapy using the ACRIN 6657 dataset.		We further incorporated the I-SPY 1 datasets into our analyses of robustness.	Will extend our radiomics features to be evaluated over all the I-SPY 1 data for response to therapy.	Will extend our radiomics features to be evaluated over all the I-SPY 1 data for monitoring therapy.	We will finalize our system for automatically analyzing (segmentation and feature extraction) for response prediction.
Aim 3. We will conduct preliminary, initial stratification and association of the MRI tumor features with cancer subtype and other clinical and histopathological data from the ACRIN dataset using unsupervised techniques.			We will investigate the merging of radiomics and genomic features for response to therapy by further conducting unsupervised discovery.		

Table 1: Tasks to be performed within the specific aims of the project.

PLANS FOR NEXT YEAR

We are now evaluating our MRI features relative to pathologic response to treatment on the full I-SPY 1 dataset of over 220 cases. First, we will evaluate the predictive value of the pre-treatment MR image-based tumor features in the "prediction of response to therapy". We will evaluate the image-based tumor features, calculated on the pre-treatment images, in terms of their ability to predict patient pathological response (pCR). Performance for the predictive task of distinguishing between patients that responded to the treatment and those that did not will be assessed quantitatively through ROC analysis with the area under the ROC curve (AUC) as the performance metric.

We will evaluate the predictive value of the pre-treatment MRI-based tumor features, the post-treatment MRI-based tumor features, and the changes in the pre- and post-treatment MR image-based tumor features in terms of "monitoring treatment response". We will

calculate the change in tumor characteristics obtained from the ratio of the tumor characteristic feature from before neoadjuvant therapy to that after therapy.

We will continue translating our findings from the TCGA Breast Phenotype Group to predicting response to therapy. Our research with the Breast Phenotype Group was conducted for "discovery" of relationships, so that we can then assess which ones are complimentary and thus could potentially be merged to yield an improved predictive imaging-genomics signature. In this study, we will determine optimal dimensional reduction methods for use with deep learning to yield these signatures using both CAD-extracted features and CNN-extracted features.

Through this QIN grant, our participation in the QIN community will yield deliverables including an open-platform system that will provide tools for linking segmentation with feature extraction and classification and for comparing performance metrics across acquisition systems and/or image analysis systems.

LIST OF REFERENCES FROM OUR SUBMITTED QIN GRANT

1. Giger ML, Huo Z, Kupinski MA, Vyborny CJ. Computer-aided diagnosis in mammography. Handbook of Medical Imaging, Volume 2. Medical Imaging Processing and Analysis, (Sonka M, Fitzpatrick MJ, eds) SPIE, pp. 915-1004, 2000.
2. Giger ML. Computer-aided diagnosis In: Haus AG Yaffe MJ eds. Technical Aspects of Mammography, Oakbrook IL: RSNA Publications, 1993, p 283.
3. Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography. AJR 162:699-708, 1994. PMID: 8109525
4. Vyborny CJ, Giger ML, Nishikawa RM. Computer-aided detection and diagnosis. Radiologic Clinics of North America 38:725-740, 2000. PMID: 10943274
5. Giger ML. Computerized image analysis in breast cancer detection and diagnosis. Seminars in Breast Disease 5: 199-210, 2002.
6. Giger ML, Chan H-P, Boone J: Anniversary Paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM. Medical Physics 35: 5799-5820, 2008. PMID: 19175137
7. Jiang Y, Nishikawa RM, Schmidt RA, et al: Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. Radiology 220: 787-794, 2001. PMID: 11526283
8. Behrens S, Laue H, Althaus M, Boehler T, Kuemmerlen B, Hahn HK, Peitgen HO. Computer assistance for MR based diagnosis of breast cancer: present and future challenges. Comput Med Imaging Graph 31:236-247, 2007. PMID: 17369019
9. Lucht R, Delorme S, Brix G. Neural network-based segmentation of dynamic MR mammographic images. Magn Reson Imaging 20:147-154, 2002. PMID: 12034335
10. Meinel LA, Stolpen AH, Berbaum KS, Fajardo LL, Reinhardt JM. Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. J Magn Reson Imaging 25:89-95, 2007. PMID: 17154399
11. Yu HJ, Chen JH, Mehta RS, Nalcioglu O, Su MY. MRI measurements of tumor size and pharmacokinetic parameters as early predictor of response in breast cancer patients

undergoing neoadjuvant AC-chemotherapy. J Magn Reson Imaging 26:615-623, 2007. PMID: 17729334
12. Chou YH, Tiu CM, Hung GS, Wu SC, Chang TY, Chiang HK. Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis. Ultrasound in Med & Biol 27: 1493-1498, 2001. PMID: 11750748
13. Chen DR, Chang RF, Chen CJ, Ho MF, Kuo SJ, Chen ST, Hung SJ, Moon WK. Classification of breast ultrasound images using fractal feature. J of Clinical Imaging 29: 235-245, 2005. PMID: 15967313
14. Huo Z, Giger ML, Vyborny CJ: Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis. IEEE Transactions on Medical Imaging 20: 1285-1292, 2001. PMID: 11811828
15. Stavros AT, Thickman D, Rapp CL, et al. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology 196:123-134, 1995. PMID: 7784555
16. Horsch K, Giger ML, Vyborny CJ, Huo Z, Venta LA. Performance of CAD in the interpretation of lesions on breast sonography. Academic Radiology 11: 272-280, 2004. PMID: 15035517
17. Giger ML, Al-Hallaq H, Huo Z, Moran C, Wolverton DE, Chan CW, and Zhong W. Computerized analysis of lesions in us images of the breast. Acad Radiol 6:665-674, 1999. PMID: 10894069
18. Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, and Zeman RK. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. Ultrason Imaging, 15:267--285, 1993. PMID: 8171752
19. Chen DR, Chang RF, and Huang YL. Computer-aided diagnosis applied to us of solid breast nodules by using neural networks. Radiology, 213:407-412, 1999. PMID: 10551220
20. Horsch K, Giger ML, Venta LA, and Vyborny CJ. Automatic Segmentation of Breast Lesions on Ultrasound. Med Phys, 28:1652-1659, 2001. PMID: 11548934
21. Horsch K, Giger ML, Venta LA, Vyborny CJ. Computerized diagnosis of breast lesions on ultrasound. Medical Physics 29: 157-164, 2002. PMID: 11865987
22. Sahiner B, LeCarpentier GL, Chan HP, and et al. Computerized characterization of breast masses using three-dimensional ultrasound images. Proceedings of the SPIE 3338: 301-312, 1998.
23. Drukker K, Giger ML, Vyborny CJ, Mendelson EB: Computerized detection and classification of cancer on breast ultrasound, Academic Radiology 11: 526-535, 2004. PMID: 15147617
24. Drukker K, Giger ML, Metz CE: Robustness of computerized lesion detection and classification scheme across differenct breast ultrasound platforms. Radiology. 237: 834-840, 2005. PMID: 16304105
25. Muller-Schimpfle M, Stoll P, Stern W. et al.: Do mammography, sonography, and MR mammography have a diagnostic benefit compared with mammography and sonography? AJR 168: 1323-1329, 1997. PMID: 9129436
26. Brinck U, Fischer U, Korabiowska M, et al. The variability of fibroadenoma in contrast-enhanced dynamic MR mammography. AJR 168: 1331-1334, 1997. PMID: 9129437

27. Gilhuijs KGA, Giger ML, Bick U. Automated analysis of breast lesions in three dimensions using dynamic magnetic resonance imaging. Medical Physics 25:1647-1654, 1998. PMID: 9775369
28. Nunes LW, Schnall MD, Orel SG, et al. Breast MR imaging: interpretation model. Radiology 202: 833-841, 1997. PMID: 9051042
29. Penn AI, Bolinger L, Schnall MD, Loew MH. Discrimination of MR images of breast masses with fractal-interpolation function models. Acad Radiol 6: 156-163, 1999. PMID: 10898034
30. Rose CJ, Mills SJ, O'Connor JPB, et al. Quantifying spatial heterogeneity in dynamic contrast-enhanced MRI parameter maps. Mag Reson Med 62:488-499, 2009. PMID: 19466747
31. Chen W, Giger ML, Bick U, Newstead G. Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI. Medical Physics, 33: 2878-2887, 2006. PMID: 16964864
32. Chen W, Giger ML, Lan L, Bick U. Computerized interpretation of breast MRI: Investigation of enhancement-variance dynamics. Medical Physics 31: 1076-1082, 2004. PMID: 15191295
33. Chen W, Giger ML, Li H, Bick U, Newstead. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. Magnetic Resonance in Medicine, 58:562-571, 2007. PMID: 17763361
34. Chen W, Giger ML, Bick U. A fuzzy c-means (FCM) based approach for computerized segmentation of breast lesions in contrast-enhanced MR images. Academic Radiology 13: 63-72, 2006. PMID: 16399033
35. Bhooshan N, Giger ML, Jansen S, Li H, Lan L, Newstead G. Cancerous breast lesions on dynamic contrast-enhanced MR images: computerized characterization for image-based prognostic markers. Radiology 254: 680-690, 2010. PMID: 20123903
36. Chen W, Giger ML, Newstead GM, Bick U, Sanaz SA, Li H, Lan L. Computerized assessment of breast lesion malignancy using DCE-MRI: Robustness study on two independent clinical datasets from two manufacturers. Academic Radiology 17: 822-829, 2010. PMID: 20540907
37. Sullivan DC. Imaging as a quantitative science. Radiology 248:328-332, 2008. PMID: 18641239
38. Luciani A, Dao TH, Lapeyre M, Schwarzinger M, Debaecque C, Lantieri L, Revelon G, Bouanane M, Kobeiter H, Rahmouni A. Simultaneous bilateral breast and high-resolution axillary MRI of patients with breast cancer: preliminary results. Am J Roentgenol 182:1059-1067, 2004. PMID: 15039188
39. Dees EC, Shulman LN. Does information from axillary dissection change treatment in clinically node-negative patients with breast cancer? an algorithm for assessment of impact of axillary dissection. Ann Surg 226:279-287, 1997. PMID: 9339934
40. Wang X, Liu L, Fackenthal J, Cummings S, Olopade O, Hope K, Silverstein JC, Olopade OI. Translational integrity and continuity: Personalized biomedical data integration. J of Biomedical Informatics 42: 100-112, 2009. PMID: 18760382
41. Huo Z, Giger ML, Vyborny CJ, Metz CE. Breast cancer: effectiveness of computer-aided diagnosis observer study with independent database of mammograms. Radiology 224:560-568, 2002. PMID: 12147857

42. Huo Z, Giger ML, Vyborny CJ, et al. Analysis of spiculation in the computerized classification of mammographic masses. Medical Physics 22: 1569-1579, 1995. PMID: 8551981
43. Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Schmidt RA, Doi K. Automated computerized classification of malignant and benign mass lesions on digitized mammograms. Academic Radiology 5: 155-168, 1998. PMID: 9522881
44. Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Metz CE. Computerized classification of benign and malignant masses on digitized mammograms: a robustness study. Academic Radiology 7:1077-1084, 2000. PMID: 11131052
45. Kupinski MA, Giger ML. Automated seeded lesion segmentation on digital mammograms. IEEE Trans on Medical Imaging, 17: 510-517, 1998. PMID: 9845307
46. Yuan Y, Giger ML, Li H, Suzuki K, Sennett C. A dual-stage method for lesion segmentation on digital mammograms. Med Phys 34:4180-4193, 2007. PMID: 18072482
47. Li H, Giger ML, Yuan Y, Chen W, Lan L, Jamieson A, Sennett CA, Jansen SA. Evaluation of Computerized Diagnosis Scheme on a Large Clinical Full-Field Digital Mammographic Dataset. Academic Radiology 15:1437-1445, 2008. PMID: 18995194
48. Yuan Y, Giger ML, Li H, Sennett CA. Correlative feature analysis on FFDM. Medical Physics 35: 5490-5500, 2008. PMID: 19175108
49. Drukker K, Giger ML, Horsch K, Kupinski MA, Vyborny CJ, Mendelson EB. Computerized lesion detection on breast ultrasound. Medical Physics 29:1438-1446, 2002. PMID: 12148724
50. Drukker K, Giger ML, Mendelson EB. Computerized analysis of shadowing on breast ultrasound for improved lesion detection. Medical Physics 30: 1833-1842, 2003. PMID: 12906202
51. Drukker K, Gruszauskas NP, Sennett CA, Giger ML. Breast US computer-aided diagnosis workstation: performance with a large clinical diagnostic population. Radiology. 248:392-397, 2008. PMID: 18574139
52. Gruszauskas NP, Drukker K, Giger ML, Sennett CA, Pesce LL. Performance of Breast Ultrasound Computer-aided Diagnosis Dependence on Image Selection. Acad Radiol 15:1234-45, 2008. PMID: 18790394
53. Gruszauskas NP, Drukker K, Giger ML, Chang R-F, Sennett CA, Moon WK, Pesce LL. Breast US computer-aided diagnosis system: Robustness across urban populations in South Korea and the United States. Radiology 253: 661- 671, 2009. PMID: 19864511
54. Drukker K, Horsch K, Giger ML. Multimodality computerized diagnosis of breast lesions using mammography and sonography. Academic Radiology 12: 970-979, 2005. PMID: 16087091
55. Giger ML, Huo Z, Lan L, Vyborny CJ. Intelligent search workstation for computer-aided diagnosis. Proc. of Computer Assisted Radiology and Surgery (CARS'2000), pp. 822-827, 2000
56. Giger ML, Huo Z, Vyborny CJ, et al. Results of an observer study with an intelligent mammographic workstation for CAD. Proceedings of IWDM 297-303, 2002.
57. Horsch K, Giger ML, Vyborny CJ, Lan L, Mendelson EB, Hendrick RE. Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set. Radiology 240: 357-368, 2006. PMID: 16864666

58. Drukker K, C.A. Sennett CA, and M.L. Giger ML. Automated method for improving system performance of computer-aided diagnosis in breast ultrasound. IEEE-TMI 28: 122-128, 2009. PMID: 19116194
59. Shimauchi A, Giger ML, Bhooshan N, Lan L, Chen W, Lee J, Abe H, Newstead G. Evaluation of clinical breast MR imaging performed with prototype computer-aided diagnosis breast MR imaging workstation: Reader study. Radiology 258: 696-704, 2011. PMID: 21212365
60. Bhooshan N, Giger ML, Drukker K, Yuan Y, Li H, McCann S, Newstead G, Sennett C. Performance of triple-modality CADx on breast cancer diagnostic classification. Proceedings of IWDM 6136: 9-14, 2010.
61. Lehman CD, Peacock S, DeMartini WB, Chen X. A new automated software system to evaluate breast MR examinations: Improved specificity without decreased sensitivity. AJR 187: 51-56, 2006. PMID: 16794155
62. Hylton N. Dynamic contrast-enhanced magnetic resounance imaging as an imaging biomarker. J Clinical Oncology 24: 3293-3298, 2006. PMID: 16829653
63. Tourassi GD, Floyd CE, Frederick ED, Markey MK. Application of the mutual information criterion for feature selection in computer-aided diagnosis. Medical Physics 28: 2394-2402, 2001. PMID: 11797941
64. Sahiner B, Chan HP, Petrick N, Wagner RF, Hadjiiski L. Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size. Medical Physics 27: 1509-1522, 2000. PMID: 10947254
65. Maaten LVD, Postma E,van den Herik H. Dimensionality Reduction: A Comparative Review. (Tilburg University: 2009).
66. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol 24: 498-520, 1933.
67. Kirby M. Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns. (John Wiley & Sons, Inc.: New York, 2000).
68. Drukker K, Gruszauskas NP, Giger ML. Principal component analysis, classifier complexity, and robustness of sonographic breast lesion classification. Proceedings of the SPIE 7260: 72602B-6, 2009.
69. Levman JE, Causer P, Warner E, Martel AL. Effect of the Enhancement Threshold on the Computer-Aided Detection of Breast Cancer using MRI. Academic Radiology 16, 1064-1069, 2009. PMID: 19515584
70. van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 9: 2579-2605,2008.
71. Bengio Y, Delalleau O, Le Roux N, et al. Learning Eigenfunctions Links Spectral Embedding and Kernel PCA. Neural Comput 16: 2197-2219, 2004. PMID: 15333211
72. van der Maaten L. Learning a Parametric Embedding by Preserving Local Structure. Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics 5: 384-391, 2009.
73. Bengio Y. Learning Deep Architectures for AI. Foundations and Trends in Machine Learning 2:1-127, 2009.
74. Larochelle H, Bengio Y, Louradour J, Lamblin P. Exploring Strategies for Training Deep Neural Networks. J Mach Learn Res 10: 1-40, 2009.
75. Min R, van der Maaten L, Yuan Z, Bonner A, Zhang Z. Deep Supervised t-Distributed Embedding. Proceedings of the International Conference on Machine Learning (ICML)

2010.
76. Belkin M, Niyogi P, Sindhwani V. Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. J Mach Learn Res 7: 2399-2434, 2006.
77. Chan H, Wei D, Helvie MA, et al. Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space. Phys Med Biol 40, 857-876, 1995. PMID: 7652012
78. Kupinski MA, Edwards DC, Giger ML, Metz CE. Ideal observer approximation using Bayesian classification neural networks. IEEE-Trans Medical Imaging, 20: 886-899, 2001. PMID: 11585206
79. Drukker K, Edwards DC, Giger ML, Metz CE. Computerized 3-way classification of lesions using breast ultrasound and mammography. IWDM 2004, 540-545, 2004.
80. Chen W, Zur R, Giger ML. Bayesian neural network with ARD priors for joint feature selection and classification: potential use in CAD of medical imaging. Proceedings SPIE 2007.
81. Kupinski MA, Giger ML. Feature selection with limited datasets. Medical Physics 26: 2176-2182, 1999. PMID: 10535635
82. Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE, Jr., Marks JR, Winchester DP, Bostwick DG. Artificial neural networks improve the accuracy of cancer survival prediction. Cancer 79: 857-862, 1997. PMID: 9024725
83. Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. Breast Cancer Res Treat 22: 285-293, 1992. PMID: 1391994
84. Li CI, Malone KE, Daling JR. Differences in breast cancer stage, treatment, and survival by race and ethnicity. Arch Intern Med 163: 49-56, 2003. PMID: 12523916
85. Gevaert O, De Smet F, Timmerman D, Moreau Y, De Moor B. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. Bioinformatics 22: e184-190, 2006. PMID: 16873470
86. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347: 1999-2009, 2002. PMID: 12490681
87. Chi CL, Street WN, Wolberg WH, Application of artificial neural network-based survival analysis on two breast cancer datasets. AMIA Annu Symp Proc 11: 130-134, 2007. PMID: 18693812
88. Metz CE, Jiang Y, MacMahon H, Nishikawa RM, Pan X. Proproc v2.2.0. Department of Radiology, University of Chicago. http://radiology.uchicago.edu/?q=MetzROCsoftware (02/03/2011)
89. Metz CE, Herman BA, Shen J-H. Maximum-likelihood estimation of ROC curves from continuously-distributed data. Stat Med 17: 1033-1053, 1998. PMID: 9612889
90. Pan X, Metz CE. The "proper" binormal model: parametric ROC curve estimation with degenerate data. Academic Radiol 4: 380-389, 1997. PMID: 9156236
91. Metz CE, Pan X. “Proper" binormal ROC curves: theory and maximum-likelihood estimation. J Math Psych 43: 1-33, 1999. PMID: 10069933
92. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 8: 283-298, 1978. PMID: 112681

93. Pee, Longton, Anderson Schummer. Selecting differentially expressed genes from microarray experiments. Biometrics 59: 133-142, 2003. PMID: 12762450
94. Elfron B, Tibshirani R. Improvements on Cross Validation: The .632+ Boostrap Method. Journal Amer Stat Soc 92: 548-560, 1997.
95. Platt R, Hanley J, Yang H. Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test. Statist Med 19: 313-322, 2000. PMID: 10649298
96. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA 299: 1335-1345, 2008. PMID: 18349094
97. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. Science 314: 268-274, 2006. PMID: 16959974
98. Parmigiani G, Lin J, Boca S, et al. Statistical methods for the analysis of cancer genome sequencing data. Johns Hopkins University Dept. Biostatistics Working Paper, COBRA Collection of Biostatistics Research Archive, Berkeley Electronic Press 2007 Paper 126.
99. Wagner RF, Metz CE, Campbell G. Assessment of medical imaging systems and computer aids: a tutorial review. Academic Radiology 14:723-748, 2007. PMID: 17502262
100. Chan HP, Doi K, Vyborny CJ et al. Improvements in radiologist's detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis. Invest Radiol 25:1102-1110, 1990. PMID: 2079409
101. Jiang Y, Nishikawa RM, Schmidt RA, et al. Improving breast cancer diagnosis with computer-aided diagnosis. Academic Radiology 6: 22-23, 1999. PMID: 9891149
102. Jamieson A, Giger ML, Drukker K, Li H, Yuan Y, Bhooshan N. Exploring non-linear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE. Medical Physics 37: 339-351, 2010. PMID: 20175497
103. Jamieson AR, Giger ML, Drukker K, Pesce LL. Enhancement of breast CADx with unlabeled data. Medical Physics 37: 4155-4172, 2010. PMID: 20879576
104. Klein Zeggelink WFA, Deurloo EE, Bartelink H, Rutgers EJT, Gilhuijs KGA. Reproducibility of the assessment of tumor extent in the breast using multiple image modalities. Medical Physics 30: 2919-2926, 2003. PMID: 14655939
105. Drukker K, Pesce L, Giger ML. Repeatability in computer-aided diagnosis: Application to breast cancer diagnosis on sonography. Medical Physics 37: 2659-2669, 2010. PMID: 20632577
106. McKean-Cowdin R, Kolonel LN, Press MF, Pike MC, Henderson BE. Germ-line HER-2 variant and breast cancer risk by stage of disease. Cancer Research 61:8393-8394, 2001. PMID: 11731415
107. Bhooshan N, Giger ML, Lan L, Li H, Marquez A, Shimauchi A, Newstead GM: Combined use of T2-weighted MRI and T1-weighted dynamic contrast-enhanced MRI in the automated analysis of breast lesions. Magnetic Resonance in Medicine 66: 555-564, 2011. PMID: 21523818
108. Bhooshan N, Giger ML, Edwards E, Yuan Y, Jansen S, Li H, Lan L, Sattar H, Newstead G: Computerized three-class classification of MRI-based progrnostic markers for breast cancer. PMB 45: 5995-6008, 2011. PMID: 21860079
109. Yi M, Krishnamurthy S, Kuerer HM, Meric-Bernstam F, Bedrosian I, Ross MI, Ames FC, Lucci A, Hwang RF, Hunt KK. Role of primary tumor characteristics in predicting positive sentinel lymph nodes in patients with ductal carcinoma in situ or microinvasive breast cancer. The American Journal of Surgery 196: 81-87, 2008. PMID: 18436181

110. Zaugg K, Bodis S. Is there a role for molecular prognostic factors in the clinical management of ductal carcinoma in situ (DCIS) of the breast? Radiotherapy & Oncology 55:95-99, 2000. PMID: 10799720
111. Smith BL. Clinical application of breast pathology: Management of in situ breast carcinomas and sentinel node biopsy issues. Modern Pathology 23: 533-535, 2010. PMID: 20436500
112. McCaskill-Stevens W. National Institutes of Health State-of-the-Science Conference on the management and diagnosis of ductal carcinoma in situ: A call to action. J NCI Monographs 41: 111112, 2010. PMID: 20956812
113. Lee YS, Mathew J, Dogan BE, Resetkova E, Huo L, Yang WT. Imaging features of micropapillary DCIS: Correlation with clinical and histopathological findings. Acad Radiol 18: 797-803, 2011. PMID: 21419669
114. Tamimi RM, Baer HJ, Marotti J, Galan M, Galaburda L, Fu Y, Deitz AC, Connolly JL, Schnitt SJ, Colditz GA, Collins LC. Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. Breast Cancer Research 10:R67, Epub 2008 Aug 5. PMID: 18681955
115. Kristensen VN, Vaaske CJ, Ursini-Siegel J, Looo PV, et al. Integrated molecular profiles of invasive breast tumors and ductal carcinoma in situ (DCIS) reveal differential vascular and interleukin signaling. Proc Natl Acad Sci, 2011. PMID: 21908711
116. Emster VL, Barclay J, Kerlikowske K, Grady D, Henderson C. Incidence of and treatment for ductal carcinoma in situ of the breast. JAMA 275: 913-918, 1996.
117. Gibbs P, Turnbull LW. Textural analysis of contrast-enhanced MR images of the breast. Magn Reson Med 2003;50:92–98.
118. Jansen SA, Conzen SD, Fan X, Krausz T, Zamora M, Foxley S, River J, Newstead GM, Karczmar GS. Detection of in situ mammary cancer in a transgenic mouse model: in vitro and in vivo MRI studies demonstrate histopathologic correlation. Phys Med Biol 53: 5481-5493, 2008.
119. Jamieson AR, Alam R, Giger ML: Exploring deep parametric embeddings for breast CADx, Proc. SPIE Medical Imaging Conference, 2011.
120. Jamieson, AR, Drukker K, Giger ML: Breast image feature learning with adaptive deconvolutional networks. Proc. SPIE Medical Imaging Conference, 2012.
121. Siwa Chan, Jeon-Hor Chen, Garima Agrawal, et al., "Characterization of Pure Ductal Carcinoma In Situ on Dynamic Contrast-Enhanced MR Imaging: Do Nonhigh Grade and High Grade Show Different Imaging Features?," Journal of Oncology, vol. 2010, Article ID 431341, 2010. doi:10.1155/2010/431341
122. Raza S, Vallejo M, Chikarmane SA, Birdwell RL. Pure ductal carcinoma in situ: a range of MRI features. American Journal of Roentgenology. 2008; 191:689–699.
123. Page DL, Dupont WD, Rogers LW et al. Continued local recurrence of carcinoma 15-25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. Cancer 1995;76:1197-1200.
124. Allegra CJ, Aberle DR, Ganschow P, Hahn SM, Lee CN, Millon-Underwood S, Pike MC, Reed SD, Saftlas AF, Scarvalone SA, Schwartz AM, Slomski C, Yothers G, Zon R. National Institutes of Health State-of-the-Science Conference Statement: Diagnosis and Management of Ductal Carcinoma In Situ September 22–24, 2009. J Natl Cancer Inst. 2010; 102(3):161–169.

125. Armato SG, Gruszauskas NP, MacMahon H, Tomo MD, Li F, Engelmann RM, Starkey A, Pudela CL, Marion JS, Santiago F, Chang PJ, Giger ML: Research Imaging in an Academic Medical Center. Academic Radiology 19: 762-711, 2012.
126. Bhooshan N, Giger ML, Medved M, Wood A, Yuan Y, Lan L, Marquez A, Li H, Newstead G, Karczmar G. Potential of computer-aided diagnosis of high spectral and spatial resolution (HiSS) MRI in the classification of breast lesions. JMRI 39: 59-67, 2014. PMID: 24023011
127. Howlader N, Noone AM, Krapcho M, et al., eds. *SEER Cancer Statistics Review, 1975-2008*. Bethesda, MD: National Cancer Institute; 2011. http://seer.cancer.gov/csr/1975_2008/, based on November 2010 SEER data submission, posted to the SEER web site.
128. Elkin EB, Hudis C, Begg CB, Schrag D. The effect of changes in tumor size on breast carcinoma survival in the U.S.: 1975-1999. Cancer 104:1149-57, 2005. PMID: 16088887
129. Shao R, Cao QJ, Arenas RB, Bigelow C, Bentley B, Yan W. Breast cancer expression of YKL-40 correlates with tumor grade, poor differentiation, and other cancer markers. Br J Cancer 105:1203-9, 2011. PMID: 21934681
130. Nixon AJ, Schnitt SJ, Gelman R, Gage I, Bornstein B, Hetelekidis S, Recht A, Silver B, Harris JR, Connolly JL. Relationship of tumor grade to other pathologic features and to treatment outcome of patients with early stage breast carcinoma treated with breast-conserving therapy. Cancer 78:1426-31, 1996. PMID: 8839547
131. Yerushalmi R, Hayes MM, Gelmon KA. Breast carcinoma--rare types: review of the literature. Ann Oncol 20:1763-1770, 2009. PMID: 19602565
132. Marc E. Lippman, *Breast Cancer*, in HARRISON'S PRINCIPLES OF INTERNAL MEDICINE, pt. 5 § 76, at 516-523 (Dennis L. Kasper, M.D. et al., eds, 16th ed 2005).
133. Rosenberg J, Chia YL, Plevritis S. The effect of age, race, tumor size, tumor grade, and disease stage on invasive ductal breast cancer survival in the U.S. SEER database. Breast Cancer Res Treat 89:47-54, 2005. PMID: 15666196
134. Michaelson JS, Silverstein M, Sgroi D, Cheongsiatmoy JA, Taghian A, Powell S, Hughes K, Comegno A, Tanabe KK, Smith B. The effect of tumor size and lymph node status on breast carcinoma lethality. Cancer 98:2133-2143, 2003. PMID: 14601082
135. Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. Breast Cancer Res 9:R6, 2007. PMID: 17239243
136. Wiseman SM, Makretsov N, Nielsen TO, Gilks B, Yorida E, Cheang M, Turbin D, Gelmon K, Huntsman DG. Coexpression of the type 1 growth factor receptor family members HER-1, HER-2, and HER-3 has a synergistic negative prognostic effect on breast carcinoma survival. Cancer 103:1770-1777, 2005. PMID: 15770691
137. Cheang MC, Chia SK, Voduc D, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009 May 20;101(10):736-50. PMID: 19436038
138. Rudolph P, Alm P, Olsson H, Heidebrecht HJ, Fernö M, Baldetorp B, Parwaresch R. Concurrent overexpression of p53 and c-erbB-2 correlates with accelerated cycling and concomitant poor prognosis in node-negative breast cancer. Hum Pathol 32:311-9, 2001. PMID: 11274641

139. Bertucci F, Finetti P, Birnbaum D. Basal Breast Cancer: a Complex and Deadly Molecular Subtype. Curr Mol Med 2011. PMID: 22082486
140. Abramson VG, Troxel AB, Feldman M, Mies C, Wang Y, Sherman L, McNally S, Diehl A, Demichele A. Cyclin D1b in human breast carcinoma and coexpression with cyclin D1a is associated with poor outcome. Anticancer Res 30:1279-1285, 2010. PMID: 20530440
141. Keyomarsi K, Tucker SL, Buchholz TA, et al. Cyclin E and survival in patients with breast cancer. N Engl J Med 347:1566-1575, 2002. PMID: 12432043
142. Robson ME, Chappuis PO, Satagopan J, et al. A combined analysis of outcome following breast cancer: differences in survival based on BRCA1/BRCA2 mutation status and administration of adjuvant treatment. Breast Cancer Res 6:R8-R17, 2004. PMID: 14680495
143. Silva J, Domínguez G, Silva JM, et al. Analysis of genetic and epigenetic processes that influence p14ARF expression in breast cancer. Oncogene 20:4586-4590, 2001. PMID: 11494155
144. Taneja P, Maglic D, Kai F, Zhu S, Kendig RD, Fry EA, Inoue K. Classical and Novel Prognostic Markers for Breast Cancer and their Clinical Significance. Clin Med Insights Oncol. 2010 Apr 20;4:15-34. PMID: 20567632
145. Venna J, Peltonen J, Nybo K, Aidos H, Kaski S. Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization. Journal of Machine Learning Research, 11(Feb.):451–490, 2010.
146. Hylton NM, Blume JD, Bernreuter WK, et al. Locally advanced breast cancer: MR imaging for predictin of response to neoadjuvant chemotherapy – Results from ACRIN 6657/I-SPY Trial. Radiology. 263: 663-672, 2012. PMID: 22623692
147. Sadinski M, Giger ML, Drukker K, Yamaguchi K, Lan L, Li H: Pilot study on consistency in size metrics for a multimodality PEM/MR breast imaging approach. Presented AAPM 2012.
148. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44(3):837–845, 1988.
149. Budreau D, Giger ML, Lan L: Breast MRI-based feature analysis in predicting neoadjuvant therapy response. Accepted for presentation at RSNA 2012.
150. Obuchowski NA. Sample size tables for receiver operating characteristic studies. AJR 175: 603-608, 2000. PMID: 22021528
151. Toledano A, Gatsonis CA. Regression analysis of correlated receiver operating characteristic data. Acad Radiol 1995; 2:S30-36; discussion S61-34, S70-31. PMID: 9419703
152. Pepe MS. The statistical evaluation of medical tests for classification and prediction. New York: Oxford University Press, 2003.
153. Wilkie J, Giger ML, Chinander MR, Engh CA, Hopper RH, Martell JM: Temporal radiographic texture analysis in the detection of periprosthetic osteolysis. Medical Physics 35: 377-387, 2008. PMID: 18293592
154. Newstead G, Abe H, Shimauchi A, Schmidt RA, Karczmar GS. Pure Ductal carcinoma in situ: kinetic and morphologic MR characteristics compared with mammographic appearance and nuclear grade. Radiology 3: 684-691, 2007. PMCID 18024450

155. Jansen SA, Fan X, Karczmar GS, Abe H, Schmidt RA, Giger M, Newstead GM. DCEMRI of breast lesions: is kinetic analysis equally effective for both mass and nonmass-like enhancement? Med Phys 2008;35:3102-3109.
156. Fan X, Karczmar GS. A new approach to analysis of the impulse response function (IRF) in dynamic contrast-enhanced MRI (DCEMRI): a simulation study. Magn Reson Med 2009;62(1):229-239.
157. Jansen SA, Fan X, Medved M, Abe H, Shimauchi A, Yang C, Zamora M, Foxley S, Olopade OI, Karczmar GS, Newstead GM. Characterizing early contrast uptake of ductal carcinoma in situ with high temporal resolution dynamic contrast-enhanced MRI of the breast: a pilot study. Phys Med Biol. 2010 Oct 7;55(19):N473-85. Epub 2010 Sep 21. PubMed PMID: 20858914.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

Below is a list of peer-reviewed publications. We also regularly presented on our QIN activities at AAPM, SPIE Medical Imaging, and RSNA.

1. Schacht D, Drukker K, Pak I, Abe H, Giger ML. Using quantitative image analysis to classify axillary lymph nodes on breast MRI: A new application for the Z 0011 Era. European Journal of Radiology 84: 392-397, 2015.
2. Guo W, Li H, Zhu Y, et al.. Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data. J Medical Imaging 2(4): 041007, 2015. PMID: 26835491
3. Burnside E, Drukker K, Li H, et al. Using computer-extracted image phenotypes from tumors on breast MRI to predict breast cancer pathologic stage. Cancer 2015 Nov 30. PMID: 26619259
4. Li H, Zhu Y, Burnside ES, et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA Dataset. npj Breast Cancer 2, 16012; doi:10.1038/npjbcancer.2016.12; 2016. PMID: 27853751
5. Li H, Zhu Y, Burnside ES, et al. MRI radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of gene assays of MammaPrint, Oncotype DX, and PAM50. Radiology doi:10.1148/radiol.2016152110, 2016. PMID: 27144536
6. Zhu Y, Li H, et al. Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma. Nature Scientific Reports 5:17787, 2015. PMID: 26639025
7. Huynh B, Li H, Giger ML: Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Medical Imaging 3(3), 034501, 2016.

U01CA187947: Computing, Optimizing, and Evaluating Quantitative Cancer Imaging Biomarkers

Stanford University Department of Radiology (2)

Sandy Napel, Ph.D.
Daniel L. Rubin, M.D., M.S.

INTRODUCTION

The Quantitative Imaging Network (QIN) is a consortium of many centers where researchers are developing and evaluating many different quantitative imaging methods to assess cancer. Among them, quantitative image features that can be computed from medical images are proving to be valuable biomarkers of underlying cancer biology that can be used for assessing response to treatment and predicting clinical outcome. It is now important to discover the best quantitative imaging features for each cancer type and imaging modality that characterize cancers to detect response to new therapeutics, to identify subtypes of cancer, and to correlate with cancer genomics. However, progress is thwarted by the lack of shared software algorithms, architectures, and tools required to compute, compare, evaluate, and disseminate these quantitative imaging features within the QIN and the broader community. Our project will tackle these challenges by developing and evaluating a publicly available executable and open source software platform, the Quantitative Imaging Feature Pipeline (QIFP), which will give researchers these capabilities for characterizing images of tumors and surrounding tissues for use in multi-center clinical trials and patient monitoring in general. It will also allow researchers to add their own algorithms for computing novel quantitative image features for their own studies, and for the benefit of the community as appropriate. In this way, the QIFP will facilitate assessment of the incremental value of new vs. existing feature sets for these purposes.

The QIFP will have the following key attributes that are needed to propel quantitative imaging research forward in the QIN and in the broader research community:

- Web-based, graphical user interface for development of configurable quantitative image feature processing pipelines that will enable researchers to explore combinations of quantitative imaging features
- Expandable and sharable library of quantitative image features algorithms
- Support for a variety of languages for quantitative image feature algorithms, e.g., Matlab, Java, and C/C++, via Docker containers
- Connectivity to images and other data stored in
    - the Cancer Imaging Archive (TCIA)
    - ePAD systems (another QIN project for image annotation/curation)
    - local data stores
    - PACS systems via DICOM
- Cloud-based cache of data and software

- Machine learning algorithms that permit researchers to efficiently establish how well a quantitative image feature or combination of features predicts a clinical or molecular variable.

The QIFP will fill a substantial gap in the science currently being carried out in the QIN by providing the tools and infrastructure to assess the value of novel quantitative imaging features of cancer, and thereby accelerate incorporation of new imaging biomarkers into single- and multi-center clinical trials. It will also have additional impact by providing a means to disseminate and to promote the use of the quantitative imaging methods being developed within QIN to the broader community.

DISCUSSION OF PROGRESS

§ Specific Aims

Aim 1: Create an expandable library of quantitative imaging feature algorithms capable of comprehensive characterization of the imaging phenotype of cancer. QIFP will accept DICOM image sets linked to regions of interest (ROIs) specifying the locations of tumors and other tissues of interest in the image sets, and compute from them vectors of quantitative features of the objects. The QIFP will be applicable to several modalities (e.g., CT, MR, PET), and it will support algorithms developed using a variety of languages, including Matlab, IDL, C++, Python, and Java. We will initially populate the library with a broad set of algorithms, including those that provide volumetric and time-varying assessment of lesion size, shape, edge sharpness, and pixel statistics, developed by our team and by QIN and other researchers. A plug-in architecture will allow the community to add and share novel algorithms developed for their own research.

Aim 2: Build a cloud-based software architecture for creating and executing quantitative image feature-generating pipelines, and for using and comparing image features to predict clinical/molecular features. The QIFP will allow researchers to configure workflows that extract a selection of quantitative features from regions of interest on images, to upload private imaging datasets with associated ROIs, or to utilize images and ROIs stored in our system or linked to public repositories such as The Cancer Imaging Archive. Researchers will use a web-based interface to configure image processing pipelines including algorithms in the library (Aim 1) and/or those supplied by themselves and/or others as plug-ins. Quantitative features extracted from images can then be integrated with other data (e.g., gene expression, RNA sequence data, clinical data, outcomes), thereby making imaging data accessible for modern biological study, including the discovery of image biomarkers of specific disease subtype (a.k.a. radiogenomics), outcome, or response to treatment. Two novel aspects of QIFP will be (a) a **Predictive Model Generation Engine** that uses machine learning to let the user specify a dependent variable (e.g., progression-free survival) that the quantitative image features can be used to predict, and (b) a **Feature Evaluation Engine** that determines the values of particular features for predicting the dependent variable. Users will be able store and share their pipelines, promoting dissemination and widespread use of the feature computation algorithms.

Aim 3: Assess the QIFP's ability to facilitate the development and evaluation of novel quantitative imaging biomarkers of cancer in multi-center clinical trials in four ways. First, we will apply the QIFP to the multi-center clinical trial data (ECOG: E2408) used in our existing QIN project, and assess its ability to reproduce the known result that SUV_{max} predicts survival in this cancer, while showing that using QIFP improves efficiency. Second, within this same trial, we will evaluate the ability of QIFP to facilitate investigations of novel quantitative imaging features by comparing linear measurement, metabolic tumor burden and novel combinations of the features in our library (Aim 1) for predicting one-year progression-free survival. Though this evaluation will be in a specific cancer and will assess particular imaging biomarkers, QIFP will be generalizable and easily applied to image datasets in other cancers for assessment of many other quantitative imaging biomarkers. Third, we will utilize QIFP to merge imaging features with known host-, drug- and tumor-based follicular lymphoma biomarkers in order to develop the most robust and integrative predictive model for patient outcomes. Fourth, we will show benefit to the community by using the QIFP to combine image feature algorithms developed by another QIN team and our own NCI-funded team in the study of radiogenomics of non-small cell lung cancer.

§ Progress against Specific Aims

Our specific objectives and progress against these Aims for Years 1 & 2 were to: (labels C.n.m refer to our grant proposal and the Gantt Chart (Fig. 1)):

AIM 1: Develop a suite of configurable image feature characterization algorithms:

C.1.1 Begin the development of configurable image feature characterization algorithms; initial version complete, refinement may follow,

C.1.2 Begin the development of new pre- and post-processing algorithms; several modules complete; additional ones may follow,

C.1.3 Complete the specification of input/output and parameter block requirements; initial specification complete and functional. However, we are currently considering a more robust design based on the Slicer Execution Model,

C.1.4 Begin the development of a set of simulated DICOM objects with known features; in progress (Fig. 2); Digital phantom objects are currently internally generated by our initial 3D quantitative image feature engine, and we plan to externalize them as a set of DICOM objects.

AIM 2: Build a cloud-based software architecture: The architecture specification is complete (Fig. 3). Although not anticipated in the proposal:

1. QIFP is based on Docker modules as plug-ins, facilitating community participation by allowing algorithms to be developed using any language on any computer platform,

2. We have implemented connectivity to the ePAD web-based image viewing and annotation software (http://epad.stanford.edu), facilitating QIFP-based image feature computations to be included in annotations.

C.2.2.1 Build QIFP cache of images, segmented regions, and clinical molecular data; complete; ahead of schedule,

C.2.2.2 Build a library of quantitative image feature algorithms:

1. The capability to store and select from a collection of Dockerized feature generation engines is complete,

2. A lung field segmentation tool that generates a DSO for the complete lung field, given a chest CT DICOM series is complete,

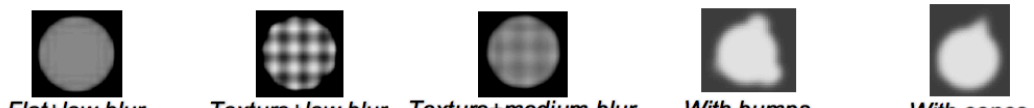
3. A lung nodule segmentation tool that generates a DSO for a lung nodule, given a chest CT DICOM series, a DSO defining the lung field in the series, and an AIM file containing a "seed circle" identifying a subset of the nodule, is complete,

4. A 3D feature computation engine, that generates a spreadsheet where each column corresponds to a DSO for a lung nodule in a CT series and each row corresponds to one of 198 computed radiomics features (including intensity, edge sharpness, shape, and texture) given a collection of DICOM series and DSOs, is complete. Refinement may follow.

C.2.2.3 Build a tool for selecting input data; complete (Fig. 4)
C.2.2.3 Build a tool for configuring processing pipelines; deferred; we are starting with several preconfigured workflows (Fig. 4).
C.2.2.3 Design/build provenance architecture; deferred
C.2.2.3 Build predictive model engine; LASSO algorithm, implemented as a Docker container, is complete ahead of schedule
C2.2.4 Build Web-based user interface; begun ahead of schedule, we are working on a more modern and flexible front end design (Fig. 5) using Omnigraffle 6 for website wireframe development.

Section	Task	Year 1	Year 2	Year 3	Year 4	Year 5
	Aim 1: Library of quantitative imaging feature algorithms					
C.1.2	Develop suite of configurable image feature characterization algorithms					
C.1.2	Develop new pre- and post-processing algorithms					
C.1.2	Develop a specification of input/output and parameter block requirements					
C.1.2	Create a set of simulated DICOM objects with known features					
	Test individual algorithms					
	Aim 2: Build a cloud-based software architecture					
C.2.2.1	Build QIFP cache of images, segmented regions, and clinical/molecular data					
C.2.2.2	Build library of quantitative image features algorithms					
C.2.2.3	Build components of QIFP computing platform:					
C.2.2.3	Build tool for selecting input data					
C.2.2.3	Build tool for configuring processing pipelines					
C.2.2.3	Design/build provenance infrastructure					
C.2.2.3	Build predictive model generation engine					
C.2.2.3	Build feature evaluation engine					
C.2.2.4	Build Web-based user interface					
	Test infrastructure and tools; starting in Year 3, half-yearly distribution of current version to early adopters for feedback and refinement.					
	Aim 3: Evaluation of QIFP					
C.3.1.1	Validation of existing qualitative follicular lymphoma PET data					
C.3.1.2	Development of novel quantitative imaging features via QIFP					
C.3.1.3	Integration of novel imaging features and patient/tumor biomarkers					
C.3.1.4	Using QIFP to enable radiogenomics research					
C.2.3	User support and dissemination of resources					

Figure 1: Gantt chart showing planned developments per Specific Aims. Red line is current point in time.



all spheres: flat or-- hi intensity high texture	Low Blur: Texture vs. NoTexture	Texture: LowBlur vs MedBlur
Intensity Mean	37.43%	52.88%
Intensity Median	30.30%	70.37%
Intensity Mode	10.26%	10.26%
Intensity Trimmed Mean (25%)	38.28%	168.35%
Intensity Interquartile Difference	7.79%	44.28%
Intensity Mean Absolute Difference	9.19%	11.58%
Intensity Range	0.43%	6.59%
Intensity Standard Deviation	6.28%	6.09%
Intensity Kurtosis	22.56%	6.42%
Intensity Skewness	13.36%	6.59%
Intensity Max	56.41%	35.56%
Intensity Min	4.68%	10.75%

Figure 2: Example of cross-sections through simulated objects in DICOM images, showing different textures, shapes, and edge-sharpness; spreadsheet shows features computed using prototype feature computation engine for the three textured objects.

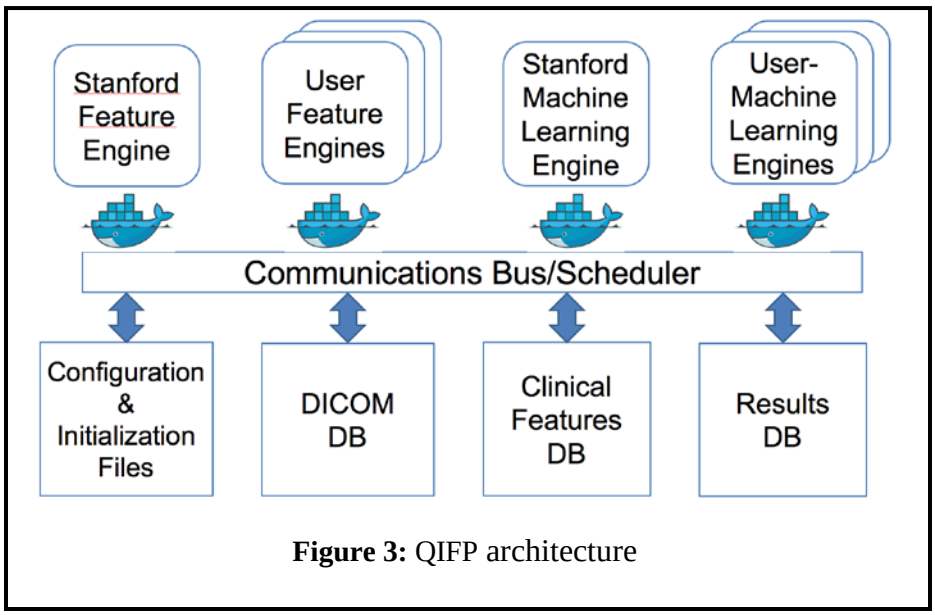


Figure 3: QIFP architecture

Prototype functionality:

The current QIFP prototype allows the user to upload data sets, consisting of a set of subjects' DICOM Series Data (DSD) + DICOM Segmentation Objects (DSOs), as well as clinical data for the subjects, in a variety of ways (Fig. 4):

- Web services connection to TCIA
- Web services connection to ePAD instances
- DICOM connection to local and remote PACS systems
- Direct upload from filesystem

The currently preconfigured workflows support the following use cases:

- Process DSD+DSOs using our 3D feature computation engine to produce quantitative features for each subject from volumes defined by the intersection of the DSOs with the DSDs
- Process stored DSD+DSOs together with clinical data as described above using our LASSO machine learning engine to produce a predictive model for a clinical variable and evaluate its performance on the training cohort using cross-validation. Process stored DSD+DSOs together with clinical data as described above and a previously computed and stored predictive model to predict a clinical variable for an uploaded cohort of DSD+DSOs and clinical data.

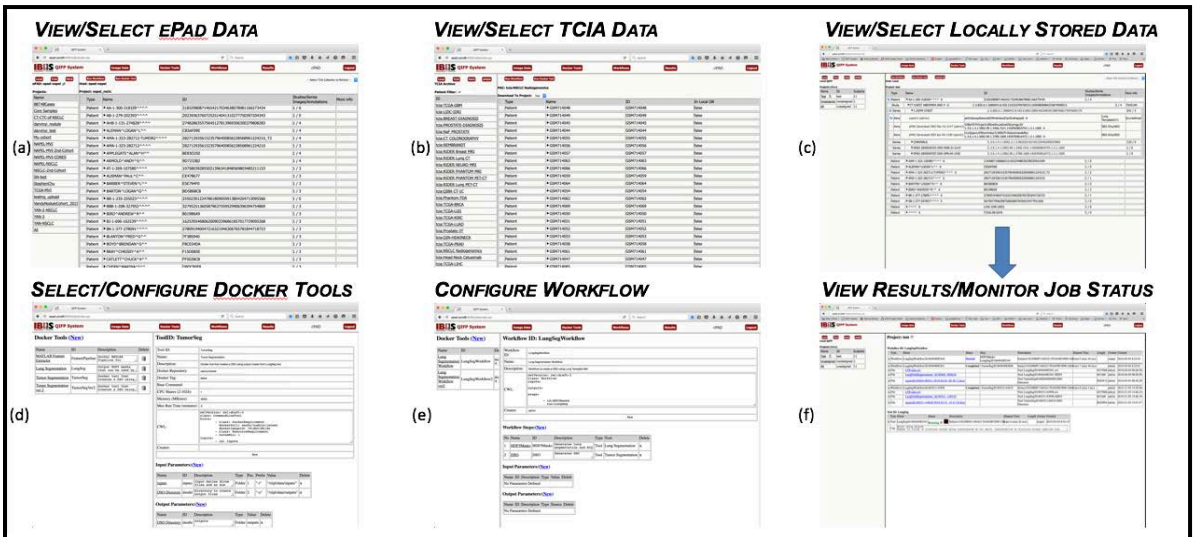


Figure 4: QIFP functional prototype implementation screens: allows selection of datasets (DICOM series, DSOs) from (a) ePAD instances, (b) TCIA, and (c) local filesystem. User can (d) select from available Docker tools (currently (1) lung field segmentation, (2) lung nodule segmentation, (3) 3D feature computation, (e) configure a workflow involving one or more Docker tools and (e) monitor job status and retrieve results.

PLANS FOR NEXT YEAR

We will continue our software developments as follows (labels C.n.m refer to our grant proposal and the Gantt Chart (Fig. 1)):

C.1.3 Complete the specification of input/output and parameter block requirements; We will finalize the specification of parameter passing to Docker modules and implement this for our existing modules (3D feature engine and LASSO machine learning engine),

C.1.4 Begin the development of a set of simulated DICOM objects with known features. We will try to complete the externalization of the digital phantoms by the end of the following year,

C.2.2.2 Build a library of quantitative image feature algorithms:

1. We will engage with the QIN community, specifically the BIDS Working Group, to test interoperability of Dockerized processing modules, and deploy additional feature engines and machine learning engines.

2. We will deploy at least one externally developed algorithm on the QIFP prototype

C.2.2.3 Build a tool for configuring processing pipelines: We will begin the design graphical use interface for user configuration of workflows,

C.2.2.3 Design/build provenance architecture: We will begin this design and implementation of a prototype system for recording experimental provenance

C.2.2.4 Build Web-based user interface: We will complete this design and a working prototype implementation.

In addition:

1. We will make a working prototype available to interested QIN participants,

2. We will release a working prototype by RSNA 2017 that will allow us to begin to train the broader community regarding the use of the QIFP.

Fig. 5: QIFP front-end design

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

§ Accepted Manuscripts

1. S. Echegaray, O. Gevaert, R. Shah, A. Kamaya, J. Louie, N. Kothary, S. Napel, “"Core Samples" for Radiomics Features that are Insensitive to Tumor Segmentation: Example in CT Images of Hepatocellular Carcinoma,” J. of Med. Imag, 2(4):041011, 2015. PMID: 26587549. PMCID: PMC4650964.
2. J. Kalpathy-Cramer, B. Zhao, D. Goldgof, Y. Gu, X. Wang, H. Yang, Y. Tan, R. Gillies, S. Napel, “A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study,” J Digit Imaging, 29(4):476–487, 2016. PMID: 26847203. PMCID: PMC4942386.
3. J. Wu, M. F. Gensheimer, X. Dong, D. L. Rubin, S. Napel, M. Diehn, B. W. Loo, R. Li, “Robust Intra-tumor Partitioning to Identify High-risk Subregions in Lung Cancer: a Pilot Study,” International Journal of Radiation Oncology, Biology, Physics 95(5):1504-12, 2016. PMID: 27212196. PMCID: PMC4969127.
4. S. Echegaray, V. Nair, M. Kadoch, A. N. C. Leung, D. L. Rubin, O. Gevaert, S. Napel, “A rapid segmentation-insensitive "digital biopsy" method for radiomic feature extraction; method and pilot study using CT images of non-small cell lung cancer,” *in press, Tomography, August 2016.*
5. J. Kalpathy-Cramer, A. Mamomov, B. Zhao, L. Lu, D. Cherezov, S. Napel, S. Echegaray, M. McNitt-Gray, P. Lo, J. C. Sieren, J. Uthoff, S. K. N. Dilger, B. Driscoll, I. Yeung, D. Goldgof, “Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features,” *in press, Tomography, October 2016.*
6. K. Lekadir, A. Galimzianova, À. Betriu, L. Igual, D. L. Rubin, E. Fernández, P. Radeva, and S. Napel, “A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound,” *in press, IEEE Journal on Biomedical and Health Informatics, October 2016 NIHMSID 832055.*
7. A. Hoogi, J. W. Lambert, Y. Zheng, D. Comaniciu, and D. L. Rubin, “A Fully Automated Pipeline for Detection and Segmentation of Liver Lesions and Pathological Lymph Nodes,” *in press, Thirtieth Annual Conference on Neural Information Processing Systems 2016.*

§ Submitted Manuscripts

1. M. Zhou, A. N. C. Leung, S. Echegaray, J. B. Shrager, K. C. Jensen, G. G. Berry MD, S. K. Plevritis, D. L. Rubin, S. Napel, O. Gevaert, “Radiogenomics mapping of non-small cell lung cancer identifies prognostic relationships between semantic image features and metagenes captured using RNA sequencing, *revised and re-submitted to Radiology August 2016.*
2. R. Minamimoto, M. Jamali, O. Gevaert, S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, S. K. Plevritis, D. L. Rubin, A. N. C. Leung, S. Napel, A. Quon, “Prediction of EGFR and KRAS Mutation in non-small cell lung cancer using

advanced quantitative 18F FDG-PET/CT metrics," *submitted to Oncotarget, September 2016.*

3. O. Gevaert, S. Echegaray, R. Shah, A. Kamaya, J. Louie, S. Napel, N. Kothary, "Non-invasive radiomics signature is predictive of microvascular invasion in hepatocellular carcinoma," *revised and re-submitted to Radiology, October 2016.*
4. O. Gevaert, S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, K. C. Jensen, G. J. Berry, S. K. Plevritis, D. L. Rubin, S. Napel, A. N. Leung, "Predictive radiogenomics modeling of EGFR mutation status in lung cancer," *revised and re-submitted to Nature Scientific Reports, October 2016.*
5. A. Hoogi, C. F. Beaulieu, G. M. Cunha, E. Heba, C. B. Sirlin, S. Napel and D. L. Rubin, "Adaptive Local Window for Level Set Segmentation of CT and MRI Liver Lesions," *revised and resubmitted, Medical Image Analysis, November 2016.*
6. S. Echegaray, S. Bakr, D. L. Rubin, S. Napel, "Quantitative Image Feature Engine (QIFE): An open-source, modular engine for 3D quantitative feature extraction from volumetric medical images," *submitted to Medical Physics, Nov. 2016.*

§ Presentations

1. O. Gevaert, S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, S. K. Plevritis, S. Napel, A. N. Leung, "Predictive modeling of epidermal growth factor receptor mutation status using semantic image features in non-small cell lung cancer (NSCLC)," Radiological Society of North America 101st Scientific Sessions, Chicago, December 2015.
2. O. Gevaert, S. Napel, S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, S. K. Plevritis, A. N. Leung, "Radiogenomics mapping of non-small cell lung cancer (NSCLC) identifies prognostic relationships between semantic image features and metagenes captured using RNA sequencing," Radiological Society of North America 101st Scientific Sessions, Chicago, December 2015.
3. V.S. Nair, A. Garcia, H. Chen, Y. Balagurunathan, T. Atwater[3] O. Gevaert, S. Antic, M. Schabath, S. Napel, R. Walker, R. Gillies, P. P. Massion, "Validating a radiomic classifier for improved lung cancer prediction of indeterminate pulmonary nodules," American Thoracic Society Annual Meeting, San Francisco, May 2016.
4. M. Zhou, S. Napel, S, Echegaray, A. N. Leung, O. Gevaert, "Radiogenomics Mapping of Non-small Cell Lung Cancer Shows Strong Correlations between Semantic Image Features and Metagenes," Radiological Society of North America 101st Scientific Sessions, December 2016.
5. S. Napel, S. Echegaray, D. Gude, O. Gevaert, D. L. Rubin, "The Quantitative Image Feature Pipeline (QIFP) for Discovery, Validation, and Translation of Cancer Imaging Biomarkers," Radiological Society of North America 102nd Scientific Sessions, Chicago, December 2016.
6. D. L. Rubin, C. Altindag, E. Alkim, "New developments in the ePAD platform to support quantitative imaging assessment in the research workflow," Scientific Exhibit in the Quantitative Imaging Reading Room of the Future (QIRR), Radiological Society of North America 102nd Scientific Sessions, Chicago, December 2016.

Contributions to QIN from Canadian Research Teams

In the fall of 2015, the Quantitative Imaging Network (QIN) welcomed two research teams from Canada. The team from the University of British Columbia and the team from the University Health Network of Toronto each submitted applications to NCI for review in study section earlier that summer. Both received favorable scores and funding was provided by the Institute of Canadian Research of the Canadian Institutes of Health Research and Genome BC.

Dr. Francois Benard leads the group from the University of British Columbia. His research topic has been *Integrating Quantitative Imaging Methods and Genomic Biomarkers to Assess the Therapeutic Response of Cancers*. This team is developing improved methods to measure the volume of tumors and to measure their accumulation of PET radiotracers. They will then correlate this measure of tumor mass to the amount of circulating tumor DNA in the blood. The 2016 report of progress is included in this section of the overall QIN report.

Dr, David Jaffrey and his team from the University Health Network is looking at Image-Based Quantitative Assessment of Tumor Hypoxia. Using PET imaging, this group is detecting hypoxic regions of tumors non-invasively. The goal is to be able to predict response to radiation therapy. Translation of their methods to a clinical setting is also being studied.

The Quantitative Imaging Network has been enriched by the participation of these two teams in the various network activities. The reports provided indicate the progress each has made and the network activities of each.

U01CA190232: Integrating Quantitative Imaging Methods and Genomic Biomarkers to Assess the Therapeutic Response of Cancers

University of British Columbia

François Bénard, M.D.
Anna Celler, Ph.D., Ghassan Hamarneh, Ph.D., Ryan Morin, Ph.D.

INTRODUCTION

The overall purpose of this project is to investigate methods for accurate quantitative measurements of functional tumor burden in cancer patients and develop robust criteria to define tumor boundaries, tumor uptake parameters and measure response in clinical trials. These data will be correlated with patient-specific circulating tumor DNA (ctDNA) levels measured serially over time in cohorts of subjects enrolled in clinical trials.

In the first year of this research program, we focused our research efforts in developing physical and digital phantoms to assess quantitative measurement methods, and started testing segmentation algorithms using phantom data. We also conducted preliminary experiments to implement quantitative serial ctDNA measurements in patients, to eventually correlate these values to quantitative measurements of functional tumour burden.

This year, additional work was done to optimize processing of PET data and generate models to better evaluate different algorithms and parameters. There has been further development of our segmentation technology, of tumour volume, and metabolic activity estimation from PET scans. Progress was made in ctDNA measurement and improvements were made to the underlying technology; ctDNA collection and analysis is ongoing with several new projects stated for 2017 with ctDNA correlation endpoints. Preliminary analysis of data from a pilot study evaluating the EF5 tracer has also been completed.

PROGRESS OVER THE PREVIOUS YEAR

One of the objectives of this project is to optimize processing of the PET data in order to obtain low-noise, high contrast quantitatively accurate images of activity distribution in the patient body. Since the recent studies strongly emphasize the critical need for standardization of PET/CT-derived metrics of tumor burden [1, 2] we focused our research on development of techniques which will allow us to objectively investigate accuracy and consistency of PET quantification for different imaging tasks when using different PET cameras and different reconstruction techniques. To this end, we investigated a number of image parameters and figures of merit which characterize the performance of the imaging system.

The spatial resolution of PET images, which is defined by both the scanner characteristics and the reconstruction algorithm, is one of such crucial parameters as it determines the ability of an observer to identify and diagnose the disease. Considering

quantitative measures, image spatial resolution will affect the sensitivity of the texture-based biomarkers. Additionally, image reconstruction algorithms introduce noise texture into PET/CT images. It may therefore be important to characterize and harmonize both the spatial resolution and noise texture of PET/CT images. Our objective is to develop a framework for measuring, comparing, and standardizing the spatial resolution and noise texture of PET/CT images across different scanners and reconstruction algorithms and to correlate them with quantitative accuracy.

Using digital and physical phantoms, we study the relationship between image spatial resolution and data quantification, and the properties of the imaged object, such as its size, shape, signal-to-background (SBR) ratio, noise, and its activity distribution. The analysis is being performed for different PET cameras, reconstruction algorithms (using a wide range of adjustable parameters) and different radioisotopes.

Spatial resolution is typically characterized in terms of the system point spread function (PSF), edge spread function (ESF) or their Fourier transform, the modulation transfer function (MTF). Measuring the PSF and/or MTF requires that the images are linear and shift-invariant, neither of which are satisfied for iteratively reconstructed PET images. However, it is possible to measure the MTF for a particular object under the assumption that linearity and stationarity are satisfied locally within the image. Spatial resolution must therefore be studied as a function of signal to background ratio (SBR), noise levels, and location.

For this task, we have developed an experimental framework which allows to measure image spatial resolution for objects scanned using PET imaging systems. We adapted the method of Richard et al [3] for measuring the MTF of CT images to the measurement of the task-based MTF of PET images. This algorithm was used to extract the transaxial MTF from PET images of a NEMA International Electrotechnical Commission (IEC) phantom containing F-18-filled spheres placed in a “hot” (radioactive) background.

Using this framework, we have already characterized image spatial resolution and quantification accuracy for the two different PET reconstruction algorithms. These are: the ordered subsets expectation maximization (OSEM) method, which is routinely used in PET image reconstruction and the block sequential regularized expectation maximization (BSREM), which is the penalized-likelihood algorithm recently introduced by GE Healthcare (under the name Q-Clear). The BSREM algorithm has two user-defined parameters β and γ which control the smoothness and edge preservation, respectively. The analysis was performed for the standard and the time-of-flight (TOF) versions of both algorithms and for a number of reconstruction parameters and digital filters.

This study was performed for two different radioisotopes:

1.) ^{18}F which is a pure positron emitter routinely used in a wide range of PET diagnostic studies and,
2.) ^{90}Y, an isotope used for liver radioembolization therapies, which rely on its pure β^- decay (emission of electrons). PET imaging of ^{90}Y uses an extremely weak (32×10^{-6}) $\beta^+ \rightarrow \beta^-$ internal pair production from the $0^+ \rightarrow 0^-$ monopole transition in ^{90}Zr [4]. Imaging of ^{90}Y

is very challenging, because low intensity of positrons combined with high background of Bremsstrahlung photons (created by β^- emissions) results in high noise in the reconstructed images.

In order to be able to compare different images, we decided to adjust reconstruction parameters and/or digital filters so that the background variability remains the same. To this end, the background variability (BV) figure of merit was used. The BV for each reconstruction was matched so that the image contrast could be objectively compared.

The BV is defined as:

$$BV = \frac{\sigma_{bkg}}{\mu_{bkg}} \qquad (2.$$

where σ_{bkg} and μ_{bkg} are the standard deviation and mean of counts in the ROI, respectively, averaged over six spherical ROIs (each ROI volume equal to 80 mL) drawn at different locations in the phantom with uniform activity concentration (outside the region with spheres).

The image quality was analyzed using the contrast-to-noise ratio (CNR) defined as:

$$CNR = \frac{S_{pk} - \mu_{bkg}}{\sigma_{bkg}} \qquad (2.$$

where S_{pk} is the peak activity, μ_{bkg} is the mean activity in the background, and σ_{bkg} is the standard deviation of the background. S_{pk} is defined as the most intense region within the analyzed volume. According to RECIST 1.1 (Response evaluation criteria in solid tumours), this region is defined as a 1.25 cm diameter spherical ROI [5]. Here, the background ROI is measured using a 200 mL spherical ROI drawn in the region with uniform activity.

The quantification accuracy was analyzed using the recovery coefficient (RC) defined as:

$$RC = \frac{A_{img}}{A_{DC}} \qquad (2.1$$

where A_{img} is the activity or activity concentration of the imaged isotope in the specified ROI, and A_{DC} is the true activity measured using a dose calibrator (DC).

The data used in these investigation were acquired experimentally using IEC phantoms and Discovery 690 PET/CT camera (GE Healthcare). Two series of experiments were performed.

§ Experiments using 18F

Methods

An NEMA-IEC phantom containing four F-18-filled spheres (two spheres had 3.7 cm diameter and the other two 2.8 cm) placed in a “hot” background was scanned twice using a

GE Discovery 690. The activity concentrations in the spheres were equal to 44.0kBq/mL and 24.9kBq/mL, resulting in SBRs for these two scans equal to 8.3 and 3.5, respectively. For each scan, list-mode data was collected and sorted into multiple frames with increasing duration. Each dataset was reconstructed using the OSEM algorithm with and without resolution recovery (RR), TOF and post-reconstruction Gaussian filtering (FWHM: 6.4 mm). The images were reconstructed with 32 subsets and 2 or 4 iterations. Additionally, the second series of reconstructions was performed using BSREM algorithm without TOF. The parameters used in these reconstructions were: β = 200, 350, and 500 and γ = 2.

The analysis was performed for the two largest spheres and for each combination of reconstruction algorithm and SBR.

Highlights of results

Figure 1 shows the ESF and MTF for OSEM with and without RR and TOF. As expected, both RR and TOF individually improve image spatial resolution and combining RR and TOF show the greatest improvement. However, RR introduces edge artifacts that appear as a 'horn' in the ESF and result in MTF values greater than 1. While applying RR and TOF improves the spatial resolution of the image, this improvement is less apparent when a post-reconstruction filter is applied.

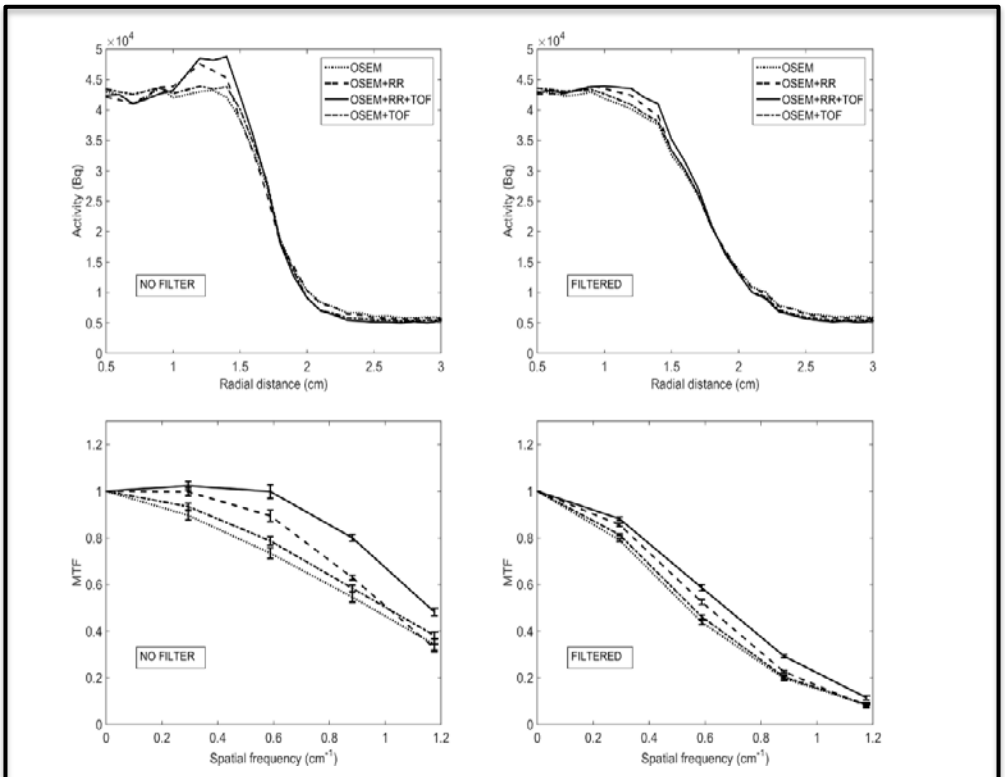


Figure 1: The ESF and MTF obtained from images reconstructed with OSEM, OSEM + RR, OSEM + RR + TOF, and OSEM + TOF reconstructions. The left and right columns show the results with and without post-reconstruction filtering with a 6.4 mm Gaussian, respectively.

Figure 2 shows a comparison of the ESF (left) and MTF (right) for images reconstructed with OSEM and with BSREM using different β values. Although BSREM reconstructed imagers have better spatial resolution compared to OSEM+RR+6.4 mm filtering, OSEM+RR (without filtering) produces images with the best image spatial resolution. Furthermore, as expected, the BSREM reconstructed images using increasing β (smoother), show a decrease in spatial resolution.

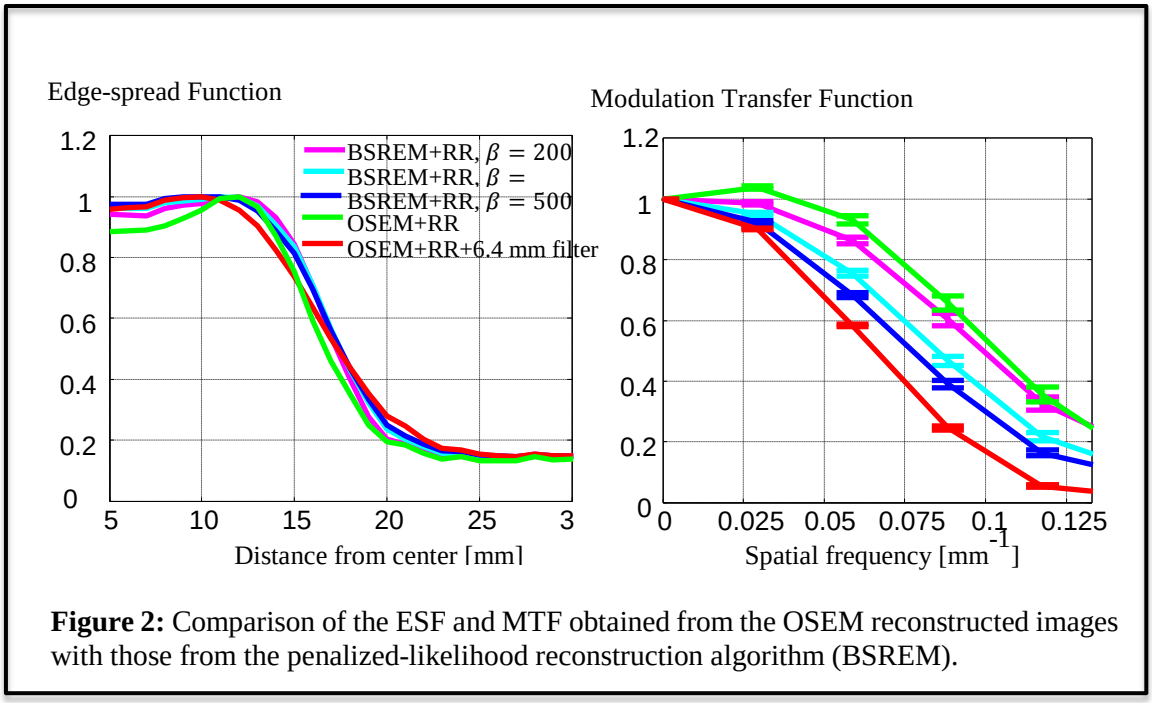


Figure 2: Comparison of the ESF and MTF obtained from the OSEM reconstructed images with those from the penalized-likelihood reconstruction algorithm (BSREM).

§ Experiments using 90Y

Methods

To investigate image quality and quantification accuracy in the presence of high levels of randomness and noise in the data, a series of PET scans of the NEMA-IEC phantom filled with ^{90}Y was performed. We adapted the ^{90}Y experimental protocol used in [6] which was a multi-center trial designed to investigate ^{90}Y PET quantitation. In particular, to obtain data with different levels of noise, the phantom once filled with ^{90}Y activity was scanned four times, on days 0, 3, 5 and 7. Additionally, for comparison, a separate ^{18}F scan with the same phantom and geometry was performed. In these experiments the phantom contained six fillable spheres (with diameters ranging from 10mm to 37mm) and a cylindrical lung insert filled with Styrofoam beads. Total activities, activity concentrations and SBR for these experiments are listed in Table 1. Subsequently, all these datasets were reconstructed using both OSEM and BSREM with RR and with and without TOF information.

Isotope used	Total Activity	Sphere activity concentration	Background activity concentration	SBR
^{18}F	55 ± 1MBq	38 ± 1kBq/mL	5.3 ± 0.2 kBq/mL	7.2 ± 0.3
^{90}Y	3180 ± 30 MBq	2.45 ± 0.02 MBq/mL	0.325 ± 0.003 MBq/mL	7.54 ± 0.09

Table 1: Values of total activity, activity concentration, and SBR used in the 18F and 90Y (Day 0) experiments.

The OSEM and TOF-OSEM reconstructions were performed using 2 iterations and 24 subsets with 5 mm and 8 mm Gaussian post-smoothing filter. These parameters were adapted from [6], while the post-smoothing filter size was chosen so that both images have the same BV. The BV value for the TOF-OSEM reconstruction of the ^{90}Y day 0 data was equal to 0.6.

The BSREM and TOF-BSREM reconstructions followed the procedure described in [7]. For BSREM and TOF-BSREM, the β was chosen to be 950 and 1300, respectively so that BV of the reconstructed images matched that of TOF-OSEM. Moreover, γ was chosen to be 2 for both BSREM and TOF-BSREM, which was adapted from [8]. Additionally, the β parameter for TOF-BSREM was adjusted at a value of 4000, which created smoother images (denoted as adj-TOF-BSREM).

Highlights of results

Images reconstructed using OSEM, BSREM, TOF-OSEM and TOF-BSREM for the day 0 scan (highest activity) are shown in Figure 3. Figure 4 represents the CNR for the four largest spheres reconstructed using OSEM, BSREM, TOF-OSEM and TOF-BSREM methods for day 0. Scans on days 3, 5 and 7 were not used in this quantitative analysis because the spheres were difficult to detect visually. Figure 4 shows that CNR for the all spheres, except the largest one, appears to be higher for images reconstructed with BSREM than those from OSEM. This result, for images with matched BV, suggests that the improvement of image contrast for objects of small sizes can be substantially improved when using penalized reconstruction algorithm with properly adjusted parameters.

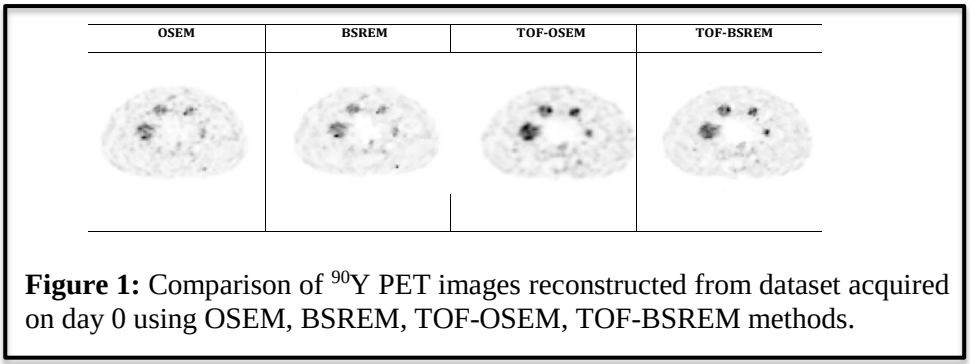


Figure 1: Comparison of ^{90}Y PET images reconstructed from dataset acquired on day 0 using OSEM, BSREM, TOF-OSEM, TOF-BSREM methods.

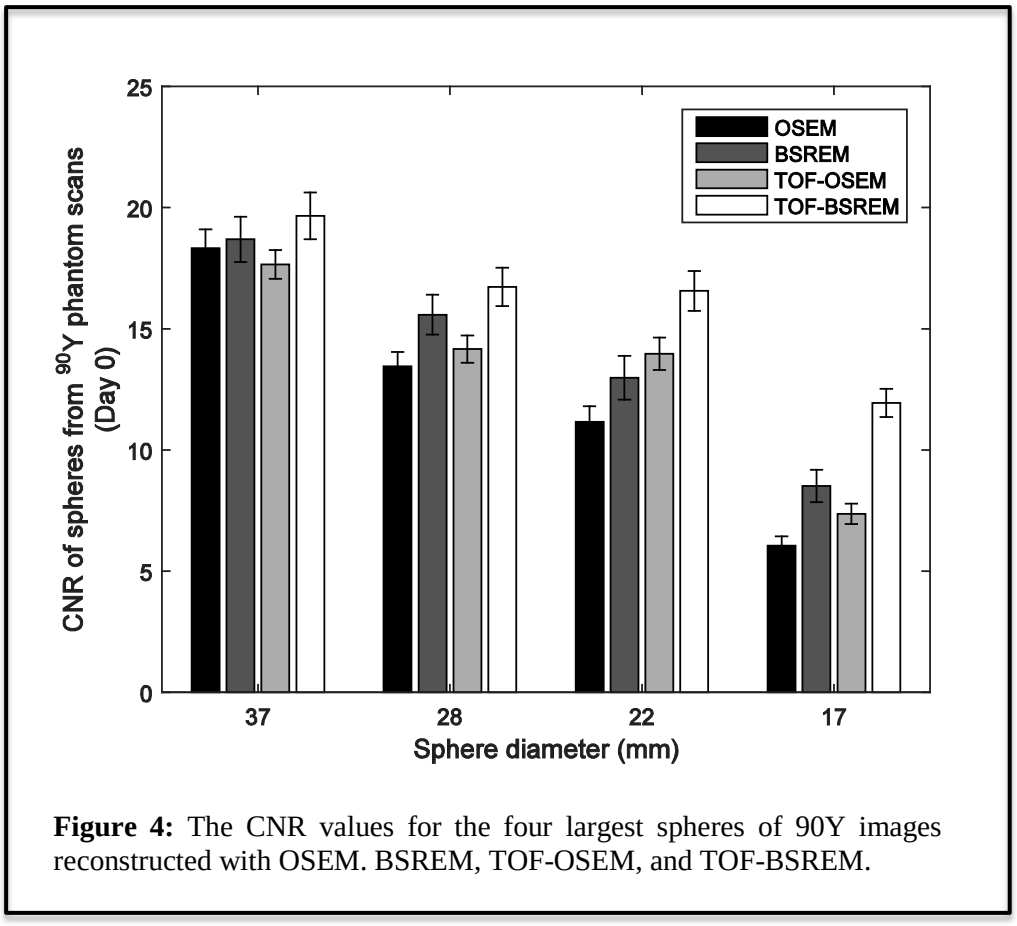


Figure 4: The CNR values for the four largest spheres of 90Y images reconstructed with OSEM. BSREM, TOF-OSEM, and TOF-BSREM.

Figure 5 shows the RC curves for the four largest spheres for images obtained using the OSEM and BSREM reconstruction methods calculated using datasets from the day 0 scan. Reconstructions with TOF information show an average improvement by 10% compared to the non-TOF reconstructions, while BSREM improves quantification by 10% for S4 (diameter of 17 mm) compared to OSEM reconstructions with and without TOF information.

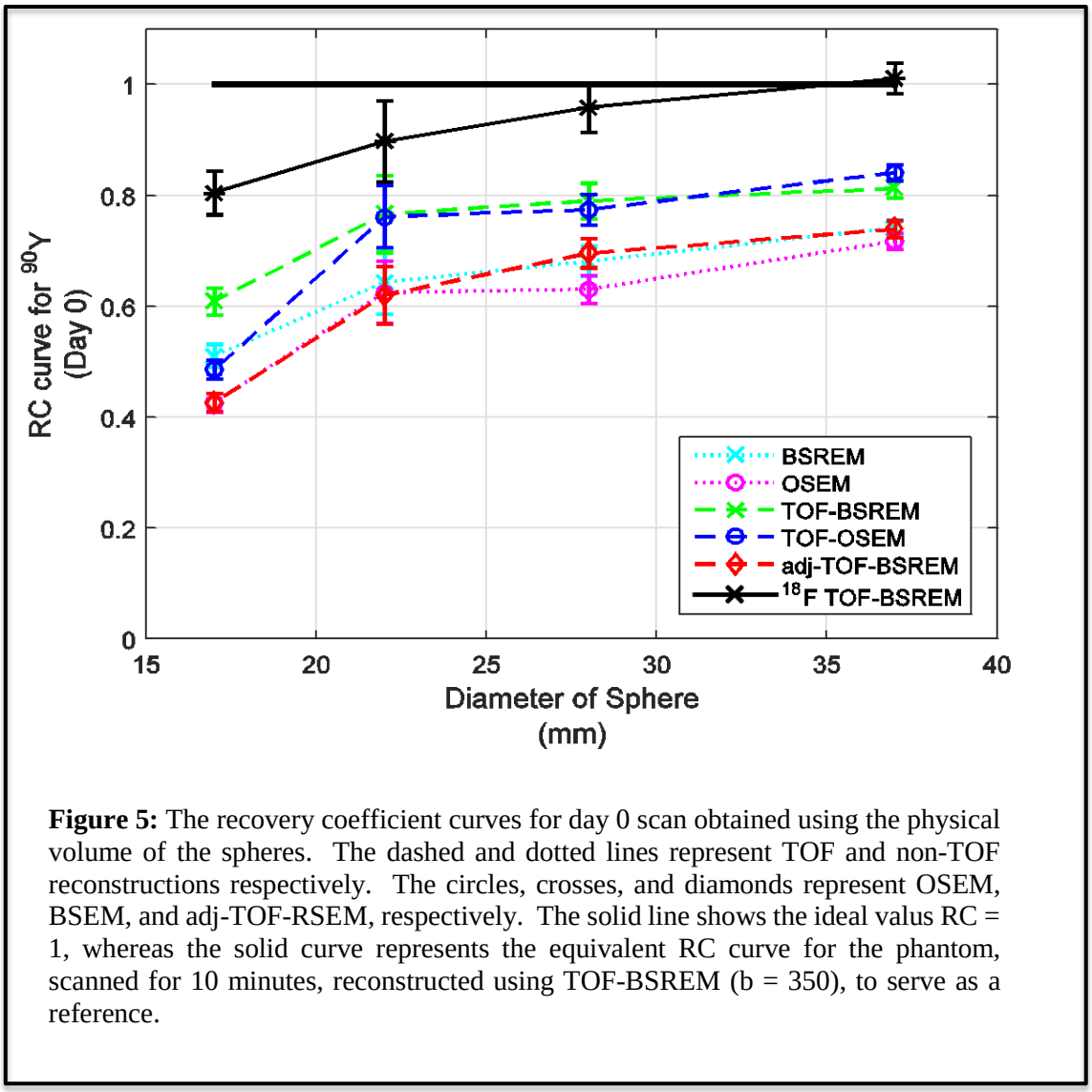


Figure 5: The recovery coefficient curves for day 0 scan obtained using the physical volume of the spheres. The dashed and dotted lines represent TOF and non-TOF reconstructions respectively. The circles, crosses, and diamonds represent OSEM, BSEM, and adj-TOF-RSEM, respectively. The solid line shows the ideal valus RC = 1, whereas the solid curve represents the equivalent RC curve for the phantom, scanned for 10 minutes, reconstructed using TOF-BSREM (b = 350), to serve as a reference.

§ Monte Carlo Simulations

Additionally, Monte Carlo (MC) simulations using GATE v7.1 software were performed in order to investigate the relationship between quantification accuracy and system spatial resolution for objects with different characteristics (sizes, shapes, activity levels, SBR…). In this case, simulations are particularly useful, because experimentally study of numerous sizes, shapes and locations of objects would be very time consuming and inefficient.

Hence, our GE Discovery 690 PET imaging system has been used in these MC experiments. The modeled system geometry and data acquisition process have been tested using simulations of two line sources filled with ^{18}F activity positioned in air and in water

filled cylindrical phantom. Additionally, techniques for generating voxelized objects with different shapes have been developed.

Finally, for experiments aiming to evaluate the quantification accuracy of different objects in clinical studies, we propose to create hybrid data sets which will combine simulated (known) lesions with clinical patient scans. To achieve this, we have already developed a technique to merge simulated data, from GATEv7.1, with patient datasets retrieved from our clinical system. We successfully reconstructed such hybrid dataset with the OSEM algorithm obtained from STIR (Open Source software for use in tomographic imaging) [9].

§ Tumor Lesion Segmentation from 3D PET using a Machine Learning driven Active Surface

We developed a fully automatic method for lesion delineation, which does not require user-initialization or parameter-tweaking, to segment novel PET images. To achieve this, we trained a machine learning system on anatomically and physiologically meaningful imaging cues, to distinguish normal organ activity from tumorous lesion activity. The inferred lesion likelihoods are then used to guide a convex segmentation model, which guarantees reproducible results. We evaluated our approach on datasets from The Cancer Imaging Archive trained on data from the Quantitative Imaging Network challenge that were delineated by multiple users. Our method produces more accurate segmentation than state-of-the art segmentation results, and does so without user interaction. We published this work in the peer-reviewed Machine Learning for Medical Imaging workshop, a satellite event of the 2016 International Medical Image Computing and Computer Assisted Intervention conference the premier conference in the area of medical image analysis [10].

§ Multi-site 3D FDG PET Segmentations study

Our team contributed to this study, which led to a publication in Medical Physics Journal titles “Multi-site Quality and Variability Analysis of 3D FDG PET Segmentations based on Phantom and Clinical Image Data”. To assess PET segmentation quality and consistency at the multi-institutional level, we were part of the study of seven institutional members of the National Cancer Institute Quantitative Imaging Network. For the study, members (including our team) were asked to segment a common set of phantom PET scans acquired over a range of imaging conditions as well as a second set of head and neck cancer (HNC) PET scans. Segmentations were 55 generated at each institution using their preferred approach. In addition, participants were asked to repeat segmentations with a time interval between initial and repeat segmentation. This procedure resulted in overall 806 phantom insert and 641 lesion segmentations. Subsequently, the volume was computed from the segmentations and compared to the corresponding reference volume by means of statistical analysis [11].

§ Tumor Volume and Metabolic Activity Estimation from PET Scans

We developed a PET quantification to estimate the volume and activity of the lesions which are basic measures needed for other important quantification metrics such as

standardized uptake value (SUV) and total lesion glucose (TLG). For validation, we used a set of 55 PET scans of the Elliptical Lung-Spine Body Phantom™ with different levels of noise, four different reconstruction methods, and three different background activities, namely air, water, and hot background. Our preliminary results are very promising, showing a relative absolute error (RAE) of 5.11% ± 3.5% and 5.7% ± 5.25% for volume and activity estimation, respectively, which represented improvements of over 20% and 6% respectively, compared with the best competing methods. We are currently finalizing the results and the write-up of this work to be submitted for publication within the next 2 months.

§ OPTIMIZING circulating tumor DNA (ctDNA) measurements for quantitative tumour tracking

Measuring ctDNA abundance using OnTarget

OnTarget (Boreal Genomics) is a highly sensitive multiplexed mutation detection platform. This technology pre-enriches DNA for mutant alleles prior to sequencing to enhance sensitivity. A 96-plex assay that covers common hotspot mutations affecting numerous genes including KRAS, TP53, and PIK3CA was used to analyze ctDNA levels and explore the relationship between ctDNA and various clinical characteristics in multiple patient cohorts. Specifically, we applied OnTarget to tumours and plasma samples collected pre- and post-operatively from a cohort of early-stage pancreatic ductal adenocarcinoma (PDAC). We detected mutations in *KRAS* or *GNAS* in 29 of 32 PDAC tumours. Using OnTarget, we then detected concordant ctDNA in 8 of 25 cases of pre-operative plasma and 5 of 22 post-operative cases. The presence of ctDNA in post-operative samples was significantly associated with shorter recurrence free survival time ($p < 0.001$). We also applied OnTarget to a diverse cohort of melanoma, lung, and colorectal cancer cases and found that mutations concordant with those observed in matched archival tumours could be detected in a high proportion of cases. Discordant mutations were observed in 15% of cases but were not associated with greater time between tumour and archival blood draw nor difference in survival.

Measuring ctDNA abundance using ddPCR

Earlier this year, we published our new digital drople PCR (ddPCR) assay that assess mutations at two common hot spots in non-Hodgkin lymphoma (NHL) [12], namely those affecting EZH22 and STAT6. Each hot spot mutation has four common alleles and each allele requires a separate TaqMan assay. We showed that each performs well on dilutions of cell-line and tumour DNA containing specific mutations. We also demonstrated the utility of these on plasma samples from lymphoma patients with compatible mutations and have detected ctDNA levels as low as one mutant copy in a background of 3000 wild type molecules. These assays were used to aid in detecting ctDNA in DLBCL cases with levels below the sensitivity of our other assays. Some of these results were used in a paper describing a clinical trial on DLBCL published earlier this year [13].

Profiling ctDNA using hybridization capture with molecular barcoding

A recent method to quantify ctDNA that may offer greater sensitivity as well as other benefits, termed CAPP-Seq, has recently been described and we have made enhancements to this style of approach in our lab [14]. In general, this style of approach uses biotinylated DNA baits spanning previously identified somatic mutations that are ordered from a commercial manufacturer (Integrated DNA technologies) or generated in-house through a custom method. DNA baits are, during targeted hybridization capture experiments, aimed at the enrichment of cfDNA libraries in our targets of interest. Specifically, we have implemented a novel molecular barcoding approach that incorporate a set of barcoding DNA adapters to tag every cfDNA molecule initially present in the blood sample at the ligation step. This allows error-correction and "duplex" sequencing to combat errors deriving from DNA damage. Using this method, we have investigated ctDNA levels in 41 blood samples drawn from 40 personalized oncogenomics study (POG) patients, including 4 adult and 36 pediatric cases. We conducted non-invasive personalized assays targeting a priori known mutations, as determined by the sequencing of tumour and normal samples, in 34 patients diagnosed with a broad variety of disease conditions (see Table 2). Commercially available gene panels or in-house generated pools of molecular probes targeting recurrently mutated regions of the genome were used in 6 patients in an effort to directly detect cancer-related genetic aberrations from blood. Personalized assays relying on digital PCR and targeted hybrid capture experiments coupled with next-generation sequencing (NGS) revealed the presence of ctDNA in 62.5% of the cases investigated. The abundance of ctDNA in blood was highly variable between samples, from allele frequency ratios of mutant DNA versus wild-type DNA below 0.1% in osteosarcoma and brain tumour cancers to more than 50% in neuroblastomas and rhabdomyosarcomas. We are obtaining additional samples from some of these patients and will compare ctDNA levels to existing PET-CT data where available.

POG ID	DISEASE	TARGETED LOCI	VAF RANGE (ctDNA)
POG-064	NEUROBLASTOMA	5-May	0.029 - 0.543
POG-067	NEUROBLASTOMA	1-Jan	0.378
POG-062	NEUROBLASTOMA	2/4 + Transl	0.107 - 0.212 (SNVs)
POG-173	NEUROBLASTOMA	5-Apr	0.20-0.309
POG-589	NEUROBLASTOMA	TOP PANEL	ALK AMPLIFICATION
POG-156	NEUROBLASTOMA	TOP PANEL	0
POG-020 (V1)*	OVARIAN GRANULOSA	6-Jan	0.0006
POG-020 (V2)*	OVARIAN GRANULOSA	6-Apr	0.255 - 0.298
POG-106*	OVARIAN CARCINOMA	5-May	0.09-0.025
POG-184*	CLEAR CELL CARCINOMA OF OVARY	1/1 (ddPCR)	0.009
POG-608*	ADENOCARCINOMA OF LUNG	6-May	0.001-0.002
POG-047	INFANTILE FIBROSARCOMA	1/1 Transl	0.0058
POG-524	EWING'S SARCOMA	PAN-CANCER PANEL	0.608
POG-194	OSTEOSARCOMA	4-Mar	0.0005 - 0.001
POG-642	OSTEOSARCOMA	6-Feb	0.006-0.007
POG-144	OSTEOSARCOMA	0/5	0
POG-079	SARCOMA	1/1 Transl	0.0016
POG-651	RHABDOMYOSARCOMA	6-Apr	0.01-0.61
POG-540	NUT MIDLINE CARCINOMA	3/5 + Transl	0.0013-0.003
POG-288	PINEOBLASTOMA	1-Jan	0.0007
POG-533	GLIOBLASTOMA	6-Feb	0.001-0.015
POG-145	FIBROVASCULLAR BRAIN TUMOUR	1/1 (ddPCR)	0.0008
POG-159	CRANIOPHARYNGIOMA	0/1	0
POG-172	PAPILLARY THIROID CARCINOMA	0/3	0
POG-168	MALIGNANT GRANULLAR CELL TUMOUR	1-Jan	0.001
POG-146	ANGIOSARCOMA OF LIVER	5-Apr	0.001-0.008
POG-565	ANGIOSARCOMA	6-Mar	0.003-0.012
POG-499	ACUTE LYMPHOBLASTIC LEUKEMIA	PAN-CANCER PANEL	0.358-0.436
POG-161	DIFFUSE LARGE B-CELL LYMPHOMA	5-Apr	0.223-0.290
POG-564	HODKING LYMPHOMA	AN-PAN-CANCER PANEL + LYMPHOMA PANE	0
POG-380	HODKING LYMPHOMA	PAN-CANCER PANEL + LYMPHOMA PANEL	0.014-0.083
POG-407	NEUROFIBROMATOSIS	0/4; PAN-CANCER PANEL	0
POG-417	NEUROFIBROMATOSIS	0/4; PAN-CANCER PANEL	0
POG-240	AGGRESSIVE FIBROMATOSIS	PERSONALIZED ASSAYS - 6 PROBES	In Progress
POG-531	EPENDYMOMA	PERSONALIZED ASSAYS - 6 PROBES	In Progress
POG-541	RHABDOID TUMOUR	PERSONALIZED ASSAYS - 6 PROBES	In Progress
POG-659	NEUROBLASTOMA	PERSONALIZED ASSAYS - 6 PROBES	In Progress
POG-355	GLIOBLASTOMA	PERSONALIZED ASSAYS - 6 PROBES	In Progress
POG-454	GLIOMA	PERSONALIZED ASSAYS - 6 PROBES	In Progress
POG-644	NEUROBLASTOMA	PERSONALIZED ASSAYS - 6 PROBES	In Progress
POG-629	GLIOMA	PERSONALIZED ASSAYS - 6 PROBES	In Progress

Table 2: ctDNA analysis of 41 blood samples drawn from 40 personalized oncogenomics studies (POG).

We have also demonstrated that the de-novo discovery of somatic mutations directly from blood is possible in those cases where tumor biopsies are hard to collect or when low tumor content deems these biopsies unsuitable for next-generation sequencing. We reported several mutations directly from plasma in a Hodgkin lymphoma patient (POG-380) after using a commercial gene panel targeting 128 cancer-related genes. We also uncovered compelling evidence of ALK amplification in the plasma of a neuroblastoma patient (POG-589). ALK status in this patient was unknown and our findings could have encouraged the treatment of this particular patient with crizotinib, a drug that specifically targets and inhibits ALK activity. The evolution of ctDNA levels could only be evaluated in a patient diagnosed with cancer of the ovary (POG-020). We observed a dramatic increase in ctDNA between two plasma samples drawn with a difference of 9 months, a finding that was interpreted as a clear indicator of disease progression in this patient. Through the expansion of our project into the adult POG population we expect to greatly increase the application of this approach over the coming year.

§ EF5 tracer for diagnosing hypoxia in head and neck squamous cell carcinomas

This is a pilot study to assess feasibility, safety and use of 18F-EF5 PET/CT at BCCA-Vancouver in untreated patients with head and neck squamous cell carcinomas planned for radical radiation therapy. The objective was to quantify hypoxia and compare differential uptake of the EF5 tracer before and after treatment in 20 patients (13 patients had pre-therapy and post-therapy PET/CT so far), using as a positivity threshold a tumor-to-muscle ratio ≥1.50 (SUVpeak of the tumor divided by average SUV of the contralateral muscle). There was no adverse reaction to EF5 administration. Preliminary results show that PET imaging with EF5 is feasible, safe, and that images of adequate quality can be obtained. Our data indicate that the hypoxia present in most primary tumors and metastatic lymph nodes on the initial EF5 PET/CT study regresses post radiation therapy. See Figures 6, 7, 8, and 9. Further analysis of data pending.

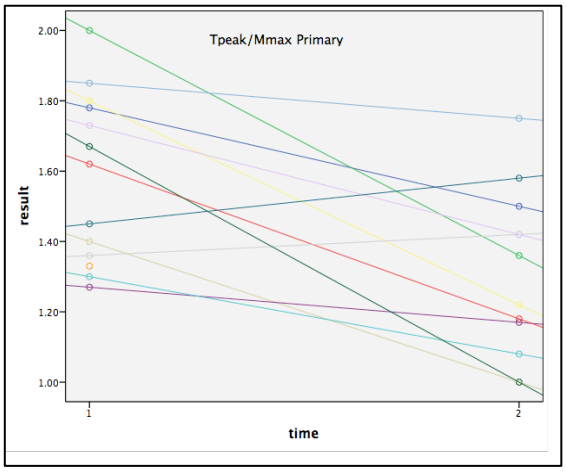


Figure 6: Quantification of primary tumour to muscle uptake ratio of EF5 PET/CT before treatment (time point 1) and after treatment (time point 2).

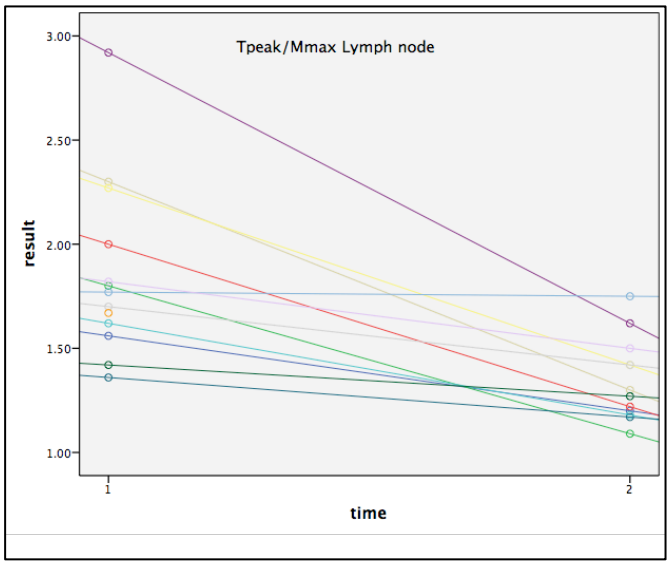


Figure 7: Quantification of lymph node metastasis to muscle uptake ratio of EF5 PET/CT before treatment (time point 1) and after treatment (time point 2).

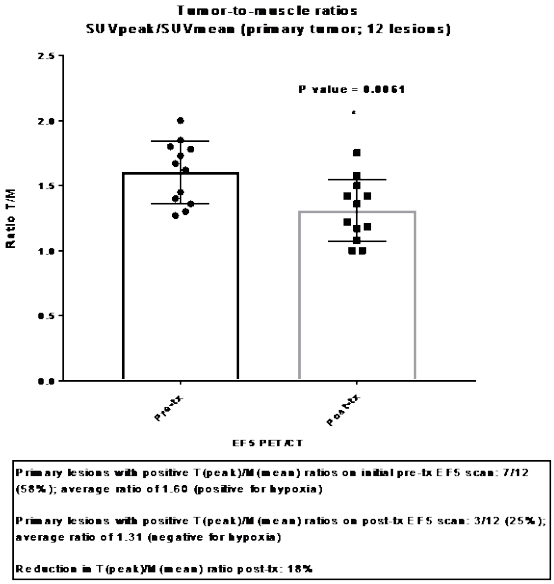


Figure 8: Comparison of pre-treatment SUV_{peak}/SUV_{mean} to post-treatment values for the primary tumour.

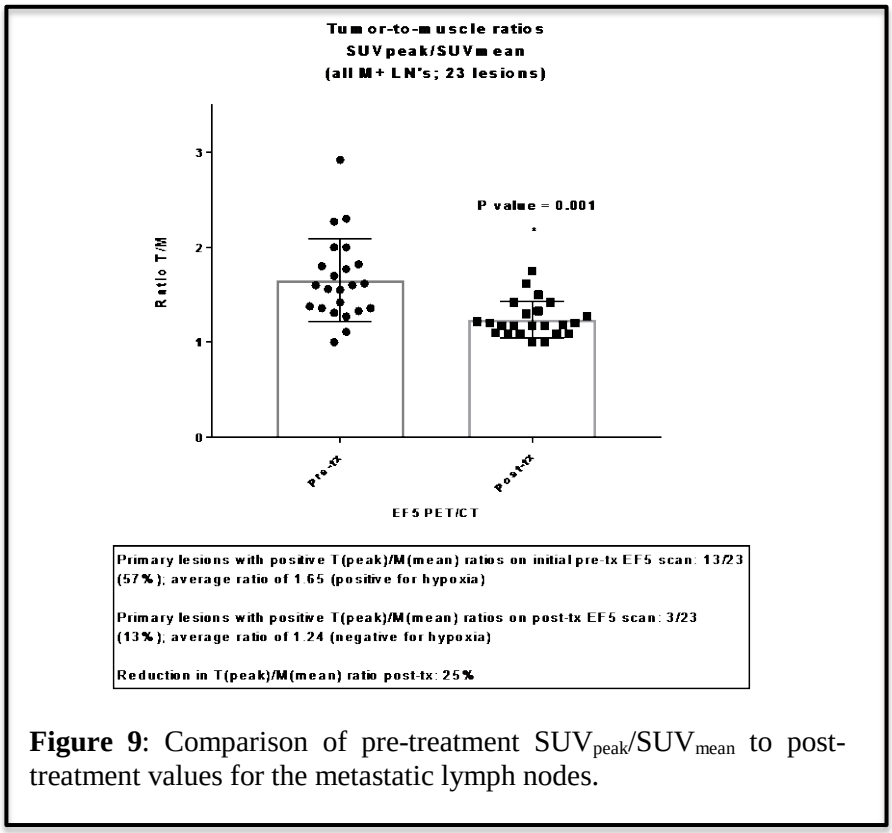


Figure 9: Comparison of pre-treatment SUV_{peak}/SUV_{mean} to post-treatment values for the metastatic lymph nodes.

COLLABORATIONS WITHIN THE NETWORK

The images obtained in all these studies were shared with the Computer Science team of Dr. G. Hamarneh and were used in their segmentation and machine learning tests.

PLANS FOR NEXT YEAR

These studies will be expanded onto investigation of the relationship between image spatial resolution (using the MTF) and quantification accuracy for object with different sizes and shapes, as well as for different levels of signal-to-background ratio (SBR) and activity distributions. In particular, we plan to investigate the behavior of different algorithms for objects with different characteristics. Because non-linearity and non-shift invariance of the iterative algorithms, it is expected that their behavior will depend on the size and shape of object that is being imaged and on its position in the field of view of the camera. In parallel, MC simulations of lesions incorporated into the patient datasets will continue. Quantitative accuracy of these data will be tested for different reconstruction algorithms.

We also will continue to collect matched sets of PET/CT data and blood samples in patients enrolled in several clinical trials.

To that end, we have the following clinical trials scheduled to start in 2017:

i. **"Evaluation of the relationship of levels of circulating ctDNA in plasma to the presence of detectable disease on 18F-FDG PET/CT for metastatic colorectal carcinoma with a comparison to classical biomarkers"**. This trial aims to establish correlation of ctDNA marker levels with tumour burden quantification techniques such as metabolic tumour volume (MTV), total lesion glycolysis (TLG), as well as new segmentation algorithms being developed. Comparison will also be made to standard criteria such as RECIST1.1 and PERCIST.

ii. **"Phase II Trial: Evaluation of the safety and efficacy of 68Ga-DOTATOC PET/CT for imaging NET patients"** is a trial aimed at the evaluation of a new tracer for NET patients and has optional sub-studies that have quantitative imaging and ctDNA endpoints:
    a. **"Quantification of change in planned therapy as result of 68Ga-DOTATOC imaging and evaluation of prognostic value of early imaging"**. This sub study evaluates quantitative imaging parameters at 12-week to predict 40-week progression free survival (PFS). Quantitative reduction of tumour burden at 12-weeks, as evaluated by MTV, TLG and new segmentation algorithms (in development), will also be compared with 40-week PFS in patients initiating systemic therapy).
    b. **"Development of novel quantitative NET ctDNA markers from archival DNA of biopsy specimens already collected by pathology and serial blood samples"**; This project aims to collect blood samples at imaging time points of the main project and to develop ctDNA assays that target mutations specific for NET patients.

iii. **"A prospective two-arm study of the efficacy and safety or 177Lu-DOTATATE for treatment of patients with SSR positive NETs"**; this trial has quantitative imaging endpoints:
- Correlation of ctDNA levels to tumour burden on 68Ga-DOTATOC, as evaluated by MTV and TLG equivalents for that tracer.
- Reduction in SUVmax between baseline 68Ga-DOTATOC PET/CT and scan done at 4-month post treatment will be compared with proportion of each category of RECIST1.1 response on CT at 6 months [1, 2, 12 PD].

iv. **"A single-blind study to evaluate the efficacy and safety of 18F-Fluorodeoxygalactose (18F-FDGal) compared to 18F-Fluorodeoxyglucose (18F-FDG) to detect hepatocellular carcinoma via PET/CT in patients with cirrhosis or chronic liver disease."**; This is a new tracer that will allow, akin to FDG, quantification of tumour metabolism and calculation of MTV.

PUBLICATIONS AND PRESENTATIONS

Tanguay, J., Ma, H., Hou, X., Benard, F., & Celler, A. (2016). An experimental framework for task-based assessment of spatial resolution in quantitative PET/CT. *Journal of Nuclear Medicine*, *57*(supplement 2), 1940-1940.

Ma, H., Ross, S., Iwata, K., Stearns, C., Benard, F., & Celler, A. (2015). Regularized iterative reconstruction of 90Y PET; a phantom study. *Journal of Nuclear Medicine*, *56*(supplement 3), 564-564.

LIST OF REFERENCES

1. Antunovic, L., et al., *Standardization and quantification in PET/CT imaging: tracers beyond FDG.* PET Clin, 2014. **9**(3): p. 259-66.
2. Boellaard, R., *Need for standardization of 18F-FDG PET/CT for treatment response assessments.* J Nucl Med, 2011. **52 Suppl 2**: p. 93S-100S.
3. Richard, S., et al., *Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms.* Med Phys, 2012. **39**(7): p. 4115-22.
4. D'Arienzo, M., *Emission of β+ Particles Via Internal Pair Production in the 0+ – 0+ Transition of 90Zr: Historical Background and Current Applications in Nuclear Medicine Imaging.* Atoms, 2013. **1**(1): p. 2-12.
5. Wahl, R.L., et al., *From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors.* J Nucl Med, 2009. **50 Suppl 1**: p. 122S-50S.
6. Willowson, K.P., M. Tapner, and D.L. Bailey, *A multicentre comparison of quantitative (90)Y PET/CT for dosimetric purposes after radioembolization with resin microspheres : The QUEST Phantom Study.* Eur J Nucl Med Mol Imaging, 2015. **42**(8): p. 1202-22.
7. Asma, E., et al., 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC), 2012(M23-7): p. 4062-4066.
8. Ahn, S., et al., *Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET.* Phys Med Biol, 2015. **60**(15): p. 5733-51.
9. Thielemans, K., S. Mustafovic, and C. Tsoumpas, *STIR: Software for tomographic image reconstruction release 2.* IEEE Nuclear Science Symposium Conference Record, 2007. **4**: p. 2174-2176.
10. Ahmadvand, P., et al., *Segmentation from 3D PET using a Machine Learning driven Active Surface.* Medical Image Computing and Computer-Assisted Intervention Workshop on Machine Learning in Medical Imaging (MICCAI MLMI), 2016. **10019**: p. 271-278.
11. Beichel, R., et al., *Multi-site Quality and Variability Analysis of 3D FDG PET Segmentations based on Phantom and Clinical Image Data.* Medical Physics, Accepted 2016.
12. Alcaide, M., et al., *Multiplex Droplet Digital PCR Quantification of Recurrent Somatic Mutations in Diffuse Large B-Cell and Follicular Lymphoma.* Clin Chem, 2016. **62**(9): p. 1238-47.

13. Assouline, S.E., et al., *Phase 2 study of panobinostat with or without rituximab in relapsed diffuse large B-cell lymphoma.* Blood, 2016. **128**(2): p. 185-94.
14. Newman, A. M. et al. *An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage*. Nat Med (2014). doi:10.1038/nm.3519

U01CA190256: Image-based quantitative assessment of tumor hypoxia

University Health Network/Lawson Health Research Institute

David Jaffray Ph.D., Ting-Yim Lee Ph.D., Ivan Yeung Ph.D., Edward Taylor Ph.D., Harald Keller Ph.D., Catherine Coolens Ph.D., Jennifer Gottwald M.S., Brandon Driscoll M.S., Tina Shek M.S., Julia Publicover M.S., Viktor Iakovenko Ph.D.

INTRODUCTION

Hypoxia in solid tumours correlates strongly with the presence of metastases and leads to enhanced resistance to radiation and chemo therapies [1, 2]. There is a strong clinical need to reliably determine the location and extent of hypoxia in order to provide targeted therapies. We seek to develop quantitative, multi-parametric approaches to hypoxia imaging to increase the predictive capacity of the hypoxia markers and improve the stratification of patients for hypoxia-targeted treatment strategies.

This work includes several aims, including the development of standardized acquisition methodology, integrating perfusion imaging methods to create a more robust tracer kinetic model for hypoxia imaging, and developing a software application to solve these models and produce quantitative metrics of hypoxia. These developments will be validated in on-going clinical trials, some of which include oral pimonidazole to produce a histology gold standard against which to compare the imaging results. In future years, we will evaluate the predictive capacity of the imaging metrics. On Figures 7-9 the detailed timeline for these activities are presented.

Our aim to standardize hypoxia imaging protocols will provide a guideline for the imaging community to design clinical studies in hypoxia imaging with PET tracers. The development of advanced hypoxia tracer models coupled with perfusion will provide an understanding of the interplay between hypoxia and perfusion in tumors. These models can be readily adapted by other researchers in the imaging community. In addition, analysis of clinical studies of different anatomical sites will provide baseline data which can also be used for the design of future clinical trials.

DISCUSSION OF PROGRESS

§ Specific aim 1. Establish a robust and reliable methodology for PET hypoxia imaging

Sub-aim 1.1: Quantifying hypoxia using static PET imaging

PET imaging with F-18-labelled hypoxia-sensitive tracers such as fluoroazomycin arabinoside (FAZA) has emerged as a promising non-invasive way of detecting hypoxia in tumours. A key challenge in using static PET data to quantify hypoxia is that the activity of tracer in a region of interest (ROI) is sensitive not only to the presence of hypoxia, but also to transport properties – perfusion, diffusion, permeability, and blood volume – which vary from

voxel-to-voxel and patient-to-patient. Compartmental modelling of dynamic PET data has the potential to correct for these properties, enhancing sensitivity of PET imaging to hypoxia.

A compartmental model based on a reaction-diffusion equation was used to study FAZA pharmacokinetics (binding and transport). Assuming that local diffusive equilibrium is achieved rapidly within the ROI, a simple expression was derived for the tracer activity in terms of the arterial input function (AIF), the area under the curve (AUC) for the diffusive (unbound) compartment, and a quantity K3 which is argued to be proportional to the volume fraction of space in which the local oxygen tension is below ~ 10 mmHg; i.e., hypoxic. This expression was used to study PET data from twenty patients with pancreatic cancer who were injected with FAZA. Comparing activities in tumour ROIs with values taken from two choices of reference tissue, spinal muscle and blood, information about the sensitivity of PET imaging to transport inhomogeneities was quantified within the framework of our transport model.

Dividing activity in the ROI by a reference value taken from the same patient leads to a reduction in the sensitivity to inter-patient transport inhomogeneities: differences in blood volumes, clearance rates, and time post-injection at which the PET scan is taken. Because it is poorly perfused, spinal muscle exhibits significant variability in the uptake of FAZA. In contrast, using blood as a reference leads to a substantial reduction in the sensitivity to transport inhomogeneities, as shown in the middle panel of Figure 1, where the AUC divided by the AIF—directly related to the hypoxic proportion – exhibits a reduced variance as compared to the AUC (left) and the AUC divided by the activity in the spinal muscle (right).

By dividing the measured tracer activity in a region-of-interest contained inside a tumour by a reference value taken from the same patient, PET measurements are sensitive to the presence of hypoxia. The choice of blood as a reference tissue optimizes this sensitivity although our work also delineates the circumstances under which muscle can also reliably be used.

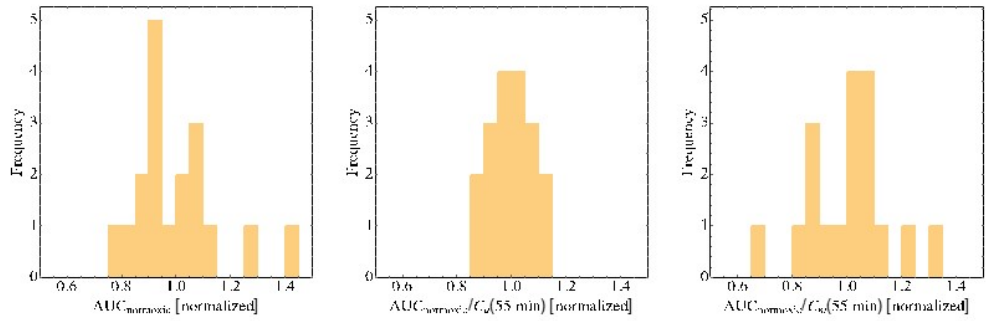


Figure 1: Normalized AUC for nineteen pancreas tumours.

Sub-aim 1.2: Quantifying hypoxia using dynamic PET imaging

In the past year, we have developed novel compartmental models to study FAZA-hypoxia PET imaging in Princess Margaret Cancer Centre patients with pancreatic ductal adenocarcinoma (PDAC) [3]. The goal of this research program was to develop a reliable

analysis method to quantify hypoxia from dynamic PET imaging and to compare the results to values obtained from static PET imaging [3] and immuno-histochemical staining of resected pancreas tumours. Our major results so far are:

1. Static PET imaging of FAZA at two hours after injection reliably quantifies hypoxia as long as the imaged tissue is devoid of substantial necroses, ductal lumen, or fat, and the metric used for hypoxia quantification is the tumour-to-blood FAZA uptake ratio and not the tumour-to-muscle value [3].

2. For tumours exhibiting necroses, ductal lumen (as in the case of PDAC), or necroses, a novel compartmental model of dynamic PET data was developed to correct for the impaired uptake of tracer into these regions. For tumours exhibiting substantial ductal lumen, the hypoxic fractions calculated from this scheme differed appreciably from those calculated using static PET imaging.

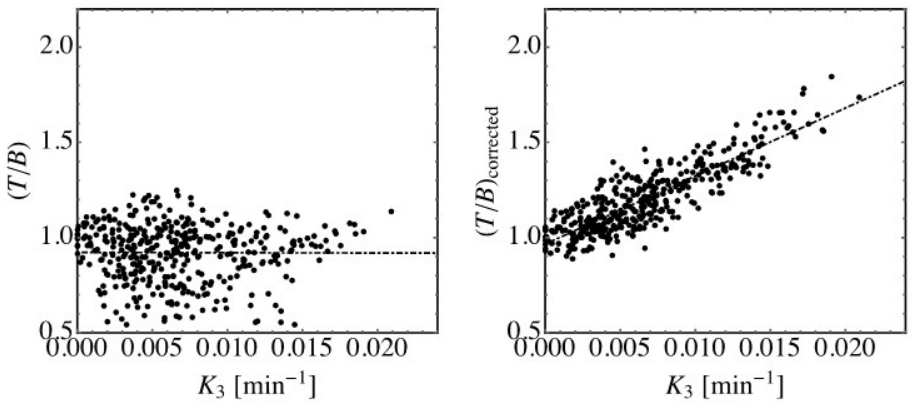


Figure 2. Left: FAZA tumour-to-blood uptake ratio (T/B) versus the net trapping rate K_3 for all voxels in a representative tumour. Right: Uptake ratio corrected for partitioning.

The effect of having necroses, ductal lumen, or fat present in the imaged tissue can be understood from Figure 2. The left panel shows the FAZA tumour-to-blood uptake ratio at two hours (static PET image) versus the FAZA "trapping rate" K_3 derived from a compartmental model analysis of voxel-scale dynamic PET data for a single tumour. The correlation between these two quantities is weak since the presence of either necroses, ductal lumen, or fat represent spatially inhomogeneous regions in which tracer is slow to reach diffusive equilibrium; i.e., it is "partitioned". This leads to variability in the voxel-scale FAZA uptake that overwhelms the signal arising from tracer bound by hypoxia, which is sensitive to K_3.

The degree to which tracer is partitioned in a given voxel was quantified by a compartmental model that treats slow-equilibrating regions as a separate compartment. Using this model to correct for partitioning gives the tumour-to-blood uptake ratio that would have arisen in a voxel had there been no partitioning: this is shown in the right panel of Figure 2. The strong correlation between voxel-scale trapping rates and uptake values corrected for partitioning

as compared to the weak correlation between the uncorrected values shows the substantial effect that partitioning has on static PET imaging of hypoxia.

Our model also distinguishes the FAZA binding rate—a direct measure of radiobiologically relevant levels of hypoxia—from the rate of equilibration in the trapping rate in K_3, allowing us to quantify hypoxia using the former quantity.

Sub-aim 1.3: Tracer kinetic models for dynamic PET imaging analysis

In the past year we have extended the conventional closed three-compartment model (Figure 3 (A)) to include the perfusion through the vasculature of the tissue (Figure 3 (B)). Unlike the conventional closed three-compartment model, the modification with the Johnson-Wilson-Lee model allows for determination of blood flow if the arterial ($C_a(t)$) and tissue time-activity curves are measured with sufficient time resolution.

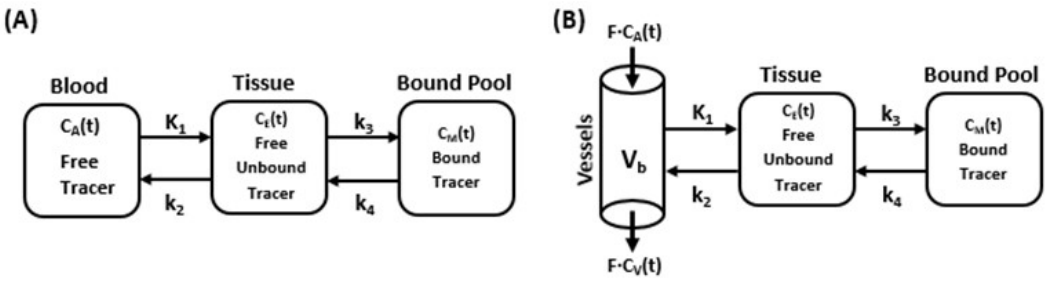


Figure 3. Kinetic models to describe the transport of tracer from blood to uptake in tissue. (A) conventional closed three compartment model where the transport through the vasculature is not modeled. (B) the modified closed three compartment model by using the Johnson-Wilson-Lee model to replace the blood compartment. K_1 is the transfer rate of tracer from blood to tissue, k_2 is the backflux rate constant from tissue to blood, k_3 is binding rate constant of tracer to target in the tissue, and k_4 is the dissociation rate constant from the target, F is blood perfusion and V_b is the tissue blood volume.

To evaluate how the modified three compartment model (Figure 3(B)) can be used to differentiate the uptake kinetics of different tracers, as a prelude to testing it on hypoxia tracer, we analyzed the time-activity curves of prostate cancer (PCa) obtained with two different tracers: ^{18}F-FCH and ^{18}F-DCFPyL- the first one is a marker for lipogenesis in PCa while the second one is a ligand for prostatic membrane specific antigen. The rationale to choose these two tracers is that PET imaging with 18F-DCFPyL has been used to localize and detect prostate cancer (PCa) nodules with high contrast to background normal prostatic tissue. In contrast many studies with 18F-fluorocholine (18F-FCH), which previously were widely used for imaging PCa, showed that choline uptake is not always higher in tumor region. This PET imaging difference provided a good test case to investigate whether their kinetic behavior in PCa as modeled by Figure 3(B) is different. If the kinetic model passes this test, then our next step would be to image with hypoxia tracer.

Two groups of seven patients each underwent dynamic PET imaging with either ^{18}F-FCH or ^{18}F-DCFPyL. The dynamic data from each group was analyzed using the modified model. Figure 4 shows the ^{18}F-DCFPyL time-activity curves of PCa and normal prostatic tissue from a patient and the corresponding model fits. Table 1 gives the model parameter values for the time-activity curves shown in Figure 4.

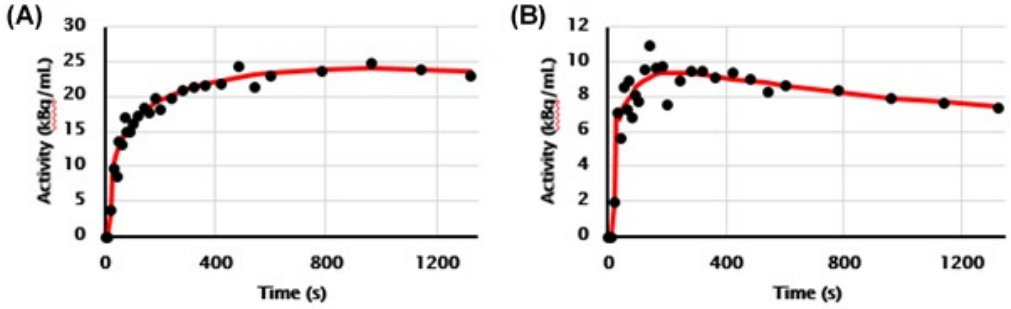


Figure 4. ^{18}F-DCFPyL time-activity curves from a PCa patient. (A) Tumor time-activity curve and model fit. (B) Normal prostatic tissue time-activity curve and model fit. Measured time-activity curve is shown as black dots while model fit curve as red line.

	F (mL/min/g)	V_b (mL/g)	K_1* (mL/min/g)	k_2 (1/min)	k_3 (1/min)	k_4 (1/min)	SUV_{avg}* (g/mL)	K_i* (mL/min/g)	V_e (mL/g)
Tumor	0.43	0.074	0.41	0.53	0.230	0.048	5.36	0.087	2.28
Normal	0.37	0.038	0.27	0.33	0.070	0.080	2.04	0.034	0.91

Table 1. Model parameter values for PCa and normal prostatic tissue from a patient. Statistical significant difference ($P<0.05$) is indicated by *.

Table 2 shows the model parameter values for ^{18}F-FCH and ^{18}F-DCFPyL in PCa and normal prostatic tissue from seven patients in each tracer group. The normalized washout rate constant from the bound pool, as estimated by the inverse of binding potential (k4/k3), of ^{18}F-DCFPyL from normal tissue was greater than tumour while both normal tissue and tumour had similar normalized washout rate constant for ^{18}F-FCH. The binding rate constant (k3) of ^{18}F-FCH was higher than ^{18}F-DCFPyL for both normal tissue and tumour.

These results suggest that the ^{18}F-DCFPyL contrast between tumour and normal tissue is due to the differential normalized washout. In contrast, the lack of ^{18}F-FCH contrast between tumour and normal tissue is due to similar normalized washout. The large binding rate constant of ^{18}F-FCH vs ^{18}F-DCFPyL suggested that the former has a faster uptake rate and at time interval when binding dominates, SUV could be used to differentiate sensitively tumour from normal tissue.

	Tumor			'Normal'		
	k_3 (min^{-1})	k_4 (min^{-1})	$\frac{k_4}{k_3}$	k_3 (min^{-1})	k_4 (min^{-1})	$\frac{k_4}{k_3}$
^{18}F-DCFPyL	0.31±0.52	0.09±0.08	0.49±0.55	0.08±0.05	0.10±0.08	1.16±0.74
^{18}F-FCH	0.74±1.20	0.09±0.14	0.18±0.39	0.30±0.52	0.07±0.12	0.10±0.20

Table 2. Mean ± SD model parameter values for ^{18}F-FCH and ^{18}F-DCFPyL in PCa and normal prostatic tissue.

The above analyses demonstrate that the proposed model (Figure 3B) is able to differentiate tracers of disparate in-vivo behavior. This gives confidence that the model has properly modeled the essential processes of the uptake of tracers, including hypoxic tracers. We will start application of the model to hypoxic tracers as the next step.

Sub-aim 1.4: Measurement of the AIF with kinetic analysis of dynamic PET imaging

The resolution of PET imaging is limited, as beta-particles annihilate with electrons at a certain distance (up to 2 mm) from the original vertex. In addition, partial volume averaging and spill-over effects should be taken into account in order to recover the true radioactivity concentration in the blood flowing through the artery selected. For this purpose, a flow phantom has been built to calibrate and standardize the imaging protocols for a PET scanner. First measurements are planned to carry out in February 2017.

Preliminary studies to investigate the accuracy and robustness of magnitude and phase-derived arterial input function (AIF) in PET-MR were performed. The results were compared to the "gold standard" volumetric DCE-CT. The impact of individualized magnitude and phase signal AIF measurements on resulting perfusion parameter maps using a common 4D temporal dynamic analysis (TDA) method in metastatic brain cancer patients treated with stereotactic radiosurgery was performed. This data highlights the stability of DCE-CT calculations as well as susceptibility of DCE-MRI K_{trans} measurements to various imaging factors, including AIF selection and T10 values used in the model. Using the same voxel-based analysis platform for both DCE-CT and MR significantly improved correlation values confirming the need to take into account tumor heterogeneity when assessing functional data.

Sub-aim 1.5: Monte Carlo models of dynamic PET

Monte Carlo simulations are used to evaluate scatter correction for quantitative PET imaging of hypoxia. Photon scattering contributes significantly to the imaging degrading effects in 3D PET imaging. It results in a loss of contrast and overall image quality which makes accurate tracer quantification challenging. The effects of scatter are particularly important in regions where two adjacent tissues have vastly different tracer concentrations (see Figure 5). Scattered events from photons originating from the intense uptake region contaminate the low

uptake region. This “cross-talk” changes the linearity, noise level and reconstruction accuracy of PET.

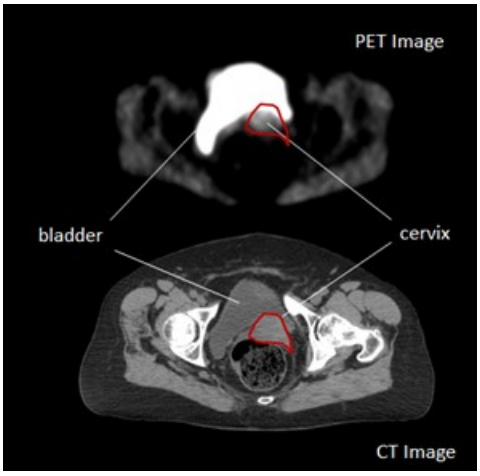


Figure 5. Clinical observation in cervix cancer show a scatter contamination effect in the tumor, originating from the bladder which makes the quantification of hypoxia challenging. UHN REB#: 14-8648-C.

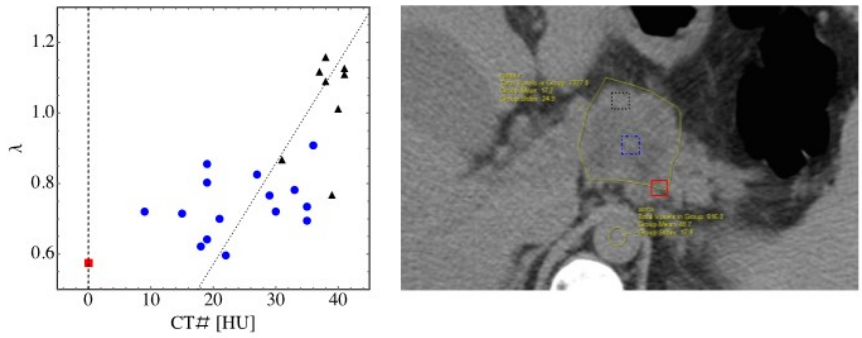


Figure 6. Partition coefficient versus CT number (Hounsfield units) for ROIs in a pancreatic tumours.

The first objective was to develop a physics-based model to simulate the scatter contamination in PET projection data using a Monte Carlo (MC) method. To set up the MC model for PET imaging a MC simulation package in GEANT4 Application for Tomographic Emission (GATE) v.7.2 is used [4]. The geometry of the MC model is a pre-defined GE Discovery 610 PET/CT model and the following technical details were implemented to match our clinical scanner - an energy resolution of 425-650 keV, a dead time of 650 ns and a coincidence window of 4.875 ns. To generate and implement individualized voxelized material phantoms and sources in GATE, regions of interest (ROIs) are defined based on CTAC images. Based on the range of image pixel values, CT numbers are converted into a voxelized material

density map by using pre-defined look up tables in GATE. A similar approach is used to assign individual activity values in the voxelized source. To set up the MC model standardized GE water cylinder quality assurance measurements were used. The MC model allows the scatter distribution component of the signal to be isolated and the image reconstruction process is integrated into the standardized dhPET method. STIR [5] is the current image reconstruction software used. In addition, we have established a research contract with GE Medical through which we will have access to the GE PET Toolbox for PET image reconstruction. The GE PET Toolbox offers the opportunity to be as close as possible to the clinical scanner reconstruction while evaluating the quantitative performance of an accurate concentration recovery.

Sub-aim 1.6: Standardization of imaging technique and characterization of scanner performance

The PET-CT QA procedure, developed based on work from the National Cancer Institute (NCI) and the American College of Radiology Imaging Network, was successfully utilized for the commissioning of PET-MR at Toronto General Hospital. All tests were processed and analyzed on the Siemens molecular MR console using the Siemens NEMA 2007 software. In addition to these NEMA tests, a cross-calibration of the scanner to the dose calibrator reading as well as a daily quality check were performed.

Sub-aim 1.7: Advanced metrics of hypoxia

We have found that the contrasting hypoxic fractions as calculated from the reference muscle of one individual with those calculated by grouping individuals' muscles together, as performed by Mortensen et al. [6], can be susceptible to the noise characteristics of a scanner at specific muscle locations [7]. Hence, the method of grouping individuals is somewhat controversial, and we hope to gain more evidence to help settle the debate at the conclusion of our multi-site FAZA-Metformin trial for cervical cancer patients and FAZA-Prostate trial. For FAZA-Metformin trial in particular, a subsample of patients can volunteer for a blood draw to complement the FAZA-PET imaging results. Also, there will be biopsied tissue samples to complement hypoxic fraction calculations for all enrolled patients after the trial concludes.

The progress chart and milestones for Specific aim 1 is presented on Figure 7.

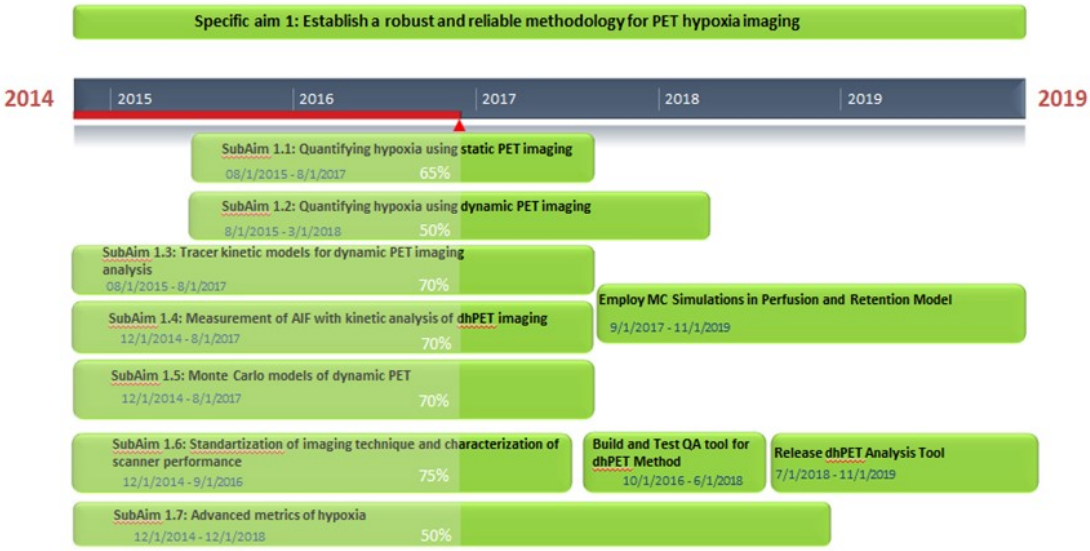


Figure 7: Progress chart with the milestones for the Specific Aim 1.

§ Specific aim 2. Validation of FAZA-PET imaging

Sub-aim 2.1: Pimonidazole correlation in pancreatic cancer

Accrual of pancreas cancer patients suitable for curative-intent surgery to our study of pre-operative FAZA-PET plus pimonidazole staining has continued. Current accrual is 8 patients out of 30.

Sub-aim 2.2: FAZA PET/MR imaging as a biomarker of hypoxia in rectal cancer

The study of stereotactic lung radiation therapy plus surgery hasn't recruited patients for FAZA imaging. As an alternative the images from a FAZA rectum pilot study will be used to achieve the goal of the sub-aim.

The data from a pilot FAZA-rectum trial will be used to measure FAZA uptake against a standard reference and study the correlation of FAZA-PET and blood oxygen level-dependent MRI to pimonidazole staining in locally advanced rectal cancer. The ability to preoperatively predict the patient subpopulation that will respond best to chemoradiotherapy will help to identify the "complete pathological" responders and avoid unnecessary surgery. The primary goal of this pilot trial is to validate FAZA-PET as a biomarker of hypoxia by correlating its uptake in rectal tumors to pimonidazole staining in histopathology specimens. PET imaging also enables quantification of radiotracer uptake which may in future help identify the clinically relevant threshold for hypoxia and identify patients at risk for resistance to CRT. The pilot study is planned for 10 patients. Current accrual is 1 patient out of 10.

The progress chart and milestones for this Specific aim 2 is presented on Figure 8.

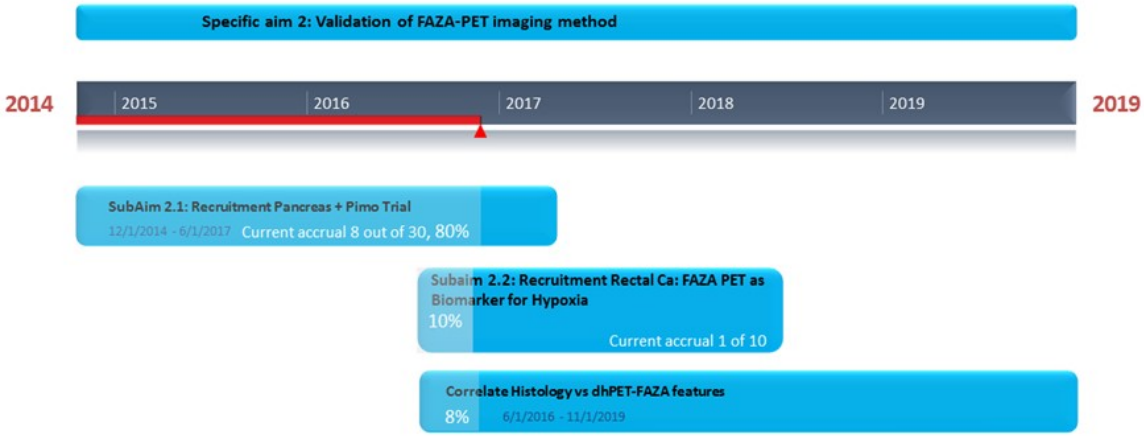


Figure 8: Progress chart with the milestones for the Specific Aim 2.

§ Specific aim 3. Quantitative methods on image-based biomarkers to predict and assess response

Sub-aim 3.1: Data handling and informatics team

Over the past year the Quantitative Imaging for Personalized Cancer Medicine (QIPCM) infrastructure has doubled in size from three to six servers providing computational power for as many as 120 simultaneous virtual desktops. An additional 80 TB storage capacity was allocated to offset the increasing size and complexity of modern imaging intensive clinical trials. VMWare horizon view and Unidesk were installed in an effort to speed up the infrastructure and simplify user creation and desktop management. The platform currently serves 9 internal and 8 multi-center active clinical trials spanning 14 hospitals and imaging centers across the world.

Sub-aim 3.2: Conventional hypoxia imaging analysis

Assess the impact of progressively more quantitative hypoxia imaging methods on the predictive capacity of hypoxia biomarkers in a 4 clinical trials:

1. A Feasibility Study of Hypoxia Imaging in Patients With Cervix Cancer Using Positron Emission Tomography (PET) With 18F-Fluoroazomycin Arabinoside (18F-FAZA) – current accrual 8 out of 48;
2. The Potential for Metformin to Improve Tumor Oxygenation in Locally Advanced Cervix Cancer : A Phase II Randomized Trial – current accrual 27 patients out of 30;
3. 18F-Fluoroazomycin Arabinoside (FAZA) Positron Emission Tomography/Magnetic Imaging Resonance (PET/MRI) as a Biomarker of Hypoxia in Rectal Cancer: A Pilot Study – current accrual 1 patient out of 10;
4. A Feasibility Study of Hypoxia Imaging in Patients With Prostate Cancer Using Positron Emission Tomography (PET) With 18F-Fluoroazomycin Arabinoside (18F-FAZA) – current accrual 12 patients out of 20.

Head& Neck trial (H&N: RT/Sx) and SBRT in NSCLC didn't start recruiting patients. Due to this the expected number of total patients accrued for this aim is now 108 patients. All hypoxia analysis for these trials is in progress and performed by QIPCM.

The progress chart and milestones for Specific aim 3 is presented on Figure 9.

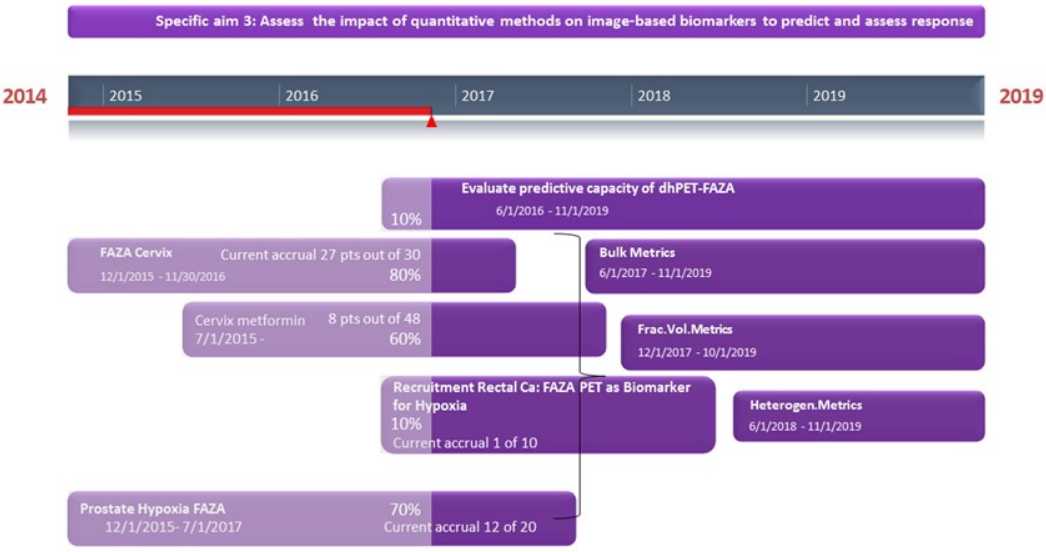


Figure 9: Progress chart with the milestones for the Specific Aim 3.

COLLABORATIONS WITHIN THE NETWORK

In the past year, we participated in two QIN challenges in the PET-CT subgroup and one challenge in the MR subgroup. The two challenges in the PET-CT subgroup are the CT nodule radiomic feature challenge and the Dynamic PET Analysis challenge. The CT nodule radiomic feature challenge concluded with a publication in Tomography [8], whereas the Dynamic PET Analysis challenge also led to a publication in JNM currently under review. Our DCE-MRI work also was published in the special issue of Tomography [9].

New challenge participation is to look at DCE-MRI brain immunology trials. Collaborative with MD Anderson and MGH [Dr Clifford Fuller, Caroline Chung and Jaysharee K-P].

PLANS FOR THE NEXT YEAR

§ Specific aim 1

Our results from dynamic PET imaging of hypoxia in pancreatic tumours emphasized that tracer uptake must be corrected for transport—notably partitioning—in order to reliably quantify hypoxia. Dynamic PET imaging is resource-intensive, however, and it would be

extremely useful if transport could be quantified using static, non-contrast CT or MRI imaging. Figure 6 shows our aim to extend this work by combining CT-based radiomics analyses with static PET data sets to improve hypoxia quantification.

In the next steps of our research the calculated scatter distribution will be validated against the measured scatter estimates with the help of the NEMA NU 2- 2012 standard [10]. To evaluate the influence of imaging parameters on the shape and magnitude of the scatter distribution, different activity concentration ratios (1:5 to 1:10), varying distances between localized tracer concentrations and changes in object size will be investigated and validated with the help of the NEMA IEC Body Phantom.

§ Specific aim 2

Future work will focus on validating this approach against PIMO uptake in resected pancreatic tumours. We also aim to extend our methods to other tumour sites by developing new imaging biomarkers for hypoxia that combine CT-based radiomics metrics with static PET images, validating with outcome data.

§ Specific aim 3

We will continue analyzing the images from 4 clinical trials for validation studies. New QIN challenge participation is to look at DCE-MRI brain immunology trials.

The milestones for the coming year for specific aims are presented on Figures 7-9.

LIST OF REFERENCES

1. Evans SM and Koch CJ Prognostic significance of tumor oxygenation in humans Cancer Lett (2003).
2. Dhani N, Fyles A, Hedley DW and Milosevic M The clinical significance of hypoxia in human cancers. Semin. Nucl. Med. (2015), 45:110–21.
3. Taylor E, Yeung I, Keller H, Wouters GB, Milosevic M, Hedley WD and Jaffray DA Quantifying hypoxia in human cancers using static PET imaging. Physics in Medicine and Biology (2016), Volume 61, Number 22.
4. Jan S, Santin G, Strul D, et al. GATE: a simulation toolkit for PET and SPECT. Phys. Med. Biol. (2004), 49:4543-4561.
5. Thielemans K, Tsoumpas C, Mustafovic S, Beisel T, Aguiar P, Dikaios N, Jacobson MW STIR: software for tomographic image reconstruction release 2. Phys. Med. Biol (2012), 57:867-883.
6. Mortensen L S, Johansen J, Kallehauge J, Primdahl H, et al. FAZA PET/CT hypoxia imaging in patients with squamous cell carcinoma of the head and neck treated with radiotherapy: Results from the DAHANCA 24 trial. Radiotherapy and Oncology (2012), 105(1):14-20.
7. Kueng R., Driscoll B., Manser P., Fix M., Stampanoni M., Keller H. Quantification of local image noise variation in PET images for standardization of noise-dependent analysis metrics. Biomedical Physics & Engineering Express, Accepted: 2016, November 25.

8. Kalpathy-Cramer J, Mamomov A, Zhao B, et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. Tomography (2016), Vol. 2, # 4:430-437.
9. Coolens C, Driscoll B, Foltz W, Pellow C, Menard C, Chung C Comparison of Voxel-Wise Tumor Perfusion Changes Measured With Dynamic Contrast- Enhanced (DCE) MRI and Volumetric DCE CT in Patients With Metastatic Brain Cancer Treated with Radiosurgery, Tomography (2016), Vol. 2, # 4:325-333.
10. National Electric Manufactures Association, NEMA NU 2-2012: Performance Measurements of Positron Emission Tomographs.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

§ Posters

1. Lin A, Vines D, Driscoll B, Le WL, Breen S, Sun A Positron Emission Tomography (PET) With 18F-Fluoroazomycin Arabinoside (FAZA) to Assess Tumor Hypoxia in Non-Small Cell Lung Cancer (NSCLC). Canadian Association of Radiation Oncology Annual Scientific Meeting (2016).
2. Vines CD, Driscoll B, Yeung I, Publicover J, Sun A and Jaffray DA Effects of respiratory gated ^{18}F-FAZA PET-CT on hypoxic fraction in patients and phantom. Imaging Network Ontario Annual Symposium (2016).
3. Yeung I, Metran-Nascente C, Vines D, Metser U, Dhani D, Green D, Milosevic M, Jaffray DA, Hedley DW Measurement of tumor hypoxia in patients with advanced pancreatic cancer based on 18F-fluoroazomyin arabinoside (18F-FAZA) uptake. Princess Margaret Cancer Centre's Annual Personalizing Cancer Medicine Conference (2016).
4. Taylor E, Yeung I, Keller H, Milosevic M, Hedley DW, Jaffray DA Measurement of tumor hypoxia in patients with advanced pancreatic cancer based on 18F-fluoroazomyin arabinoside (18F-FAZA) uptake. Princess Margaret Cancer Centre's Annual Personalizing Cancer Medicine Conference (2016).
5. Taylor E, Metran-Nascente C, Yeung I, Vines CD, Metser U, Dhani CN, Green D, Milosevic M, Hedley DW, Jaffray DA Optimal strategy for quantifying hypoxia from static PET imaging. Imaging Network Ontario Annual Symposium (2016).
6. Driscoll B, Yeung I, Coolens C, Keller H, Disney G, Svistoun I, Shek T, Publicover J and Jaffray DA An Update from the Quantitative Imaging for Personalized Cancer Medicine (QIPCM) Initiative. Imaging Network Ontario Annual Symposium (2016).
7. Taylor E, Gottwald J, Yeung I, Keller I, Milosevic M, Dhani NC, Hedley DW, Jaffray DA Overcoming The Complexity Of Molecular Transport In Neoplastic Tissue: Using A Dynamic Analysis Of Faza-Pet Imaging To Quantify Hypoxia In Human Tumours. Terry Fox Research Institute Ontario Node Research Symposium (2016).

§ Oral Presentations

1. Coolens C, Foltz W, Driscoll B, Pellow C, Chung C Comparison of Arterial Input Functions by Magnitude and Phase Signal Measurement in DCE MRI of brain cancer patients. Princess Margaret Cancer Centre's Annual Personalizing Cancer Medicine Conference (2016).

2. Vines DC, Driscoll B, Keller H, Lin A, Sun A, Jaffray DA Respiratory gated hypoxia imaging using ^{18}F-FAZA PET-CT. Society of Nuclear Medicine and Molecular Imaging Annual Meeting (2016).
3. Driscoll B, Yeung I, Coolens C, Keller H, Disney G, Svistoun I, Shek T, Publicover J and Jaffray DA An Update from the Quantitative Imaging for Personalized Cancer Medicine (QIPCM) Initiative. International Conference on the use of Computers in Radiation Therapy (2016).
4. Welch M, Jaffray DA Automated Spatial Dose Prediction Using Contextual Atlas Regression Forests. International Conference on the use of Computers in Radiation Therapy (2016)..
5. Yang DM, Li F, Bauman G, Valliant J, He W, Lee TY Comparison of Dynamic 18F-DCFPyL and 18F-FCH PET Imaging in Patients with Prostate Cancer. Ontario Institute of Cancer Research Imaging Applications in Prostate Cancer Workshop (2016).

§ Publications

1. Coolens C, Driscoll B, Foltz W, Pellow C, Menard C, Chung C Comparison of Voxel-Wise Tumor Perfusion Changes Measured With Dynamic Contrast-Enhanced (DCE) MRI and Volumetric DCE CT in Patients With Metastatic Brain Cancer Treated with Radiosurgery. Tomography (2016). Vol.2, #4:325-333.
2. Coolens C, Driscoll B, Moseley J, Brock KK, Dawson AL Feasibility of 4D perfusion CT imaging for the assessment of liver treatment response following SBRT and sorafenib. Advances in Radiation Oncology (2016). Vol. 1, # 3:194-203.
3. Han K, Croke J, Foltz W, Metser U, Xie J, Shek T, Driscoll B, Menard C, Vines D, Coolens C, Simeonov A, Beiki-Ardakani A, Leung E, Levin W, Fyles A, Milosevic M A prospective study of DWI, DCE-MRI and FDG PET imaging for target delineation in brachytherapy for cervical cancer. Radiotherapy and Oncology (2016). Vol. 120, #3:519-525.
4. Taylor E, Yeung I, Keller H, Wouters GB, Milosevic M, Hedley WD, Jaffray DA Quantifying hypoxia in human cancers using static PET imaging. Physics in Medicine and Biology (2016). Volume 61, Number 22.
5. Kueng R, Driscoll B, Manser P, Fix M, Stampanoni M, Keller H Quantification of local image noise variation in PET images for standardization of noise-dependent analysis metrics. Biomedical Physics & Engineering Express, Accepted: 2016, November 25.
6. Metran-Nascente C, Yeung I, Vines CV, Metser U, Dhani CN, Green D, Milosevic M, Jaffray DA, Hedley WD Measurement of Tumor Hypoxia in Patients with Advanced Pancreatic Cancer Based on ^{18}F-Fluoroazomyin Arabinoside Uptake. Journal of Nuclear Medicine (2016). Vol. 57:361-366.
7. Kalpathy-Cramer J, Mamomov A, Zhao B, et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. Tomography (2016), Vol. 2, # 4:430-437.

The QIN Working Groups

Clinical Trials Design and Development Working Group

Hui-Kuo Shu, MD, PhD
Ella Jones, PhD

MISSION

Develop, validate and harmonize methods and tools of quantitative imaging for use in cancer clinical trials to predict outcome and tumor response to therapy.

GOALS

1. Identify challenges and opportunities in clinical trial design and development particularly in trials using quantitative imaging (QI).
2. Identify best practices for clinical trial design, analysis and reporting.
3. Facilitate and introduce QIN-developed methods into cancer imaging trials through collaboration with other QIN working groups.
4. Disseminate the best clinical trial design and development method through publications and guidelines.
5. Outreach to cooperative groups and organizations to apply QIN methods in multicenter trials through cross-membership and presentations.
6. Translate relevant and mature QIN methods into clinical practice settings as appropriate.

ACCOMPLISHMENTS/ACTIVITIES FOR THE YEAR (2016-2017)

§ Manuscripts

QIN Accrual Survey

Led by Brenda Kurland, the past Chair of the CTDD WG, this effort was performed to gain a better understanding of factors that affect enrollment for QI clinical trials and to identify potential barriers to accrual of such studies. This work was recently published in a special QIN issue of Tomography with abstract and reference listed below.

Patient accrual is essential for the success of oncology clinical trials. Recruitment for trials involving the development of quantitative imaging biomarkers may face different challenges than treatment trials. This study surveyed investigators and study personnel for evaluating accrual performance and perceived barriers to accrual and for soliciting solutions to these accrual challenges that are specific to quantitative imaging-based trials. Responses for 25 prospective studies were received from 12 sites. The median percent annual accrual attained was 94.5% (range, 3%-350%). The most commonly selected barrier to recruitment (n = 11/25, 44%) was that "patients decline participation," followed by "too few eligible patients" (n = 10/25, 40%). In a forced choice for the single greatest recruitment challenge, "too few eligible patients" was the most common response (n = 8/25, 32%). Quantitative

analysis and qualitative responses suggested that interactions among institutional, physician, and patient factors contributed to accrual success and challenges. Multidisciplinary collaboration in trial design and execution is essential to accrual success, with attention paid to ensuring and communicating potential trial benefits to enrolled and future patients.

- Kurland, B.F., Aggarwal, S., Yankeelov, T.E., Gerstner, E.R., Mountz, J.M., Linden, H.M., Jones, E.F., Bodeker, K.L. and Buatti, J.M., 2016, Accrual Patterns for Clinical Studies Involving Quantitative Imaging: Results of an NCI Quantitative Imaging Network (QIN) Survey, Tomography, 2:276-82.

STandard In Reporting Quantitative Imaging (STIRQI*)*

Increasingly, imaging methods are used in clinical trials both as primary as well as secondary or correlative endpoints. Interpretation of imaging has also been moving from qualitative and subjective to quantitative and objective. As these methods become more sophisticated, basic information regarding the acquisition of QI data must be provided to the reader so that the validity and reliability of these results can be determined and generalized. This initiative seeks to define a set of criteria that should be presented in QI-related publications to ensure that quantitative data extracted from images are reported in a meaningful, consistent, and repeatable manner.

Rich Wahl has taken input from members of the CTDD WG and leads the effort of this project. A draft of the manuscript with preliminary checklist of standards for reporting has been circulated. This checklist is based on the STAndards for Reporting of Diagnostic Accuracy (STARD) criteria first reported in 2003 (Bossuyt, et al., Ann Int Med 138:W1-12, 2003). There has since been an update of the STARD criteria in 2015 (Bossuyt, et al., BMJ 351:h5527, 2015). The manuscript will be revised accordingly and will be circulated more widely within the QIN for comments.

Quantitative Imaging in Radiation Oncology

Radiation oncology is increasingly reliant on both high-resolution anatomic-based imaging (CT and MRI) as well as functional imaging (PET, DWI, MRSI etc.). This evolution in the field has arisen because of the improved ability to localize the radiation treatment delivery accurately through stereotactic guidance as well as image-guided radiation therapy (IGRT) with daily image-based alignment. As this becomes possible on a routine basis, the margins of error in delivery have decreased to sub-millimeter accuracy in intracranial applications and on the order of 1-2 millimeters in body treatments. A review of QI, in particular, focusing on its utility for the radiation oncologist would be an important addition to the literature and help highlight the increasing range of advanced imaging modalities for this clinical field.

John Buatti and Hui-Kuo Shu will be leading the effort to produce a manuscript reviewing imaging modalities used by radiation oncologists in routine patient management and highlighting potential utility of QI techniques in treatment planning and patient follow-up. The current plan is to complete a draft in the next 4-6 months and circulate to the CTDD WG

for comments before the final revision and submission to a radiation oncology journal.

§ Outreach activities

Panel Session at the 2016 Annual Meeting of the American Society for Therapeutic Radiology and Oncology (ASTRO)

A panel session co-chaired by John Buatti and Hui-Kuo Shu entitled "Advanced Quantitative Imaging for the Radiation Oncologist: Response Assessment and Targeting for Clinical Trials and Practice, A View from the NCI's Quantitative Imaging Network" was presented at ASTRO on September 25, 2016. This session awarded CME credit. Overall, the session was well-received with an attendance of above 100. The overall course rating was 4.2 (on a scale of 1-5 w/higher being better) and nearly 83% of respondents desired to have this session repeated in the future. The summary, learning objectives and presentations of this panel session were as follows:

Summary: The Quantitative Imaging Network (QIN) is an effort by the National Cancer Institute (NCI) to develop novel approaches, including tool development, for QI and clinical decision support in oncology. It consists of a number of multi-disciplinary research teams from top institutions across the country that are developing quantitative cancer imaging methods and tools that can be applied in unique ways in the context of clinic trials. This session introduced the structure and aims of NCI's QIN program and described some of the work being done by specific groups to develop and validate QI tools for common cancers that are of interest to the membership of ASTRO. The panel described how advanced imaging modalities can complement more conventional ones for oncologic evaluations and some of the tools that are in various stages of development for this assessment. It is our hope that this panel will raise awareness of the potential utility of advanced QI for both research and general clinical practice. Several barriers have limited wider adoption of advanced QI techniques in the clinic. This panel may address some of these barriers by educating oncologists about the value of QI. Other barriers are being addressed as central goals of the QIN to help develop and validate new tools and methods that benefit the utility of QI in the clinical setting.

Learning Objectives:

1. Demonstrate knowledge of the structure and goals of the NCI's QIN program for developing and applying advanced QI in oncology.
2. Demonstrate knowledge of and apply various areas of advanced QI research including validation of methodologies and development of tools that will eventually be translated to the clinic.
3. Demonstrate an understanding of how the developing QI tools can be used in general oncology practice.

PRESENTATIONS

1. QI-based tools for radiation therapy targeting and response assessment in head and neck cancer (John Buatti).
2. The potential of radiomic-based phenotyping for precision medicine (Hugo Aerts).
3. DCE-perfusion and diffusion-weighted MR imaging for clinical decision support in head and neck cancer (Yue Cao).
4. Feasibility of whole brain, high-resolution spectroscopic MRI for glioblastoma brain tumor imaging (Hui-Kuo Shu).

CROSS-INSTITUTIONAL QIN TOOLS VALIDATION

§Auto-PERCIST variance test

PET Response Criteria in Solid Tumors (PERCIST) was initially proposed by Wahl et al. (J Nucl Med, 50 (suppl 1):122S-150S, 2009) as an approach to standardize interpretation of FDG-PET results using a consistent PET protocol. Richard Wahl's group has subsequently developed a software (AutoPERCIST™) to semi-automatically identify and measure reference tissue (liver), set disease threshold values and calculate SUVs (peak, max, mean, volume and total lesion glycolysis) based on PERCIST criteria. AutoPERCIST™ has been used to evaluate 30 test cases and demonstrated its robust performance. Results from this work has recently been accepted for publication in the *Journal of Nuclear Medicine*. A subsequent variance test was proposed to determine whether lesions quantified by AutoPERCIST™ is consistent across institutions.

Joo Hyun O, Clinical Assistant Professor (Department of Nuclear Medicine, Seoul St. Mary's Hospital, Seoul, Korea) and Richard Wahl are leading this effort to determine the variance of lesion quantification by AutoPERCIST™. The latest version of this software was installed by 15 participating institutions (from United States, Asia and Europe) through a materials transfer agreement (MTA) and 30 paired sets of anonymized FDG PET-CT images were downloaded for evaluation. Instructions for this study were recently sent to participating institutions, up to 5 tumor lesions from each PET image will be selected. All selections will be recorded and sent to the central database at Johns Hopkins Image Response Assessment Team for quality control. Initial results are expected in the second quarter of 2017.

§ Pathways to Clinical Trials (PathCT) initiative

A major focus of the CTDD WG in 2016-2017 is to help facilitate the translation of QIN-developed tools to clinical trials, particularly those in the NCI's National Clinical Trials Network (NCTN). These goals have been advanced in several ways over the past year as listed below.

2016 QIN Face-to-Face Annual meeting (April 11-12, 2016)

A framework for clinical engagement based on the perspectives of clinical trialists who actively participate in the design and conduct of imaging clinical trials was introduced in multiple presentations by Larry Schwartz, Hannah Linden and Dave Mankoff on the 1st day of this meeting. Further discussions on the role of the CTDD Working Group in this engagement were the topics of subsequent breakout sessions at this meeting.

QIN-NCTN Planning Meeting (December 13, 2016)

This one-day meeting was held in Philadelphia to discuss ideas and opportunities where QI could play a key role in the NCTN trials and to consider which tools from the QIN portfolio may have merit. The meeting minutes will be made available on Sharepoint.

Updated list of QIN tools and their level of readiness for deployment (January, 2017)

Lori Henderson compiled this updated list after communications with each QIN group to obtain the latest status on their respective tools. This step was critical to gain an understanding of where current tools stand so that further customized recommendations can be formulated.

PathCT Summary Report

Lori Henderson led the effort to prepare this document summarizing the activities (listed above) in 2016 and the goals of the PathCT initiative. In addition, this report discussed future directions to further the goal of advancing the development and maturation of the novel QI technologies from the network through a targeted and appropriately balanced approach for clinical validation in trials. This report will be made available on Sharepoint.

PathCT Focus Group

This group, comprised of members of the CTDD WG, is being tasked to review the readiness of QIN-developed tools and facilitate in-depth discussions with the corresponding PIs to bring these tools to clinical translations. The core members of this focus group currently consist of John Buatti, Lori Henderson, Ella Jones, Hui-Kuo Shu and Richard Wahl with the inaugural meeting on Monday, 2/13/2017. Membership may be modified later including the addition of those that can serve in specific ad hoc or advisory roles. The current plan is to meet at least once monthly and to report progress at each CTDD WG teleconference.

PLANS FOR THE COMING YEAR (2017-2018)

§ Goals for the coming year include the following:

1. Completion of two manuscripts (STIRQI criteria and Quantitative Imaging in Radiation Oncology) that are currently under preparation.

2. Continued outreach efforts at national oncology and cooperative group meetings to educate about the utility and promise of QI and the role of the QIN in developing these techniques. Specific efforts will include the following:

    a. NRG Oncology will be targeted for a presentation at the Imaging subcommittee at their July 2017 meeting. The goal is to make a presentation introducing the QIN and some mature QIN tools, in particular, that are poised for potential inclusion in developing clinical trials. Plan to have handouts available and contact information so that communications between clinical trialists and imaging scientists can be facilitated.

    b. 2018 ASTRO will be targeted for another panel session similar to the one presented in 2016. This proposal will be prepared for submission to the ASTRO organizing committee by late 2017 (Nov to Dec).

3. Completion of the AutoPERCIST™variance test.

4. Continue to advance the goals of the PathCT initiative primarily through the activity of the PathCT Focus Group and additional potential ad hoc activities such as with meetings similar to the QIN-NCTN planning meeting held on 12/13/2017.

Bioinformatics and Data Sharing Working Group

Bradley Erickson, M.D., Ph.D.
Ashish Sharma, Ph.D.

INTRODUCTION

There are 2 main purposes of the Bioinformatics and Data Sharing (BIDS) working group: to promote and facilitate data sharing within the QIN and for other cancer researchers; and to promote and facilitate the sharing of computational tools within the QIN and for other cancer researchers. In the past, much of the work was on enabling data sharing and TCIA. In the past year, we have continued that work, in the form of advancing the discussion of common data elements for the non-image data that is associated with images. However, the greater focus of the past year was to improve the ability to share tools.

DISCUSSION OF PROGRESS

§ Improved Tool Sharing and Pipeline Creation

Every form of quantitative imaging requires some computational approach to the data. It may be as simple as computing the mean of a region of interest (ROI). However, it is rare that this is sufficient, and even computation of ROIs has been found to have its subtleties—for instance, are the pixels under the line defining the ROI included or not? Furthermore, one must usually match the ROI with some structure of interest, and that requires accurate and reproducible segmentation, and may also require registration onto an anatomic image. In many cases, the mean does not contain all (or even most) of the useful information. Recent studies have shown textures can reveal information about the genomic makeup of tumors even when visual inspection reveals no such information.

The BIDS group does not develop those measurement tools, nor does it evaluate them. However, in nearly all cases, research groups create ‘pipelines’ for image analysis. A pipeline is a series of well-defined steps, and each step is a certain class of algorithm. Some example pipeline steps include image registration, image segmentation, image classification, and perhaps some ROC or other measurement of performance. When properly constructed, one may ‘swap in’ some new algorithm that performs one of those steps and evaluate the impact on accuracy or computational efficiency.

A challenge is that there has not been any standard or convention for creating pipelines, let alone sharing the modules that make up each step. This results in duplication of effort as each group creates their own pipeline methodology and their own algorithms that go into the pipeline. The BIDS Pipeline project is focused on identifying best practices for creating and sharing the modules used to make pipelines as well as the best way to create and maintain the pipelines themselves.

The Stanford Group has been developing the Quantitative Imaging Feature Pipeline. At present, they have largely focused on development of the computational modules (e.g. computation of image features or a classifier that consumes to computed

features) and have not focused much effort on the actual pipeline technology. They have selected CWL—the common workflow language as the method for describing a pipeline.

The Mayo and Emory groups have been working for about 2 years on pipeline technology, mostly focusing on how best to connect them. A rudimentary demonstration was given at last year's face to face meeting. The basic construct of this effort is that all computational modules should exist as a docker. Some examples have been published on Dockerhub by the Mayo group. Docker is an open source technology that allows one to capture a complete execution environment as a file that can then be executed on any Docker host, much like virtual machine technology. To leverage Dockers in a versatile and flexible way we extended an open source software tool called Grunt (https://github.com/Mayo-QIN/grunt) to simplify the creation and deployment of modules in a pipeline.

REST may provide advantages in its ability to natively support cloud execution (including mixed and local models), is commonly used outside of research and thus may allow us to access tools for pipeline creation, and also provides a way to be somewhat more secure than what docker file access may allow.

We plan to demonstrate a pipeline that uses Grunt for connections, and how one can insert/replace docker modules when one wishes to use alternative algorithms for certain steps. There are a variety of properties of pipeline technology that are important, including: Ease of Use (including creating and maintaining a pipeline, and ease of making a docker compatible with the pipeline), Computational Efficiency, Security, and Scalability (ability to add 'cloud' resources). Most of these properties are subjective, and not amenable to traditional 'challenge' evaluations, but the BIDS group is committed to developing challenges to identify and promote the better pipeline options, and most importantly, to promote compatibility of the modules that compose a pipeline. It is those modules that are the greatest value produced by the QIN.

§ Improved Data Sharing and Common Data Elements

An important activity of QIN is data sharing. The TCIA is a tremendous resource that enables data sharing. However, it is not always easy to contribute data, as there is much greater value if there is metadata associated with the images. Furthermore, that metadata needs to be represented in a standard way, if data collections are to be aggregated. Therefore, it is important to establish Common Data Elements for TCIA data sets. When TCGA constituted the source of metadata for the images, this problem was well-handled.

PLANS FOR NEXT YEAR

The BIDS group will continue to work on better pipeline technology to make it easier for QIN community to leverage the tools that others develop. This includes making it easy to get the tools, to connect them together in a reliable and efficient way, to update/maintain the pipeline, and to have efficient computation. We plan will be demonstrating this at the face-to-face meeting, and hope that will both educate the QIN community and give us actionable feedback about the relative importance of the above pipeline properties.

As more data sets come into TCIA that are not associated with TCGA, the challenge of determining which CDEs are required to accept an image, as well as representations of CDEs is needed. This will be another activity for BIDS in the next year.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

1. Fedorov A, Clunie D, Ulrich E, Bauer C, Wahle, Brown B, Onken M, Riesmeier J, Pieper S, Kikinis R, Buatti J, Beichel RR. 2016. DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research. PeerJ 4:e2057. DOI: 10.7717/peerj.2057.
2. Paul R, Hawkins SH, Balagurunathan Y, Schabath MB, Gillies RJ, Hall LO, Goldgof DB. Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma. Tomography : a journal for imaging research. 2016;2(4):388-95. doi: 10.18383/j.tom.2016.00211. PubMed PMID: 28066809; PubMed Central PMCID: PMC5218828
3. Kelm ZS, Korfiatis P, Lingineni RK, Daniels JR, Buckner JC, Lachance DH, Parney IF, Carter RE, Erickson BJ. Variability and accuracy of different software packages for dynamic susceptibility contrast magnetic resonance imaging for distinguishing glioblastoma progression from pseudoprogression. J Med Imaging. 2015 Apr; 2(2): 026001 doi: 10.1117/1.JMI.2.2.026001. PMID 26158114, PMCID:PMC4478857
4. Akkus Z, Sedlar J, Coufalova L, Korfiatis P, Kline TL, Warner JD, Agrawal J, Erickson BJ. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging. Cancer Imaging. 2015 Aug 14;15(1):12. doi: 10.1186/s40644-015-0047-z. PMID: 26268363, PMCID:PMC4535671
5. Korfiatis P, Kline TL, Coufalova L, Lachance DH, Parney IF, Carter RE, Buckner JC, Erickson BJ. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. Med Phys. 2016 Jun;43(6):2835. doi: 10.1118/1.4948668. PMID: 27277032, PMCID:PMC4866963
6. Korfiatis P, Kline TL, Blezek DJ, Langer SG, Ryan WJ, Erickson BJ. PESSCARA: A Content Management System for medical image analysis research. Radiographics. 2015 Sep-Oct; 35 (5): 1461-8. PMCID:PMC26284301
7. Korfiatis P, Kline TL, Erickson BJ. Automated Segmentation of Hyperintense Regions in FLAIR MRI Using Deep Learning Tomography.2016;2(4)334. DOI: 10.18383/j.tom.2016.00166
8. Korfiatis P, Kline TL, Kelm ZS, Hu LS, Erickson BJ. Dynamic Susceptibility Contrast MRI Quantification Software Tool: Development and Evaluation. Tomography. PMC Journal - 2016;2(4)448. DOI: 10.18383/j.tom.2016.00172
9. Erickson BJ, Korfiatis P, Akkus Z, Kline T. Machine Learning For Medical Imaging. Radiographics, March 2017. PMC Journal - In Press.
10. Korfiatis P, Erickson BJ. PESSCARA: An Example Infrastructure for Big Data Research In: Big Data on Real-World Applications. (Book chapter) DOI: 10.5772/63815
11. Clunie D.A., Fedorov A. Knowledge Representation of Prostatic Sector Anatomy from PI-RADS in Standard Lexicons. SSG07-08 Informatics (Results and Reporting). The Radiological Society of North America 102nd Scientific Assembly and Annual Meeting, 2016 November, Chicago, IL, USA.
12. Fedorov A., Rubin D., Kalpathy-Cramer J., Kirby J., Clunie D., Onken M., Flade D., Mongkolwat P., Venkateraman R., Bertling J., Pieper S., Kikinis R. The Radiological

Society of North America 101nd Scientific Assembly and Annual Meeting, 2015. November, Chicago, IL, USA. Interoperable communication of quantitative image analysis results using DICOM standard. DOI: 10.6084/m9.figshare.1619877.v1.
13. Pujol,S, Pieper,S, Fedorov,A, Kikinis,R, The 3D Slicer Open-source Platform for Segmentation, Registration, Quantitative Imaging and 3D Visualization of Multi-modal Image Data. Radiological Society of North America 2016 Scientific Assembly and Annual Meeting, November 27 - December 2, 2016, Chicago IL. archive.rsna.org/2016/16006433.html Accessed December 5, 2016
14. Fedorov A, Rubin D, Clunie D, Pieper S, Kikinis R. Standardized communications of quantitative image analysis results using DICOM: Establishing interoperability through outreach and community engagement. AMIA Joint Summits on Translational Science. March 27-30, 2017, San Francisco, CA.
15. Korfiatis P. Kline TL. Erickson BJ. Residual Deep Convolutional Neural Network Predicts Management Methylation Status. Machine Intelligence in Medical Imaging, Alexandria, Virginia (2016).
16. Akkus Z, Sedlar J., Ali I., Giannini G., Parney I, Erickson B.J.. Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images using Machine Intelligence. Conference on Medical Intelligence in Medical Imaging. Sep 2016
17. Akkus Z, Ali I, Sedlar J., Kline TL., Agrawal J.P, B. Erickson J.. Predicting Response of Low Grade Gliomas to Therapy from MRI Images using Convolutional Neural Networks (CNNs). RSNA 2016
18. Erickson BJ, Korfiatis P, Akkus Z, Kline T. Man vs Machine: Hands-on tutorial on Machine Learning. RSNA 2016

Data Acquisition Working Group

Stefanie Hectors, Ph.D. and Octavia Bane, Ph.D.

INTRODUCTION

The role of the Data Acquisition Working Group is to identify, characterize, and ameliorate sources of variance and bias in image data acquisition, thereby enhancing the value of advanced oncologic quantitative imaging methods used in clinical trials. Toward this end, we work with the QIN and system manufacturers to develop standardized system test procedures to enable objective assessment of quantitative image performance across sites and platforms. As of September 1st 2016, the Data Acquisition Working Group has been merged with the PET/CT and MRI working groups. Ongoing projects will continue within those groups. This report therefore gives an overview of the accomplishments of the previous year, while future plans will be given in the reports of the PET/CT and MRI working groups.

ACCOMPLISHMENTS OF THE PREVIOUS YEAR

§ PET-CT

Cross-Calibration (X-Cal) demonstration project

PET/CT scanner sensitivity is frequently determined empirically by filling and imaging aqueous phantoms. This calibration procedure is believed to lead to approximately 5% variability over time, for a single scanner, when no mistakes are made. However, overall stability of scanner bias due to calibration has not been well characterized for networks of hospitals, in which sites may differ in their hardware, software and calibration procedures. The X-Cal demonstration project used long-lived sources that were implicitly traceable by the standard of the National Institute of Standards and Technology (NIST) to characterize scanner calibration variability. The project has now been concluded with nearly 250 calibration scans on 19 scanners. An analysis of the PET scanner and dose calibrator bias data has been completed at the University of Washington, and the manuscript is being finalized. Results have been presented to the QIN (1) and RSNA (2). Key results include the larger-than-expected instability of PET scanner calibration bias (**Figure 1**) and the lack of significant correlation between scanner and dose calibrator bias.

PET Protocol Acquisition Survey project

Bias and variability of PET SUV values are known to be influenced by many factors including patient handling, data acquisition and image reconstruction techniques. In particular, user-selected parameters such as acquisition duration, injected dose, uptake time, iterative reconstruction updates, and post-reconstruction smoothing lead to a spectrum of possible noise and resolution properties, which directly influence bias and variability. The Protocol Acquisition Survey Project distributed questionnaires to and compiled data from 44 PET scanners at 36 separate sites, focusing on the patient requirements and clinical scanner

settings that are most likely to affect SUVs. The survey project has now been concluded. Key results were the large reported ranges for factors affecting SUV bias, including reconstruction settings (Figure 2), uptake time, and other factors. The manuscript for this project is being finalized.

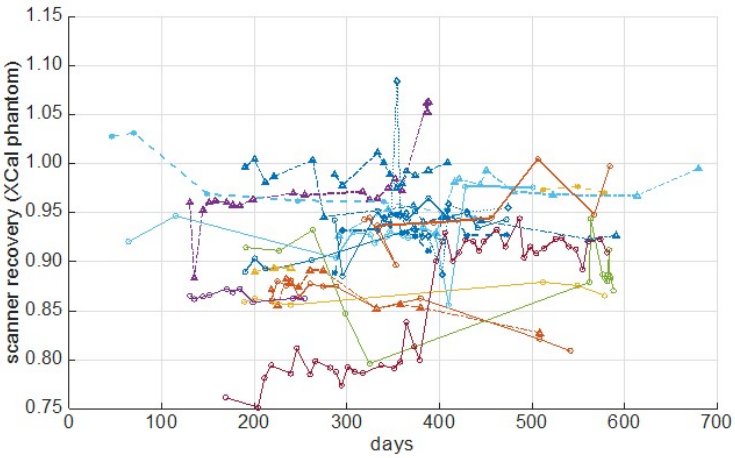


Figure 1: For 19 scanners, signal recovery (the ratio of known activity to mean ROI signal) versus time, as determined by scanning of long-lived solid phantoms distributed in 2013.

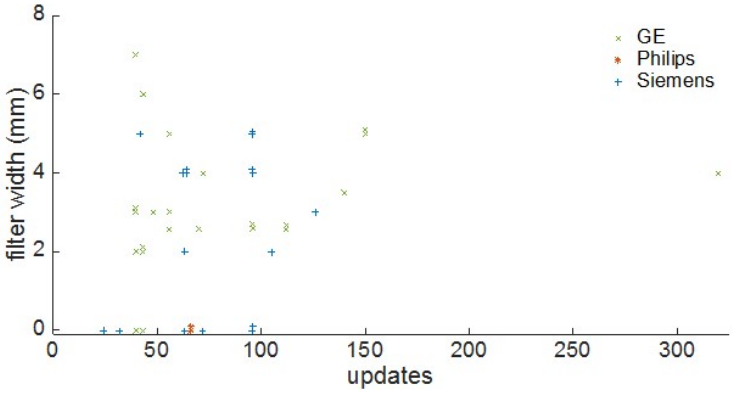


Figure 2: Reconstruction parameters reported by survey sites. Because iterative updates are commonly computed with subsets, we report the total updates as the product of iterations and subsets on the x-axis. The y-axis shows the reported post-filter, specified by the full-width-at-half-maximum. Some sites used non-iterative reconstruction and were excluded from this plot.

MRI

DWI gradient nonlinearity project (phase 2): The objective of this collaborative project was to demonstrate feasibility of centralized retrospective system-specific correction of gradient nonlinearity (GNL) bias for quantitative diffusion weighted imaging (DWI) across diverse scanners independent of scanned object, and therefore, applicable in multi-site clinical trials. Six representative MR scanner models were selected (two from each vendor: GE, Philips and Siemens). Using corrector maps generated from gradient system characterization by ice-water phantom in the previous project phase (3), GNL bias correction was performed for test ADC measurements from an independent DWI phantom (room-temperature agar). The pre-computed three-dimensional GNL correctors (4) were retrospectively applied to test DWI scans by the central analysis site. The correction was blinded to reference DWI of the agar phantom acquired by sites at magnet isocenter where GNL bias is negligible. The performance was evaluated from changes in ADC ROI histogram statistics before and after correction with respect to the unbiased reference ADC values provided by sites (Figure 3). Both absolute error and non-uniformity of ADC map induced by GNL (median: 12%, range: -35% to +10%) were substantially reduced by correction (seven-fold in median and three-fold in range). Correction of systematic GNL bias resulted in two-fold decrease of technical variability across scanners (down to site temperature range).

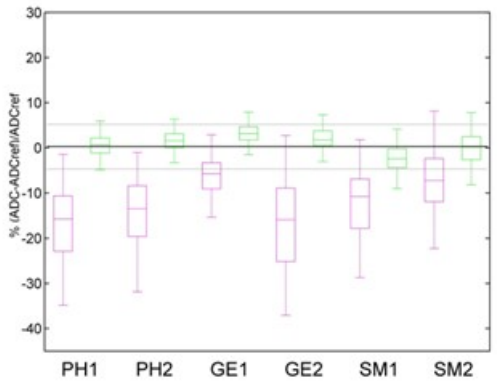


Figure 3: Percent-bias box-plot summary for ADC ROI histograms measured in respect to reference ADC for six studied systems (PH1, PH2, GE1, GE2, SM1, SM2) before (magenta) and after (green) "scaled" GNL bias correction illustrates substantial improvement of ADC precision and uniformity, as well as reduction of cross-scanner variability post correction. Median ADC bias is marked with the central line inside the box. The edges of the box correspond to 25th and 75th percentiles, while whiskes encompass the 5th to 95th percentile data points. The dashed horizontal lines delineate +5% error ranges.

This work has demonstrated that centralized retrospective correction of GNL bias in diffusion weighting, obtained from one-time empiric characterization of system GNL, is warranted by the stability of gradient channel characteristics, is desired for substantial reduction of ADC map bias, and is clearly feasible for multi-center clinical trial setting. In the absence of the preferred, prospective GNL correction using system design coefficients or independent 3D gradient field mapping, available (approximate) empiric correctors provide a practical solution for substantial improvement by removing systematic non-uniformity bias at off-center locations and reducing technical variability across multiple scanner systems. When not corrected, this technical bias both shifts and artificially broadens the corresponding ADC ROI histograms, and increases cross-system variability of the quantitative DWI metrics. The reduction of systematic ADC map errors using the proposed technology will have a positive impact on clinical trials that utilize quantitative parametric ADC maps in diagnostic and treatment response metrics. This collaborative project was successfully concluded and results published in QIN special issue of Tomography (5).

T_1 mapping data acquisition project: The objectives of this project were to 1) to determine the accuracy and test-retest precision of several T_1 mapping protocols used in DCE-MRI studies and 2) to measure interplatform variability in T_1 quantification by multicenter testing of common protocols on a dedicated phantom. Between April 2015 and June 2016, the phantom was circulated among 8 sites for data collection with three types of T_1 mapping protocols: IR-SE and VFA with protocol standardized among sites, and site-specific protocols. We used a model selection procedure to identify the independent predictors of accuracy and precision (repeatability) errors (Figure 4**)**. The general linear models showed that the accuracy of the common VFA protocol with respect to reference NMR T_1 is dependent on field strength, with measurements at 3T, less accurate. The test-retest repeatability of the common VFA protocol depends on the scanner used. For site-specific protocols, accuracy of T_1 measurements depended on site, while test-retest repeatability depended on the type of protocol used. Look-Locker inversion recovery protocols (4a, 9a) at both field strengths provided the most repeatable measurements. Among VFA protocols, protocols 2 and 3 for the brain, which use multiple flip angles, were more repeatable that VFA protocols for the liver (4b, 9b) and prostate (4c, 9c, 9d), which use 2-3 flip angles. The results of this project have been presented at QIN (6) and RSNA (7). The manuscript is being finalized.

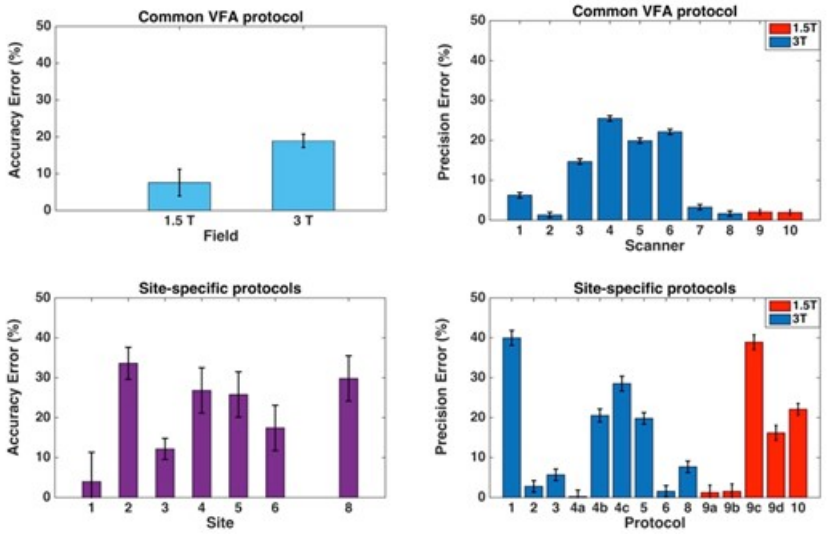


Figure 4: Significant predictors of accuracy and test-retest precision errors of T_1 measurements with the common VFA protocol and site-specific protocols. The results of general linear mixed models are presented as least square means ± standard error. Smaller numbers represent better accuracy/repeatability. Site-specific protocols, by scanner number: 1= Prostate VTR, 2= Brain VFA I, 3= Brain VFA II, 4a= Liver Look-Locker 3T, 4b= Liver VFA 3T, 4c=Prostate VFA 3T, 5=Sarcoma PD, 6= Brain VTR, 8=Breast VFA 3T, 9a= Liver Look-Locker 1.5T, 9b= Liver VFA 1.5T, 9c= Prostate VFA I 1.5T, 9d= Prostate VFA II 1.5T, 10= Breast VFA 1.5T. Site/scanner 7 did not provide site-specific data.

LIST OF REFERENCES

1. Byrd D. Data Acquisition Working Group XCal Demonstration Project. QIN annual meeting 2016.
2. Kinahan PE, Byrd D. Longitudinal Variability of PET SUVs in the NCI Quantitative Imaging Network (QIN). Proceedings of the RNSA 2016 Annual Meeting.
3. Malyarenko DI, Newitt D, L JW, Tudorica A, Helmer KG, Arlinghaus LR, Jacobs MA, Jajamovich G, Taouli B, Yankeelov TE, Huang W, Chenevert TL. Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials. Magnetic resonance in medicine. 2016;75(3):1312-23 (MRM March 2016 Editor's Pick). doi: 10.1002/mrm.25754. PubMed PMID: 25940607; PMCID: 4630210.
4. Malyarenko DI, Chenevert TL. Practical estimate of gradient nonlinearity for implementation of apparent diffusion coefficient bias correction. Journal of magnetic resonance imaging : JMRI. 2014;40(6):1487-95. PubMed PMID: 25667948; PMCID: 4324556.

5. Malyarenko DI, Wilmes LJ, Arlinghaus LR, Jacobs MA, Huang W, Helmer KG, Taouli B, Yankeelov TE, Newitt D, Chenevert TL. QIN DAWG Validation of Gradient Nonlinearity Bias Correction Workflow for Quantitative Diffusion-Weighted Imaging in Multicenter Trials. Tomography : a journal for imaging research. 2016;2(4):396-405. doi: 10.18383/j.tom.2016.00214. PubMed PMID: 28105469; PMCID: 5241082.
6. Bane O, Hectors S, Wagner M, Arlinghaus LR, Aryal M, Boss M, Cao Y, Chenevert TL, Fennessy F, Huang W, Hylton N, Kalpathy-Cramer J, Keenan K, Malyarenko D, Mulkern R, Newitt D, Wilmes L, Yen Y, Yankeelov TE, Taouli B. Data Acquisition Working Group Challenge: Assessment of interplatform variability of T1 quantification methods used for DCE-MRI in a multicenter phantom study. QIN annual meeting 2016.
7. Bane O, Hectors S, Wagner M, Arlinghaus LR, Boss M, Cao Y, Taouli B. Assessment of Interplatform Variability of T1 Quantification Methods Used for DCE-MRI in a Multicenter Phantom Study. Proceedings of the RNSA 2016 Annual Meeting.

QIN MRI Subgroup

C. Chad Quarles, Ph.D. and Melissa Prah, Ph.D.

INTRODUCTION

The mission for the MRI subgroup of the Image Analysis & Performance Metrics Working Group (IAPM) of the Quantitative Imaging Network (QIN) is to provide guidance, coordination, consensus building, and awareness regarding the development of acquisition and post-processing methods for quantitative analysis of tumors, related tissues and organs, and changes in response to disease progression and treatment, as well as to influence the development of sharable objective methods and metrics for assessment of image analysis accuracy, reproducibility, and robustness. The working group will coordinate the collaboration between members in this area.

Newly restructured in 2016, members with an MRI focus have now merged from the Data Acquisition Working Group (DAWG) into the MRI subgroup as one cohesive group. MRI subgroup activities now comprise participation from 28 cancer imaging centers, which include Barrow Neurological Institute (BNI), Brigham and Women's Hospital (BWH), Columbia University (CU), Icahn School of Medicine at Mount Sinai (MS), Johns Hopkins University (JHU), Maastricht University Medical Centre (MUMC), Mayo Clinic in Phoenix (MCP), Mayo Clinic in Rochester (MCR), Massachusetts General Hospital (MGH), Medical College of Wisconsin (MCW), Memorial Sloan Kettering Cancer Center (MSKCC), Moffitt Cancer Center (MCC), Ohio State University (OSU), Oregon Health and Science University (OHSU), Princess Margaret Cancer Centre (PMCC), Stanford University (SU), University of California San Francisco (UCSF), University of Iowa (UI), University of Michigan Center #1 (UM1), University of Michigan Center #3 (UM3), University of Pennsylvania (PENN), University of Pittsburgh Medical Center (UPMC), University of South Florida (USF), University of Texas at Austin (UTAUS), University of Washington (UWA), Vanderbilt University (VU), and Washington University in St. Louis (WUSTL).

DISCUSSION OF PROGRESS

There are currently five active projects, not including those that arose from the former DAWG, as they are included in the IAPM WG annual report. Four projects are nearing completion or in the final stages of analysis, while one project has newly commenced. The status and time-lines for these projects are presented in Figure 1.

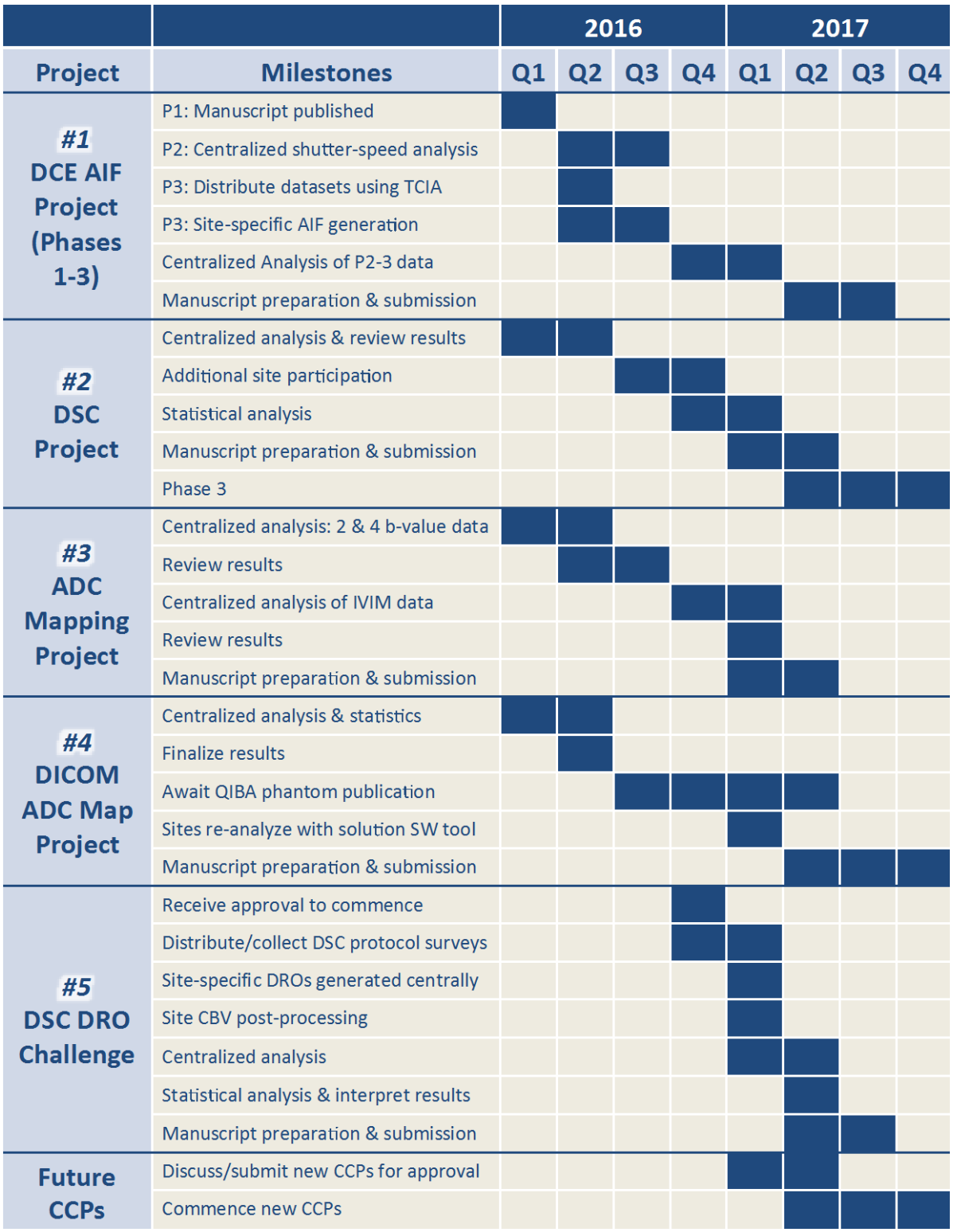


Figure 1: CCP milestones of MRI subgroup.

§ DCE Arterial Input Function Project 1 (lead institution: OHSU)

This 3-phase project developed out of previous challenge results published in 2014 [1], in which a static arterial input function (AIF) was used to compare pharmacokinetic (PK) parameters produced from various software (SW) algorithms. A manuscript [2] was published in 2016 detailing the results of the first phase of this project, where AIF was varied for the calculation of PK parameters utilizing a Tofts model [3] with centralized analysis performed using only one SW platform. The second phase of this project expanded on the first phase by comparing the results of AIF variation using the Shutter-Speed model [4]. In parallel with the second phase, the third phase examines the PK parameters using clinical therapy response endpoints to evaluate AIF selection method in the context of clinical efficacy.

Purpose

A large source of error in DCE PK parameter modeling is due to the uncertainty or error in the determination of an arterial input function. Therefore, the goal of this project is to evaluate variation in PK parameters based on AIF selection method, within the context of assessing cancer response to therapy.

Methods

There were 9 sites that participated in all phases of this challenge using dynamic imaging data that was hosted through The Cancer Imaging Archive (TCIA) [5]. For both phase 2 and 3, sites generated AIFs through use of their local SW tools and a centralized site (OHSU) performed PK analysis with the submitted AIFs, including reference tissue amplitude-adjusted AIFs [6]. The tumor ROI and pre-contrast T_1 were kept constant to measure PK parameter variations due solely to AIF estimation. In phase 2, PK parameters were evaluated in 11 DCE prostate datasets using a shutter-speed model at a single time-point. In phase 3, PK parameters were evaluated using a Tofts model in 7 DCE datasets of soft tissue sarcoma in the thigh, where the AIF was measured from the femoral artery by the QIN centers. Time-points included a baseline, visit 1 (V1) and visit 2 (V2, after one cycle of chemotherapy). In both phase 2 and 3, PK parameters were evaluated using Lin's Concordance Correlation Coefficient (LCCC). For phase 3 alone, V1, V2, and V21% PK parameter values were calculated and correlated with clinical response end points, where optimal response was set at greater than or equal to 95% necrosis in the surgical specimen (suboptimal <95%).

Results and Progress

Phase 2 results using a shutter-speed model were largely consistent with those achieved in the phase 1 evaluation, which utilized a Tofts model. There were considerable, and largely systematic, PK parameter variations observed due to AIF uncertainties, where K^{trans} had the largest and v_e the smallest AIF-caused variations. These variations were reduced when the AIF was adjusted to reference tissue. The initial AIF curve strongly influenced K^{trans} with extensive contrast agent extravasation. K_{ep} was less sensitive to AIF

uncertainty than K^{trans}, suggesting that it is a more robust biomarker. These results stress the importance of minimizing PK parameter variations in multicenter studies, which can be mitigated through central analysis using a fixed AIF, for single-time point studies, or using parameter percent changes, for longitudinal studies.

Phase 3 results of preoperative therapy time-points confirmed findings from both phase 1 and phase 2, in that v_e had the smallest AIF-induced variations and k_{ep} had less variation than K^{trans} (Figure 2). Interestingly, there were no decreases in variations for V21% of PK parameters. This suggests that random errors in AIF quantification may occur in a longitudinal study, such as occurs with ROI or voxel placement in the artery, partial volume averaging, and inflow effects, among others. It was encouraging that the uniform sign of correlations between visit 2 and visit 1 (V21%) metrics, especially V21% k_{ep}, and surgical specimen necrosis percentage across all AIF measurements exist, as this demonstrates the robustness of DCE-MRI for prediction of soft tissue sarcoma therapy response despite uncertainties in AIF determination. These results are promising but need to be evaluated in a larger cohort. Next steps include comparing individually measured AIFs using a population-averaged AIF within the context of DCE-MRI evaluation of therapy response.

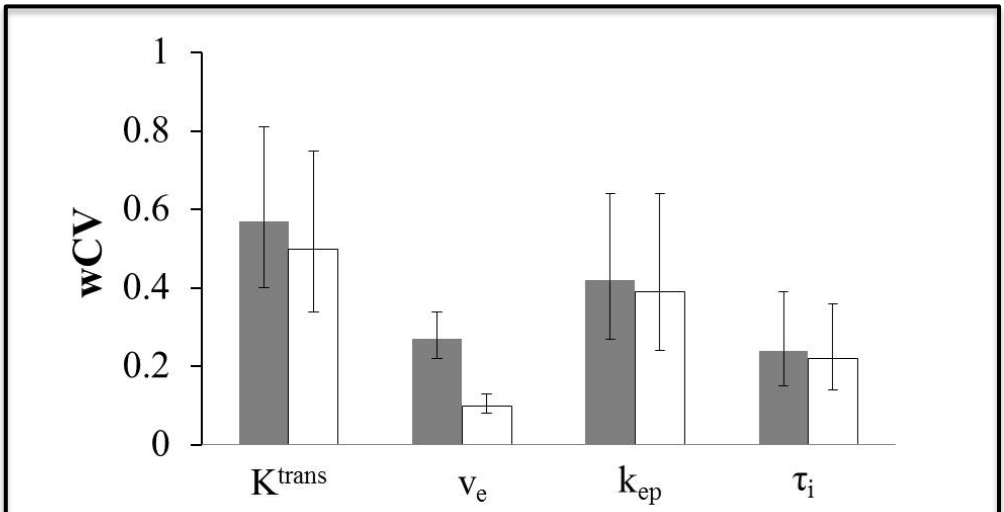


Figure 2: Column graph of wCV for the tumor mean K^{trans}, v_e, k_{ep}, and τ_i parameters obtained with the unadjusted (grey) and adjusted (white) AIFs and calculated using the shutter-speed model. The 95% confidence interval (CI) for each value is shown as error bars.

§ DSC Challenge Project 2 (lead institution: MCW)

The overall goal of this project is to reach a data-driven consensus for acquisition and post-processing of dynamic susceptibility contrast (DSC) MRI data in the evaluation of brain tumors. This project contains several phases, each using a different dataset, to not only compare post-processing algorithms, but to also address differences in the context of clinical relevance. Results were published from the first phase and focused on the

repeatability of post-processing and scaling methods [7]. This second phase focuses on data processing with respect to ground truth outcomes.

Purpose

The purpose of this phase of the project was to compare the software (SW) platforms used at several sites to achieve a consensus regarding post-processing of DSC-MRI data obtained in brain tumor patients. Each site was given the same dataset to process so that differences in acquisition would not contribute to any variations in the output parameter maps or ability to predict outcomes.

Methods

Co-registered DSC, anatomical images, and ROIs, including tumor, normal appearing cerebral cortex (NACC), reference normal appearing white matter, whole brain, and AIFs were uploaded to TCIA [5] in DICOM format for 49 pathologically confirmed low (13) and high-grade (36) brain gliomas, with outcomes blinded to all sites during processing. All datasets were acquired with a preload dose of contrast agent to diminish leakage effects [8]. Individual sites generated DSC-derived parameter maps, including relative cerebral blood volume (RCBV) and flow (CBF), using various SW platforms, including those that incorporated or excluded leakage correction algorithms [9]. There was some overlap of SW platforms among sites, where overall the sites produced 19 RCBV and 12 CBF parameter maps for each glioma case. Linn's Concordance Correlation Coefficient (LCCC) was used to assess agreement (good: $0.8<LCCC<0.89$; excellent: $LCCC>0.9$) between pairs of SW, and a ROC analysis was performed to identify the threshold that gave the best sensitivity and specificity for each method. Lastly, it was determined if one threshold existed that could provide a minimum sensitivity and specificity of 0.8 across all SW platforms and sites.

Results and Progress

Good or excellent agreement was observed for normalized RCBV (nRCBV) in tumor, in 19% or 75% of SW platforms, respectively. For normalized CBF (nCBF), good or excellent agreement in tumor was observed for 35% or 59% of SW platforms, respectively. Agreement was worse for NACC, where good or excellent agreement was 19% or 35% for RCBV, and 24% or 18% for nCBF, respectively. All SW platforms were able to distinguish low and high-grade glioma ($P<0.0001$), yet the thresholds resulting from a ROC analysis varied between 1.24-1.75 for nRCBV, and between 1.26-2.26 for nCBF. However, it was determined that all SW platforms could distinguish low and high-grade glioma with a sensitivity of at least 0.8 if a threshold of 1.45 is used, while a nCBF threshold of 1.84 could at best provide a sensitivity and specificity of only 0.64 (Figure 3). Overall, these results show that there is substantial consistency of DSC post-processing among SW platforms and sites for both nCBF and nRCBV, which was even greater when those SW platforms also incorporated a leakage correction algorithm, ultimately increasing clinical confidence in evaluation of perfusion MRI, with the condition that data are acquired as described. A manuscript is being written describing these results.

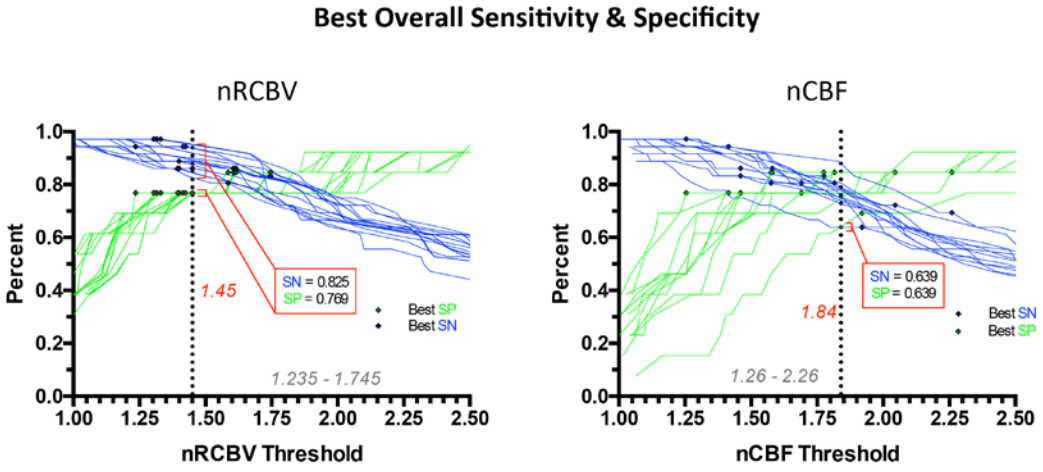


Figure 3: ROC curves for all SW entries for both nRCBV and nCBF. The best sensitivity and specificity for each individual method is indicated by the black diamonds shown on each ROC curve. A threshold of 1.45 for which all nRCBV entries maintained a SN/SP of at least 0.8 was identified. At best, all nCBF entries maintained a SN and SP greater than 0.639 for a threshold of 1.84.

§ ADC Mapping Project 3 (lead institution: UCSF)

Purpose

Reproducibility of diffusion metrics is essential given the increasing role quantitative diffusion weighted imaging plays in diagnosis and treatment monitoring. In addition, for validation and reproduction of results and meta-analyses in multi-center studies, it is essential that different implementations produce consistent results. The ADC Mapping CCP was undertaken to examine the variability in apparent diffusion coefficient (ADC) measures resulting from different post-processing software implementations utilized by researchers across the NCI Quantitative Imaging Network. A secondary aim was to evaluate the feasibility and practical challenges involved in centralized analysis of multi-center ADC data.

Methods

TCIA [5] was utilized to host 13 human [10] and NCIP-Hub [11] to host 4 phantom datasets, which included 2 and 4 b-value breast, 16 b-value liver, and multi-manufacturer 4 b-value polyvinylpyrrolidone (PVP) phantom [12] (provided by the sites from Quantitative Imaging Biomarker Alliance (QIBA) RSNA) DWI. There were 11 QIN sites that utilized 12 SW platforms to generate DWI-derived parameters that included mono-exponential ADC with 2, 4, or 16 b-values, perfusion minimized (excluding b=0) mono-exponential ADC and extrapolated perfusion fraction using 3 b-values, and IVIM models of perfusion fraction,

fast, and slow ADC using 16 b-values. Centralized analysis (UCSF) was performed using standardized regions of interest and included: 1) cataloguing capabilities of SW tools for processing multi-vendor quantitative DWI data, 2) evaluating linearity of ADC fit SW over order of magnitude ADC range provided by PVP phantom, and 3) evaluating concordance of parametric diffusion maps from different SW tools, including vendor-provided tools, for standardized ROIs. Concordance was evaluated from the percent difference of each measurement from the median value for all QIN sites. Pairwise within-subject coefficient of variation (wCV) was calculated for all site pairs and metrics to establish groupings of similar (wCV<0.1%) results.

Results and Progress

Preliminary analysis was completed in Fall 2016, and submitted as an abstract for the 2017 ISMRM meeting. Sample results for the 4 b-value breast ADC (all b-values) are shown in Figure 4. Inter-site wCV tables revealed eight of the sites were grouped into 2 separate groups: sites (1, 4, 13) with wCV<0.01% and sites (3, 5, 6, 8, 9) wCV<0.1%, while the other 4 sites and the scanner-generated maps showed more individualistic behavior. ADC values differed 2.8±0.2% between the two groups and up to 5% for non-grouped sites. The Philips scanner map had a 28% error due to inaccurate scaling information in the DICOM. Phantom results showed similar groupings amongst analysis implementations, though with smaller differences between the groups: RMS percent difference in ADC values for all phantom ROI of 0.29%, 0.30%, 0.62% for GEMS, Siemens, and Philips scans respectively. Full results are expected to be prepared for publication in 2017.

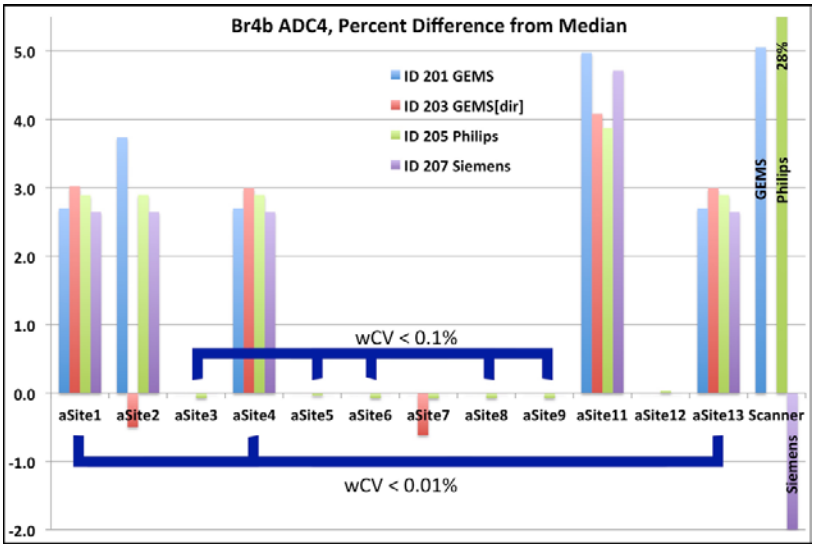


Figure 4: Sample results: Percent difference from median values for all sites for the mono-exponential ADC from the 4 b-value breast scans. Horizontal bars indicate the 2 groups of sites with close to identical results as measured by the wCV. Scanner-generated results are shown at the far right – note that the Phillips map had a large bias (28%) traced to problems with DICOM metadata.

§ DICOM ADC Parametric Map Project 4 (lead institutions: UM1, BWH, and UCSF)

Parametric map DICOM standards are being enforced to achieve uniformity and portability of QI metrics across vendors, sites and software tools in multi-site clinical trials. This project was launched as a supplemental effort, related to parent ADC Mapping challenge, to assess current capability of QIN participants and provide future guidelines to generate DICOM-compliant parametric diffusion maps. This project is a collaboration of the Bioinformatics/IT and Data Sharing (BIDS) WG in collaboration with the QIBA-RSNA DWI task force and major MRI equipment manufacturers.

Purpose

The purpose of this project is to demonstrate the ability to generate consistent quantitative parametric ADC maps and relevant DICOM metadata across vendors, sites, and SW tools.

Methods

Participating sites used a single multi-vendor DWI DICOM data set from a polyvinylpyrrolidone (PVP) phantom [12], imaged by three QIBA RSNA DWI task force members, to generate parametric ADC maps. The required attributes of parametric map DICOM were defined in collaboration with BIDS WG according to ITCR QIICR recommendations. Eight participating QIN sites (and two vendors) generated DICOM format ADC maps using a mono-exponential fit between zero and highest b-value, as well as all b-values. The resulting ADC map DICOM header metadata were evaluated by central site analysis and compared to general QIICR recommendations [13] and DWI DICOM macro [14].

Results & Progress

The CCP analysis has confirmed that (a) current vendor DWI DICOMs deviate from standard (Table 1), and (b) scanner-console (vendor-specific) ADC analysis software is not capable of parsing cross-vendor DWI DICOM. A majority of the participating QIN sites have resorted to using home-built DWI DICOM converter/parsers to derive ADC (**Table 2**). Only half of QIN sites that participated in the parent ADC mapping challenge have demonstrated an ability to generate parametric map DICOM files “off-line”. The CCP results (Table 2) show that no standard parametric ADC DICOM has been implemented by the community, and the standard source-image reference and ADC units/scale tags are mostly missing from the QIN site implementations. Furthermore, ADC fit parameters (e.g., b-values) and models (e.g., linear versus non-linear), deemed important for multi-site analysis in parent CCP, are missing both from the parametric map standard [13] and implementations, or are stored in private DICOM attributes. DICOM bit-depth of 12-16 has provided sufficient ADC precision across implementations. GE (vendor) ADC DICOM has been the closest to the standard and would provide a good starting template for on-site implementation.

Guided by current CCP findings and identified limitations, the final stage of the CCP will focus on inclusion of ADC fit parameters and models in parametric map standard and site DICOM dictionaries, and evaluation of recently implemented ITCR solution (DCMQI tools [15]) for uniform ADC DICOM generation across the QIN sites.

Table 1: Vendor-specific DICOMs lack uniformity & compliance with standard required DWI attributes.

Attribute\Vendor (SW)	Siemens (B17,B19,D13)	General Electric (v.12,15,23,24)	Philips (v.3,4,5)
(0018,9087)	private	private	public
Diffusion b-value	(0019,100C)	(0043,1039)	(0018,9087)
(0018,9075)			
Diffusion Directionality	(0019,100D)	(0043,1030); (0019,10BB/BC/BD)	(0018,9075)

Table 2: ADC map DICOM header comparison for participating sites shows inconsistent implementations.

Institution	NIST	MSKCC	VU	UM1	MCW	BNI	MS	UCSF	UWA	JHU
ADC SW Source	scanner	scanner(GE)/ functool(SM/PH)	scanner	off-line	off-line	off-line	off-line	off-line	off-line	off-line
Proc SW Name	TrioTim 3T	Discovery 3T/ Functool	Achieva 3T	QibaPhan (m-based)	IB Diffusion	Matlab/OSIRIX	Matlab	adcmap	ADCmap	Parametric Map Maker
SW version	Syngo MR B17	24\DX\MR; FCTL 11.3(PH) 14.3(SM)	5.1.7.1	1.3	2.0.1041	R2015b 8.6/7.0.2	R2015b 7.1	3 (IDL-based)	1.9	1
DWI parse (GE/SM/PH)	y/y/y (b only for SM)	y/y/y (b only for GE)	n/n/y	y/y/y	y/y/y	y/y/y	y/y/y	y/y/y	y/y/y	y/y/y
ADC maps (GE/SM/PH)	na/y/na	y/y/y (man b/ order for PH and SM)	n/n/y	y/y/y	y/y/y	y/y/y	y/y/y	y/y/y	y/y/y	y/y/y
2b/all-b ADC map	y/n	y/n	y/y	y/y	y/y	y/y	y/y	y/y	y/y	n/y
DCM SourceApplication	OFFIS_DCMTK_353	AW4 FCTL	Philips MR 51.0	MATLAB IPT 7.1	OSIRIX (GDCM/ITK 4.8.2)	OSIRIX	MATLAB IPT 7.1	MRSC_IMAGE_2 19	OSIRIX	orig(DWI)
ADC DCM (GE/SM/PH)	na/y/na	y/y/y	n/n/y	y/y/y	y/y/y	y/y/y	y/y/y	y/y/y	y/y/y	n/n/y
DERIVED/ADC (ImageType)	y/y	y/y	n/y	n/y	y/n	n/n	n/n	y/y	n/n	n/n
Format /BitsStored	uint16/12	int16/15(+/-)	uint16/12	uint16/16	uint16/15	int16/15(+/-)	uint32/31	int16/15(+/-)	uint16/15	int16(GE);uint16 /12 (PH/SM)
Input Scale (GE/SM/PH)	na/1e3/na	1e3/1e3/1e3	na/na/1.6e3	1e4/1e4/1e4	1e4/1.5e4/1.3e4	(2-8)e5/(4-5)e5/(4-5)e5	1e3/1e3/1e3	1e3/1e3/1e3	1e3/1e3/1e3	1e3/1e3/1e3
Scale Tag	na	(0008,0100)	(0028,1052/3)	(0008,103e)	(0028,1052/3)	(0028,1052/3)	na	na	(0028,1052/3)	(0028,1052/3)
Units	na	(0040,08ea)	(0028,1054)	(0008,103e)	na	na	na	na	na	na
Source Image	na	na	na	(0020,4000)	na	na	na	(0020,4000)	na	na
b-values	(0018,0024)	na	(0018,1030)	(0008,103e)	na	na	na	(0020,4000)	(0008,103e) no-vals	na
DWI Geometry (GE/SM/PH)	na/y/na	y/n/n	na/na/y	y/y/y	n/y/y	n/n/y	y/y/n	y/y/y	n/y/n	y/y/y

§ DSC Digital Reference Object Challenge Project 5 (lead institution: BNI)

As cerebral blood volume (CBV) values calculated from DSC-MRI are dependent on acquisition and post-processing methods, a validated DSC digital reference object (DRO) of the brain, containing both normal regions and tumor, was developed to serve as the first of it's kind test-bed for DSC-MRI data. In phase 2, optimal methods for post-processing and acquisition will be examined using both site-specific and standardized image parameter DROs.

Purpose

The primary goals of this work are two-fold: 1) to determine optimal post-processing methods and to identify post-processing steps that introduce variability to CBV

analysis and 2) to determine the variability in imaging scan parameters across sites and their influence on CBV analysis.

Methods

There are currently 7 sites participating in this challenge. To test the influence of imaging parameters on CBV, site-specific DROs are being generated based on survey responses, which identified site-specific acquisition protocols. Another DRO, containing standardized imaging parameters, will also be distributed to all sites to test the influence of post-processing using site-established methods. Therefore, all individual sites will perform post-processing on two DROs, which are being hosted by TCIA [5]. Central analysis (BNI) will be performed to determine accuracy and consistency. To evaluate accuracy between ground truth CBV and site-specific CBV (due to both imaging parameters and post-processing methods) Lin's Concordance Correlation Coefficient (CCC) will be calculated. To evaluate reproducibility across sites for CBV measurements a repeated ANOVA statistical test will be accomplished.

Results & Progress

A preliminary analysis of two acquisition methods demonstrates the influence of preload dosing schemes on CBV accuracy and precision. **Figure 5** illustrates a comparison of the accuracy (as determined by the CCC between the estimated and input CBV) and precision (as determined by the coefficient of variation (CV) between the estimated and input CBV) for a dosing scheme that delivers a standard full dose (0.1 mmol / kg) over two injections (½, ½) and one that uses two full dose injections.

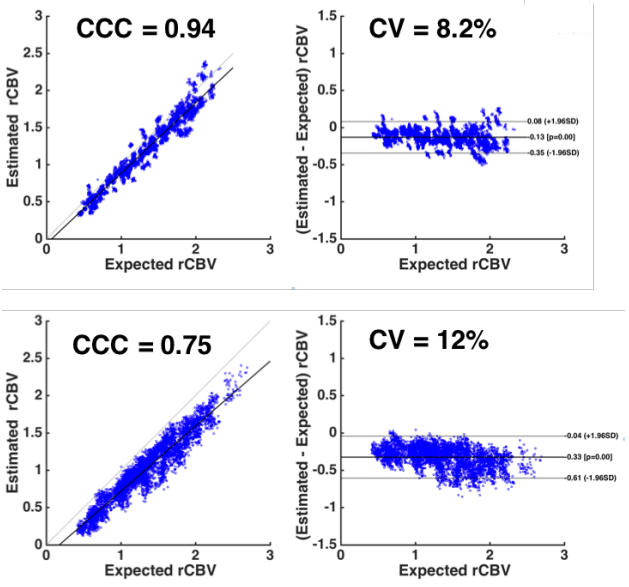


Figure 5: CBV accuracy for a dosing scheme that delivers a standard full dose (0.1 mmol / kg) over two injections (½, ½) and one that uses two full dose injections.

PLANS FOR NEXT YEAR

A priority in the current year is to finalize and submit manuscripts for the challenges that are currently completed or will be finished in Q1-2 of 2017. Specifically, there are six manuscripts that should be submitted for publication by Q2. Going forward, the MRI subgroup has discussed developing the following studies into CCPs to be initiated in 2017:

1. Undertake a new DSC-MRI challenge that focuses on the acquisition of the DSC-MRI data, rather than post-processing, which was the focus of the first challenge. Using the same approach described for the first DSC-MRI challenge we will make available to participating sites DSC-MRI datasets obtained using a variety of acquisition protocols with regard to contrast agent dosing and image settings such as flip-angle, TR and/or TE.
2. Assess influence of site- and/or vendor specific under-sampling pulse sequences on the fidelity of DCE-MRI.
3. Evaluate the influence of manually defined regions of interest on multi-site consistency of ADC measures.
4. Examine the repeatability and multi-site consistency of ADC measures in prostate cancer.
5. Challenge QIN sites to use their site-specific MRI post-processing methodologies to distinguish tumor from treatment effect or progression, using perfusion and diffusion datasets from patients with known outcomes. Any and all parameter types can be used in an attempt to determine which one(s) prove most predictive.

Beyond these specific action items, the MRI-subgroup commits to prioritizing initiatives and CCPs that focus on establishing and characterizing imaging acquisition and analysis methods that inform clinical decision-making (e.g. establishing a threshold for imaging biomarker changes that enable reliable treatment response detection). Validating the multi-site consistency of such tools is critical for increasing confidence in their utility and fostering their adoption into clinical use.

LIST OF REFERENCES

1. Huang, W., et al., *Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge.* Transl Oncol, 2014. **7**(1): p. 153-66.
2. Huang, W., et al., *The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge.* Tomography, 2016. **2**(1): p. 56-66.
3. Tofts, P.S., et al., *Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols.* J Magn Reson Imaging, 1999. **10**(3): p. 223-32.
4. Yankeelov, T.E., et al., *Variation of the relaxographic "shutter-speed" for transcytolemmal water exchange affects the CR bolus-tracking curve shape.* Magn Reson Med, 2003. **50**(6): p. 1151-69.

5. Clark, K., et al., *The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository.* J Digit Imaging, 2013. **26**(6): p. 1045-57.
6. Li, X., et al., *Cell membrane water exchange effects in prostate DCE-MRI.* J Magn Reson, 2012. **218**: p. 77-85.
7. Prah, M.A., et al., *Repeatability of Standardized and Normalized Relative CBV in Patients with Newly Diagnosed Glioblastoma.* AJNR Am J Neuroradiol, 2015. **36**(9): p. 1654-61.
8. Schmainda, K.M., et al., *Characterization of a first-pass gradient-echo spin-echo method to predict brain tumor grade and angiogenesis.* AJNR Am J Neuroradiol, 2004. **25**(9): p. 1524-32.
9. Boxerman, J.L., K.M. Schmainda, and R.M. Weisskoff, *Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not.* AJNR Am J Neuroradiol, 2006. **27**(4): p. 859-67.
10. *American College Of Radiology Imaging Network (ACRIN) 6698 Trial, Diffusion Weighted MR Imaging Biomarkers for Assessment of Breast Cancer Response to Neoadjuvant Treatment*. [cited 2016; Available from: https://www.acrin.org/6698_protocol.aspx
11. *NCIPHub*. Available from: https://nciphub.org.
12. *Boss M, Chenevert T, et al. Multicenter Study of Reproducibility of Wide Range of ADC at 0C. Proc. RSNA (2016) , Chicago, IL*

.
13. *Parametric map DICOM attributes*. [cited 2017 January 20]; Available from: http://dicom.nema.org/medical/dicom/2016c/output/chtml/part03/sect_C.8.32.2.html.
14. *MR Diffusion DICOM macro*. [cited 2017 January 20]; Available from: http://dicom.nema.org/medical/dicom/2016c/output/chtml/part03/sect_C.8.13.5.9.html.
15. *DICOM for Quantitative Imaging library*. [cited 2017 January 20]; Available from: https://github.com/QIICR/dcmqi - readme.

LIST OF QIN PUBLICATIONS AND PRESENTATIONS

§ Publications

1. Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, O'Sullivan F, Abramson RG, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Besa C, Kalpathy-Cramer J, Mountz JM, Laymon CM, Muzi M, Kinahan PE, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy FMM, Li X. The impact of arterial input function determination variations on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: a multicenter data analysis challenge. Tomography 2016;2:56-66.

§ Presentations

1. Huang W, Chen Y, Fedorov A, Li X, Jajamovich G, Malyarenko DI, Aryal M, LaViolette PS, Oborski MJ, O'Sullivan F, Abramson RG, Muzi M, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Besa C, Kalpathy-Cramer J, Mountz JM, Laymon CM, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy F, Li X. The impact of arterial input function determination variation on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: a multicenter data analysis challenge. Proc Intl Soc Magn Reson Med 2016;24:580.
2. Kathleen M Schmainda, Melissa A Prah, Scott D Rand, Mark Muzi, Swati D Rane, Xiao Da, Yi-Fen Yen, Jayashree Kalpathy-Cramer, Thomas L Chenevert, Dariya Malyarenko, Benjamin Hoff, Brian Ross, Yue Cao, Madhava P Aryal, Bradley Erickson, Panagiotis Korfiatis, Laura Bell, Leland Hu, and Christopher Chad Quarles. Multi-Site Concordance of DSC-MRI Analysis for Brain Tumors: Results of a NCI Quantitative Imaging Network DSC-MRI Collaborative Project. International Society of Magnetic Resonance in Medicine Annual Conference. Oral Presentation. 2017
3. Wei Huang, Yiyi Chen, Andriy Fedorov, Xia Li, Guido H. Jajamovich, Dariya I. Malyarenko, Madhava P. Aryal, Peter S. LaViolette, Matthew J. Oborski, Finbarr O'Sullivan, Richard G. Abramson, Kourosh Jafari-Khouzani, Aneela Afzal, Alina Tudorica, Brendan Moloney, Sandeep N. Gupta, Cecilia Besa, Jayashree Kalpathy-Cramer, James M. Mountz, Charles M. Laymon, Mark Muzi, Paul E. Kinahan, Kathleen Schmainda, Yue Cao, Thomas L. Chenevert, Bachir Taouli, Thomas E. Yankeelov, Fiona Fennessy, and Xin Li. Effects of AIF Quantification Variations on Shutter-Speed Pharmacokinetic Modeling of Prostate DCE-MRI Data: A Multicenter Data Analysis Challenge, Part II International Society of Magnetic Resonance in Medicine Annual Conference. Oral Presentation. 2017
4. Kimberly Li, Yiyi Chen, Yun Yu, Xia Li, Andriy Fedorov, Guido H. Jajamovich, Dariya I. Malyarenko, Madhava P. Aryal, Peter S. LaViolette, Matthew J. Oborski, Finbarr O'Sullivan, Richard G. Abramson, Kourosh Jafari-Khouzani, Aneela Afzal, Alina Tudorica, Brendan Moloney, Sandeep N. Gupta, Cecilia Besa, Jayashree Kalpathy-Cramer, James M. Mountz, Charles M. Laymon, Mark Muzi, Paul E. Kinahan, Kathleen Schmainda, Yue Cao, Thomas L. Chenevert, Bachir Taouli, Fiona Fennessy, Thomas E. Yankeelov, Xin Li, Christopher Ryan, Wei Huang. The Effects of AIF Quantification Variations on DCE-MRI Prediction of Soft Tissue Sarcoma Response to Preoperative Therapy: A Preliminary Multicenter Study International Society of Magnetic Resonance in Medicine Annual Conference. Poster Presentation. 2017
5. David C Newitt, Dariya Malyarenko, Thomas Chenevert, C.Chad Quarles, Laura Bell, Andrey Fedorov, Fiona Fennessy, Michael A Jacobs, Meiyappan Solaiyappan, Stefanie Hectors, Bachir Taouli, Kathleen M Schmainda, Melissa A Prah, Yi-Fen Yen, Jayashree Kalpathy-Cramer, Erin Taber, Christopher Kroenke, Yue Cao, Madhava Aryal, Mark Muzi, Paul Kinahan, Thomas E Yankeelov, Lori R Arlinghaus, Michael A Boss, Amita Shukla-Dave, and Nola Hilton. Multi-site Concordance of DWI Metrics: Results of the NCI Quantitative Imaging Network

ADC Mapping Collaborative Project" International Society of Magnetic Resonance in Medicine Annual Conference. Oral Presentation. 2017

QIN PET-CT subgroup

Lubomir Hadjiyski, PhD
Dmitry Goldgof, PhD

INTRODUCTION

Mission Statement

The PET-CT subgroup is a subgroup of the Image Analysis and Performance Metrics Working Group, whose mission statement is:

> *The mission of the Image Analysis & Performance Metrics Working Group (IAPMWG) is to provide guidance, coordination, consensus building, and awareness regarding the development of algorithms and methods for quantitative analysis of tumors, related tissues and organs, and changes in response to disease progression and treatment, as well as to influence the development of sharable objective methods and metrics for assessment of image analysis accuracy, reproducibility, and robustness. The IAPMWG will coordinate the collaboration between QIN members in this area.*

Subgroup activities focus on quantitative image analysis applications to CT, PET-CT, and dynamic PET data in several clinical domains, including lung cancer and head & neck cancer. Our major efforts to stimulate the collection and sharing of tools, and analysis and evaluation methods, has been through the development of "challenges" , of which there are four active at this time (described below). In addition, we have participated in cross-WG activities with the Bioinformatics and Data Sharing (BIDS) WG which are also described below.

Accomplishments of the Previous Year

§ CT Feature Comparison Challenge using Moist Run Data (Hosted by USF/Moffitt CC)

The goal of the CT Feature Comparison Challenge is to investigate the sensitivity of quantitative descriptors of pulmonary nodules to segmentations and to illustrate comparisons across different feature types and features computed by different implementations of feature extraction algorithms.

Radiomics is to provide quantitative descriptors of normal and abnormal tissues during classification and prediction tasks in radiology and oncology. Quantitative Imaging Network members are developing radiomic "feature" sets to characterize tumors, in general, the size, shape, texture, intensity, margin, and other aspects of the imaging features of nodules and lesions. Efforts are ongoing for developing an ontology to describe radiomic features for lung nodules, with the main classes consisting of size, local and global shape descriptors, margin, intensity, and texture-based features, which are based on wavelets, Laplacian of Gaussians, Law's features, gray-level co-occurrence matrices, and run-length features.

Methods: The features generated from different computer segmentations will be evaluated for repeatability between repeated runs of each algorithm, and reproducibility across segmentation algorithms [5]. This is being done by using the 5 collections of DICOM CT images that were used for the Moist Run Challenge [1] and the segmentations (3 algorithms with 3 repeat trials) created there.

This work expands on that of the Moist Run segmentation challenge [1] as well as the work on feature reproducibility on test-retest data [2] [3] and numerous papers on various measures [4], [5] for feature stability measures, including repeatability and reproducibility. It should be noted that this challenge is being run on C-BIBOP (a U24 project funded by the ITCR associated with the QIN) and that R-scripts will be run to compute commonly used metrics of repeatability and reproducibility. We will coordinate these submissions with separate QIN effort to compare and harmonize features across sites (Daniel Rubin).

Seven QIN teams (Columbia, Moffitt, Stanford, UCLA, Iowa, Princess Margaret, and Michigan) obtained the data set, including all images and segmentations, from TCIA, and each computed their own set of features for each of 468 segmentations ((4*10+12)*3*3) and uploaded them to a NCIP HUB using a standard format.

Results: We calculated the concordance correlation coefficients of the features as a measure of their stability with the underlying segmentation; 68% of the 830 features in this study had a concordance CC of $\geq$0.75. Pairwise correlation coefficients between pairs of features were used to uncover associations between features, particularly as measured by different participants. A graphical model approach was used to enumerate the number of uncorrelated feature groups at given thresholds of correlation (Fig. 1). At a threshold of 0.75 and 0.95, there were 75 and 246 subgroups, respectively, providing a measure for the features' redundancy.

Conclusion: Having a common set of reference images, well-specified objects and existing object masks allowed to focus on the very specific task of feature computation, its sensitivity to segmentation results, and the associations among specific features. High correlations between certain groups of features calculated across participants were observed. There is substantial value in comparing feature values among different groups, even when the feature values are expected to be the same or very similar, allowing to uncover subtle differences and even errors in approach and calculations that may not have been discovered otherwise. This study also showed the value of using phantom images or synthetic images where there are objects with known values such as known density or known volume. These provide users the ability to gain confidence that their methods and calculations are performing in a manner similar to some reference methods.

This challenge is successfully completed and a manuscript is published in the *QIN special issue of Tomography*, December 2016, [10].

§ Lung Nodule Interval Segmentation Challenge using NLST data (Hosted by USF/Moffitt CC)

The goal of this project is to study the variability of segmentation methods in estimating the size of the pulmonary nodules on scans of the same patient at two different time instances.

Lung cancer has been one of the leading cancer deaths in the US. It has been shown that early detection of the lung cancer improves survival and patient prognosis. Estimation of size of these lung nodules during screening is an important clinical factor in the determination of patient follow-up procedures. In an effort to quantify variability among different users in segmentation of these nodules over time, we conducted a friendly challenge among the QIN members to segment and estimate size of lung nodules using their preferred methods.

Methods: We have downloaded the National Lung Screening Trial (NLST) data table (6) and assembled patients with reported abnormalities (incident cancer cohort or just cancer cohort: CC) at one of the three time point scans and epidemiologically matched against diagnosed as normal (7,8). The CT images for the identified cohorts have been downloaded from NLST trial. Scans were collected across the NLST (some 50+ institutions around the US) and patients enrolled in the trial were known to be at elevated risk of lung cancer (as required by NLST inclusion criteria) due to smoking history and age (9).

In this challenge CT image data was assembled from 100 subjects imaged at two time points (baseline and follow up) approximately one year apart (200 total CT datasets). Fifty subjects were confirmed to have cancer and 50 subjects were with nodules which were determined to be non-cancer. Nodules in both cancer and non-cancer cases were identified and approximate locations were supplied to the sites along with image data (Figure 2). The final diagnosis and exact coordinates or nodule slice centers were not shared to avoid any potential bias between participants. In patients with multiple nodules, we selected one that was largest (by univariate diameter) at baseline. Participants were asked to segment the nodules in each of the 100 cases using their preferred segmentation approach (automated, semi-automated, etc.). Each participating site agreed to submit at least one set of segmentation results where no editing of the resulting segmentations was performed. Sites were provided an option to submit additional results where editing of the nodule boundaries was allowed and these will be analyzed separately.

We have five participating sites (Moffitt, Columbia, Michigan, Harvard, UCLA) with two additional sites participating with their analytics expertise (Stanford, MGH). All teams have segmented the challenge data sets and reported the segmentation masks. The details of the challenge have been made available for public view via the NCIPHUB, while the results and data are restricted to the participants who comply and signed the NCI's DTA to access the NLST data. URL for the Interval Challenge:

https://nciphub.org/publications/20/versions?v=1.

Results: In our preliminary analysis we find similarity in segmentation (DICE coefficients) between all five participants has wide ranges, with a mean of 0.48 [Range: 0 (no overlap) to 0.97]. The concordance in volume estimates between any two sites range from [0.71 to 0.95] while the volume change range from [0.15 to 0.89]. Using a logistic regression, we found that the ideal volume increase threshold for predicting cancer status ranged from 15% to 35% in well-performing sites. While the prediction accuracy (AUC) of the sites based on volume change estimates ranges from [0.64 to 0.82]. Further, we find categorizing these nodules by baseline size (≥ 8mm) improve participants' prediction with an AUC range of

[0.75 to 0.9]. While for smaller nodules (baseline ≤ 8mm) the prediction accuracy shows slightly lower performance, with AUC range of [0.57 to 0.8]. While overall prediction accuracy was comparable between both manual and automatic segmentations, agreement of cancer status classification on specific nodules varied from algorithm to algorithm. Highest classification agreement as measured by Cohen's kappa statistic was found between the two manual segmentation submissions (0.68), while lower agreement was found between automatic segmentations (0.17 to 0.54).

Conclusion: We are at the advanced inference phase and we expect to submit a peer-reviewed publication on this challenge by middle of 2017. We have also developed a visualization platform for data derived from this competition using RStudio's Shiny package, which can currently be publicly viewed at http://cbibop.cloudapp.net:3838/Interval_Lung_Challenge_ShinyApp/. Note that the contents of this visualization may change in the coming weeks. An illustration of the visualization platform is shown in Figures 3, 4, and 5.

§ PET Segmentation Challenge (Hosted by Iowa)

The goal of this project is to perform segmentations on PET scans of several objects, starting from Digital Reference Objects (DROs) progressing to phantoms and ultimately to patient scans of head and neck tumors, to assess the bias and variability and to determine the impact on derived Quantitative Imaging measures.

Radiomics utilizes a large number of image-derived features for quantifying tumor characteristics that can in turn be correlated with response and prognosis. Unfortunately, extraction and analysis of such image-based features is subject to measurement variability and bias. The challenge for radiomics is particularly acute in Positron Emission Tomography (PET) where limited resolution, a high noise component related to the limited stochastic nature of the raw data, and the wide variety of reconstruction options confound quantitative feature metrics. Extracted feature quality is also affected by tumor segmentation methods used to define regions over which to calculate features, making it challenging to produce consistent radiomics analysis results across multiple institutions that use different segmentation algorithms in their PET image analysis. Understanding each element contributing to these inconsistencies in quantitative image feature and metric generation is paramount for ultimate utilization of these methods in multi-institutional trials and clinical oncology decision making.

Methods: To assess segmentation quality and consistency at the multi-institutional level, we conducted a study of 7 institutional members of the National Cancer Institute Quantitative Imaging Network (Columbia, Moffitt, MSKCC, Simon Fraser University, Pittsburgh, Iowa, and Washington). For the study, members were asked to segment a common set of phantom PET scans (Figure 6) acquired over a range of imaging conditions as well as a second set of head and neck cancer (HNC) PET scans (Figure 7). Segmentations were generated at each institution using their preferred approach. In addition, participants were asked to repeat segmentations with a time interval between initial and repeat segmentation. This procedure resulted in overall 806 phantom insert and 641 lesion segmentations.

Subsequently, the volume was computed from the segmentations and compared to the corresponding reference volume by means of statistical analysis.

Results: On the two test sets (phantom and HNC PET scans), the performance of the seven segmentation approaches was as follows. On the phantom test set, the mean relative volume errors ranged from 29.9 to 87.8% of the ground truth reference volumes, and the repeat difference for each institution ranged between -36.4 to 39.9%. On the HNC test set, the mean relative volume error ranged between -50.5 to 701.5%, and the repeat difference for each institution ranged between -37.7 to 31.5%. In addition, performance measures per phantom insert/lesion size categories are given in the paper. On phantom data, regression analysis resulted in coefficient of variation (CV) components of 42.5% for scanners, 26.8% for institutional approaches, 21.1% for repeated segmentations, 14.3% for relative contrasts, 5.3% for count statistics (acquisition times), and 0.0% for repeated scans. Analysis showed that the CV components for approaches and repeated segmentations were significantly larger on the HNC test set with increases by 112.7% and 102.4%, respectively.

Conclusion: Analysis results underline the importance of PET scanner reconstruction harmonization and imaging protocol standardization for quantification of lesion volumes. In addition, to enable a distributed multi-site analysis of FDG PET images, harmonization of analysis approaches and operator training in combination with highly automated segmentation methods seems to be advisable. Future work will focus on quantifying the impact of segmentation variation on radiomics system performance.

This challenge is successfully completed and a manuscript is published in *Medical Physics*, February 2017, [11].

§ Dynamic PET FMISO Challenge (Hosted by Sadek Nehmeh/MSKCC)

The goal of this project is to assess the inter-observer variability in the compartmental kinetic analysis (CKA) of ^{18}F-Fluoromisonidazole (FMISO) dynamic positron emission tomography (PET) images. Specifically, MSKCC will share static FDG PET/CT and dynamic FMISO PET data from five NSCLC patients with the institutions that participate in this challenge. Each institution will use its own approaches to conduct kinetic analysis. The variability in target volume definition, input function, output function, and kinetic rate constants computed by the participating institutions will be compared.

Methods: The inter-operator variability in CKA due to; (1) differences in mathematical modeling; and, (2) difference in full CKA process was assessed. In (1), twenty-three tumor Time-Activity-Curves (TACs) with the corresponding input functions were deduced from dynamic FMISO PET studies in patients diagnosed with non-small cell lung cancer (NSCLC), and shared with five experts in CKA from four institutions members of QIN (MSKCC, Johns Hopkins, Pittsburgh, Princess Margaret). Each of the operators carried out, independently, CKA for each of the datasets. In (2), CKA of FMISO dynamic PET images from four NSCLC patients was carried out by each of four of the operators. Target volumes, input functions, output functions, and the deduced KRCs deduced by the four operators were compared. In all cases, an irreversible one-plasma two-tissue compartment model was used.

Results: For study-I, strong Interclass correlation (ICC>0.9) was measured for all KRCs. Similarly, strong Pearson correlation (R>0.75, P<0.001) was observed among the operators for all KRCs (V_b, K_1, $K1/k_2$, and k_3). No systematic or proportional biases could be identified for any of the four KRCs. The average changes in K_1 and k_3 were 0.0114 and 0.00021 respectively, and the corresponding 95% Limits of Agreement (LoA) were (-0.163 to 0.118) and (-0.00575 to 0.00617) respectively.

For study-II, a weak ICC = 0.36 for the segmented tumor volumes was observed; the coefficients of variability in the segmented tumor volumes among the four operators were 30.2%, 92.8%, 52.2%, and 103.3% respectively. The ICC's for the KRCs were: ICC-V_B= 0.53; ICC-K_1= 0.91; ICC-K_1/k_2=0.25; ICC-k_3= 0.32. However, Passing-Bablok analysis showed those results to be interchangeable, but for k_3, among the four operators with no systematic or proportional biases. The average changes in K_1 and k_3 were ~0.016 and ~0.00075 respectively, and the corresponding 95% LoA were (-0.582 to 0.357) and (-0.00485 to 0.00634) respectively.

Conclusions: KRCs were mostly reproducible when CKA was carried out by multiple operators. A major source of error is the target volume definition which yields modifying the corresponding Time Activity Curve.

This challenge is successfully completed and a manuscript will be submitted to *Journal of Nuclear Medicine* by middle of 2017.

Plans for the Next Year

Expected contributions to the literature, to shared data, and to shared tools include:

- Dynamic PET w/FMISO paper (submitted by 2nd Q 2017).
- Lung Nodule Interval Segmentation Challenge (submitted by 2nd Q 2017).

Expected new projects include:

- Extend the interval change challenge – extract features, test their efficiency for prediction of two classes (normal, abnormal), study the stability of the features across different institutions.
- PET Image feature challenge – study the stability of the PET features across different institutions.
- Community-based terminology standards – definition of feature dictionaries, ontologies, lexicons.

Activities Across Working Groups

We worked closely with Daniel Rubin (and the BIDS group) on the feature nomenclature project. For the Feature Comparison Challenge, we have developed a two level description of features (a feature class and a feature subclass). This project will give us an opportunity to determine how robust these descriptions are and to assess how this can be improved. This project is important as image-based features described in the literature are

often described imprecisely and this makes repeating published results difficult, so that successes can be hard to replicate. With a more consistent naming and description approach, some of these issues will be addressed.

Other expected contributions to the QIN will be derived from the studies being performed, which are expected to provide robust guidance with respect to approaches to minimize bias and variance in image segmentation and feature calculations.

List of References

[1] J. Kalpathy-Cramer, B. Zhao, D. Goldgof, Y. Gu, X. Wang, S. Napel, H. Yang, Y. Tan, "A platform for the comparison of lung nodule segmentation algorithms: methods and preliminary results" Radiology Society of North America (RSNA) Annual Meeting, Chicago, IL, 12/2013.

[2] Y. Balagurunathan, V. Kumar, Y. Gu, J. Kim, H. Wang, S. Basu, R. Korn, B. Zhao, D. Goldgof, L. Hall, L. Schwartz, S. Eschrich, R. Gatenby, R. Gillies, Test-Retest reproducibility analysis of lung CT image features. Journal of Digital Imaging, 27(6), 805 - 823, Dec 2014.

[3] Y. Balagurunathan, Y. Gu, H. Wang, V. Kumar, O. Grove, S. Hawkins, J. Kim, D. Goldgof, L. Hall, R. Gatenby, R. Gillies, "Reproducibility and Prognosis of Quantitative features extracted from CT images", Translational Oncology, vol. 7(1), pp. 72-87, 2014.

[4] Huiman X Barnhart, Eric Yow, Anna Lisa Crowley, Melissa A Daubert, Dawn Rabineau, Robert Bigelow, Michael Pencina, and Pamela S Douglas, “Choice of agreement indices for assessing and improving measurement reproducibility in a core laboratory setting”, Stat Methods Med Res 0962280214534651, first published on May 14, 2014 doi:10.1177/0962280214534651

[5] N. Obuchowski, A. Reeves, E. Huang, X. Wang, A. Buckler, H. Kim, H. Barnhart, E. Jackson, M. Giger, G. Pennello, A. Toledano, J. Kalpathy-Cramer, T. Apanasovich, P. Kinahan, K. Myers, D. Goldgof, Daniel P. Barboriak, R. Gillies, L. Schwartz, D. Sullivan, "Quantitative Imaging Biomarkers: A Review of Statistical Methods for Computer Algorithm Comparisons", Statistical Methods in Medical Research 0962280214537390, first published on June 11, 2014 as doi:10.1177/0962280214537390

[6] Kramer BS, Berg CD, Aberle DR, Prorok PC. Lung cancer screening with low-dose helical CT: results from the National Lung Screening Trial (NLST). Journal of medical screening. 2011;18(3):109-11.

[7] Schabath M, Thompson, Balagurunathan, Goldgof, Eschrich, Gillies. Survival of patients with incident lung cancer following screening by computed tomography in National Lung screening Trial. Annual American Association for Cancer Research (AACR). San Diego2014.

[8] Schabath M, Thompson, Balagurunathan, Goldgof, Eschrich, Gillies. Behavior of incident lung cancer in CT arm of National Lung screening Trial. JAMA (in review).

[9] NLST. https://biometry.nci.nih.gov/cdas/nlst/learn/trial_sum/

[10] Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren J C, Uthoff J, Dilger S K N, Driscoll B, Yeung I, Hadjiiski L, Cha K, Balagurunathan Y, Gillies R, Goldgof D. “Radiomics of Lung

Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features". Tomography. 2016; 2(4): 430-437.

[11] Beichel RR, Smith BJ, Bauer C, Ulrich EJ, Ahmadvand P, Budzevich MM, Gillies RJ, Goldgof D, Grkovski M, Hamarneh G, Huang Q, Kinahan PE, Laymon CM, Mountz JM, Muzi JP, Muzi M, Nehmeh S, Oborski MJ, Tan Y, Zhao B, Sunderland JJ, Buatti JM. "Multi-site quality and variability analysis of 3D FDG PET segmentations based on phantom and clinical image data". Medical Physics. 2017; 44(2): 479-496.

Figures

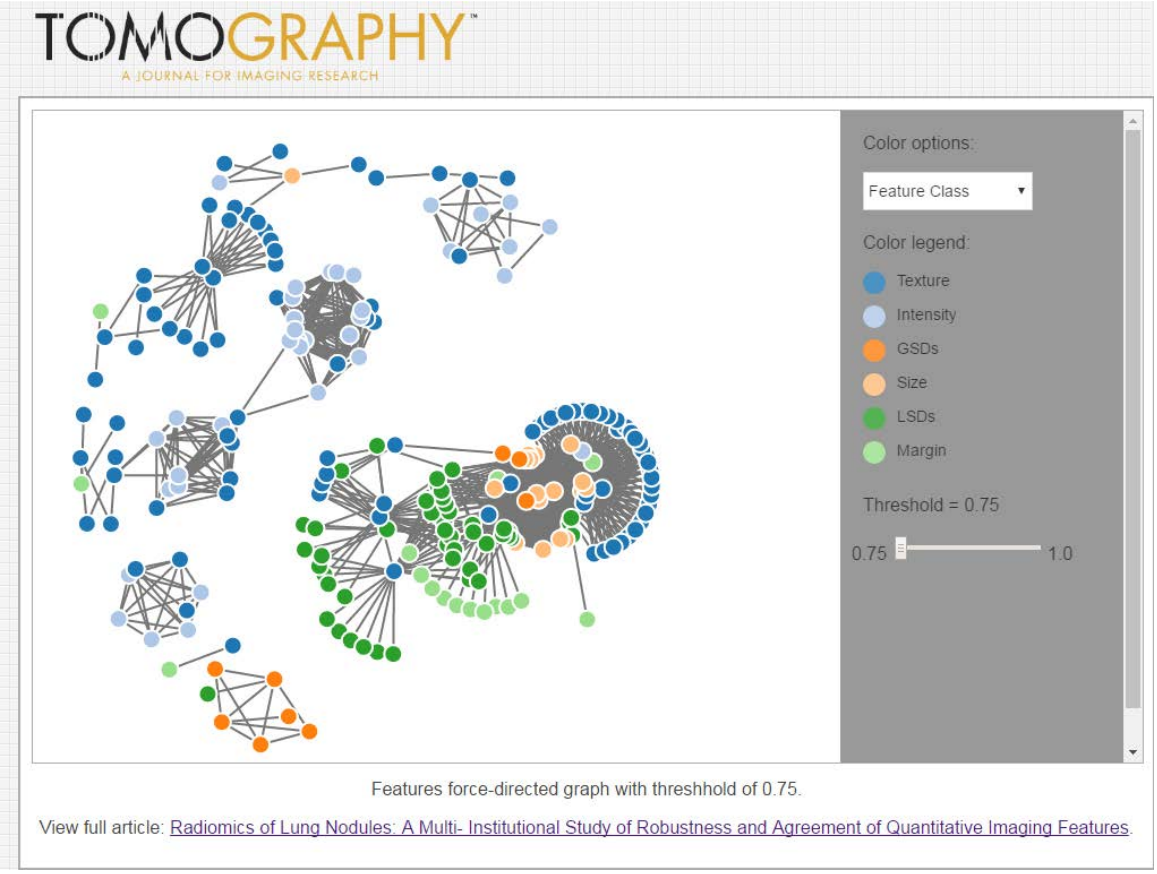


Figure 1. Graphical model of connected components at a CC of 0.75.

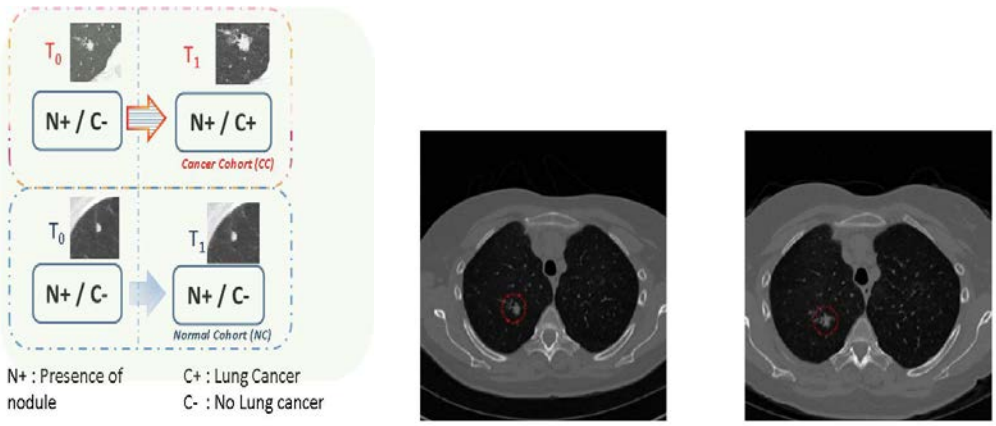


Figure 2. Illustration of the Interval challenge with an example patient scan at two time points.

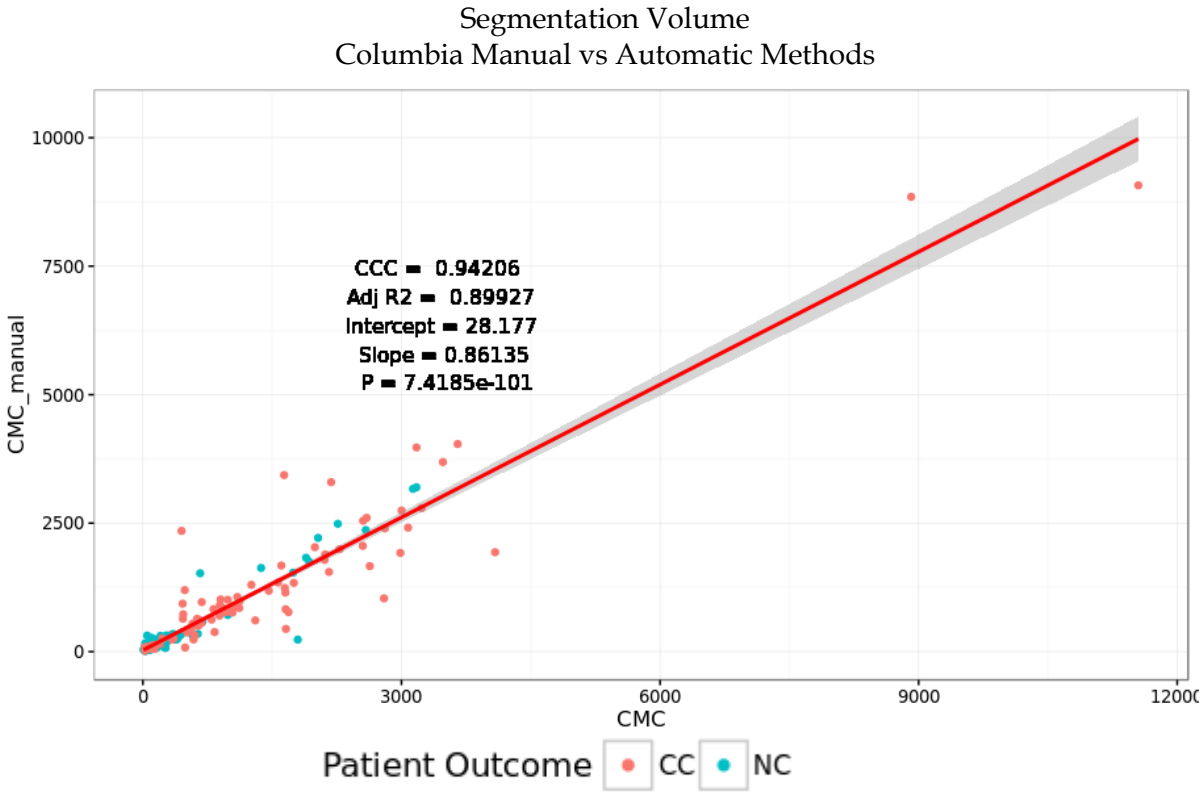


Figure 3. Chart of the correlation between volume estimates in Columbia's automatic and manual nodule segmentations. CC indicates lung cancer, while NC indicates no cancer.

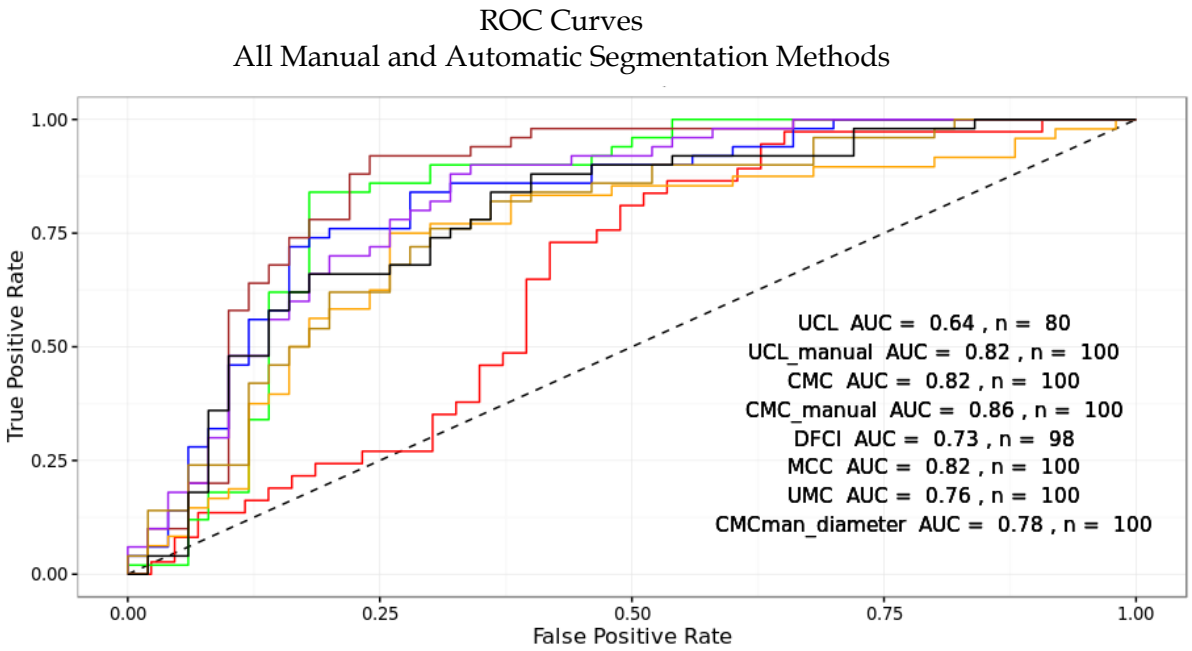


Figure 4. ROC Curves for all manual and automatic segmentation methods.

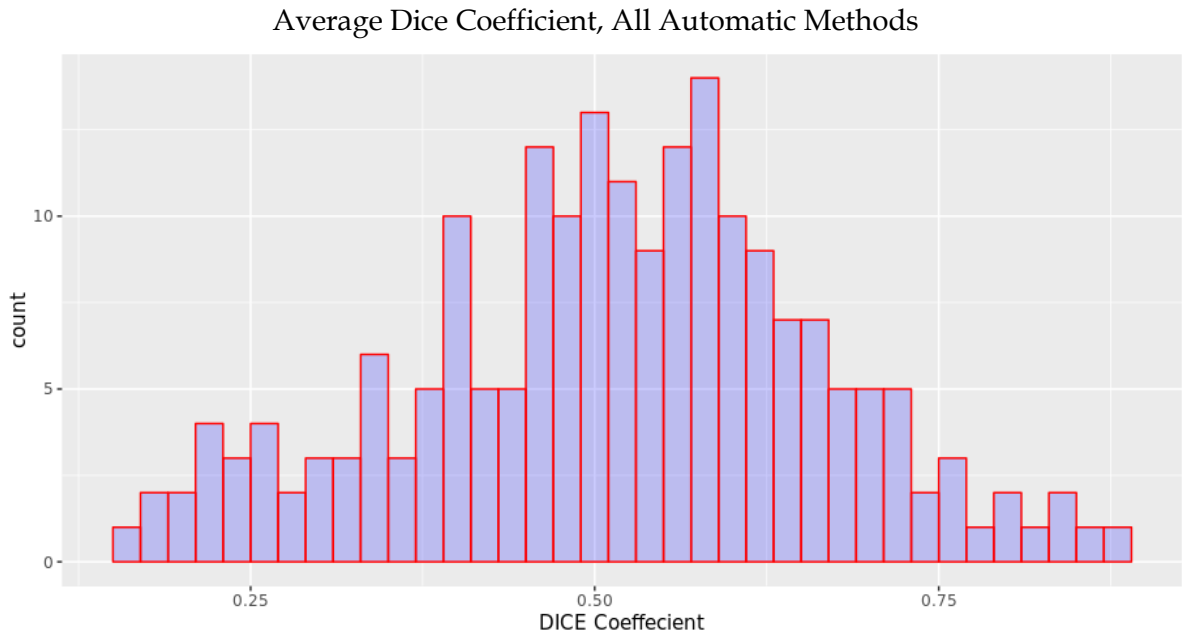


Figure 5. A histogram of DICE coefficients aggregated from all combinations of segmentations from all nodules.

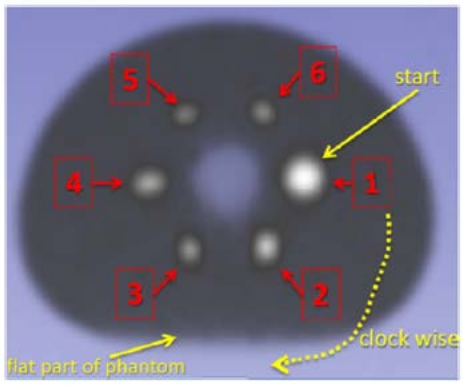


Figure 6. QIN PET-CT Phantom Challenge

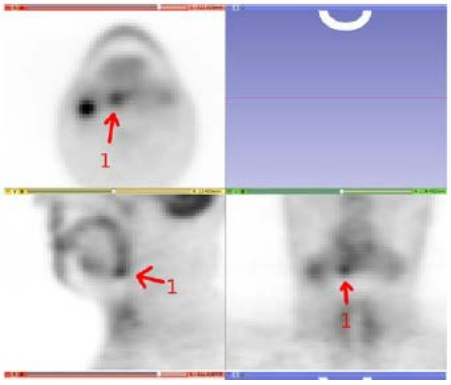


Figure 7 – QIN PET-CT Head and Neck Lesion Challenge