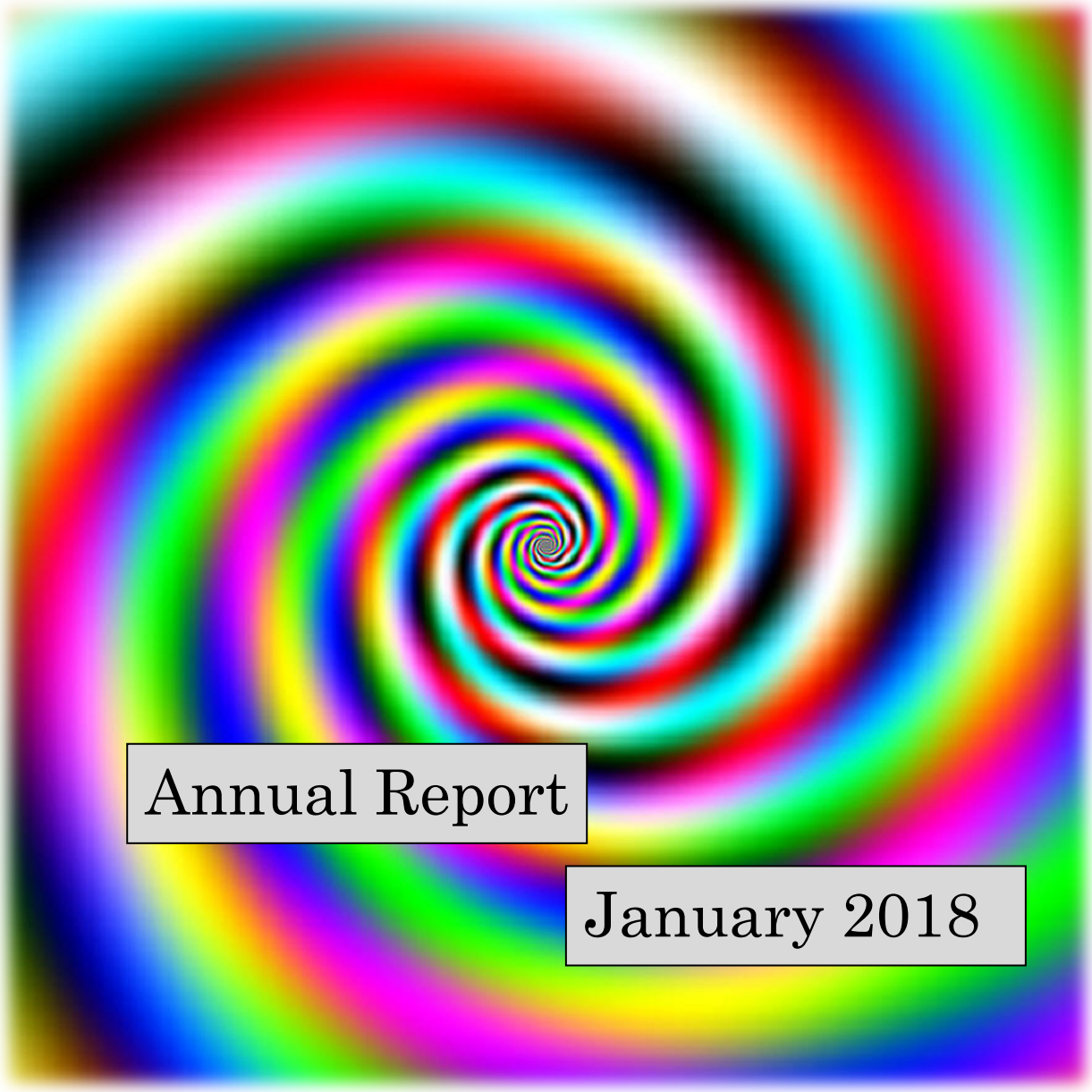
National Cancer Institute

The Cancer Imaging Program

The Quantitative Imaging Network



U.S. Department of Health & Human Services | National Institutes of Health

INTRODUCTION TO THE SIXTH ANNUAL REPORT

OF THE

QUANTITATIVE IMAGING NETWORK (QIN)

BACKGROUND

For the past nine years, the goal of the Quantitative Imaging Network (QIN) has been to develop novel approaches for quantitative imaging methods and clinical decision support systems. Over the past year, however, a major redirection of goals is taking place within the Network. While development and validation of quantitative imaging clinical decision tools remains important, new emphasis is being placed on insertion of completed tools in clinical trial environments. This has required close discussion with National Clinical Trial Network (NCTN) teams such as ECOG-ACRIN, Alliance, and NRG. QIN members have participated in meetings where the use of decision tools in specific trials has been discussed. While initial results are encouraging, more work will have to be done before wide acceptance of quantitative imaging tools in clinical trials is a reality.

This evolution from development and validation to promoting clinical utility of QIN tools requires a new thought process for the members of the Network. The technical challenges of developing robust algorithms for reducing measurement errors and extracting measurable information from clinical images must proceed along with plans for integrating the tools into clinical utility without disrupting the flow of care.

AIMS OF THE NETWORK

The traditional aims of the Network have been

- To develop advanced and harmonized methods for data collection across different commercial imaging platforms for a range of imaging modalities (CT, PET CT, DW or DEC MRI, or SPECT). These methods are required to reduce the physical measurement uncertainty across different clinical trial sites. This will have the potential to:
    - o Reduce the size of a planned clinical trial for a statically significant result, and
    - o Improve the ability of data collection methods to support robust data integration from each modality and thus enhance the performance of clinical decision support systems.

- To develop advanced methods for data modeling, analysis, display, and workflow pipelines. An informatics platform is required to compare the relative performance of these methods developed by each QIN team, namely for each targeted cancer problem. An important goal is to reach a consensus on the selection of the best QI analysis methods, and provide an informatics structure to share the results of these methods

among imaging physicians involved in the optimization of QI methods in clinical trials. The latter includes the newly formed National Clinical Trial Network (NCTN), planned U10's or U24's imaging centers, other clinical trials supported by NCI Cancer Centers, or other nationally or internationally supported imaging trial networks.

- The strategy to reach these aims is to provide an "open science network" where various oversight committees and working groups are tasked to help create a consensus on best practices to reach these objectives. The network "open science goals" are designed to allow these translational research objectives to be met while permitting the network teams to maintain their intellectual property for the clinical decision support systems being developed by each QIN team. The latter is a critical requirement to permit the QIN industrial partners to seek commercial and FDA approved QI solutions as required to support Phase 3 radiation therapy or drug trials.

- To develop a public resource for data collected by QIN teams, posted on The Cancer Imaging Archive (TCIA), and a depository for exchanging informatics tools generated by QIN. These resources are required to support planned "Grand Challenges" to compare the relative performance of the clinical decision support systems by QIN Teams. The data and tools will initially be restricted to QIN members, until methods developed are suitable for other researchers in the broader research community to evaluate, and participate in these Grand Challenges. The image archives and informatics resources are supported by CIP and NCI CBIIT.

In addition to these aims dealing with development and validation new Network aims are formulating along the lines of clinical utility of quantitative imaging tools. These are still in the formative stage, and different groups within the Network are looking into connection with NCTN components.

GOAL OF THIS REPORT

The generation of this sixth annual report from QIN, covering the period from January 2017 to January 2018, serves the following purposes:

- As a detailed scientific report or reference document for the academic and industrial community, to serve as a resource for how to translate an important technology into the clinical trial and clinical practice setting. It underpins the fact that as technologies move toward multimodal molecular imaging, the validation strategies become more complex than the development of the technology itself. Thus, the research community needs to be more fully aware of their translational complexity and seek common solutions using test beds created by QIN and other research networks.

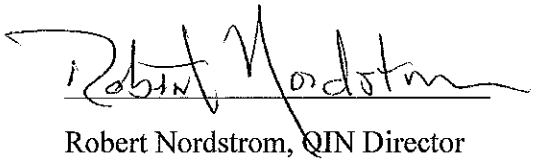
- To serve as a comprehensive reference report for imaging research to the other federal agencies such as the following:

    - FDA, where this report may help the development of guidance documents for enhanced approval for emerging technologies related to oncology imaging; or

- NIST, exploring the development of physical imaging standards for cancer and other diseases, and
- CMS, with a potential interest in encouraging improved standards for both physical and clinical evaluation of emerging biomedical imaging technologies to ensure they meet their proposed clinical objectives.

A MODEL FOR OPEN SCIENCE AND BIG DATA

The QIN is a very timely effort to promote "open science and public data access" within the oncology imaging community. This is an important development given the advances made in open science and open data developed by other cancer research domains such as genomics and proteomics. QIN addresses the importance of the quality of data collected and the need for consensus based best practices approach to data analysis and clinical decision support systems. The NIH and NCI are now exploring the BIG DATA initiatives, including cloud computing, for data collected across all research domains within NIH IC's. Thus, QIN may prove to be a timely initiative to position the oncology imaging community to participate more effectively in NIH BIG DATA initiatives in the future, where imaging results need to be integrated with results of other research domains.



Robert Nordstrom, QIN Director

TABLE OF CONTENTS

Section I

Technical Reports from Individual QIN Team Members

U01 CA143062: Radiomics of NSCLC

H. Lee Moffitt Cancer Center

Robert J. Gillies, Ph.D. and Matthew B. Schabath, Ph.D.
Moffitt Cancer Center
and
Yoganand Balagurunathan, Ph.D. Moffitt Cancer Center
Dmitry Goldgof, Ph.D. Univ. S. Florida
Andre Dekker, Ph.D. Univ. Maastricht
Zhaoxiang Ye, M.D. Tianjin Medical Hospital

INTRODUCTION

"Radiomics" is the process of converting radiographic images into mineable data. The overarching hypothesis of Radiomics is that image features describing size, shape and texture, reflect the underlying tumor pathophysiology and hence, can be developed and qualified as biomarkers for prediction, prognostication or response monitoring. Radiomics is designed to use standard-of-care images, allowing the development and curation of large data sets that are needed for statistical power. In the first cycle of the award we addressed challenges to all steps in the "radiomic pipeline", *viz.* (1) defining the impact of acquisition and image reconstruction on the quality of radiomics data; (2) curating to maintain high data quality; (3) qualifying semi-automated segmentation tools; (4) statistical qualification of radiomic features; (5) developing database sharing tools to allow rapid hypothesis testing; and (6) developing and applying informatics approaches to these datasets. With this pipeline, we have identified specific features from CT (and PET) images that accurately predicted overall survival (OS) in lung cancer patients treated with surgery or chemo-radiation. Importantly, these features have been validated in completely independent data sets.

In the current award, we will build on this prior work to incorporate radiomics into a multi-parametric "Risk-of-Recurrence" score for patients with resectable non-small cell lung cancer (NSCLC). NSCLC is the leading cause of cancer deaths worldwide and hence, even incremental improvements in decision support can have a profound impact on patients' lives. We will use and extend the radiomics framework that we have developed to address a compelling and focused question in lung cancer care: whether to treat post-surgery (stages I-IIIa) patients with adjuvant chemotherapy. Virtually all NSCLC surgical candidates receive high-quality diagnostic CTs. Early stage NSCLC patients are commonly resected with lobectomy and complete mediastinal lymph node removal. Of these, up to 35% will experience distant recurrence within 5 years. Recurrence can be reduced with adjuvant therapy (AT), yet the decision whether or not to use AT is not trivial, as it is associated with significant morbidities and even mortality. This decision is currently ill-informed by stage or gene expression patterns alone. Predictive models are lacking that can accurately identify which patients have the highest likelihood of recurrence, thus requiring most aggressive adjuvant follow-up. Our proposal addresses this important clinical problem, by using radiomics

pipeline to develop a Risk-of-Recurrence score for pre-operative patients using radiomic, clinical and genomic data.

PROGRESS OVER THE PAST YEAR

§ Aim 1. Develop a Risk-of-Recurrence Score.

(1.1) Assemble a two-institution cohort into a radio(geno)mic database.

Moffitt:

We assembled a retrospective cohort of surgically resected lung cancer patients with the following inclusion criteria: had a pre-surgical CT at Moffitt within three months prior to surgery, stage IA to IIIA disease, and a primary non-small cell lung cancer diagnosis that was surgically resected at Moffitt between Jan 2008 and December 2016. As such, we curated a dataset of 800 patients that met the aforementioned criteria. The pre-surgical CTs were identified, downloaded and databased, and the lung lesions are currently being segmented by Moffitt's IRAT (Image Response Assessment Team) using the HealthMyne PACs. Patient demographic information and clinical data from the Cancer Registry were obtained, curated, and QA/QCed. Additionally, manually abstraction was initiated and completed to obtain progression and recurrence (date and location), tumor mutation data (e.g., KRAS, EGFR, BRAF, etc), and IHC protein biomarkers, where available. Among the 1,093 patients, 285 also had PET images which were downloaded and curated for future PET-CT analyses. The data and images from a test set of 10 patients were uploaded to the Translation Research IT (TraIT) infrastructure in the Netherlands (see Aim 2.1 below).

Tianjin:

We retrospectively reviewed all the lung cancer cases in the PACS system from Jan 2013 to Dec 2016. According to the criteria in our study, 500 cases were selected finally. All the CT images have been stored and shared. The clinical information is uploading via Open Clinica. Another 100 new cases were prospectively collected. The CT images have been stored separately. The clinical characteristics, pathological results, and treatment information were recorded one by one.

(1.2) Analyze the acquired radiomic-clinical-genomic data in a Bayesian framework to develop a "risk-of-recurrence" score for individual patients to support a decision whether to treat with AT. **This has not yet started as we have not yet completed 1.1**

(1.3) Investigate the influence of acquisition conditions (kVp, recon kernel, slice thickness, FOV, contrast). **This has not yet started as we have not yet completed 1.1**

§ Aim 2. Improve Radiomics Data Sharing through the NCI QIN. Image sharing: share specific high quality, curated image sets for validation and share image sets from aim 1 as open data.

(2.1) Develop platform to house images and CDEs.

To support the work in Aim 1, a platform was developed which can host all image sets and clinical data elements. For the image upload and archive a combination of open source tools (local CTP clients for de-identification and a cloud based instance of NBIA for storage, Figure 1), similar to the setup used by The Cancer Imaging Archive, is hosted by the Translation Research IT (TraIT) infrastructure in the Netherlands.

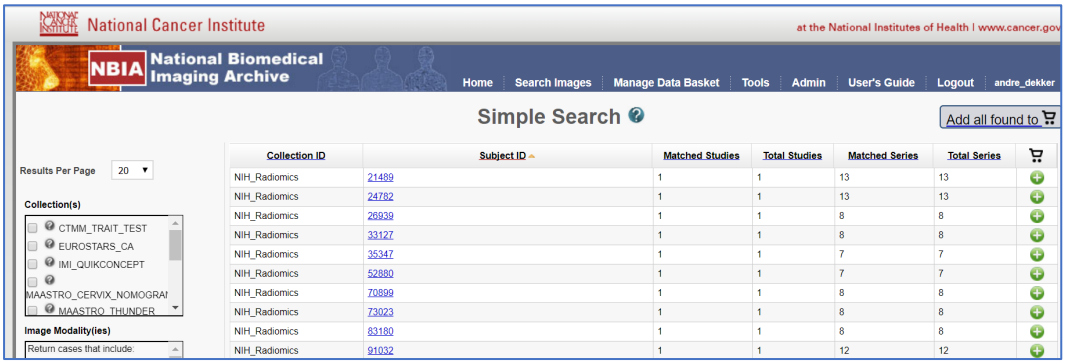


Figure 1: Screen capture of the NBIA image archive to support Aim 1

For the clinical data elements, the open source tool OpenClinica is used – again cloud-based and hosted by TraIT. An eCRF designed for this project is given in the below figure. The eCRFs underwent multiple iterations and testing phases, with a final version accepted in Q3 2017 (Figure 2).

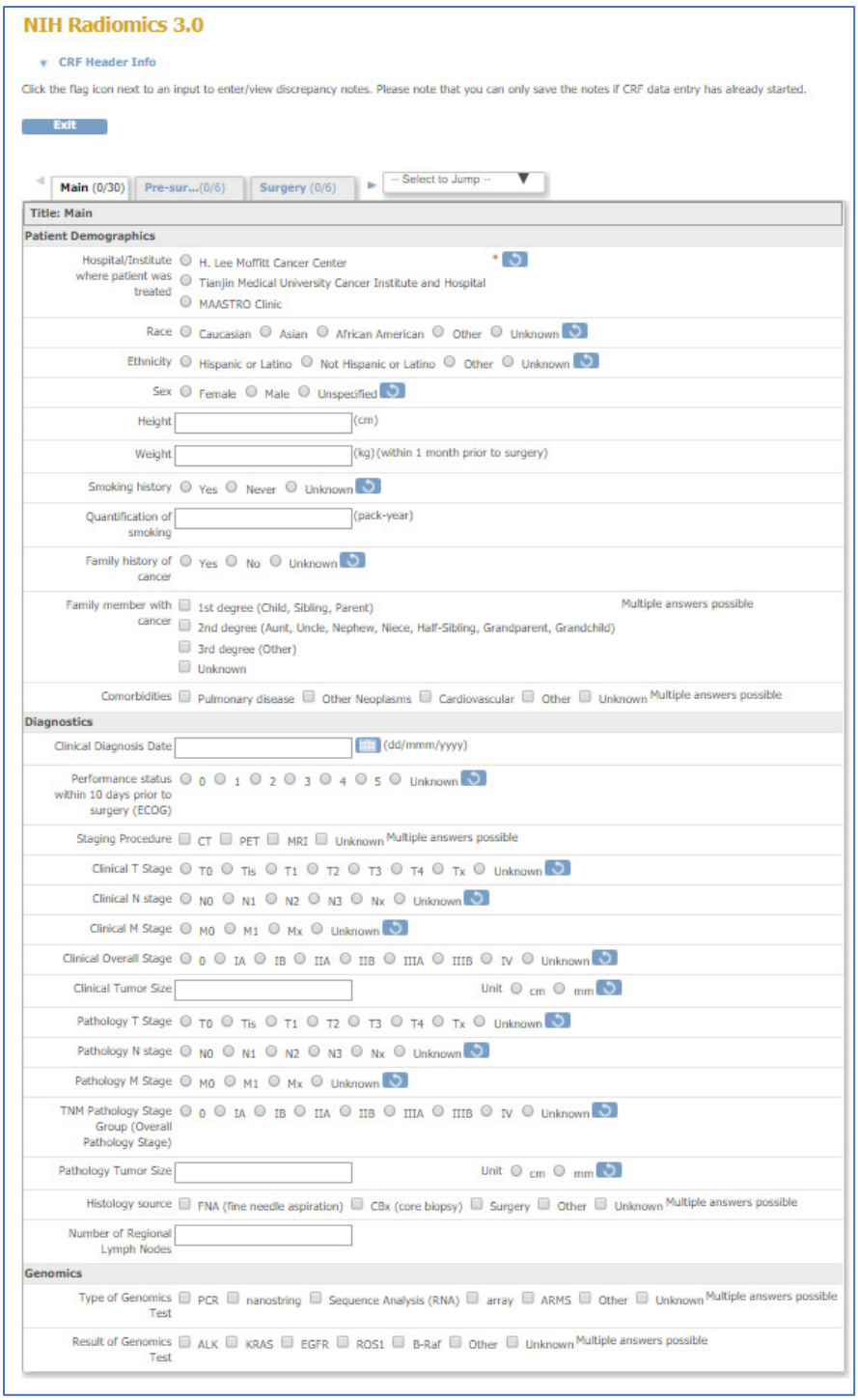


Figure 2: Screen capture of one of the CDE forms designed in OpenClinica to support Aim 1. After extensive testing the first 10 patients are now live with 500 patients ready to be de-identified and uploaded (Figure 3).

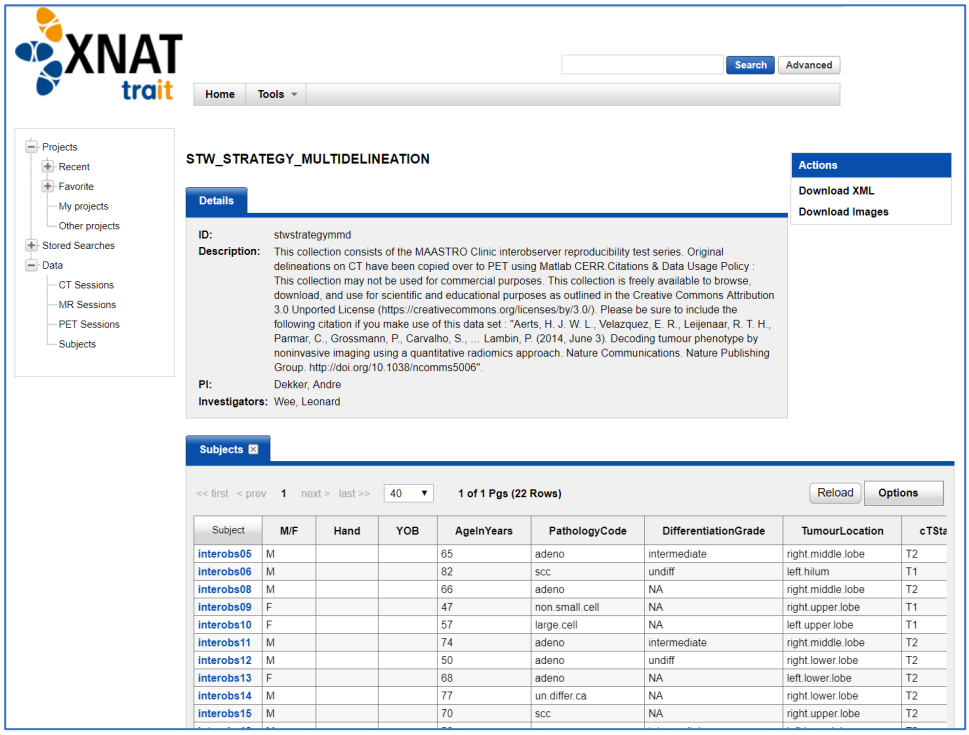


Figure 3: Screen capture of Trait Dbase showing de-identified patient data.

(2.2) Share high quality multiple segmentation sets for validation and provide tools to the QIN to compare segmentations.

CT-PET image sets and clinical data elements from 22 NSCLC patients segmented by 5 human observers were curated and published open-access in 2017 (https://tinyurl.com/RadiomicsMultipleDelineation) for validation use and comparison of various segmentation tools. These are the data sets which were used in same Radiomics publication mentioned above (http://doi.org/10.1038/ncomms5006).

An important contribution was made to the AAPM (American Association of Physicist in Medicine) Task Group No. 263 - Standardizing Nomenclature for Radiation Therapy. This task group has standardized over 700 segmentation names used in radiation oncology and importantly link these the the Foundational Model of Anatomy ontology and thereby to the Radlex ontology which is the most common ontology in Radiology. The AAPM task group report will be published in 2018. When implemented, the recommendations if this task group will be very instrumental in comparing segmentations.
Finally a thoracic segmentation grand challenge was co-organized with Jayashree Kalpathy-Kramer and others at AAPM 2017:
(https://www.aapm.org/meetings/2017AM/PRAbs.asp?mid=127&aid=35318).

The aim of the challenge was to compare various (semi-)automatic segmentation

methods. The data used in the challenge consisted of images and segmentations of 60 thoracic cancer patients (incl. contributions from this project) were made publically available at TCIA (http://doi.org/10.7937/K9/TCIA.2017.3r3fvz08) after the challenge (see below: Network Collaborations).

(2.3) Define ontology for Radiomics features, provide tools to publish Radiomics features independent of the specific Radiomics application and provide tools to compare Radiomics features as a first step to harmonization.

The first version of the Radiomics Ontology (RO) was published to the NCBO Bioportal on May 11, 2017. It has been visited over 700 times in 2017. A journal manuscript detailing the development and initial testing of the RO is in preparation. A further update of the RO is currently in progress, in collaboration with LaTIM in Brest, France to align the RO with the international biomarker standardization initiative (IBSI - https://arxiv.org/abs/1612.07003). Based on Semantic Web technology, the Radiomics Ontology allows image features to be published independent on the specific Radiomics implementation. This is an important step towards the ultimate aim of FAIR (Findable, Accessible, Interoperable, Reusable: https://www.nature.com/articles/sdata201618) imaging datasets.

The open source pyRadiomics was extended with open source components (https://github.com/zhenweishi/Py-rex). These extra components allow pyRadiomics to handle DICOM RTSTRUCT segmentations, apply the standardized nomenclature described in Aim 2.2 and subsequently export the Radiomics features using Semantic Web technology (RDF) using the concepts defined in the Radiomics Ontology. This extended pyRadiomics serves as a reference implementation of the Radiomics Ontology. Ongoing work is to implement the Radiomics Ontology for other Radiomics implementations (e.g. IBEX, Philips proprietary, OncoRadiomics proprietary).

Finally, a systematic literature review (*submitted*) was performed to establish which factors influence the repeatability and reproducibility of radiomics features. This review serves as an important driver of the concept to be included in the Radiomics Ontology so that radiomics features can be compared and ultimately harmonized. A synthesis is given in Figure 4 (Figure 2 in MS) with details on segmentation, image reconstruction, acquisition and pre-processing shown to be important to capture in Radiomics studies.

	FIRST ORDER	SHAPE METRICS	TEXTURE ANALYSIS	COMMENTS
ROI SEGMENTATION				
MANUAL DELINEATION	♦	♦♦♦	♦♦♦	Mainly PET studies and one multi-centre CT study. Shape metrics from PET may be less subject to inter-observer differences. Semi-automated methods generally improves reproducibility.
SEMI-AUTO / AUTO	♦	♦♦	♦♦	
IMAGE RECONSTRUCTION				
RECONSTRUCTION FILTER	♦	♦♦	♦♦♦	Consistent in a few CT and PET studies of NSCLC.
VOXEL SAMPLING	♦♦	♦♦	♦♦♦	
IMAGE ACQUISITION SETTINGS				
RESPIRATORY MOTION	♦♦	♦♦	♦♦	Consistent over single-institution PET and CBCT studies of NSCLC.
SCATTERED RADIATION	♦♦	?	♦♦	In one CBCT study of NSCLC, but did not evaluate shape metrics.
CT SCANNER	♦♦	♦♦	♦♦	In one multi-institutional CT study in NSCLC, effects were similar in magnitude to inter-patient differences.
DIGITAL IMAGE PRE-PROCESSING				
NOISE AND SMOOTHING	♦♦	?	♦♦	Single-centre CBCT and planning CT study in H&N; smoothing and noise has less effect than high-pass and logarithmic filters.
INTENSITY DISCRETIZATION	♦♦	♦♦	♦♦	Consistent in H&N studies of perfusion CT and PET, bin size may have less impact in PET.
CONSENSUS ABOUT MOST STABLE OR LEAST STABLE RADIOMIC FEATURES	Entropy was consistently among the most repeatable/reproducible first-order features. There were inconsistent findings for skewness and kurtosis. Certain shape metrics may be reproducible in PET, and slightly less reproducible in CT, though it is unclear which individual features prove to be stable. No emergent pattern nor consensus for highly reproducible textural features. Coarseness and contrast were among the least reproducible.			

Figure 2. Qualitative synthesis of radiomic feature classes, indicating processing steps that are either highly likely (●●●), probable (●●) or less likely (●) to exert an adverse effect on repeatability and reproducibility, for each class of radiomic feature. Feature classes for which no information was available is marked as unknown (?).

Figure 4: Synthesis of Radiomic feature classes.

(2.4) Host Annual workshops.

On October 23-24, 2017, we hosted the 8th annual Radiomics workshop in Clearwater Beach, FL. This workshop was supported by Moffitt Cancer Center and the U01 provided travel support for investigators on this grant. There were 98 attendees (**Figure 5**) from 35 different institutions from 6 different countries (USA, China, Netherlands, Canada, Germany, France). There was significant participation from industrial academic investigators as well (e.g. IBM, GE, HealthMyne, Draper, Oncoradiomics, Pulsar). As in the past, there was no travel support, and all who attended gave talks, if they wanted to. There were three invited talks:

Plenary 1: "Adaptive Therapy" (Bob Gillies for Bob Gatenby)
Plenary 2: "How Pathomics can Compliment Radiomics: (Joel Saltz)
Plenary 3: "Prediction of Glioma Molecular Markers from MRI using Deep Learning" (Bradley Erickson)
and the following sessions:
Session One: LONGITUDINAL MONITORING OF THERAPY RESPONSE (5 speakers)
Session Two: EARLY DETECTION (6 speakers)
Session Three: HABITAT IMAGING (8 speakers)
Session Four: RESPONSE PREDICTION AT BASELINE (15 speakers)
Session Five: DEEP LEARNING (6 speakers)
Session Six: STANDARDIZATION EFFORTS (6 speakers)

Session Seven: CHALLENGES and TOOLBOXES (6 speakers)
Session Eight: PHANTOM STUDIES (3 speakers)

Figure 5: Attendees at the 2017 Radiomics Workshop in Clearwater Beach (taken indoors because of inclement weather)

COLLABORATIONS WITHIN THE QIN

We had an active collaborative research year working with leaders in medical imaging through QIN working groups and outside the network. Our members are active participants in the PET/CT subgroup and frequently attend the Bioinformatics working group. Below is the outline of our participation and research activity for the year.

- Dr. Dmitry Goldgof has taken responsibility as chair of the PET-CT working group for year 2016-17; previously he served as a Co-Chair. He actively stimulates discussion in the monthly conference call and documents discussion. Both monthly agendas and minutes are available on QIN Sharepoint site. He personally takes effort to push group research effort to promote institutional collaboration. Dr. Goldgof is responsible for producing the PET-CT WG annual report. He is also a member of QIN Coordinating Committee and attends monthly meetings. Drs. Schabath and Balagurunathan are both members of the PET-CT working group.

- We continue to contribute in the harmonizing and creating feature ontology. The group effort is led by Dr. McNitt-Gray (UCLA) and Kalpathy-Cramer (MGH) and the effort is titled: "Quantitative Feature Standardization – Creating a Feature Ontology". This is in high-bandwidth communication with the international IBSI effort, described above.

- We played an active role in co-organizing community wide challenge with the NCI and QIN members, which will be conducted under the IEEE's International Symposium on Biomedical Imaging (ISBI 2018, http://biomedicalimaging.org/2018/challenges/). The challenge will involve using medical image data (CT lung) previously provided by the Moffitt team, used by PET-CT subgroup along with the segmentation mask created by the team. The challenge will have multiple phases and will go from January to April 2018.

- We participate in inter-working group Quarterly meetings (IAPM). In the past year, we have presented our feature challenge and interval challenge collaborative research.

- Dr. Matt Schabath has served as an *ad hoc* reviewer for two QIN Grant Review Study sections.

Collaborative Research Projects: We actively work with the QIN member teams and IBSI to participate in collaborative challenges with a goal to develop, validate reproducible imaging biomarkers. In the last year we participated and lead following collaborative research projects, which had been published/ accepted in peer-reviewed journals.

Feature Variability Study Across Sites: Our members participated and provided data and logistic support for the Radiomics feature computation challenge in the last year, which we published in the Special issue of Tomography December 2016. The goal of the study was to compare the Radiomics features generated at different institutional sites with different segmentations.

- The study goal was to investigate sensitivity of radiomic descriptors on pulmonary nodules with different segmentations and different feature types using institutional specific feature implementations. We calculated the concordance correlation coefficients of the features as a measure of their stability with the underlying segmentation; 68% of the 830 features had concordance of over 75%. We then found groups of features using graph tree method (cutoff of 0.75) there were 75 subgroups of features and it increased to 246 groups with a larger cutoff (0.95). The study illustrates the diversity in the types of quantitative features extracted by different groups on the same set of patients. Some categories of features show lower variability but few others show broad differences. Few key details of this research projects are briefly described below:

- We used 41 CT images (52 lung nodules) from 5 different collections of images with feature extraction from different institutions. The collections were: Phantom images & RIDER images from Columbia University, Lung Image Database Collection (LIDC) from Stanford University, and Images from TCIA (with additional IRB requirement waived).

- We had seven participating group that included, Moffitt Cancer Center, Stanford University, Columbia University, University of California at Los Angeles, University of Iowa, Princess Margaret Cancer Center, and the University of Michigan. Each team computed features on the nodule segmentations (which was about 468) and uploaded features extracted with their own pipeline.

- Extracted features include: size, intensity, global shape, local shape descriptors, margin and texture features: gray level co-occurrence matrices, shape descriptors (LSD: local shape descriptor and GSD: global shape descriptor), margin, followed by Texture features across sites. Size based features shows lowest variability across sites, followed by shape descriptors (LSD and GSD), margin. Figure 6 shows different ways to visualize the intergroup variability.

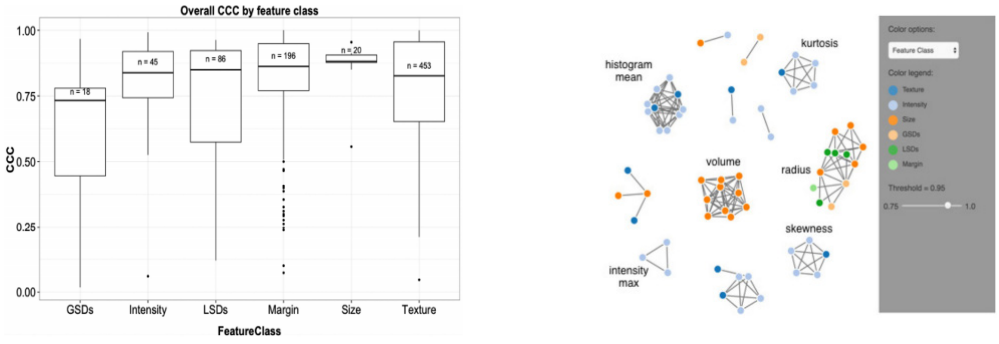


Figure 6: Concordance correlation of features category across sites.

This study illustrates the spread of quantitative features extracted across sites. The variability could arise from implementational difference, different formulation followed at the sites for categories that warrants further investigation.

Semi-automated Pulmonary Nodule Interval Segmentation Challenge:

Lung nodule size estimation is an important measurement that can trigger an array of clinical treatments. Nodules are considered positive for follow-up if they measure 6 to 8mm in recent guidelines (LungRads and NCCN). In most cases, a trained Radiologist, who may use a suite of semi-automated tools, bases the decision on measured size. Recently, change in size and volume has shown to be related to aggressiveness of the disease. Since size and change in size/volume are important prognostic factors for the oncologist, we investigated the measurement variability across participating centers through the group challenge.

Study Goal: We proposed to study the variability of size measurements in the screening intervals, across teams. We believe any bias that may exist in each teams size measurements (obtained via segmentations) would offset by repeating the segmentation at a later time point, for the same nodule. In this study, we obtained 100 patient image datasets from the National Lung Screening Trial (NLST) that had a nodule detected on each of two consecutive low dose computed tomography (LDCT) scans, with an equal proportion of malignant and benign cases (50 malignant, 50 benign). The teams were asked to provide segmentation masks for each nodule at both time points. From these masks, the volume was estimated for the nodule at each time point; the change in volume (absolute and percent change) across time points was estimated as well (Figure 7).

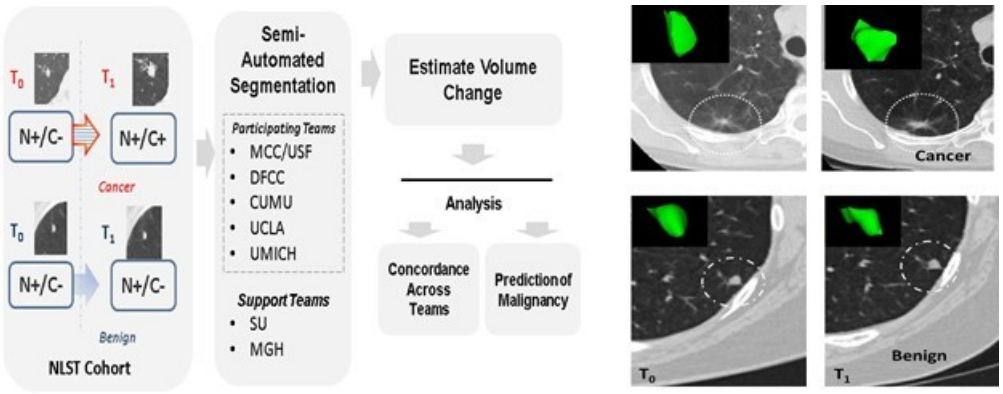


Figure 7: Work flow followed in the challenge and sample nodules detected at different time points.

We used the concordance correlation coefficient (CCC) to compare the similarity of computed nodule volumes (absolute and percent change) across algorithms. We used Logistic regression model on the change in volume (absolute change and percent change) of the nodules to predict the malignancy status, the area under the receiver operating characteristic curve (AUROC) and confidence intervals were reported. Because the size of nodules was expected to have a substantial effect on segmentation variability, analysis of change in volumes was stratified by lesion size, where lesions were grouped into those with a longest diameter of <8mm and those with longest diameter ≥ 8mm.

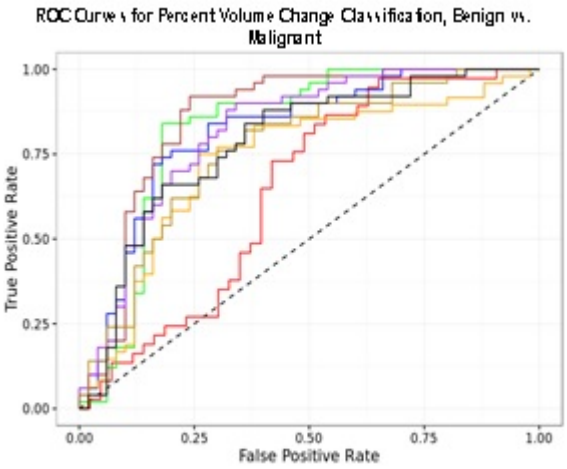


Figure 8. ROC curve for predicting cancer status using change in volume estimates across different teams.

We observed that segmentation of the nodules showed substantial variability across algorithms, with the CCC ranging from 0.56 to 0.95 for change in volume (percent change in volume range was [0.15 to 0.86]) across the nodules. When examining nodules based on their longest diameter, we find the CCC had higher values for large nodules with a range of [0.54 to 0.93] among the algorithms, while percent change in volume was [0.3 to 0.95]. The AUROC generated from change in volume ranged from 0.65 to 0.89 (Percent change in volume was 0.64 to 0.86) for entire nodule range (Figure 8). Prediction improves for large nodule range (≥ 8mm) with AUC range 0.75 to 0.90 (percent change in volume was 0.74 to 0.92).

When considering the entire nodule size range, Team 1 and 2 showed statistically comparable AUCs with overlapping confidence range [0.78, 0,86] and [0.73, 0.83]. Team 3A's (corrected) AUC was superior to any other teams with a confidence limits of [0.86, 0.92], while their uncorrected AUC showed slightly lower performance that was comparable with other teams [0.82, 0.90]. Team 4's corrected estimates' AUC was in the range of [0.79, 0.87], while their uncorrected AUC showed lower average AUC with a confidence range of [0.59, 0.71]. Team 5's average AUC was in the middle compared to others uncorrected estimates, with confidence limits of [0.72, 0.83]. It is interesting to note that, most semi-automated AUC's showed slightly superior performance compared to radiologist delineated contours, whose average AUC was 0.78 with a confidence range of [0.73, 0.83]. When nodule sizes were restricted to smaller size (<8mm), Team 1, 3 and 5's predictor AUC confidence ranges are comparable. Teams 2 and 4 AUC performances were lower compared to other teams.

Lessons Learned from the challenge:

- It is important to avoid biases in size and volume measurements. Some common biases include use of any clinical diagnostics and or radiological observational intuition prior to delineate the region of interest.
- Some known variations are attributed to the segmentation algorithms and the imaging suites methodologies, which show differences
- There are few others variability sources caused by scanner parameters and reconstruction methods which influence the image intensity.

(4.0) Targets of Opportunity

To expand our presence in the radiomics community, we are continually and actively pursuing new and orthogonal research directions. Over the last year, two specific studies arose as targets of opportunity that align with our current efforts to identify radiomic signatures that predict lung cancer outcomes. Both studies were spearheaded by a promising doctoral student (Mr. Ilke Tunali) that are described below.

In the first study that was published in Oncotarget (Tunali *et al.* at https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/29221183/), a set of image features extracted from radial gradient (RG) and radial deviation (RD) maps generated from thoracic CT images. It's worth noting that these are features were not part of our radiomics pipeline and have since been included in our toolbox. In this study, we assessed whether these RD/RG features were associated with clinical outcomes in patients with lung adenocarcinoma in a training and test cohort. Following feature reduction and model building analyses, we identified two combinatorial features discriminates indolent lung cancers vs. aggressive lung cancer. Additionally, we explored the potential biological underpinnings of these features by analyzing the association between radial gradient and radial deviation image features with semantic radiological features. We found that three RD/RG radiomic features that were statistically significantly associated with three semantic features: lobulation, pleural attachment, and border definition.

In the second study, we curated a dataset of lung cancer patient treated with immunotherapy to develop clinical and radiomic models to predict treatment response. The manuscript from these efforts is in preparation and expected to be submitted for review to Cancer Discovery in February 2018. Among 235 NSCLC patients treated with single agent or double agent immunotherapies, we identified parsimonious models containing highly informative clinical covariates and radiomic image features with modest to high ability to predict rapid disease progression with AUROCs ranging from 0.821 to 0.865. These models have potential important translational implications to identify highly vulnerable patients that experience disease progression, rapid tumor growth, and poor patient outcomes.

Finally, based on the aforementioned immunotherapy work, we have a review manuscript in preparation to synthesize the current knowledge of *Clinical, Genomic, and Radiomic Predictors of Lung Cancer Immunotherapy Response.*

PLANS FOR NEXT YEAR

We will complete curation of the surgical cases with CDEs by the end of 2018. We will perform another query of surgical cases from Moffitt to include patients from 2017 through 2018. The total accrual goal is to exceed 2,200 patients deposited into TRAIT. With these in hand, data mining will actively begin by the end of Q4 YR03.

- Finish development automated upload tools for clinical data elements (Q2 2018)
- Upload the current 1,300 image and clinical data elements for Aim 1 to the central infrastructure (Q2 2018)
- Finish curation of the datasets (Q4 2018)
- Complete segmentation of lung lesions, extract radiomic features, and QA/QC radiomic datasets Q4 2018)
- Analyze demographic and clinical descriptors of our data set, and prepare first publication.

We will also continue to collaborate within the QIN

- We will continue to work closely with the QIN working group teams to harmonize the quantitative features across centers and work with the group to formulate ontology for feature definition.
- We will validate newly discovered biomarkers by our center through with other QIN member teams.
- We will work with the QIN and NCI team members to disseminate knowledge and participate in wider community challenges through leading conferences, such as the ISBI 2018.
- We will continue to play an active to catalytic role to promote the use of quantitative imaging biomarkers for oncology research through the QIN working groups.
- Develop Radiomics Ontology implementations for additional Radiomics software implementation (IBEX, Philips, OncoRadiomics)

Publications Planned

- APPM grand challenge
- AAOM TG 263 report
- Update on the Radiomics Ontology
- Review on Repeatability and reproducibility of radiomics features
- Review on Radiomic Predictors of Immune Therapy

LIST OF REFERENCES

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, Gillies R, Goldgof D. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. Tomography : a journal for imaging research. 2016;2(4):430-7. doi: 10.18383/j.tom.2016.00235. PubMed PMID: PMC5279995.

Tunali I, Stringfield O, Guvenis A, Wang H, Liu Y, Balagurunathan Y, Lambin P, Gillies RJ, Schabath MB. Radial gradient and radial deviation radiomic features from pre-surgical CT scans are associated with survival among lung adenocarcinoma patients. Oncotarget. 2017;8(56):96013-96026. doi: 10.18632/oncotarget.21629. PubMed PMID: 29221183; PubMed Central PMCID: PMCPMC5707077

PUBLICATIONS FROM QIN EFFORTS (2017)

Following is a list of Radiomics-related publications from investigators affiliated and supported by our QIN grant. They are grouped according to those Directly related to the aims of our current grant, Collaborative Studies within the QIN, Phantom Studies, Lung Nodules, and other cancers. These are included as the methods they develop have impacted the U01.

NSCLC (U01 directly associated)

Yip SSF, Liu Y, Parmar C, Li Q, Liu S, Qu F, Ye Z, Gillies RJ, Aerts H. Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer. Scientific reports. 2017;7(1):3519. doi: 10.1038/s41598-017-02425-5. PubMed PMID: 28615677; PubMed Central PMCID: PMCPMC5471260.

Rios-Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, Ye Z, Makrigiorgos GM, Fennessy FMM, Mak RH, Gillies RJ, Quackenbush J, Aerts H. Somatic mutations drive distinct imaging phenotypes in lung cancer. Cancer research. 2017. doi: 10.1158/0008-5472.CAN-17-0122. PubMed PMID: 28566328.

Li Q, Kim J, Balagurunathan Y, Liu Y, Latifi K, Stringfield O, Garcia A, Moros EG, Dilling TJ, Schabath MB, Ye Z, Gillies RJ. Imaging features from pretreatment CT scans are associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy. Med Phys. 2017. doi: 10.1002/mp.12309. PubMed PMID: 28464316.

Tunali I, Stringfield O, Guvenis A, Wang H, Liu Y, Balagurunathan Y, Lambin P, Gillies RJ, Schabath MB. Radial gradient and radial deviation radiomic features from pre-surgical CT scans are associated with survival among lung adenocarcinoma patients. Oncotarget. 2017;8(56):96013-96026. doi: 10.18632/oncotarget.21629. PubMed PMID: 29221183; PubMed Central PMCID: PMCPMC5707077.

Li Q, Kim J, Balagurunathan Y, Qi J, Liu Y, Latifi K, Moros EG, Schabath MB, Ye Z, Gillies RJ, Dilling TJ. CT imaging features associated with recurrence in non-small cell lung cancer patients after stereotactic body radiotherapy. Radiat Oncol. 2017;12(1):158. doi: 10.1186/s13014-017-0892-y. PubMed PMID: 28946909; PubMed Central PMCID: PMCPMC5613447.

Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, Leijenaar RT, Haibe-Kains B, Lambin P, Gillies RJ, Aerts HJ. Defining the biological basis of radiomic phenotypes in lung cancer. Elife. 2017;6. doi: 10.7554/eLife.23421. PubMed PMID: 28731408; PubMed Central PMCID: PMCPMC5590809.

QIN collaborative studies

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, Gillies R, Goldgof D. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. Tomography : a journal for imaging research. 2016;2(4):430-7. doi: 10.18383/j.tom.2016.00235. PubMed PMID: PMC5279995.

Y. Balagurunathan, A. Beers, J. K.Cramer, M.McNitt-Gray, L.Hadjiiski, B.Zhao, J. Zhu, H. Yang, S.S.F. Yip, H.J.W.L. Aerts, S. Napel, D. Cherezov, K.Cha, H. Chan, C. Flores, A.Garcia, R.Gillies, D.Goldgof. Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST data, Medical Physics (acccepted) Jan 2018.

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. Med Phys. 2017 Feb;44(2):479-496. doi: 10.1002/mp.12041.

Phantoms and General (direct)

Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, Abdalah MA, Schabath MB, Goldgof DG, Mackin D, Court LE, Gillies RJ, Moros EG. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys. 2017;44(3):1050-1062. doi: 10.1002/mp.12123. PubMed PMID: 28112418; PubMed Central PMCID: PMCPMC5462462.

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. Nat Rev Clin Oncol. 2017;14(3):169-186. doi: 10.1038/nrclinonc.2016.162. PubMed PMID: 27725679; PubMed Central PMCID: PMCPMC5378302.

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. Med Phys. 2017;44(2):479-496. doi: 10.1002/mp.12041. PubMed PMID: 28205306.

Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu s, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. Nature Reviews Clinical Oncology 14, 749–762 (2017) PMID: 28975929

Lung Nodules (indirect)

Liu Y, Wang H, Li Q, McGettigan MJ, Balagurunathan Y, Garcia A, Thompson ZJ, Heine JJ, Ye Z, Gillies RJ, Schabath MB. Radiologic Features of Small Pulmonary Nodules and Lung Cancer Risk in the National Lung Screening Trial: A Nested Case-Control Study. Radiology. 2017;(in press).

Liu Y, Balagurunathan Y, Atwater T, Antic S, Li Q, Walker RC, Smith GT, Massion PP, Schabath MB, Gillies RJ. Radiological Image Traits Predictive of Cancer Status in Pulmonary Nodules. Clinical cancer research : an official journal of the American Association for Cancer Research. 2017;23(6):1442-1449. doi: 10.1158/1078-0432.CCR-15-3102. PubMed PMID: 27663588.

Li Q, Balagurunathan Y, Liu Y, Qi J, Schabath MB, Ye Z, Gillies RJ. Comparison Between Radiological Semantic Features and Lung-RADS in Predicting Malignancy of Screen-Detected Lung Nodules in the National Lung Screening Trial. Clinical lung cancer. 2017. doi: 10.1016/j.cllc.2017.10.002. PubMed PMID: 29137847.

Other Cancers (indirect)

Zhou M, Chaudhury B, Hall LO, Goldgof DB, Gillies RJ, Gatenby RA. Identifying spatial imaging biomarkers of glioblastoma multiforme for survival group prediction. Journal of magnetic resonance imaging : JMRI. 2017;46(1):115-123. doi: 10.1002/jmri.25497. PubMed PMID: 27678245.

Permuth JB, Choi JW, Chen DT, Jiang K, DeNicola G, Li JN, Coppola D, Centeno BA, Magliocco A, Balagurunathan Y, Merchant N, Trevino JG, Jeong D. A pilot study of radiologic measures of abdominal adiposity: weighty contributors to early pancreatic carcinogenesis worth evaluating? Cancer Biol Med. 2017;14(1):66-73. doi: 10.20892/j.issn.2095-3941.2017.0006. PubMed PMID: 28443205; PubMed Central PMCID: PMCPMC5365183.

Permuth JB, Chen DT, Yoder SJ, Li J, Smith AT, Choi JW, Kim J, Balagurunathan Y, Jiang K, Coppola D, Centeno BA, Klapman J, Hodul P, Karreth FA, Trevino JG, Merchant N, Magliocco A, Malafa MP, Gillies R. Linc-ing Circulating Long Non-coding RNAs to the Diagnosis and Malignant Prediction of Intraductal Papillary Mucinous Neoplasms of the Pancreas. Scientific reports. 2017;7(1):10484. doi: 10.1038/s41598-017-09754-5. PubMed PMID: 28874676; PubMed Central PMCID: PMCPMC5585319.

Caudell JJ, Torres-Roca JF, Gillies RJ, Enderling H, Kim S, Rishi A, Moros EG, Harrison LB. The future of personalised radiotherapy for head and neck cancer. Lancet Oncol. 2017;18(5):e266-e273. doi: 10.1016/S1470-2045(17)30252-8. PubMed PMID: 28456586.

Chang YC, Ackerstaff E, Tschudi Y, Jimenez B, Foltz W, Fisher C, Lilge L, Cho H, Carlin S, Gillies RJ, Balagurunathan Y, Yechieli RL, Subhawong T, Turkbey B, Pollack A, Stoyanova R. Delineation of Tumor Habitats based on Dynamic Contrast Enhanced MRI. Scientific reports. 2017;7(1):9746. doi: 10.1038/s41598-017-09932-5. PubMed PMID: 28851989; PubMed Central PMCID: PMCPMC5575347.

Cusumano D, Dinapoli N, Boldrini L, Chiloiro G, Gatta R, Masciocchi C, Lenkowicz J, Casà C, Damiani A, Azario L, Van Soest J, Dekker A, Lambin P, De Spirito M, Valentini V. Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. Radiol Med. 2017 Dec 11. doi: 10.1007/s11547-017-0838-3. [Epub ahead of print]PMID: 29230678

U01 CA140206: QUANTITATIVE IMAGING TO ASSESS RESPONSE IN CANCER THERAPY TRIALS

UNIVERSITY OF IOWA

John M. Buatti, M.D.
Thomas L. Casavant, Ph.D.
Michael M. Graham, Ph.D., M.D.
Milan Sonka, Ph.D.

INTRODUCTION

The University of Iowa QIN team continues 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."

This completes our 7th year of participation and significant progress building both developed infrastructure and through multi-institutional working group teams as part of QIN. Several new publications highlight this progress. Our group continues to move forward on 4 specific aims in a highly innovative fashion to 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.

PROGRESS OVER THE PAST YEAR

§Aim 1

Over the past year, we have been working to develop new algorithms to compensate for several sources of error in the TCGA HNSC clinical data as well as our pre-processing of such data to enhance feature selection and power. The pre-processing improvements have shown progress in reducing the class imbalance problem within the dataset, thereby correctly identifying more cases with undesirable outcomes. This had a slight negative effect on classifier performance likely due to the reduction of class imbalance within the data. We continue to develop in this area.

Additionally, it became clear that the TCGA HNSC clinical dataset is subject to anticipated fundamental problems associated with multi-institutional clinical data collection; e.g., data inconsistency/sparsity. In a large fraction of cases, this sparsity strongly affects treatment information—features that are important to the goals of our research. To combat this issue, we have experimented with several forms of feature imputation, filling in missing entries using complex modeling software (Multivariate Imputation by Chained Equations 2011)[1].

This imputation yielded improved classifier performance, as well as notable increases in measures of feature importance. Before imputation, the treatment fields in question ranked at 7^{th} and 12^{th} most important among 24 clinical features. After imputation, these features ranked 4^{th} and 5^{th} respectively.

In addition, processes to mine linked tissue samples for whole exome analysis have begun for our internal database for fully curated head and neck cancer cases with metadata and PET/CT image analysis. Working with our head and neck cancer pathologist and Institute for Human Genetics we are moving forward with a planned set of 100 cases with full imaging and genomic data. Finally, we have requested tissue samples and imaging from both the NRG and ECOG-ACRIN cooperative groups. We are hopeful some of this data will become available in the coming year.

§ Aim 2

Progress to apply highly and fully automated quantitative image analysis methods to a cooperative group data set of H&N cancers has been active in that all forms necessary to acquire images and tissue data have been submitted to the NRG and ECOG-ACRIN groups. The access to the ECOG-ACRIN 6685 data has been approved but access will still await completion of the primary data analysis and initial publication of results. We anticipate gaining access to actual images and metadata in the coming year and will continue to monitor progress in the coming year. Access to the NRG 0522 data is pending committee review.

Progress in developing unique new tools for FLT/PET in H&N cancer including assessment of 4D data has made significant progress. Initial data suggests heterogeneity may be an important finding in the FLT HNC subgroup although final analysis and publication of results are pending. Some work in analysis of bone marrow uptake for FLT has also been undertaken. We have also made progress in identifying advanced quantitative imaging biomarkers (QIBs) for HNC treatment outcome prediction based on both FDG as well as FLT PET-CT scans. For this purpose, we have developed a framework for early stage QIB discovery. Advanced features investigated include texture- and shape-based markers. We are currently working on publishing our results to disseminate our findings.

We have updated and further enhanced our publicly released quantitative image analysis tools, which are based on 3D Slicer (Figure: 1). Specifically, we have added DICOM-based Structured Reporting (SR) capabilities, enabling archiving of quantitative analysis results together with acquired image data on PACS systems, which will help increasing transparency/reproducibility and sharing of research results.

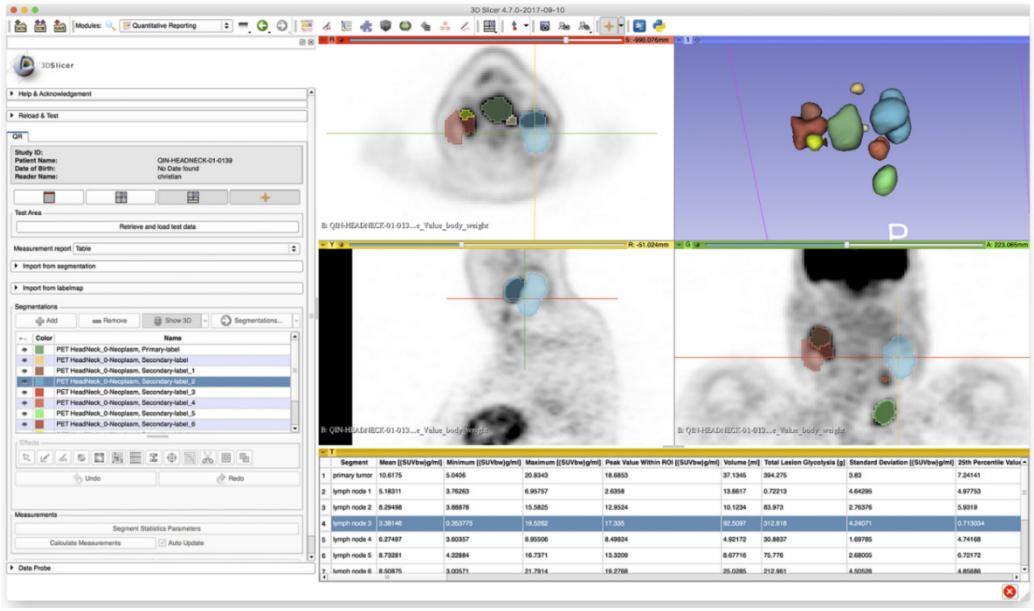


Figure 1: 3D Slicer-based quantitative analysis tool for PET-CT scans, enabling DICOM SR based reporting of QIBs.

Research in liver and liver-tumor segmentation from CT-PET multimodality images is ongoing. We have developed a quantitative approach for determination of liver tumor load in DOTATOC images. This approach consists of two main steps: 1) liver segmentation primarily using volumetric CT images, 2) tumor segmentation primarily using PET images within the liver volume, and 3) quantification of tumor load.

Liver segmentation based on an Alpha-Path-Moves strategy introduced in a general form as part of 3D and 4D volumetric image segmentation methodology (NIH R01 EB004640, Sonka PI) and further developed and modified for liver segmentation in CT data. Tumor segmentation uses graph-cut approach considering global SUV and 3D spatial context as well as the liver volume segmentation. Quantitative analysis of tumor load provides mean liver and tumor SUV information and volumetrically quantifies percent liver subjected to tumor presence – see Figs. 2-4 below.

In a preliminary fashion, 16 liver PET/CT datasets were analyzed providing good-quality segmentation of liver and tumor volumes Quantitative validation on a larger dataset is ongoing.

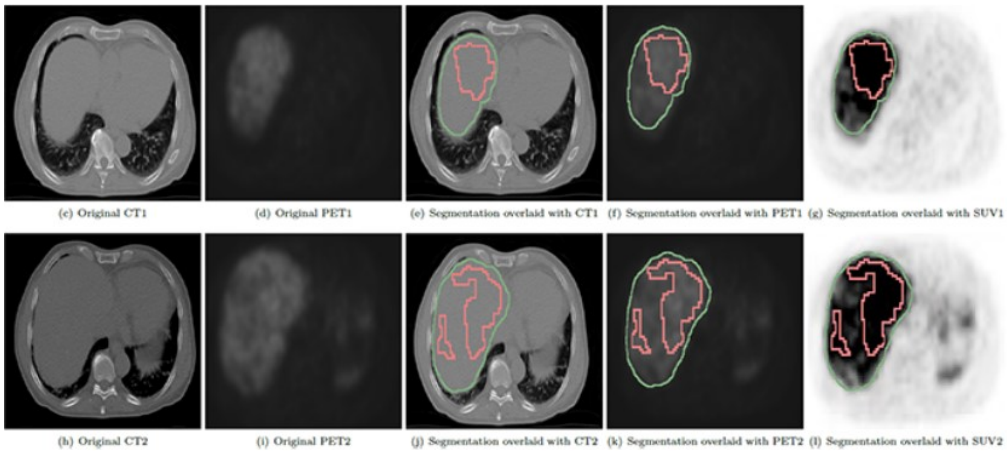


Figure 2: Liver and liver tumor segmentation. Left to right: Original CT, original PET, liver/tumor segmentation overlaid on CT, liver/tumor segmentation overlaid on PET, liver/tumor segmentation overlaid on PET after PET image enhancement for visibility. Two slices of a 3D volume shown.

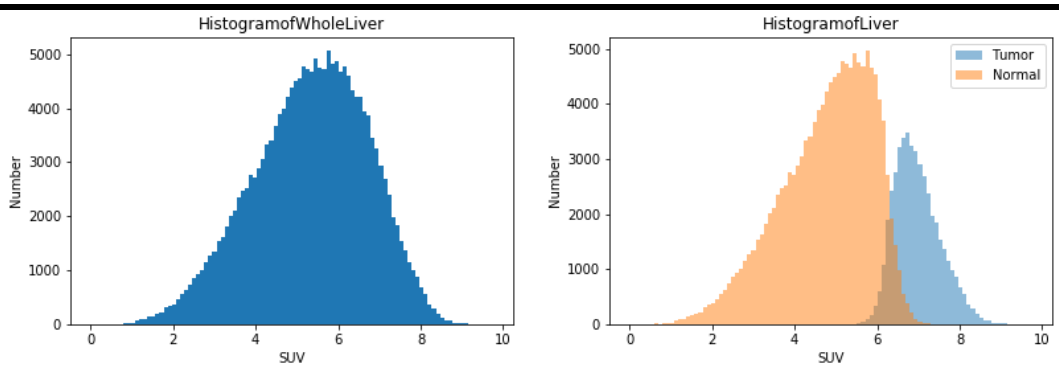


Figure 3: Quantitative analysis of liver tumor load in patient from Fig. 1. (a) Global histogram of PET SUV in the liver. (b) SUV histogram of normal liver and liver tumor tissue. Normal liver SUV: 4:77±1.11, tumor SUV: 6:97±0:56. Tumor-to-liver

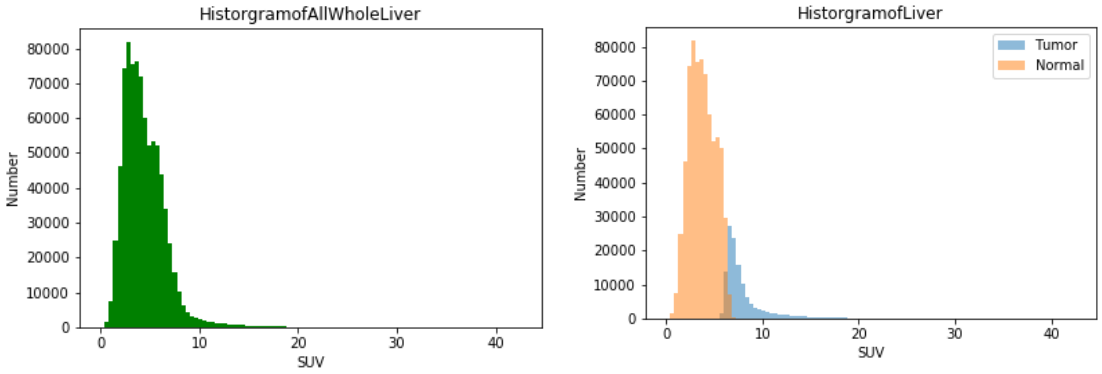


Figure 4: Quantitative analysis of liver tumor load in a pilot cohort of 16 patients. (a) Global histogram of PET SUV in the cohort. (b) SUV histogram of normal liver and liver tumor tissue. Normal liver SUV: 3.78±1.35, tumor SUV: 8.47±3.54. Average tumor-to-liver volume ratio: 15.5%.

§ Aim 3

The developed fully-automated PET phantom analysis software for PET scanner quality control was evaluated. A paper describing our approach as well as evaluation results was published in Medical Physics [2]. It was highlighted under the Editors' Choice column for the Medical Physics citation and medphys.org websites for the January 2018 issue. This was the third time that a Medical Physics paper stemming from our QIN project was highlighted as Editors' Choice ([2-4]; see below). In addition, we are actively working with SNMMI on disseminating our work by development and implementation of a Joint-Commission compliant phantom program for PET/CT with fully-automated cloud-based analysis enabled by our analysis algorithm.

In 2015 The Joint Commission published new "Diagnostic Imaging Requirements" that mandated phantom-based PET/CT scanner performance evaluations by diagnostic medical physicists. Image uniformity, high contrast resolution/system spatial resolution, low contrast resolution or detectability, and artifact evaluation are the newly required components. No guidance is given as to which phantoms or what criteria are to be used in these assessments. Since most PET/CT scanners are hospital-based in the US, and 82% of hospitals are currently Joint Commission accredited, these new required evaluations impact the majority of PET/CT scanners in the US. We developed a phantom-based PET/CT scanner evaluation program designed to be compliant with these Joint Commission requirements while providing meaningful and actionable scanner performance information.
The proposed phantom program requires sites to perform two phantom scans. The first scan images a standard uniform 20 cm diameter cylinder phantom filled with aqueous [^{18}F]FDG. The cylinder is centered in the gantry, but tilted at a slightly oblique angle with respect to the *z*-axis and is imaged for two bed positions at 20-30 minutes per bed position. Images reconstructed using the site's standard clinical reconstruction yields data for in-plane uniformity, axial uniformity, assessment of quantitative calibration, and reconstructed

resolution in the radial and axial directions. Spatial resolution is calculated using measurement of the edge response function using the method of Lodge and Leal[5]. The second phantom scan uses *either* the NEMA Image Quality (IQ) phantom, or the new version of the SNMMI Clinical Trials Network (CTN) oncology phantom with NEMA sized spheres (Figure 5). A 4:1 sphere-to-background ratio is used in either case. The phantom data are acquired using the site's standard oncology protocol including time per bed position and reconstruction parameters (advanced reconstructions allowed). Image data yields a contrast recovery coefficient curve, clinically relevant noise assessment, and assessment of low-contrast lesion detectability.

The two phantom image sets are uploaded to a cloud-based server along with phantom-fill documentation. After a manual image quality control check, the datasets sets are exported to a folder that is continuously interrogated (Python). Upon dataset detection, images are automatically downloaded and analyzed in a fully automated fashion (C++, ITK) and a full scanner performance report (PDF) is generated. The report is designed to demonstrate Joint Commission compliance and record other performance metrics (Fig 6). In the final step, a medical physicist reviews the images and PDF report (Fig. 7), performs an assessment for lesion detectability and artifacts, and approves the report.

Robustness of the automated phantom analysis software for the CTN and NEMA IQ phantoms has been validated across a number of PET/CT systems over a range of orientations and statistical image quality (Ulrich, Med Phys 23 Nov 2017). The uniform phantom algorithm for spatial resolution has been similarly validated. Analysis of three phantom datasets (uniform, NEMA IQ, and CTN) is benchmarked at 5 minutes 20 seconds.
This efficient Joint Commission compliant phantom program for PET/CT has been developed and tested using commonly available phantoms and a robust cloud-based analysis tool. Dissemination of the program to the general public is planned in the coming year.

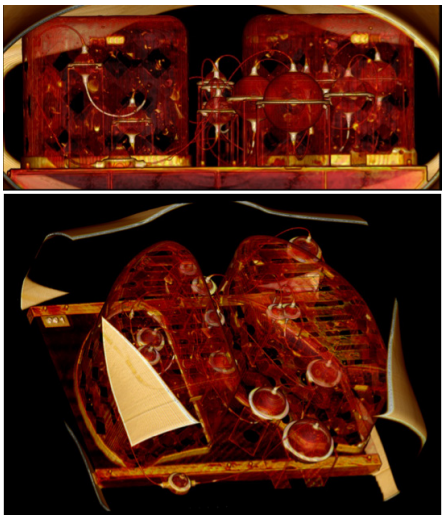


Figure 5: Example Phantoms

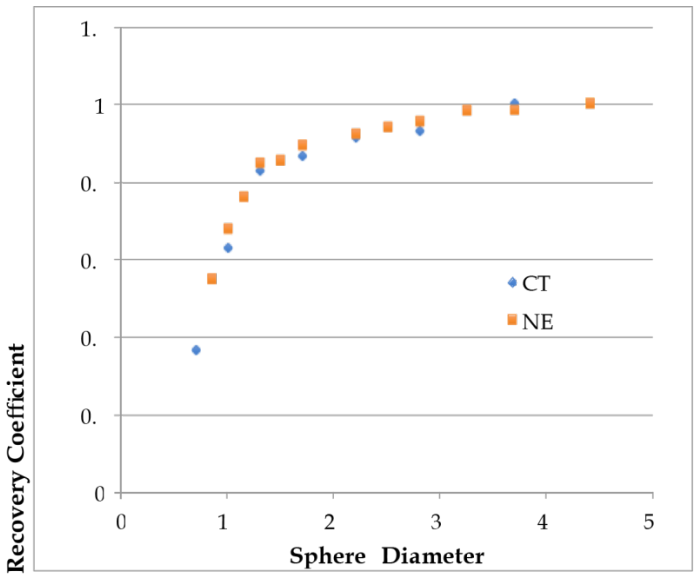


Figure 6: Recovery coefficients for CTN spheres and modified NEMA IQ Phantom (8.5--- 44mm)

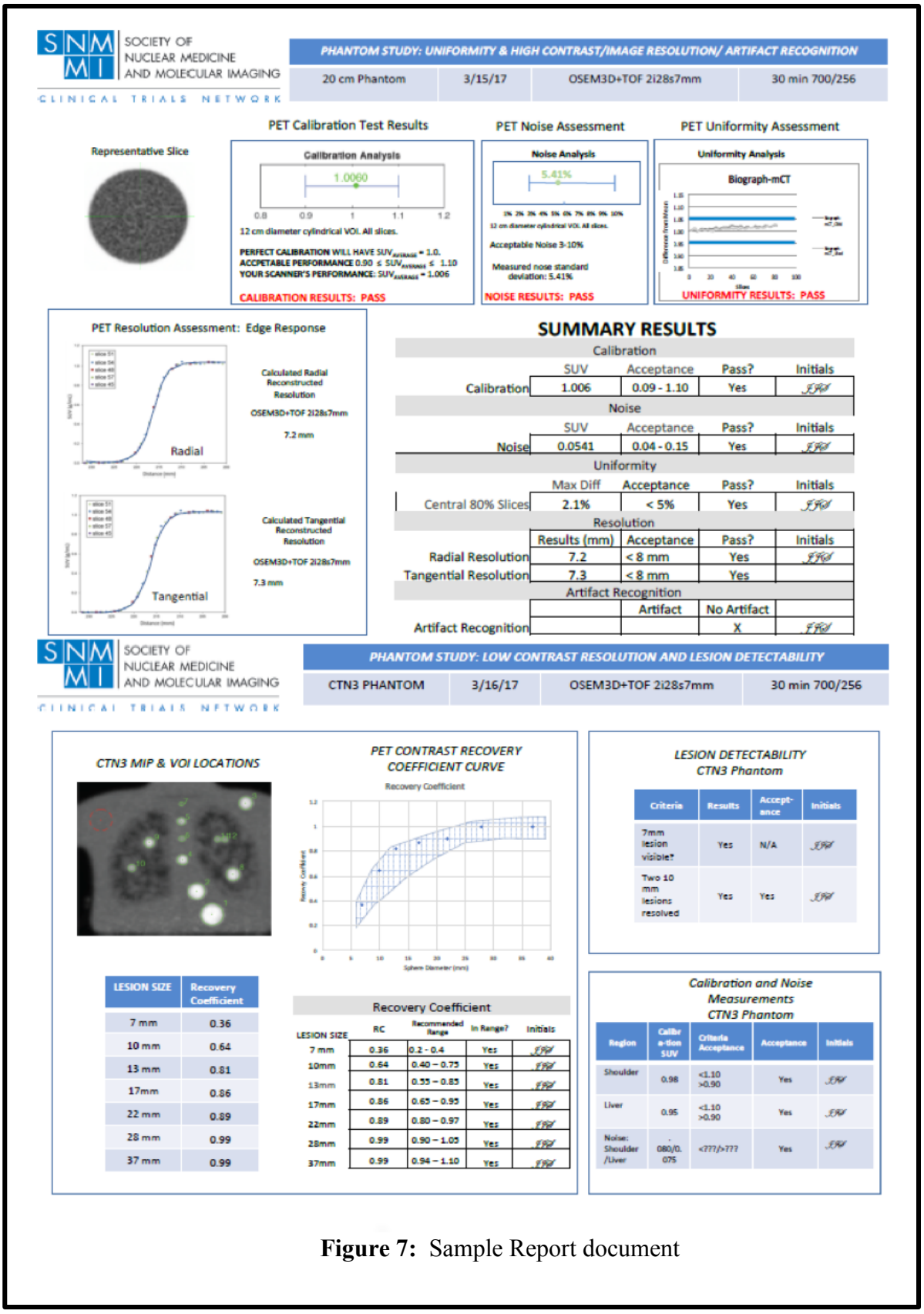


Figure 7: Sample Report document

§ Aim 4

The focus in Aim 4 over the past year was to validate the changes in T_2*/QSM measurements in response to administration of intravenous high dose vitamin C in gliomas. Labile iron has been demonstrated *in vitro* to be one of the mechanisms by which pharmacological ascorbate is selectively toxic to tumor cells. Labile ferric iron (Fe^{3+}) is proposed to be reduced to ferrous iron (Fe^{2+}). This reduction is what we aim to image using T_2* and QSM. To demonstrate first that T_2* relaxation times and susceptibility are affected specifically by Fe^{3+} and not by Fe^{2+}, phantoms consisting of physiologically relevant concentrations of iron were created in a agar based phantom. The iron concentrations ranged from 0 to 175 uM. The phantom was scanned at on a Siemens Tim TRIO scanner using a 12 channel head coil. A multi-echo gradient-echo sequence was collected with the following parameters: TE [8, 16, 24, 32, 40, 48, 56, and 64 ms], TR 300 ms, flip angle 30 degrees. Both the T_2* relaxation times and the magnetic susceptibility maps were found to be linear with Fe^{3+} concentrations while the Fe^{2+} had little effect (Figure 8).

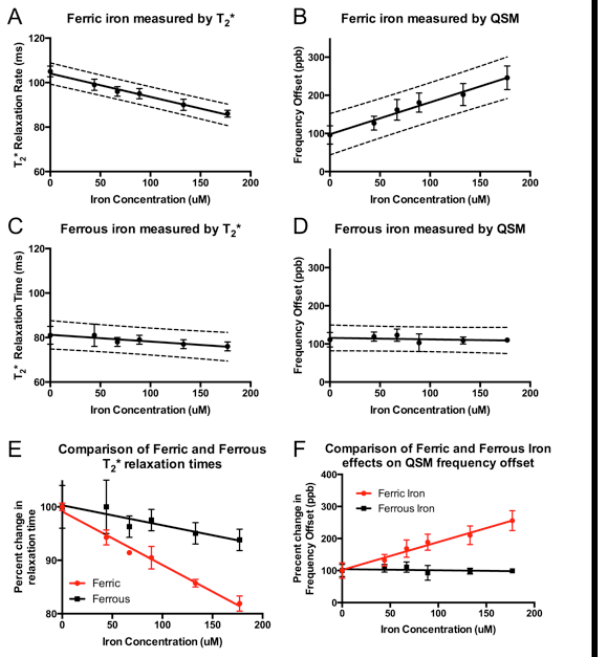


Figure 8: Effects of Fe^{3+} (ferric) and Fe^{2+} (ferrous) concentrations on T_2* relaxation times and magnetic susceptibility as measured by QSM

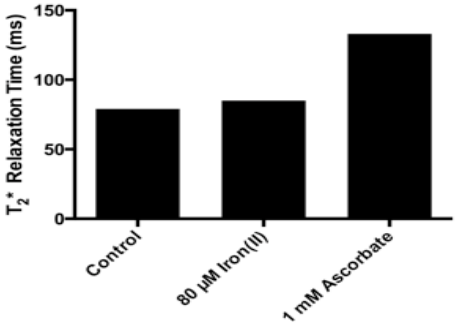


Figure 9: Effects of ascorbate on H1299 cells

This study then expanded into cancer cell lines to observe the affect that ascorbate had on these cells. H1299 cells were exposed to either 80 µM ferrous sulfate, 1 mM ascorbate, or left untreated. Cells were pelleted and loaded into an agar phantom. The phantom was scanned using a multi-echo gradient echo sequence with echoes times of 4.6, 10, 20, and 40 ms. Relaxation maps were calculated by fitting a mono-exponential decay curve to the echoes on a voxel-by-voxel basis. The data shown in Figure 9 suggest that Fe^{3+} is reduced *in vitro*, and that this reduction can be measured *via* T_2* relaxometry.

To demonstrate the alteration of iron oxidation state *in vivo,* Patients were recruited from a Phase II trial of pharmacological ascorbate as an adjuvant to standard of care for glioblastoma. 10 subjects undergoing radiotherapy and temozolamide treatment for glioblastoma received concurrent ascorbate. The imaging study was conducted on the same day as boost simulation. Ascorbate therapy was administered between morning and noon imaging sessions. Tumors were contoured to gross tumor volume using FLAIR and contrast-enhanced T1. Tumor median relaxation rates increase after therapy and continue to increase for an additional four hours (Figure 10). This change indicates that the steady-state level of Fe^{3+} is decreased. This suggests enhanced redox cycling of labile iron, one of the key mechanisms of the selective toxicity of ascorbate to tumors. No significant change is seen in normal tissue, suggesting that, as seen in biochemical assays, normal tissue has a smaller redox-active labile iron pool, leading to nearly zero toxicity in normal tissue.

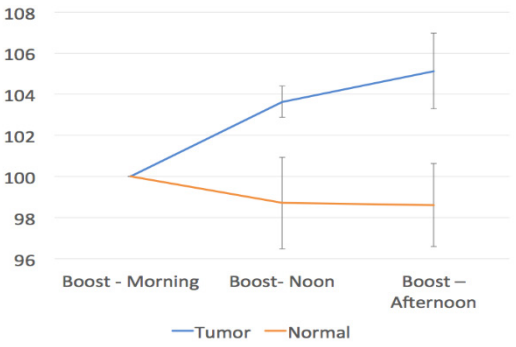


Figure 10: T2* relaxation times shown before (morning), immediate after (noon), and 4 hours post (afternoon) ascorbate administration.

COLLABORATIONS WITHIN THE NETWORK

Our efforts in comparing PET segmentation approaches on a national level have led to first results. In the first analysis phase of the QIN PET segmentation challenge (seven QIN sites participated), we have assessed the bias and variability of PET phantom and clinical HNC scan segmentations. Findings were summarized in a Medical Physics paper [4], which was highlighted as Editors' Choice in the February 2017 issue of Medical Physics. We continue our efforts with a second analysis phase, where we will focus on assessing the variability of PET segmentation derived quantitative imaging biomarkers. For this purpose we have established a statistical analysis framework that was published in Statistical Methods in Medical Research[6]. Quantitative biomarkers derived from medical images are being used increasingly to help diagnose disease, guide treatment, and predict clinical outcomes. Measurement of quantitative imaging biomarkers is subject to bias and variability from multiple sources, including the scanner technologies that produce images, the approaches for identifying regions of interest in images, and the algorithms that calculate biomarkers from regions. Moreover, these sources may differ within and between the quantification methods employed by institutions, thus making it difficult to develop and implement multi-institutional standards. We present a Bayesian framework for assessing bias and variability in imaging biomarkers derived from different quantification methods, comparing agreement to a reference standard, studying prognostic performance, and estimating sample size for future clinical studies. The statistical methods are illustrated with data obtained from a positron emission tomography (PET) challenge conducted by members of the NCI's Quantitative Imaging Network program, in which tumor volumes were measured manually and with 7 different semi-automated segmentation algorithms. Estimates and comparisons of bias and variability in the resulting measurements are provided along with an R software package for the technical performance analysis and an online web application for sample size and power analysis (Figure11).

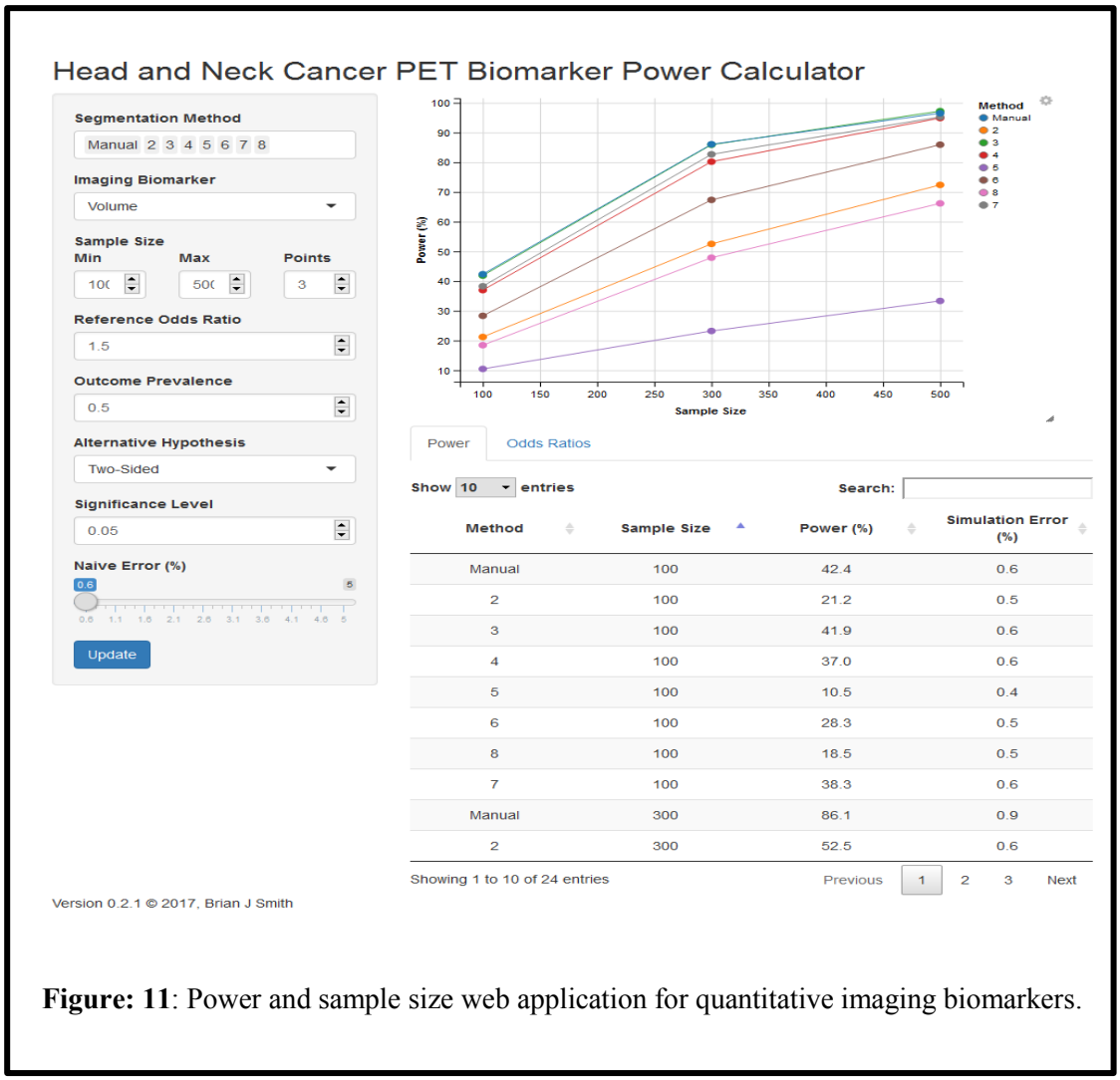


Figure: 11: Power and sample size web application for quantitative imaging biomarkers.

In addition, the group has participated in two papers under review from the Clinical Trials Design and Development Working Group. One involves Quantitative Imaging in Radiation Oncology and the other Standards in Reporting in Quantitative Imaging. Initial contribution evaluating the ability to use Auto-percist in a multi-group collaboration is also under analysis.

PLANS FOR NEXT YEAR

- Finish Joint-Commission SNMMI dissemination project
- Publish results of FDG and FLT QIB analysis
- Finish and evaluate FLT PET based quantification of bone marrow.
- Finish second phase of PET Segmentation challenge by summarizing results in a paper and submit to a journal for review.
- Participate in QIN phantom study of 'hypoxic fraction' measurement
- Initiate analysis of whole exome sequence data from University of Iowa
- Begin analysis of cooperative group clinical data from NRG 0522 and ECOG-ACRIN 6685

PUBLICATIONS FROM QIN EFFORTS

a) 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. *Med Phys.* Feb 2017;44(2):479-496. PMID: 28205306.

b) Smith BJ, Beichel RR. A Bayesian framework for performance assessment and comparison of imaging biomarker quantification methods. *Stat Methods Med Res.* Jan 1 2017:962280217741334. PMID: 29271301.

c) Yusung K, Patwardhan KA, Beichel RR, Smith BJ, Mart C, Plichta KA, Chang T, Sonka M, Graham MM, Magnotta V, Casavant T, Junyi X, Buatti JM. Development of a radiobiological evaluation tool to assess the expected clinical impacts of contouring accuracy between manual and semi-automated segmentation algorithms. *Conf Proc IEEE Eng Med Biol Soc.* Jul 2017;2017:3409-3412. PMID: 29060629.

d) Ulrich EJ, Sunderland JJ, Smith BJ, Mohiuddin I, Parkhurst J, Plichta KA, Buatti JM, Beichel RR. Automated model-based quantitative analysis of phantoms with spherical inserts in FDG PET scans. *Med Phys.* Jan 2018;45(1):258-276. PMID: 29091269.

e) Graham MM, Gu X, Ginader T, Breheny P, Sunderland JJ. (68)Ga-DOTATOC Imaging of Neuroendocrine Tumors: A Systematic Review and Metaanalysis. *J Nucl Med*. 2017 Sep;58(9):1452-1458. Erratum in: J Nucl Med. 2017 Oct;58(10):1707. PMID: 28280220

f) Schoenfeld JD, Sibenaller ZA, Cramer-Morales KL, Mapuskar KA, Wagner BA, Furqan M, Sandhu S, Carlisle TL, Smith MC, AbuHejleh T, Berg DJ, Zhang J, Keech J, Parekh KR, Bhatia S, Monga V, Bodeker KL, Ahmann L, Vollstedt S, Brown, H, Shanahan Kauffman EP, Schall ME, Hohl RJ, Clamon GH, Greenlee JD, Howard MA, Schultz MK, Smith BJ, Riley DP, Domann FE, Cullen JJ, Buettner GR, Buatti JM, Spitz D, Allen, BG. $O_2^{\cdot-}$ and H_2O_2-Mediated Disruption of Fe Metabolism Causes the Differential Susceptibility of NSCLC and GBM Cancer Cells to Pharmacological Ascorbate. Cancer Cell. 2017 Apr 10;31(4):487-500. PMID: 28366679.

g) Byrd D, Christopfel R, Buatti J, Moros E, Nehmeh S, Opanowski A, Kinahan P. Multicenter survey of PET/CT protocol parameters that affect standardized uptake values. J Med Imaging (Bellingham). 2018 Jan;5(1):011012. PMID: 29250567; PMC5722234

REFERENCES

1. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* 2011-12-12 2011;45(3):67.

2. Ulrich EJ, Sunderland JJ, Smith BJ, Mohiuddin I, Parkhurst J, Plichta KA, Buatti JM, Beichel RR. Automated model-based quantitative analysis of phantoms with spherical inserts in FDG PET scans. *Med Phys.* Jan 2018;45(1):258-276. PMID: 29091269.

3. 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):2948-2964. PMID: 27277044. PMCID: PMC4874930.

4. 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. *Med Phys.* Feb 2017;44(2):479-496. PMID: 28205306.

5. Lodge M, Leal J. A PET cylinder phantom positioned at an oblique angle. *J Nucl Med.* May 1, 2017 2017;58(supplement 1):701.

6. Smith BJ, Beichel RR. A Bayesian framework for performance assessment and comparison of imaging biomarker quantification methods. *Stat Methods Med Res.* Jan 1 2017:962280217741334. PMID: 29271301.

U01 CA190214: Qualification and Deployment of Imaging Biomarkers of Cancer Treatment Response

Stanford University Department of Radiology (team 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 endpoints 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.

PROGRESS OVER THE PAST YEAR

§ 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 goal of generalizability, we will develop and deploy two 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.

§ Progress on the Specific Aims

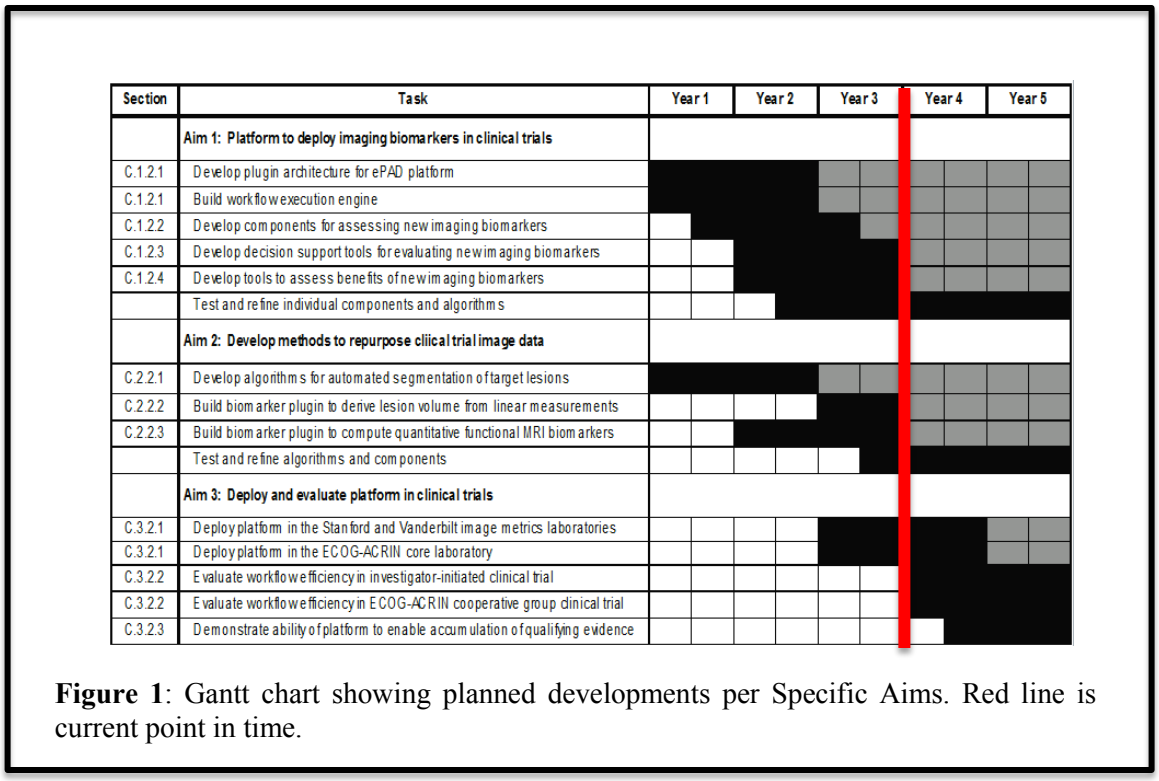


Figure 1: Gantt chart showing planned developments per Specific Aims. Red line is current point in time.

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)):

AIM 1: Develop a suite of configurable image feature characterization algorithms:

C.1.2.1 **Plugin architecture and workflow execution engine for deploying new imaging biomarkers**: We improved our plugin mechanism to accommodate the expanding number of plugins with a plugin store. The plugin store is a more extensive and flexible way of installing/managing the default local plugins and remote plugins developed by ePad Team or third parties. Local plugins are embedded in ePAD and can be activated or deactivated. The remote plugins, on the other hand are downloaded from our repository with the installation request. The plugins can specify parameters in addition to the executable files.

There are three components of the plugin architecture of ePAD:

- **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. We have improved ePAD to aggregate biomarkers automatically from each annotation and save them to the produced

AIM file. The biomarkers that are collected automatically for all the annotations are minimum, maximum, standard deviation and mean for all the pixels that are in the region of interest. If the region of interest is marked using a line, ePAD also calculates the length of the line and stores it. If the region of interest is marked using perpendicular lines, ePAD will calculate the length of the long axis and short axis, and store the values as long axis and short axis respectively in the AIM file.

We have also added new biomarker plugins:

Quantitative Image Feature Engine (QIFE): An Open-Source, Modular Engine for 3D Quantitative Feature Extraction from Volumetric Medical Images: QIFE is an open-source feature-extraction framework that works on Quantitative Imaging Feature Pipeline (QIFP), a separate QIN project undertaking developing a workflow engine. It computes 3D radiomics features for the region of interests that are marked as DICOM Segmentation Objects [Echegaray S, Bakr S, Rubin DL, Napel S. Quantitative Image Feature Engine (QIFE): An Open-Source, Modular Engine for 3D Quantitative Feature Extraction from Volumetric Medical Images. Journal of Digital Imaging 2017]. ePAD can run QIFE through QIFP, get the feature values and store them in an AIM file for future reference.

Quantitative Feature Explore (QFExplore) plugin suite - Feature Extraction, Comparison and Classifier: The Quantitative Feature Explore (QFExplore) plugin suite for ePAD platform enables the exploration and validation of imaging biomarkers in a clinical environment [Schaer R, Dicente Cid Y, Alkim E, John S, Rubin DL and Depeursinge A, Web-Based Tools for Exploring the Potential of Quantitative Imaging Biomarkers in Radiology: Intensity and Texture Analysis on the ePAD Platform, in: Biomedical Texture Analysis: Fundamentals, Applications and Tools, Elsevier, 2017]. The latter include:

- the extraction, visualization and comparison of intensity- and texture-based quantitative imaging features (Fig. 2),
- regional division of Regions of Interests (ROI) to reveal tissue diversity
- the construction, use and sharing of user-personalized statistical machine learning models,
- helper tools for image segmentation are also available (Exampler #3c).

Imaging features that can be extracted using QFExplore are:

- histogram bins of Pixel Intensity Distributions (PID),
- statistical moments of PIDs (i.e., mean, standard deviation, skewness, kurtosis),
- Gray-Level Co-occurrence Matrices (GLCMs),
- Riesz wavelets

The machine learning model that is used is based on linear Support Vector Machines (SVMs).

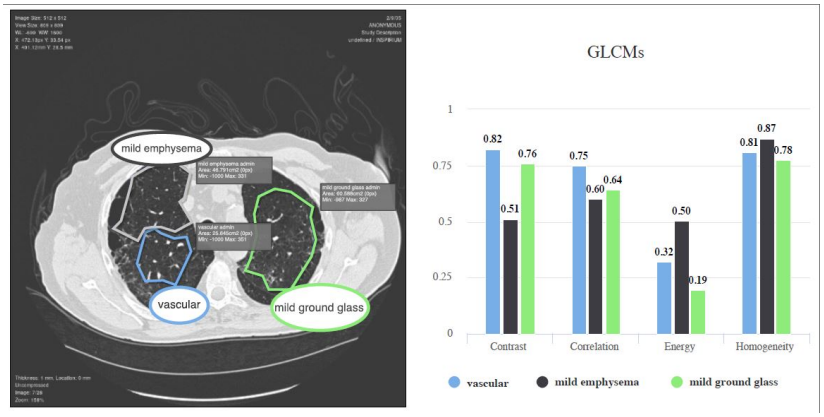


Figure 2: QF Explore Plugin Suite: GLCM feature extraction and comparison chart. The user can compare the feature's values for various ROIs. GLCM contrast and correlation is higher for vascular ROIs.

- **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 the RECIST, Longitudinal and Waterfall reports as application modules for a modular design and independence from platform. See section C.1.2.3 for more information on reporting tools.

- **Workflow execution engine**: We have implemented a prototype of workflow engine by running the QIFE feature extraction via Quantitative Imaging Feature Pipeline (QIFP). The prototype runs the QIFE to extract 3D quantitative features for the segmentation objects that are created in ePAD with default parameters, waits for the results to be ready and writes the features to the AIM file when the results are received.

  We have implemented running plugins on the existing annotations after the fact. There are two types of plugins, first type works on a single annotation whereas the second one gets a list of annotations and extracts information by comparing/processing all of them cumulatively. Depending on the type of the plugin, ePAD can run the plugin in parallel by queuing the execution on each annotation or sending all the annotations to the plugin.

  During the interoperability efforts, we have developed a plugin wrapper to run a Slicer plugin, Lung segmentation plugin in particular. The wrapper sends the

request to the Slicer via SlicerChronicle module, retrieves the created DICOM Segmentation Object when it is ready and saves it in ePAD.

We also made enhancements to the core ePAD functionality, including DicomRT migration support, DicomSR migration and export support for segmentation annotations and more extensive DICOM Segmentation Object (DSO) support with probability masks. We are also participating actively on Supp 200-Transformation of NCI Annotation and Image Markup (AIM) and DICOM SR Measurement Templates Standard efforts to extend the DICOM-SR support to all annotation types.

C.1.2.2 **Image viewing to facilitate assessment of quantitative imaging biomarkers**: We made progress on several of the tasks:

Facilitating image biomarker assessments by clinical centers
The waterfall plot tool shows the summary of the biomarker assessment taking into account the quantitative measurements of the annotations across a study for a cohort of patients. More information about the waterfall plot tool can be found below (see "Waterfall Report").

Facilitating oversight of image readings by clinical trial researchers and sponsors: ePAD associates radiologists with the images they interpret. We developed 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. The module enables a user to follow the image annotations made in multiple studies by group of users assigned to a particular study. It can also track the progress of the annotation process by identifying which subjects are fully annotated by all the annotators, which annotators have completed the annotation process for each subject and which subjects/studies have not been annotated yet (Figure 3).

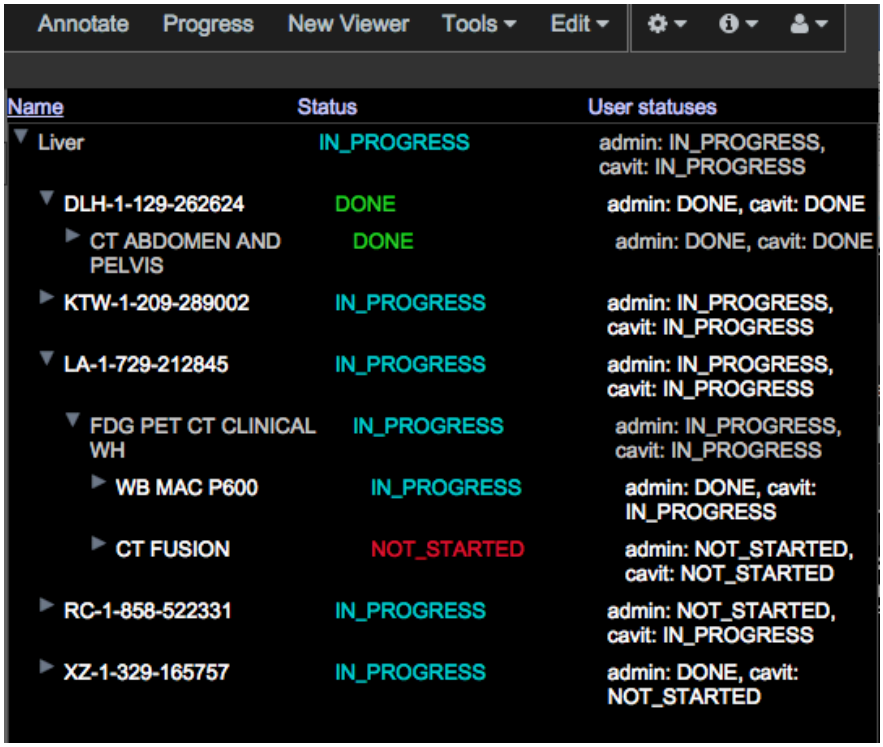


Figure 3: Progress View

This functionality has been helpful in a project that adopted ePAD: the MGH/HST Martinos Center for Biomedical Imaging) used this for MEDICI project.

In addition, we have added a view in ePAD to show Multi-planar reconstruction (MPR) view of DICOM images and to support drawing of 3D annotations on the MPR view (Figure 4). To support the MPR development, we have leveraged the open source AMI JavaScript toolkit in ePAD (https://github.com/FNNDSC/ami).

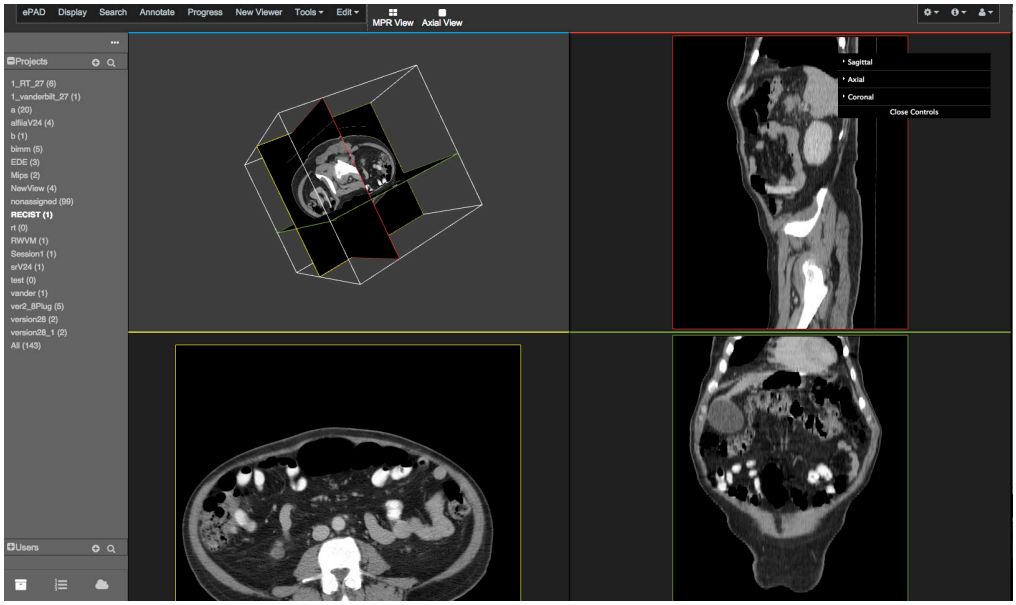


Figure 4: MPR view of a series implemented using AMI.js in ePAD.

C.1.2.3 **Decision support in assessing treatment response**: We have developed tools to assist decision making based on image biomarker assessments in two major clinical trial tasks: (1) determine treatment response in patients (RECIST, Longitudinal annotation report), and (2) evaluate treatment effectiveness by determining the cohort-based treatment response (Waterfall).

RECIST Report: We have improved our RECIST report by re-implementing it as an application module (Figure 5). It analyzes all the annotations of a patient, calculates sum of lesion dimensions (SLD) on each time point. RECIST report generation depends on the usage of the RECIST templates to annotate the lesions and supports the geometric shape annotations of line and perpendicular lines. The report checks the consistency of the annotations, if the location is specified differently in the annotation of the same lesion on different time points, the cell will be marked as error to notify the user. The report also marks missing annotations for a lesion as error.

The user can open the annotation in the image viewer by clicking on the annotation measurement on a specific time point. The user can also open all annotations of the lesions on all time points by clicking the lesion name. The

populated report can be exported as a word document that can be filed in the clinical workflow.

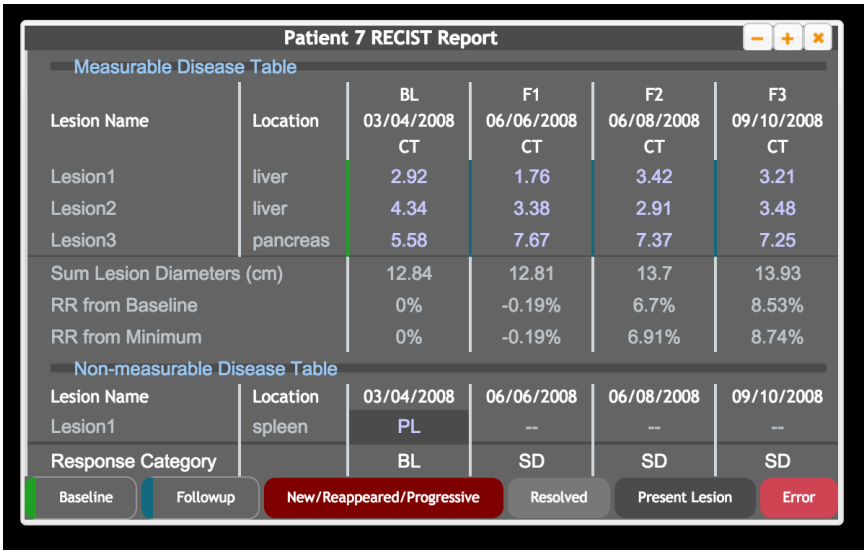


Figure 5: RECIST Report with measurable and non-measurable disease tables.

Longitudinal Annotation Reporting: We have implemented a more generic Longitudinal annotation reporting tool that analyzes all the annotations of a subject and doesn't filter automatically for any template (Figure. 6). The report populates three dropdown menus by analyzing the annotations: shape, template and measurement type. The user can filter using the shape and/or template or choose to see all annotations. The table will be populated using the selected measurement type. If the measurement is not present in a specific time point of a lesion, the table display it as NA. The summary section will be filled automatically for the measurement type in a similar manner to the RECIST report but by using the selected measurement type. The access to the annotations on the image viewer works in the same manner with RECIST report.

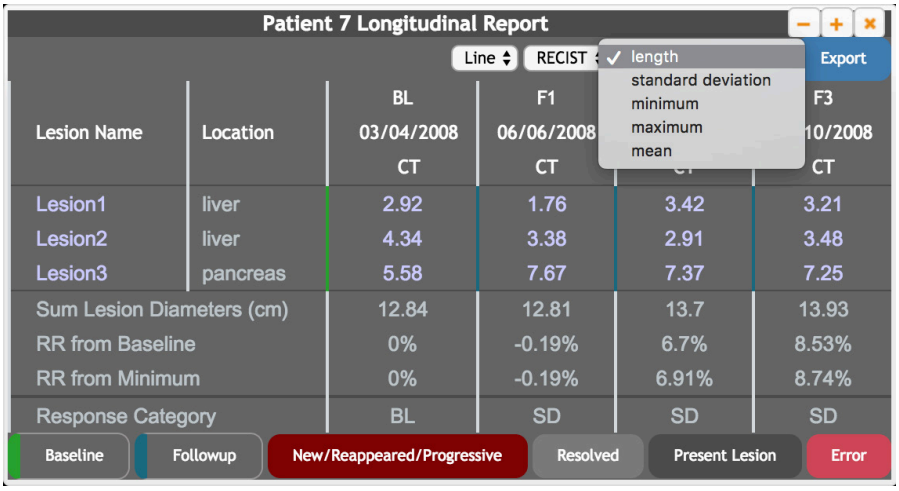


Figure 6: Longitudinal annotation report

Waterfall Report: We have implemented a waterfall report plotting application module to evaluate treatment effectiveness by determining the cohort-based treatment response (Figure 7).

The report can use either RECIST and ADLA as the imaging biomarker used to calculate the best response rate of a subject. If the user selects to use RECIST, Waterfall report module analyzes every subject in the cohort, generates the RECIST tables, gets the best response for each subject and plots it in a decreasing order forming a waterfall plot. If the user selects to use ADLA, Waterfall report module generates an ADLA table for each subject by using the Longitudinal report and filtering the shape to Line and using the standard deviation as the measurement type. Then, the best response from the ADLA table for each patient is used to plot the waterfall graph. The waterfall plot is responsive, the user can access the table that is used to make the best response rate computation by clicking the specific bar in the waterfall plot.

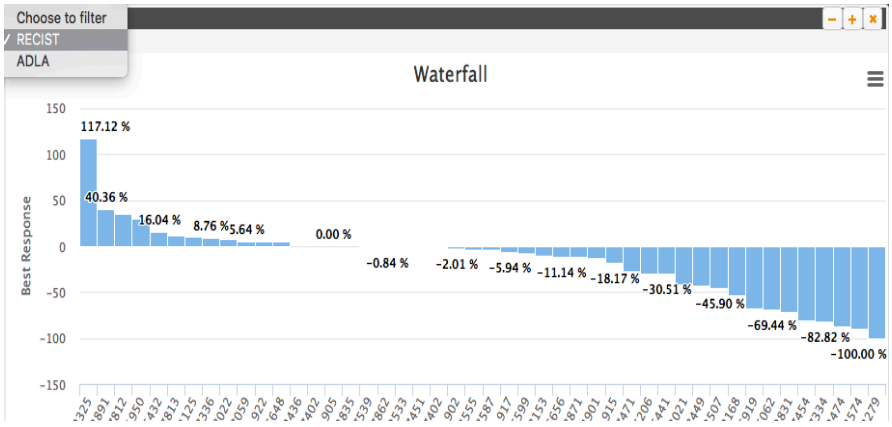


Figure 7: Waterfall report plot

Dissemination: We have kept our public website for ePAD up-to-date. Our public website, http://epad.stanford.edu/ contains introductory material, a demo movie, documentation, a detailed description of the developer interface, download information in addition to user and download statistics. 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 **Automated segmentation in PET-Tedseg**: We have integrated an automatic segmentation plugin which works on PET images. The plugin, which was developed by Edward Graves [Technol Cancer Res Treat, 6(2):111-21, 2007], is triggered with a seed point. It analyzes the image volume to find the whole

region of the lesion and creates a DICOM Segmentation object marking the volume of the lesion (Figure 8).

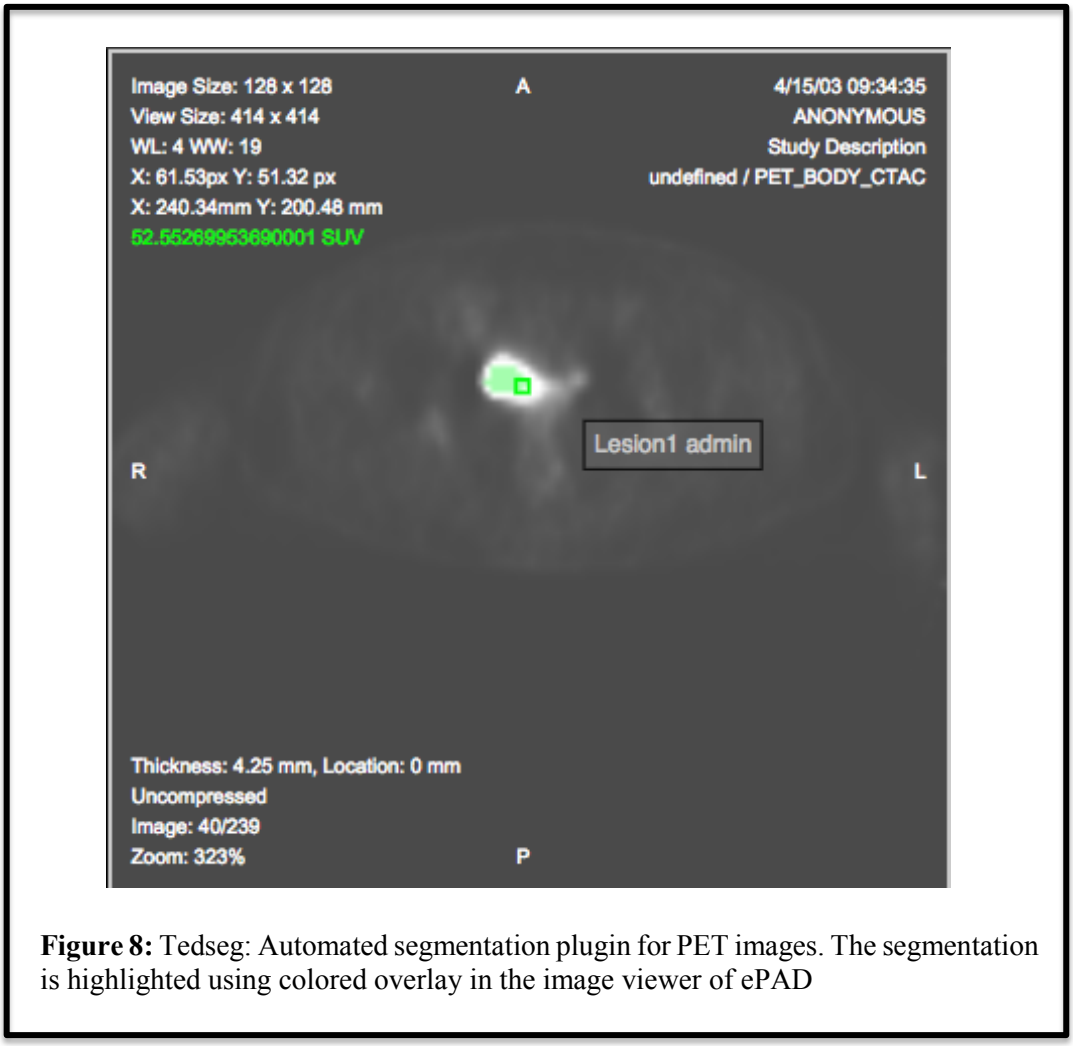


Figure 8: Tedseg: Automated segmentation plugin for PET images. The segmentation is highlighted using colored overlay in the image viewer of ePAD

C.2.2.2 **Automated 2D lesion segmentation in MR-LesionSeg**: We have integrated an automatic 2D lesion segmentation plugin which works on MR images. The plugin, is triggered with drawing a polygon inside the region of interest. It analyzes the image and expands the polygon ROI to

the edges by adding more points [Ref 12; Med Image Anal 2017].

C.2.2.3 **Automated image segmentation of QF Explore Plugin Suite**: The QF Explore plugin suite has a plugin for automatically segmenting lungs in a DICOM volume (Figure 9).

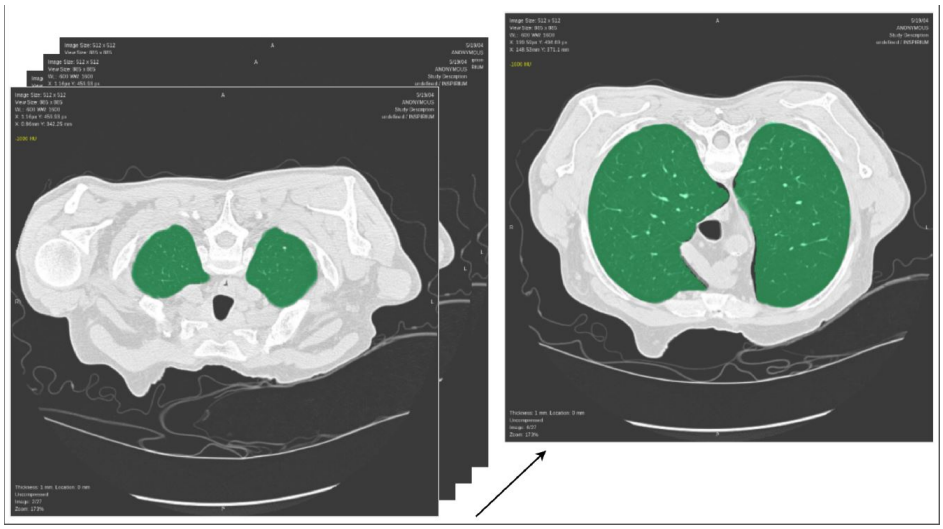


Figure 9: QF Explore Plugin Suite – Automated image segmentation plugin: Two images of automatically segmented lungs in a DICOM volume. The detected lung regions are highlighted with green image overlays in the front-end of ePAD.

C.2.2.4 **ADLA biomarker plugin:** We have integrated Attenuation Distribution across the Long Axis (ADLA) plugin. ADLA is a semi-quantitative imaging biomarker for assessing treatment response in solid malignancies and a measure of intralesional heterogeneity [Nikita Lakomkin, Hakmook Kang, Bennett Landman, Radiology, Volume 23, Issue 6, 2016, Pages 718-723, ISSN 1076-6332]. ePAD calculates the standard deviation along the long axis to compute the ADLA and saves in the AIM

file to be used for further analysis. ePAD also draws an ADLA histogram when the long axis is selected (Figure 10) and an ADLA change report (Figure 11)

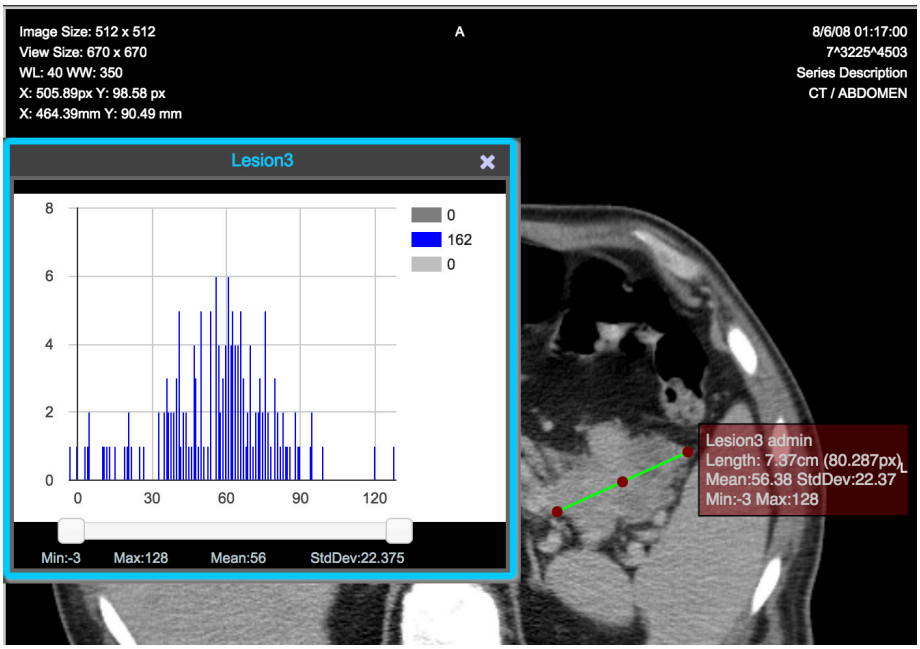
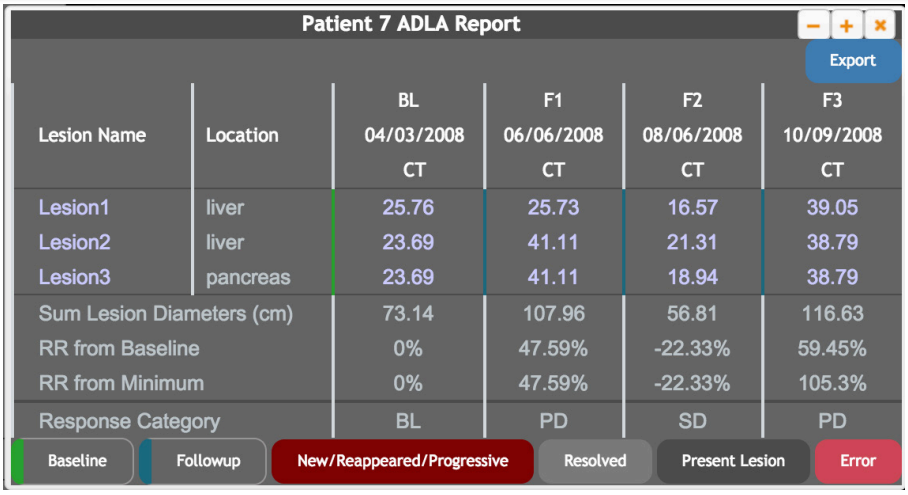


Figure 10: ADLA histogram



Lesion Name	Location	BL 04/03/2008 CT	F1 06/06/2008 CT	F2 08/06/2008 CT	F3 10/09/2008 CT
Lesion1	liver	25.76	25.73	16.57	39.05
Lesion2	liver	23.69	41.11	21.31	38.79
Lesion3	pancreas	23.69	41.11	18.94	38.79
Sum Lesion Diameters (cm)		73.14	107.96	56.81	116.63
RR from Baseline		0%	47.59%	-22.33%	59.45%
RR from Minimum		0%	47.59%	-22.33%	105.3%
Response Category		BL	PD	SD	PD

Figure 11: ADLA Report

C.2.2.5 **T1 map extraction from Philips MRI images**: We have integrated a T1 map extraction plugin that is developed by Vanderbilt University [Yankeelov and Gore, Curr Med Imaging Rev. 2009 May 1;3(2):91-107]. The plugin analyses the multiframe MRI image with different phases and calculates the T1 map for

the volume. ePAD gets the T1 map volume, scales it to 8 bits to be able to save in DICOM standard, saves in a probability DICOM Segmentation object and paints the mask on the image using a color LUT (Figure 12).

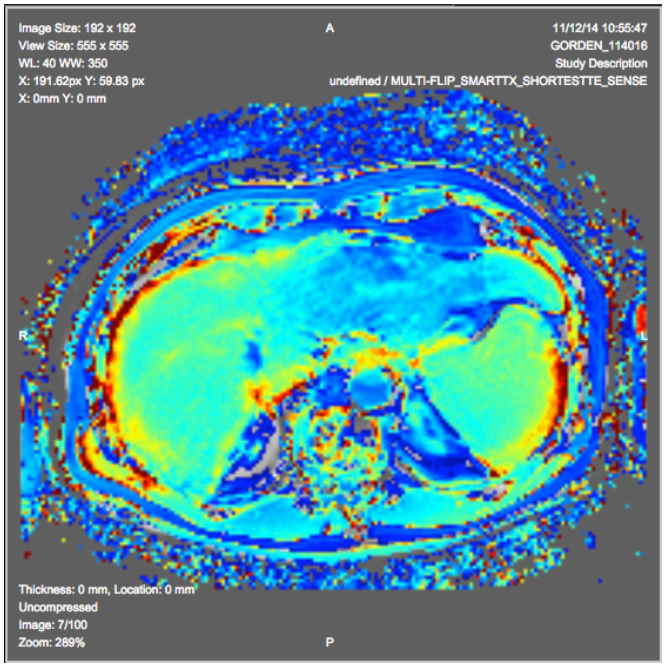


Figure 12: T1 map overlaid on MR image using color LUT

COLLABORATIONS WITHIN THE NETWORK

We have engaged actively in QIN collaboration as well as community outreach and dissemination.

Vanderbilt QIN: We worked with the Vanderbilt QIN to deploy their perfusion MRI biomarker methods as a plugin to ePAD (see C.2.2.5).

MGH/HST Martinos Center for Biomedical Imaging): This site used ePAD functionality for tracking radiologist progress in making image annotations in their MEDICI project. We interacted with them to collect requirements and address usage issues.

ECOG-ACRIN: We have begun interacting with the ECOG-ACRIN cooperative group Core Laboratory to interface it with their DART infrastructure to enable collecting image annotations as part of clinical trials in AIM format and storing that in DART.

IROC: We have begun talking with IROC about doing a pilot project of using ePAD for image management within their clinical trial workflow to facilitate collection and management of image annotations as part of clinical trial workflow in IROC.

Commercial and open source interoperability: We have worked with various laboratories to develop migration tools for ePAD to enable ePAD to leverage the existing annotations that are created by other software tools. These include ROIs exported from Osirix, ROIs and measurements collected by MINT Lesion commercial software. We have also made improvements in ePAD core capabilities to handle the laboratories' special needs, in particular to support small imaging data better.

NCIP and DICOM Committee: We are also participating actively in an NCIP-funded project to harmonize AIM with the DICOM standard, which was taken up in the past year by DICOM WG-8. This work is extending DICOM-SR support to AIM annotation types.

PLANS FOR NEXT YEAR

§ Tumor volume plugin:

We will create a plugin to compute the volume of target lesions that have been outlined using DSO objects. We will use this plugin in our evaluate Aim 3 studies as one of the imaging biomarkers to be compared against another biomarker (ADLA).

Tools to assess the benefits of new imaging biomarkers:

The Waterfall chart module was implemented for the RECIST and ADLA tumor response rules. As a subsequent step, 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, RECIST vs ADLA) 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.

Build biomarker plugin to derive lesion volume from linear measurements:

In addition:

1. In addition to the MPR view, we will develop a new viewer using AMI.js that will support display of multi-segment DICOM Segmentation Objects (DSOs) and 3D ROI annotations.

2. There are several plugins that have been implemented outside of ePAD and we will integrate these plugins as modules that can be executed within ePAD:

a. Automated lesion tracking: This plugin will use a 2D ROI and image to generate binary masks or segmentation that will be used to create a mesh that can be visualized in ePAD as a 3D volume rendering model.
The automated progress or tracking of a lesion will be triggered after selecting a baseline AIM file, which will then prompt the module to generate AIM files for follow-up segmentations. This module will have a User Interface for selection of baseline annotation and for displaying the automated lesion tracking.

b. FASR plugin: This tool will return a predictive score after a user enters inputs for questions in a template for mammography interpretations, providing the probability of malignancy for evaluated lesions. We will integrate this plugin into ePAD.

3. We will make regular public releases of ePAD and will submit an educational exhibit to RSNA 2018 that will allow us to begin to train the broader community regarding the use of the QIFP.

Specific Aim 3: We will commence work on this Specific Aim, deploying and beginning to evaluate our platform and tools in the core imaging laboratories of two cancer centers and the ECOG-ACRIN national cooperative group. We will apply ePAD retrospectively to a recently-completed ECOG-ACRIN cooperative group trial (SWOG 0518), 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. 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.

PUBLICATIONS AND PRESENTATIONS FROM QIN EFFORTS

§ Published papers:

1. 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.

2. 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.

3. 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.

4. 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.

5. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. J Digit Imaging. 2017; 30:449-459. doi: 410.1007/s10278-10017-19983-10274.

6. Banerjee I, Chen MC, Lungren MP, Rubin DL. Radiology Report Annotation using Intelligent Word Embeddings: Applied to Multi-institutional Chest CT Cohort. J Biomed Inform 2017; 23:30257-30255.

7. Banerjee I, Malladi S, Lee D, Depeursinge A, Telli M, Lipson J, Golden D, Rubin DL. Assessing treatment response in triple-negative breast cancer from quantitative image analysis in perfusion magnetic resonance imaging. J Med Imaging (Bellingham). 2018; 5:011008. doi: 011010.011117/011001.JMI.011005.011001.011008. Epub 012017 Nov 011002.

8. Diamant I, Hoogi A, Beaulieu CF, Safdari M, Klang E, Amitai M, Greenspan H, Rubin DL. Improved Patch-Based Automated Liver Lesion Classification by Separate Analysis of the Interior and Boundary Regions. IEEE J Biomed Health Inform. 2016; 20:1585-1594. doi: 1510.1109/JBHI.2015.2478255. Epub 2472015 Sep 2478211.

9. 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. 2016; 2:242-249. doi: 210.18383/j.tom.12016.00265.

10. Finlayson SG, Levy M, Reddy S, Rubin DL. Toward rapid learning in cancer treatment selection: An analytical engine for practice-based clinical data. J Biomed Inform. 2016; 60:104-13.:10.1016/j.jbi.2016.1001.1005. Epub 2016 Feb 1012.

11. Graim K, Liu TT, Achrol AS, Paull EO, Newton Y, Chang SD, Harsh GRt, Cordero SP, Rubin DL, Stuart JM. Revealing cancer subtypes with higher-order correlations applied to imaging and omics data. BMC Med Genomics. 2017; 10:20. doi: 10.1186/s12920-12017-10256-12923.

12. Hoogi A, Beaulieu CF, Cunha GM, Heba E, Sirlin CB, Napel S, Rubin DL. Adaptive local window for level set segmentation of CT and MRI liver lesions. Med Image Anal. 2017; 37:46-55.:10.1016/j.media.2017.1001.1002. Epub 2017 Jan 1013.

13. Hoogi A, Subramaniam A, Veerapaneni R, Rubin DL. Adaptive Estimation of Active Contour Parameters Using Convolutional Neural Networks and Texture Analysis. IEEE Trans Med Imaging. 2017; 36:781-791. doi: 710.1109/TMI.2016.2628084. Epub 2622016 Nov 2628011.

14. 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. 2017; 30:4-10. doi: 10.1007/s10278-10016-19916-10277.

15. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data. 2017; 4:170177.:10.1038/sdata.2017.1177.

16. Lekadir K, Galimzianova A, Betriu A, Del Mar Vila M, Igual L, Rubin DL, Fernandez E, Radeva P, Napel S. A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound. IEEE J Biomed Health Inform. 2017; 21:48-55. doi: 10.1109/JBHI.2016.2631401. Epub 2632016 Nov 2631422.

17. Rister B, Horowitz MA, Rubin DL. Volumetric Image Registration From Invariant Keypoints. IEEE Trans Image Process. 2017; 26:4900-4910. doi: 4910.1109/TIP.2017.2722689. Epub 2722017 Jul 2722683.

18. Yu KH, Berry GJ, Rubin DL, Re C, Altman RB, Snyder M. Association of Omics Features with Histopathology Patterns in Lung Adenocarcinoma. Cell Syst. 2017; 5:620-627.e623. doi: 610.1016/j.cels.2017.1010.1014. Epub 2017 Nov 1015.

§ Submitted Manuscripts

1. S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A.N.C. Leung M. Kadoch, J. Shrager, A. Quon, D.L. Rubin, S. K. Plevritis*, Sandy Napel*, "A Radiogenomic Dataset of Non-Small Cell Lung Cancer," submitted to Nature Scientific Data, Dec. 2017.

§ Presentations

1. Fedorov A, O'Donnell LJ, Rubin DL, Clunie DA, Flade D, Nolden M, et al., DICOM4QI demonstration and connectathon: Structured 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 third annual scientific meeting of the RSNA, Chicago, IL, 2017.

2. Rubin DL, John S, 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 third annual scientific meeting of the RSNA, Chicago, IL, 2017.

3. Rubin DL, Image Annotation and Semantic Labeling, in Refresher Course, "Radiomics Mini-Course: From Image to Omics," One hundred-third annual scientific meeting of the RSNA, Chicago, IL, 2017.

4. Rubin DL, Machine Learning and Radiomics in Practice: Tools and Case Example, in Refresher Course, "Platforms and Infrastructures for Accelerated Discoveries in Machine Learning and Radiomics," One hundred-third annual scientific meeting of the RSNA, Chicago, IL, 2017.

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.

7. S. Napel, D. L Rubin, S. John, D. Gude, S. Echegaray, S. Bakr, D. Gude, et al. "The Quantitative Image Feature Pipeline (QIFP): Automated Radiomic Feature Extraction to Derive Associations with and Prediction of Clinical Variables from Image Features," Radiological Society of North America 103rd Scientific Sessions, December 2017.

8. J. Kalpathy-Cramer, B. Zhao, D. Goldgof, S. Napel, D. L. Rubin, M. F. McNitt-Gray, et al, "Standardizing Radiomic Feature Descriptions for Quantitative Imaging: A Preliminary Report of the Cooperative Efforts of the NCI's QIN PET-CT Subgroup," Radiological Society of North America 103rd Scientific Sessions, December 2017.

U01 CA148131: Advanced PET/CT Imaging for Improving Clinical Trials

University of Washington
Seattle Cancer Care Alliance

Hannah Linden, M.D.
Dave Mankoff, M.D., Ph.D.
Paul Kinahan, Ph.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 and Results of Progress made over the previous year

Survey of PET/CT protocol parameters that affect standardized uptake values.
Clinical trials that evaluate cancer treatments may benefit from positron emission tomography (PET) imaging, which for many cancers can discriminate between effective and ineffective treatments. However, the image metrics used to quantify disease and evaluate treatment may be biased by many factors related to clinical protocols and PET system settings, many of which are site- and/or manufacturer-specific.

We conducted an observational study using two surveys that were designed to record key sources of bias and variability in PET imaging. These were distributed to hospitals across the United States. The first round of surveys was designed and distributed by the American College of Radiology's Centers of Quantitative Imaging Excellence program in 2011. The second survey expanded on the first and was completed by the National Cancer Institute's Quantitative Imaging Network. Sixty-three sites responded to the first survey and 36 to the second.

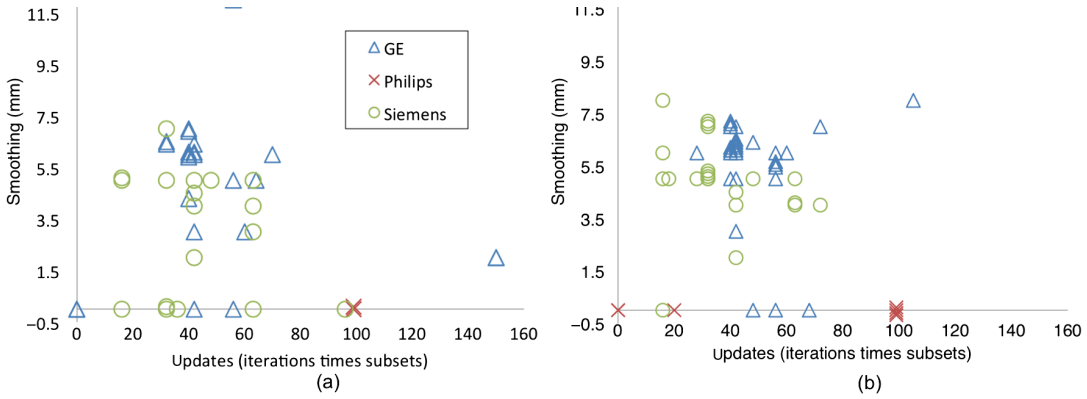


Figure 1: For body imaging, smoothing (post filter full width at half maximum) versus iterative updates for the (a) QIN survey and (b) CQIE survey. Where multiple points lie in the same location, small shifts have been introduced to make markers visible. On the x-axis, where a post-filter of 0 mm is indicated, no post-filtering was applied. Algorithms plotted on the y-axis, implying no updates, are non-iterative methods.

There are roughly 10 parameters that can affect PET SUVs, and the survey found substantial variations in all of them between sites. For example, reconstruction methods varied across scanners in the QIN survey. The most common algorithm reported was the ordered-subsets expectation maximization (OSEM) algorithm without time-of-flight data, used by 31 sites (one of these used two-dimensional OSEM). Eight sites indicated their reconstructions used time-of-flight data and three used the analytic methods. Some sites used Fourier rebinning. Figure 1. shows the reported image smoothing parameters versus the number of iterative updates (defined as the number of iterations times the number of subsets).

For scanners in the QIN survey, the reported trans-axial field-of-view diameter was 63±11 cm (range 30 to 81 cm) for body imaging and 32±10 cm (range 25 to 70 cm) for brain imaging. The surprisingly wide distributions of trans-axial voxel dimensions and slice thicknesses are shown in Figure 2.

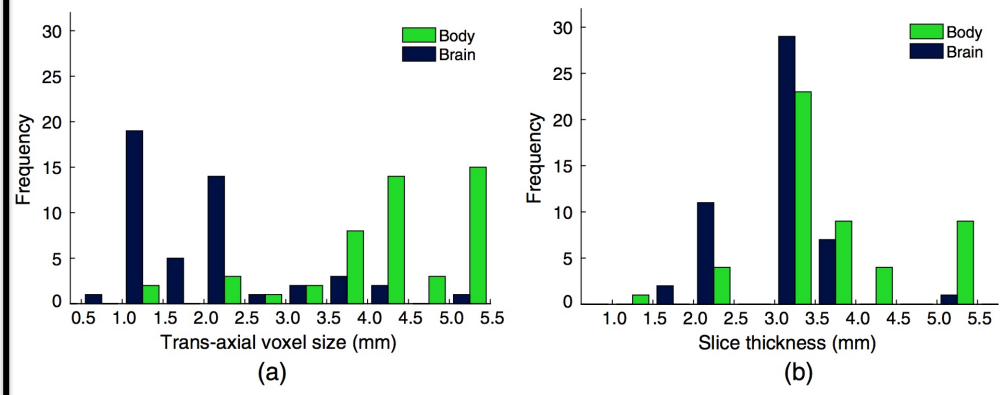


Figure 2: Histograms of (a) trans-axial voxel sizes and (b) slice thicknesses for body and brain imaging protocols at surveyed suites.

The range of reported methods for image acquisition and reconstruction suggests that signal biases are not matched between sites. Patient preparation was also inconsistent, potentially contributing additional variability. For multicenter clinical trials, efforts to control biases through standardization of imaging procedures should precede patient measurements. These results were recently published.

§ Measuring temporal stability of positron emission tomography standardized uptake value bias using long-lived sources in a multicenter network.

We have recently published a QIN-wide assessment of the variability in the calibration PET process in multicenter clinical trials. Sealed source kits containing traceable amounts of 68Ge/68Ga were distributed to 9 hospitals in the QIN (Table 1). Repeat measurements of the sources were performed on PET scanners and in dose calibrators. The measured scanner and dose calibrator signal biases were used to compute the bias in SUVs at multiple time points for each site over a 14-month period. On average, single-scanner SUV bias varied over a range of 10%. (Figure 3). Calibration factors from the image metadata were nearly as variable as scanner signal, and were correlated with signal for many scanners. This shows that SUV biases are unstable even when measurements are repeated at a single site. Long-lived sources and image metadata may provide a check on the recalibration process.

A more extensive review of these results has been published in the QIN special issue of Journal of Medical Imaging.

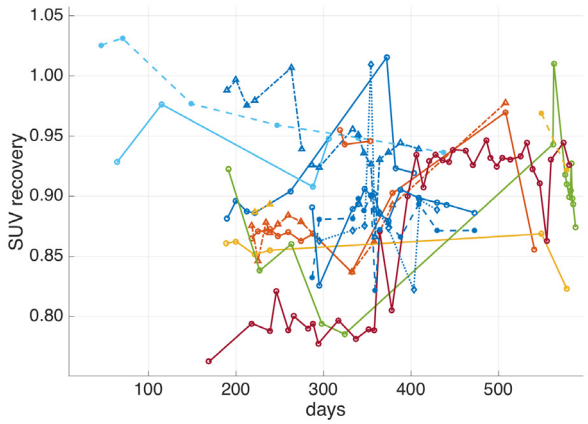


Figure 3: SUV bias versus time for 9 QIN PET sites.

Table 1: Participating QIN sites.

H. Lee Moffitt Cancer Center
Johns Hopkins University
Massachusetts General Hospital
Memorial Sloan Kettering Cancer Center
Ohio State University
University of Iowa
University of Pennsylvania
University of Pittsburgh
University of Washington

§ Virtual Clinical Trials

We are continuing to extend our 'Virtual Clinical Trial' (VCT) concept, which is used to evaluate variation 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 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 study, we were interested in measuring change in the true underlying biology of the tumor, not random change due to noise in the imaging process. Outside of the VCT framework, this would have been very difficult to do. Note that all of the data points are below the line of identity, indicating that Ki was the preferred metric for this particular patient cohort. This allowing lowering the number of patients we need to enroll a clinical trial.

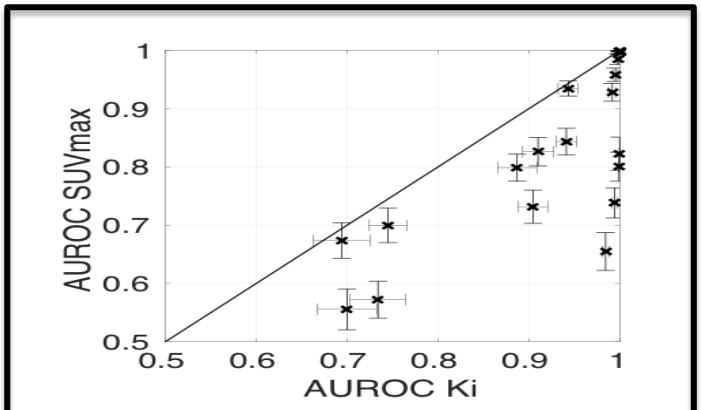


Figure 4: AUC for SUV vs AUC for Ki.

§ Test-retest reproducibility of FDG-PET/CT uptake in cancer patients within a qualified and calibrated local network

We are continuing our evaluation of the multi-center test-retest studies. Figure: 5 shows representative lesions from a 60 year-old woman (Patient 03) with Stage IV invasive ductal breast carcinoma studied in the same scanner 8 days between FDG scans. The patient had 9 evaluable lesions. SUVmax ranged from 3.4-5.1 (average 4.0) in the first scan and 3.1-4.9 (average 4.2) in the second scan. Percentage difference ranged from -16% to +16% (average 3.9%) and the absolute SUV unit difference was -0.62 to +0.64 (average 0.15).

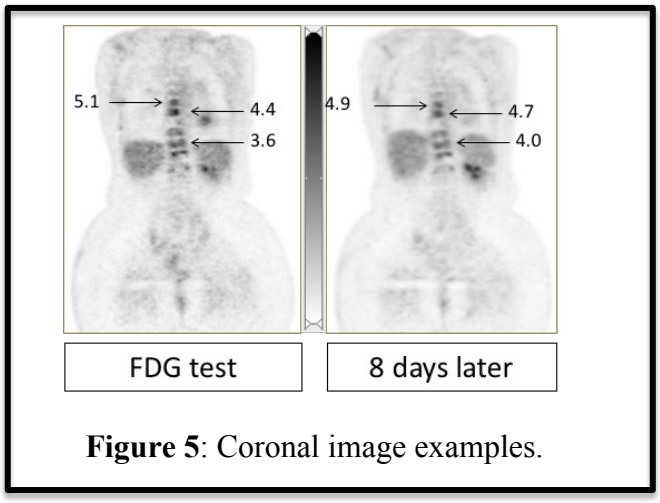


Figure 5: Coronal image examples.

Patient test/retest studies show that if PET/CT systems are carefully calibrated and monitored, and imaging protocols are consistent, then variability associated with FDG SUVmax between scans is similar to prior test/retest studies (manuscript submitted).

§ Integration of QIN tools into prospective clinical trials

We are deploying a set of the X-cal phantom kits 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.

COLLABORATIONS WITHIN THE NETWORK

We have collaborated with the QIN Network on the following projects:

1. Multicenter survey and publication of PET/CT protocol parameters that affect standardized uptake values (described above).
2. Multicenter measurement of temporal stability of PET SUV bias using long-lived sources (described above).
3. Multicenter data analysis challenge on the impact of arterial input function determination variations on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling (published).
4. Multi-site quality and variability analysis of 3D FDG PET segmentations based on phantom and clinical image data.

PLANS FOR NEXT YEAR

- Develop and implement a unified database and imaging platform for our phantoms and software tools.
- Develop PET study guidelines that incorporate instrument performance, patient variability, and protocol adherence into study design.
- Extend our biologically principled Virtual Clinical Trials tools developed for FDG to FLT (proliferation) and FES (receptor status) in multicenter studies.

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

1. 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

2. Rosen M, Kinahan PE, Gimpel JF, Opanowski A, Siegel BA, 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. Acad Radiol 24(2):232-245, 2017. PMID: 28395794. PMCID: PMC5389125.

3. Byrd DW, Doot RK, Allberg KC, MacDonald LR, McDougald WA, Elston BF, Linden HM, Kinahan PE. Evaluation of Cross-Calibrated (68)Ge/(68)Ga Phantoms for Assessing PET/CT Measurement Bias in Oncology Imaging for Single- and Multicenter Trials. Tomography 2(4):353-360, 2016. PMID: 28066807. PMCID: PMC5214172.

4. 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. Med Phys 44(2):479-496, 2017. PMID: 28205306.
5. Wangerin KA, Muzi M, Peterson LM, Linden HM, Novakova A, Mankoff DA, Kinahan PE. A virtual clinical trial comparing static versus dynamic PET imaging in measuring response to breast cancer therapy. Phys Med Biol 62(9):3639-3655, 2017. PMID: 28191877.

6. Scheuermann JS, Reddin JS, Opanowski A, Kinahan PE, Siegel BA, Shankar LK, Karp JS. Qualification of National Cancer Institute-Designated Cancer Centers for Quantitative PET/CT Imaging in Clinical Trials. J Nucl Med 58(7):1065-1071, 2017. PMID: 28254874.

U01CA190234: Tumor Genotype and Radiomic Phenotype of Lung Cancer

Harvard-Dana Farber Cancer Institute-Brigham & Women's Hospital

Joost van Griethuysen, M.D. Maastricht University
Thibaud Coroller, Ph.D. Dana Farber Cancer Institute
Fiona Fennessy, Ph.D. Brigham & Women's Hospital
Ahmed Hosny, Dana Farber Cancer Institute
John Quackenbush, Ph.D. Dana Farber Cancer Institute
Roman Zeleznik, Dana farber Cancer Institute
Stephen Yip, Ph.D. Dana Farber Cancer Institute
Andriy Fedorov, Ph.D. Brigham & Women's Hospital
Hugo Aerts, Ph.D. Dana Farber Cancer Institute

INTRODUCTION

There is overwhelming evidence that the initiation and progression of lung cancer are caused by specific genetic abnormalities, such as mutations in EGFR, KRAS, and ALK. Tumor tissues acquired from biopsies and surgical resection are used for genotyping but these procedures are invasive and are not generally repeated during treatment. The aim of this grant is to investigate if 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. To achieve these goals we aim to develop a radiomic system for the assessment of NSCLC tumors by non-invasive imaging, develop a rigorous statistical platform, develop radiomic and genomic biomarkers, and share computational resources to the community. This grant takes advantage of large cohorts of public and private non-small cell lung cancer (NSCLC) patients, using tumor samples for which we have both non-invasive imaging data (CT-PET) and mutational profiling data. One of these resources is the Profile study at the Dana-Farber Cancer Institute, a comprehensive personalized cancer medicine initiative generating mutational data of nearly all patients undergoing therapy.

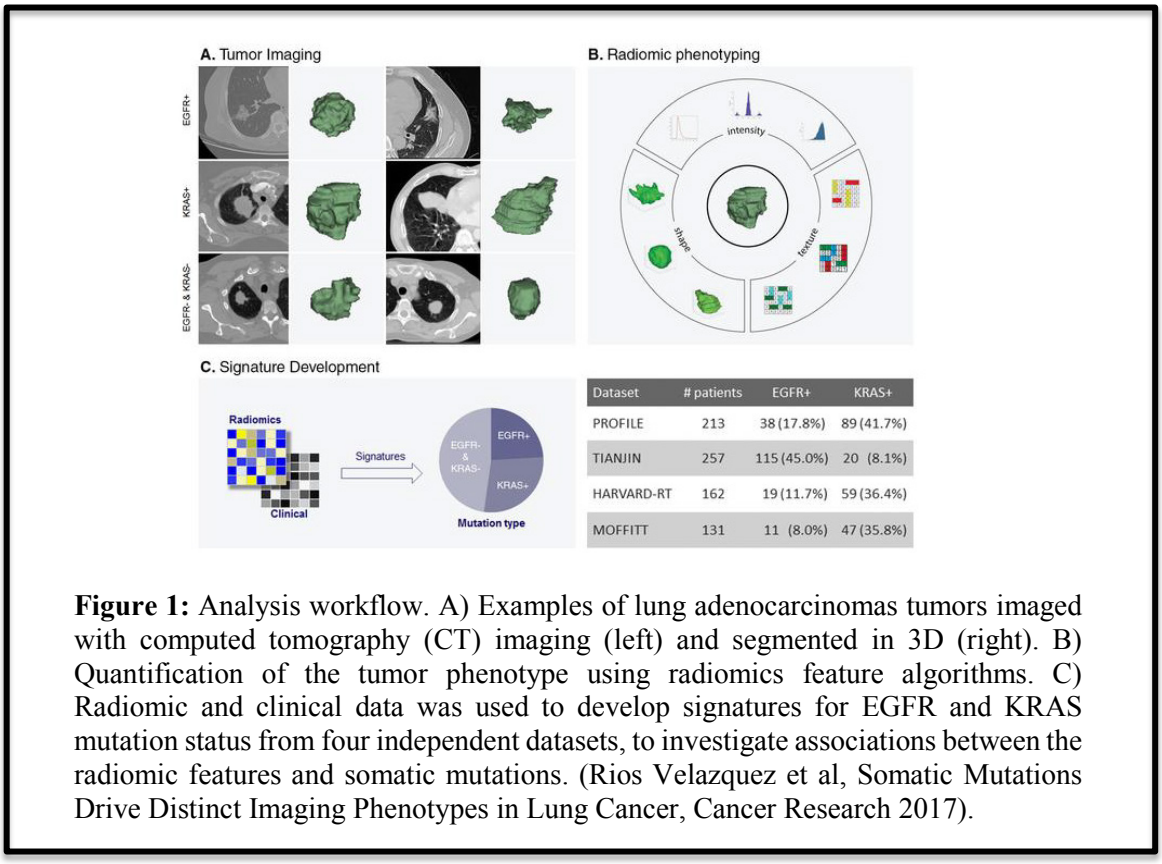
PROGRESS OVER THE PAST YEAR

The progress over the past year has been divided along the main aims. In specific, we are actively building a database of patients with imaging and mutational data, and developing platforms for the radiomic feature extraction.

A key study has been published in Cancer Research in 2017. As tumors are characterized by somatic mutations that drive biological processes ultimately reflected in tumor phenotype. With regard to radiographic phenotypes, generally unconnected through present understanding to the presence of specific mutations, artificial intelligence (AI)

methods can automatically quantify phenotypic characters by using predefined, engineered algorithms or automatic deep-learning methods, a process also known as radiomics. Here we demonstrate how imaging phenotypes can be connected to somatic mutations through an integrated analysis of independent datasets of 763 lung adenocarcinoma patients with somatic mutation testing and engineered computed tomography (CT) image analytics (Figure 1). We developed radiomic signatures capable of distinguishing between tumor genotypes in a discovery cohort (n=353) and verified them in an independent validation cohort (n=352). All radiomic signatures significantly outperformed conventional radiographic predictors (tumor volume and maximum diameter).

We found a radiomic signature related to radiographic heterogeneity that successfully discriminated between EGFR+ and EGFR- cases (AUC=0.69) (Figure 2). Combining this signature with a clinical model of EGFR status (AUC=0.70) significantly improved prediction accuracy (AUC=0.75). The highest performing signature was capable of distinguishing between EGFR+ and KRAS+ tumors (AUC=0.80) and, when combined with a clinical model (AUC=0.81), substantially improved its performance (AUC=0.86). A KRAS+/KRAS- radiomic signature also showed significant albeit lower performance (AUC=0.63) and did not improve accuracy of a clinical predictor of KRAS status. Our results argue that somatic mutations drive distinct radiographic phenotypes that can be predicted by radiomics. This work has implications for the use of imaging-based biomarkers in the clinic, as applied non-invasively, repeatedly and at low cost.



Dataset	# patients	EGFR+	KRAS+
PROFILE	213	38 (17.8%)	89 (41.7%)
TIANJIN	257	115 (45.0%)	20 (8.1%)
HARVARD-RT	162	19 (11.7%)	59 (36.4%)
MOFFITT	131	11 (8.0%)	47 (35.8%)

Figure 1: Analysis workflow. A) Examples of lung adenocarcinomas tumors imaged with computed tomography (CT) imaging (left) and segmented in 3D (right). B) Quantification of the tumor phenotype using radiomics feature algorithms. C) Radiomic and clinical data was used to develop signatures for EGFR and KRAS mutation status from four independent datasets, to investigate associations between the radiomic features and somatic mutations. (Rios Velazquez et al, Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer, Cancer Research 2017).

Another key analysis was published in eLife in 2017. While radiomics has been associated with several clinical endpoints, the complex relationships of radiomics, clinical factors, and tumor biology are largely unknown. To this end, we analyzed two independent cohorts of respectively 262 North American and 89 European patients with lung cancer, and consistently identified previously undescribed associations between radiomic imaging features, molecular pathways, and clinical factors. In particular, we found a relationship between imaging features, immune response, inflammation, and survival, which was further validated by immuno-histochemical staining (Figure 2). Moreover, a number of imaging features showed predictive value for specific pathways; for example, intra-tumor heterogeneity features predicted activity of RNA polymerase transcription (AUC = 0.62, p=0.03) and intensity dispersion was predictive of the autodegration pathway of a ubiquitin ligase (AUC = 0.69, p<10-4).

To test our hypothesis that radiomic data provide prognostic information complementary to clinical and genomic data, we built a clinical, genomic, and radiomic biomarkers to predict survival in lung cancer cases[38]. We tested previously published gene and radiomic biomarkers to predict for OS. In addition, we tested a radiomic biomarker that we recently published. Using independent datasets to train and validate the models, we observed that prognostic performance consistently increased with the addition of different data types. While the clinical model performed with a concordance index (CI) of 0.65 (Noether p=0.001), the combined radiomic-genomic-clinical model performed significantly higher (permutation test p=0.001) with a CI of 0.73 ($p=2x10^{-9}$). This radiomic-genomic-clinical model also performed significantly better than the combined clinical-radiomic model (p=0.007) and the clinical-genomic model (p=0.01). These results show the complementary value of clinical, radiomic, and genomic data for prediction of overall survival.

In conclusion, we demonstrate that radiomic approaches permit noninvasive assessment of both molecular and clinical characteristics of tumors, and therefore have the potential to advance clinical decision-making by systematically analyzing standard-of-care medical images. To enhance the scientific premise of current research, we address general strengths and weaknesses of previous results. As shown by previous studies, there have been associations found between features and somatic mutations. However, the overall sample sizes were low, and this is especially a concern as mutations in key genes (such as EGFR), occur in only a small subset of patients.

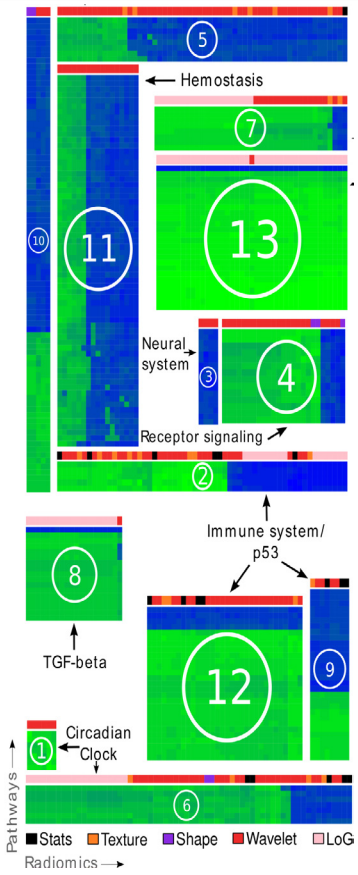


Figure 2: Radiomic pathway-clinical associations. Clustering of selected radiomic-pathway association modules. Normalized enrichment scores (NESs) have been biclustered to coherently expressed modules. Every heatmap in this figure corresponds to a module (M1 - M13) with radiomic features in columns and pathways in rows. Module sizes are proportional to heatmap sizes. Elements are NESs as Z-scores across features, and are displayed in blue when positive and green when negative. (Grossmann et al, Defining the biological basis of radiomic phenotypes in lung cancer, Elife. 2017)

COLLABORATIONS WITHIN THE NETWORK

We have been active within the network and have contributed to a number of community efforts. Specifically, we actively participated in the activities of the BIDS working group by contributing to the development of consensus on the process of tool catalog collection, and the development of the idea of a challenge. Following up on those efforts, we contributed the tools developed by our project to the QIN catalog, and initiated a challenge for PyRadiomics. We developed PyRadiomics, a flexible open-source platform capable of extracting a large panel of engineered features from medical images. PyRadiomics is implemented in Python and can be used standalone or using 3D-Slicer. Source code, documentation, and examples are publicly available at www.radiomics.io. With this platform, we aim to establish a reference standard for radiomic analyses, provide a tested and maintained resource, and to grow the community of radiomic developers addressing critical needs in cancer research. Within this CCP project, we evaluate the application of our radiomic informatics platforms at several QIN sites and assess the performance on a reference dataset as well as their own datasets. We will evaluate metrics, such as dissemination, training, evaluation, and reporting. The process of evaluating this tool is currently ongoing with the participation of sites inside and outside the QIN, including Stanford: lung cancer, Moffitt: lung cancer, PMH: H&N cancer, NKI: rectum cancer, and MD Anderson: lung cancer. Other interested QIN members are invited to participate as well.

PLANS FOR NEXT YEAR

During the next year, we plan to continue our research program, pursuing both the refinement of existing methods and the development of new datasets. Our results generated in the first years of our proposal, demonstrated strong associations between imaging phenotypes and somatic mutations, in an integrated analysis of several external and internal cohorts. However, to evaluate the association of less frequent mutations, even larger cohorts are needed. The next years will focus on generating and analyzing these. First, we will curate and analyze additional large datasets by collecting and analyzing, imaging, genomic, and clinical outcome data. We will build on important analyses that are recently published in Cancer Research 2017 (somatic mutation and radiomics) and in eLife 2017 (biological basis of imaging phenotypes). Second, we are extent our machine-learning framework to include novel methods for classification, such as deep learning. We will refine these models over time by including additional sources of data as they become available and as our methodologies are refined. Open source toolboxes are being developed and will be shared to the community. Third, we will further validate developed radiomic, genomic, and integrated biomarkers. For this purpose, we have access to multiple novel datasets and techniques to validate developed signatures. Several machine-learning techniques have been evaluated and will be applied to build novel predictors for mutational status as well as clinically relevant outcomes, such as overall survival, local control, and distant metastasis. Fourth, 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. While the focus of this project is on NSCLC patient data, 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 as we recognize that the data-generation landscape in oncology is rapidly evolving. Therefore, we aim to make

all software as independent as possible for disease site, imaging modality, and genomic data. Our radiomic systems are integrated within 3D-Slicer and shared with the public using a free installable application. 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.

RELEVANT PUBLICATIONS

1. Huynh E, Coroller TP, Narayan V, Agrawal V, Romano J, Franco I, Parmar C, Hou Y, Mak RH, Aerts HJ. "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 Jan 3;12(1):e0169172

2. Shaikh F, Kolowitz B, Awan O, Aerts HJ, von Reden A, Halabi S, Mohiuddin S, Shrestha R, Deible C, "Technical Challenges in the Clinical Application of Radiomics", accepted JCO-CCI

3. Vishesh Agrawal, Thibaud Coroller, Ying Hou, Stephanie Lee, John Romano, Aileen Chen, David Kozono, Elizabeth Baldini, Scott Swanson, David Jackman, Jon Wee, Aerts HJ, Ray Mak, "Tumor Volume Changes Following Chemoradiation Predict Pathologic Complete Response in Resectable Locally Advanced NSCLC", accepted Clinical Lung Cancer

4. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, Quesne JL, Moore DA, Veeriah S, Rosenthal R, Marafioti T, Kirkizlar E, Watkins TBK, McGranahan N, Ward S, Martinson L, Riley J, Fraioli F, Bakir MA, GrÖnroos E, Zambrana F, Endozo R, Bi WL, Fennessy FM, Sponer N, Johnson D, Laycock J, Shafi S, Czyzewska-Khan J, Rowan A, Chambers T, Matthews N, Turajlic S, Hiley C, Lee SM, Forster MD, Ahmad T, Falzon M, Borg E, Lawrence D, Hayward M, Kolvekar S, Panagiotopoulos N, Janes SM, Thakrar R, Ahmed A, Blackhall F, Summers Y, Hafez D, Naik A, Ganguly A, Kareht S, Shah R, Joseph L, Quinn AM, Crosbie P, Naidu B, Middleton G, Langman G, Trotter S, Nicolson M, Remmen H, Kerr K, Chetty M, Gomersall L, Fennell DA, Nakas A, Rathinam S, Anand G, Khan S, Russell P, Ezhil V, Ismail B, Irvin-Sellers M, Prakash V, Lester JF, Kornaszewska M, Attanoos R, Adams H, Davies H, Oukrif D, Akarca AU, Hartley JA, Lowe HL, Lock S, Iles N, Bell H, Ngai Y, Elgar G, Szallasi Z, Schwarz RF, Herrero J, Stewart A, Quezada SA, Van Loo P, Dive C, Lin CJ, Rabinowitz M, Aerts HJ, Hackshaw A, Shaw JA, Zimmermann BG; TRACERx consortium.; PEACE consortium., Swanton C." Phylogenetic ctDNA analysis depicts early stage lung cancer evolution.", Nature. 2017 Apr 26. doi: 10.1038/nature22364.

5. Grossmann P, Narayan V, Chang K, Rahman R, Abrey L, Reardon DA, Schwartz LH, Wen PY, Alexander BM, Huang R, Aerts HJWL., "Quantitative Imaging Biomarkers for Risk Stratification of Patients with Recurrent Glioblastoma Treated with Bevacizumab.", Neuro Oncol. 2017 May 11. doi: 10.1093/neuonc/nox092.

6. Agrawal V, Coroller TP, Hou Y, Lee SW, Romano JL, Baldini EH, Chen AB, Kozono D, Swanson SJ, Wee JO, Aerts HJWL, Mak RH., "Lymph node volume predicts survival but not nodal clearance in Stage IIIA-IIIB NSCLC.", PLoS One. 2017 Apr 20;12(4):e0174268

7. Rios-Velazquez E., Parmar C., Liu Y., Coroller T.P., Cruz G., Stringfield O., Ye Z., Makrigiorgos M., Fennessy F., Mak R.H., Gillies R., Quackenbush J., Aerts HJ, "Somatic mutations drive distinct imaging phenotypes in lung cancer", Cancer Res. 2017 Jul 15;77(14):3922-3930.

8. Patrick Grossmann, Olya Grove, Nehme El-Hachem, Emmanuel Rios-Velazquez, Chintan Parmar, Ralph T.H. Leijenaar, Benjamin Haibe-Kains, Philippe Lambin, Robert J. Gillies, Aerts HJ, "Identification of Molecular Phenotypes in Lung Cancer by Integrating Radiomics and Genomics", Elife. 2017 Jul 21;6. pii: e23421.

9. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmer C, Bakers FCH, Peters NHGM, Beets-Tan RGH, Aerts HJWL. "Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR.", Sci Rep. 2017 Jul 13;7(1):5301.
10. Yip SSF, Parmar C, Blezek D, Estepar RSJ, Pieper S, Kim J, Aerts HJWL."Application of the 3D slicer chest imaging platform segmentation algorithm for large lung nodule delineation.", PLoS One. 2017 Jun 8;12(6):e0178944.

11. Yip SSF, Liu Y, Parmar C, Li Q, Liu S, Qu F, Ye Z, Gillies RJ, Aerts HJWL." Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer.", Sci Rep. 2017 Jun 14;7(1):3519.

12. van Griethuysen J.J.M., Fedorov A., Parmar C., Hosny A., Aucoin N., Narayan V., Beets-Tan R.G.H., Fillion-Robin J.C., Pieper S., Aerts HJWL, "Computational Radiomics System to Decode the Radiographic Phenotype", Cancer Research 2017 Nov 1;77(21):e104-e107

13. Aerts HJWL,Data Science in Radiology: A Path Forward.. Clin Cancer Res. 2017 Nov 2. doi: 10.1158/1078-0432.CCR-17-2804.

14. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, Birkbak NJ, Veeriah S, Van Loo P, Herrero J, Swanton C; TRACERx Consortium. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution., Cell. 2017 Nov 30;171(6):1259-1271.e11. doi: 10.1016/j.cell.2017.10.001.

15. Coroller TP, Bi WL, Huynh E, Abedalthagafi M, Aizer AA, Greenwald NF, Parmar C, Narayan V, Wu WW, Miranda de Moura S, Gupta S, Beroukhim R, Wen PY, Al-Mefty O, Dunn IF, Santagata S, Alexander BM, Huang RY, Aerts HJWL. Radiographic prediction of meningioma grade by semantic and radiomic features. PLoS One. 2017 Nov 16;12(11):e0187908.

16. Impact of experimental design on PET radiomics in predicting somatic mutation status. Yip SSF, Parmar C, Kim J, Huynh E, Mak RH, Aerts HJWL., Eur J Radiol. 2017 Dec;97:8-15. doi: 10.1016/j.ejrad.2017.10.009.

17. Balagurunathan Y, Beers A, Kalpathy-Cramer J, McNitt-Gray M, Hadjiiski L, Zhao B, Zhu J, Yang H, Yip SSF, Aerts HJWL, Napel S, Cherezov D, Cha K, Chan HP, Flores C, Garcia A, Gillies R, Goldgof D. Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST data., Med Phys. 2018 Jan 24. doi: 10.1002/mp.12766.

18. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, Khaouam N, Nguyen-Tan PF, Wang CS, Sultanem K, Seuntjens J, El Naqa I.Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep. 2017 Aug 31;7(1):10117. doi: 10.1038/s41598-017-10371-5.

19. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmar C, Bakers FCH, Peters NHGM, Beets-Tan RGH, Aerts HJWL.Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR. Sci Rep. 2017 Jul 13;7(1):5301. doi: 10.1038/s41598-017-05728-9.

20. Agrawal V, Coroller TP, Hou Y, Lee SW, Romano JL, Baldini EH, Chen AB, Kozono D, Swanson SJ, Wee JO, Aerts HJWL, Mak RH.Lymph node volume predicts survival but not nodal clearance in Stage IIIA-IIIB NSCLC. PLoS One. 2017 Apr 20;12(4):e0174268.

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

Columbia University

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

INTRODUCTION

Methodology used to assess tumor response to a given therapy is critical to success of the treatment. Ideally, response to therapy would be determined with the highest accuracy as early as possible, so that a lack of response would prompt a quick change of the inefficient, toxic treatment. Conventional methods utilizing tumor diameter and the unproven "arbitrarily" determined response cut-off thresholds to estimate change in tumor burden are outdated and may delay detection of tumor progression or underestimate tumor response to therapy. 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 (i.e., 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.

PROGRESS OVER THE PAST YEAR

§ A statement of progress made towards the achievement of originally stated aims

Our project demonstrated the feasibility and value of the volumetric CT as a new quantitative imaging biomarker for early and more accurate assessment of tumor responses to therapies especially novel targeted therapies in clinical trials.

A list of major activities, significant results, and key outcomes

- We developed three robust computer algorithms for semi-automated segmentation of tumors in the lungs (REF #1), liver (REF #2) and lymph nodes (REF #3; minor revision under review), the three most common sites of cancer metastases. These three algorithms will allow total tumor burden to be efficiently and reproducibly measured for response assessment in clinical trials evaluating new therapies.

- Using the three segmentation algorithms, we explored the variability in measuring total tumor burden in a metastasis setting on CT scan images and revealed the low variation magnitudes of 22%, 19%, and 11% for the volumetric, bi-dimensional, and uni-dimensional measurements, respectively (Ref #4).

- Based on an open source, Weasis, we developed a client-based user-friendly response assessment platform that has integrated the three segmentation algorithms and an efficient editing tool (Ref #5). We used this system to complete the tumor measurements for the two proposed clinical trial studies, SARC 011 and HCC CALGB 80802. Since it was developed, the Weasis-based platform has been used in numerous studies to help validate the volumetric technique as a better technique for tumor response assessment including the FNIH-sponsored VOL-PACT study (Ref #6).

- We completed the data analysis of SARC 011 (A phase II trial of R1507, a recombinant human monoclonal antibody to the insulin-like growth factor-1 receptor for the treatment of patients with recurrent or refractory Ewing's sarcoma). In this study, we analyzed 101 sarcoma patients (303 scan time points). We compared prediction power of the response assessment metrics of volume, bi- and uni- to the overall survival (OS) and found that the volumetric technique was superior to WHO and RECIST methods in identifying tumor response. Our result, along with the PET part of this study, was reported at ASCO 2015 annual meeting and published in JCO (Ref #7).

- We completed the tumor measurement for HCC CALGB 80802 (A Phase II/III randomized study of Sorafenib plus Doxorubicin versus sorafenib in patients with advanced hepatocellular carcinoma). In this study, we measured target tumors in 207 HCC patients (683 scan time points). When investigating the density-based response assessment metrics, we noticed considerable differences in the measured density values due to varying timing of contrast-enhanced CT acquisition. This was a not yet studied area, though tumor density change has been suggested to add a new dimension for improved response assessment in the era of targeted therapies. We thus decided to first explore the impact of the density variability due to imperfect portal venous phase (PVP) acquisition timing in response assessment. Using a machine learning approach, we developed a computer-aided, semi-automated quality control algorithm to gauge scans' timing (early/optimal/late PVP) based on the acquired images and found that density of lesions metastatic to liver could be decreased by 16.7% at early-PVP and 12.6% at late-PVP (Ref #8). Latterly, we developed a fully-automated QC program for CT scan timing using artificial intelligence technology.

- We actively participated in the four CT and PET segmentation and characterization challenges within the QIN. These challenges were designed to explore possible variability existed in measuring tumors and tumor changes. All of the four challenge results are published (Refs #9 – #12).

- As one of the four QIN teams (the other three sites: MGH (PI site), Stanford, and Moffitt), we jointly applied and received a U24 grant entitled "Informatics Tools for Optimized Imaging Biomarkers for Cancer Research & Discovery" (U24 CA180927). The purpose of this new grant is to widely support the development and validation of quantitative imaging biomarkers for the oncology community.

- We extended our volume- and necrosis-based imaging metrics to more complex radiomic biomarkers and, besides the clinical studies (Refs #13, #14), we intensively explored the reproducibility and variability in radiomic features and biomarkers, an important area in the rapidly emerging field of radiomics (Refs #15 – #18).
- Our other activities at QIN include:

  Dr. Schwartz (contact PI) served as the Chair for QIN Executive Committee (EC) and organized monthly EC t-cons between May 2016 – April 2017.
  Dr. Zhao (Co-PI) served as Co-chair for QIN Image Analysis & Performance Metrics Working Group between May 2011 – April 2013 (two terms).
  We contributed to the QIN joint publication on Quantitative Imaging in Cancer Clinical Trials (Ref #19)
  We contributed multiple times to QIN Specific Issues articles (Refs #5, #10, #15, #18)

A LIST OF PUBLICATIONS RESULTING FROM THE PROJECT

1. 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. PubMed PMID: 23556926; PubMed Central PMCID: PMC3618093.

2. 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. **(Article chosen as 2015 Editor's Picks)**. PubMed PMID: 26520721; PubMed Central PMCID: PMC4600084.

3. Tan Y, Lu L, Bonde A, Wang D, Qi J, Schwartz LH and Zhao B. Lymph node segmentation by dynamic programming and active contours. Med Phys. Minor revision submitted.

4. 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. **(Article featured in Highlights of This Issue).** PubMed PMID: 24780294; PubMed Central PMCID: PMC4337392.

5. 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. QIN Special Issue.

6. Dercle L, Connors DE, Tang Y, Adam SJ, Gönen M, Hilden P, Karovic S, Maitland M, Moskowitz CS, Kelloff G, Zhao B, Oxnard GR, Schwartz LH. Vol-PACT: An FNIH public-private partnership supporting sharing of clinical trial data for development of improved imaging biomarkers in oncology. JCO Clinical Cancer Informatics (in press).

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. PubMed PMID: 27573658; PubMed Central PMCID: PMC5065114.

8. Dercle L, Lu L, Lichtenstein P, Yang H, Wang D, Zhu J, Wu F, Piessevaux H, Schwartz HL, Zhao B. Impact of Variability in Portal Venous Phase Acquisition Timing in Tumor Density Measurement and Treatment Response Assessment: Metastatic Colorectal Cancer as a Paradigm. JCO Clinical Cancer Informatics (in press)

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. PubMed PMID: 26847203; PubMed Central PMCID: PMC4942386.

10. 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. QIN Special Issue. PubMed PMID: 28149958; PubMed Central PMCID: PMC5279995.

11. 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. 2017; 44(2):479-496. **(Article chosen as 2017 Editor's Picks)**. PMID not yet available

12. Balagurunathan Y, Beers A, Cramer JK, McNitt-Gray M, Hadjiiski L, Zhao B, Zhu J, Yang H, Yip SSF, Aerts HJWL, Napel S, Cherezov D, Cha K, Chan H, Flores C, Garcia A, Gillies R, Goldgof D. Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST data, Med Phys (accepted).

13. Lee HJ, Kim YT, Kang CH, Zhao B, Tan Y, Schwartz LH, Persigehl T, Jeon YK and Chung DH, EGFR Mutation in Lung Adenocarcinomas: relationship with CT Characteristics and Histologic Subtypes. Radiology 2013; 268:254-264. PubMed PMID: 23468578

14. Aerts H, Grossmann P, Tan Y, Oxnard G, Schwartz LH, Zhao B. Defining a Radiomic Response Phenotype: A Pilot Study using TKI therapy in NSCLC. Nat Sci Rep 6; 33860, 2016. PubMed PMID: 27645803; PubMed Central PMCID: PMC5028716.

15. 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. QIN Special Issue. PubMed PMID: 24772211; PubMed Central PMCID: PMC4000020.

16. 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. PubMed PMID: 27009765; PubMed Central PMCID: PMC4806325.

17. Lu L, Ehmke R, Schwartz LH, Zhao B. Assessing Agreement between Radiomic Features Computed for Multiple Imaging Settings. PLoS One. 2016 Dec 29;11(12):e0166550. PubMed PMID: 28033372; PubMed Central PMCID: PMC5199063.

18. Huang Q, Lu L, Dercle L, Lichtenstein P, Li Y, Yin Q, Zong M, Schwartz HL, Zhao B. Inter-observer variability in tumor contouring affects the use of Radiomics to predict mutational status. Journal of Medical Imaging. QIN Special Issue to honor Dr. Larry Clarke (in press).

19. 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. PubMed PMID: 26773162; PubMed Central PMCID: PMC4717912.

U01CA187013-05- Resources for Development and Validation of Radiomic Analyses & Adaptive Therapy

University of Arkansas for Medical Sciences and Emory University

Fred Prior, Ph.D. University of Arkansas
Ashish Sharma, Ph.D. Emory University

INTRODUCTION

Imaging data and in particular quantitative features extracted by image analysis have been identified as a critical source of information particularly for cohort classification (imaging phenotypes) and tracking response to therapy (1-3). Radiomics and Pathomics, where quantitative features are extracted from Radiology and Pathology imaging studies, provide valuable diagnostic and prognostic indicators of cancer (4-11) (12). For example, Aerts and Gillies (6-8, 13-16) have shown that Radiology image features can be linked to patient outcomes, Pathologist generated classifications and genomic signatures. Such methodologies require large collections of well-curated data for development, validation and to ensure research reproducibility. (17-19)

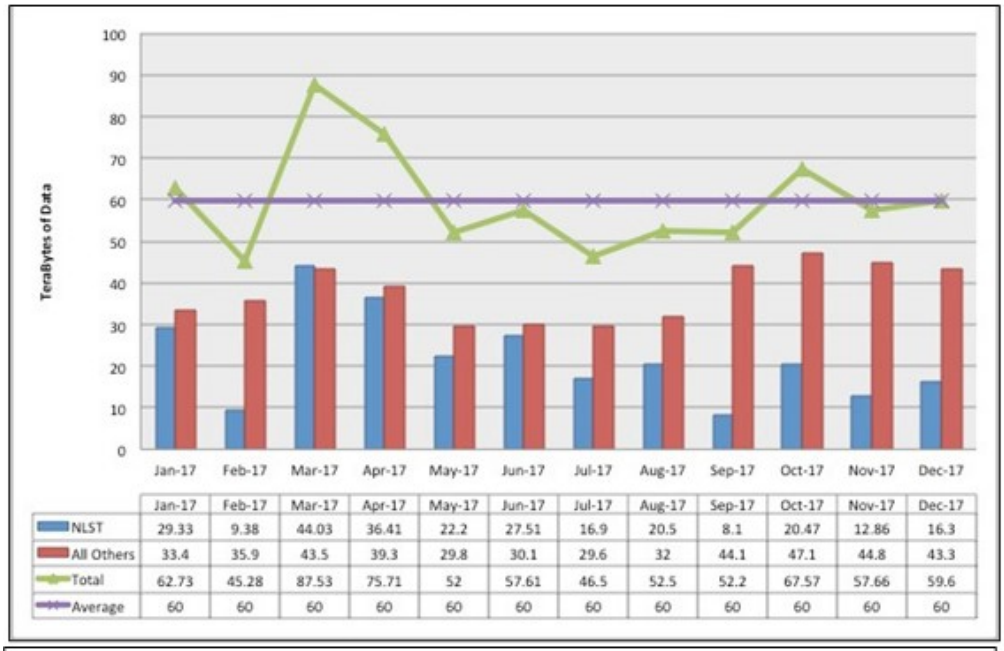
The Cancer Imaging Archive (TCIA) continues to be NCI's primary resource for acquiring, curating, managing and distributing images and related data to support Cancer Research and the primary image repository for the Quantitative Imaging Network (20). Integrative imaging studies enable a highly data- driven approach to diagnosis and outcome prediction (21), and are a key component of precision medicine. Locating and accessing datasets with the relevant information, is frequently cited as one of the major hurdles to such Integrative Imaging Studies. It is therefore no surprise that TCIA has been the image-source of many such studies (13, 22-25) Tools and procedure developed by our QIN team that enhance TCIA capabilities enable new innovations for the cancer imaging research community.

PROGRESS OVER THE PAST YEAR

The Cancer Imaging Archive provides both data to support our investigations in radiomics and an informatics resource to which we add capabilities in support of QIN and the broader user community. During 2017, 80,946 users from 166 countries visited TCIA. As of December 31, 2017 the registered user community was 9258 and TCIA data had been used in 553 peer reviewed publications and graduate theses. TCIA was recognized by the Cancer Moonshot Blue Ribbon Panel's Enhanced Data Sharing Working Group as an example of the type of multimodal data repositories that are needed to develop a Cancer Data Ecosystem (26). Figure 1 summarizes the volume of data downloaded from TCIA during each month of 2017.

Expert curation and quality control of incoming data sets significantly contribute to the success of TCIA. Curation workflows (see Figure 2) for radiology and pathology

images and radiation therapy data have been continuously improved during the past year in large part due to the use and extension of the Posda open source toolkit (27, 28). To support acquisition of data from the Veteran's Administration's Precision Oncology Pilot as a contribution to the Cancer Moon Shot Apollo program, Posda tools and TCIA's variant of the Clinical Trial Processor (CTP) de-identification and secure transport package, have been approved for use on the VA's secure internal network. Fully de-identified VA data is now flowing into TCIA. A publication on Posda and curation workflow was submitted in 2017 and is still in review (Bennett et al., Reengineering Workflow for Curation of DICOM Datasets, Journal of Digital Imaging, 2018, in review.)



	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17
NLST	29.33	9.38	44.03	36.41	22.2	27.51	16.9	20.5	8.1	20.47	12.86	16.3
All Others	33.4	35.9	43.5	39.3	29.8	30.1	29.6	32	44.1	47.1	44.8	43.3
Total	62.73	45.28	87.53	75.71	52	57.61	46.5	52.5	52.2	67.57	57.66	59.6
Average	60	60	60	60	60	60	60	60	60	60	60	60

Figure 1: TCIA Data Distribution to Researchers in 2017. Users download data from TCIA via the web portal, API and removable media distribution (for large downloads most frequently from NLST). The National Lung Screening Trial (NLST) data set is the largest managed by TCIA (21,082,265 images for 26,254 subjects). A total of **716.9 TB** of data were distributed from TCIA during 2017.

The TCIA infrastructure has been enhanced by extending the application programming interface (API) to better support data mashups. We have deployed a visual query and analytics environment, DataScope (Figure 3) that gives users an interactive environment to create scientific mashups and graphically explore TCIA image data and associated clinical data. DataScope utilizes the TCIA API and has been extended through mashups of TCIA image metadata with clinical data from The Cancer Genome Atlas (TCGA) and Clinical Proteomic Tumor Analysis Consortium (CPTAC) collections.

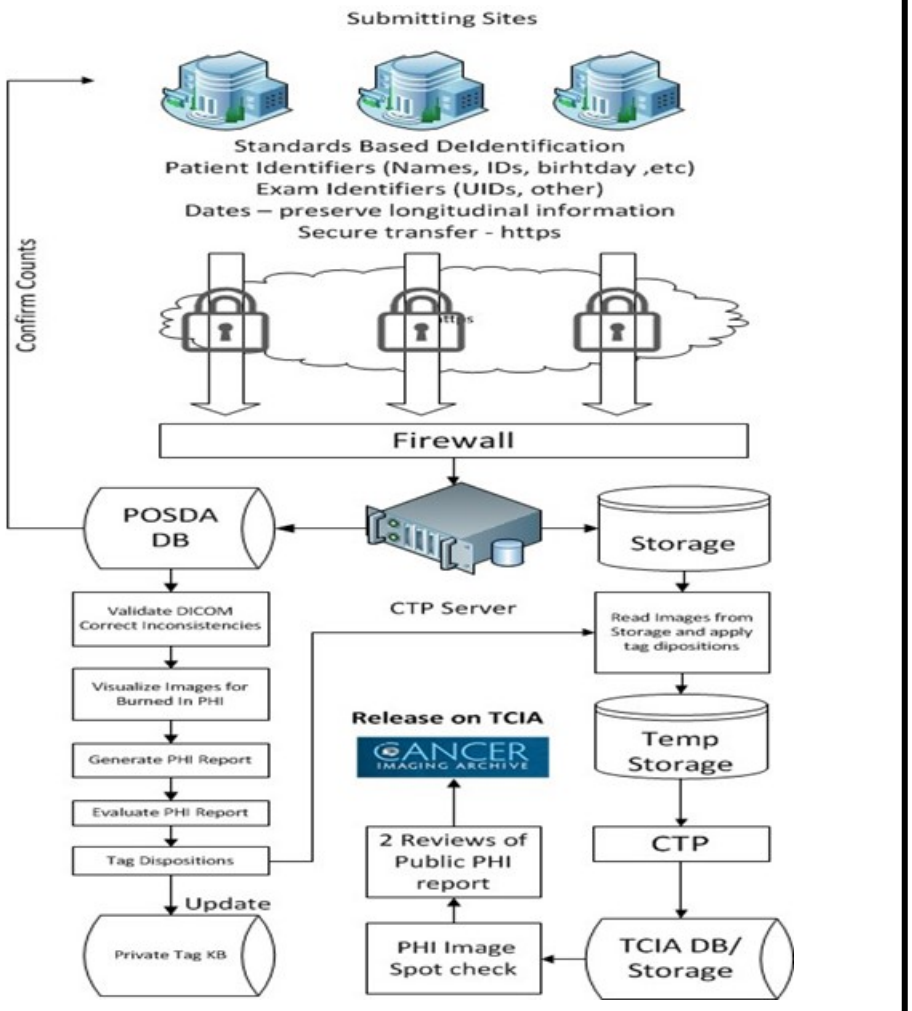


Figure 2: TCIA Curation Workflow. Posda has been extended and used to increase the automation and thereby the throughput of TCIA data curation including pathology data objects.

Work toward a QIN Portal that supports radiomics pipelines on both high performance and cloud computing platforms continues on two fronts. The Galaxy web-based scientific workflow and data integration package (29) continues to be a productive platform into which we have added custom processing tools for radiomic analysis and data visualization. We have integrated our lung nodule segmentation pipeline into Galaxy and liked it to the visualization tools we incorporated during the prior reporting period (Figure 4). In parallel we have expanded our efforts to containerize radiomics pipelines, now using the singularity (30) container technology. UAMS has deployed a large-scale high-performance computing (HPC) environment providing our team access to both the UAMS and Washington University HPC environments as well as the Amazon and Google clouds for testing and performance analysis. Thus, we incorporate containerized pipelines into Galaxy and explore running them independently as well. Our goal remains to develop a

radiomics portal for QIN researchers that allows execution of QIN developed radiomics tools on TCIA image data without an explicit download of data.

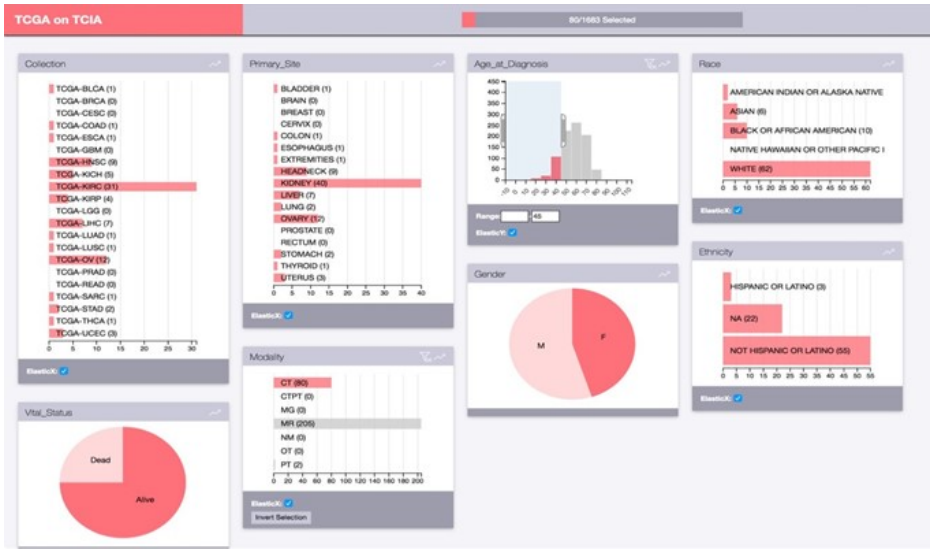


Figure 3: DataScope provides an interactive exploration of TCGA images, on TCIA, using a combination of image metadata and clinical data.

Figure 4: QIN Portal is based on Galaxy's plugin capability. We have implemented a screening radiomics pipeline and linked it to the Papaya visualization tool we previously incorporated into Galaxy. This figure illustrates the output of the lung nodule segmentation stage of the pipeline.

Work continues on radiomic analysis of lung CT images. Our attention has shifted from graph theoretic approaches to machine learning (random forests, support vector machines) and most recently on deep learning (convolutional neural networks). A publication on our work applying radiomics and deep learning to enhance lung screening was submitted in 2017 and is still in review (Causey et al., Highly accurate model for prediction of lung nodule malignancy with CT scans, Nature Scientific Reports, 2018, in review, available as a preprint on arXiv).

Although prohibited from competing in the Data Science Bowl challenge competition because much of the data was provided by TCIA, our team was allowed to participate. Our algorithm, DeepScreener, ranked 16th (out of 1972 teams) in the competition. The algorithm is based on a novel deep learning approach and does not need lung nodule annotations to conduct cancer prediction. DeepScreener takes as input complete lung CT image sets and can predict a patient's cancer status with an AUC of 0.885 an AUPRC of 0.866.

We believe from the challenge and subsequent analysis of NLST data not used in the competition (manuscript in preparation) that the deep learning approach has the potential to reach to a performance comparable to human experts for lung cancer screening with low-dose CT and could have great clinical impact.

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 2017: Course number RCB54 (*Using Publicly Accessible 'Big Data' from the NIH/NCI's Cancer Imaging Archive (TCIA) to Research Quantitative Radiomics, Proteomics, Genetics and Pathology*). In addition, Dr. Prior participated in RSNA Course RCC45A where he presented a talk based on our radiomics research (*Computer* Science 'Deep Learning' Research by the Academic Community). The complete list of publications, presentations and training courses is presented in Table 1.

The imphub open source community support and software development environment continues to support TCIA operations, Posda and API development, and new projects being developed as part of our team's participation in ITCR.

COLLABORATIONS WITHIN THE NETWORK

During the past year Dr. Sharma chaired the BIDS working group with Dr. Prior participating as a member. In conjunction with BIDS we have explored containerizing community pipelines and improving data curation and accessibility. One or both PIs participate in the QIN Executive Committee calls.

§Collaborations Outside the Network

Dr. Sharma collaborated with the ISB Cancer Genomics Cloud Pilot and Amazon

to host TCIA collections in the cloud pilot. The TCIA team is discussing further cloud hosting options with Google.

Dr. Bosch is an active member of DICOM WG-7 (RT Information Objects) and Dr. Tarbox is an active member of several DICOM working groups. UAMS is currently the only academic institution to have full voting membership in DICOM.

Dr. Bosch continues to serve as Connectathon Test Manager for the IHE Radiation Oncology (IHE- RO) Domain, and in AAPM TG-263, which seeks to standardize nomenclature for radiotherapy treatment planning.

Dr. Bosch (lead PI) and Dr. Prior continue to 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. TCIA is currently collecting data from this trial.

Dr. Prior and Dr. Bosch have been working with PCORI on an Open Science Pilot project to help PCORI and their contractors to understand how to acquire and manage data from PCORI funded clinical trials, drawing on our TCIA experience and use of TCIA and QIN developed curation tools and processes.

In September 2017 the UAMS-Emory QIN team in collaboration with Dr. Saltz's team from Stony Brook, joined NCI's Informatics Technology for Cancer Research (ITCR) network. The new ITCR award, on which Dr. Prior, Dr. Sharma and Dr. Saltz are co-PIs, focuses on the evolution of the TCIA technology stack into a new framework called PRISM. PRISM will incorporate tools and technologies from this QIN, Dr. Saltz' ITCR (QuIP) and other funded projects as illustrated in Figure 5. This offers many new opportunities for cross-network collaboration particularly between QIN and ITCR researchers building on TCIA.

The UAMS-Emory QIN team collaborated with a group of QIN and ITCR investigators to conduct the Crowds Cure Cancer event at RSNA 2017. Radiologists analyzed a large selection of TCIA data to identify malignant lesions and make quantitative measurements. This large sample allows statistical models to be used to define the most probable measurements to be used as truth for future research.

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.

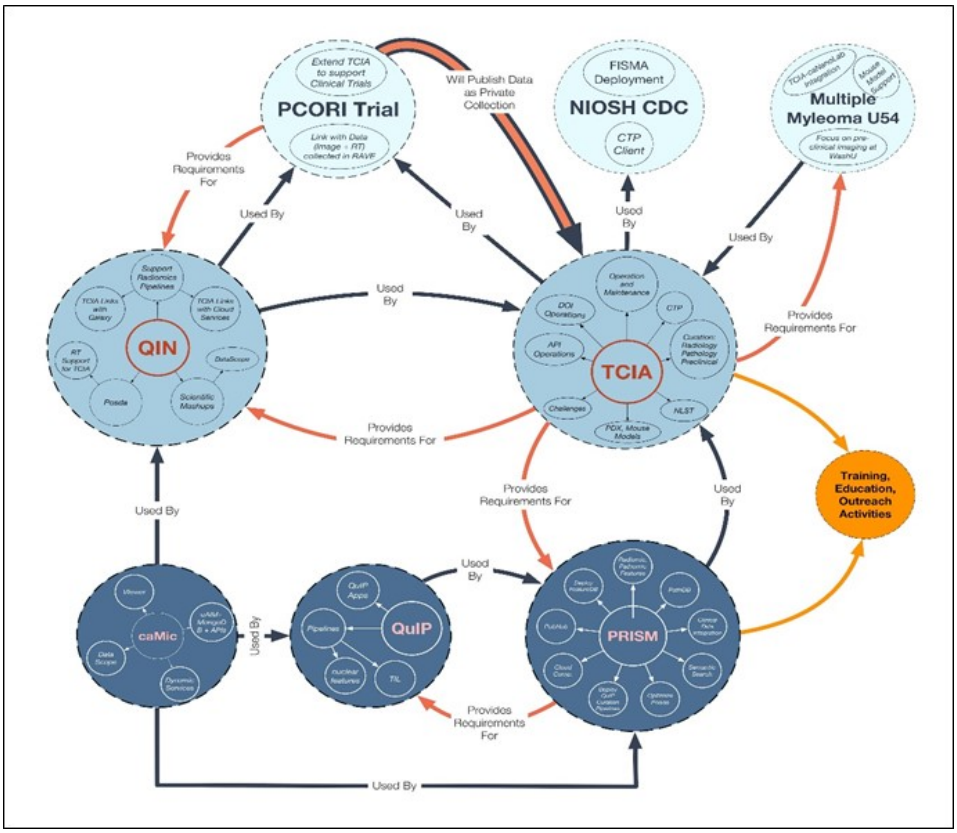
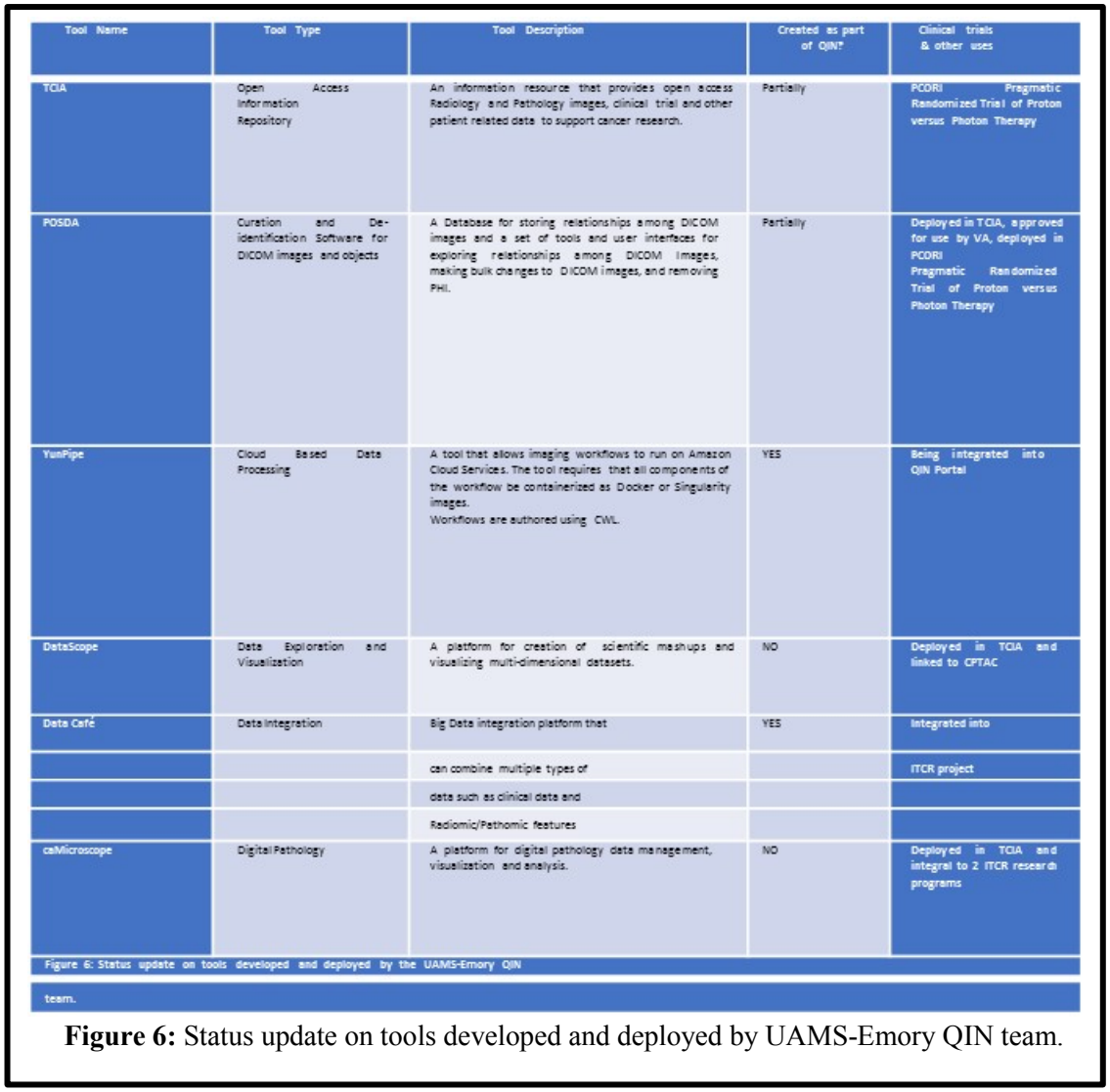


Figure 5: TCIA forms the basis of a cross-linked set of research initiatives all stemming from the UAMS-Emory QIN team's research initiatives.



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, approved for use by VA, deployed in PCORI Pragmatic Randomized Trial of Proton versus Photon Therapy
YunPipe	Cloud Based Data Processing	A tool that allows imaging workflows to run on Amazon Cloud Services. The tool requires that all components of the workflow be containerized as Docker or Singularity images. Workflows are authored using CWL.	YES	Being integrated into QIN Portal
DataScope	Data Exploration and Visualization	A platform for creation of scientific mashups and visualizing multi-dimensional datasets.	NO	Deployed in TCIA and linked to CPTAC
Data Café	Data Integration	Big Data integration platform that	YES	Integrated into
		can combine multiple types of		ITCR project
		data such as clinical data and		
		Radiomic/Pathomic features		
caMicroscope	Digital Pathology	A platform for digital pathology data management, visualization and analysis.	NO	Deployed in TCIA and integral to 2 ITCR research programs

Figure 6: Status update on tools developed and deployed by the UAMS-Emory QIN team.

Figure 6: Status update on tools developed and deployed by UAMS-Emory QIN team.

Chennubhotla C, Clarke LP, Fedorov A, Foran D, Harris G, Helton E, Nordstrom R, Prior F, Rubin D, Saltz JH, Shalley E. **An Assessment of Imaging Informatics for Precision Medicine in Cancer**. Yearbook of Medical Informatics. 2017:26(01): 110-119.
Prior F, Smith K, Sharma A, Kirby J, Tarbox L, Clark K, Bennett W, Nolan T, Freymann J, **Cancer Imaging Data – the Public Collections of The Cancer Imaging Archive**. Nature Scientific Data, 2017:4; doi:10.1038/sdata.2017.124
Kathiravelu P, Chen Y, Sharma A, Galhardas H, Van Roy P, Veiga L. **On-Demand Service-Based Big Data Integration: Optimized for Research Collaboration.** InVLDB Workshop on Data Management and Analytics for Medicine and Healthcare 2017 Sep 1 (pp. 9-28). Springer, Cham.
Post AR, Ai M, Pai AK, Overcash M, Stephens DS. **Architecting the Data Loading Process for an i2b2 Research Data Warehouse: Full Reload versus Incremental Updating**. AMIA Annu Symp Proc 2017
Iyer, G. R., Duttaduwarah, S., & Sharma, A. (2018). **DataScope: Interactive visual exploratory dashboards for large multidimensional data**. Presented at IEEE Visual Analytics in Healthcare Workshop 2017. (No. e26441v1). PeerJ Preprints.
Causey J, Zhang J, Ma S, Jiang B, Qualls J, Politte DG, Prior F, Zhang S, Huang X, "**Highly accurate model for prediction of lung nodule malignancy with CT scans**", Abstract and poster presentation, Quantitative Imaging Network (QIN) Annual Meeting, Rockville, MD, April 10, 2017.
Post A, Sharma A, Prior F, "**Eureka! Clinical Analytics**," Abstract and poster presentation, Quantitative Imaging Network (QIN) Annual Meeting, Rockville, MD, April
Prior F, Tobias M, Nolan T, Moore S, "**Data Acquisition and Management for Nanotherapeutics**", Abstract and poster presentation, 2017 NCI Alliance for Nanotechnology in Cancer, Bethesda, MD, October 3-5, 2017.
Nolan T, Kirby J, Prior F, "**Data Management, Sharing, and Citation Strategies in The Cancer Imaging Archive (TCIA),"** Abstract and poster presentation, AMIA Annual Symposium, Washington DC, November, 2017.
"**Imaging Informatics and PMI**," Presented by F.W. Prior, Fourth Catholic University International Symposium on Medical Informatics, Seoul, South Korea, September 9, 2016
"**Overview of biomedical imaging and precision medicine, shared informatics challenges**," Presented by F.W. Prior, Keynote lecture, Quantitative Imaging and Imaging Informatics in the Era of Precision Medicine pre-Symposium, AMIA 2016, Chicago, Il, November 13, 2016.
"**Radiomics and Imaging Informatics to Track Response to Cancer Therapy**," Presented by F.W. Prior, MCBIOS XIV Session Featured Speaker, Little Rock, AR, March 24, 2017

Table 1: Publications and Presentations

PLANS FOR NEXT YEAR

During the coming year we plan to complete the following work on each of our aims.

§ AIM 1

Complete the production deployment of DataScope as a new capability of 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 has been submitted.

We will continue to accumulate data from the PCORI funded RT clinical trial and refine our curation processes accordingly. Posda tools are continuously evolving to meet new TCIA curation demands and to improve automation.

§ 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 radiomics pipelines and their integration with the other Galaxy tools that we are developing. We will also integrate tools for doing automated lung nodule characterization based on our preliminary experience with CNNs. A draft publication on this work is nearing completion.

We plan to continue development of CNN based radiomic analysis of the NLST data set to learn how to identify patients who developed lung cancer from those who did not using the entire lung as input to the analysis. This work is based on our successful participation in the Data Science Challenge.

§ AIM 3

We will focus our efforts on advanced machine learning based imaging phenotypes that combine clinical data and imaging features. 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.

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.

§ AIM 4

A short course on the advanced features added to TCIA has been proposed for presentation at RSNA 2018. Presentations are planned for the 2018 QIN annual meeting. Dr. Prior has been invited to give two lectures on big data in cancer research and the use of radiomics in lung cancer screening at the 2018 meeting of the International Cancer Imaging Society in Menton, France. Both talks are based on QIN funded research.

LIST OF REFERENCES

1. Thrall JH. Personalized Medicine. Radiology. 2004;231(3):613-6. doi: doi:10.1148/radiol.2313040323. PubMed PMID: 15163802.

2. Thrall JH. Trends and Developments Shaping the Future of Diagnostic Medical Imaging: 2015 Annual Oration in Diagnostic Radiology. Radiology. 2016;279(3):660-6. doi: doi:10.1148/radiol.2016160293. PubMed PMID: 27183401.

3. Herold CJ, Lewin JS, Wibmer AG, Thrall JH, Krestin GP, Dixon AK, Schoenberg SO, Geckle RJ, Muellner A, Hricak H. Imaging in the Age of Precision Medicine: Summary of the Proceedings of the 10th Biannual Symposium of the International Society for Strategic Studies in Radiology. Radiology. 2015:150709.

4. Cooper L, Kong J, Gutman D, Wang F, Cholleti S, Pan T, Widener P, Sharma A, Mikkelsen T, Flanders A, Rubin D, Van Meir E, Kurc T, Moreno C, Brat D, Saltz J. An Integrative Approach for In Silico Glioma Research. IEEE Transactions on Biomedical Engineering Letters. 2010;57(10):2617-21.

5. Cooper LA, Kong J, Gutman DA, Wang F, Gao J, Appin C, Cholleti S, Pan T, Sharma A, Scarpace L, Mikkelsen T, Kurc T, Moreno CS, Brat DJ, Saltz JH. Integrated morphologic analysis for the identification and characterization of disease subtypes. Journal of the American Medical Informatics Association : JAMIA. 2012;19(2):317-23. Epub 2012/01/27. doi: 10.1136/amiajnl-2011-000700. PubMed PMID: 22278382; PubMed Central PMCID: PMCPmc3277636.

6. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications. 2014;5:4006. Epub 2014/06/04. doi: 10.1038/ncomms5006.

PubMed PMID: 24892406; PubMed Central PMCID: PMCPmc4059926.

7. Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, Rietbergen MM, Haibe-Kains B, Lambin P, Aerts HJ. Radiomic feature clusters and prognostic signatures specific for Lung and Head & Neck cancer. Scientific reports. 2015;5:11044. Epub 2015/08/08. doi: 10.1038/srep11044. PubMed PMID: 26251068.

8. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, Mitra S, Shankar BU, Kikinis R, Haibe-Kains B, Lambin P, Aerts HJ. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. PloS one. 2014;9(7):e102107. Epub 2014/07/16. doi: 10.1371/journal.pone.0102107. PubMed PMID: 25025374; PubMed Central PMCID: PMCPmc4098900.

9. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ. Radiomics: the process and the challenges. Magnetic resonance imaging. 2012;30(9):1234-48. Epub 2012/08/18. doi: 10.1016/j.mri.2012.06.010. PubMed PMID: 22898692; PubMed Central PMCID: PMCPmc3563280.

10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. European journal of cancer (Oxford, England : 1990). 2012;48(4):441-6. Epub 2012/01/20. doi: 10.1016/j.ejca.2011.11.036. PubMed PMID: 22257792; PubMed Central PMCID: PMCPmc4533986.

11. 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.

12. Singanamalli A, Rusu M, Sparks RE, Shih NN, Ziober A, Wang LP, Tomaszewski J, Rosen M, Feldman M, Madabhushi A. Identifying in vivo DCE MRI markers associated with microvessel architecture and gleason grades of prostate cancer. Journal of Magnetic Resonance Imaging. 2016;43(1):149-58.

13. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, Rubin DL, Napel S, Plevritis SK. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data--methods and preliminary results. Radiology. 2012;264(2):387-96. Epub 2012/06/23. doi: 10.1148/radiol.12111607. PubMed PMID: 22723499; PubMed Central PMCID: PMCPMC3401348.

14. Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, Lewis JH, De Ruysscher D, Kikinis R, Lambin P, Aerts HJ. Volumetric CT-based

segmentation of NSCLC using 3D-Slicer. Scientific reports. 2013;3:3529. Epub 2013/12/19. doi: 10.1038/srep03529. PubMed PMID: 24346241; PubMed Central PMCID: PMCPmc3866632.

15. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, Velazquez ER, Lambin P, Gu Y, Balagurunathan Y, Eikman E, Gatenby RA, Eschrich S, Gillies RJ. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. PloS one. 2015;10(3):e0118261. Epub 2015/03/05. doi: 10.1371/journal.pone.0118261. PubMed PMID: 25739030; PubMed Central PMCID: PMCPmc4349806.

16. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine Learning methods for Quantitative Radiomic Biomarkers. Scientific reports. 2015;5:13087. Epub 2015/08/19. doi: 10.1038/srep13087. PubMed PMID: 26278466; PubMed Central PMCID: PMCPMC4538374.

17. Chatellier G, Varlet V, Blachier-Poisson C. “Big data” and “open data”: What kind of access should researchers enjoy? Thérapie. 2016;71(1):107-14.

18. Tenenbaum JD, Avillach P, Benham-Hutchins M, Breitenstein MK, Crowgey EL, Hoffman MA, Jiang X, Madhavan S, Mattison JE, Nagarajan R. An informatics research agenda to support precision medicine: seven key areas. Journal of the American Medical Informatics Association. 2016:ocv213.

19. Yankeelov TE, Mankoff DA, Schwartz LH, Lieberman FS, Buatti JM, Mountz JM, Erickson BJ, Fennessy FM, Huang W, Kalpathy-Cramer J. Quantitative Imaging in Cancer Clinical Trials. Clinical Cancer Research. 2016;22(2):284-90.

20. 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.

21. Colen R, Foster I, Gatenby R, Giger ME, Gillies R, Gutman D, Heller M, Jain R, Madabhushi A, Madhavan S, Napel S, Rao A, Saltz J, Tatum J, Verhaak R, Whitman G. NCI Workshop Report: Clinical and Computational Requirements for Correlating Imaging Phenotypes with Genomics Signatures. Translational Oncology. 2014;7(5):556-69. doi: http://dx.doi.org/10.1016/j.tranon.2014.07.007.

22. McCann SM, Jiang Y, Fan X, Wang J, Antic T, Prior F, VanderWeele D, Oto A. Quantitative Multiparametric MRI Features and PTEN Expression of Peripheral Zone Prostate Cancer: A Pilot Study. American Journal of Roentgenology. 2016;206(3):559-65.

23. Feldman M, Piazza MG, Edwards NA, Ray-Chaudhury A, Maric D, Merrill MJ, Zhuang Z, Chittiboina P. Somatostatin Receptor Expression on VHL-Associated

Hemangioblastomas Offers Novel Therapeutic Target. Neurosurgery. 2015;62:209-10.

24. Gutman DA, Dunn Jr WD, Grossmann P, Cooper LA, Holder CA, Ligon KL, Alexander BM, Aerts HJ. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. Neuroradiology. 2015:1-11.

25. Katrib A, Hsu W, Bui A, Xing Y. "Radiotranscriptomics": A synergy of imaging and transcriptomics in clinical assessment. Quantitative Biology. 2016:1-12.

26. NCI. Enhanced Data Sharing Working Group Recommendation: The Cancer Data Ecosystem Rockville, MD: National Cancer Institute; 2016 [cited 2017 6/12/2017]. Available from: https://http://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/blue-ribbon-panel/ enhanced-data-sharing-working-group report.pdf

27. 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.

28. 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-.

29. Afgan E, Baker D, Van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic acids research. 2016;44(W1):W3-W10.

30. Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. PloS one. 2017;12(5):e0177

U01CA166104-05: Advancing Quantification of Diffusion MRI for Oncologic Imaging

University of Michigan (team 1)

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

INTRODUCTION

The overarching goal of this U01 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 thereby improve use in multi-site cancer imaging trials. Our QIN team research approach is based on development and integration of tools for the three specific aspects of the qDWI workflow from MR system characterization and DWI acquisition QC (toward removal of technical bias and improved measurement precision) to statistical analysis of qDWI biomarkers with quantitative software testing. We currently have several components of the outlined workflow developed in parallel and either undergoing validation through QIN CCPs or being used by ongoing clinical trials to fulfill the specific project Aims.

PROGRESS OVER THE PAST YEAR

§ 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.

Toward development of robust software solution for quantitative DWI analysis in clinical practice, we demonstrated integration of the semi-automated workflow including image co-registration, segmentation, and PRM classification of diffusion scans for glioma clinical trial data set [1]. The developed application provides connectivity to existing picture archive and communication system (PACS) and allows analysis of tumor response to therapy by volumetric changes post-treatment using local hardware or remotely in the cloud for generation of a clinical report document. The performed tests ensured that the developed platform can be seamlessly plugged into clinical decision workflow for any cancer therapy response trial utilizing parametric response mapping (PRM) of Apparent Diffusion Coefficient (ADC) functional diffusion metrics (fDM) from qDWI imaging end points.

To further benchmark PRM fDM results with an alternative ADC metrics, we performed analysis of the 80-patient glioma qDWI data set, using lower ADC histogram portions (excluding necrotic values) to reflect solid tumor volume change. This histogram-

based metrics was intrinsically less sensitive to image registration errors compared to conventional PRM analysis. Kaplan-Meier analysis results (Figure 1) indicated comparable predictive power of both metrics confirming common bio-physical interpretation of the PRM biomarker (increasing with shrinking dense tumor volume). In addition, we are currently awaiting data release from the ISPY-2 trial to allow for processing and evaluation of DW-MR scans of breast cancer patients undergoing therapy.

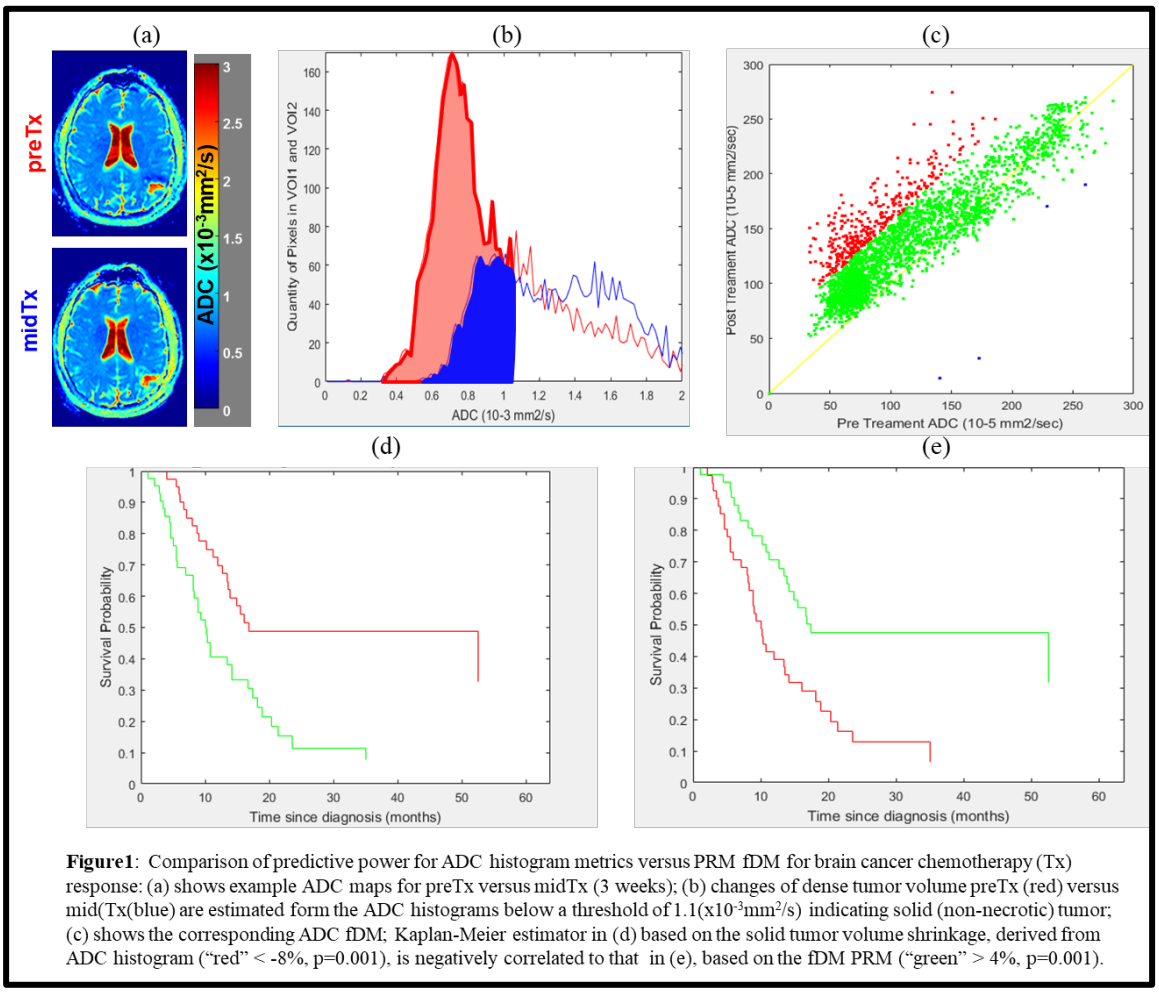


Figure1: Comparison of predictive power for ADC histogram metrics versus PRM fDM for brain cancer chemotherapy (Tx) response: (a) shows example ADC maps for preTx versus midTx (3 weeks); (b) changes of dense tumor volume preTx (red) versus mid(Tx(blue) are estimated form the ADC histograms below a threshold of 1.1(x10^{-3}mm^2/s) indicating solid (non-necrotic) tumor; (c) shows the corresponding ADC fDM; Kaplan-Meier estimator in (d) based on the solid tumor volume shrinkage, derived from ADC histogram ("red" < -8%, p=0.001), is negatively correlated to that in (e), based on the fDM PRM ("green" > 4%, p=0.001).

§ Aim 2: To devise the next generation DWI phantom for absolute quantitation over tissue ADC range that incorporates internal MR-thermometry; and to extend QA/QC metrics to include characterization of systematic bias for ongoing multi-center cancer trials.

Our U01 team continued efforts toward sharing and improving QA/QC tools [URL1] that provide quantitative performance metrics (wCV, SNR, bias) for multi-center qDWI trials. The major improvements focused on interoperability for variable DICOM input formats supplied by an ongoing NRG-BN001 (brain) clinical trial. The Matlab-based (platform-independent) p-code libraries are being shared [URL1a] with interested clinical trial

community (QIBA RSNA, NIST, ECOG-ACRIN, QIN) for centralized qDWI QA based on ice-water and polyvinylpyrrolidone (PVP) quantitative diffusion phantoms that provide an array of calibrated ADC values. The QA SW tools were validated through the participation in QIN ADC mapping CCP [6].

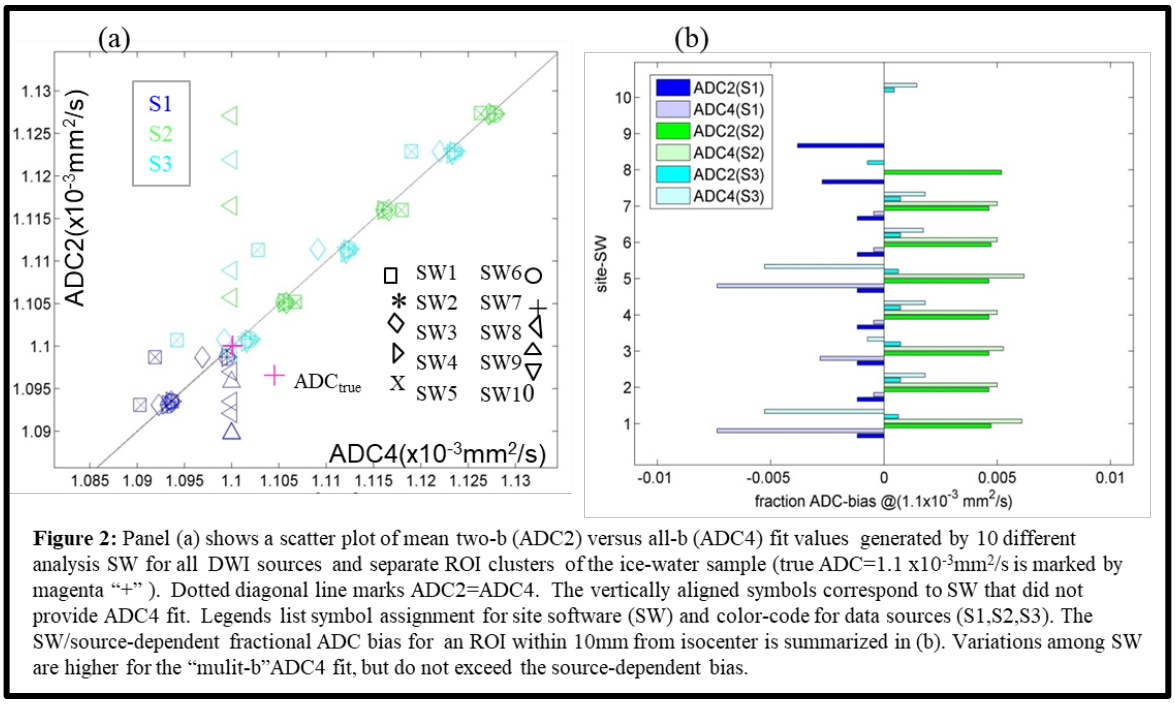


Figure 2: Panel (a) shows a scatter plot of mean two-b (ADC2) versus all-b (ADC4) fit values generated by 10 different analysis SW for all DWI sources and separate ROI clusters of the ice-water sample (true ADC=1.1 x10^{-3}mm^2/s is marked by magenta "+"). Dotted diagonal line marks ADC2=ADC4. The vertically aligned symbols correspond to SW that did not provide ADC4 fit. Legends list symbol assignment for site software (SW) and color-code for data sources (S1,S2,S3). The SW/source-dependent fractional ADC bias for an ROI within 10mm from isocenter is summarized in (b). Variations among SW are higher for the "mulit-b"ADC4 fit, but do not exceed the source-dependent bias.

A sample multi-vendor PVP phantom DWI DICOM were used by two QIN-wide CCPs [6,7] to evaluated the SW algorithms for the sources of bias and variability in multi-site trials starting with the common qDWI data set. This CCP analysis reconfirmed the limited DWI DICOM interoperability and multi-vendor GNL bias as major sources of discrepancy in the derived ADC metrics [6,7]. The consistency between different SW analysis was found to be higher than that between multi-vendor source data (e.g., different ROI locations and gradient systems, Figure 2). The detected absolute variations (< 1%) provided guidance for the desired precision and accuracy of the phantom ADC value calibration to better than 0.5%. To achieve such calibration precision, the temperature measurement accuracy of better than 0.2˚C is desired.

Substantial progress was made by our QIN team toward development of accurate in situ thermometers for ambient temperature diffusion phantoms. To increase chemical shift sensitivity to temperature, we explored paramagnetic systems doped with europium (Eu), thulium (Tm) and Dysprosium (Dy), and devised a novel in-situ thermometer based on bulk magnetic susceptibility (BMS). The designed BMS temperature probe consists of para- and diamagnetic (un-doped) solutions placed in the inside and outside compartments of a coaxial NMR tube (Fig.3a inset). Both temperature sensitivity and absolute chemical shift between probe targets can be tuned using different concentration of the paramagnetic ions. Our preliminary results (Figure 3) indicate 5-7-fold enhancement of ppm/˚C sensitivity (compared

to baseline t-butanol probe) preserving moderate chemical shift range (< 20 ppm, practical for clinical MRI scanners) and allowing for improved accuracy (<0.2˚C) of the temperature probe. For independent verification of ADC calibration of qDWI phantom carried out by NIST, we designed and implemented the dynamic qDWI measurement protocol with optical probe to double ADC measurement accuracy compared to conventional methods (CI(ADC)>1%).

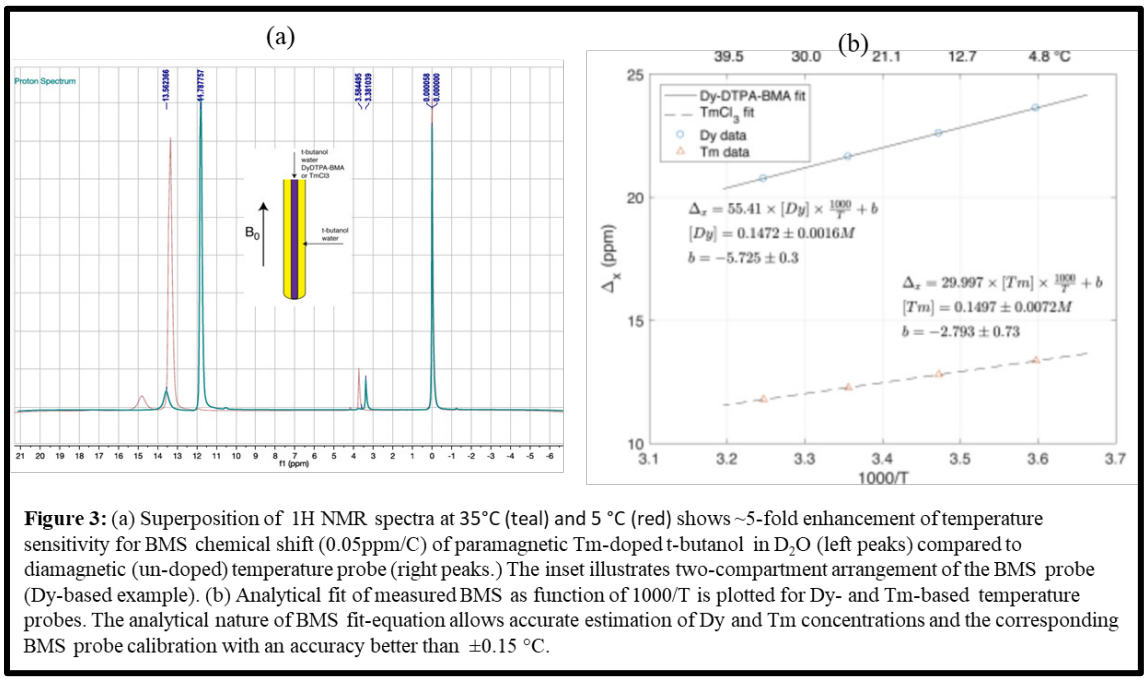


Figure 3: (a) Superposition of 1H NMR spectra at 35°C (teal) and 5 °C (red) shows ~5-fold enhancement of temperature sensitivity for BMS chemical shift (0.05ppm/C) of paramagnetic Tm-doped t-butanol in D_2O (left peaks) compared to diamagnetic (un-doped) temperature probe (right peaks.) The inset illustrates two-compartment arrangement of the BMS probe (Dy-based example). (b) Analytical fit of measured BMS as function of 1000/T is plotted for Dy- and Tm-based temperature probes. The analytical nature of BMS fit-equation allows accurate estimation of Dy and Tm concentrations and the corresponding BMS probe calibration with an accuracy better than ±0.15 °C.

§ AIM 3: Enhancement of predictive power for quantitative diffusion metrics by retrospective correction of DW-MRI gradient nonlinearity errors in multi-center therapy-response trials.

To circumvent data sharing logistics issues, our team is developing vendor-agnostic correction tools to share with the UCSF central analysis lab that hosts data repository for the breast cancer therapy response trial. Since trace-only DWI DICOM was archived for this trial, the trace-DWI direction-average (isotropic) GNL correctors need to be constructed for direct application to the ADC maps [4, URL2]. Toward this aim, through our AIP collaboration with the MRI vendors and UCSF and JHU QIN academic partners, our QIN team started generating the GNL correction maps for the MRI scanner model inventory compiled for the ACRIN 6698 trial. These maps are based on vendor-provided gradient system design coefficients. As a result of this effort, the pre-developed Matlab p-code libraries for GNL map generation were pre-validated by the AIP participants and shared with the wider research community [URL3a].

To finalize and deploy the stand-alone tools for retrospective GNL correction "on-site", our team continued integration of the developed correction algorithms [2-4] using available multi-system, multi-vendor phantom and human volunteer DWI DICOM. The ADC ROI histogram statistics for phantom and healthy parenchyma compared before and after

correction confirms improved uniformity and reproducibility across scanners [3, 5, 10, 11]. Major hurdle for implementation of vendor-agnostic retrospective "off-line" GNL bias correction was found to be in generalization for the vendor-specific DICOM conventions. ADC DICOM mapping QIN CCP led by our team in collaboration with UCSF and Brigham Women's QIN centers supplied requirements for this critical component [7]. Our combined AIP and QIN efforts [2-7] have already encouraged two vendors to implement prototype prospective (on-line) correction tools on their scanner systems.

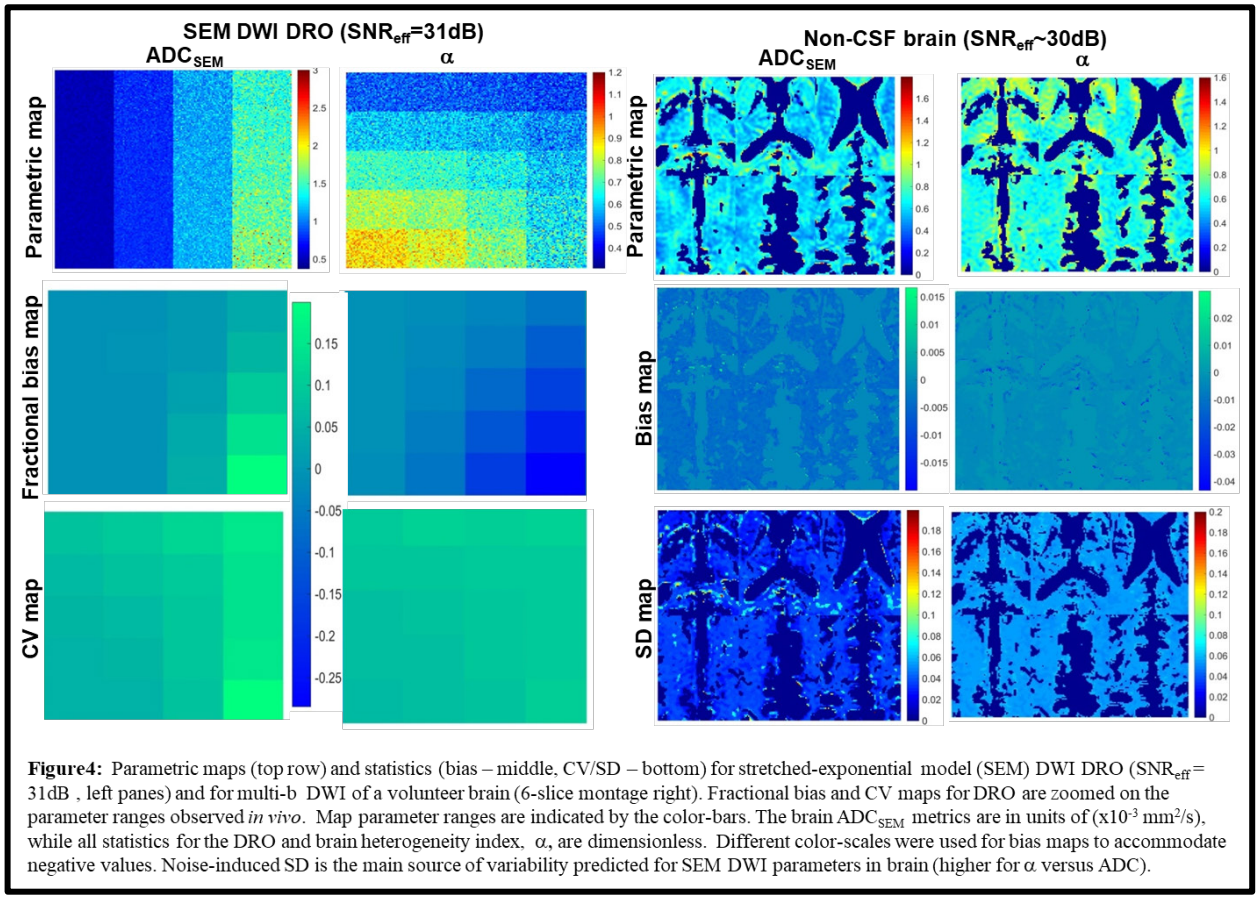


Figure4: Parametric maps (top row) and statistics (bias – middle, CV/SD – bottom) for stretched-exponential model (SEM) DWI DRO (SNR$_{eff}$= 31dB , left panes) and for multi-b DWI of a volunteer brain (6-slice montage right). Fractional bias and CV maps for DRO are zoomed on the parameter ranges observed *in vivo*. Map parameter ranges are indicated by the color-bars. The brain ADC$_{SEM}$ metrics are in units of (x10^{-3} mm^2/s), while all statistics for the DRO and brain heterogeneity index, α, are dimensionless. Different color-scales were used for bias maps to accommodate negative values. Noise-induced SD is the main source of variability predicted for SEM DWI parameters in brain (higher for α versus ADC).

For quantitative testing of multi-b diffusion model fit algorithms, a set of synthetic DWI digital reference object (DRO) DICOM was generated for several forward diffusion models including physical noise. By providing the ground-truth diffusion parameters, these DROs allow evaluation of noise-induced biases and standard error in derived quantitative diffusion metrics (Figure 4) for the typical DWI acquisition protocols of ongoing clinical trials (e.g., ACRIN 6698, NRGBN001). Several DRO DWI DICOM data sets were shared through ResearchGate [URL1c] and QIBA RSNA QIDW (https://goo.gl/yYPG0W), and offered for SW tool pre-screening by the ongoing multi-b qDWI QIN MRI CCP (led by MCW).

COLLABORATIONS WITHIN THE NETWORK

Our team continued close collaboration with two other QIN sites (UCSF and JHU) on the AIP project with three dominant MRI vendors to implement the GNL correction of DWI bias on clinical MRI scanners. The tools and resources developed through this project are being shared with the QIN for validation and wider implementation [2-5, URL2, URL3].

With the MSKCC QIN site and NIST collaborators, our team continues to contribute to QIBA RSNA activities of the DWI Task Force (co-chaired by Dr. Chenevert) toward development of technical consensus profile (http://qibawiki.rsna.org/index.php/Profiles) for qDWI biomarkers in brain, liver and prostate, detailing standardized acquisition/processing procedures and system performance metrics (including GNL bias effects) for confident measurement of the quantitative ADC metrics in multi-center clinical trials. Through these collaborations, our team shares and improves QC/QA tools [URL1; QIBA QIDW: https://goo.gl/xjHc6G] for clinical trials utilizing quantitative DWI. Our QC tools are also used by several QIN teams (UWASH, ECOG-ACRIN, MSKCC, Moffit).

In addition, our team has been an active member of the QIN MRI Subgroup and participated in multiple QIN-wide Challenges and Collaborative Projects (CCPs) that resulted in four publications this year [6-9]:

1) "ADC Challenge" – led by David Newitt, UCSF (completed 2017, published JMI 2018 [6])

2) "ADC DICOM Challenge" – led by Dariya Malyarenko, UMich – Team1 (completed, published JMI 2018 [7])

3) "DCE AIF Challenge Part 2" – led by Wei Huang, OHSU (completed 2017 [12,13], manuscript in prep)

4) "DSC Challenge" – led by Kathleen Schmainda, MCW (completed 2017, accepted AJNR [8])

5) "T1-Mapping Challenge" led by Octavia Bane, Mount Sinai (completed, published MRM 2018 [9])

6) "DSC DRO" – led by Chad Quarles, Barrows Neurological Institute (ongoing)

7) "Multi-b DWI Challenge" – led by Peter LaViolette, MCW (ongoing)

PLANS FOR NEXT YEAR

Our future efforts will be focused on sharing and validating (through QIN CCPs) the newly developed qDWI analysis, QC and GNL bias correction tools, as well as, on testing local PRM SW application on the subset of ACRIN6698 breast cancer trial data, shared by the UCSF central analysis lab. The local SW evaluation will also utilize the pre-built DWI DRO data sets. We plan to deploy the first prototype tools for DWI DRO generation and bias analysis in the 2nd-3rd quarter of 2018. On a similar time scale, we plan to finalize p-code libraries for the GNL correction tools to share with the UCSF QIN team for retrospective application to multi-vendor ISPY-2 breast cancer DWI data on site.

Temperature calibration and development of practical imaging protocol for BMS-based in-situ thermometer on 3T and 1.5T is planned for the 3rd quarter of 2018. In collaboration with NIST and QIBA DWI, we will continue efforts for calibration of ADC values for qDWI PVP-based phantom with the target precision of CI < 0.5%. We also plan to expand our collaboration with MSKCC QIN team to develop new qDWI physical phantoms for non-Gaussian diffusion and optimization of multi-b DWI acquisition. Our QIN team will continue improvement of existing quantitative DWI QA/QC libraries and sharing validated tools with interested users through the ongoing AIP, QIBA RSNA, NIST, and ECOG-ACRIN collaborations.

LIST OF REFERENCE

1. Keith L, Ross BD, Galbán CJ, Luker GD, Galbán S, Zhao B, Guo X, Chenevert TL, Hoff BA. Semiautomated Workflow for Clinically Streamlined Glioma Parametric Response Mapping. *Tomography*, 2016; 2(4):267-275

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; PMCID: PMC4630210.

3. 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(4):56-66. PMCID: PMC 5241082

4. Chenevert TL, Ross BD, Malyarenko DI. Error Analysis and Correction of Mean Diffusivity Measurements for Gradient Nonlinearity. Patent US9851426, issued 12/2017.

5. Newitt DC, Tan ET, Wilmes LJ, Chenevert TL, Kornak J, Marinelli L, Hylton N. "Gradient nonlinearity correction to improve apparent diffusion coefficient accuracy

and standardization in the american college of radiology imaging network 6698 breast cancer trial," *J Magn Reson Imaging* **42**(4), 908-919 (2015).

PUBLICATIONS FROM QIN 2017 EFFORTS:

6. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, et.al.: Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. J Med Imaging (Bellingham) 5(1): 011003, 2018.

7. Malyarenko D, Fedorov A, Bell L, Prah M, Hectors S, Arlinghaus L, Muzi M, Solaiyappan M, Jacobs M, Fung M, Shukla-Dave A, McManus K, Boss M, Taouli B, Yankeelov TE, Quarles CC, Schmainda K, Chenevert TL, Newitt DC: Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. J Med Imaging (Bellingham) 5(1): 011006, 2018

8. Schmainda K, Prah M, Rand S, Muzi M, Rane S, Da X, Yen Y-F, Kalpathy-Cramer J, Chenevert T, Hoff B, Ross B, et.al. Multi-Site Concordance of DSC-MRI Analysis for Brain Tumors: Results of a NCI Quantitative Imaging Network DSC-MRI Collaborative Project. Accepted AJNR, 2018.

9. Bane O, Hectors SJ, Wagner M, Arlinghaus LL, Aryal MP, Cao Y, Chenevert TL, Fennessy F, Huang W, Hylton NM, Kalpathy-Cramer J, Keenan KE, Malyarenko DI, Mulkern RV, Newitt DC, Russek SE, Stupic KF, Tudorica A, Wilmes LJ, Yankeelov TE, Yen YF, Boss MA, Taouli B: Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: Results from a multicenter phantom study. Magn Reson Med: 2017. PM28913930

§ Conference Proceedings:

10. Malyarenko D, Pang Y, Wilmes L, Tan ET, Kirsch J, Sénégas J, Jacobs M, Newitt D, Chenevert T. Correction of nonuniform diffusion weighting in DWI using vendor-provided gradient characteristics. *TP05: 1780 proceedings ISMRM'25* Annual Conference, Honolulu HI, April 2017.

11. Pang Y, Malyarenko D, Schar M, Wilmes L, Newitt D, Jacobs M, Chenevert T. Characterization of B0 Shim-Induced Bias in Diffusion Weighting Gradients. *EP03: 3360 proceedings ISMRM'25* Annual Conference, Honolulu HI, April 2017.

12. Li K, Chen Y, Yu Y, Li X, Fedorov A, Jajamovich G, Malyarenko D, Aryal M, LaViolette P, et.al. The Effects of AIF Quantification Variations on DCE-MRI Prediction of Soft Tissue Sarcoma Response to Preoperative Therapy: A Preliminary Multicenter Study. *EP11: 4375 proceedings ISMRM'25* Annual Conference, Honolulu HI, April 2017.

13. Huang W, Chen Y, Fedorov A, Li X, Jajamovich G, Malyarenko D, Aryal M, LaViolette P, et.al. Effects of AIF Quantification Variations on Shutter-Speed Pharmacokinetic Modeling of Prostate DCE-MRI Data: A Multicenter Data Analysis Challenge, Part II. *O41: 625 proceedings ISMRM'25* Annual Conference, Honolulu HI, April 2017.

§ URLs for shared tools and resources:

URL1: Quantitative DWI QC resources on ResearchGate (to share pre-validated tools):
https://www.researchgate.net/project/Quantitative-DWI-tools
The project resource links:
a. Quantitative DWI phantom QC analysis tool p-code and manual (1/2018): https://www.researchgate.net/publication/322405528
b. Video tutorial for DWI phantom QC analysis for a clinical trial (5/2016): https://www.researchgate.net/publication/322405550
c. Multi-b three-model DWI DRO DICOM and manual (1/2018): https://www.researchgate.net/publication/322404473

URL2: Empiric GNL tensors and isotropic corrector maps (MAT-structures) for six QIN DAWG CCP (MRI) systems (3/2017):
https://umich.box.com/s/w0grq3p91l3dhv2f39d037dyagn7bked

URL3: GNL correction resources on ResearchGate (to share validated protocols and tools):
https://www.researchgate.net/project/Prospective-and-retrospective-correction-of-gradient-nonlinearity-bias-in-diffusion-weighting-for-multi-center-trials-using-ADC
The project resource links:
a. Bz and GNL map generating p-code, manual and demo (12/2016): https://www.researchgate.net/publication/316933059
b. DWI scan instructions for GNL correction validation (Sup. Mat. PMC5241082): https://www.researchgate.net/publication/316930132
c. DWI scan instructions for empiric GNL characterization (Sup. Mat. PMC4630210): https://www.researchgate.net/publication/312301728

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

University of Michigan (team 2)

Lubomir Hadjiyski, Ph.D.

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.

PROGRESS OVER THE PAST YEAR

§ Specific Aim 1: Data Collection

During the current time period of the project we have collected additionally 60 CTU bladder cancer cases from CTU examinations performed at University of Michigan. This includes 50 pre- and post- neoadjuvant chemotherapy treatment cases with clinical stage larger than T1, and 10 cases of which the clinical stage were called T1 and did not underwent neoadjuvant chemotherapy treatment. As a result, we have collected in total 286 CTU bladder cancer cases: 182 pre- and post- neoadjuvant chemotherapy treatment cases and 104 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 191 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 are continuing the data collection. We have full access to the collected data. So far we have collected 14 cases. This will be very valuable dataset allowing us to develop tools for very early prediction of response to treatment.

In addition, with IRB approval we also are collecting molecular biomarker data for the above bladder cancers. The molecular biomarker data includes DNA alterations such as

somatic variants, copy number gains, copy number losses. So far we have collected molecular data for 25 cases. This also will be very valuable dataset allowing us to develop multimodality tools combining molecular, clinical and radiomics information for 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 continued exploring 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. We have applied the method on the enlarged test set of 92 cases, which also were hand outlined for both the inner and outer wall and used as reference standard. The results are very similar to the one reported in the previous year report on a smaller test set of 37 cases. This is a promising result demonstrating the robustness of the system. We are in the final stages of preparation for submission of a manuscript to Medical Physics.
(**Publications from QIN Efforts: #11**)

§ 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. The inaccuracy in staging may be partly attributed to the subjectivity and variability of clinicians in utilizing available diagnostic information. The pathological staging after cystectomy (the removal of the bladder) is considered very reliable. An objective decision support system

trained with the correct staging information may be useful for assisting 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 classifier trained with improved pathological stage labels obtained after cystectomy. The details of the CDSS-S design and evaluation are presented in the following:

Methods: We have used a data set of 84 bladder cancers from 76 CTU cases collected as described in Specific Aim 1, that were all clinically staged prior to treatment. The bladder lesions were segmented using our previously developed auto-initialized cascaded level sets (AI-CALS) method. For each lesion, a bounding box provided by a radiologist was used as input for the AI-CALS segmentation system. Each lesion was categorized into to one of two classes: (1) pathologically staged below T2 or (2) pathologically staged at or above T2. Of the 84 bladder cancers, 43 of the lesions were below stage T2, and 41 were stage T2 or above.

Ten of the 84 bladder cancers were upstaged to stage T2 or above after cystectomy (bladder removal) and pathological staging, which is considered to be the ground truth. We corrected the labels for these 10 understaged lesions. After correcting stage labels for the understaged lesions, 33 of the 84 lesions were below stage T2, and 51 were stage T2 or above.

Eighty-nine radiomics features were extracted from each of the segmented lesions including 26 morphological [10] and 63 texture features [11, 12]. The morphological features included gray level, contrast, Fourier descriptor, gradient magnitude profile 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.

A leave-one-case-out cross-validation was conducted with the 76 cases. Feature selection was performed to identify key features within the data set. Three features were selected including the Fourier descriptor, texture, and gradient magnitude profile features. We trained a linear discriminant analysis (LDA) classifier to distinguish between bladder lesions that were diagnosed as stage T2 or above and those below stage T2. For each training/test leave-one-case-out partition the trained classifier outputted likelihood of stage T2 or above score for the lesions in the left-out test case.

As a comparison to our computerized decision support system (CDSS-S) trained and tested with the corrected post-cystectomy labels, a second CDSS-S was trained with the understaged pre-treatment labels and tested on the lesions corrected for the understaged labels. Again, feature selection and leave-one-case-out cross-validation were performed for training and testing the LDA classifier. The selected features were texture, gradient magnitude, and gradient magnitude profile features.

The CDSS-S performance was evaluated by the receiver operating characteristics (ROC) analysis, and the classification performance was quantified by the area under the ROC curve (AUC).

Results: For the CDSS-S trained with the corrected stage labels, the test AUC was 0.89 ± 0.04. For the CDSS-S that was trained with the understaged pre-treatment labels and tested on the data set corrected for the understaged lesions, the test AUC was 0.86 ± 0.04. The test ROC curves for the two CDSS-S are shown in Figure 1. The differences did not reach statistical significance ($p > 0.05$). When the CDSS-S was trained with the corrected stage labels, the CDSS-S produced correctly higher likelihood scores of stage T2 or above for 9 of the 10 understaged lesions. Examples of bladder cancers with stages ≥ T2 or < T2 with the corresponding computer outlines and classifier scores are presented in Figure 2.

Conclusions: We demonstrate the promise of using radiomic features automatically extracted from CTU and correct staging information to build a statistical predictive model for staging of bladder cancer. The improvement of the AUC scores when the decision support system was trained with more accurate labels affirms the effect of correct labels on the statistical modeling. Further work includes improvement in the radiomic features and in the predictive model through the inclusion of clinical and molecular data and the collection of a larger data set. (**Publications from QIN Efforts: #14 and #15**)

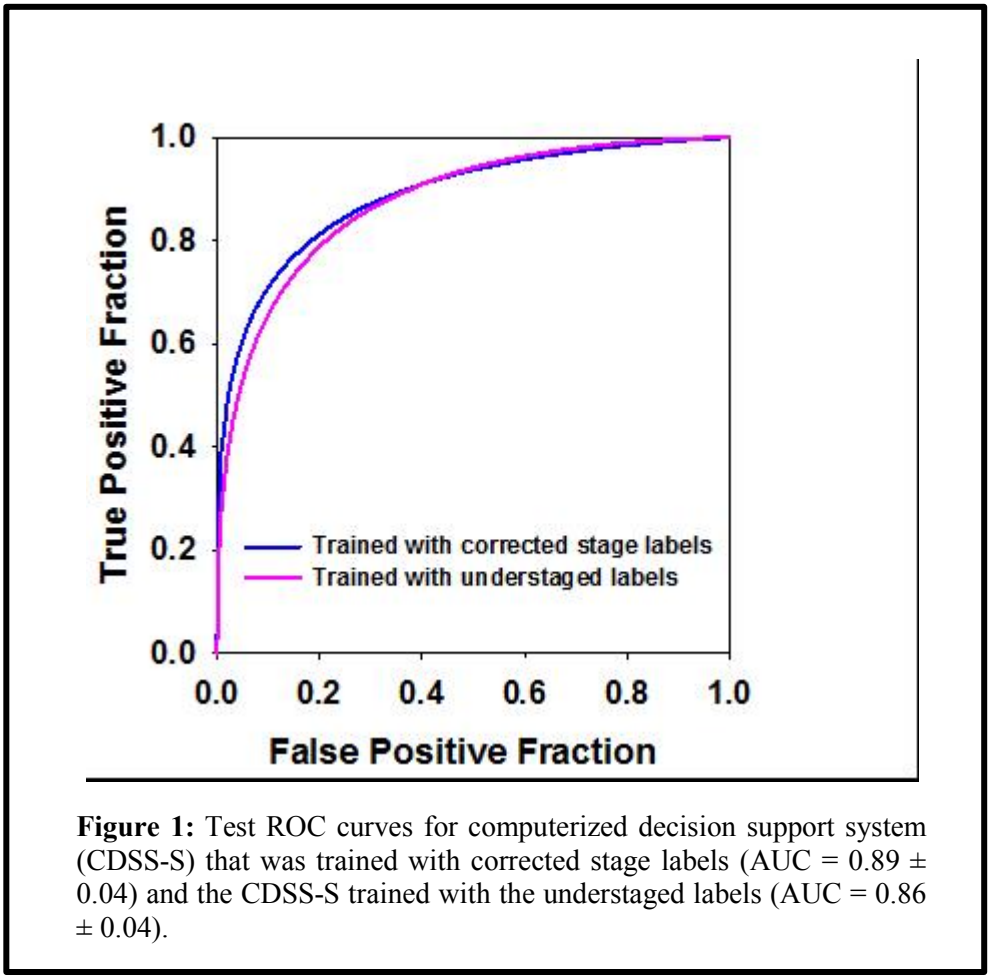


Figure 1: Test ROC curves for computerized decision support system (CDSS-S) that was trained with corrected stage labels (AUC = 0.89 ± 0.04) and the CDSS-S trained with the understaged labels (AUC = 0.86 ± 0.04).

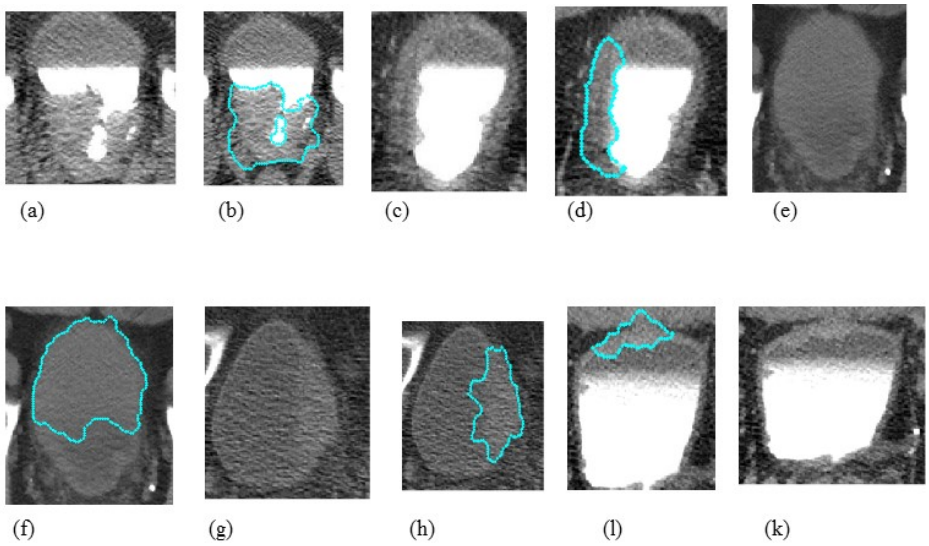


Figure 2: Here are examples of CTU scans of bladder cancers with stages ≥ T2 or < T2. The blue outlines are the AI-CALS segmentation. The following reported scores are the test scores. (a) and (b) show a T1 stage cancer that was properly classified with a score of -4.47. (c) and (d) show a T1 stage case that was properly classified with a score of -3.13. (e) and (f) show a T2 stage case that was properly classified with a score of 1.37. (g) and (h) show a T3 stage case that was properly classified with a score of 1.41. (k) and (l) show a case that was clinically identified as Ta (< T1) pre-treatment but was identified as a T2 stage cancer post-cystectomy. The classifier trained with understaged labels identified the cancer as <T2 with a score of -1.11 and the classifier trained with corrected labels identified the cancer correctly as ≥T2 with a higher score of 0.094.

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. In addition, if a patient can be reliably identified as having complete response to treatment, the treatment option of preserving the bladder may be considered, which would drastically reduce the morbidity of the patient and improve his/her quality of life as compared to the current standard treatment by cystectomy.

During the current time period of the project we have continued the development of the computer decision support system (CDSS-T) for bladder cancer treatment response monitoring. We have continued exploring further the use of a deep learning convolution neural network (DL-CNN) in CDSS-T. We also combined radiomic features and clinically estimated feature as input to CDSS-T. We have started a pilot observer performance study evaluating whether CDSS-T can assist radiologists in identifying patients who have complete response after neoadjuvant chemotherapy. The details of the CDSS-T design and evaluation are presented in the following:

Bladder Cancer Treatment Response Assessment in CT Urography using Two-Channel Deep-learning Network

We have continued the design of CDSS-T by using a 2-Channel Deep-Learning Convolution Neural Network (2Ch-DCNN) to recognize the pattern changes indicative of treatment response. The 2Ch-DCNN may be useful as decision support for bladder cancer treatment response assessment, which is crucial for identifying non-responders and stopping treatment early to preserve their physical condition.

Methods: Pre- and post-neoadjuvant chemotherapy CTU scans of 82 patients (87 lesions) were collected as described in Specific Aim 1. The cancer stage after treatment, as determined by cystectomy, was used as reference standard for treatment response. Bladder lesions in the CTU scans were segmented with our auto-initialized cascaded level sets (AI-CALS). Paired ROIs were extracted from the segmented lesions in the pre- and post-treatment scans and partitioned by case for 2-fold cross validation. Partition 1 contained 42 cases (45 lesions, 13 stage T0) and 9,000 training ROI pairs after data augmentation. Partition 2 contained 40 cases (42 lesions, 11 stage T0) and 9,600 training ROI pairs. Examples of the ROI pairs are shown in Figure 3. The paired pre- and post- treatment ROIs were input to the two parallel channels, respectively, of the 2Ch-DCNN to distinguish between bladder lesions of stage T0 and >T0 after treatment. A flow chart of the 2Ch-DCNN is shown in Figure 4. During training, the number of ROIs between the 2 classes was balanced. During testing, the trained network was applied directly to the ROIs in the test set. We compared the 2Ch-DCNN with our hybrid pre-post-treatment ROI DCNN method, and the assessments by 2 experienced GU radiologists. The radiologist estimated the likelihood of stage T0 after viewing each pre-post-treatment CTU pair. Receiver operating characteristic (ROC) analysis was performed and the area under the curve (AUC) and AUC at sensitivity >90% ($AUC^{0.9}$) were compared.

Results: The test AUCs were 0.76±0.07 and 0.75±0.07 for the 2 partitions, respectively, for the 2Ch-DCNN, and were 0.75±0.08 and 0.75±0.07 for the hybrid ROI method. The AUCs for Radiologist 1 were 0.67±0.09 and 0.75±0.07 for the 2 partitions, respectively, and were 0.79±0.07 and 0.70±0.09 for Radiologist 2. For the 2Ch-DCNN, the $AUC^{0.9}$s were 0.43 and 0.39 for the 2 partitions, respectively, and were 0.19 and 0.28 for the hybrid ROI method. For Radiologist 1, the $AUC^{0.9}$s were 0.14 and 0.34 for partition 1 and 2, respectively, and were 0.33 and 0.23 for Radiologist 2. The differences in the performance

did not achieve statistical significance. Test ROC curves for prediction of T0 stage after neoadjuvant chemotherapy are presented in Figure 5.

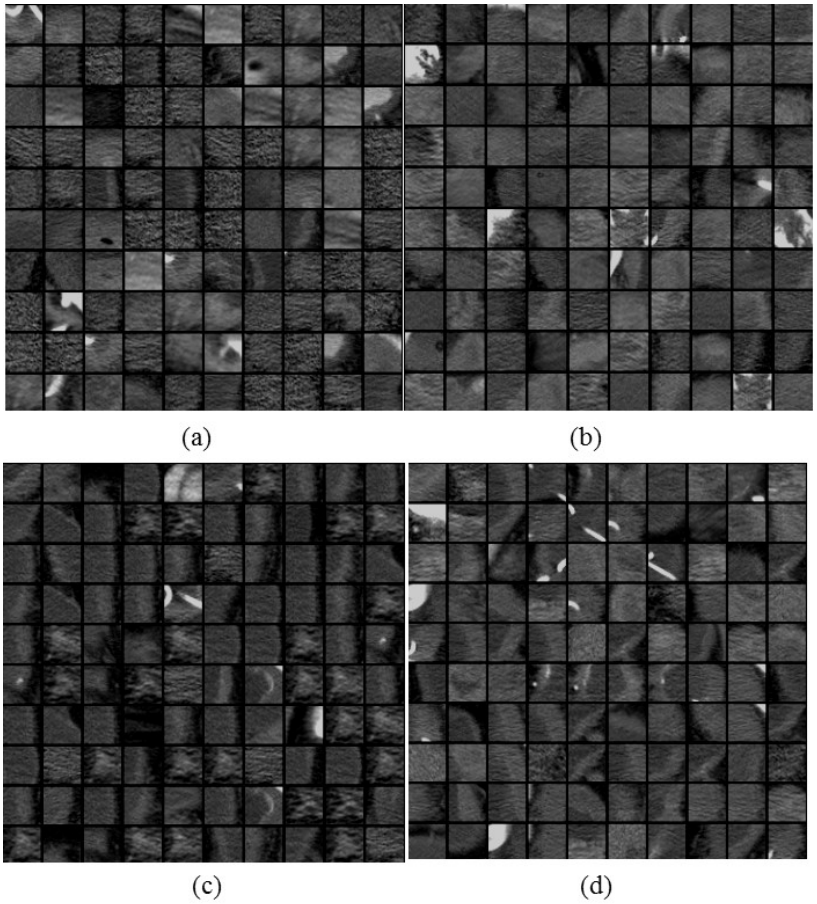


Figure 3: Subset of ROIs used to train the 2 channel DCNN. Each ROI is 32x32 pixels. (a) Pre-treatment ROIs that were labeled as being stage T0 after treatment. (b) Pre-treatment ROIs that were labeled as being greater than stage T0 after treatment. (c) Post-treatment ROIs that were labeled as being stage T0 after treatment. (d) Post-treatment ROIs that were labeled as being greater than stage T0 after treatment.

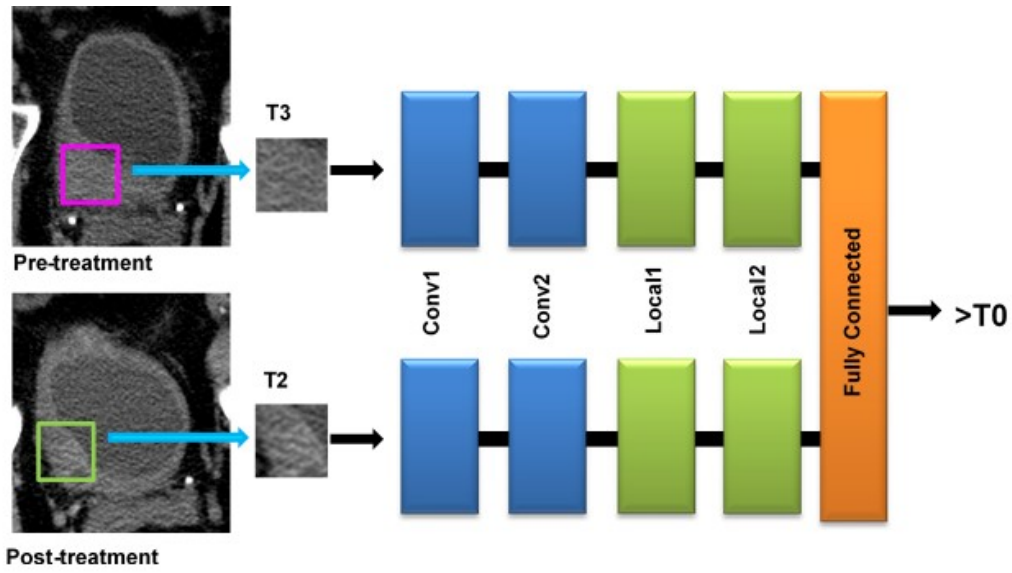


Figure 4: Flow chart of the 2Ch-DCNN. ROIs from the pre- and post-treatment scans are inputted into two separate channels of the DCNN, which is trained to distinguish ROI channel pairs that have fully responded to treatment and those that have not.

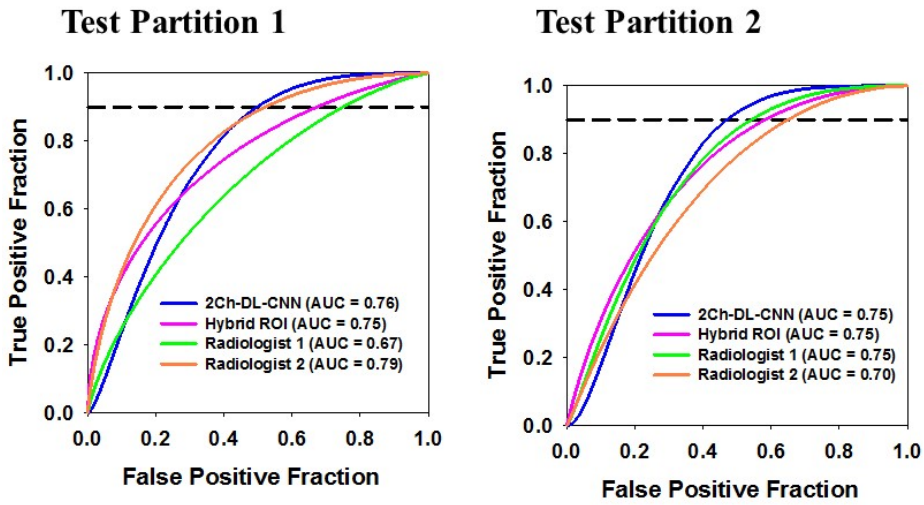


Figure 5: Test ROC curves for prediction of T0 stage after neoadjuvant chemotherapy using ROIs generated from pre- and post-treatment pairs for the test partitions using the 2Ch-DCNN, 1-channel hybrid ROI DCNN, and radiologists for the two partitions.

Conclusions: Our study demonstrated the feasibility of using 2Ch-DCNN for the estimation of bladder cancer treatment response in CT. The 2Ch-DCNN performed comparably to the 1-channel hybrid ROI DCNN, with higher $AUC^{0.9}$. The 2Ch-DCNN also performed comparably to the two radiologists.
(**Publications from QIN Efforts: #17**)

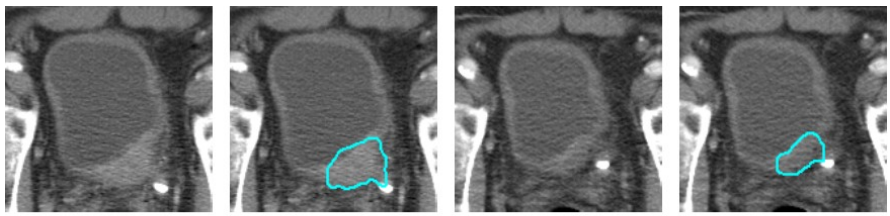
Analysis of Treatment Response of Bladder Cancer on CT Scans by Synergistic Combination of Radiomic Features and Clinically Estimated Feature

We have continued the design of CDSS-T also by addition of clinically estimated feature to the automatically extracted radiomic features from CT scans as input to the CDSS-T for treatment response assessment of bladder cancer.

Methods: Our Auto-Initialized Cascaded Level Set (AI-CALS) system is designed to extract 3D lesion boundary based on level sets. Forty-seven radiomic features (RF) based on pre- and post- treatment changes in volume (V), 5 gray level (GL) and 9 shape (S) descriptors and 32 texture features (RLS) were extracted from the segmented lesions. A clinically estimated feature, the bimanual exam under anesthesia (EUA), was also collected from the clinical reports. Linear discriminant analysis was used to generate two combined response indices: one by the radiomic features (RFs) alone (CRI-RF), and the other with both radiomic features and EUA (CRI-RF-EUA). Pre- and post-chemotherapy treatment CT scans of 98 patients with bladder cancers were collected as described in Specific Aim 1. Examples of CT slices of carcinoma on pre- and post-treatment CT scans with AI_CALS segmentation are presented in Figure 6. For all cases, cystectomy was performed after treatment and the disease outcome was available as reference standard of treatment response. Twenty-five percent of patients had pT0 disease (complete response) at cystectomy. A radiologist marked 122 temporal pairs of primary site cancers. Stepwise feature selection and leave-one-case out cross-validation and receiver operating characteristic (ROC) analysis were performed. The area under the ROC curve (AUC) was calculated to estimate the accuracy for predicting pT0 stage (complete response) at cystectomy by V, CRI-RF and CRI-RF-EUA methods. Two radiologists also provided the likelihood of pT0 stage of the tumor by reading the pre- and post-treatment paired CT scans.

Results: For the 122 cancers, the AUC for prediction of pT0 disease at cystectomy was 0.70±0.05 for volume (V). The AUC for CRI-RF based on 2 Contrast and 2 RLS features was 0.74±0.05 and increased to 0.78±0.05 when EUA was added (CRI-RF-EUA). Prediction of pT0 disease by radiologists resulted in AUCs of 0.77±0.05 and 0.75±0.05, respectively. The differences did not reach statistical significance ($p>0.05$). The test ROC curves for prediction of T0 stage after neoadjuvant chemotherapy are presented in Figure 7.

Conclusions: Both CRI-RF and CRI-RF-EUA performed similar to the radiologists and better than V for estimation of treatment response. The addition of EUA further improved the accuracy of treatment response assessment. The combined response index using both the radiomic features and clinically estimated EUA has the potential to provide accurate treatment response assessment and is superior to volume change alone.



(a) (b) (c) (d)

Figure 6: CT slices of carcinoma on pre- and post-treatment CT scans with AI_CALS segmentation: (a) pre-treatment scan, (b) pre-treatment scan with AI_CALS segmentation, (c) post-treatment scan, (d) post-treatment scan with AI_CALS segmentation. The carcinoma (white arrow) is shown on the best slice marked by radiologist.

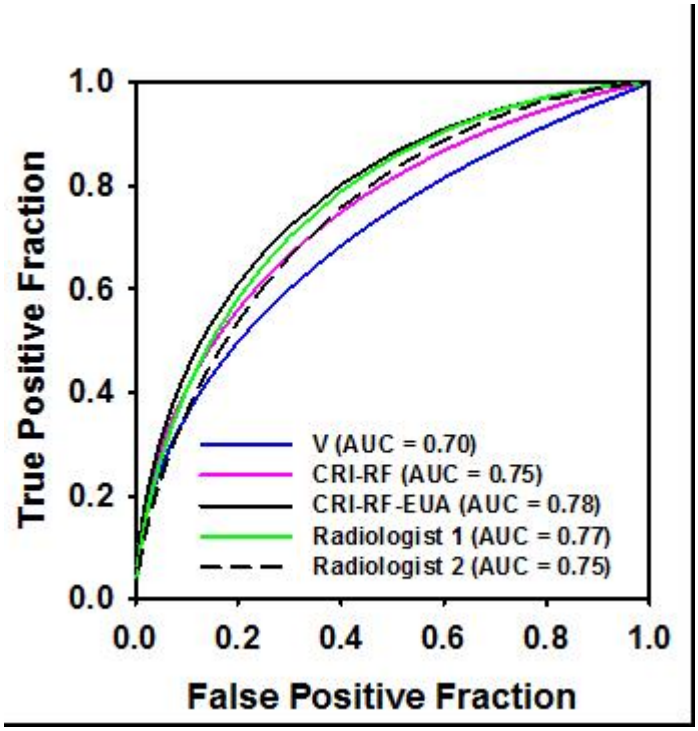


Figure 7: ROC curves for prediction of pT0 disease at the time of cystectomy for AI-CALS estimated V (AUC=0.70±0.05), the combined response index (CRI-RF) based on 2 Contrast and 2 RLS features (AUC=0.74±0.05), the combined response index (CRI-RF-EUA) based on 2 Contrast, 2 RLS features and clinically estimated EUA (AUC=0.78±0.05), the Radiologist1 estimation (AUC=0.77±0.05), and the Radiologist 2 estimation (AUC=0.75±0.05).

Observer Performance Study for Bladder Cancer Treatment Response Assessment in CT Urography with and without computerized decision support

We have started evaluation of our computerized decision support system for bladder cancer treatment response assessment (CDSS-T) for whether it can assist radiologists in identifying patients who have complete response after neoadjuvant chemotherapy by a pilot observer performance study. If a patient can be reliably identified as having complete response to treatment, the treatment option of preserving the bladder may be considered, which would drastically reduce the morbidity of the patient and improve his/her quality of life as compared to the current standard treatment by cystectomy. More accurate estimation of response to treatment is also vital for identifying non-responders and allowing them to seek alternative therapy.

Methods: Pre- and post-chemotherapy CTU scans of 123 patients were collected as described in Specific Aim 1, resulting in 158 pre- and post-treatment lesion pairs. The pathological cancer stage after treatment, as determined by cystectomy, was collected as the reference standard of whether a patient fully responded to treatment. Twenty-five percent of the lesion pairs (40/158) had T0 cancer stage after chemotherapy, which corresponds to a complete response. We have developed a CDSS-T system that uses a combination of DL-CNN and radiomics features to distinguish between cases that have fully responded to treatment and those that have not. Two abdominal radiologists and 4 residents trained in abdominal radiology estimated the likelihood of stage T0 disease after treatment by viewing each pre-post-treatment CTU pair displayed side by side on a specialized graphic user interface designed for CDSS-T (Figure 8). The observer provided an estimate without CDSS-T first (Figure 8a) and then might revise the estimate, if preferred, after the CDSS-T score was displayed (Figure 8b). The cases were randomized differently for each observer. The observers' estimates with and without CDSS-T were analyzed with multi-reader, multi-case (MRMC) receiver operating characteristic (ROC) methodology. The area under the curve (AUC) and the statistical significance of the difference were calculated.

Results: The AUC for prediction of T0 disease after treatment was 0.80±0.04 for the CDSS-T alone. Each observer's performance increased with the aid of CDSS-T. The average AUC for the observers were 0.75 (range: 0.70-0.79) without CDSS-T, and increased to 0.78 (range: 0.73-0.81) with CDSS-T. The individual AUCs without and with CDSS-T are presented in Table 1 and Figure 9. The differences in the average AUC values between without CDSS-T and with CDSS-T were statistically significant ($p < 0.01$).

Conclusions: Our pilot study demonstrated that CDSS-T system for bladder cancer treatment response assessment in CTU has the potential to improve radiologists' performance in identifying patients who fully responds to treatment. The improved radiologists' accuracy in bladder cancer treatment response assessment is vital for identifying non-responders and allowing them to seek alternative therapy. It is also vital for identifying complete responders as candidates for organ preservation therapy.

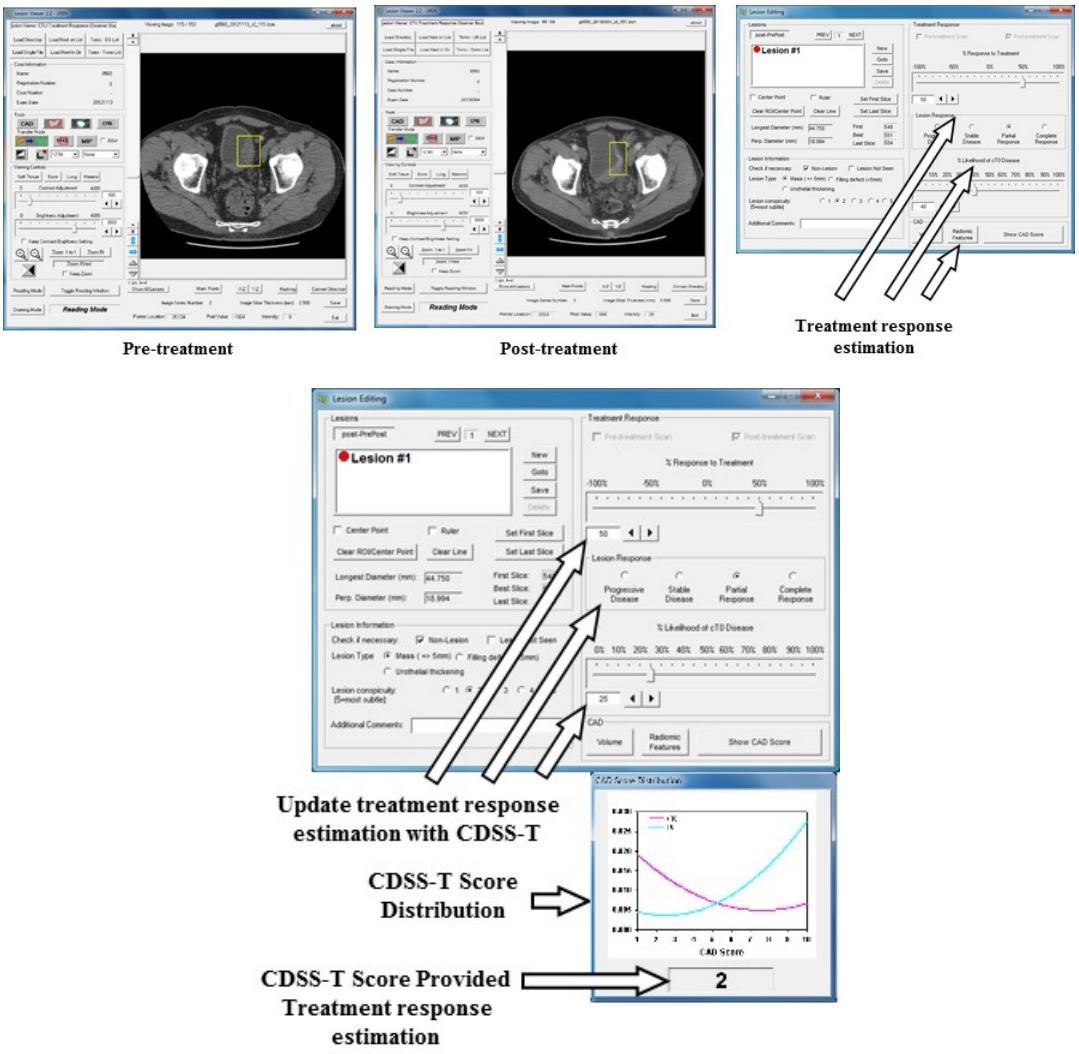


Figure 8: The graphical user interface for reading with and without our computer-aided diagnosis (CAD) system designed for supporting treatment response assessment (CDSS-T). (a) The pre- and post-treatment scans are shown side-by-side, and the observer estimates the treatment response. (b) The observer is shown the CAD score. The score distribution of the two classes is displayed for reference. The observer may revise their treatment response assessment after considering the CAD score.

Table1. AUC of observers with and without CAD.

	Observers					
	1	2	3	4	5	6
No CAD	0.77 ± 0.04	0.70 ± 0.05	0.74 ± 0.04	0.73 ± 0.04	0.76 ± 0.04	0.79 ± 0.04
With CAD	0.81 ± 0.04	0.73 ± 0.04	0.78 ± 0.04	0.76 ± 0.04	0.80 ± 0.04	0.81 ± 0.04

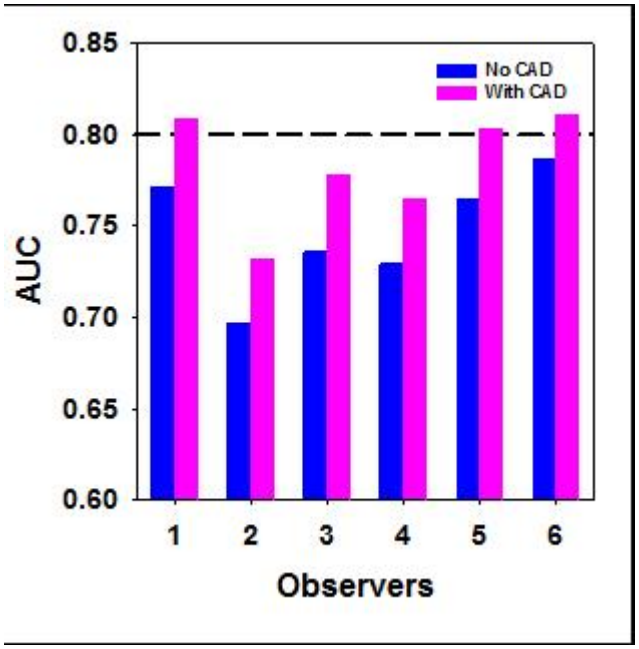


Figure 9: AUC values for the 6 observers with and without CAD. The performance of the CAD system is shown with the dashed line. The performance of all observers increased with using CAD.

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 served as a chair of PET-CT working subgroup. He serves as a co-chair of Image Analysis Performance Metrics working group. Dr. Hadjiyski is one of the Guest Editors of the Journal of Medical Imaging QIN special issue honoring Dr. Larry Clarke.

QIN Grand Challenges

We also have participated 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 it 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 participated in the data analysis and the publications preparation related to the challenges, which resulted in joint publications in the QIN Special issue of Tomography and in the Medical Physics journal. (**Publications from QIN Efforts: #18 and #19**). We also participate in the Feature Ontology Project – The Community-based Terminology Standards. We also are members of the QIN team organizing the 2018 ISBI Challenge: Lung Nodule Malignancy Prediction based on Sequential CT Scans.

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 2017 and have demonstrated our GUI, QIBC segmentation tool, and the CDSS-T decision support 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. We also will continue collecting molecular biomarker data for our bladder cancer dataset. 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, include clinical and molecular biomarkers 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. We also will continue evaluating the performance of the CDSS-S and CDSS-T as decision support to the physicians for assessment of the stage and the treatment response of bladder cancer.

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. 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).

11. 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).

12. 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).

PUBLICATIONS FROM QIN EFFORTS

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. PMCID: PMC4791536.

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(4): 1882-1896. PMCID: PMC4808067.

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(4):421-429. PMCID: PMC5241049.

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; 7(1):8738, 1-12. PMCID: PMC5562694.

5. Garapati SS, Hadjiiski L, Cha KH, Chan HP, Caoili EM, Cohan RH, Weizer A, Alva A, Paramagul C, Wei J, Zhou C. Urinary Bladder Cancer Staging in CT Urography using Machine Learning. *Med Phys*. 2017; 44(11): 5814-5823. PMCID: PMC5689080.

6. 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.

7. Cha K, Hadjiiski L, Samala RK, Chan HP, Cohan RH, Caoili EM. Comparison of bladder segmentation using deep-learning convolutional neural network with and without level sets. *Proc. SPIE Medical Imaging*, 2016; 9785: 978512-1,7.

8. Garapati SS, Hadjiiski L, Cha K, Chan HP, Caoili EM, Cohan RH, Weizer A, Alva A, Paramagul C, Wei J, Zhou C. Automatic staging of bladder cancer on CT urography. *Proc. SPIE Medical Imaging*, 2016; 9785: 97851G-1,6.

9. 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.

10. Cha K, Hadjiiski L, Chan HP, Samala RK, Cohan RH, Caoili EM, Paramagul C, Alva A, and Weizer AZ. Bladder Cancer Treatment Response Assessment using Deep Learning in CT with Transfer Learning. *Proc. SPIE Medical Imaging*, 2017; 10134: 1013404-1,6.

11. Gordon M, Hadjiiski L, Cha K, Chan HP, Samala R, Cohan RH, Caoili EM. Segmentation of inner and outer bladder wall using deep-learning convolutional neural network in CT urography. *Proc. SPIE Medical Imaging*, 2017; 10134: 1013402-1,7.

12. Cha K, Hadjiiski L^{SA}, Chan HP, Cohan R, Caoili E, Samala RK, Shampain K, Meyer N, Barkmeier D, Woolen S, Weizer A, Alva A. Observer Performance Study for Bladder Cancer Treatment Response Assessment in CT Urography with and without computerized decision support. Oral presentation at the *103rd Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA)*, Chicago, IL, Nov 26-Dec 1, 2017.

13. Hadjiiski L, Cha K, Chan HP, Cohan R, Caoili E, Weizer A, Samala RK, Alva A, Zhou C, Wei J. Analysis of Treatment Response of Bladder Cancer on CT Scans: Improved Assessment by Synergistic Combination of Radiomic Features and Clinically Estimated Feature. Oral presentation at the *103rd Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA)*, Chicago, IL, Nov 26-Dec 1, 2017.

14. Hadjiiski L, Cohan RH, Garapati S, Cha KH, Chan HP, Caoili EM, Alva A, Weizer A, Paramagul C, Wei J, Zhou C. Bladder Cancer Staging in CT Urography Using Radiomic Biomarkers. Oral and poster presentation at the *103rd Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA)*, Chicago, IL, Nov 26-Dec 1, 2017.

15. Gandikota D, Hadjiiski L, Cha KH, Chan HP, Caoili EM, Cohan RH, Weizer A, Alva A, Paramagul C, Wei J, Zhou C, "Bladder Cancer Staging in CT Urography: Effect of Stage Labels on Statistical Modeling of a Decision Support System". Presented at the *SPIE Medical Imaging, Houston*, TX, February 10- 15, 2018.

16. Gordon MN, Cha KH, Hadjiiski L, Chan HP, Cohan RH, Caoili EM, Paramagul C, Alva A, Weizer AZ, "Bladder Cancer Treatment Response Assessment with Radiomic, Clinical and Radiologist Semantic Features." Presented at the *SPIE Medical Imaging, Houston*, TX, February 10- 15, 2018.

17. Cha KH, Hadjiiski L, Chan KP, Samala RK, Cohan RH, Caoili EM, Weizer AZ, Alva A, "Bladder Cancer Treatment Response Assessment in CT Urography using Two-Channel Deep-learning Network." Presented at the *SPIE Medical Imaging, Houston*, TX, February 10- 15, 2018.

18. 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. PMCID: PMC5279995.

19. Balagurunathan Y, Beers A, Cramer JK, McNitt-Gray M, Hadjiiski L, Zhao B, Zhu J, Yang H, Yip SSF, Aerts HJWL, Napel S, Cherezov D, Cha K, Chan HP, Flores C, Garcia A, Gillies R, Goldgof D. Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST data, *Medical Physics* (in press) 2018.

U01 CA183848: Quantitative MRI models of HN cancers for physiological adaption of RT

University of Michigan (team 3)

Yue Cao, Ph.D.

INTRODUCTION

Current state-of-art therapy of poor prognosis, advanced head-and-neck cancers (HNC) (e.g., HPV-, and/or smoker), 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 a challenge due to lack of technical validation, quantitative image analysis tools, individual patient-based quality control (QC), and understanding of tumor heterogeneity and image phenotypes. To address these challenges, we have implemented an individual patient-based QC procedure to assess reproducibility of quantitative image parameters during a clinical trial. Also, we investigated tumor heterogeneity and heterogeneous response by comparing tumor characteristics of FDG uptake, blood volume distribution, and restricted water diffusion. We have two manuscripts accepted for publication. In addition, we have two manuscripts, of which one will be submitted within a week and another will be submitted to the QIN Special Issue in June. Also, we have been participating multi-site collaborative projects.

PROGRESS OVER THE PAST YEAR

We have acquired dynamic contrast enhanced (DCE) MRI, diffusion MRI and FDG PET from approximately 60 patients with poor prognosis head and neck cancers. Basic quantitative analysis of these images, producing parametric maps, is done when the patients are recruited into the clinical trial. Further analysis is conducted to address scientific questions of interest. In this funding year, we focused on 1) developing metrics to assess individual patient-based QC for quantitative images, and 2) understanding tumor image phenotypes.

§ Individual patient-based QC

It is important to have an individual patient-based QC procedure for quantitative images during a clinical trial. Importance of validating quantitative image analysis tools using DROs and validating image acquisition using physical phantoms has been recognized in recent years. However, during a clinical trial, quality and reliability of quantitative images from individual patients need to be assessed quantitatively. We developed a procedure to assess reliability of quantitative image metrics of individual patients during the clinical trial, and investigate variables affecting reliability of quantitative metrics.

In the head and neck cancer clinical trial that we are acquiring dynamic contrast enhanced (DCE) and diffusion MRI before and after 2 weeks of chemoradiaiton therapy (CRT). Blood volume (BV) maps are quantified from DCE MRI to guide clinical decision making. BV in the tumor cannot be used for assessing reproducibility since the tumor is received radiation therapy. In general, we use a large field of view (FOV), from the base of scout to shoulder, to cover primary and nodal cancers. Therefore, cerebellum is in the FOV and is not irradiated with doses > 5 Gy. Therefore, we chose cerebellum as a reference region to assess reproducibility of BV. Volumes of interest in cerebellum were created to calculate percentage changes in BV for each individual patient from pre-CRT to after 2 weeks of CRT, see Figure 1. Then, a repeatability coefficient (RC) of BV was calculated. We identified three patients who had less reproducibility in BV than other patients with 95% confidence using the RC. This analysis creates a warning signal to us to re-check the data. Interesting, there was no obvious quality issue related to the three DCE datasets. In the clinical trial, the BV maps from the two patients who had percentage changes in BV far away from the reproducible range of BV defined by RC were re-scaled based upon the group mean of BV in the cerebellum VOI.

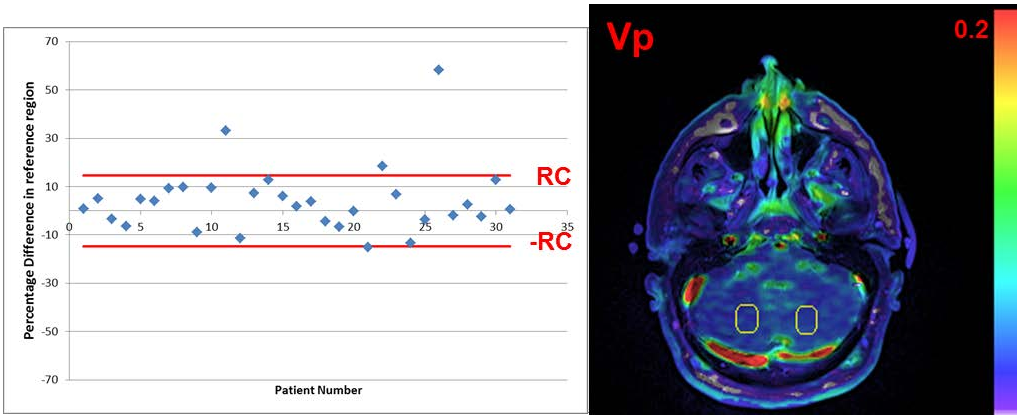


Figure 1: Left: Percentage differences of BV between pre-RT and 2 weeks after starting RT in the reference regions from 31 patients. Red lines represent the reproducibility range of BV defined by RC. Right: Plasma volume maps, Vp, (color coded) quantified from DCE MRI and two VOIs placed in the cerebellum for assessment of reproducibility. Vp can be converted to BV by accounting for hematocrit.

We further investigate whether reproducibility of arterial input function (AIF) peak is a major factor affecting reproducibility of BV. We found that reproducibility of AIF peak is four times worse than reproducibility of BV, see Figure 2. Two of the three patients who had percentage changes in BV beyond the reproducibility range had reproducible AIFs. Furthermore, one of the two patients who had worse reproducibility in AIF peak (orange circles) had reproducible BV. These data suggest that reproducibility of AIF peak may not be the major factor affecting reproducibility of BV.

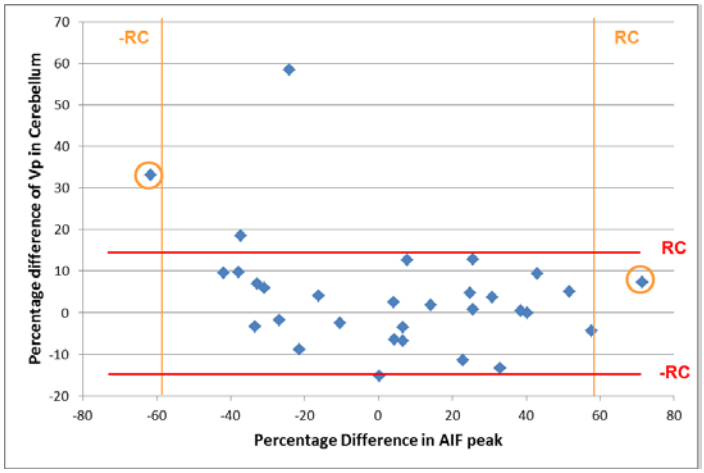


Figure 2: Scatter plot of percentage changes in BV vs percentage changes in AIF peak. Red line represents RC of Vp and orange lines depict RC of AIF peak.

We will update these data to submit a manuscript to the QIN Special Issue in *Tomography* for publication.

Spatial correspondence among quantitative high-risk image biomarkers

FDG PET, DCE-MRI and diffusion MRI each identify unique risk factors for treatment outcomes in head-and-neck cancer (HNC). Clinical trials in HNC largely rely on a single imaging modality to define targets for radiation boosting. We investigated spatial correspondence of FDG uptake, perfusion and apparent diffusion coefficient (ADC) in HNC and their response to chemoradiation therapy (CRT), and to determine implication of this overlap or lack thereof for adaptive boosting.

Forty patients with HNC enrolled in a clinical trial had FDG-PET/CT pre-CRT, and DCE and DW MRI scans pre and during CRT. Gross tumor volume (GTV) of primary tumor was contoured on post-Gd T1-weighted images. Tumor subvolumes with high FDG uptake, low BV, and low ADC were created by using previously established thresholds. Spatial correspondences between subvolumes were analyzed using Dice coefficient and between each pair of image parameters at voxel-level were analyzed by Spearman's rank correlation coefficient. Prior to CRT, median subvolumes of high FDG, low BV and low ADC relative to primary GTV were 20%, 21% and 45%, respectively. Voxel-by-voxel correlation analysis yielded Spearman's correlation coefficients between BV and ADC from -0.47 to 0.22, between BV and FDG from -0.08 to 0.59, and between ADC and FDG from -0.68 to 0.25, see Figure 3. Dice coefficients between subvolumes of FDG and BV, FDG and ADC, and BV and ADC were 10%, 46%, and 15%, respectively. The union of the three parameters was 64% of GTV. The union of the subvolumes of BV and ADC was 56% of GTV pre-CRT, but reduced significantly by 57% after 10 fractions of RT.

This analysis indicates that high FDG uptake, low BV and low ADC as imaging risk biomarkers of HNC identify largely distinct tumor characteristics. A single imaging modality may not define the boosting target adequately.

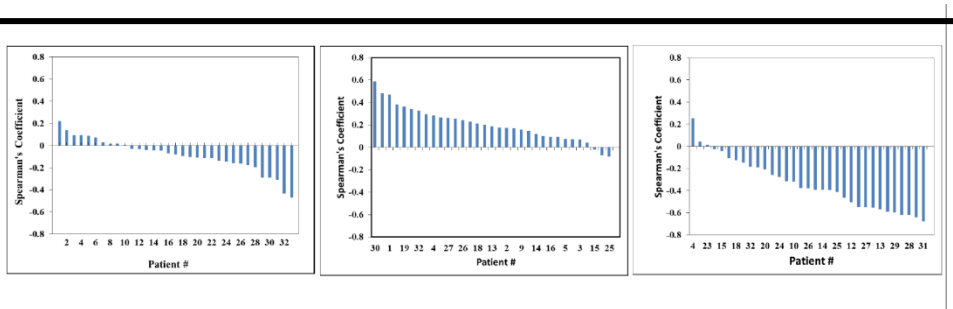


Figure 3: Waterfall plots of Spearman's rank correlation coefficients between BV and ADC (left), between BV and FDG uptake (middle), and between ADC and FDG uptake (right).

This work was presented in ASTRO 2017. A paper is accepted for publication in Int J Rad Onc Biol Phys. Also, this work is invited to be present in the NCI QIRT working group.

§ Other Works

Restricted diffusion model

We developed a restricted diffusion model that explicitly considers restricted diffusion of intracellular water in cells. We combined this model with bi-polar diffusion gradients that reduces eddy current effects on diffusion weighted images. This model estimates four parameters, cell radius (R), diffusivities of intracellular (D_{in}) and extracellular (D_{ex}) water, and the fractional volume of intracellular water (V_{in}). We applied this model to high b-values diffusion weighted images in 30 patients with glioblastoma (GBM) to assess whether this technique can differentiate GBM from surrounding edema and normal brain tissue. We found that three parameters (R, V_{in} and D_{ex}) were significantly different between tumor, and edema, normal white matter and grey matter, see Figure 4. Also, for comparison, we implemented mono-exponential and bi-exponential diffusion models. However, both mono-exponential and bi-exponential diffusion models could not robustly differentiate GBM from edema and normal tissue.

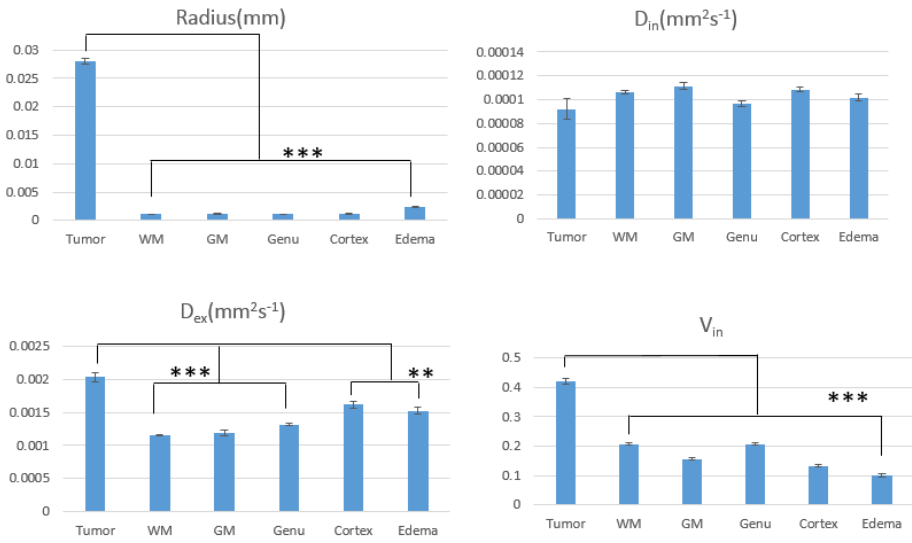


Figure 4: Bar graphs of estimated parameters of *R* (top left panel), D_{in} (top right panel), D_{ex} (bottom left panel), and V_{in} (bottom right panel) in all tissue types using the restricted diffusion model. **: $0.001<p<0.01$; ***: $p<0.001$. Error bar is SEM.

An abstract was submitted to ISMRM 2018. A manuscript will be submitted within a week.

COLLABORATIONS WITHIN THE NETWORK

We have participated in several collaborative projects, including DCE AIF challenges of part I and part II led by Huang, diffusion quantification changes led by Newitt, T1 measurement challenges led by Bane and DSC challenges led by Schmainda. These collaborative projects have led to three papers and one manuscript under review.

We are participating in a new prostate diffusion imaging challenge led by LaViolette.

We are proposing a collaborative project with Buatti to evaluate our quantitative image analysis tool, called *imFIAT*.

PLANS FOR NEXT YEAR

1. Image data collection. We will continue collecting DCE and diffusion MRI, and FDG PET from the patients with head and neck cancers who are enrolled into two clinical trials. We have acquired image data from approximately 60 patients from the first trial. We will continue acquiring image data from this trial. There is a new trial that will start patient recruitment in March 2018. This provides us an opportunity to acquire new data to validate what we developed quantitative image tools in the first clinical trial.

2. We will finish the analysis of individual patient QC of quantitative images during a clinical trial and complete a manuscript.
3. We will explore a population based analysis of quantitative image parameters in head and neck tumors to assess whether we can learn patterns from the data to differentiate patients who have disease progression from those who have disease control.
4. We will test the utility of quantitative images for assessing normal tissue toxicity after radiaiton, by analyzing whether early changes in quantitative BV in normal neck structures are correlated with dysphagia.

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.

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.

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.

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.

17. O. Bane, S. Hectors, M. Wagner, L. R. Arlinghaus, M. Aryal, **Y. Cao**, T. L. Chenevert, F. Fennessy, W. Huang, N. Hylton, J. Kalpathy-Cramer, K. E. Keenan, D. Malyarenko, R. Mulkern, D. Newitt, S. E. Russek, K. F. Stupic, A. Tudorica, L. Wilmes, T. Yankeelow, Y. Yen, M. Boss and B. Taouli. Accuracy, Repeatability and Interplatform reproducibility of T1 quantification methods used for DCE-MRI: results from a multicenter phantom study. *Magnetic Resonance in Medicine*, Sept 14, (Epub ahead of print) 2017.

18. David C. Newitt, Dariya Malyarenko, Thomas L. Chenevert, C. Chad Quarles, Laura Bell, Andriy Fedorov, Fiona Fennessy, Michael A. Jacobs, Meiyappan Solaiyappan, Stefanie Hectors, [f] Bachir Taouli, Mark Muzi, Paul Kinahan, Kathleen M Schmainda, Melissa A Prah, Erin N. Taber, Christopher Kroenke, Wei Huang, Lori R. Arlinghaus, Thomas E. Yankeelov, Yue Cao, Madhava Aryal, Yi-Fen Yen, Jayashree Kalpathy-Cramer, Amita Shukla-Dave, Maggie Fung, Jiachao Liang, Michael Boss, Nola Hylton. Multi-site concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. *Journal of Medical Imaging* 5(1): 011003, (2017). doi: 10.1117/1.JMI.5.1.011003.

19. Daekeun You, Michelle Kim, Madhava Aryal, Hemant Parmar, Morand Piert, Theodore Lawrence, and Yue Cao. Tumor Image-Signatures and Habitats: A Processing Pipeline of Multi-modality Metabolic and Physiological Images. *Journal of Medical Imaging* 5(1): 011009, (2017). doi: 10.1117/1.JMI.5.1.011009.

19. Feifei Teng, Jae Lee, Choonik Lee, Madhava Aryal, Peter Hawkins, Michelle Mierzwa, Avraham Eisbruch, Yue Cao. Adaptive boost target definition in high-risk head and neck cancer based on multi-imaging risk biomarkers. Int J Rad Onc Biol Phys (ePrint), 2017.

20. Yuan Li, Michelle Kim, Theodore S Lawrence, Hemant Parmar, Yue Cao. The restricted diffusion model for differentiation of tumor from normal tissue in glioblastoma. ISMRM 2018 (abstract).

U01CA187947: Computing, Optimizing, and Evaluating Quantitative Cancer Imaging Biomarkers

Stanford University Department of Radiology (team 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 1-3 were to: (labels C.n.m refer to our grant proposal and the Gantt Chart (Figure 1)):

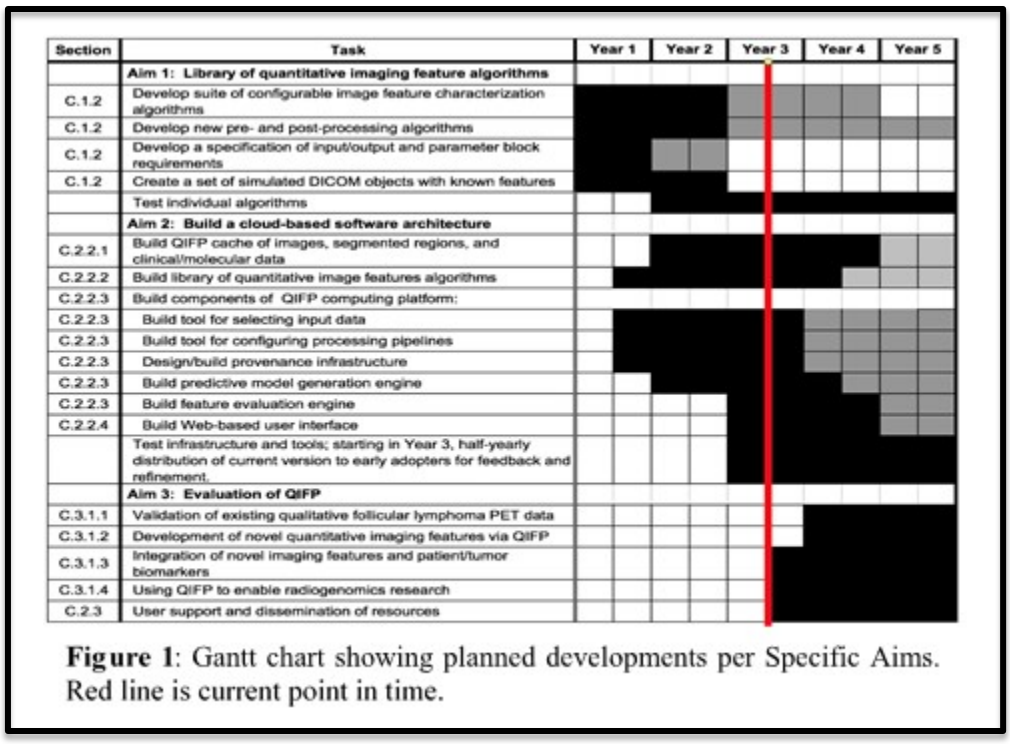


Figure 1: Gantt chart showing planned developments per Specific Aims. Red line is current point in time.

AIM 1: Library of quantitative imaging feature algorithms:

C.1.2 Develop suite of configurable image feature characterization algorithms: *Aims complete with additional algorithms and refinement ongoing.* Examples:

- We have finished the initial version of the "Quantitative Image Feature Engine (QIFE)" (published manuscript #13 below), packaged it in a Docker container, and deployed it to Dockerhub for use by others, and to the QIFP. It is fully configurable via a graphical user interface or via an unloadable configuration file.
- We are in the process of adding additional features to the QIFE, including Laws texture features and a suite of 2D features for slice-by-slice analysis.
- We have made another feature characterization package, Pyradiomics (van Griethuysen et al., Cancer Research, 77(21), 2017), available on the QIFP.
- We have also built a SIFT feature extractor, packaged it in a Docker container, and made it available on the QIFP.
- At least one additional feature engine from a Stanford lab is being ported, and additional feature engines may follow.

C.1.2 Begin the development of new pre- and post-processing algorithms: *Aims complete with additional algorithms and refinement ongoing.*
Examples:

- We have completed and deployed a "delta-feature" post-processing module, which takes features computed from different scans at different times and computes difference and ratio features.
- We have completed and deployed a post-processing module that integrates semantic features (stored as AIM files) with quantitative image features computed by any of our quantitative feature engines.
- We have completed and deployed a pre-processing module for indexing DICOM files prior to processing with the Pyradiomics feature engine.
- We have completed and deployed a pre-processing module that converts DICOM image and segmentation data to the NiFTI format, enabling use of Pyradiomics on the QIFP.

C.1.3 Complete the specification of input/output and parameter block requirements; *Done; further work possible but not expected at this time.*

C.1.4 Begin the development of a set of simulated DICOM objects with known features; *in progress: We now have code that can generate these phantoms as DICOM data, we have visualized them using ePAD, and processed them using the QIFE on the QIFP (Fig. 2). However, we want to make the code that produces these phantoms user-friendly and publicly available.*

AIM 2: Build a cloud-based software architecture: As reported last year, the architecture specification and working framework is complete.

C.2.2.1 Build QIFP cache of images, segmented regions, and clinical molecular data: *reported last year complete; ahead of schedule*

C.2.2.2 Build a library of quantitative image feature algorithms: Complete: all algorithms described in AIM 1 are available on the QIFP.

C.2.2.3 Build a tool for selecting input data: *reported last year, complete*

C.2.2.3 Build a tool for configuring processing pipelines; *in process; this can now be done by our software developer, and we decided to defer implementation for users until it can be made available in the new web-based user interface (see C.2.2.4 below).*

C.2.2.3 Design/build provenance architecture: *in process. The QIFE outputs a log file and a feature dictionary that can be used to reconstruct the process. But this is not yet implemented for other feature algorithms or in a user-friendly manner.*

C.2.2.3 Build predictive model engine; *reported last year, LASSO algorithm, implemented as a Docker container, is complete ahead of schedule*

C.2.2.4 Build Web-based user interface; *begun ahead of schedule, but then delayed due to staffing leave of absence. Many features are available, however, and all screenshots in the Figures were generated with the new interface.*

In addition to the above:

- we have made a QIFP working prototype available and demonstrated it art RSNA 2017
- we have worked with one commercial entity and other QIN users to get feedback on the operation and improvement of the QIFP.

Current 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:

- 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 (see Fig. 3 for example workflows):

- Process DSD+DSOs using our 3D feature computation engine (QIFE), Pyradiomics, and/or SIFT, 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 following with 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.
- Process multi-phasic CT to produce delta-radiomic features, which can be followed by LASSO for model building or testing of an existing predictive model.

COLLABORATIONS WITHIN THE NETWORK

- Participated in tool interoperability demonstration at the 2017 Face-to-Face meeting: shared QIFE Docker container and demonstrated interoperability with pipeline developed at Mayo (Bradley Erickson).
- Active member of IAPM, PET-CT, and BIDS working groups.
- Collaborated with other QIN PIs on several papers, e.g., published manuscripts #2, #5, and #21 and presentations #12 and #15, below.
- Collaborating with Keyvan Farahani on QIN benchmarks project
- Collaborating with Keyvan Farahani on QIN Challenge Task Force
- Collaborating with QIN members McNitt-Gray, Kalpathy-Cramer, Hadjiyski et. al on Feature Ontology project

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.1.4 Finish the development of a set of simulated DICOM objects with known features. *We will complete a set of digital phantoms and make them available on the QIFP and TCIA for testing and radiomics feature development.*

C.2.2.2 Build a library of quantitative image feature algorithms: *We will continue to engage with the QIN community, specifically the BIDS Working Group, to test interoperability of our Dockerized processing modules, and deploy additional feature engines and machine learning engines as they become available.*

C.2.2.3 Complete and deploy on QIFP a tool for user-configuration of processing pipelines: *We will complete the graphical use interface on our new web-based interface for user configuration of workflows*

C.2.2.3 Design/build provenance architecture: *We will integrate the run-time parameters and other data we are already collecting into a user-friendly system for recording experimental provenance and re-running completed workflows.*

C.2.2.4 Build Web-based user interface: *We will complete this design and a working prototype implementation.*

C.3 We will continue working with colleagues who are using QIFP in radiomics research. For example:

- Drs. Kotharty and Gevaert (Stanford): Predicting microvascular invasion in HCC using quantitative image features
- Dr. Patel (Stanford): Radiomics in pancreatic and liver cancer
- Dr. Nair (Moffitt Cancer Center): Radiomics analysis of indeterminate pulmonary nodules using CT and PET-CT
- Dr. Itakura (Stanford): Radiomics of pancreatic cancer
- Dr. Evens (now at Rutgers): Radiomics of follicular lymphoma (previously collected data, ECOG 2408)
- Drs. Plevritis, Gevaert, Ann Leung (Stanford): Radiogenomics of NSCLC

- Dr. Rubin (Stanford, U01 CA190214, *Qualification and Deployment of Imaging Biomarkers of Cancer Treatment Response*): This project, while internal to Stanford, is another opportunity for QIFP synergy. It leverages SWOG 0518, *(Prospective Randomized Comparison Trial of Depot Octreotide Plus Interferon Alfa-2b Versus Depot Octreotide Plus Bevacizumab in Patients with Advanced Carcinoid Tumors)*.

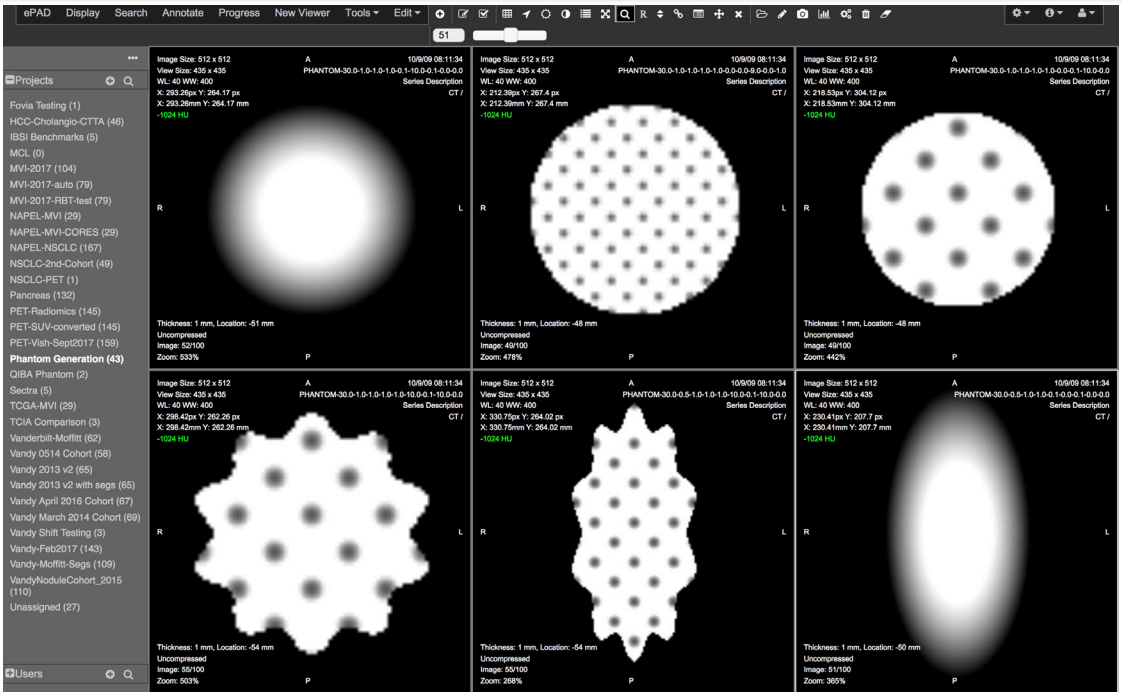


Figure 2: ePAD display of selection of DICOM images of cross-sections of 3D digital radiomics phantoms, exhibiting different shapes, textures, and edge sharpness. These and other images stored on an ePAD system are directly selectable from within the QIFP user interface, as are images from TCIA and any configured PACS system.

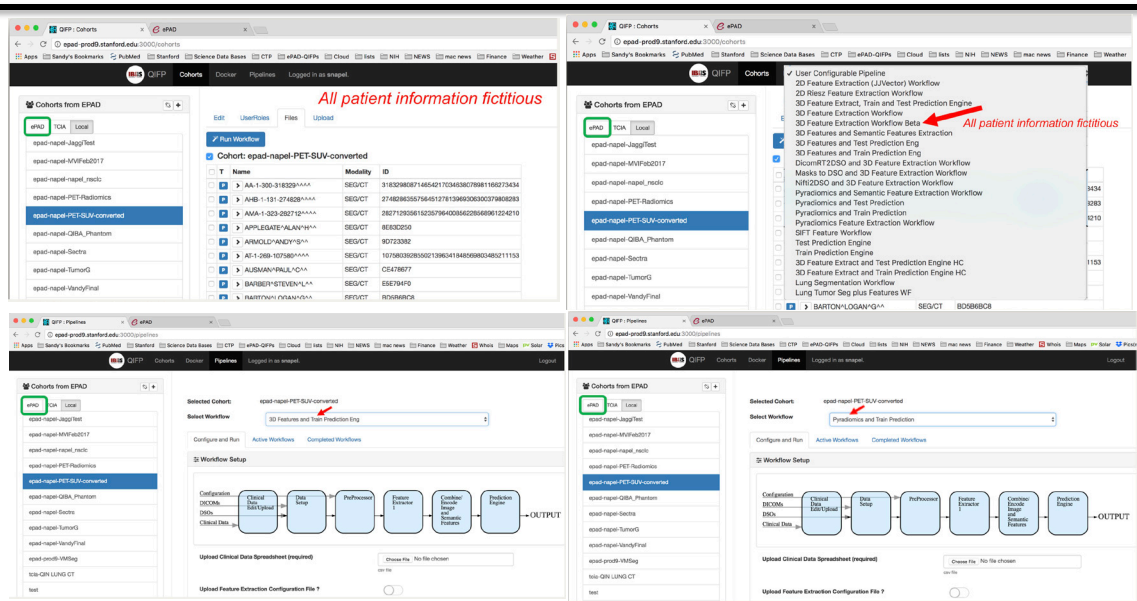


Figure 3: Screenshots of new QIFP user interface, showing: upper left: cohort selection; upper right: workflow selection; lower left: workflow to compute 3D features and train a prediction engine using clinical data; (lower right) workflow to compute 3D features using externally-develop module (Pyradiomics) and train a precdiction engine using clinical data.

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,” Tomography 2(4) 283-292, 2016. PMID: 28612050; PMCID: PMC5466872.

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,” Tomography 2(4) 430-437, 2016. PMID: 28149958, PMCID: PMC5279995.

6. 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,” Sci Rep. 2017 Jan 31;7:41674, PMID: 28139704, PMCID: PMC5282551.

7. 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,” IEEE J Biomed Health Inform 2017 Jan 21(1):48-55, PMID:27893402, PMCID: PMC5293622.

8. 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,” Med Image Anal. 2017 Jan 37:46-55. PMID: 28157660, PMCID: PMC5393306.

9. 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," Oncotarget 10;8(32):52792-52801, 2017. PMID:28538213. PMCID: PMC5581070.

10. J. Wu, B. Li, X. Sun, G. Cao, D. L. Rubin, MD, S. Napel, D. M. Ikeda, A. W. Kurian, R. Li, "Heterogeneous enhancement patterns of tumor-adjacent parenchyma on MRI are associated with dysregulated signaling pathways and poor survival in breast cancer," Radiology 205(2):401-413, 2017. PMID:28708462. PMCID: PMC5673053.

11. S. Bakr, S. Echegaray, R. Shah, A. Kamaya, J. Louie, S. Napel, N. Kothary, O. Gevaert, "Non-invasive radiomics signature based on quantitative analysis of computed tomography images for prediction of microvascular invasion in hepatocellular carcinoma: a pilot study," J Med Imaging (Bellingham). 2017 Oct;4(4):041303. doi: 10.1117/1.JMI.4.4.041303. Epub 2017 Aug 21. PMID:28840174. PMCID: PMC5565686.

12. M. Zhou, A. N. C. Leung, S. Echegaray, J. B. Shrager, K. C. Jensen, G. G. Berry, S. K. Plevritis, D. L. Rubin, S. Napel, O. Gevaert, "Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic Implications," Radiology 286(1):307-315, 2018. PMID:28727543.

13. 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," J Digit Imaging, 2017 Oct 6 doi: 10.1007/s10278-017-0019-x. [Epub ahead of print] PMID: 28993897. PMC-Journal in Process.

14. M. Zhou, J. Scott, B. Chaudhury, L. Hall, D. Goldgof, K. W. Yeom, M. Iv, Y. Ou, J. Kalpathy-Cramer, S. Napel, R. Gillies, O. Gevaert, R. Gatenby, "Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors and Machine-learning Approaches," AJNR Am J Neuroradiol. 2017 Oct 5. doi: 10.3174/ajnr.A5391. [Epub ahead of print] PMID:28982791.

15. Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, B. J. Erickson, "Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions," J Digit Imaging. 2017;30(4):449-59. doi: 10.1007/s10278-017-9983-4. PMID:28577131. PMCID:PMC5537095.

16. I. Banerjee, M. C. Chen, M. P. Lungren, D. L. Rubin, "Radiology Report Annotation using Intelligent Word Embeddings: Applied to Multi-institutional Chest CT Cohort," J Biomed Inform. 2017;23(17):30257-5. PMID:29175548.

17. I. Banerjee, S. Malladi, D. Lee, A. Depeursinge, M. Telli, J. Lipson, D. Golden, D. L. Rubin, "Assessing treatment response in triple-negative breast cancer from

quantitative image analysis in perfusion magnetic resonance imaging," J Med Imaging (Bellingham). 2018;5(1):011008. doi: 10.1117/1.JMI.5.1.. Epub 2017 Nov 2. PMID:29134191. PMCID:PMC5668126.

18. K. Farahani, J. Kalpathy-Cramer, T.L. Chenevert, D. L. Rubin, J. J. Sunderland, R. J. Nordstrom, J. Buatti, N. Hylton, "Computational Challenges and Collaborative Projects in the NCI Quantitative Imaging Network," Tomography. 2016;2(4):242-9. doi: 10.18383/j.tom.2016.00265. PMID:28798963. PMCID:PMC5548142

19. K. Grain, T. T. Liu, A S. Achrol, E. O. Paull, Y. Newton, S. D. Chang, G. R. Harsh, S. P. Coredero, D. L. Rubin, J. M. Stuart, "Revealing cancer subtypes with higher-order correlations applied to imaging and omics data," BMC Med Genomics. 2017;10(1):20. doi: 10.1186/s12920-017-0256-3.PMID: 28359308. PMCID:PMC5374737.

20. A. Hoogi, A. Subramaniam, R. Veerapaneni, D. L. Rubin, "Adaptive Estimation of Active Contour Parameters Using Convolutional Neural Networks and Texture Analysis," IEEE Trans Med Imaging. 2017;36(3):781-91. doi: 10.1109/TMI.2016.2628084. Epub 2016 Nov 11. PMID:28113927. PMCID:PMC5510759.

21. Y. Balagurunathan, A. Beers, J. Kalpathy-Cramer, M. McNitt-Gray, L. Hadjiiski, B Zhao, J. Zhu, H. Yang, S.F. Yip, H. J.W.L. Aerts, S. Napel, D. Cherezov, K. Cha, H-P Chan, C. Flores, A. Garcia, R. Gillies, D. Goldgof, "Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST Data," *in press, Medical Physics, January 2018.*

22. J. Wu, G. Cao, X. Sun, J. Lee, D.L. Rubin, S. Napel, A.W. Kurian, B. Daniel, R. Li, "Intratumoral spatial heterogeneity by perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy," *in press, Radiology, December 2017.*

Submitted Manuscripts:

1. W. Zhang, G. Bouchard, A. Yu, M. Shafiq, M. Jamali, J. Shrager, K. Ayers, S. Bakr, A. Gentles, Max. Diehn, A. Quon, R. West, V. S. Nair, M. van de Rijn, S. Napel, S. K. Plevritis, "PET-FDG uptake in human lung adenocarcinoma associated with invasion mediated through GFPT2-expressing cancer-associated fibroblasts," accepted pending revision, Cancer Research, September 2017.

2. S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A.N.C. Leung M. Kadoch, J. Shrager, A. Quon, D.L. Rubin, S. K. Plevritis*, S. Napel*, "A Radiogenomic Dataset of Non-Small Cell Lung Cancer," submitted to Nature Scientific Data, Dec. 2017.

3. S. Napel, W. Mu, B. Jardim-Perassi, H. J. W. L. Aerts, R. Gillies, "Quantitative Imaging of Cancer in the Post-genomic Era: Radio(geno)mics, Deep Learning and Habitats," submitted to Cancer Research, January 2018.

4. J. Wu, X. Li, X. Teng, D. L. Rubin, S. Napel, B. L. Daniel, R. Li, "MR imaging and molecular features associated with tumor infiltrating lymphocytes in breast cancer**,"** *submitted to Radiology, February 2018.*

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.

7. E. Lee, M. Zhou, O. Grove, Y. Balagurunathan, S. Echegaray, R. Gillies, N. Gamboa, B. Murmann, S. Napel, S. Wong, O. Gevaert, "A deep learning framework to predict survival from medical images of lung cancer patients," Neural Information Processing Society, December 5-10 2016, Barcelona, Spain.

8. W. Zhang, G. Bouchard, A. Yu, M. Shafiq, M. Jamali, J. Shrager, K. Ayers, V. S. Nair, A. Gentles, M. Diehn, A. Quon, S. Napel, S. K. Plevritis, "FDG uptake in human lung adenocarcinoma associated with invasion through the hexosamine biosynthesis pathway, AACR Annual Meeting, Washington D.C., April 2017.

9. M. Shafiq, W. Zhang, A. Gentles, K. Ayers, V.S. Nair, J. Shrager, C. Hoang, O. Gevaert, S. Napel, S. Plevritis, "Using Tissue Gene Expression to Predict Survival Following 'Curative' Surgical Resection in Lung Adenocarcinoma, American Thoracic Society Annual Meeting, Washington D.C., May 19-24, 2017.

10. I. Cetin, G. Sanroma, S.E. Petersen, S. Napel, O. Camara, M-A G. Ballester1, K. Lekadir, A Radiomics Approach to Computer-Aided Diagnosis with Cardiac Cine-MRI, Medical Image Processing and Computer-Aided Intervention STACOM (Statistical Atlases and Computational Modeling of the Heart), September 10-14, 2017.

11. S. Napel, D. L Rubin, S. John, D. Gude, S. Echegaray, S. Bakr, D. Gude, et al. "The Quantitative Image Feature Pipeline (QIFP): Automated Radiomic Feature Extraction to Derive Associations with and Prediction of Clinical Variables from Image Features," Radiological Society of North America 103rd Scientific Sessions, December 2017.

12. J. Kalpathy-Cramer, B. Zhao, D. Goldgof, S. Napel, D. L. Rubin, M. F. McNitt-Gray, et al, "Standardizing Radiomic Feature Descriptions for Quantitative Imaging: A Preliminary Report of the Cooperative Efforts of the NCI's QIN PET-CT Subgroup," Radiological Society of North America 103rd Scientific Sessions, December 2017.

13. S. Bakr, S. Mattonen, PhD; A. L. Garcia, S. Antic, Y. Balagurunathan, T. Atwater, H. Chen, O. Gevaert, R. Walker, M. Schabath, R. Gillies, P. Massion, V. Nair, S. Napel, "A Size-independent Radiomics Model for Classification of Indeterminate Pulmonary Nodules Seen at CT," Radiological Society of North America 103rd Scientific Sessions, December 2017.

14. S. Mattonen, S. Bakr, G. A. Davidzon, V. S. Nair, S. Napel, "Positron Emission Tomography (PET) Tumor Penumbra Radiomics For Prediction of Survival in Non-Small-Cell Lung Cancer: A Pilot Study*,"* Radiological Society of North America 103rd Scientific Sessions, December 2017.

15. A. Fedorov, L. J. O'Donnell, D. L. Rubin, D. A. Clunie, D. Flade, M. Nolden, et al., "DICOM4QI demonstration and connectathon: Structured communication of quantitative image analysis results using the DICOM standard," Scientific Exhibit in the Quantitative Imaging Reading Room of the Future (QIRR), Radiological Society of North America 103rd Annual Meeting, December 2017.

16. D. L. Rubin, S. John, 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 103rd Annual Meeting, December 2017.

17. S. Napel, "Image Feature Computation and Considerations," in Refresher Course, Radiomics Mini-Course: From Image to Omics, Radiological Society of North America 103rd Annual Meeting, December 2017.

18. D. L. Rubin, "Image Annotation and Semantic Labeling," in Refresher Course, Radiomics Mini-Course: From Image to Omics, Radiological Society of North America 103rd Annual Meeting, December 2017.

19. D. L. Rubin, "Machine Learning and Radiomics in Practice: Tools and Case Example," in Refresher Course, Platforms and Infrastructures for Accelerated Discoveries in Machine Learning and Radiomics, Radiological Society of North America 103rd Annual Meeting, December 2017.

20. S. A. Mattonen, G. A. Davidzon, S. Bakr, M. Vasanawala, G. Horng[5], S. Napel, V. S. Nair, "Positron Emission Tomography (PET) Tumor Penumbra Imaging Features Predict Outcome in Non-Small Cell Lung Cancer," *accepted for oral presentation, American Thoracic Society Annual Meeting, San Diego, CA, May 18-23 2018.*

U01 CA181156: Quantitative CT Imaging for Response Assessment when Using Dose Reduction Methods

David Geffen School of Medicine at UCLA

Michael McNitt-Gray, Ph.D.
Matthew Brown, Ph.D.
Jonathan Goldin, M.D. Ph.D.
Grace Kim, Ph.D.

INTRODUCTION

CT continues to be widely used for assessing response to therapy in many clinical trials settings. There have been significant developments that 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 and especially radiomics features, are unclear and need to be investigated before they are deployed in clinical trials. That is, while many radiomics features have been shown to be predictive or useful in assessing response to therapy, some features might be significantly affected by the acquisition and reconstruction conditions under which the images were acquired. For example, when reducing the dose in the CT acquisition can increase the image noise which may affect some radiomics feature values such as texture [1] while not having a tremendous effect on size related features (as shown previously in [2].

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.

PROGRESS OVER THE PAST YEAR

Over the past year, we have made progress on a number of aspects of our project. These are reported below.

§ Extended Infrastructure for Software and Data Collection Efforts

Extension of Reconstruction Tools

During the current project period, we have extended our toolsets and infrastructure capabilities in a few ways. The first is that we have developed some offline iterative reconstruction tools that are based on the Iterative Coordinate Descent (ICD) method that allows us to reconstruct raw projection data using a model-based iterative reconstruction method in a way that is similar to the methods the manufacturers use in their clinical scanners. This added step provides an strong complement to our previous capabilities of simulating reduced dose scans (described in [1]) as well as being able to reconstruct offline using the (more conventional) weighted Filtered Backprojection (wFBP) tools that we have developed previously [3]. When taken as a whole, this suite of tools allow us to take raw projection data that has been collected (see below) and:

(a) Simulate reduced dose conditions; and
(b) Reconstruct original raw projection data *and* simulated reduced dose scan data under a variety of slice thickness and kernel conditions using conventional wFBP; and
(c) Reconstruct original raw projection data *and* simulated reduced dose scan data under a variety of slice thickness and kernel conditions using an iterative reconstruction (ICD) method.

The end result is that when we have raw projection data available, we now have tremendous flexibility to create image datasets that represent an extremely wide range of acquisition and reconstruction conditions. This is extremely useful in our assessments of the robustness of quantitative imaging features over a range of acquisition and reconstruction conditions and other investigations described below.

Extension of the Pipeline Capabilities

In addition to the extension of the simulated reduced dose scan tools and the reconstruction algorithm tools, we have extended our underlying infrastructure tools 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 and ICD reconstruction engines have been organized into a fully automated pipeline (**Figure 1**) that takes the raw data files and creates the desired set of image datasets that represent a range of dose levels, slice thicknesses and reconstruction conditions (wFBP and ICD). These

datasets have been used in large-scale investigations into the effects of acquisition and reconstruction parameters on quantitative imaging tasks.

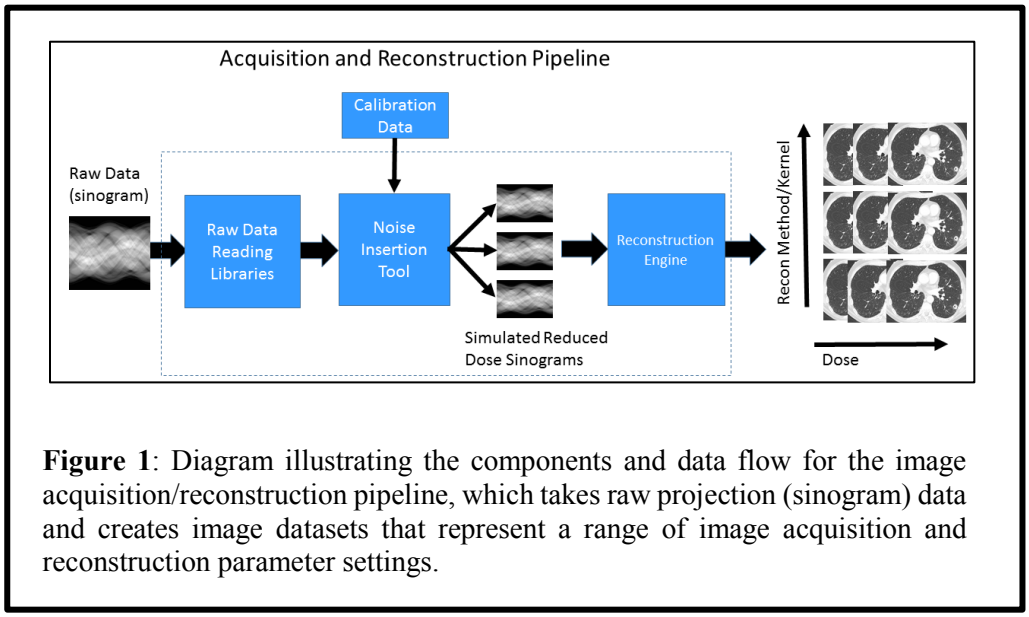


Figure 1: Diagram illustrating the components and data flow for the image acquisition/reconstruction pipeline, which takes raw projection (sinogram) data and creates image datasets that represent a range of image acquisition and reconstruction parameter settings.

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. This past year we extended these capabilities by 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. Based on our past experience, this new high throughput, batch mode processing allows us more than 2 orders of magnitude increase in throughput, which provide a much broader exploration of the acquisition and reconstruction parameter space than was previously achievable.

Raw Data Collection Efforts

During the past year we have also extended our collection of raw projection data. We are now collecting data from 7 different outpatient scanners (all from Siemens). The cases we are collecting are primarily focused on either oncology or screening cases. We have collected nearly 1000 low dose lung cancer screening studies now (956 at last count; we expect to exceed 1000 cases next month). With this raw projection data inventory and the tools described above, we will be able to create image datasets that represent a wide range of acquisition and reconstruction conditions required for our studies into the robustness of quantitative imaging features.

§ Image DeNoising Software Efforts

One of the primary effects of reducing the radiation dose in CT is to increase the noise in the resulting reconstructed image. While Iterative Reconstruction methods seek to reduce the noise while preserving the structural integrity (e.g. edges) of the image data, another approach to reducing noise is to apply algorithms that operate on the reconstructed image data itself. While methods to reduce noise (such as Gaussian smoothing) are well known, the challenge for these methods is also to reduce noise while preserving the edge content of the image. In quantitative imaging, the additional challenge is to determine if these denoising software tools can preserve quantitative imaging features over a range of acquisition and reconstruction conditions.

To begin investigations into the effects of these denoising methods, we have first added to our pipeline the capability of performing image denoising with several different methods including bilateral filtering [4]and BM3D denoising[5]. To date, the bilateral filtering approach has turned out to be more computationally efficient and so we have used that approach to date. We are still working on making the BM3D approach more computationally efficient so that it can be incorporated into our pipeline.

§ Lung Nodule Detection for Reduced Dose CT Scanning

Previously we have published work on the detection of lung nodules on original and simulated reduced dose scans from the NLST. In the work we have performed over the past year, we have now extended that work to cases collected as part of the UCLA clinical lung cancer screening program (as mentioned above, nearly 1000 cases to date). These cases are different from the NLST cases in that they have: (a) a scanning protocol on current CT scanners with more modern radiation dose reduction capabilities such as tube current modulation and iterative reconstruction (SAFIRE or ADMIRE from Siemens); (b) results of the interpretation of the scan that are reported in a structured environment that allows easier identification of nodules identified clinically; and (c) as a result these cases can be classified into a LungRADS classification which did not exist when NLST was being performed.

Therefore, the goal of this study was to assess the impact of dose reduction and various reconstruction parameters, on the detection of lung nodules by CAD software applied to low dose Lung Cancer Screening CT exams.

Our dataset consisted of data from 59 subjects who underwent low dose CT for lung cancer screening at our institution. Scans were performed using a 64-slice MDCT with Tube Current Modulation (TCM) according to the AAPM lung cancer screening protocols, which yielded a CTDIvol of approximately 2 mGy. For each case, image data and raw projection data were collected. A total of 71 solid nodules larger than 4mm (30 in LungRad1, 25 in LungRAD2, 11 in LungRAD3, 5 in LungRAD4) were identified by radiologists (on original scans with 1mm slice thickness and B45 kernel). Noise was added to the raw projection data to simulate reduced dose scans at 3 dose levels of 10%, 25%, and 50% (corresponding to CTDIvol of 0.2, 0.5 and 1 mGy). All original and simulated raw data were reconstructed with slice thicknesses of 0.6, 1, and 2mm and at three reconstruction kernels (smooth, medium and

sharp). Lung regions were segmented using original image data and were mapped to all other image datasets. An in-house CAD software was used to detect nodules on simulated images. Subject level mean sensitivity (in LungRADS categories) and mean false-positive for each of acquisition conditions were calculated.

Our results showed that subject level mean sensitivity values ranged between 40% to 80% for LungRAD4, 27% to 55% for LungRAD3, 22% to 50% for LungRAD2, and 10% to 53% for LungRAD1. Sensitivity was stable at all dose levels in LungRAD3 and LungRAD4 categories (even down to 10% of screening dose), with a medium kernel and either 1mm or 0.6 mm slice thickness. Mean false-positive was between 2-9 per patient with most conditions yielding < 4.

Our study concluded that CAD detection sensitivity was reasonably robust to dose, slice thickness and kernel, though the sharper kernel yielded the most variable performance. False positives were also surprisingly stable except at very high noise (low dose, thin slice, sharp kernel) conditions. This work was presented at both AAPM and RSNA conferences this past year [6], [7].

COLLABORATIONS WITHIN THE NETWORK

§ QIN Feature Challenge

The UCLA QIN team participated in the CT Image Feature Challenge (coordinated by Moffit QIN and executed with the PET-CT group). 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 was 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).[8]

§ QIN Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST data.

The UCLA QIN team also participated in the Nodule Interval Segmentation project that was again executed within the QIN by the PET-CT group. The purpose of this study was to investigate the variability in volume change estimates of pulmonary nodules due to

segmentation approaches used across several algorithms and to evaluate these effects on the ability to predict nodule malignancy.

The study used 100 patient image datasets from the National Lung Screening Trial (NLST) that had a nodule detected on each of two consecutive low dose computed tomography (LDCT) scans, with an equal proportion of malignant and benign cases (50 malignant, 50 benign). Information about the nodule location for the cases was provided by a screen capture with a bounding box and its axial location was indicated. Five participating Quantitative Imaging Network (QIN) institutions performed nodule segmentation using their preferred semi-automated algorithms with no manual correction; teams were allowed to provide additional manually corrected segmentations (analyzed separately). The teams were asked to provide segmentation masks for each nodule at both time points. From these masks, the volume was estimated for the nodule at each time point; the change in volume (absolute and percent change) across time points was estimated as well. We used the concordance correlation coefficient (CCC) to compare the similarity of computed nodule volumes (absolute and percent change) across algorithms. We used Logistic regression model on the change in volume (absolute change and percent change) of the nodules to predict the malignancy status, the area under the receiver operating characteristic curve (AUROC) and confidence intervals were reported. Because the size of nodules was expected to have a substantial effect on segmentation variability, analysis of change in volumes was stratified by lesion size, where lesions were grouped into those with a longest diameter of <8mm and those with longest diameter ≥ 8mm.

The results showed that segmentation of the nodules shows substantial variability across algorithms, with the CCC ranging from 0.56 to 0.95 for change in volume (percent change in volume range was [0.15 to 0.86]) across the nodules. When examining nodules based on their longest diameter, we find the CCC had higher values for large nodules with a range of [0.54 to 0.93] among the algorithms, while percent change in volume was [0.3 to 0.95]. Compared to that of smaller nodules which had a range of [-0.0038 to 0.69] and percent change in volume was [-0.039 to 0.92]. The malignancy prediction results showed consistent results across the institutions, the AUC using change in volume ranged from 0.65 to 0.89 (Percent change in volume was 0.64 to 0.86) for entire nodule range. Prediction improves for large nodule range (≥ 8mm) with AUC range 0.75 to 0.90 (percent change in volume was 0.74 to 0.92). Compared to smaller nodule range (<8mm) with AUC range 0.57 to 0.78 (percent change in volume was 0.59 to 0.77).

The study concluded that there is a high concordance in the size measurements for larger nodules (≥8mm) than the lower sizes (<8mm) across algorithms. We find the change in nodule volume (absolute and percent change) were consistent predictors of malignancy across institutions, despite using different segmentation algorithms. Using volume change estimates without corrections shows slightly lower predictability (for two teams). This work was recently published in Medical Physics (Med Phys. 2018 Jan 24. doi: 10.1002/mp.12766) [9]

PLANS FOR NEXT YEAR

During the next year, we will extend our work in several different ways as described below.

§ Extend Data Collection Efforts

We will continue our data collection efforts for raw projection data from our outpatient CT scanners. In addition, we will be working with our Oncology Imaging team within the Department of Radiology along with Drs. Toni Ribas and Jonathan Goldman from the Department of Oncology to identify scans of patients who are under specific clinical trials, with an emphasis on immunotherapy trials (or those who are having immunotherapy treatments as part of their clinical management).

§ Further Investigate the Effects of Mitigation Measures on Quantitative Imaging Tasks: Iterative Reconstruction and DeNoising Methods

In previous years, we have focused on investigations into the effects of acquisition and reconstruction parameters on quantitative features. In the coming year, we will extend these investigations into the effects of mitigation measures – such as limiting the imaging protocol space or image denoising – on these quantitative imaging features. More specifically, we will examine the effects on quantitative imaging features such as size (e.g. volume), density, shape and texture under several acquisition/reconstruction/denoising conditions (see table below). Using a reference imaging condition for the basis of comparison, we will examine the effects on features extracted from a lung lesion over this wide range of conditions and determine: (a) the range of conditions over which the feature value can be maintained within a specified value (say 5%) of the value obtained from the reference condition and then (b) determine the range of conditions this could be expanded to either by using iterative reconstruction techniques or other denoising methods. We fully expect this "range of acceptable conditions" to vary from feature to feature (as we have seen with results previously investigated). We believe this will inform clinical trials using quantitative endpoints (tumor volume or other radiomic features) of both: (a) the range of conditions that should be specified as "acceptable" for a clinical trial and (b) describe methods to extend that range by denoising (which could also serve as a "corrective method" if a study is performed out of protocol and the alternative is to remove the scan from consideration because it was not performed according to study protocol procedures. We think this would be a substantial contribution to clinical trials using CT and quantitative endpoints to help assess treatment response. Please see the collection of images created for this purpose that are shown in the Appendix of this report.

§ QIN Feature Ontology Effort

This effort has been initiated by the PET-CT Group and has begun only very preliminary work to date. The motivation is that medical imaging is one of the largest sources of "Big Data" in the world. Yet most of the data is in the form of unstructured objects, making these data difficult to access in an organized manner and/or utilize in any subsequent computerized analysis. Radiomics, "the high-throughput extraction of large amounts of image features from radiographic images," has been used to provide quantitative characterizations

of regions of interest in images that are the basis for the classification and prediction tasks in radiology and oncology (e.g., tumor characterization for diagnostic purposes). However, there are substantial challenges to comparing and reproducing results across sites and studies for several reasons. The purpose of this work was to begin to address one of those issues by starting to standardize the terminology used to describe imaging features. This will facilitate the specific definitions of quantitative features, which will allow direct comparisons between features collected by different investigators or in different datasets or even domains and will ultimately increase the confidence in integrating these quantitative features into clinical practice in the future.

Over the coming year, the group will investigate several key aspects, starting with:

1. Identifying the mission of this effort. Specifically, this will involve performing a gap analysis to determine what other entities in our community have done and what activities are underway and to identify how we can interact with those communities and not duplicate efforts, but provide complementary activities that are within the QIN's expertise.
2. Define different types/definitions of Region of Interest (ROIs) that are widely used. This will be very helpful as a pre-requisite for many feature definitions (that is, many times the feature calculation depends heavily on the definition of the region being used to describe the object of interest)
3. The group has also had preliminary discussions about possible Phantoms (Digital and Physical), Reference Datasets and a testing procedure (including possibly a suggested progression). These discussions have included possibly progressions where one can start with analytical phantoms, move to physical phantoms (or rather digital image representations of physical phantoms) and then to patient datasets.
4. The group has even had preliminary discussions about providing reference implementations of features (not an exhaustive list of features, but "representative implementations" from different feature groups that could be shared publicly).

Together, the possibility of a reference set of feature values from a reference dataset using a reference (and well described) reference implementation of quantitative features (e.g. annotated source code) could be a substantial resource for the radiomics community that could facilitate more direct comparisons of methods and results.

REFERENCES FOR THIS REPORT

[1] P. Lo, S. Young, H. J. Kim, M. S. Brown, and M. F. McNitt-Gray, "Variability in CT lung-nodule quantification: Effects of dose reduction and reconstruction methods on density and texture based features," *Med. Phys.*, Aug. 2016.

[2] S. Young, H. J. G. Kim, M. M. Ko, W. W. Ko, C. Flores, and M. F. McNitt-Gray, "Variability in CT lung-nodule volumetry: Effects of dose reduction and reconstruction methods," *Med. Phys.*, vol. 42, no. 5, 2015.

[3] J. Hoffman, S. Young, F. Noo, and M. McNitt-Gray, "Technical Note: FreeCT_wFBP: A robust, efficient, open-source implementation of weighted filtered backprojection for helical, fan-beam CT," *Med. Phys.*, vol. 43, no. 3, pp. 1411–1420, Mar. 2016.

[4] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," in *Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India*, 1998.

[5] D. . Zhao, T.; Hoffman, J.M.; McNitt-Gray, M.F.; Ruan, "Low-Dose CT Image Denoising Using An Optimized Wiener Filter in the BM3D Algorithm.," in *AAPM Annual Meeting (Oral presentation)*, 2017.

[6] N Emaminejad, M Wahi-Anwar, J Hoffman, A Sultan, K Ruchalski, G Kim, J Goldin, M Brown, McNitt-Gray M. "Evaluation of CAD Nodule Detection Performance in Low Dose CT Lung Cancer Screening Across a Range of Dose Levels, Slice Thicknesses and Reconstruction Kernels," in *AAPM Annual Meeting (Oral presentation)*, 2017.

[7] N Emaminejad, M Wahi-Anwar, J Hoffman, K. Ruchalski, J. Goldin, A Sultan, HG Kim, M Brown, McNitt-Gray M. "Assessing the Performance of CAD in Lung Nodule Detection from Low-Dose Lung Cancer Screening CT Exams Under Different Combinations of Radiation Dose Level, Slice Thickness, and Reconstruction Kernel," in *RSNA 2017*, 2017.

[8] J. Kalpathy-cramer *et al.*, "Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features," *Tomography*, vol. 2, no. 4, pp. 430–437, 2016.

[9] Balagurunathan Y, Beers A, Kalpathy-Cramer J, McNitt-Gray M, Hadjiiski L, Zhao B, Zhu J, Yang H, Yip SSF, Aerts HJWL, Napel S, Cherezov D, Cha K, Chan HP, Flores C, Garcia A, Gillies R, Goldgof D. "Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST data," *Med. Phys.*, 2018.

LIST OF PUBLICATIONS AND PRESENTATIONS OVER THE PAST YEAR

§ Published in Peer-Reviewed Journals (* Denotes joint QIN effort)

1. Young S, Lo P, Kim GHJ, Brown M, Hoffman JM, Hsu W, Wahi-Anwar M, Flores C, Lee G, Noo F, Goldin JG, McNitt-Gray MF. The Effect of Radiation Dose Reduction on Computer-Aided Detection (CAD) Performance in a Low-Dose Lung Cancer Screening Population. Med Phys. 2017 Apr;44(4):1337-1346. doi: 10.1002/mp.12128.
2. *Balagurunathan Y, Beers A, Kalpathy-Cramer J, McNitt-Gray M, Hadjiiski L, Zhao B, Zhu J, Yang H, Yip SSF, Aerts HJWL, Napel S, Cherezov D, Cha K, Chan HP, Flores C, Garcia A, Gillies R, Goldgof D. Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST data. Med Phys. 2018 Jan 24. doi: 10.1002/mp.12766.
3. Martin T, Hoffman J, Alger JR, McNitt-Gray M, Wang DJ. Low-dose CT perfusion with projection view sharing. Med Phys. 2018 Jan;45(1):101-113. doi: 10.1002/mp.12640.
4. *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. doi: 10.18383/j.tom.2016.00235.

§ Submitted, Under Review

1. Hoffman J, Noo F, Young S, Hsieh S, McNitt-Gray MF. Technical Note: FreeCT_ICD: An Open Source Implementation of a Model-Based Iterative Reconstruction Method using Coordinate Descent Optimization for CT Imaging Investigations. Submitted to Medical Physics January 2018. Under review.

§ In preparation (submission 1st quarter 2018)

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 2018.
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 2018.
3. Hoffman J, Emaminejad N, Wahi-Anwar M, Kim HG, Brown MS, Young S, McNitt-Gray MF. Technical Note: Design and Implementation of a High Throughput Pipeline for Reconstruction and Quantitative Analysis of CT Image Data. To be submitted to Medical Physics, 1st quarter 2018.

§ Conference Abstracts, Presentations and Posters

1. Emaminejad, N, Lo, P, Ghahremani, S, Kim, GH, 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. Proc. SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis, 101340B (March 27, 2017); doi: 10.1117/12.2255000

2. J Hoffman, H Kim , J Goldin , M Brown , M McNitt-Gray. A Pilot Study Evaluating the Robustness of Density Mask Scoring (RA-950), a Quantitative Measure of Chronic Obstructive Pulmonary Disease, to CT Parameter Selection Using a High-Throughput, Automated, Computational Research Pipeline. Snap Oral presentation at AAPM annual meeting 2017.

3. http://www.aapm.org/meetings/2017AM/PRAbs.asp?mid=127&aid=38039 . accessed July 15, 2017.

4. J Hoffman, M Wahi-Anwar , N Emaminejad , H Kim , M Brown , M McNitt-Gray. A Fully-Automated, High-Throughput, Reconstruction and Analysis Pipeline for Quantitative Imaging in CT. ePoster presentation at AAPM annual meeting 2017. http://www.aapm.org/meetings/2017AM/PRAbs.asp?mid=127&aid=38146 . accessed July 15, 2017.

5. T Zhao, J Hoffman , M McNitt-Gray , D Ruan, Low-Dose CT Image Denoising Using An Optimized Wiener Filter in the BM3D Algorithm. Oral presentation at AAPM annual meeting 2017. http://www.aapm.org/meetings/2017AM/PRAbs.asp?mid=127&aid=36953 . accessed July 15, 2017.

6. N Emaminejad, M Wahi-Anwar, J Hoffman, A Sultan, K Ruchalski, G Kim, J Goldin, M Brown, M McNitt-Gray. Evaluation of CAD Nodule Detection Performance in Low Dose CT Lung Cancer Screening Across a Range of Dose Levels, Slice Thicknesses and Reconstruction Kernels. Oral presentation at AAPM annual meeting 2017. http://www.aapm.org/meetings/2017AM/PRAbs.asp?mid=127&aid=36990 accessed July 15, 2017.

7. J Hoffman, F Noo , M McNitt-Gray. Influence of Tube Current Modulation On Noise Statistics of Reconstructed Images in Low-Dose Lung Cancer CT Screening . Oral presentation at AAPM annual meeting 2017.

8. http://www.aapm.org/meetings/2017AM/PRAbs.asp?mid=127&aid=37956 .accessed July 15, 2017.

9. N Emaminejad, M Wahi-Anwar, J Hoffman, K. Ruchalski, J. Goldin, A Sultan, HG Kim, M Brown, M McNitt-Gray Assessing the Performance of CAD in Lung Nodule Detection from Low-Dose Lung Cancer Screening CT Exams Under Different Combinations of Radiation Dose Level, Slice Thickness, and Reconstruction Kernel. Oral presentation at RSNA 2017.

10. JM Hoffman, HJ Kim, JG Goldin, MS Brown, MF McNitt-Gray, Robustness Evaluation of RA-950 Scoring in a Cohort of CT Lung Screening Patients Across a Large Range of CT Acquisition and Reconstruction Conditions. Oral presentation at RSNA 2017.

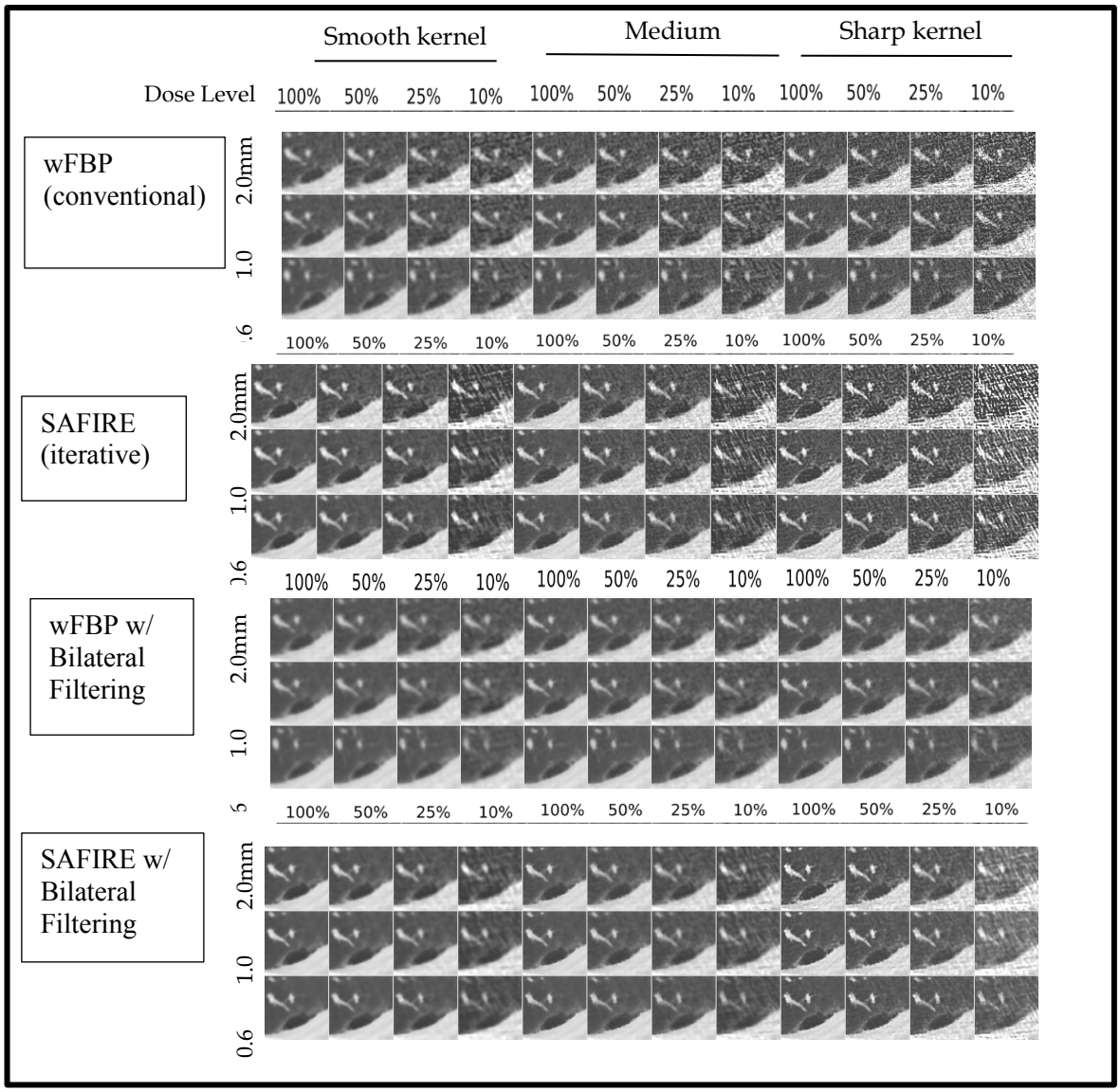
11. *J Kalpathy-Cramer, B Zhao, D Goldgof, S Napel, DL Rubin, MF McNitt-Gray, J Sieren, I Yeung, LM Hadjiiski, Y Balagurunathan, A Beers. Standardizing Radiomic Feature Descriptions for Quantitative Imaging: A Preliminary Report of the Cooperative Efforts of the NCI's QIN PET-CT Subgroup. Quantitative Reading Room poster presentation at RSNA 2017.

APPENDIX

§ Representative Images from Different Acquisition/Reconstruction Conditions.

The below images all originated from the same raw projection dataset. They represent a small section (with some emphysematous destruction in the lung) of a lung cancer screening exam. Raw projection data was used to first simulate 3 reduced dose level scans (100%, 50%, 25% and 10% of original dose) which were then reconstructed with: (a) different slice thicknesses (0.6, 1.0 and 2.0 mm), (b) different kernels for wFPB and SAFIRE (Siemens' iterative reconstruction); and then (c) applied bilateral filtering to each dataset. The reference condition is 100% dose, 1.0mm thickness, wFPB using the medium kernel (outlined in red).

These images demonstrate the substantial variation in appearance due to acquisition (dose) and reconstruction conditions. These image datasets will be used to assess the effects of these acquisition/ reconstruction and mitigation measures (denoising) on quantitative imaging features.


Smooth kernel
Medium
Sharp kernel
Dose Level
100% 50% 25% 10% 100% 50% 25% 10% 100% 50% 25% 10%
wFBP (conventional)
SAFIRE (iterative)
wFBP w/ Bilateral Filtering
SAFIRE w/ Bilateral Filtering
2.0mm
1.0
0.6

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 for 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. Also, 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, we are working with industry partners to develop an integrated image analysis platform for use in large-scale multi-center clinical trials and daily clinical care.

PROGRESS OVER THE PAST YEAR

§ Specific Aim1

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

We spearheaded a QIN collaborative project to assess the consistency of DSC-MRI analysis methods across sites and platforms.

The initial results from this study were presented at the 2017 International Society of Magnetic Resonance in Medicine Meeting [1], and recently submitted as a manuscript to the American Journal of Neuroradiology. The results showed excellent cross-site agreement for rCBV. All metrics were capable of distinguishing low-grade from high-grade tumor. Optimum thresholds were determined using pooled data of both normalized rCBV and CBF (Figure 1). This study demonstrates that both normalized relative cerebral blood volume (nRCBV) and cerebral blood flow (nCBF) can be used to distinguish low-grade from high-grade brain tumor, in a consistent fashion and using a single consensus nRCBV or nCBF threshold.

Significance: This result should lend confidence and consistency for the use of nRCBV on a routine basis, potentially motivating its incorporation into an updated RANO criteria for the assessment of brain tumor response.

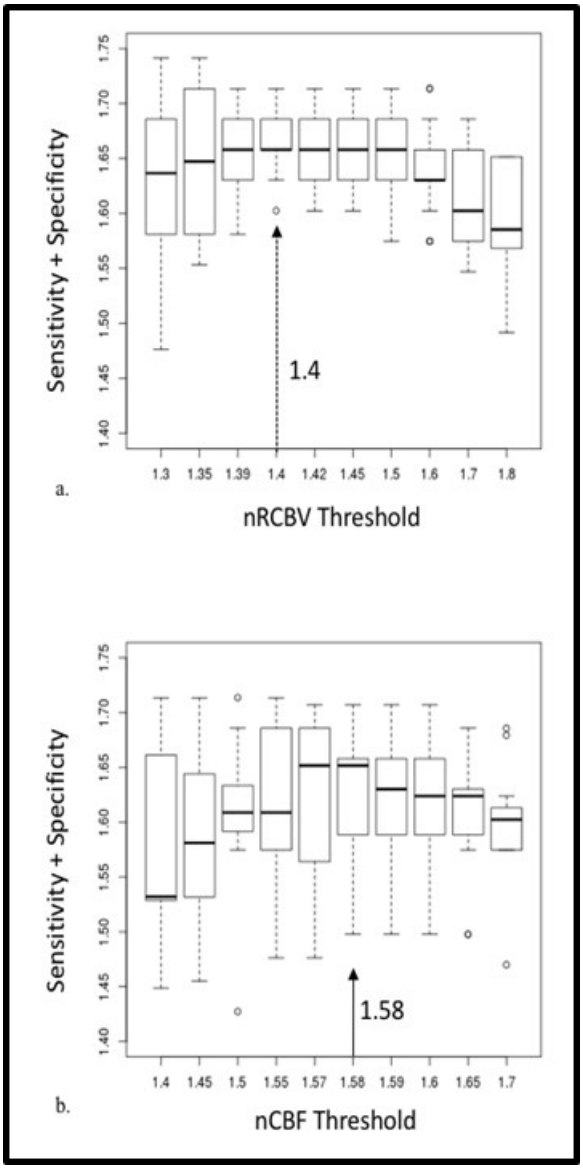


Figure 1: Box plots showing sum of sensitivity and specificity values for nRCBV (a) and nCBF (b) for all sites and metrics. From these data, values that gave a maximum sum and minimal variance were chosen as the consensus threshold values.

Optimization of DSC MRI Echo Times for CBV Measurements.

Using a combination of simulations and DEGES data acquired for our U01 study an error analysis was performed to determine the TE that would minimize the variance in CBV measurements made with DSC MRI. The results of this study, recently published [2], demonstrate that the optimal TE for a typical single-echo DSC MRI acquisition is a weighted average of T2* values that occur before and after the passage of contrast agent.

Significance: The results confirm that the TE = 30ms, that we have been using in our studies is ideal for the analysis of brain tumor. However, it is less ideal for other tissues, such as white matter, or when an arterial input function (AIF) is to be determined, an important factor to keep in mind for ongoing analyses.

§ Specific Aim 2

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

Histologic validation of perfusion and diffusion metrics to distinguish tumor from treatment effect.

The aim of this study was to determine the diagnostic accuracy of several MRI-derived diffusion and perfusion parameters to distinguish pure treatment effect (TE) from pure glioblastoma (GBM) using spatially-correlated biopsy samples. Histopathologic diagnosis of pure TE (n=10) or pure GBM (n=34) was confirmed in tissue samples from 15 consecutive subjects with analyzable data. Perfusion thresholds distinguished TE from GBM ($P<0.05$), whereas ADC, PSR, and PH could not ($P>0.05$). A manuscript describing this study was recently published [3].

Significance: These validated thresholds can be used to determine the fractional tumor burden (FTB) within enhancing lesions. This should prove useful for surgical guidance and increased confidence in the assessment of treatment response.

MRI-derived fractional tumor burden (FTB) is predictive of overall and progression free survival in newly diagnosed GBM following concomitant chemoradiation therapy and recurrent GBM following bevacizumab therapy.

These two studies were selected as oral presentations at the 25th Annual International Society of Magnetic Resonance in Medicine meeting [4, 5]. Using the deltaT1 method and the validated perfusion thresholds described above, FTB maps were created and demonstrated the ability to predict response to therapy using two different therapeutic strategies. Example FTB maps are shown in Figure 2. Additional patient data is being analyzed with the plans to submit two papers on this topic within the next three months.

Significance: This new imaging biomarker, fractional tumor burden (FTB), may provide the answer to the longstanding unmet need of being able to distinguish tumor from treatment effect.

§ Specific Aim 3

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

IB Rad Tech development

The DWI outputs added to the core processing library used by IB Rad Tech were revised in response to the feedback received from the diffusion MRI challenge led by Dr David Newitt [6, 7]. A base PWI/DWI comparison workflow was designed and added to IB Rad Tech, and workflow features continued to be added and enhanced. In addition, after extensive testing by QIN members and others, versions 2.0 of IB Rad Tech (workflow engine), IB Neuro (DSC-PWI processor), IB Diffusion (DWI processor), and IB DCE (DCE-PWI processor) were all released as FDA-cleared and CE-Marked products in November 2017.

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.

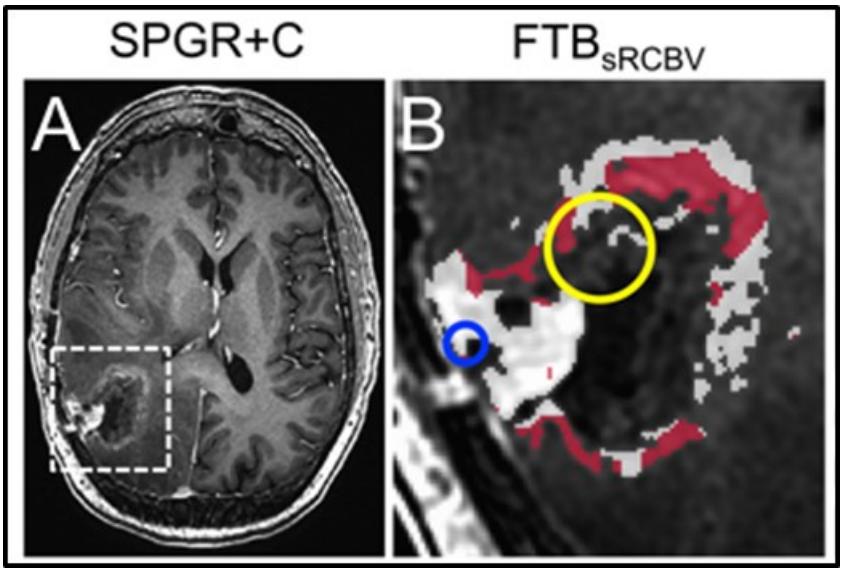


Figure 2: Example fractional tumor burden (FTB) maps in a 58-year-old female. The tissue histopathology of the region encircled in blue is treatment effect (TE) with sparse neoplastic astrocytes, and encircled in yellow contains sections of both TE and GBM. Using the respective thresholds, GBM (red) and TE (white) are shown in areas within enhancement on post-contrast SPGR (A) using sRCBV thresholds (B).

COLLABORATIONS WITHIN THE NETWORK

§ We participated in a QIN collaborative project to assess the concordance of DWI metrics across sites.

The initial results from this study were presented at the 2017 International Society of Magnetic Resonance in Medicine [6]. The manuscript describing the study was recently

published [7]. The conclusions of the study were that while agreement between the majority of ADC mapping implementations was good, the biases in in vivo ADC measures both between different offline implementations and between vendor-generated and offline maps are significant. Furthermore, these differences may, in some cases, be large enough to adversely affect the analysis of multi-site diffusion data.

Significance: The results of this study indicate that for any given longitudinal (e.g. treatment response) study all analyses should be performed on a common platform. Therefore, using DWI for our brain tumor response studies should not rely on each vendor's calculated ADC maps. Instead the raw data should be imported into a single platform and analyzed there.

We led a QIN collaborative project to assess the consistency of DSC-MRI analysis methods for the evaluation of brain tumors. This was performed across sites and platforms as described above under Aim 1 results.

We participated in a DICOM (Digital Imaging and Communication in Medicin) Challenge.

The results are described in the recent publication [8] . The goal of the collaborative project was to assess the current capability and provide future guidelines for generating a standard parametric diffusion map Digital Imaging and Communication in Medicine (DICOM) in clinical trials that utilizes quantitative diffusion-weighted imaging (DWI). Participating sites used a multivendor DWI DICOM dataset of a single phantom to generate parametric maps (PMs) of the apparent diffusion coefficient (ADC) based on two models. The results were evaluated for numerical consistency among models and true phantom ADC values, as well as for consistency of metadata with attributes required by the DICOM standards. This analysis identified missing metadata descriptive of the sources for detected numerical discrepancies among ADC models. Instead of the DICOM PM object, all sites stored ADC maps as DICOM MR objects, generally lacking designated attributes and coded terms for quantitative DWI modeling. Source-image reference, model parameters, ADC units and scale, deemed important for numerical consistency, were either missing or stored using nonstandard conventions.

Significance: Guided by the identified limitations, the DICOM PM standard has been amended to include coded terms for the relevant diffusion models. Open-source software has been developed to support conversion of site-specific formats into the standard representation.

PLANS FOR NEXT YEAR

§ Specific Aim1: Characterize the repeatability of DSC and DEGES PWI and DWI (fDM) parameters in primary brain tumors.

Additional SPICE data has been collected and undergoing analysis.

This data will be combined with data previously collected (for this U01 study) at Barrow Neurological Institute. Comparison of single-echo DSC-MRI data to dual-echo DSC MRI data, and the resulting rCBV maps is underway, with journal submission expected within the next 6 months.

Implementation of an EPI-based multiple echo sequence.

While we have made great progress with our dual-echo spiral-based sequence for perfusion imaging (ie SPICE), spiral-based methods are not currently well-supported. Therefore, we continue to search for EPI-based dual-echo options. In this context, we have initiated a relationship with developers at GE Healthcare who have a prototype multi-echo EPI (MEPI), which they will make available within the next few months. We plan to install this and test it with data collection in 2018.

§ Specific Aim 2: To prospectively determine the ability of pMRI and DWI to predict treatment response in glioblastoma patients.

The creation and validation of dT1 maps

This is being written with plans for journal submission within the next six months.

Access has been granted to pre/post-contrast T1-weighted MR images from RTOG 0625.

We computed dT1 maps for this data, which will be compared to the central reader analysis previously published. The goal is to determine if automated delineation of enhancing lesion, obtained with dT1 maps, is as good as central reader analysis. If proven this would be a paradigm shift in how brain tumor clinical studies are performed. Data analysis is underway.

MRI-derived fractional tumor burden (FTB) is predictive of overall and progression free survival in newly diagnosed GBM following concomitant chemoradiation therapy and recurrent GBM following bevacizumab therapy.

Using the deltaT1 method and the validated perfusion thresholds, FTB maps were created and demonstrated the ability to predict response to therapy using two different therapeutic strategies. Patient data is being analyzed with the plans to submit two papers on this topic within the next three months.

Implementation of RSI (restriction spectrum imaging) data collection and processing.

RSI is a promising new way to collect and analyze diffusion MRI data. Initial results demonstrate its ability to more directly detect tumor cellularity without being confounded by other treatment related changes such as edema. We have recently implemented this sequence and have begun collecting data. This advancement directly addresses the original goal of this aim to assess the value of diffusion in treatment monitoring of patients with brain tumors.

§ Specific Aim 3: To develop a commercial integrated PWI/DWI image analysis platform for use in large-scale multi-center multi-platform clinical trials.

Additional enhancements are planned for IB Rad Tech for year 5.

These include direct calculation of DWI parameters; additional longitudinal statistical calculations, retention and reporting features; and automation of some currently manual processing activities.

The IB Rad Tech 2.0 protocol will continue to be enhanced

This will occur during year five to accommodate IB Rad Tech enhancements. Work has begun on the workflow designer/editor, and efforts on this will continue throughout year five.

For data sharing Imaging Biometrics putting resources into development of features

This will address data access and processing needs as expressed in QIN discussions, teleconferences and collaborative projects.

LIST OF REFERENCES

1. Schmainda, K.M., et al. Multi-site Concordance of DSC-MRI Analysis for Brain Tumors: Results of a NCI Quantitative Imaging Network DSC-MRI Collaborative Project. in Proc International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii.
2. Bell, L.C., et al., Optimization of DSC MRI Echo Times for CBV Measurements Using Error Analysis in a Pilot Study of High-Grade Gliomas. AJNR Am J Neuroradiol, 2017. **38**(9): p. 1710-1715.
3. Prah, M.A., et al., Spatial discrimination of glioblastoma and treatment effect with histologically-validated perfusion and diffusion magnetic resonance imaging metrics. J Neurooncol, 2017.
4. Prah, M.A., J.M. Connelly, and K.M. Schmainda. MRI-perfusion derived fractional tumor burden (FTB) is predictive of overall and progression free survival in newly diagnosed glioblastoma following concomitant chemoradiotherapy. in Proceedings International Society of Magnetic Resonance in Medicine. 2017. Honolulu, Hawaii.

5. Prah, M.A., et al. MRI-perfusion derived Fractional Tumor Burden (FTB) stratifies survival in recurrent glioblastoma following treatment with bevacizumab. in Proceedings of the International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii.
6. Newitt, D.C., et al. Multi-Site Concordance of DWI Metrics: Results of teh NCI Quantitative Imaging Network ADC Mapping Collaborative Project. in Proceedings of the International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii.
7. Newitt, D.C., et al., Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. J Med Imaging (Bellingham), 2018. **5**(1): p. 011003.
8. Malyarenko, D., et al., Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. J Med Imaging (Bellingham), 2018. **5**(1): p. 011006.

PUBLICATIONS FROM QIN EFFORTS

§ Abstracts

1. Hurrell SL, Cochran E, McGarry Sean D, Kaczmarowski AL, Connelly J, Mueller W, Rand SD, **Schmainda KM**, LaViolette PS. Predictive cytological topography highlights regions of pathologically confirmed non-enhancing hypercellular tumor in glioblastoma patients. in Proceedings of the International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii. P 256.

2. **Schmainda, KM**, Prah MA, Rand SD, Muzi M, Rane SD, Da X, Yen YF, Kalpathky-Cramer J, Chenevert TL, Malyarenko D, Hoff B, Ross B, Cao Y, Aryal MP, Erickson B, Korfiatis P, Bell L, Hu L, Quarles CC. Multi-site Concordance of DSC-MRI Analysis for Brain Tumors: Results of a NCI Quantitative Imaging Network DSC-MRI Collaborative Project. in Proc International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii. P 261.

3. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, Bell L, Fedorov A, Fennessy F, Jacobs MA, Solaiyappan M, Hectors S, Taouli B, **Schmainda KM**, Prah MA, Yen YF, Kalpathy-Cramer J, Taber Erin, Kroenke C, Cao Y, Madhava A, Muzi M, Kinahan P, Yankeelov TE, Arlinghaus LR, Boss MA, Suukla-Dave A, Hylton N. Multi-Site Concordance of DWI Metrics: Results of the NCI Quantitative Imaging Network ADC Mapping Collaborative Project. in Proceedings of the International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii. P 600.

4. Huang W, Chen Y, Fedorov A, Li X, Jajamovich G, Malyarenko D, Aryal M, LaViolette P, Oborski M, O'Sullivan F, Abramson R, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta S, Besa C, Kalpathy-Cramer J, Mountz J, Laymon C, Muzi M, Kinahan P, **Schmainda K**, Cao Y, Chenevert T, Taouli B, Yankeelov T, Fennessy F, Li X. Effects of AIF quantification variations on shutter-speed pharmacokinetic modeling of prostate DCE-MRI data: a multicenter data analysis challenge. in Proceedings of the International

Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii. P 625.

5. Prah MA, Connelly JM, and **Schmainda KM**. MRI-perfusion derived fractional tumor burden (FTB) is predictive of overall and progression free survival in newly diagnosed glioblastoma following concomitant chemoradiotherapy. in Proceedings International Society of Magnetic Resonance in Medicine. 2017. Honolulu, Hawaii. P 707.

6. Prah MA, Connelly JM, and **Schmainda KM**. MRI-perfusion derived Fractional Tumor Burden (FTB) stratifies survival in recurrent glioblastoma following treatment with bevacizumab. in Proceedings of the International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii. P 708

7. Li K, Chen Y, Yu Y, Li X, fedorov A, Jajamovich G, Malyarenko D, Aryal M, LaViolette P, Oborski M, O'Sullivan F, Abramson R, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta S, Besa C, Kalpathy-Cramer J, Mountz J, Laymon C, Muzi M, Kinahan P, **Schmainda K**, Cao Y, Chenevert T, Taouli B, Fennessy F, Yankeelov T, Li X, Huang W. The effects of AIF quantification variations on DCE-MRI prediction of soft tissue sarcoma response to preoperative therapy: a preliminary multicenter study. in Proceedings of the International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii. P 4375.

8. **Schmainda KM**, Prah MA, Rand SD, Muzi M, Rane SD, Da X, Yen YF, Kalpathy-Cramer J, Chenevert TL, Malyarenko D, Hoff B, Ross B, Cao Y, Aryal MP, Erickson B, Korfiatis P, Bell L, Hu L, Quarles CC. "Multi-site concordance of DSC-MRI analysis for brain tumors: Results of a NCI Quantitative Imaging Network DSC-MRI Collaborative Project". in Proceedings of the International Society of Magnetic Resonance in Medicine, 25th Annual Meeting. 2017. Honolulu, Hawaii. P 261.

§ Papers

1. Bell LC, Does MD, Stokes AM, Baxter LC, Schmainda KM, Dueck AC, Quarles CC. "Optimization of dSC MRI Echo Times for CBV measurements using error analysis in a pilot study of high-grade gliomas" Am J NeuroRadiology 38:1710-15 (2017).

2. Malyarenko D, Fedorov A, Bell L, Prah M, Hectors S, Arlinghaus L, Muzi M, Solaiyappan M, Jacobs M, Fung M, Shukla-Dave A, McManus K, Boss M, Taouli B, Yankeelov TE, Quarles CC, **Schmainda K**, Chenevert TL, Newitt DC. Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. J Med Imaging (Bellingham), 2018 5(1): Epub 2017 Oct 30.

3. Newitt DC, Malyarenko D, chenevert TL, Quarles CC, Bell L, Fedorov A, Fennessy F, Jacobs MA, Solaiyappan M, Hectors S, Taouli B, Muzi M, Kinahan PE, **Schmainda KM**, Prah MA, Taber EN, Kroenke C, Huang W, Arlinghaus LR, Yankeelov TE, Cao Y, Aryal M, Yen YF, Kalpathy-Cramer J, Shukla-Dave A, Fung M, Liang J, Boss M, Hylton N.

Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. J Med Imaging (Bellingham), 2018. 5(1): p. 011003.

4. Prah MA, Al-Gizawiy MM, Mueller WM, Cochran EJ, Hoffmann RG, Connelly JM, **Schmainda KM**. Spatial discrimination of glioblastoma and treatment effect with histologically-validated perfusion and diffusion magnetic resonance imaging metrics. J Neurooncol, 136:13-21 (2018).

5. **Schmainda KM**, Prah MA, Rand SD, Liu Y, Logan B, Muzi M, Rane SD, Da X, Yen YF, Kalpathy-Cramer J, Chenevert TL, Hoff B, Ross B, Cao Y, Aryal MP, Erickson B, Korfiatis, Dondlinger T, P, Bell L, Hu L, Kinahan P, Quarles CC. "Multi-site concordance of DSC-MRI analysis for brain tumors: Results of a NCI Quantitative Imaging Network DSC-MRI Collaborative Project". Am J Neurorad (In Press).

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 lower grade gliomas, they do not show the entirety of 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 metabolite 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

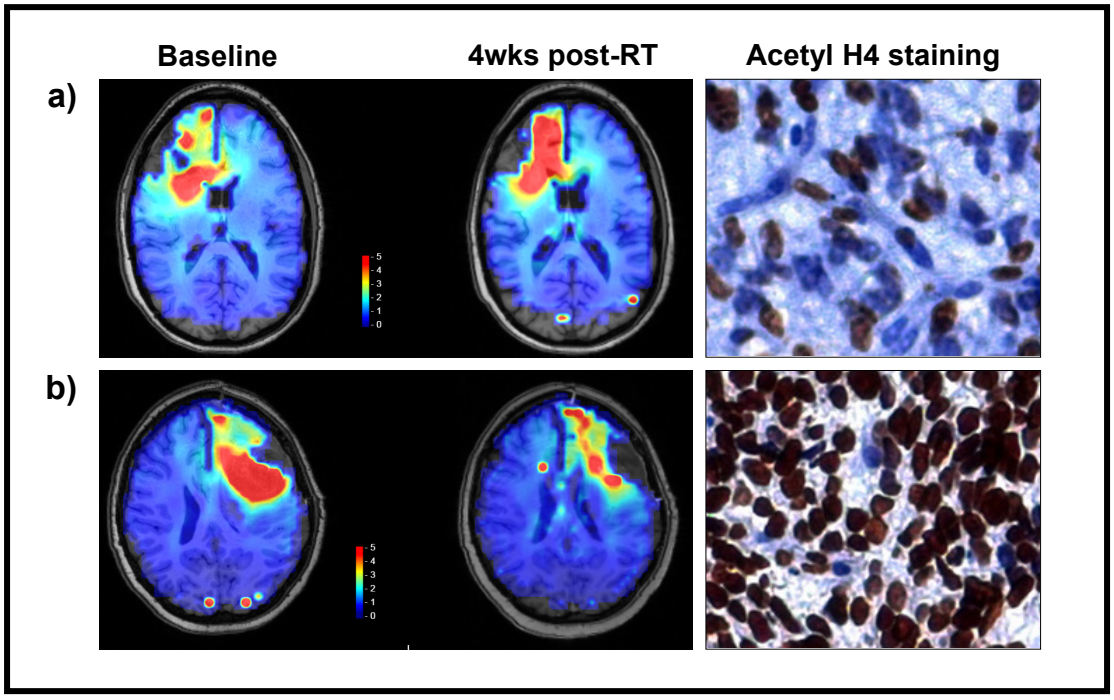
§ Development of Web-based spectroscopic MRI-dedicated Clinical Interface for Brain Tumor Imaging (BrICS: Brain Imaging Collaboration Suite)

There is widespread agreement that MR spectroscopy can provide valuable information without the need for exogenous contrast agents, however the infrastructure needed to incorporate MRSI into the clinical workflow is lacking. We have been developing a web-based application, the Brain Imaging Collaboration Suite to facilitate use of MRSI in the clinical workflow for radiation therapy planning. This "scanner-to-clinician" platform is designed to provide quantitative, expedient, and objective analysis to integrate spectroscopic MRI (sMRI) into routine clinical usage, including diagnosis and therapy planning (radiation or surgery). It also includes editing tools to modify the treatment volume and automatic segmentation tool for defining residual contrast enhanced tumor volume. 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. A video clip briefly introducing its capabilities is enclosed.

§ Tumor response to HDAC inhibitor

Patients with GBM enrolled in NCT02137759 were given intravenous belinostat (Spectrum Pharmaceuticals) at a dose of either 750mg/kg/m^2 or 500mg/kg/m^2 (3 cycles, 5 days, every third week). Patients underwent maximal safe tumor resection followed by daily 75mg/m^2 temozolomide plus 60 Gy radiation to residual contrast-enhancing tissue. Tissue samples collected during resection were stained with anti-acetylated histone H4 antibody (Acetyl H4, #Ab15823, Abcam). sMRI scans were performed at baseline and one-month post-radiation, specifically looking at the choline-to-NAA ratio (CHO/NAA). The volume of metabolically active tissue (2x abnormal in Cho/NAA compared to contralateral metabolism) was calculated for quantitative assessment of tumor response. The Cho/NAA ratio is shown for two patients assessed at baseline and 4 weeks' post-chemoradiation. Patient #1 is a 51-year-old female with poor therapeutic response (Figure a). Her tissue stained weakly for acetyl H4; her metabolically-active tumor volume increased 64.8cm^3 to 80.5cm^3. Patient #2 is a 28-year-old female (Figure b) whose tissue stained strongly for acetyl H4, and whose metabolically-active tumor decreased 81.6cm^3 to 50.1cm^3. These responses are consistent with progression-free survival, as patient #1 progressed while patient #2 has not (9 months post-chemoradiation). These cases suggest that acetyl H4 may be a good biomarker for predicting HDACi treatment efficacy, and that metabolic response can be monitored non-invasively using sMRI. So far, we have enrolled 13 patients in HDAC inhibitor treatment cohort while 15 matching control (standard care) subjects. We anticipate to enroll 20 more patients in HDAC inhibitor treatment cohort in next 6 months at two sites.



COLLABORATIONS WITHIN THE NETWORK

Our current project is a two-site clinical study, with the Emory team collaborating with Johns Hopkins. BrICS has been shared between these sites, and is expanding to the University of Miami. We have a plan to expand to include Cedars Sinai, Mount Sinai, Memorial Sloan Kettering, and New York University late this year. We hope to deploy this technology with several QIN sites later in 2018/2019. Hui-Kuo Shu, MD, PhD has been serving as the chair of CTDD Working Group and involved in many activities to promote QIN to various national clinical trial consortiums. He also serves as a co-chair of NCI Quantitative Imaging for Radiation Therapy (QIRT) Working Group.

PLANS FOR NEXT YEAR

We plan to continue with patient enrollment for our clinical study at two sites for 6 more months and collect follow-up data. We will compile the data and present the results at the national meetings. We are preparing a manuscript to be submitted to Cancer Research.

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 two new Siemens Prisma 3T scanners with 32 channel head coil array that are well-calibrated. We installed the same advanced sequence on two Prisma scanners with help from Dr. Maudsley at University of Miami (consultant) and Siemens.

Hopefully, we can get funding (BRP U01 CA225462) for further developing our sequence and analysis program plus hardware to transform sMRI a standard imaging.

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

§ Manuscripts published directly as a result of this grant:

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. The impact of integrating volumetric whole-brain spectroscopic MRI into radiation treatment planning for glioblastoma. Tomography, 2(4): 366-373. PMC5241103

Gurbani, S., Schreibmann, E., Maudsley, A., Cordova, J.S., Soher, B.J., Poptani, H., Verma, G., Barker, P.B., Shim, H, Cooper LAD A convolutional neuronal network to filter artifacts in spectroscopic MRI. Final revision, Magnetic Resonance in Medicine.

Gurbani, S., Olson, J., Shu, H., Shim, H. Assessing treatment response of glioblastoma to an HDAC inhibitor belinostat. Will be submitted to Cancer Research

NATIONAL PRESENTATIONS DIRECTLY AS A RESULT OF THIS GRANT

§ **Invited Lectures at the National Meetings:**

Shim et al. "spectroscopic MRI for the management of brain tumor patient" – Siemens Healthcare Webinar, Feb 2017.

Shim et al. "spectroscopic MRI to guide radiation therapy dose escalation in GBM patients" – Eastern Cooperative Oncology Group – American College of Radiology Imaging network (ECOG-ACRIN) meeting, Invited lecture "future technology in pipeline" Washington DC, May 2017

Shim et al. "Lead compound discovery and optimization, on the example of CXCR4" – International Society of Radiopharmaceutical Sciences, Plenary Lecture, Dresden, Germany, May 2017.

Shim et al. "Spectroscopic MRI guided dose escalation for GBM" in CE session, "MRI-Guided Adaptation: From Anatomy To Biology", ASTRO, San Diego, Sept 2017

Weinberg et al. "spectroscopic MRI to guide radiation therapy dose escalation in GBM patients" – Eastern Cooperative Oncology Group – American College of Radiology Imaging network (ECOG-ACRIN) meeting, Invited lecture "future technology in pipeline" Orlando, FL, October 2017

Shim et al. "Molecular Imaging mini-course: Clinical application of molecular imaging – Neuro MRS & PET" RSNA Refresher Course, Chicago, December 2017

Shu et al. "Implementing Quantitative Imaging Network (QIN) Tools in NCTN trials: QIN Clinical Trials Design & Development working group," presentation at the Imaging Working Group, NRG Oncology Semi-Annual Meeting, Philadelphia, PA, July 2017

Shu et al. Discussant for Oral Scientific Session SS 5 Biology 1 - Innovative Biologic Approaches to Improve Risk Stratification and Treatment Outcomes, 59th Annual Meeting of the American Society for Therapeutic Radiology and Oncology (ASTRO), San Diego, CA, September 2017

Shu et al. Keynote speaker for the CNS—Science Session with Keynote, 103rd Annual Meeting of the Radiological Society of North America, Chicago, IL, November 2017

Shu et al. "Velocity for Radiosurgery and Radiotherapy Multimodality Tracking and Response Assessment," Invited presentation at Varian Australasian Oncology Summit, Grand Hyatt Hotel, Melbourne, Victoria, Australia, February 2017

INVITED LECTURES AT THE ACADEMIC CENTERS:

Shim et al. "The use of high resolution 3D whole brain MR spectroscopic imaging in the management of brain tumor patients", BIRI Seminar, Cedars Sinai Hospital, September 2017

Shim et al. "Use of whole brain spectroscopic MRI for the management of GBM patients", Spectrum Pharmaceuticals, February 2016

Shim et al. "Improved whole brain spectroscopic MRI to guide radiation dose escalation for glioblastomas", Memorial Sloan Kettering Cancer Center, October, 2017

Shim et al. "The significant potential of CXCR4 as an imaging/therapeutic target" Seoul National University Hospital, Seoul, Korea, December 2017

Shu et al. "Brain tumor patients: Critical unmet needs for treatment planning imaging and spectroscopic MRI," Invited seminar, Radiation Therapy Section, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia, February 2017

NATIONAL MEETING ABSTRACTS SELECTED FOR PRESENTATIONS

Schreibmann E, Gurbani S., and Shim H. (2017) Integrating 3D whole brain sMRI into GBM treatment planning, Varian Conference, May 2017, Chicago, IL

Gurbani S, Schreibmann S, Sheriff S, Holder, CA, Cooper L, Maudsley A, Shim H. (2017) Rapid internal normalization of spectroscopic MRI maps using a Gaussian mixture model. American Association of Physicists in Medicine (AAPM) Annual Meeting, Denver, CO.

Schreibmann E, Gurbani S., and Shim H. (2017) Building an anatomical filter to facilitate 3D whole brain spectroscopic MRI into GBM treatment planning. American Association of Physicists in Medicine (AAPM) Annual Meeting, Denver, CO.

Gurbani S, Schreibmann S, Sheriff S, Cooper L, Maudsley A, Shim H. (2017) A software platform for collaborative radiation therapy planning using spectroscopic MRI. American Society for Radiation Oncology Annual Meeting, San Diego, CA.

Gurbani S, Sengupta S, Voloschine A, Liang Z, Yoon Y, Olson J, Shu H, Shim H (2017) Assessing treatment response of glioblastoma to an HDAC inhibitor Belinostat. Society of Neuro-Oncology Annual Meeting, San Francisco, CA.

Gurbani S, Kleinberg L, Zhong J, Olson J, Mellon E., Maudsley A, Shu H, Shim H (2017) Spectroscopic MRI predicts recurrence patterns in glioblastoma. Society of Neuro-Oncology Annual Meeting, San Francisco, CA.

Special Note:

Hyunsuk Shim, Ph.D. has taken a role as the senior/deputy editor of Cancer Research (the major AACR journal) who is in charge of Cancer Imaging.

Hui-Kuo Shu, M.D., Ph.D. was selected as the Top Atlanta Doctor in 2017

UO1CA172320: Evaluation of HCC Response to Systemic Therapy with Quantitative MRI

Icahn School of Medicine at Mount Sinai, New York
Bachir Taouli, MD

INTRODUCTION

In this proposal, we have developed a multiparametric MRI (mpMRI) protocol (**Specific Aim 1)** quantifying different features of hepatocellular carcinoma (HCC) tumor biology and pathology, including tumor cellularity, grade, angiogenesis and degree of hypoxia. The multiparametric MRI (mpMRI) approach includes intravoxel incoherent motion (IVIM) DWI, DCE-MRI, T2* and T1 mapping using oxygen and carbogen challenge.

This protocol has been performed in patients with HCC, including patients undergoing hepatic resection. Routine and advanced histopathologic methods have been performed in a subset of patients (**Specific Aim 2)**.

In the 30 initial patients, our team has showed that the R_1 and R_2* parameters in HCC are reproducible, and that there is a statistically significant decrease in R_2* values in HCC before and after O2 and increase in R_1 after O_2. No significant effect was observed with carbogen.

In addition, our team assessed the correlation between IVIM-DWI and DCE-MRI in HCC tumors and liver parenchyma. DCE-MRI derived arterial fraction and arterial flow were significantly negatively correlated with IVIM-DWI-derived perfusion fraction and pseudodiffusion in the liver, while IVIM-DWI parameters did not correlate with DCE-MRI parameters in HCC. These results indicate that IVIM-DWI and DCE-MRI provide non-redundant information in HCC.

Recently, our group assessed intratumor heterogeneity in HCC using mpMRI and histogram quantification (central tendency parameters mean and median and heterogeneity parameters standard deviation, kurtosis and skewness) in HCC lesions. The imaging findings were correlated with histopathology and gene expression levels. We observed that central tendency and heterogeneity parameters were largely complementary in terms of the assessed correlations. The proposed histogram analysis is therefore promising for noninvasive HCC characterization on the functional, immunohistochemical and genomics level.

Finally, we are currently acquiring data assessing the prediction of HCC response to yttrium 90 radioembolization in unresectable HCC (**Specific Aim 3)**.

§ Discussion and Results of Progress made over the previous year

During this grant period, we have assessed HCC response to Yttrium 90 radioembolization (one of the most efficient locoregional therapy currently available), where patients were assessed pre- and post-treatment to assess changes in MRI parameters and test the predictive value of MRI for HCC response at 6 months. We have preliminary data, and the final data analysis is still in progress. For this aim, we have also added HCC stiffness quantification using MR elastography.

The use of these techniques will eventually result in a better assessment of HCC tumor biology, that could be used to orient treatment and predict response to new therapies, such as immune check-point inhibitors.

SIGNIFICANT RESULTS

- In HCC lesions, for DWI parameters: test-retest repeatability was good to excellent except for D* (pseudodiffusion coefficient). R_2* measurements in HCC and liver parenchyma are more reproducible at 1.5T than at 3.0T, and with oxygen than with carbogen challenge. R_1 measurement was highly repeatable. Repeatability of DCE-MRI parameters was variable ranging from poor to excellent, depending on the parameter studied.
- Our preliminary experience with BOLD and TOLD MRI demonstrates variable response of HCC to O_2 and CB challenges, with a proportion of tumors with O_2 uptake that may possibly be considered hypoxic, depending on their perfusion pattern.
- IVIM-DWI and DCE-MRI provide non-redundant information in HCC, while they correlate in liver parenchyma. These findings may be secondary to the predominant portal inflow in the liver and tortuous vasculature and tissue heterogeneity in tumors.
- HCCs exhibit high tumor heterogeneity as assessed with histogram quantification in comparison with liver parenchyma.
- Central tendency MRI parameters (mean/median) show significant correlations between MRI methods and with histopathology and gene expression; while heterogeneity parameters (histogram derived) exhibit 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. Thus, histogram analysis combining central tendency and heterogeneity mpMRI features is promising for non-invasive HCC characterization on the imaging, histologic and genomics levels.
- Preliminary results suggest that HCC tumor stiffness measured with MR elastography increase early after radioembolization, while there is a significant drop in tumor perfusion (measured with DCE-MRI and IVIM blood-flow related parameter D*) and an increase in tumor ADC.

COLLABORATIONS WITHIN THE NETWORK

We have performed a multicenter QIN study assessing T1 quantification variability across sites using a dedicated phantom. This has been the subject of a recent publication in

MRM (Bane et al, see below). We have also contributed to several MRI projects, which have been now published (see below).

PLANS FOR NEXT YEAR

NA (the grant period has ended on 1/31/18)

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

1. 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. PMID 24772219

2. Malyarenko D, Newitt D, 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. PMID 25940607

3. Malyarenko D, 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(4):396-405. PMID: 28105469

4. Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, Muz 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 Mar;2(1):56-66. PMID: 27200418

5. Bane O, Hectors S, Wagner M, Arlinghaus, Madhava Aryal, Yue Cao, Thomas Chenevert, Fiona Fennessy, Wei Huang, Nola Hylton, Jayashree Kalpathy-Cramer, Kathryn Keenan, Dariya Malyarenko, Robert Mulkern, David Newitt, Stephen Russek, Karl Stupic, Alina Tudorica, Lisa Wilmes, Thomas Yankeelov, Yi-Fei Yen, Michael Boss, Bachir Taouli. Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: Results from a multicenter phantom study. Magnetic Resonance in Medicine 2017. doi: 10.1002/mrm.26903. Pubmed PMID: 28913930

6. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, Bell L, Fedorov A, Fennessy F, Jacobs MA, Solaiyappan M, Hectors S, Taouli B, Muzi M, Kinahan P, Schmainda KM,

Prah MA, Taber EN, Kroenke C, Huang W, Arlinghaus LR, Yankeelov TE, Cao Y, Aryal M, Yen Y, Kalpathy-Cramer J, Shukla-Dave A, Fung M, Liang J, Boss M, Hylton N. Multi-site concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. J Med Imaging. 2018;5(1):011003.

7. Malyarenko D, Fedorov A, Bell L, Prah M, Hectors S, Arlinghaus L, Muzi M, Solaiyappan M, Jacobs M, Fung M, Shukla-Dave A, McManus K, Boss M, Taouli B, Yankeelov T, Quarles CC, Schmainda K, Chenevert T, Newitt D. Toward uniform implementation of parametric map DICOM in multi-site quantitative diffusion imaging studies. Journal of Medical Imaging (Bellingham). 2018 Jan;5(1):011006). PMID: 29021993

LIST OF REFERENCES

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 S, Wagner M, Corcuera-Solano I, Kang M, Boss M, Taouli B. Comparison between three-scan-trace and diagonal body DWI acquisitions: a phantom and volunteer study. Tomography. 2016;2:411-420. PMID: 28480331

5. 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. Scientific Reports. 2017 26;7(1):2452. PMID: 28550313

6. Bane O, Hectors S, Wagner M, Arlinghaus, Madhava Aryal, Yue Cao, Thomas Chenevert, Fiona Fennessy, Wei Huang, Nola Hylton, Jayashree Kalpathy-Cramer, Kathryn Keenan, Dariya Malyarenko, Robert Mulkern, David Newitt, Stephen Russek, Karl Stupic, Alina Tudorica, Lisa Wilmes, Thomas Yankeelov, Yi-Fei Yen, Michael Boss, Bachir Taouli. Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: Results from a multicenter phantom study. Magnetic Resonance in Medicine 2017. doi: 10.1002/mrm.26903. Pubmed PMID: 28913930

7. Hectors S, Wagner M, Besa C, Huang W, Taouli B. Multiparametric FDG-PET/MRI of hepatocellular carcinoma: Initial experience. Submitted to PLOS ONE

U01 CA190254-01: ECOG-ACRIN-Based QIN Resource for Advancing Quantitative Cancer Imaging in Clinical Trials

ECOG-ACRIN

Mitchell D. Schnall, M.D., Ph.D.
David Mankoff, M.D., Ph.D.
Paul E. Kinahan, Ph.D.
Mark Rosen, M.D., Ph.D.

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, outlined in Figure 1, 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.

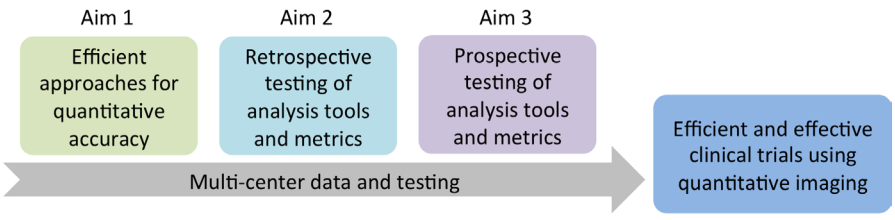


Figure 1: Progression through the specific aims of the project.

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. The relationship between QIN anad the various components of ECOG-ARIN are shown in Figure 2.

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).

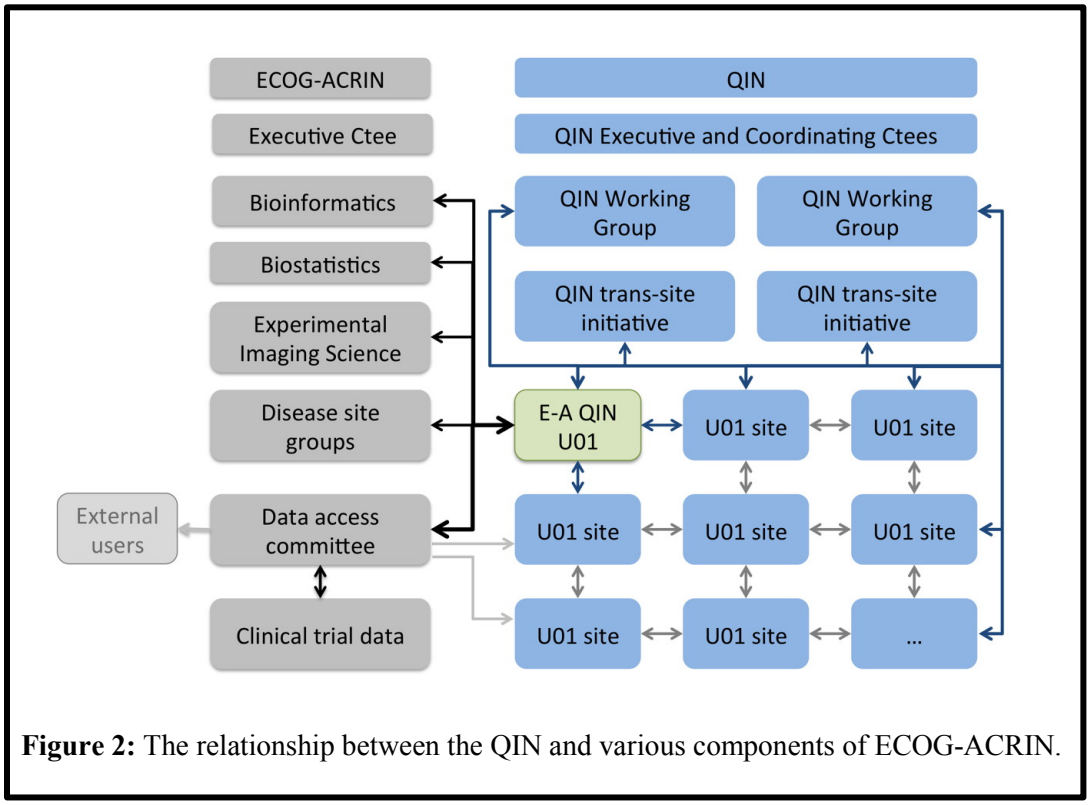


Figure 2: The relationship between the QIN and various components of ECOG-ACRIN.

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:

A modified version of the Centers of Quantitative Imaging Excellence (CQIE) database for QIN sites was used to create a site profile in the Qualification Utility for Imaging Clinical Trials (QUIC) of qualified QIN sites. This allows for dynamic updating as QIN sites change. This is being used to feed automated data quality evaluation methods running Dokerized versions of software tools for phantom analysis that have been developed by several QIN members. These are intended to test respectively CT, MRI-DWI, and PET-SUV metrics of QIN sites. All three sub-aims (CT, MRI-DWI, and PET-SUV) have validated software tools and are being tested with the phantom data from QIN sites that is already contained in QUIC utility.

§ Aim 2:

Continued collaboration with the QIN leadership and the TCIA technical staff to transfer datasets that are intended to provide a resource for method testing and validation from completed ACRIN research for assessment of QIN metrics, validation methods, and challenge projects. The previously established list of prioritized datasets was refined based on the time-lines for availability of each data set, ownership, and discussions with NCI-CIP program staff on appropriateness and broadest applicability. Transfer of the 6688 dataset was completed and 6684 is near completion. The remaining datasets are in queue to be transferred.

In addition, outcomes and progression data are being be made available for correlation with computational findings.

§ Aim 3:

Formalized a process to leverage the clinical trial development structure to enable prospective testing for methods developed by the QIN. Formal discussions between NCTN oncology leaders and QIN members were initiated in the form of a highly successful QIN-NCTN planning meeting in December 2016. This meeting was directed to both the NCTN leaders as well as the QIN investigators. This effort has been expanded, with the help of other QIN investigators who also hold NCTN leadership positions, with the goal to include QIN sessions and disease-site discussions in the NCTN groups on a rotating basis.

At the fall, 2017 Alliance meeting, the EA U01 worked closely with Dr. Larry Schwartz and others in the planning and execution of a QIN plenary session and presentations in key disease site committees that were highly successfully and very well received. Planning

for the summer NRG meeting is in progress. We have advanced concepts for new trials that include exploratory objectives focused on QIN tool testing. Developing concepts include a study (EA1172) of FDG PET/CT in metastatic breast cancer imaging response assessment and developing concept in EA8171 trial "Multiparametric MRI (mpMRI) for Preoperative Staging and Treatment Planning for Newly-diagnosed Prostate Cancer" will have an exploratory objective focused on QIN tools that are designed to lead to prospective multi-center testing of QIN tools for scanner calibration and image analysis for this task.

COLLABORATIONS WITHIN THE NETWORK

We successfully navigated numerous logistical issues related to data transfer and prepared several datasets to be transferred to the TCIA. We continued efforts focused on building the infrastructure for sharing imaging datasets and refined the previously developed prioritized dataset transfer list based on the timelines for availability of each dataset, ownership, and discussions with NCI-CIP program staff on appropriateness and broadest applicability.

With the help of other QIN investigators who also hold NCTN leadership positions within the NCTN and NCI leadership, the EA U01 played a key role in the planning and executing QIN presentations and discussions at the fall Alliance group meeting. Planning for the summer NRG meeting is already well underway.

PLANS FOR NEXT YEAR

Our plans for the coming year are as follows:

Aim 1: We will establish a timeline for testing the validated software tools for all three sub-aims (CT, MRI-DWI, and PET-SUV) with phantom data from QIN sites that are already in the QUIC utility with expectation that will lead to QIN multi-center results suitable for presentation at the next QIN annual meeting.

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 as a resource for method testing and validation from completed ACRIN research for assessment of QIN metrics, validation methods, and challenge projects. Outcomes and progression data are being be made available for correlation with computational findings.

Aim 3: The EA QIN Resource Center along with other QIN investigators who also hold NCTN leadership positions will hold QIN sessions and disease-site discussions in the NCTN groups on a rotating basis. A formal session and specific disease site participation is planned at the summer 2018 NRG Meeting.

PUBLICATIONS AND PRESENTATIONS FROM QIN INVOLVEMENT

The team wrote a white paper summarizing the December 2016 planning meeting was written and. The possibility of further developing this document into a manuscript is being explored.

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-candidate Analysis 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 third annual report on our research, we continue to pursue the goal of our research, which 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 are investigating 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. We have modified our 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, as well as those from deep learning with convolutional neural networks (CNN). Also, we evaluated one of our specific radiomic feature – the most-enhancing tumor volume -- to predict recurrence-free survival "early on" in neoadjuvant treatment of breast cancer. The C-statistics for the association of METV with recurrence-free survival were 0.69 with 95% confidence interval of [0.58; 0.80] at pre-treatment and 0.72 [0.60; 0.84] at early treatment. In conclusion, we showed that the performance of the automatically-calculated METV rivaled that of a semi-manual model described for the ACRIN 6657 study (published C-statistic 0.72 [0.60; 0.84]), which involved the same dataset but required semi-manual delineation of the functional tumor volume (FTV) and knowledge of the pre-surgical residual cancer burden. We also investigated the robustness of our radiomic features across MRI magnet strength (1.5 T and 3 T), continued our radiogenomics investigation, and further developed deep learning techniques for breast MRI.

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 had 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, as well as now, those from deep learning with convolutional neural networks (CNN).

We conducted further investigations on radiomics, specifically the most-enhancing tumor volume by MRI radiomics to predict recurrence-free survival "early on" in neoadjuvant treatment of breast cancer. The hypothesis of this study was that MRI-based radiomics has the ability to predict recurrence-free survival "early on" in breast cancer neoadjuvant chemotherapy. A subset, based on availability, of the ACRIN 6657 dynamic contrast-enhanced MR images was used in which we analyzed images of all women imaged at pre-treatment baseline (141 women: 40 with a recurrence, 101 without) and all those imaged after completion of the first cycle of chemotherapy, i.e., at early treatment (143 women: 37 with a recurrence vs. 105 without). Our method was completely automated apart from manual localization of the approximate tumor center. The most enhancing tumor volume (METV) was automatically calculated for the pre-treatment and early treatment exams. Performance of METV in the task of predicting a recurrence was evaluated using ROC analysis. The association of recurrence-free survival with METV was assessed using a Cox regression model controlling for patient age, race, and hormone receptor status and evaluated by C-statistics. Kaplan-Meier analysis was used to estimate survival functions. The C-statistics for the association of METV with recurrence-free survival were 0.69 with 95% confidence interval of [0.58; 0.80] at pre-treatment and 0.72 [0.60; 0.84] at early treatment. The hazard ratios calculated from Kaplan-Meier curves were 2.28 [1.08; 4.61], 3.43 [1.83; 6.75], and 4.81 [2.16; 10.72] for the lowest quartile, median quartile, and upper quartile cut-points for METV, respectively. In conclusion, we showed that the performance of the automatically-calculated METV rivaled that of a semi-manual model described for the ACRIN 6657 study (published C-statistic 0.72 [0.60; 0.84]), which involved the same dataset but required semi-manual delineation of the functional tumor volume (FTV) and knowledge of the pre-surgical residual cancer burden. This work is currently under review for publication.

§ Relationship of MRI phenotypes to genomics

We have extended our previously-reported research, which involved the Breast Phenotype Group and cases 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. 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. Among the many novel findings, we discovered some highly specific imaging-genomic associations, 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.

We report here, our research now with UChicago cases on radiogemonics of breast cancer using DCE-MRI and gene expression profiling. Imaging techniques can provide information about the tumor non-invasively and have been shown to provide information about the underlying genetic makeup. Correlating image-based phenotypes (radiomics) with genomic analyses is an emerging area of research commonly referred to as "radiogenomics" or "imaging-genomics". The purpose of this study was to assess the potential for using our automated, quantitative radiomics platform on magnetic resonance (MR) breast imaging for inferring underlying activity of clinically relevant gene pathways derived from RNA sequencing of invasive breast cancers prior to therapy. We performed quantitative radiomic analysis on 47 invasive breast cancers based on dynamic contrast enhanced 3 Tesla MR images acquired within three months of surgery and obtained gene expression data by performing total RNA sequencing on corresponding fresh frozen tissue samples. We used gene set enrichment analysis to identify significant associations between the 186 gene pathways and the 38 image-based features that have previously been validated. All radiomic size features were positively associated with multiple replication and proliferation pathways and were negatively associated with the apoptosis pathway. Gene pathways related to immune system regulation and extracellular signaling had the highest number of significant radiomic feature associations, with an average of 18.9 and 16 features per pathway, respectively. Tumors with upregulation of immune signaling pathways such as T-cell receptor signaling and chemokine signaling as well as extracellular signaling pathways such as cell adhesion molecule and cytokine-cytokine interactions were smaller, more spherical, and had a more heterogeneous texture upon contrast enhancement. Tumors with higher expression levels of JAK/STAT and VEGF pathways of had more intratumor heterogeneity in image enhancement texture. Other pathways with robust associations to image-based features include metabolic and catabolic pathways. We provide further evidence that MR imaging of breast tumors can infer underlying gene expression by using RNA sequencing. Size and shape features were appropriately correlated with proliferative and apoptotic pathways. Given the high number of radiomic feature associations with immune pathways, our results raise the possibility of using MR imaging to distinguish tumors that are more immunologically active, although further studies are necessary to confirm this observation. We are currently submitting this work for publication.

§ Robustness of MRI phenotypes

This past year, we focused on MRI magnet strength as we continued to investigate the robustness of our computer-extracted MRI lesion phenotypes. Radiomics features extracted from breast lesion images have shown potential in diagnosis and prognosis of breast cancer. As clinical institutions transition from 1.5 T to 3.0 T magnetic resonance imaging (MRI), it is helpful to identify robust features across these field strengths. In this study, dynamic contrast-enhanced MR images were acquired retrospectively under IRB/HIPAA compliance, yielding 738 cases: 241 and 124 benign lesions imaged at 1.5 T and 3.0 T and 231 and 142 luminal A cancers imaged at 1.5 T and 3.0 T, respectively. Lesions were segmented using a fuzzy C-means method. Extracted radiomic values for each group of lesions by cancer status and field

strength of acquisition were compared using a Kolmogorov-Smirnov test for the null hypothesis that two groups being compared came from the same distribution, with p-values being corrected for multiple comparisons by the Holm-Bonferroni method. Two shape features, one texture feature, and three enhancement variance kinetics features were found to be potentially robust. All potentially robust features had areas under the receiver operating characteristic curve (AUC) statistically greater than 0.5 in the task of distinguishing between lesion types (range of means 0.57-0.78). The significant difference in voxel size between field strength of acquisition limits the ability to affirm more features as robust or not robust according to field strength alone, and inhomogeneities in static field strength and radiofrequency field could also have affected the assessment of kinetic curve features as robust or not. Vendor-specific image scaling could have also been a factor. These findings will contribute to the development of radiomic signatures that use features identified as robust across field strength. This work is currently under review for publication.

§ Role of deep learning in assessing response to therapy

We are investigating the role of different CNN structures along with methods to extract features from CNN image analysis pipelines. Deep learning methods for radiomics/computer-aided diagnosis (CADx) are often prohibited by small datasets, long computation time, and the need for extensive image preprocessing. We reported on a methodology that extracts and pools low- to mid-level features using a pre-trained convolutional neural network and fuses them with handcrafted radiomic features computed using conventional CADx methods. Our fusion-based method demonstrates significant improvements to previous breast cancer CADx methods across three clinical imaging modalities (dynamic contrast-enhanced MRI, full-field digital mammography, and ultrasound) in terms of predictive performance in the task of estimating lesion malignancy. This work was published in MEDICAL PHYSICS.

PLANS FOR NEXT YEAR

We will continue evaluating our MRI radiomic features relative to pathologic response to treatment on the I-SPY 1 dataset, a UChicago dataset, and potentially the I-SPY 2 dataset (for which we have requested access). 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 using multiple breast MRI datasets. Within all our efforts, we use rigorous statistical techniques such leave-one-out-by-case jacknifing and boot strapping to assess performance levels.

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, RSNA, and the IWDM/IWBI.

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.
8. Li H, Giger ML, Huynh BQ, Antropova, NO: Deep learning in breast cancer risk assessment: evaluation of convolutional neural networks on a clinical dataset of full-field digital mammograms. J Med Imaging 4(4), 041304 (2017), doi: 10.1117/1.JMI.4.4.041304, 2017
9. Antropova N, Huynh BQ, Giger ML: A deep fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Medical Physics online doi.org/10.1002/mp.12453, 2017.
10. Mendel K, Li H, Lan L, Cahill C, Rael V, Abe H, Giger ML: Quantitative texture analysis: robustness of radiomics across two digital mammography manufacturers' systems. J Med Imaging 5(1), 011002 (2017), doi: 10.1117/1.JMI.5.1.011002, 2017.
11. Sutton EJ, Huang EP, Drukker K, Burnside ES, Li H, Net JM, Rao A, Whitman GJ, Zuley M, Ganott M, Bonaccio E, Giger ML, Morris EA and TCCGA Group. Breast MRI radiomics: comparison of computer- and human-extracted imaging phenotypes. European Radiology Experimental 22:1-10, 2017.

U01 CA142565: Quantitative MRI for Predicting Response of Breast Cancer to Neoadjuvant Therapy

Vanderbilt University and the University of Chicago

Richard Abramson, M.D. Vanderbilt University
Greg Karczmar, Ph.D. University of Chicago
Rita Nanda, M.D. University of Chicago
Thomas E. Yankeelov, Ph.D. University of Texas at Austin

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. 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 seek to validate and then extend these results in two multi-site clinical trials. Success would contribute to the development of personalized, adaptive treatment strategies for breast cancer patients undergoing NAT.

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. Furthermore, knowing the optimal timing for image acquisition during NAT would help maximize the predictive ability of quantitative MRI.

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. More recently we have deployed these techniques in a multi-site clinical trial executed between Vanderbilt University (R. Abramson) and The University of Chicago (G. Karczmar and R. Nanda), with data analytics provided by The University of Texas at Austin (T. Yankeelov) and are currently accruing patients at both institutions.

PROGRESS OVER THE PAST YEAR

§ Data Acquisition

The data acquisition methods for this application require scanning patients up to four times before and early during the course of neoadjuvant therapy. At the time of this study, we have acquired data on 20 cases (13 at Vanderbilt University and seven at The University of Chicago). The data acquired during these imaging sessions include: 1) quantitative T_1 maps, 2) quantitative DCE-MRI data appropriate for pharmacokinetic modeling, 3) ADC maps, and 4) quantitative magnetization transfer data. Please see **Figure 1** (next page) for illustrative examples of these data in a responder and non-responder. Analysis of these data is ongoing. Our first step is focused on updating the population based AIF given the higher temporal resolution data that we are now collecting.

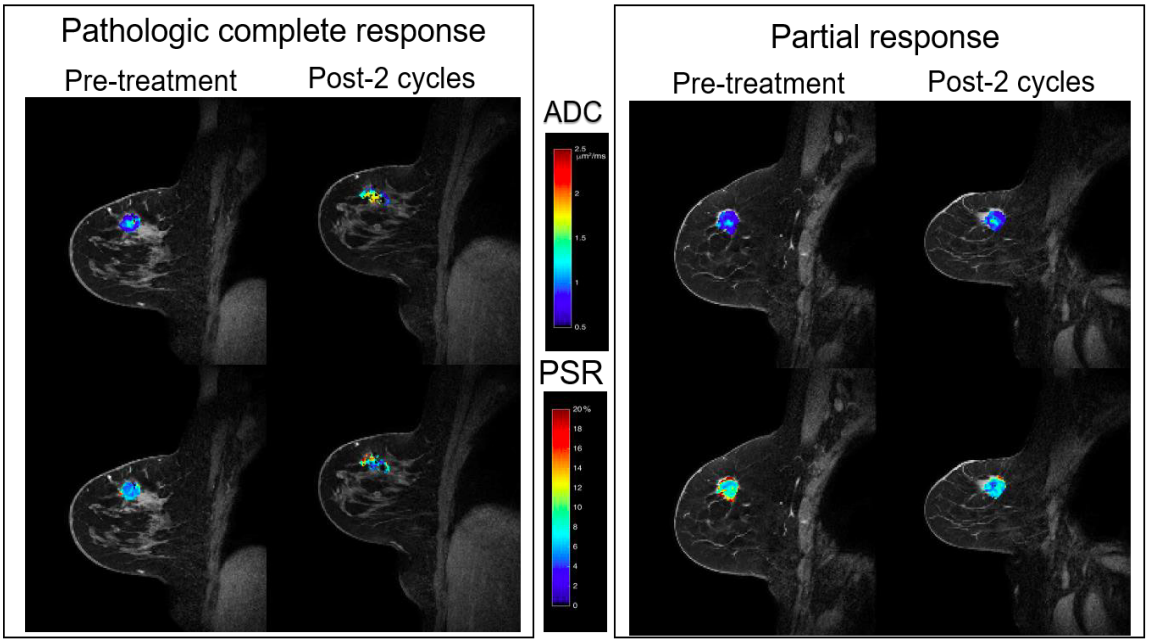


Figure 1. Example ADC and PSR maps for a central tumor slice are shown for a patient achieving a pCR (left) and a patient with residual disease (right panel) at the conclusion of NAT. After two cycles of NAT, mean tumor ADC increased for the patient who went on to pCR (0.95 to 1.4 $\mu m^2/ms$) and remained relatively the same for the patient who achieved only a partial response (0.85 to 0.90 $\mu m^2/ms$). The PSR is macromolecular-to-free water proton pool size ratio and represents the first application of this technique to assessing the changes in breast tumors in response to NAT. After two cycles of NAT, mean tumor PSR increased for the patient who went on to pCR (6.6% to 8.9%) and decreased for the patient who achieved only a partial response (11.3% to 7.3%).

We performed a meta-analysis to assess the prognostic value of quantitative DCE-MRI and DW-MRI performed during neoadjuvant therapy (NAT) of locally advanced breast cancer. A systematic literature search was conducted to identity studies of quantitative DCE-MRI and DW-MRI performed during breast cancer NAT that report the sensitivity and specificity for predicting pathological complete response. Across ten studies which met the

inclusion criteria for this meta-analysis (out of 325 initially identified studies), we find that MRI had a pooled sensitivity of 0.86 (95% CI, 0.71-0.94) and specificity of 0.82 (95% CI, 0.68-0.91) when adjusted for covariates. Quantitative DCE-MRI exhibits greater specificity for predicting pathological complete response than semi-quantitative DCE-MRI ($p < 0.001$). However, there is a high degree of heterogeneity in published studies highlighting the lack of standardization in the field. This effort was published in the *Journal of Medical Imaging*.
We have recently completed an effort to combine multi-parametric MRI with tumor cell receptor information to predict pathologic complete response. In particular, we employed ADC and k_{ep}, along with the status of three hormonal receptors (i.e., ER, PR, and HER2) in our predictive model. The overfitting-corrected area under the curve (AUC) and Brier score of our proposed logistic ridge model was 0.92 (95% CI: 0. 67, 0.99) and 0.11 (95% CI: 0.049, 0.19), respectively. Also, the same statistics computed via 10-fold cross-validation were 0.98 (95% CI: 0.69, 0.99) and 0.076 (95% CI: 0.054, 0.17), respectively. This effort was published in the Journal of Medical Imaging.

We published our effort on using a Block-Siegert B_1 mapping technique to improve the accuracy and precision of T_1 measurements of the breast at 3T. This was published in Tomography and is implemented in all of our ongoing patient studies. We tested the reproducibility and repeatability of quantitative magnetization transfer MRI in healthy volunteers.

In addition to the above studies, we have also begun applying the experience and expertise we have gained in the breast studies to implement these techniques in other disease sites. We are pleased to note that this was done in a Phase 1 clinical trial involving vorinostat and chemoradiation and capecitabine in pancreatic cancer. It was published in Radiotherapy and Oncology. We also participated in a Phase 1 clinical trial involving Dual SRC and EGFR inhibition in pancreatic cancer that was published in Investigational New Drugs. Furthermore, we have leveraged the resources and techniques from this program to assist in the quantitative imaging of brain tumor patients working with an affiliate member QIN team form the Barrow Neurological Research Institute which resulted in two abstracts presented at the 2017 ISMRM conference.

§ Data Sharing

We take the notion of data sharing very seriously and towards this end have always made it a priority to upload our data to The Cancer Imaging Archive (TCIA). At the time of this submission, we have uploaded 67 data sets and are waiting to upload an additional 13 data sets once we received the new CTP software from TCIA.

COLLABORATION WITH OTHER NETWORK SITES

We continue to have extensive interactions with our sister QIN sites. We now summarize those efforts.

1) We have submitted a manuscript for publication in collaboration with Oregon Health Sciences University and the University of Washington (led by our team). The manuscript is a comparative analysis of high temporal resolution dynamic contrast-enhanced MRI (DCE-MRI) data collected in the International Breast MR Consortium (IBMC) 6883 multicenter trial was performed to distinguish benign and malignant breast tumors. We found that both quantitative pharmacokinetic (K^{trans} and k_{ep}) and semi-quantitative signal intensity (SER_{mean}) metrics discriminated benign and malignant suspicious lesions, with receiver operating characteristic (ROC) area under the curve (AUC) values of 0.71, 0.70, and 0.82 for K^{trans}, k_{ep}, and SER_{mean}, respectively ($p < 0.05$). At equal 94% sensitivity, the specificity and positive predictive value (PPV) of SER_{mean} (53% and 63%, respectively) were higher than clinical MRI interpretation (32% and 54%).

2) We contributed to the impact of arterial input function determination on the analysis of DCE-MRI data from prostate cancer (led by Dr. Huang of OHSU). This was published in Tomography.

3) We contributed to developing multi-site concordance of DW-MRI metrics (led by Dr. Newitt from UCSF). This was presented at the 2017 ISMRM conference and the manuscript is currently being composed.

4) We contributed to an assessment of the inter-platform reproducibility of T_1 quantification methods for DCEMRI (led by Dr. Bain from Mt. Sinai). This was presented at the 2017 ISMRM conference and the manuscript is in revision at Magnetic Resonance in Medicine following a favorable review.

5) We contributed to an assessment of the effects of AIF quantification variations in DCE-MRI for predicting soft tissue sarcoma response (led by Dr. Huang of OHSU). This was presented at the 2017 ISMRM conference.

6) We contributed to an assessment of AIF estimation on shutter-speed analysis (a model which Dr. Yankeelov developed during his Ph.D. thesis) in kinetic modeling of DCE-MRI data from prostate cancer (led by Dr. Huang of OHSU). This was presented at the 2017 ISMRM conference.

7) We contributed to an effort focused on a retrospective system-specific correction of gradient nonlinearity bias for quantitative DW-MRI data obtained across diverse scanners (led by Dr. Chenevert from the University of Michigan). This was published in Tomography.

8) Members of our Vanderbilt team currently hold a subaward site for Stanford's QIN U01 award (U01CA190214; PI: Daniel Rubin)

We have also implemented, as exploratory methods, related to 1) imaging cell size, and 2) ultra-fast DCE-MRI. We now provide brief updates on both of these techniques.

§ ECOG-ACRIN study

The ECOG-ACRIN 1141 trial is a national trial comparing abbreviated breast MRI and digital breast tomosynthesis in breast cancer screening in women with dense breasts. This trial offers the option of including high-temporal-resolution imaging during the early post-contrast phase of the abbreviated MR. Because of our expertise with quantitative DCE-MRI and specifically ultrafast breast DCE-MRI at the University of Chicago, we were asked to coordinate between the advanced imaging sites, and standardize protocols this portion of the abbreviated MRI. We worked with sites across the country to develop imaging protocols for each of the major vendors, and reviewed the images from these sites to ensure their quality. After the trial is completed we will coordinate the analysis of the high-temporal-resolution images across the advanced imaging sites, with the goal of determining the value that ultrafast imaging may add to breast cancer screening with Ab-MRI, and use of DCE-MRI to evaluate response to therapy. Our leadership role in this national trial was made possible by the U01 funding to develop and test quantitative MRI methods.

§ Ultra-fast DCE-MRI

It is well established that DCE-MRI provides important information on tissue vascular properties and lesion morphology. Our team has been investigating the advantages of ultrafast (temporal resolution of 1-3 sec) DCE-MRI to 1) segment the vascular tree within the breast, 2) identify the direction of flow in the vessels feeding and draining the tumor, 3) employ computational fluid dynamics to model flow within the tumor. **Figure 2** (next page) shows an illustrative example of segmenting the vascular tree within the breast (left panel) and a color coding to indicate the direction of flow within each segment of the vascular tree. Next steps are to quantify the flow within each section of the vascular tree. We respectfully note that this is an entirely new approach to analyzing DCE-MRI data and that changes in the major vessels feeding and draining cancers may prove to be an early marker for response to therapy.

§ Additional Collaborations

Our team has been an extremely active participant in the QIN from its earliest days. In particular, Dr. Yankeelov served as co-Chair of the Executive Committee from 2012-2013, Chair from 2013- 2014, and co-Chair again from 2014-2015. He has led the organization and execution of two special issues of journals dedicated to QIN efforts (1,2), and facilitated the QIN's emphasis on inter-site collaborations leading to joint publications focused on building consensus standards; we are particularly proud of this latter achievement as it substantially accelerated the degree of interaction amongst QIN members. (Please note we are not, in any way, suggesting that our efforts were solely responsible for this development, but we do feel that we significantly enhanced those efforts.) In addition to our leadership role, our team has been actively engaged with the QIN working-groups and data analysis challenges. We have previously collaborated with QIN teams from the University of Washington (3-5), Oregon Health Sciences University (5-7), Stanford University (8), University of Michigan (9-12), UCSF (13), Mt. Sinai (14), and Brigham and Women's Hospital (15). More specifically, we have contributed to consensus efforts related to imaging and communication standards in medicine (16), multi-site concordance of ADC measurements (10-13), accuracy, repeatability,

and interplatform reproducibility of T_1 quantification (14), multi-site DCE-MRI quantification efforts (6,7), patient accrual patterns in imaging trials (4), as well as several other efforts in quantitative imaging in clinical trials (3-5,8,15). Though not a QIN-effort, we were also involved in the larger field's effort at building an imaging biomarker roadmap (15).
We also have an strong history of sharing our data. Our current QIN data set is located at https://wiki.cancerimagingarchive.net/display/Public/QIN-Breast and includes PET, MRI, and PET+MRI data sets. In fact, through The Cancer Imaging Archive (TCIA), our data has been shared with > 40 unique users from eight countries (Canada, China, Egypt, France, Germany, Spain, Taiwan, and the United States).

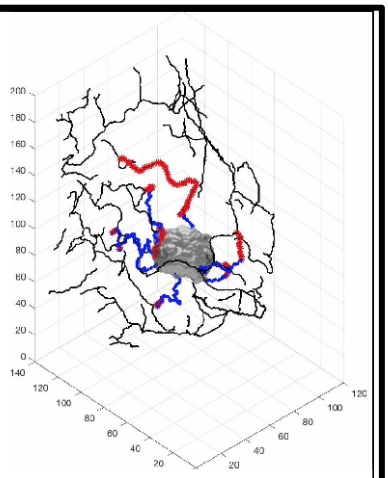


Figure 2. The black lines indicate the segmented vasculature from the entire affected breast (the chest wall is towards the back and the nipple is in the lower right of the image). The gray volume is the lesion. The red sections are identified as tumor-associated vessels, and the blue dots are the most possible paths for vessels connecting to the lesion.

PLANS FOR NEXT YEAR

In the coming year, we will pursue the following studies:

1) We will complete the data analysis of the first 14 patients we currently have in our database. We will apply the logistic ridge regression model we developed (in the proposal) using parameter values from the training data set obtained in the first period of funding.

2) We will continue to enroll patients into our study. In particular, we have a plan to increase our accrual from The University of Chicago we are doing outreach to other hospitals

in the UC system (e.g., Silvercross and Ingalls) and have added effort for a clinical trials coordinator specifically targeted for recruitment.

3) We are currently working out the logistics of sending five healthy subjects between Vanderbilt University and The University of Chicago. The goal is to perform quantitative T1 and ADC mapping so as to quantify the variances in these measurements that are included in the logistic ridge regression predictive model.

4) We will also have each institution analyze the patient data sets separately to determine if our software is robust enough to be shared with multiple sites.

5) We will begin collecting the imaging microstructural parameters using limited spectrally edited diffusion (IMPUSLED) data developed by the Vanderbilt team as part of our acquisition protocol. These data promise to provide us with new and uncorrelated potential biomarkers to work into the predictive framework.

6) We will compare T_1 maps obtained with two different methods based on variable flip angle spoiled gradient echo sequences. First T_1 maps will be obtained by fitting the VFA data corrected with B_1 maps measured using the Bloch-Siegert shift method. The second method we will use is a reference tissue method, using fat as the reference tissue. In this method, flip angle correction factors (proportional to B_1) are calculated for each fat voxel, by comparing the uncorrected T_1 value from fitting to the VFA data to a population average value of the T_1 of fat in the breast. Previous results have shown low inter- and intra-patient variability in the T_1 of fat, making it an ideal tissue to use as a reference for these measurements. A whole-breast B_1 map is then generated by interpolating from the values measured in the fat voxels. The VFA data are then fit again for the entire breast, this time using the correct flip angle for each voxel. The accuracy of both methods will be evaluated by comparison to the gold-standard values obtained from a single slice inversion recovery.

LIST OF REFERENCE

1. Yankeelov TE and Gore JC. Preface to the special issue on quantitative imaging in cancer. 2012;30:1201-2.

2. Laurence P. Clarke, Robert J. Nordstrom, Huiming Zhang, Pushpa Tandon, Yantian Zhang, George Redmond, Keyvan Farahani, Gary Kelloff, and others The Quantitative Imaging Network: NCI's Historical Perspective and Planned Goals. Translational Oncology, 2014;7:1–

3. Williams JM, Rani SD, Li X, Arlinghaus LR, Lee TC, MacDonald LR, Partridge SC, Kang H, Whisenant JG, Abramson RG, Linden HM, Kinahan PE, Yankeelov TE. Comparison of prone versus supine 18F-FDG-PET of locally advanced breast cancer: Phantom and preliminary clinical studies. Med Phys. 2015;42:3801-13.

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 2016;4:276-82.

5. Sorace AG, Partridge SC, Li X, Barnes SL, Hippe DS, Huang W, Yankeelov TE. Distinguishing benign and malignant breast tumors: preliminary comparison of kinetic modeling approaches using multi-institutional dynamic contrast-enhanced MRI data from the International Breast MR Consortium 6883 trial. Journal of Medical Imaging 2018;5:accepted.

6. Huang W, Li X, Chen Y, Li X, Chang M-C, Oborski MJ, Malyarenko DI, Muzi M, Jajamovich GH, FederovA, Tudorica A, Gupta S, Laymon CM, Marro KI, Dyvorne HA, Miller JV, 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. Translational Oncology, 2014;7:153-66.

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. 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.

9. Malyarenko D, Fedorovb A, Bell L, Prahd M, Hectorse S, Arlinghaus L, Muzig M, Jacobs M, Fungi M, Davei A, McManus K, Boss M, Tauolie B, Yankeelov TE, Quarles C, Schmainda K, Cluniel D, Chenevert T, Newitt D. Toward uniform implementation of parametric map DICOM in multi-site quantitative diffusion imaging studies. Journal of Medical Imaging, 2018;5:011006.

10. 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.

11. Malyarenko D, Newitt D, Wilmes L, Tudorica A, Helmer KG, Arlinghaus LR, Jacobs MA, Jajamovich G, Taouli B, Yankeelov TE, Huang W, Hylton N, Chenevert TL. Demonstration of Nonlinearity Bias in the Measurement of the Apparent Diffusion Coefficient in Multicenter Trials. Magnetic Resonance in Medicine, 2016;75:1312-23.

12. Chenevert TL, Malyarenko DI, Newitt D, Hylton N, Huang W, Li X, Tudorica A, Fedorov A, Fennessy F, Kikinis R, Arlinghaus L, Li X, Yankeelov TE, Muzi M, Marro KI, Kinahan PE, Jajamovich GH, Dyvorne HA, Taouli B, Kalpathy-Cramer J, Oborski MJ, Laymon CM, Mountz JM, Ross BD. Error in Quantitative Image Analysis Due to Platform-Dependent Image Scaling. Translational Oncology, 2014;7:65-71.

13. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, Bell L, Fedorov A, Fennessy F, Jacobs MA, Solaiyappan M, Hectors S, Taouli B, Muzi M, Kinahan P, Schmainda KM, Prah MA, Taber EN, Kroenke C, Huang W, Arlinghaus LR, Yankeelov TE, Cao Y, Aryal M, Yen Y-F, Kalpathy-Cramer J, Shukla-Dave A, Fung M, Liang J, Boss M, Hylton N. Multi-site concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. Journal of Medical Imaging, 2018;5:011003.

14. Bane O, Hectors Stefanie, Wagner M, Arlinghaus L, Aryal M, Cao Y, Chenevert T, Fennessy F, Huang W, Hylton N, Kalpathy-Cramer J, Keenan K, Maylarekno D, Mulkern R, Newitt, D, Russek S, Stupic K, Tudorica A, Wilmes L, Yankeelov TE, Yen Yi-Fen, Boss, M, Taoli B. Accuracy, repeatability and interplatform reproducibility of T1 quantification methods used for DCE-MRI: results from a multicenter phantom study. Magnetic Resonance in Medicine, 2017 Sep 14. doi: 10.1002/mrm.26903. [Epub ahead of print].

15. Fedorov A., Fluckiger J, Ayers GD, Li X, Gupta SN, Mulkern R, Yankeelov TE, Fennessy FM. A Comparison of Two Methods for Estimating DCE-MRI Parameters via Individual and Cohort Based AIFs in Prostate Cancer: A Step Towards Practical Implementation. Magnetic Resonance Imaging, 2014;32:321-9.

16. 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.

PUBLICATIONS FROM QIN EFFORTS

(Below we list the publications that have resulted from our QIN efforts only during the last year)

1. Sorace AG, Partridge SC, Li X, Virostko J, Barnes SL, Huang W, Hippe DS, Yankeelov TE. "Distinguishing Benign and Malignant Breast Tumors: Preliminary Comparison of Kinetic Modeling Approaches Using Multi-Institutional DCE-MRI Data from the IBMC 6883 Trial". Journal of Medical Imaging, 2018;5:011019.

2. Elhalawani H, Ger R, Mohamed ASR, Awan M, Ding Y, Li K, Fave X, Beers A, Driscoll B, Hormuth DA, van Houdt P, He R, Zhou S, Mathieu K, Li H, Coolens C, Chung C, Bankson J, Huang W, Wang J, Sandulache V, Lai S, Howell R, Stafford J, Yankeelov TE, van der

Heide U, Frank S, Barboriak D, Hazle J, Court L, Kalpathy-Cramer J, Fuller C. "Dynamic Contrast-Enhanced Magnetic Resonance Imaging for Head and Neck Cancers". Scientific Data, accepted, 2017.

3. Hormuth DA, Weis JA, Eldridgge SB, Miga MI, Quaranta V, Yankeelov TE. "Biophysical modeling of in vivo glioma response following whole brain radiotherapy in a murine model of brain cancer". International Journal of Radiation Oncology, Biology, Physics, accepted, 2017.

4. Virostko J, Hainline A, Kang H, Arlinghaus L, Abramson RG, Barnes SL, Blume JD, Avery S, Patt D, Goodgame B, Yankeelov TE, Sorace AG. "Dynamic Contrast-Enhanced MRI and Diffusion-Weighted MRI for Predicting the Response of Locally Advanced Breast Cancer to Neoadjuvant Therapy: A Meta-analysis". Journal of Medical Imaging, 2018;5:011011.

5. Kang H, Hainline A, Arlinghaus LR, Elderidge SL, Li X, Abramson VG, Chakravarthy AB, Abramson RG, Bingham B, Fakhoury K, Yankeelov TE. "Combining multi-parametric MRI with receptor information to optimize prediction of pathologic response to neoadjuvant therapy in breast cancer: Preliminary results". Journal of Medical Imaging, 2018;5:011015.
6. Woodall R, Eldridge SL, Hormuth DA, Sorace AG, Quarles CC, Yankeelov TE. "The effects of intra-voxel contrast agent diffusion on the analysis of DCE-MRI data in realistic tissue domains". Magnetic Resonance in Medicine, 2017 Nov 8. doi: 10.1002/mrm.26995. [Epub ahead of print].

7. Cardin DB, Goff LW, Chan E, Whisenant JG, Ayers GD, Takebe N, Arlinghaus LR, Yankeelov TE, Berlin J, Merchant N. "Dual Src and EGFR Inhibition in Combination with Gemcitabine in Advanced Pancreatic Cancer: Phase I Results". Investigational New Drugs, accepted, 2017.

8. Malyarenko D, Fedorovb A, Bell L, Prahd M, Hectorse S, Arlinghaus L, Muzig M, Jacobs M, Fungi M, Davei A, McManus K, Boss M, Tauolie B, Yankeelov TE, Quarles C, Schmainda K, Cluniel D, Chenevert T, Newitt D. "Toward uniform implementation of parametric map DICOM in multi-site quantitative diffusion imaging studies". Journal of Medical Imaging, accepted, 2017.

9. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, Bell L, Fedorov A, Fennessy F, Jacobs MA, Solaiyappan M, Hectors S, Taouli B, Muzi M, Kinahan P, Schmainda KM, Prah MA, Taber EN, Kroenke C, Huang W, Arlinghaus LR, Yankeelov TE, Cao Y, Aryal M, Yen Y-F, Kalpathy-Cramer J, Shukla-Dave A, Fung M, Liang J, Boss M, Hylton N. "Multi-site concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network". Journal of Medical Imaging, accepted, 2017.

10. Sorace AG, Harvey S, Syed A, Yankeelov TE. "Imaging Considerations and Interprofessional Opportunities in the Care of Breast Cancer Patients in the Neoadjuvant Setting". Seminars in Oncology Nursing, 2017:33:425-39.

11. Ger R, Mohamed ASR, Awan M, Ding Y, Li K, Fave X, Beers A, Driscoll B, Elhalawani H, Hormuth D, van Houdt P, He R, Zhou S, Mathieu K, Li H, Coolens C, Chung C, Bankson J, Huang W, Wang J, Sandulache V, Lai S, Howell R, Stafford J, Yankeelov TE, van der Heide U, Frank S, Barboriak D, Hazle J, Court L, Kalpathy-Cramer J, Fuller C. "A Multi-Institutional Comparison of Dynamic Contrast-Enhanced Magnetic Resonance Imaging Parameter Calculations". Scientific Reports, 2017:7:11185.

12. Bane O, Hectors Stefanie, Wagner M, Arlinghaus L, Aryal M, Cao Y, Chenevert T, Fennessy F, Huang W, Hylton N, Kalpathy-Cramer J, Keenan K, Maylarekno D, Mulkern R, Newitt, D, Russek S, Stupic K, Tudorica A, Wilmes L, Yankeelov TE, Yen Yi-Fen, Boss, M, Taoli B. "Accuracy, repeatability and interplatform reproducibility of T1 quantification methods used for DCE-MRI: results from a multicenter phantom study." Magnetic Resonance in Medicine, 2017 Sep 14. doi: 10.1002/mrm.26903.

13. Wang D, Arlinghaus LR, Yankeelov TE, Yang X, Smith D. "Quantitative Evaluation of Temporal Regularizers in Compressed Sensing Dynamic Contrast Enhanced MRI of the Breast." International Journal of Biomedical Imaging, 2017;2017:7835749.

14. 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.

CIHR-funded QIN U01: 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., Ur Metser M.D., Jennifer Gottwald M.S., Brandon Driscoll M.S., Tina Shek M.S., Julia Publicover M.S., Viktor Iakovenko Ph.D.

Investigators: Anthony Fyles M.D., Brad Wouters Ph.D., Neesha Dhani M.D., David Hedley M.D., Michael Milosevic M.D., David Palma M.D. Ph.D., Anthony Joshua MBBS Ph.D., Aaron Ward Ph.D., Leonard Luyt Ph.D., Eugene Wong Ph.D., John Valliant Ph.D., Paul Boutros Ph.D., Mattea Welch M.S.

INTRODUCTION

Hypoxia in solid tumors 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.

Over the last year we have advanced methods and models for hypoxia quantification with static and dynamic PET imaging. Our model is able to quantify the presence of slow-equilibrating regions and also the hypoxia-sensitive FAZA binding rate, potentially enhancing the sensitivity of hypoxia-PET imaging. Future work will aim to validate this model against immunohistochemical staining of resected tumors. We believe the ancillary information about tumor transport properties provided by dynamic PET imaging may prove to be clinically relevant in its own right, and we are seeking to develop MRI and CT biomarkers for tissue transport. Also we have developed a dynamic CT scanning with contrast injection method (CT Perfusion) to image all three functions – lung ventilation and perfusion plus myocardial perfusion simultaneously from one contrast injection and one ECG gated scanning session of 40 s duration without breath-hold. These results show that our new scanning and analysis method has the potential to image lung perfusion and ventilation together with myocardial perfusion conveniently in a single study. Monte-Carlo simulation for Aim 1 were modified and advanced with additional capabilities. This will guide a validation protocol. Further physics validation experiments will be performed to ensure the quantitative accuracy of the MC simulation platform. Our team has developed a phantom which can be used to improve standardization of hypoxia measurements across institutions and plans on initiating a QIN challenge where a phantom is imaged at participating sites in the QIN CT/PET subgroup. This phantom will be sent around to sites across North America to compare results obtained with

different PET scanners. Finally, we have continued the impact assessment of progressively more quantitative hypoxia imaging methods on the predictive capacity of hypoxia biomarkers in 4 clinical trials.

PROGRESS OVER THE PAST YEAR

§ Specific aim 1. Establish a robust and reliable methodology for PET hypoxia imaging

Sub-aim 1.1: Quantifying hypoxia using static PET imaging

A compartmental model based on a reaction-diffusion equation was used to study fluoroazomycin (FAZA) pharmacokinetics (binding and transport) and in turn, to assess the challenges in using a single static PET scan to quantify hypoxia [3]. Assuming that local diffusive equilibrium is achieved rapidly within the region of interest (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 tumor 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.

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).

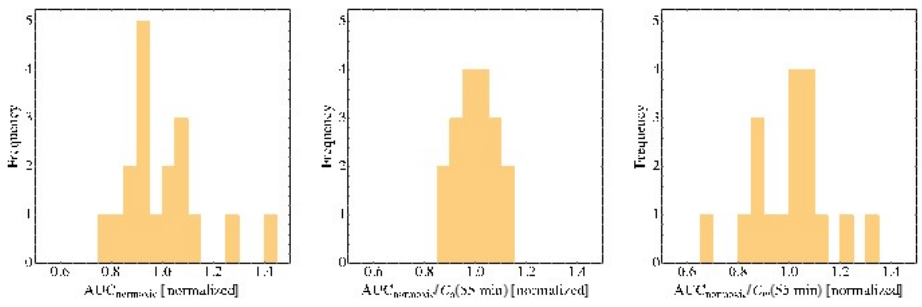


Figure 1: Normalized AUC for nineteen pancreas tumors.

Sub-aim 1.2: Quantifying hypoxia using dynamic PET imaging

While normalizing static PET-hypoxia images by uptake in blood can enhance sensitivity to hypoxia by removing inter-patient transport inhomogeneities, intra-patient variances remain. Over the past two years, we have developed a novel compartmental model to quantify the impact of such variances [4]. 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 [5] and immunohistochemical staining of resected pancreas tumors. Our major results so far are:

Static PET imaging of FAZA at two hours after injection reliably quantifies hypoxia as long as the imaged tissue is devoid of substantial mucous deposits, micro-necrosis, or fat, and the metric used for hypoxia quantification is the tumor-to-blood FAZA uptake ratio and not the tumor-to-muscle value [3].

For tumors exhibiting substantial mucous deposits, micro-necrosis, or fat, slow diffusive tracer equilibration reduces the sensitivity of static PET imaging to hypoxia and a suitably-interpreted dynamic PET scan (i.e., using our novel compartmental model) improves upon this sensitivity.

The effect of having regions in which tracer equilibrates slowly can be understood from Figure 2. The left panel shows the FAZA tumor-to-blood uptake ratio at two hours (static PET image) versus the FAZA "trapping rate" K3 derived from a standard three-tissue (blood, unbound and bound in cells) compartment model analysis of voxel-scale dynamic PET data for a single tumor. Figures analogous to this one constitutes a standard figure-of-merit for static PET imaging: assuming that the trapping rate is a sensitive metric for the biological process of interest, static PET imaging is sensitive to this biology when correlations between uptake and K3 are strong. As evident in the left panel of Figure 2, correlations for FAZA are weak. Within our model, this is attributed to regions of slow-equilibration, as quantified by a small distribution volume. Correcting for this quantity, correlations become strong (right panel).

Our model is able to quantify the presence of slow-equilibrating (likely mucous-filled) regions and also the hypoxia-sensitive FAZA binding rate, potentially enhancing the sensitivity of hypoxia-PET imaging. Future work will aim to validate this model against immunohistochemical staining of resected tumors (see below). At the same time, we believe the ancillary information about tumor transport properties provided by dynamic PET imaging may prove to be clinically relevant in its own right, and we are seeking to develop MRI and CT biomarkers for tissue transport (see Figure 3).

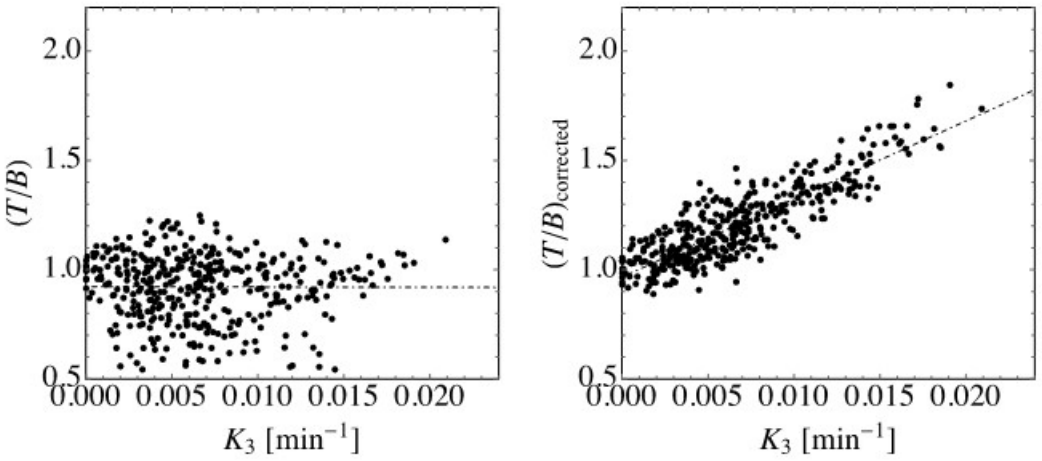


Figure 2: Left: FAZA tumor-to-blood uptake ratio (T/B) versus the net trapping rate K_3 for all voxels in a representative tumor. Right: Uptake ratio corrected for partitioning.

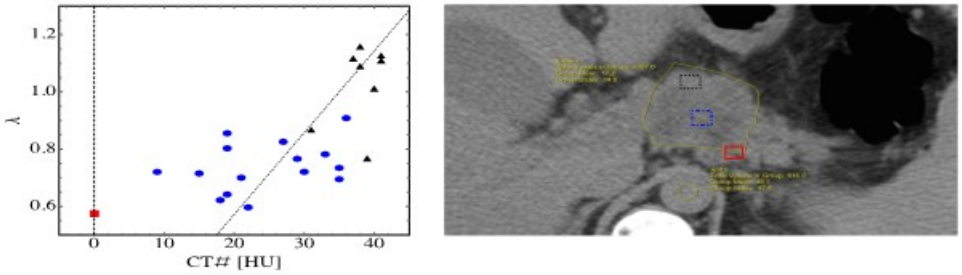


Figure 3: Partition coefficient versus CT number (Hounsfield units) for ROIs in pancreatic tumors.

Sub-aim 1.3: Tracer kinetic models for dynamic PET imaging analysis

Last year we extended the conventional closed three-compartment model (Figure 4 (A)) to include the perfusion through the vasculature of the tissue (Figure 4 (B)). Unlike the conventional closed three-compartment (C-3C) model, the modification with the Johnson-Wilson-Lee (JWL) model allows for the effect of tracer delivery by blood flow on the target binding rate constant to be included in the modelling. This is achieved by replacing the blood compartment with the JWL model and retaining the two tissue compartments to give the JWL-2TC model as shown in Figure 4 (B).

We continued to improve the stability of estimated parameters from fitting the JWL-2TC model to dynamic PET time-activity curves. We used computer simulation in this

investigation wherein theoretical time-activity curves (TACs) were generated with known JWL-2TC model parameters and Gaussian noise at the level typically observed in dynamic PET studies was added, noisy simulated TACs, up to 50 different curves, were fitted with the JWL-2TC model, to define bias (deviation from truth) and coefficient of coefficient (COV, ratio of standard deviation to mean parameter estimate) of estimated parameters. The same TACs were also fitted with C-3C model (Figure 4 (A)) to define the effect on the model parameters by leaving out the blood flow delivery effect. This latter fitting was performed with a commercially available tracer kinetics modeling software (PMOD Technologies LLC). Table 1 compares the bias and COV of the estimated parameters using the PMOD software and our in-house software based on the JWL-2TC model (Figure 4 (B)). Ignoring the blood flow delivery effect in fitting TACs generated with JWL-2TC model would lead to bias in the estimated model parameters of the order of 10% and higher. Therefore, in order to have accurate estimates of target binding (k_3) and washout (k_4) rate constant, the blood flow delivery effect has to be modeled in the fitting method. On the other hand, the COV of estimated parameters is less dependent on the blood flow delivery effect as fitting with either C-3C or JWL-2TC model gave the same COV.

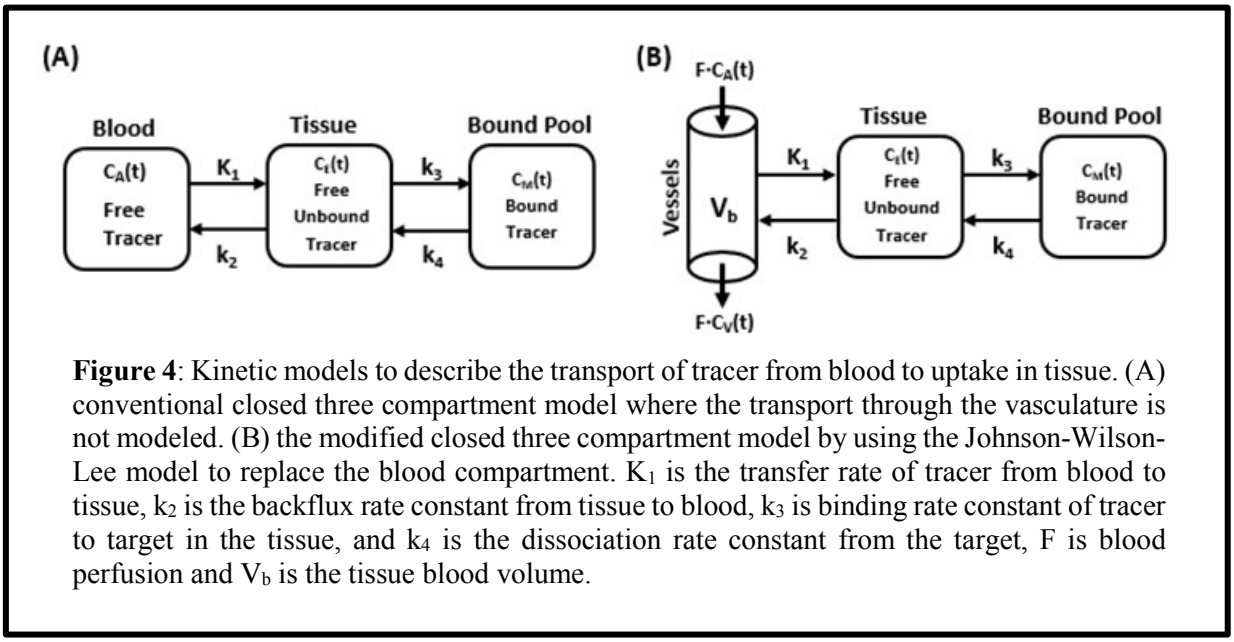


Figure 4: Kinetic models to describe the transport of tracer from blood to uptake in tissue. (A) conventional closed three compartment model where the transport through the vasculature is not modeled. (B) the modified closed three compartment model by using the Johnson-Wilson-Lee model to replace the 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.

	F		K_1		k_2		k_3		k_4	
SOFTWARE	***Bias (%)***	***COV (%)***	***Bias (%)***	***COV (%)***	***Bias (%)***	***COV (%)***	***Bias (%)***	***COV (%)***	***Bias (%)***	***COV (%)***
PMOD	-	-	21.8	2.4	25.6	9.1	-8.0	10.0	-12.5	3.2
In-house	0.28	3.9	0.16	2.6	1.1	9.1	1.2	9.8	0.17	3.9

Table 1: Bias and COV of parameters estimated from simulated TACs generated with theJWL-2TC model shown in Figure 4 (B).

One cancer site of interest to the London (Ontario) node is NSCLC and its treatment with external beam radiation therapy. In this application, besides the need of accounting for blood flow delivery effect, the definition of treatment target and monitoring of treatment response would require imaging of lung perfusion and ventilation as well as myocardial perfusion. The last functional assessment monitors for radiation induced cardiotoxicity which can manifest as myocardial ischemia at rest and/or decrease in myocardial perfusion reserve at stress. For this purpose, in the last year we have developed a dynamic CT scanning with contrast injection method (CT Perfusion) to image all three functions – lung ventilation and perfusion plus myocardial perfusion simultaneously from one contrast injection and one ECG gated scanning session of 40 s duration without breath-hold. Figure 5 shows representative myocardial perfusion and lung perfusion and ventilation maps from a pig study. The pig was rendered ischemic in the mid-septal region of the heart to simulate that from radiation induced cardiotoxicity. Figure 5 (A) shows the expected ischemia in the mid-septal region and Figure 5 (B) & (C) show the expected gravity dependence of perfusion and ventilation in normal lungs. These results show that our new scanning and analysis method has the potential to image lung perfusion and ventilation together with myocardial perfusion conveniently in a single study. Another significant consideration from the above results is that with a PET/CT scanner, a CT Perfusion study can be used to estimate F and V_b of the JWL-2TC model to reduce the number of parameters to be estimated from dynamic PET TAC from 6 to 4. This would further enhance the stability of the estimated parameters including the binding rate constant (k_3) to hypoxic targets.

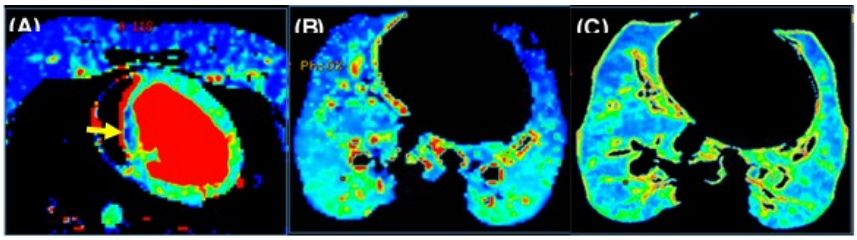


Figure 5: Functional mapping in the thorax. (A) Myocardial perfusion. (B) Lung perfusion. (C) Lung ventilation.

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. To this end, a controllable multi-modal flow phantom has been built to calibrate and standardize functional imaging protocols.

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 "gold standard" AIF measurements using Gadolinium in volumetric DCE-CT where signal to concentration relationships are linear.

The impact of individualized magnitude and phase signal AIF measurements on resulting perfusion parameter maps was assessed using a validated 4D temporal dynamic analysis (TDA) method in metastatic brain cancer patients treated with stereotactic radiosurgery. 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.

First FAZA measurements are planned to be carried out in February 2018 on both our PET-MRI and PET-CT scanner.

Sub-aim 1.5: Monte Carlo models of 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. 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. The objective is to develop a physics-based model to simulate and correct the scatter contamination using a Monte Carlo (MC) method.

Previously established workflows based on the Monte Carlo (MC) simulation tool GEANT4 Application for Tomographic Emission (GATE) v.7.2 have been extended and upgraded to the most recent version of GATE (v.8.0) see Figure 6. This enables the integration of new workflow elements such as the Insight Toolkit as well as List Mode Data output. The ITK is a cross-platform software which performs image registration and segmentation. In combination with GATE, it will be mainly used to process DICOM images in order to set up voxelized phantoms. Previously, regions of interest (ROIs) were defined on CTAC images to

achieve correct slice spacing, essential for the image reconstruction in STIR. This module of the workflow has been extended to propose the usage of clinical contours and CT images. Also, the List Mode Data output library has been compiled against GATE and the cluster tool Condor. This enables a more general data output without loss in overall simulation time.

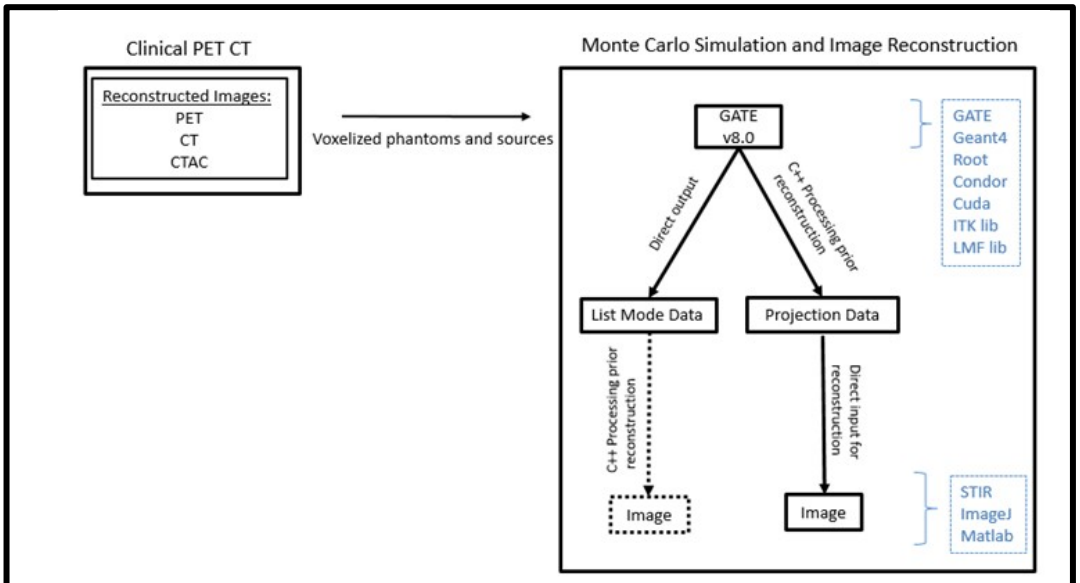


Figure 6: Overview of the Monte Carlo (MC) simulation platform integrated with the image reconstruction. Clinically reconstructed images are used to set up the voxelized phantom and source geometry in the MC simulation program GATE. The GATE platform is built upon Geant4 (physics modelling), Root (Data analysis Framework), Condor (Running GATE on a Cluster), Cuda (Parallel computing platform, GPU computing), ITK lib (Insight Toolkit-Image registration and segmentation) and LMF lib (List Mode Data). For the image reconstruction the software STIR or the GE PET toolbox can be 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. The current image reconstruction programs are STIR and the GE PET toolbox. The GE PET toolbox is provided by GE Healthcare and offers the opportunity to be as close as possible to the clinical scanner reconstruction while evaluating the quantitative performance of an accurate concentration recovery. First water cylinder phantom validation experiments confirmed the high quantitative accuracy between the clinical reconstruction and the offline reconstruction with the PET toolbox. The GE water cylinder offline reconstruction (Ordered subset expectation maximization -2-iterations, 32subsets, post-filter-6.4mm, z-filter 4mm) was performed in 10 minutes on a Linux machine with 3.5 GHz Core and 7.86 GB RAM and showed a concentration recovery within 0.1% with the injected activity on the first slice. Future steps in the research will be to develop further post-processing protocols for the List Mode Data output. This will enable the isolations of scattered events and image reconstruction in STIR as well as the GE PET toolbox. This will guide a validation protocol proposed in

Figure 7. Further physics validation experiments will be performed to ensure the quantitative accuracy of the MC simulation platform.

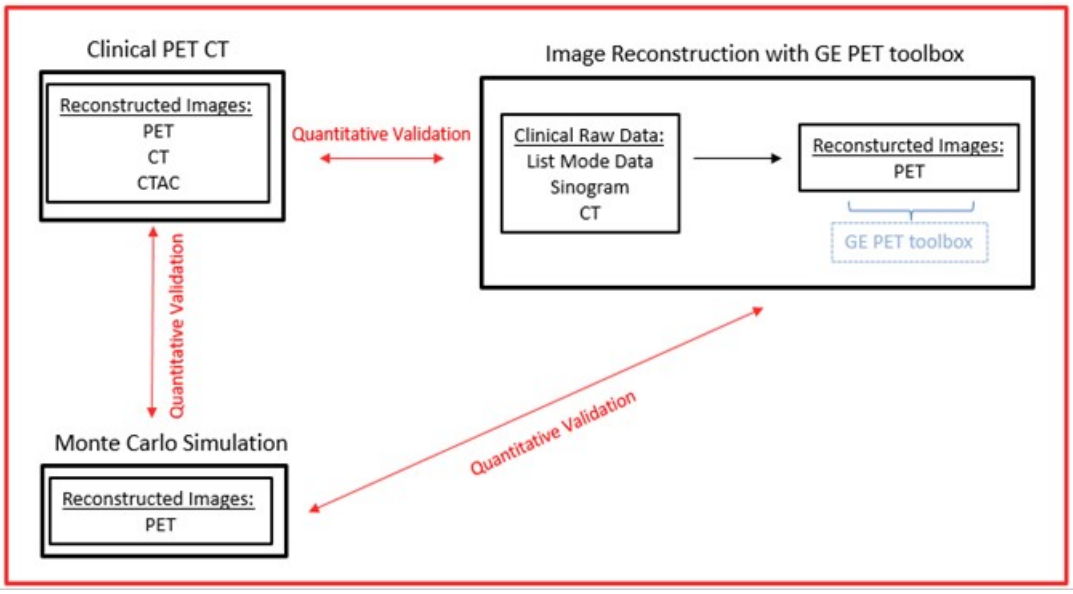


Figure 7: Validation of MC simulation after image reconstruction. MC simulation output will be compared with results obtained with the clinical scanner as well as different reconstruction software (e.g. PET toolbox).

Sub-aim 1.6: Standardization of imaging technique and characterization of scanner performance

The quantification of tumor hypoxia typically relies on a count of the number of voxels above a predetermined threshold activity over the uptake distribution in a region of interest (ROI) known as the hypoxic fraction (HF). There is currently no consensus as to how the threshold is chosen and variation in scanner characteristics will affect the uptake distribution which will directly affect HF calculation. In order to simulate various levels of tumor hypoxia as well as reference regions for comparison (i.e. blood, muscle) a phantom was developed consisting of a set of two plates separated by 2mm diameter nylon rods of different spacing densities to create negative space inside a phantom filled with F^{18} and water.

Two prototype phantoms were created, the first simulated a hypoxic tumor geometry while the second phantom had four different spatial densities of rod ranging from 10 to 40% fill. In the final version of the phantom the four regions consist of a normoxic surrogate region (40% filled with rods) and 3 regions representing various hypoxic tumor geometries.

The final version of the phantom (Figure 8) was constructed and imaged using a variety of different reconstruction parameters and scan durations. Depending on scan threshold an increase in noise can increase or decrease the hypoxic fraction even when using non- noise-

dependent thresholds. This phantom will now be used as part of a QIN challenge in the CT/PET subgroup where it is sent around to sites across North America to compare results obtained with different CT scanners. The design, fabrication and validation of a hypoxia standardization phantom is a necessary first step towards developing a standardized imaging and analysis methodology for hypoxia imaging worldwide. The phantom created is quantitative and simulates hypoxic tumors from clinical imaging. This new phantom will be able to help quantify differences in hypoxia measurements between sites as well as investigate in a more robust method of quantifying hypoxia with PET.

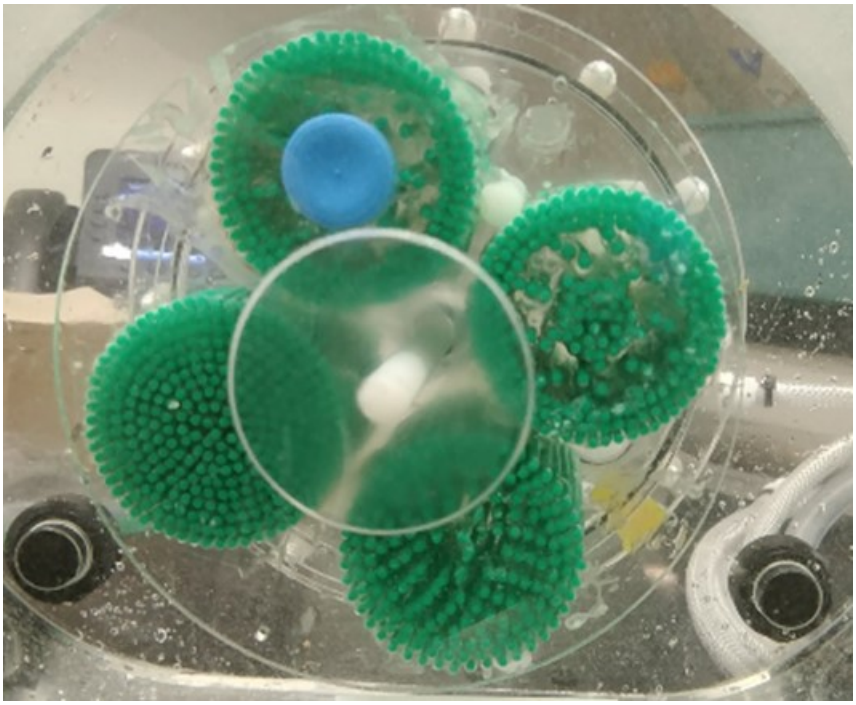


Figure 8: The hypoxic fraction standardization phantom.

Sub-aim 1.7: Advanced metrics of hypoxia

Over the last year we continued work towards standardizing of hypoxia quantification. For FAZA-Metformin trial a subsample of patients (9 out of a targeted 20) have volunteered for a blood draw to complement their FAZA-PET imaging results that include thorax PET scans. Preliminary analysis of this additional data has begun [6]. We learned that the gluteus maximus muscle from cervix cancer patients could mimic the FAZA uptake in blood draws or the left ventricle in the heart. Since blood is typically the gold standard from a kinetic perfusion analysis standpoint, a surrogate such as the gluteus maximus muscle can serve as a vehicle towards standardizing hypoxia quantification. This finding is currently part of a publication under review by IJROBP [7]. There will be biopsied tissue samples to complement hypoxic fraction calculations for all enrolled patients after the trial concludes.

§ 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 10 patients out of 30.

Sub-aim 2.2: **FAZA PET/MR imaging as a biomarker of hypoxia in rectal cancer**

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. The pilot study has started last year with current accrual 4 patient out of 10. There was a delay in patient accrual, however the PI has manage to add another site, so renewal of accrual is expected soon. Validation analysis with pimonidazole staining is planned to be done after accrual is completed.

§ Specific aim 3. Quantitative methods on image-based biomarkers to predict and assess response

Sub-aim 3.1: Data handling and informatics team

Quantitative Imaging for Personalized Cancer Medicine program has evolved to an Imaging Core Lab with a robust infrastructure to help clinical investigators achieve systematic image collection and improved collaboration. Currently we have 32 trials on our platform, 43 sites globally transferring images to our servers behind UHN firewall. These represents accrual of over 2.63 million images for approximately 600 patients in 1791 studies. QIPCM is continuously engaged in clinical imaging research and standardization initiatives in the Quantitative Imaging Network.

QIPCM, partly supported by the National Cancer Institute's QIN grant, participated in three QIN challenges; (1) kinetic modeling of head and neck patient images as well as digital reference objects, (2) comparison of radiomics analysis of lung tumors by 9 research groups and (3) comparison of dynamic FMISO-PET analysis on a set of lung tumor patient data by 4 research groups.

Also QIPCM team has developed a phantom which can be used to improve standardization of hypoxia measurements across institutions and plans on initiating a QIN challenge where the phantom is imaged at participating QIN sites.

Sub-aim 3.2: Conventional hypoxia imaging analysis

We plan to assess the impact of progressively more quantitative hypoxia imaging methods on the predictive capacity of hypoxia biomarkers in four 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 29 patients out of 30. The hypoxic volume (HV) was defined as all voxels within a tumor (T) with standardized uptake values (SUVs) greater than 3 standard deviations from the mean gluteus maximus muscle SUV value (M), or SUVs greater than 1–1.4 times the mean SUV value of the left ventricle, a blood (B) surrogate. The hypoxic fraction (HF) was defined as the ratio of the number of hypoxic voxels to the total number of tumor voxels. A 18F-FAZA PET HV could be identified in the majority of cervical tumors (89% when using T/M or T/B >1.2 as threshold) on the 2-hour static scan. The HF ranged from 0-99% (median 31%) when defined using the T/M threshold, and 0-78% (median 32%) with the T/B >1.2 threshold. HVs derived from the different thresholds were highly correlated (Spearman's correlation coefficient ρ between T/M and T/B >1–1.4 were 0.82-0.91), as were HFs (0.75-0.85). Compartmental analysis of the dynamic scans showed k3, the FAZA accumulation constant, to be strongly correlated with HF defined using the T/M (Spearman's ρ=0.72) and T/B >1.2 thresholds (0.76).

2. The Potential for Metformin to Improve Tumor Oxygenation in Locally Advanced Cervix Cancer: A Phase II Randomized Trial – current accrual 12 patients out of 48;

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 4 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.

COLLABORATIONS WITHIN THE NETWORK

- New challenge participation is related to DCE-MRI metrics in brain immunology trials. Collaborative with MD Anderson [Dr Caroline Chung].

- Our team has developed a phantom, which can be used to improve standardization of hypoxia measurements across institutions. We plan to initiate a QIN challenge in the CT/PET subgroup where the phantom will be imaged at participating QIN sites.

PLANS FOR THE NEXT YEAR

§ Specific aim 1

- Our results from dynamic PET imaging of hypoxia in pancreatic tumors emphasized that tracer uptake must be corrected for transport—quantified by the FAZA distribution volume – 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 3 shows correlations between the FAZA distribution volume and CT number in a pancreatic tumor. Future work will aim to further quantify these correlations by developing transport models for pancreatic and non-small cell lung tumors. In doing so, we can improve hypoxia quantification using only static PET images and also test and improve CT-only radiomics biomarkers of hypoxia recently proposed by the group of Philippe Lambin at Maastricht University. Currently a joint research collaboration is currently being set up. Figure 3 shows our aim to extend their work by combining CT-based radiomics analyses with static PET data sets to improve hypoxia quantification.

- Accurate measurements of AIF for dynamic PET imaging analysis will be investigated in both PET-MRI and PET-CT systems. This will allow translation of previously established CT and MRI validation data to the assessment of hypoxia accuracy in combined modality (or even PET only) imaging systems.

- Although the JWL-2TC model accounts for the blood flow delivery effect, the current implementation of this model in the fitting algorithm has a long processing time, i.e. about 30 seconds per fitting. In the coming year, we will investigate other implementations that could decrease the processing time to about 5-10 seconds per fitting.

- Our team has developed a phantom which can be used to improve standardization of hypoxia measurements across institutions and plans on initiating a QIN challenge where the phantom is imaged at participating QIN sites. This phantom will be sent around to sites across North America to compare results obtained with different CT scanners.

§ Specific aim 2.

- London site will participate in a Phase II randomize trial of "the Potential for Metformin to Improve Tumor Oxygenation in Locally Advanced Cervix Cancer". The recruited patients will undergo dynamic PET imaging with 18F-FAZA and CT Perfusion scanning. As discussed in progress report for **Sub-aim 1.3**, the CT Perfusion study will be used to estimate F and V_b of the JWL-2TC model and then the tumor TAC from the dynamic PET study will be analyzed to determine if the binding rate constant to hypoxic targets decreases with metformin treatment.

§ Specific aim 3.

- Having developed the capability to measure lung perfusion and ventilation with CT Perfusion, London site will be also investigating imaging NSCLC with ^{18}F-FAZA dynamic PET and correlate the PET findings with pimonidazole staining of explanted tumor.

- We will continue analyzing the images from 4 clinical trials for validation studies. Our group is planning to propose a QIN challenge to the network.

LIST OF REFERENCE

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. Taylor E, Gottwald J, Yeung I, Keller H, Milosevic M, Dhani NC, Siddiqui, I, Hedley WD and Jaffray DA Impact of tissue transport on PET hypoxia quantification in pancreatic tumors. *Eur J Nuc Med Mol Imag Res* (2017), 7:101.

5. Metran-Nascente C, Yeung I, Vines DC, Metser U, Dhani NC, Green D, Milosevic M, Jaffray DA, and Hedley DW. Measurement of Tumor Hypoxia in Patients with Advanced Pancreatic Cancer Based on 18F-Fluoroazomyin Arabinoside Uptake (2015). *J Nuc Med* (2015), 57(3):361.

6. Vines DC, Driscoll BD, Keller H, Shek T, Han K, Jaffray DA. 18F-FAZA PET-CT precision and accuracy of image-derived blood surrogate reference values. *J Nucl Med*. 2017;58(Suppl 1):778.

7. Kathy Han, Tina Shek, Douglass Vines, Brandon Driscoll, Anthony Fyles, David Jaffray, Harald Keller, Ur Metser, Melania Pintilie, Jason Xie, Ivan Yeung, Michael Milosevic. Measurement of Tumor Hypoxia in Patients with Locally Advanced Cervical Cancer using Positron Emission Tomography (PET) with ^{18}FFluoroazomyin Arabinoside (18F-FAZA). *International Journal of Radiation Oncology*Biology*Physics*. Submitted Dec 31 2017. Under review.

PUBLICATIONS FROM QIN EFFORTS

1. Taylor E, Gottwald J, Yeung I, Keller H, Milosevic M, Dhani NC, Siddiqui, I, Hedley WD and Jaffray DA Impact of tissue transport on PET hypoxia quantification in pancreatic tumors. *Eur J Nuc Med Mol Imag Res* (2017), 7:101.

2. Vines DC, Driscoll BD, Keller H, Shek T, Han K, Jaffray DA. 18F-FAZA PET-CT precision and accuracy of image-derived blood surrogate reference values. *J Nucl Med*. 2017;58(Suppl 1):778.

3. Kathy Han, Tina Shek, Douglass Vines, Brandon Driscoll, Anthony Fyles, David Jaffray, Harald Keller, Ur Metser, Melania Pintilie, Jason Xie, Ivan Yeung, Michael Milosevic. Measurement of Tumor Hypoxia in Patients with Locally Advanced Cervical Cancer using Positron Emission Tomography (PET) with 18FFluoroazomyin Arabinoside (18F-FAZA). *International Journal of Radiation Oncology*Biology*Physics.* Submitted Dec 31 2017. Under review.

4. Ger RB, Mohamed ASR, Awan MJ, Ding Y, Li K, Fave XJ, Beers AL, Driscoll B, Elhalawani H, Hormuth DA 2nd, Houdt PJV, He R, Zhou S, Mathieu KB, Li H, **Coolens C, Chung C**, Bankson JA, Huang W, Wang J, Sandulache VC, Lai SY, Howell RM, Stafford RJ, Yankeelov TE, Heide UAV, Frank SJ, Barboriak DP, Hazle JD, Court LE, Kalpathy-Cramer J, Fuller CD. A Multi-Institutional Comparison of Dynamic Contrast-Enhanced Magnetic Resonance Imaging Parameter Calculations. *Sci Rep*. 2017 Sep 11;7(1):11185. [IF: 4.259].

5. J Haynes, TD McKee , AC Haller , Y Wang , C Leung , DMA Gendoo , E Lima-Fernandes , A Kreso , R Wolman , Eva Szentgyorgyi , DC Vines , B Haibe-Kains , BG Wouters , U Metser, DA Jaffray , M Smith, CA O'Brien. (2018) Administration of Hypoxia-Activated Prodrug Evofosfamide after Adjuvant Therapy Enhances Therapeutic Outcome and Targets Cancer-Initiating Cells in Colorectal Cancer. *Clinical Cancer Research* - accepted, in press.

6. M Ding, T van der Kwast, R Vellanki, W Foltz, TD McKee, N Sonenberg, PP Pandolfi, M Koritzinsky, BG Wouters. The mTOR Targets 4E-BP1/2 Restrain Tumor Growth and Promote Hypoxia Tolerance in PTEN-driven Prostate Cancer. *Molecular Cancer Research* - accepted, in press.

7. Mitsuteru Yoshida, Hisashi Oishi, Tereza Martinu, David M. Hwang, Trevor D. McKee, Xiaohui Bai, Zehong Guan, Hae-Ra Cho, Stephen Juvet, Marcelo Cypel, Shaf Keshavjee, Mingyao Liu. (2018) Pentraxin 3 Deficiency Enhances Features of Chronic Lung Allograft Dysfunction in a Mouse Orthotopic Lung Transplantation Model. *Oncotarget* - accepted, In Press.

8. D Xia, R Casanova, D Machiraju, TD McKee, W Weder, AH Beck, A Soltermann. (2018) Computationally-Guided Development of a Stromal Inflammation Histologic Biomarker in Lung Squamous Cell Carcinoma. *Scientific Reports* - accepted, in press.

9. H Rezaeeyan, R Shirzad, TD McKee, N Saki. (2018) Role of Chemokines in Metastatic Niche: New insights along with a Diagnostic and Prognostic Approach. *APMIS* - accepted, in press.

10. Borst, G. R., Kumareswaran, R., Yücel, H., Telli, S., Do, T., McKee, T., Zafarana G., Jonkers J., Verheij M., O'Connor M.J., Rottenberg S., Bristow, R. G. (2017). Neoadjuvant olaparib targets hypoxia to improve radioresponse in a homologous recombination-proficient breast cancer model. *Oncotarget*, *8*(50), 87638–87646. http://doi.org/10.18632/oncotarget.20936

11. L Zhang, DC Vines, DA Scollard, T McKee, T Komal, M Ganguly, T Do, B Wu, N Alexander, R Vali, A Shammas, T Besanger, S Baruchel. (2017) Correlation of Somatostatin Receptor-2 Expression with Gallium-68-DOTA-TATE Uptake in Neuroblastoma Xenograft Models. *Contrast Media & Molecular Imaging*, vol. 2017, Article ID 9481276, 10 pages. doi:10.1155/2017/9481276

U01 CA211205: Quantitative imaging tools to derive DW-MRI oncologic biomarkers

Memorial Sloan Kettering Institute Cancer Research
Columbia University Medical Center

Amita Schukla-Dave Ph.D., MSKCC
Lawrence Schwartz, CUMC

INTRODUCTION

We propose to develop, optimize and validate novel DW-MRI acquisition and modeling methods, which address non-Gaussian water diffusion and perfusion effects through diffusion kurtosis imaging and non-Gaussian intravoxel incoherent motion imaging and provide more specific measures of tissue structure and biology. Additionally, we will develop and implement advanced image processing tools to maximize the biologic information from the tumor/tissue provided by the imaging data. The essence of our timely proposal lies in it being the first multi-center, imaging trial to identify quantitative imaging biomarkers as early response to therapy indicators, which interrogate tumor biology in accordance with the central mission of the NCI Quantitative Imaging Network (QIN). It will address an urgent, unmet need in clinical trials for recurrent/metastatic (R/M) head and neck cancers. This UO1 proposal is in response to PAR-14-116 and the specific aims outlined in the proposal are as follows: **Aim 1**: To develop and standardize a multi b-value reduced field of view (rFOV) DW-MRI acquisition method and non-mono exponential modeling DW-MRI for oncology applications; **Aim 2**: To develop and implement optimal model methodology with advanced image segmentation and image feature analysis in patients with R/M malignancies in the HN region for oncology applications; and **Aim 3**: To establish the next generation DW-MRI biomarkers as early response to therapy indicators in experimental therapies using R/M HN squamous cell carcinoma (SCC) as a proof of principle model. We hypothesize that imaging metrics derived from newer methods can be used as quantitative imaging biomarkers for assessing early therapeutic efficacy in R/M HNSCC. The principles of identifying robust, reliable and quantitative imaging biomarkers derived from DW-MRI and image feature analysis remain similar and such imaging protocols, after appropriate adaptation, can have a wider clinical application, including their use in treating other solid tumors

RESULTS

Our study has completed nearly six months and is in its first year. Our preliminary data demonstrate the feasibility of DW-MRI across sites (Memorial Sloan Kettering Cancer Center [MSKCC] and Columbia University Medical Center [CUMC]) in a reproducible manner using the same methods of MRI data acquisition. This is a prospective study of patients identified in the clinic with head and neck cancer who will undergo MRI and thereafter be treated at the respective institutions. MSKCC and CUMC have processed new institutional review board (IRB) protocols.

§ Diffusion-weighted MRI (DW-MRI) on phantom

One of the prerequisites for a multi-site study is to show that both clinical sites can acquire data using the new National Institute of Standards and Technology (NIST) and RSNA-QIBA phantom and perform multi b-value DW-MRI studies on both 1.5 and 3T GE MRI scanners. The phantom is constructed using varying concentrations of polyvinylpyrrolidone (PVP) in an aqueous solution to generate physiologically relevant ADC values. The imaging protocol was performed 4 times for repeatability, based on the guidelines provided by QIBA, and b-values of 0, 500, 900 and 2000 sec/mm^2 were used to obtain the composite ADC. Drs. Shukla-Dave (MSKCC, PI), and Jambawalikar (Chief Medical Physicist, CUMC, Co-investigator), are members of the QIBA MRI committee and have access to QIBA phantoms. The repeatability and reproducibility studies were performed on both 1.5 and 3T MRI platforms by Drs. Shukla-Dave and Jambawalikar at MSKCC and CUMC. Figure 1 shows representative repeatability results from 3T GE MRI scanners at both sites MSKCC and CUMC. The variance in ADC values within and across sites was less than 2% (at both 3T and 1.5T) and ADC mean values for the central vial in the phantom was 1.11×10^{-3} mm^2/sec at MSKCC, and 1.16×10^{-3} mm^2/sec at CUMC for 3T MRI and 1.14×10^{-3} mm^2/sec at MSKCC, and 1.10×10^{-3} mm^2/sec at CUMC for 1.5T MRI.

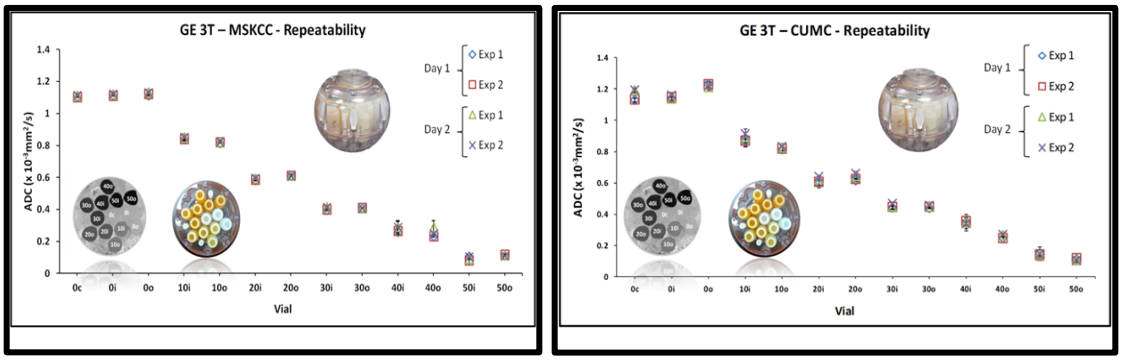


Figure 1: Repeatability results obtained using NIST/RSNA QIBA DW-MRI phantom containing vials with varying concentrations of PVP (0-50%) to generate physiologically relevant ADC values and different vial positions (c=central; o=outer; i=inner)**.** The phantom, and ADC image are shown as inserts in the graph. (A) Graph showing ADC (mean±sd) values for each vial in 4 experiments performed at MSKCC. (B) Graph showing ADC (mean±sd) values for each vial in 4 experiments performed at CUMC.

PLAN FOR THE YEAR 2018

In collaboration with Dr Chenevert (University of Michigan [UM], Co-Investigator), we are designing new phantom for Kurtosis. The first generation phantoms to accurately assess the Kurtosis phenomenon were developed in-house and they were not robust enough to be shipped for testing in multi-site trials (1). The group used dairy cream which is shown to be a simple, inexpensive, isotropic phantom useful for testing diffusional kurtosis imaging

data acquisition and postprocessing. The MR-visible protons of cream exhibit slow and fast diffusion components, attributed to the fat and water protons, respectively, which give rise to a diffusion coefficient of 1.1 $\mu m^2/ms$ and a diffusional kurtosis of 1.2. These parameter values were similar to those observed in vivo for human brain. Heating the cream was found to increase the T_2-relaxation time of the fat protons, which facilitates the evaluation of typical diffusional kurtosis imaging protocols used in clinical settings. However, this was not a robust phantom.

Thus, there is still an unmet need worldwide and we plan to design and construct robust novel Kurtosis phantom for multi-site trials.

For each of our aims in the proposal, we plan to continue data analysis and start patient volunteer accrual in the upcoming project period. We anticipate being able to prospectively enroll patient volunteer that meet our current study eligibility criteria. The novel analysis tools have been developed individually by the participating PIs group (Dr Schwartz, PI CUMC and Dr Shukla-Dave, PI MSKCC) and will be exchanged within the groups and to the QIN community.

STRENGTHS AT THE INSTITUTIONS [MSKCC, CUMC, UM]

Both groups (MSKCC and CUMC) have extensive experience in performing DW-MRI studies in cancer patients. MSKCC team led by Dr Shukla-Dave has been designated as Clinical Site I as they have published key papers in the field from testing and implementing the reduced field view of DW-MRI acquisition protocol to developing novel models for non-Gaussian modeling of the DW-MRI data for head and neck cancers. CUMC team lead by Dr Schwartz has been designated as clinical site II and will test and validate the imaging protocols developed at MSKCC and perform reproducibility and repeatability studies both in phantoms and patient volunteers. MSKCC and CUMC will also study patients going on experimental therapies. Both MSKCC and CUMC will perform MRI patient studies each as described in the proposal.

Dr. Shukla-Dave will work closely with Dr Chenevert in developing a new Kurtosis phantom that reflects properties the metrics derived from multi b-value DW-MRI acquisition and identify the best b-values selection and range for appropriate balance between SNR, scan time, and sensitivity to non-mono-exponential characteristics.

It is important to ensure that quantitative imaging biomarkers are established independent of individual site expertise, and using the same or similar data acquisition and analysis protocols for its clinical application worldwide. This proposal adheres to the vision of QIN as mentioned in PAR 14-116. As detailed in the application, the proposed study not only leverages the expertise of the laboratories of both PIs, but will also help the clinical community manage patients with (R/M) head and neck cancer by providing non-invasive, quantitative, imaging biomarkers to predict or assess early treatment response in patients being treated with experimental therapy.

MANUSCRIPTS

1. Fieremans E, Pires A, Jensen JH. A. simple isotropic phantom for diffusional kurtosis imaging.
   Magn Reson Med. 2012 Aug;68(2):537-42. doi: 10.1002/mrm.23263. Epub 2011 Dec 8.

PUBLICATIONS

1. Malyarenko D, Fedorov A, Bell L, Prah M, Hectors S, Arlinghaus L, Muzi M, Solaiyappan M, Jacobs M, Fung M, Shukla-Dave A, McManus K, Boss M, Taouli B, Yankeelov TE, Quarles CC, Schmainda K, Chenevert TL, Newitt DC. Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. J Med Imaging (Bellingham). 2018 Jan;5(1):011006. doi: 10.1117/1.JMI.5.1.011006. Epub 2017 Oct 30

2. Newitt DC, Malyarenko D, Chenevert TL, Quarles CC, Bell L, Fedorov A, Fennessy F, Jacobs MA, Solaiyappan M, Hectors S, Taouli B, Muzi M, Kinahan PE, Schmainda KM, Prah MA, Taber EN, Kroenke C, Huang W, Arlinghaus LR, Yankeelov TE, Cao Y, Aryal M, Yen YF, Kalpathy-Cramer J, Shukla-Dave A, Fung M, Liang J, Boss M, Hylton N. Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. J Med Imaging (Bellingham). 2018 Jan;5(1):011003. doi: 10.1117/1.JMI.5.1.011003. Epub 2017 Oct 10

U01CA207091: Quantitative Non-Contrast Perfusion using Arterial Spin Labeling for Assessment of Cancer Therapy Response

University of Texas Southwestern Medical Center

Ananth J. Madhuranthakam, Ph.D.
Joseph A. Maldjian, M.D.
Ivan Pedrosa, MD, Ph.D.

INTRODUCTION

The University of Texas Southwestern (UTSW) Medical Center has recently joined the quantitative imaging network (QIN), as of Sep. 2017. The primary focus of the UTSW QIN team is to validate arterial spin labeled (ASL) MR imaging as a quantitative imaging biomarker to measure non-contrast perfusion for assessment of cancer therapy response.

Tumors exhibit neo-angiogenesis, a key pathophysiological process for delivering oxygen and other essential nutrients for proliferation and metastasis (1). Cancer therapies are targeted at disrupting these vascular supplies either directly using antiangiogenic treatments or indirectly using cytotoxic chemoradiation. A significant number of these therapies have entered clinical trials in a variety of tumors. However, the radiological assessment of treatment outcomes still predominantly relies on morphological changes (e.g. RECIST), which is a limiting factor (2). An imaging technique to quantitate tumor vascular supply (or perfusion) and its response to therapy will have great importance in evaluating clinical trials by providing early, physiologically based indicators of response to therapy.

ASL-MRI has recently emerged as a quantitative imaging method to measure perfusion without the administration of exogenous contrast agents (3). ASL has the potential to not only assess therapy response (4, 5) but also to predict tumor aggressiveness based on the pre-treatment tumor vascularity (6, 7). ASL also has a number of advantages compared to dynamic contrast enhanced (DCE) and dynamic susceptibility contrast (DSC) based MRI perfusion measurements. Specifically, ASL does not require exogenous contrast administration and, unlike DCE/DSC, ASL measures absolute perfusion. In spite of these advantages, ASL has not undergone a robust and rigorous validation process to be established as a quantitative imaging method for evaluating response to therapy in oncological applications.

As part of the QIN grant, we will fulfill this goal by evaluating the ASL measurement in two known highly vascularized cancers, glioblastoma (GBM) and metastatic renal cell carcinoma (mRCC). Furthermore, we will also establish quality control protocols using a custom-built 3D printed perfusion phantom that can be used to validate the reliability of ASL measured flow. The specific aims of our proposal are:

Aim1: To demonstrate the reliability and precision of ASL based quantitative non-contrast perfusion in the brain and kidneys.

Aim2: To predict clinical outcomes based on baseline (pre-treatment) perfusion and early changes in post-treatment perfusion in patients with newly diagnosed GBM undergoing chemoradiation therapy.

Aim3: To predict long-term outcomes using baseline (pre-treatment) perfusion and early changes in post-treatment perfusion in patients with metastatic RCC undergoing antiangiogenic therapies.

DISCUSSION OF PROGRESS

In this section, we will list the methods and preliminary results acquired during the past year for each of the specific aims.

§ Aim 1: To demonstrate the reliability and precision of ASL based quantitative non-contrast perfusion in the brain and kidneys.

A1.1 Reliability of ASL Measured Flow in Phantoms:

The first goal of this aim is to develop a quality assurance (QA) protocol using our 3D printed perfusion phantom (Figure 1), that can be used across different scanners and platforms for consistent reproducibility of the quantitative ASL measured flow (8). Towards this goal, we are optimizing the imaging protocol with varying labeling durations and post-label delay with both pseudo-continuous ASL (pCASL), the primary sequence, compared against the pulsed ASL using flow alternating inversion recovery (FAIR) (Figure 2).

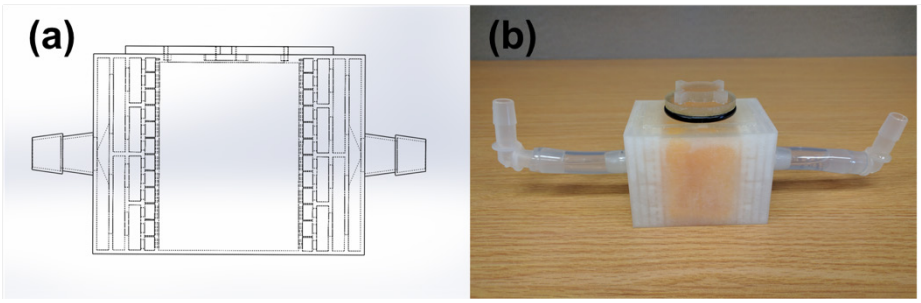


Figure 1: 3D printed perfusion phantom. Schematic (a) and the final printed perfusion phantom (b).

A1.2 Reliability of ASL Measured Perfusion in Brain:

The second goal of this aim is to measure reproducibility of ASL measured perfusion in brains of 30 normal volunteers. Initially, we proposed to use spiral based acquisitions for

ASL measurement. While the spiral acquisitions have provided high signal to noise ratio (SNR) and robust images in normal volunteers, they are prone to distortions with increased B0 inhomogeneities.

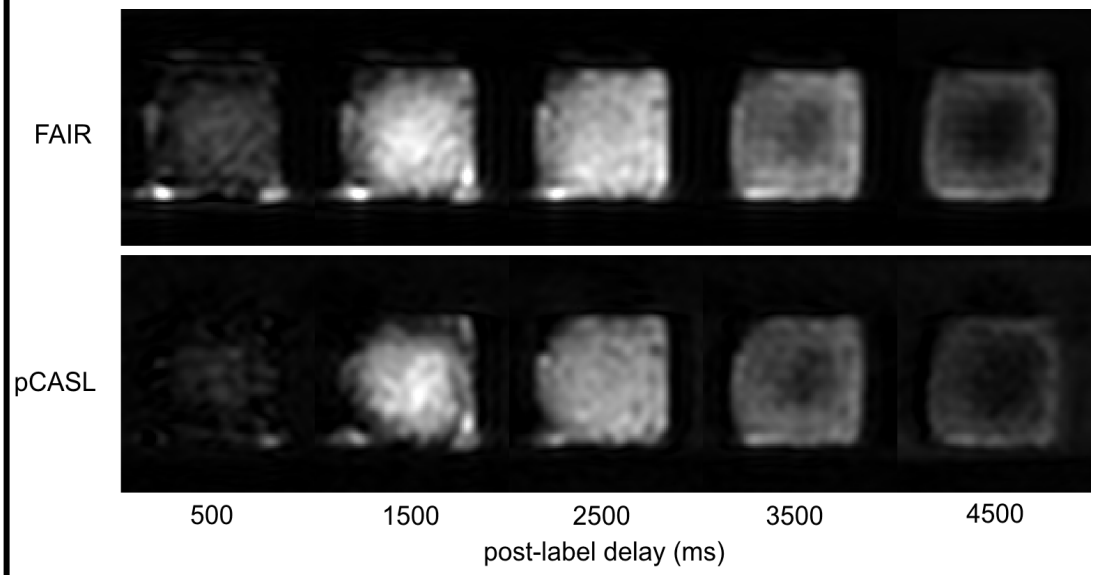


Figure 2: ASL measured flow in perfusion phantom. Top row: Images acquired with a pulsed ASL approach, FAIR at different post-label delays. Bottom row: Same images acquired with the pCASL approach using a 500 ms labeling duration and at different post-label delays. Both techniques show similar signal evolutions at different post-label delays.

In anticipation of evaluating this sequence on GBM patients, who often have craniotomy that induces increased B0 inhomogeneities, we also developed Cartesian based acquisitions for improved robustness in the presence of metal implants. Specifically, we designed a Cartesian acquisition with spiral reordering (CASPR), which has the robustness of acquiring images using Cartesian trajectory, while also increasing the robustness of ASL measured signal by using a spiral trajectory (Figure 3) (9). Using CASPR, we achieved robust image quality throughout the brain compared to the product EPI based ASL measurement (Figure 4). We have initiated the recruitment of normal volunteers to test the reproducibility of this approach in brain in both intra-session and inter-session.

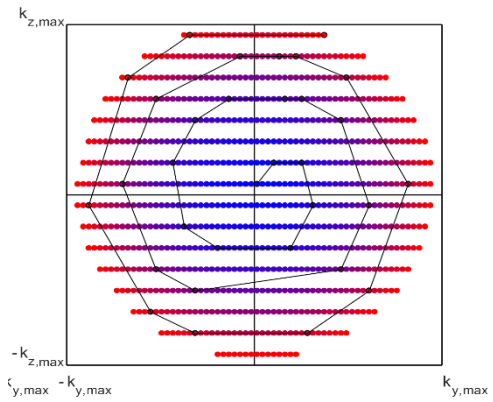


Figure 3: Cartesian Acquisition with SPiral Reordering (CASPR) samples the center of k-space during the early echoes of the turbo spin echo readout (blue dots), and the periphery of k-space towards the later part of the turbo spin echo readout (red dots) in a spiral order (black line), while still maintaining samples on the Cartesian view ordering.

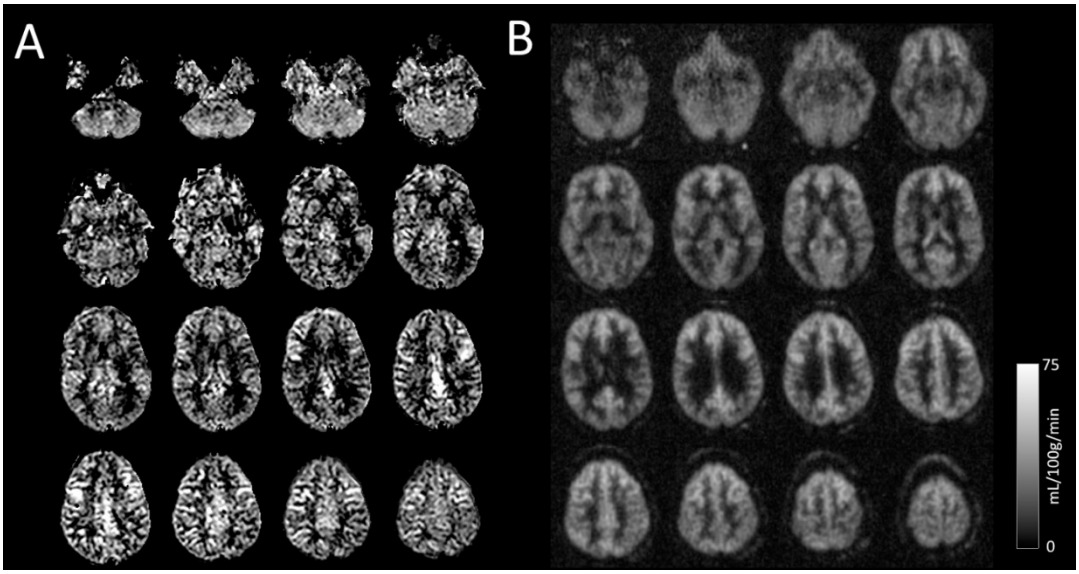


Figure 4: Brain perfusion images of a normal volunteer. Compared to standard pCASL with 2D EPI acquisition (A), the newly developed CASPR acquisition (B) generates robust perfusion images throughout out the brain with uniform signal intensity, higher SNR and robust to B0 inhomogeneities. This will generate robust images without significant artifacts due to B0 inhomogeneities in GBM patients, with craniotomy.

A1.3 Reliability of ASL Measured Perfusion in Kidneys:

The third goal of this aim is to measure reproducibility of ASL measured perfusion in kidneys of 30 normal volunteers, similar to the above sub-aim. Compared to brain perfusion imaging, which has become a standard imaging technique on most commercially available scanners, the ASL acquisitions in the body are significantly lacking. Using our newly designed CASPR approach (Figure 3), we are able to achieve robust kidney perfusion images throughout the volume (Figure 5) (9). We have initiated the recruitment of normal volunteers to test the reproducibility of this approach in kidneys in both intra-session and inter-session.

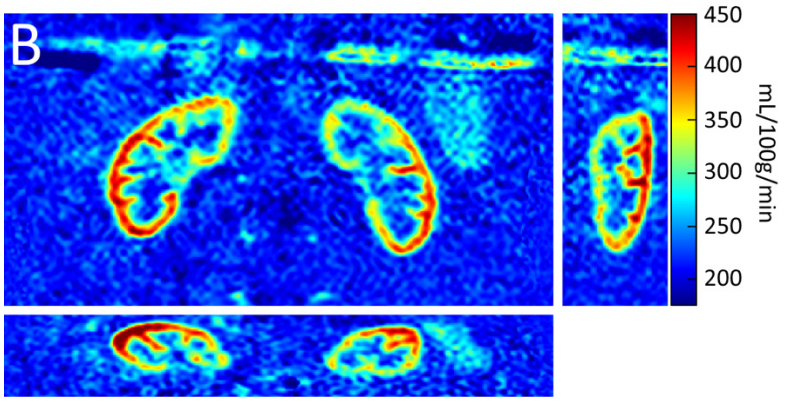


Figure 5: Kidney perfusion images of a normal volunteer. CASPR view ordering (fig. 3) generates robust images in the body that shows uniform perfusion throughout the kidneys, without being sensitive to B0 inhomogeneities. The images were acquired in the coronal orientation and reformatted into sagittal and axial orientations.

§ Aim 2: To predict clinical outcomes based on baseline (pre-treatment) perfusion and early changes in post-treatment perfusion in patients with newly diagnosed GBM undergoing chemoradiation therapy.

We have submitted our IRB protocol for the recruitment of GBM patients undergoing chemoradiation. We anticipate the IRB approval in the early part of the upcoming year. Following that, we will open the recruitment of GBM patients to evaluate ASL measured perfusion. Meanwhile, we have been working on the following:

- Established standard protocol that includes the following quantitative imaging techniques in addition to ASL – DCE-MRI, DSC-MRI, and DWI.
- Evaluating several tumor segmentation algorithms for automated segmentation of the tumors.
- Established contact with ECOG-ACRIN EAF151 trial for possible evaluation of ASL in addition to the standard clinical imaging for evaluating treatment response.

§Aim 3: To predict long-term outcomes using baseline (pre-treatment) perfusion and early changes in post-treatment perfusion in patients with metastatic RCC undergoing antiangiogenic therapies.

We have also submitted another IRB protocol for the recruitment of mRCC patients undergoing anti-angiogenic therapy. We anticipate the IRB approval in the early part of the upcoming year. Following that, we will open the recruitment of mRCC patients to evaluate ASL measured perfusion. Meanwhile, we have been working on the following:

- Established standard protocol that also includes the quantitative DWI along with the ASL for evaluating treatment response.
- Evaluating ASL in small renal mass protocol, who are under active surveillance, under the kidney cancer SPORE grant. Although, this is not measuring cancer therapy response, it provides valuable information in identifying the stable disease against progressing disease, that can be measured with ASL, without the administration of exogenous contrast agent.

COLLABORATIONS WITHIN THE NETWORK

We have begun collaborations with other QIN sites. Specifically, we are participating in two challenges – DSC/DRO organized by Barrow's Neurological Institute and Prostate DWI challenge, organized by Medical College of Wisconsin.

PLANS FOR NEXT YEAR

Aim 1

a. Establish QA protocol with the perfusion phantom. Once established, the perfusion phantom will be scanned every other week for the entire year to evaluate reproducibility of ASL measured flow.
b. Continue the recruitment of at least 15 normal volunteers for brain imaging.
c. Continue the recruitment of at least 15 normal volunteers for kidney imaging.

Aim 2

a. Start recruitment of GBM patients with a possible target of at least 10 patients for this year.
b. Refine and finalize the automated tumor segmentation tools for brain tumors.
c. Establish automated processing pipelines for image analysis including registration of multiple contrasts and feature extraction.

Aim 3

a. Start recruitment of mRCC patients with a possible target of at least 10 patients for this year.
b. Optimize the automated brain tumor segmentation tools from Aim 2 for tumor segmentation in the body.

We will also continue to participate in QIN collaborative projects.

PUBLICATIONS AND PRESENTATIONS

We are currently drafting two manuscripts, one on the perfusion phantom and the other on robust perfusion imaging using our newly designed CASPR approach.

REFERENCES

1. Folkman J. Role of angiogenesis in tumor growth and metastasis. Seminars in oncology. 2002;29(6 Suppl 16):15-8.
2. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). European journal of cancer. 2009;45(2):228-47.
3. Alsop DC, Detre JA, Golay X, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine. 2014.
4. de Bazelaire C, Alsop DC, George D, et al. Magnetic resonance imaging-measured blood flow change after antiangiogenic therapy with PTK787/ZK 222584 correlates with clinical outcome in metastatic renal cell carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;14(17):5548-54.
5. Fellah S, Girard N, Chinot O, Cozzone PJ, Callot V. Early evaluation of tumoral response to antiangiogenic therapy by arterial spin labeling perfusion magnetic resonance imaging and susceptibility weighted imaging in a patient with recurrent glioblastoma receiving bevacizumab. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(11):e308-11.
6. Lanzman RS, Robson PM, Sun MR, et al. Arterial spin-labeling MR imaging of renal masses: correlation with histopathologic findings. Radiology. 2012;265(3):799-808.
7. Fujima N, Kudo K, Yoshida D, et al. Arterial spin labeling to determine tumor viability in head and neck cancer before and after treatment. Journal of magnetic resonance imaging : JMRI. 2014;40(4):920-8.
8. Greer JS, Wang X, Hulsey K, Lenkinski RE, Madhuranthakam AJ. A 3D printed perfusion phantom for quality controlled measurement of arterial spin labeled perfusion Proceedings of the 25th Annual Meeting of ISMRM. Honolulu, HI2017; p. 3805.
9. Greer JS, Wang X, Pinho MC, Pedrosa I, Madhuranthakam AJ. Robust 3D pCASL perfusion imaging using a Cartesian acquisition with SPiral Reordering (CASPR). Proceedings of the 25th Annual Meeting of ISMRM. Honolulu, HI2017; p. 3628.

Section II

Working Group Reports

Clinical Trials Design and Development Working Group

Hui-Kuo Shu, MD, PhD, Chair
Elizabeth Gerstner, MD, Co-chair

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 (2017-2018)

§ Manuscripts

Standard in Reporting Quantitative Imaging (STIRQI)

QI methods are increasingly used in clinical trials both as primary and secondary, or correlative endpoints. With increasing sophistication of QI methodologies, a minimum standard of basic information regarding the acquisition and analysis of QI data must be provided in publications so that reader can determine the validity and reliability of these results and findings can be generalized. This initiative seeks to define a set of criteria that should be presented in QI-related peer-reviewed papers to ensure that quantitative data extracted from images are reported in a meaningful, consistent, and repeatable manner.

Richard Wahl has led this effort basing STIRQI on the STAndards for Reporting of Diagnostic Accuracy (STARD) criteria first reported in 2003 (Bossuyt, et al., Ann Int Med 138:W1-12, 2003) and subsequently updated in 2015 (Bossuyt, et al., BMJ 351:h5527, 2015). He has taken input from members of the CTDD WG and drafted a manuscript with a checklist of

standards for QI reporting. This manuscript has now been circulated among working group members for comments/edits and is in the final stages of preparation. It is expected that this manuscript will be submitted for review in the first quarter of 2018.

Quantitative Imaging in Radiation Oncology

With advances in radiation oncology, this field is now increasingly reliant on both high-resolution anatomic-based imaging (CT and MRI) as well as functional imaging (PET, DWI, MRSI etc.). Greater utilization of conventional and advanced imaging methodologies has developed because the radiation therapy delivery has become increasingly accurate 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 should 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 are leading the effort in writing a manuscript reviewing the use of QI modalities by radiation oncologists in clinical trials and routine patient management. This review will highlight, in particular, the role of the QIN in developing and advancing QI techniques and their utilization in the field of radiation oncology. A manuscript with wide input from members of the CTDD WG was submitted for review for a special imaging issue of the International Journal of Radiation Oncology, Biology, Physics (AKA The Red Journal) on 12/31/2017. This manuscript was recently returned for modification and will be re-submitted by 2/13/2018.

Pathways for Adoption of QI Tools/Techniques into Clinical Trials and Widespread Clinical Use

With the impending conclusion of the above two projects with manuscript submissions, additional ideas for new projects have been entertained. Discussions are ongoing within the working group regarding a new white paper project that matches well with the PathCT initiative. The idea is to produce a review paper that goes through the hurdles that would need to be overcome for bringing new QI methodologies into clinical use. Areas to be explored include what level of evidence is needed to warrant bringing a new QI technique forward and developing a roadmap for bringing this into clinical trial and subsequent adoption for general clinical use.

§ Outreach activities

Presentation at the July 2017 NRG Oncology meeting

Hui-Kuo Shu gave a presentation for the Imaging Working Group of NRG Oncology entitled "Implementing Quantitative Imaging Network (QIN) Tools in NCTN trials: QIN Clinical Trials Design & Development working group" at the semiannual summer meeting in Philadelphia, PA on July 14, 2017. This presentation introduced the structure and goals of the

QIN to this audience as well as promoting the idea of including QI endpoints in clinical trials and how to facilitate greater interaction with QIN groups to help accomplish this. In particular, details were given about AutoPERCIST™, an semi-automated tool for making SUL measurements on FDG-PET studies, which was touted as an example of an advanced tool developed within the QIN that may be of interest to clinical researchers within NRG Oncology.

QIN session at the 3rd AACR-SNMMI Joint Conference on State-of-the-Art Molecular Imaging in Cancer Biology and Therapy, February 14-17 in San Diego, CA

Ella Jones was instrumental in organizing a QIN session for this meeting scheduled on Friday, February 16. This session will focus on clinical translation of QIN tools. Although the meeting is generally on more pre-clinical subject matters, this session will be on a day where there are more clinical considerations. Program will include Ella Jones who will introduce the QIN as well as additional speakers including Jayashree Kalpathy-Cramer, Tom Yankeelov and Brenda Kurland talking on a variety of QIN-related topics.

Plan for a Cancer Imaging session for the 2018 Annual Meeting of the American Association for Physicist in Medicine (AAPM)

John Buatti will be chairing a session entitled "The Essential Role of Longitudinal Imaging in Cancer Decision Making" at the Annual AAPM meeting in Nashville, TN on July 30-August 3, 2018. This session will focus on advanced imaging techniques that can help provide clinical decision support for oncologic management and will include multiple QIN members. Speakers (topics) for this session will include Rujiang Li (The prediction of immune modulating effects of cancer therapy using radiomics features in lung cancer), John Bayouth (Longitudinal assessment of lung function using 4D imaging), Andrey Federov (Multiparametric MR parameters in Prostate Cancer predict response and guide therapy) and Hui-Kuo Shu (Whole Brain MRSI predicts early response to therapy and enables adaptation).

Proposal for a Panel Session at the 2018 Annual Meeting of the American Society for Therapeutic Radiology and Oncology (ASTRO)

A panel session was proposed by John Buatti and Hui-Kuo Shu entitled "NCI's Quantitative Imaging Network: Development and Integration of Novel Tools for Oncology Clinical Trials and Patient Management" the next ASTRO Annual Meeting in San Antonio, TX on October 21-24, 2018. We are currently awaiting review by the organizing committee to determine whether this panel session will be chosen for this upcoming ASTRO meeting. The proposed summary, learning objectives and presentations of this panel session are as follows:

Summary: The National Cancer Institute (NCI) has funded the Quantitative Imaging Network (QIN) since 2009 to develop novel approaches, including tool development, for quantitative imaging and clinical decision support in oncology. Multi-disciplinary research teams from top institutions across the country that make up this network are developing a number of quantitative cancer imaging methods and tools that can be applied in unique ways to add value to the management of oncology patients. This session will feature QIN members that are developing and validating quantitative imaging tools for common cancers that are of interest to the membership of ASTRO. The panel will describe how advanced imaging modalities can complement more conventional ones for oncologic evaluations. There will be a particular focus on tools that are in an advanced stage of development with presentations that address unique clinical translational issues. Another focus will involve an overview of the computing architecture/pipelines and informatics tools needed to fully exploit the wealth of data that will be generated with increased utilization of advanced quantitative imaging techniques. This panel will raise awareness of the potential utility of advanced quantitative imaging for both research and general clinical practice. Several barriers (shown below) have limited wider adoption of advanced quantitative imaging techniques in the clinic. We believe that panels such as the one proposed here will help educate oncologists about the value of quantitative imaging and further promote the development and validation of new tools and methods needed to more fully bring quantitative imaging into the clinical setting.

Learning Objectives: The following Learning objectives were achieved.

1. Differentiate conventional qualitative from advanced quantitative imaging techniques and describe the added value of the quantitative methodologies to clinical management.

2. Demonstrate knowledge of specific tools being developed by featured QIN research groups that will allow various advanced quantitative imaging techniques to be utilized on oncology clinical trials.

3. Demonstrate an understanding of computing architecture/pipeline and informatic needs to fully exploit development of quantitative imaging methods/tools for clinical use.

Presentations

1. Advanced imaging tools for head and neck cancer segmentation and evaluation (John Buatti).

2. Computing architecture/pipeline requirements for quantitative imaging assessment and application to radiogenomic mapping (Sandy Napel).

3. Informatic issues with implementation of quantitative imaging methodologies for improved digital interoperability (Andrey Federov).

4. Development of novel spectroscopic MRI technique for evaluation and follow-up of brain tumors (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 examined the variability for interpretation of FDG-PET results in 22 readers across 15 institutions on 30 test cases with scans before and after therapy to determine multicenter variability using each institution's own preferred software for this analysis. This published study (J Nucl Med, 58:1429-1434, 2017) showed high correlation across readers and institutions but did still identify some level of variability.

Dr. Wahl's team has 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. He has initiated a study to test across institutions and readers using AutoPERCIST™ to evaluate the same 30 test cases referenced above. The main goal is to determine whether variance among readers and institutions is further reduced when the AutoPERCIST™ tool is used. The latest version of this software was installed by 11 participating institutions (from United States, Asia and Europe) through a materials transfer agreement (MTA) and the 30 paired sets of anonymized FDG PET-CT images were downloaded for evaluation. Instructions for this study were given to each participating institutions with up to 5 tumor lesions from each PET image to be evaluated. All selections were recorded and sent to the central database at Johns Hopkins Image Response Assessment Team for quality control. Initial results show that very consistent SUV peak values were obtained and this had less variability than was seen when different institutional software was used. Joo Hyun O, Clinical Assistant Professor (Department of Nuclear Medicine, Seoul St. Mary's Hospital, Seoul, Korea) and Richard Wahl are leading this effort with an abstract that has been submitted for the 2018 SNMMI Annual meeting.

Pathways to Clinical Trials (PathCT) initiative

This initiative was started at the QIN Annual F2F meeting in April 2016 with significant progress including the drafting of a PathCT Summary report, updating of a list of QIN tools that documents their level of readiness for clinical trial translation, a QIN-NCTN planning meeting in Philadelphia in December 2016 and establishment of the PathCT Focus Group that is working to advance this initiative. These accomplishments were documented in the previous year's Annual Report.

The CTDD WG in 2017-2018 has continued to work to help facilitate the translation of QIN-developed tools to NCTN clinical trials. The working group have the following accomplishments towards advancing these goals:

QIN presentation for the Imaging Working Group at the July 2017 NRG Oncology meeting: See above for more details.

QIN-NCTN planning meeting at the Alliance Semiannual Meeting on November 2-4, 2017 in Chicago, IL: Through a several months planning process led by Larry Schwartz, Paul Kinahan and David Mankoff, QIN investigators were given the opportunity to participate in a significant fashion at the Alliance meeting in November 2017. The planning committee for this effort included Hui-Kuo Shu as a representative of the CTDD WG. Robert Nordstrom and Michael Knopp addressed the entire Alliance group at their Plenary session introducing the QIN and making a case for the value of incorporating quantitative imaging assessments/endpoints into clinical trials. Many QIN representatives were also given the opportunity to speak and participate in different disease site/discipline committees including Nola Hylton and Mary Ellen Giger (at the Breast committee), Paul Kinahan and Amita Dave (at the Experimental Therapeutics committee), Larry Schwartz and Hugo Aerts (at the GI committee), Mike Jacobs, Andrey Federov and Brian Ross (at the GU committee), Rich Wahl and Dave Mankoff (at the Lymphoma committee), Brad Erickson, Jatsharee Kalpathy-Cramer and Michael Knopp (at the Neuro-Oncology committee), Hui-Kuo Shu and Yue Cao (at the Radiation Oncology committee), John Buatti and Michael McNitt-Grtay (at the Respiratory committee) and Ying Xiao, Michael Knopp, Mark Rosen and T.J. Fitzgerald (at the Imaging committee).

Coordinating QIN-NCTN planning meeting for the NRG Oncology Semiannual Meeting on July 12-14, 2018 in Philadelphia, PA: The planning process has been initiated (1st teleconference in January 2018) for a QIN effort similar to the past Alliance meeting at the NRG Oncology meeting in July 2018. There has been early communications with Mitch Machtay, Deputy Group Chair for Research Strategy within NRG, about organizing this effort. Hui-Kuo Shu and John Buatti will also be taking on an increased role in the planning process for this effort. Monthly teleconferences will continue until the time of the meeting to maintain momentum for planning. In addition to the type of participation that QIN had at the Alliance meeting, one idea would be to bring specific proposals to incorporate certain advanced QIN tools/methodologies into developing protocols at specific targeted disease committees.

PathCT Focus Group: This group, which consists of members of the CTDD WG (current composition include John Buatti, Elizabeth Gerstner, Lori Henderson, Ella Jones, Hui-Kuo Shu and Richard Wahl), continues to meet on a regular basis to review the readiness of QIN-developed tools. To gain a better understanding of the most up-to-date status of QIN tools, this group developed a short questionnaire, which is now ready for distribution to QIN PIs to help assess the stage of development for individual tools. Once this is determined, this group will likely target groups with tools that are just about at the stage for clinical trials translation to see this Focus Group can help provide guidance for the next steps in advancing these specific tools.

PLANS FOR THE COMING YEAR (2018-2019)

§ Goals for the coming year include the following:

1. Completing the submission of two manuscripts (STIRQI criteria and Quantitative Imaging in Radiation Oncology) for review in appropriate peer-reviewed journals.

2. Starting a new project on discussing the requirements and hurdles for translation of QI tools into clinical trials and general clinical practice and producing a new white paper on this topic.

3. 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. A planned session on Cancer Imaging will be given at the AAPM Annual Meeting July 30-August 3, 2018 in Nashville, TN.

    b. A panel session proposal has been submitted to present at the ASTRO Annual Meeting in 2018. If this proposal is approved, the session will take place at ASTRO October 21-24, 2018 in San Antonio, TX.

4. Presentation of the results of the AutoPERCIST™variance test at the SNMMI Annual Meeting June 23-26, 2018 in Philadelphia, PA and preparation/submission of a manuscript for this effort.

5. Continue to advance the goals of the PathCT initiative through the following:

    a. Planning significant QIN participation at the NRG Oncology Semiannual Meeting July 12-14, 2018 in Philadelphia, PA.

    b. Continuing the activity of the PathCT Focus Group as detailed above.

Bioinformatics and Data Sharing Working Group

Jayashree Kalpathy-Cramer, Ph.D.
Peter S. LaViolette, Ph.D.

INTRODUCTION

The Bioinformatics and Data Sharing (BIDS) working group serves two main aims to the QIN. Aim 1 is promoting and facilitating data sharing. Aim 2 is the promotion and facilitation of tool sharing between sites and researchers. The past focus has been on data sharing, specifically through the TCIA. This year the broad focus of the BIDS group has been on tool sharing.

DISCUSSION OF PROGRESS

§ Group Communication Update

Early in the year it was decided to track both the meeting minutes and the agendas on a shared Google doc. We have also formed a Google Group called "QIN BIDS Working Group" where group communications are tracked, and links to shared documents are located. To join the group, please email either Justin Kirby justin.kirby@nih.gov or Peter LaViolette: plaviole@mcw.edu. To email the entire group, use the email address: qin-bids@googlegroups.com In addition to using resources such as Google for communication, the The National Cancer Informatics Program (NCIP) Hub has also been proposed as a mechanism for carrying out collaborative research projects. As stated on the website: www.nciphub.org, the NCIPHub is "a site for community research and collaboration in cancer research and informatics. Users can share resources, host online communities, and use collaboration tools. These resources are available to visitors from all over the world. The NCIP hub is supported by the National Cancer Institute, National Institutes of Health to foster collective innovation and democratize access to data, tools and standards across the cancer research community."

§ Tool Sharing

Much of the effort of the BIDS group this year have been focused on the creation and dissemination of a list of tools created by QIN investigators. After extensive discussion it was decided to use a google spreadsheet for the list, which is located here:
https://docs.google.com/spreadsheets/d/1YMj_KvrAYSbaK9AVzK0n_zF3GGMJ1ZfWZy5dhR4hwfg/edit#gid=0

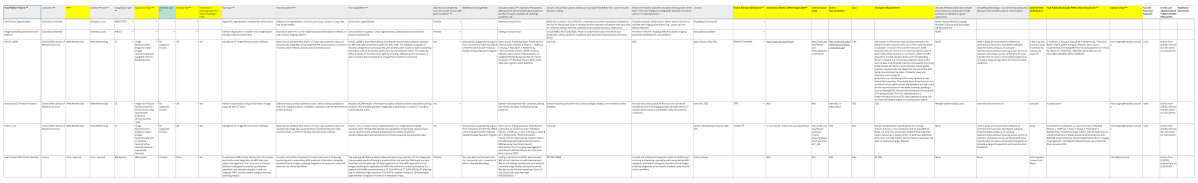


Figure 1. Snapshot of the first few tools listed in the QIN Tool List.

The tools listed range from 3D-Slicer, a multi-modal image processing platform, to histology segmenting tools geared specifically towards one specific immunohistochemistry stain. The current list of tools began the year as a spreadsheet that was passed user to user for updates. We decided early that the best way to ask everyone all at once was to use an online spreadsheet available to all for editing and maintaining. The initial Google Sheet created contained the same columns as the excel sheet that was passed user to user. To improve the information contained in the list, many new columns have been added which contain unique information including:

Tool/Platform Name **
Institution **
PI**
Contact Person **
Image/Data Type **
Application type **
Intended user **
Disease Site**
Interested in participating in the ToolX Challenge? (Y/N)
Tool Type **
Tool Description **
Tool Capabilities **
Was this tool created by your site as part of your QIN grant activities? **
Additional Funding Notes
Evaluation Status ** Examples: Prospective, Retrospective, primary/secondary endpoints, Number of cases, examples of challenge validation, etc.
List any clinical trials used to assess your tool (add Title & NCT #) or used in trial for decision making.
What is the study's objective and what is the purpose of the tool in this trial? Integral or Integrated? Used for correlative work or to make decisions?
What is the study's objective and what is the purpose of the tool in this trial? Integral or Integrated? Used for correlative work or to make decisions?
License
How is the tool distributed **
Distribution Notes / LINK if Applicable**
Link to Source Code
Link to documentation
Cost
Hardware Requirements
Industry Affiliates (bold text indicates actual industry collaborator on QIN activities as indicated on grant application)
Competing Technology - List commercial product(s) that your tool competes with or could replace
External Sites Utilizing Tool
Tool Publication(s) (List PMID) Users Should Cite **
Contact Email **
Funded Technical Support?
Entries Last Updated (Date) PLEASE UPDATE REGULARLY
Additional Comments

List items marked by ** are required entries. Many of the items in the list are organized as drop-down menus meant to standardize the entries. The list currently has 62 different tools listed and has been updated by each group. A finalized list will be available for users at the F2F group meeting. Ongoing efforts in the future will be to maintain this list through BIDS group leadership. Reminders to update entries will be sent to the tool developers listed twice annually. Eventually this list will appear in an abbreviated form on the new QIN web page.

§ ToolX Challenge

In order to facilitate and encourage cross-group collaboration using shared tools, the ToolX challenge has been proposed. This project is currently in the planning stages and will likely be rolled out summer/fall of 2018. The QIN Tool list described above has a column where tool developers can indicate whether or not they are interested in participating. Currently ~25 tools are listed as interested. The broader goals of this project are to improve the tool usage potential by matching external users with tool developers. Feedback provided by new users will then be incorporated into the documentation and tools themselves.

QIN RELATED TCIA UPDATES

The Cancer Imaging Archive (TCIA) is a web based repository for storing and sharing datasets of imaging and associated meta-data. There have been several updates over the past year to datasets located on the TCIA. Links can be found at their website: www.cancerimagingarchive.net. Table 1 shows a list of the recently updated datasets.

Table 1. List of QIN related datasets on the TCIA (Source TCIA Website).

Collection	Cancer Type	Modalities	Subjects	Location	Metadata	Access	Status	Updated
QIN LUNG CT	Non-small Cell Lung Cancer	CT	47	Lung	Yes	Public	Complete	7/31/17
QIN-Breast	Breast Cancer	MR, PT, CT	67	Breast	Yes	Limited	Ongoing	9/4/15
QIN-BRAIN-DSC-MRI	Low & High Grade Glioma	MR	49	Brain	No	Limited	Complete	8/28/15
QIN GBM Treatment Response	Glioblastoma Multiforme	MR	54	Brain	No	Limited	Complete	8/12/15
QIN-SARCOMA	Sarcomas	MR	15	Breast, Calf, Chest, Elbow, Knee, Leg, Shoulder, Thigh	No	Limited	Ongoing	9/5/14
QIN PET Phantom	PET Phantom	PT	2	Phantom	No	Public	Complete	9/4/14
QIN-HeadNeck	Head and Neck Carcinomas	PT, CT, SR, SEG, RWV	156	Head-Neck	Yes	Public	Complete	8/26/14
QIN Breast DCE-MRI	Breast Cancer	MR, KO	10	Breast	Yes	Public	Ongoing	7/31/14
QIN Prostate	Prostate Cancer	MR	22	Prostate	No	Limited	Complete	7/2/14

§ TCIA Publications

QIN Specific publications can be seen here:
https://wiki.cancerimagingarchive.net/display/Public/Publications#Publications-QIN

§ Pipeline Collaborative Research Project

A pipeline is a series of well-defined steps meant to take a dataset from its rawest form to a dataset processed to a point that statistical testing can be applied to determine group level inferences. Some example pipeline steps include raw image conversion, 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.

Currently, there are no standard guidelines for standardizing modules that make up pipelines, as well as the inputs and outputs of each "step". This unfortunately results in duplication efforts as each group creates their own pipeline methodology and their own algorithms that go into the pipeline. The BIDS Pipeline collaborative research 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. While this project is currently still in the planning phase, ongoing efforts by BIDS members developing both the modules and pipelines are ongoing. It is planned that the first projects will involve datasets from the TCIA such as the LIDC dataset.

PLANS FOR NEXT YEAR

The BIDS group will continue to work on the dissemination of datasets and tools. This will include focussing on efforts aimed at making it easier for QIN community to leverage the tools that others develop, and making it easier to share datasets. Focus will be on both the ToolX project and the Pipeline Collaborative Research Project. We plan to open dialogue 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 projects.

LIST OF BIDS QIN PUBLICATIONS 2017-2018

1. Balagurunathan Y, Beers A, Kalpathy-Cramer J, et al. Semi-automated pulmonary nodule interval segmentation using the NLST data. *Med Phys.* 2018.

2. Farahani K, Kalpathy-Cramer J, Chenevert TL, et al. Computational Challenges and Collaborative Projects in the NCI Quantitative Imaging Network. *Tomography.* 2016;2(4):242-249.

3. Malyarenko D, Fedorov A, Bell L, et al. Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. *J Med Imaging (Bellingham).* 2018;5(1):011006.

4. Newitt DC, Malyarenko D, Chenevert TL, et al. Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. *J Med Imaging (Bellingham).* 2018;5(1):011003.

5. Saltz J, Almeida J, Gao Y, et al. Towards Generation, Management, and Exploration of Combined Radiomics and Pathomics Datasets for Cancer Research. *AMIA Jt Summits Transl Sci Proc.* 2017;2017:85-94.

6. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21):e104-e107.

7. Herz C, Fillion-Robin JC, Onken M, et al. dcmqi: An Open Source Library for Standardized Communication of Quantitative Image Analysis Results Using DICOM. *Cancer Res.* 2017;77(21):e87-e90.

8. Hassanzadeh E, Alessandrino F, Olubiyi OI, et al. Comparison of quantitative apparent diffusion coefficient parameters with prostate imaging reporting and data system V2 assessment for detection of clinically significant peripheral zone prostate cancer. *Abdom Radiol (NY).* 2017.

9. Langkilde F, Kobus T, Fedorov A, et al. Evaluation of fitting models for prostate tissue characterization using extended-range b-factor diffusion-weighted imaging. *Magn Reson Med.* 2018;79(4):2346-2358.

10. Mehrtash A, Pesteie M, Hetherington J, et al. DeepInfer: Open-Source Deep Learning Deployment Toolkit for Image-Guided Therapy. *Proc SPIE Int Soc Opt Eng.* 2017;10135.

11. Fedorov A, Vangel MG, Tempany CM, Fennessy FM. Multiparametric Magnetic Resonance Imaging of the Prostate: Repeatability of Volume and Apparent Diffusion Coefficient Quantification. *Invest Radiol.* 2017;52(9):538-546.

12. Velez E, Fedorov A, Tuncali K, et al. Pathologic correlation of transperineal in-bore 3-Tesla magnetic resonance imaging-guided prostate biopsy samples with radical prostatectomy specimen. *Abdom Radiol (NY).* 2017;42(8):2154-2159.

13. 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).* 2017;42(3):918-925.

14. Hurrell SL, McGarry SD, Kaczmarowski A, et al. Optimized b-value selection for the discrimination of prostate cancer grades, including the cribriform pattern, using diffusion weighted imaging. *J Med Imaging (Bellingham).* 2018;5(1):011004.

15. Echegaray S, Bakr S, Rubin DL, Napel S. Quantitative Image Feature Engine (QIFE): an Open-Source, Modular Engine for 3D Quantitative Feature Extraction from Volumetric Medical Images. *J Digit Imaging.* 2017.

16. Zhou M, Scott J, Chaudhury B, et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *AJNR Am J Neuroradiol.* 2018;39(2):208-216.

17. Echegaray S, Nair V, Kadoch M, et al. 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.* 2016;2(4):283-294.

QIN MRI Subgroup Annual Report

2017-2018 Chairmen:
Yue Cao, PhD
Laura Bell, PhD

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 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 MRI image analysis accuracy, reproducibility, and robustness. The working group coordinates the collaboration between members in this area." Currently, 12 cancer imaging centers participate in the MRI subgroup activities: Oregon Health and Science University (OHSU), Brigham and Women's Hospital (BWH), Medical College of Wisconsin (MCW), Icahn School of Medicine at Mount Sinai (MS), University of Michigan center #1 (UM1), University of Michigan center #3 (UM3), University of Pittsburgh (UPitt), Vanderbilt University (VU), University of Washington (UW), University of California San Francisco (UCSF), Stanford University (SU), Massachusetts General Hospital (MGH).

This past year we had three challenges wrap up completely and published in peer-review journals. Fortunately, we have several challenges that have taken over in the meantime. One challenge is near completion and a manuscript is planned to be submitted to QIN special issue in Tomography 2018. Two additional challenges remain quite active and work-in-progress. One challenge idea is currently being discussed as a potential proposal. Specific status details are provided in the following sections.

§ Completed Challenges – submitted for publication/published

ADC Mapping Project (Lead: David Newitt, UCSF)

The goal of this challenge was to investigate the reproducibility of DW-MRI ADC maps generated from 11 QIN sites, 1 non-QIN site, and 3 scanner vendor online maps. Both phantom and in vivo breast data were analyzed by all participating sites for 2 and 4 b-value diffusion metrics using their own diffusion processing software.

This challenge showed that ADC metrics both in vivo and in phantoms showed significant differences between analysis implementation. Differences were the largest when comparing vendor online ADC maps to site-specific post-processed maps. For ADC maps calculated with two b-values, the differences were found to be clinically insignificant.

However, analysis of ADC maps from multi b-values showed differences large enough to be of concern, especially for multi-site, multi-vendor clinical trials.

The results from this challenge were presented at International Society of Magnetic Resonance in Medicine 25th Annual Meeting (Hawaii, 2017) and published in JMI [1].

DICOM Storage Project (Leads: Dariya Malyarenko, UMICH1 & Andrey Fedorv, BWH QIICR)

The goal of this challenge was to demonstrate the ability to generate and store relevant DICOM metadata for ADC parametric maps (PM) across vendors, sites, and software tools. Ten QIN participating sites analyzed multi-vendor DWI DICOM datasets for polyvinylpyrrolidone (PVP) diffusion phantom to generate ADC maps in DICOM format using their choice of software for two (ADC2) and four (ADC4) b-value fit.

The results from this challenge showed that standardization of DICOM formatting is needed for ADC parametric maps. Minor numerical discrepancies among sites were observed within the source (multi-vendor) DWI bias and were higher for ADC4 versus ADC2 fits. Limited ability for DWI DICOM parsing was observed across vendor software compared to site-specific software. All sites stored ADC maps as DICOM MR (non-PM) objects and were lacking standard attributes for source-image reference, model parameters, ADC units, and scale (all fields that would have been provided by a DICOM PM header). Site-specific ADC DICOM MR was back-compatible with existing research and commercial DICOM parsers/viewers. On the other hand, ADC DICOM PM had limited back-compatibility as it is not yet adopted by PACs vendors. As a solution to these issues, it's recommended to use the QIICR supported command line "dcmqi" developed to support ADC conversion to DICOM PM from site-specific software output to supply metadata required for centralized meta-analysis in muti-site trials.

The results from this challenge were presented last year at the QIN F2F 2017 and published in JMI [2].

DSC Project (Lead: Kathleen Schmainda, MCW)

The goal of this challenge was to compare multi-site/mult-platform analysis of a publically available brain tumor patient DSC-MRI dataset. Numerous studies have demonstrated the value of DSC-MRI perfusion metrics, but they have not been widely implemented due to the lack of confidence in the consistency of DSC-MRI metrics across sites, imaging platforms, and analysis software.

The results from this challenged showed that DSC-MRI derived normalized nRCBV and nCBF maps across sites are reliable in differentiating tumor grade. For nRCBV and nCBF, 93% and 94% of entries showed good or excellent cross-site agreement ($0.8 \leq LCCC \leq 1.0$). All hemodynamic metrics were able to distinguish low-grade from high-grade brain tumor. For the first time, these pooled results were also able to determine optimum thresholds for

nRCBV (1.4 threshold, 90% sensitivity, 77% specificity) and nCBF (1.58 threshold, 86% sensitivity, 77% specificity).

The results from this challenge were presented at International Society of Magnetic Resonance in Medicine 25th Annual Meeting (Hawaii, 2017), and a manuscript has been accepted to AJNR for publication [3].

§ Near Completed Challenges – analysis completed

DCE Project Part 2 (Lead: Wei Huang)

The goal of this challenge was to determine the effect of AIF determination from 11 prostrate DCE-MRI datasets shared amongst 9 QIN centers. Each participating site used their own site-specific methods to determine the AIF. These AIFs were then submitted to a managing center for pharmokinetic data analysis using the Shutter-Speed model with and without the use of a reference tissue (adjacent normal muscle). Literature population AIF (Parker et al.) was also included in the analysis.

The results from this challenge showed that normalizing the AIF to a reference tissue improved the agreement in the derived physiological oarameters. K^{trans} had the highest variation affected by AIF uncertainty across sites, while v_e and τ_i had the lowest variability. These results suggest that maybe v_e and τ_i should be the biomarkers used in clinical trials.

The results from this challenge were presented at International Society of Magnetic Resonance in Medicine 25th Annual Meeting (Hawaii, 2017) and the manuscript is currently being prepared and planned to be submitted to the special issue in Tomography in 2018.

§ Active Challenges – proposal accepted

DSC DRO Project (Lead: Chad Quarles, BNI)

The goal of this challenge is to understand how brain tumor DSC-MRI acquisition and post-processing affect the accuracy and multi-site consistency of computed hemodynamic biomarkers. In order to accomplish this, a validated digital-reference-object (DRO) was developed for each of the participating QIN sites based on their clinical DSC protocol. These site-specific DROs (that included pulse sequence parameters and dosing schemes) were provided to participating groups for processing (using their in-house tools) in order to assess accuracy and consistency of CBV values across QIN sites.

As of the beginning of February, the site-specific DROs have been simulated and distributed to 10 participating QIN sites. Analysis from this QIN challenge is pending these results.

Multi b-value Project (Lead: Peter LaViolette, MCW)

The goal of this challenge is to determine if DWI parameters are reliably calculated across QIN sites in both normal and cancerous regions in the prostate, and if these parameters can reliably differentiate malignant tumor from normal tissue. This challenge has roughly 30 prostrate DWI datasets that were acquired with 10 b-values with histology as the ground truth. Each participating site will compute any DWI biomarkers of their choosing (ie – IVIM, kurtosis, bi-exponential etc) and submitted their parametric maps back to MCW.

Nine QIN participating sites have been identified who will participate in this challenge. As of the beginning of February, IRB approval is being sought in order to distribute the datasets to the participating sites.

To evaluate performance of site-specific diffusion modeling software tools, Dariya Malyarenko from UMICH1 created DROs of IVIM and kurtosis models, which were tuned to the parameters for prostate cancers and normal tissue. DROs were provided to Peter LaViolette, MCW, the managing team, and are ready for distribution to the participating teams.

§Potential Challenges – proposal in preparation

1. imFIAT (Toolx) evaluation Project (Lead: Yue Cao, UMICH3 & John Buatti, UIowa)
   Yue Cao has submitted a request for a material transfer from University of Michigan to University of Iowa to evaluate imFIAT (a functional image analysis tool developed in UMICH3). A proposal will be submitted in April 2018.

CITATIONS

1. Newitt DC, Malyarenko D, Chenevert TL, et al. Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. J. Med. Imaging. 2017;5(1):1.

2. Malyarenko D, Fedorov A, Bell L, et al. Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies. J. Med. Imaging. 2017;5(1):1.

3. Kathleen M. Schmainda, Melissa A. Prah, Scott D. R and , Ying Liu, Brent Logan, Mark Muzi, Swati D. Rane, Xiao Da, Yi-Fen Yen, Jayashree Kalpathy-Cramer, Thomas L. Chenevert, Benjamin Hoff, Brain Ross, **Yue Cao**, Madhava P Aryal, Bradley Erickson, Panagiotis Korfiatis, Timothy Dondlinger, Laura Bell, Leland Hu, Christopher C. Quarles. Multi-site Concordance of DSC-MRI Analysis for Brain Tumors: Results of a NCI Quantitative Imaging Network Collaborative Project. AJRN, 2018 (in press).

QIN PET-CT Subgroup

Dmitry Goldgof, Chair
Sandy Napel, Co-Chair

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

Kinetic Analysis of Dynamic PET with FMISO (Hosted by Sadek Nehmeh/MSKCC)

The goal of Kinetic Analysis of Dynamic PET with FMISO challenge is to assess the inter-operator variability in compartment analysis (CA) of dynamic-FMISO (dyn-FMISO) PET.

Methods: Study-I: Five investigators conducted CA for 23 NSCLC dyn-FMISO tumor time-activity-curves. Study-II: Four operators performed CA for four NSCLC dyn-FMISO datasets. Repeatability of Kinetic-Rate-Constants (KRCs) was assessed.

Results: Study-I: Strong correlation (ICC > 0.9) and interchangeable results among operators existed for all KRCs. Study-II: Up to 103% variability in tumor segmentation, and

weaker ICC in KRCs (ICC-V_B = 0.53; ICC-K_1 = 0.91; ICC-K_1/k_2 = 0.25; ICC-k3 = 0.32; ICC-K_i = 0.54) existed. All KRCs were repeatable among the different operators.

Conclusions: Inter-operator variability in CA of dyn-FMISO was shown to be within statistical errors.

This challenge is successfully completed and a manuscript is published in the Clinical Imaging Journal, January 2018, [1].

§ Lung Nodule Interval Segmentation Challenge using NLST data (Hosted by Yoganand Balagurunathan and Dmitry Goldgof, USF/Moffitt CC)

The goal of the Lung Nodule Interval Segmentation Challenge using NLST data challenge is to study the variability in volume change estimates of pulmonary nodules due to segmentation approaches used across several algorithms and to evaluate these effects on the ability to predict nodule malignancy.

Methods: We obtained 100 patient image datasets from the National Lung Screening Trial (NLST) that had a nodule detected on each of two consecutive low dose computed tomography (LDCT) scans, with an equal proportion of malignant and benign cases (50 malignant, 50 benign). Information about the nodule location for the cases was provided by a screen capture with a bounding box and its axial location was indicated. Five participating Quantitative Imaging Network (QIN) institutions performed nodule segmentation using their preferred semi-automated algorithms with no manual correction; teams were allowed to provide additional manually corrected segmentations (analyzed separately). The teams were asked to provide segmentation masks for each nodule at both time points. From these masks, the volume was estimated for the nodule at each time point; the change in volume (absolute and percent change) across time points was estimated as well.

We used the concordance correlation coefficient (CCC) to compare the similarity of computed nodule volumes (absolute and percent change) across algorithms. We used Logistic regression model on the change in volume (absolute change and percent change) of the nodules to predict the malignancy status, the area under characteristic curve (AUC) and confidence intervals were reported. Because the size of nodules was expected to have a substantial effect on segmentation variability, analysis of change in volumes was stratified by lesion size, where lesions were grouped into those with a longest diameter of <8mm and those with longest diameter ≥ 8mm.

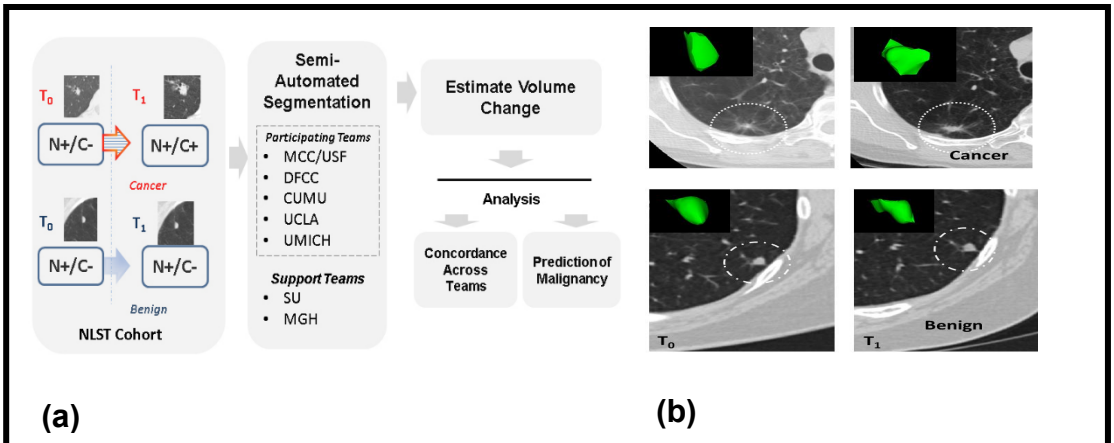


Figure 1: Describes the multi-institutional Interval Segmentation Challenge, (a) Process work flow of the study and (b) Representative patient scans (2D, center slice) at different screening intervals for nodules diagnosed to be malignant and benign (non-cancerous) at follow-up scan. The teams are: Moffitt Cancer Center/University of South Florida (MCC/USF), Dana Faber Cancer Center (MCC), Columbia University Medical Center (CUMU), and University of California at Los Angeles Medical Center (UCLA), University of Michigan Medical Center (UMICH), Stanford University (SU), and Massachusetts General Hospital (MGH).

Results: We find that segmentation of the nodules shows substantial variability across algorithms, with the CCC ranging from 0.56 to 0.95 for change in volume (percent change in volume range was [15% to 86%]) across the nodules. When examining nodules based on their longest diameter, we find the CCC had higher values for large nodules with a range of [0.54 to 0.93] among the algorithms, while percent change in volume was [30% to 95%]. Compared to that of smaller nodules which had a range of [-0.0038 to 0.69] and percent change in volume was [-3.9% to 92%]. The malignancy prediction results showed fairly consistent results across the institutions, the AUC using change in volume ranged from 0.65 to 0.89 (Percent change in volume was 64% to 86%) for entire nodule range. Prediction improves for large nodule range (≥ 8mm) with AUC range 0.75 to 0.90 (percent change in volume was 74% to 92%). Compared to smaller nodule range (<8mm) with AUC range 0.57 to 0.78 (percent change in volume was 59% to 77%).

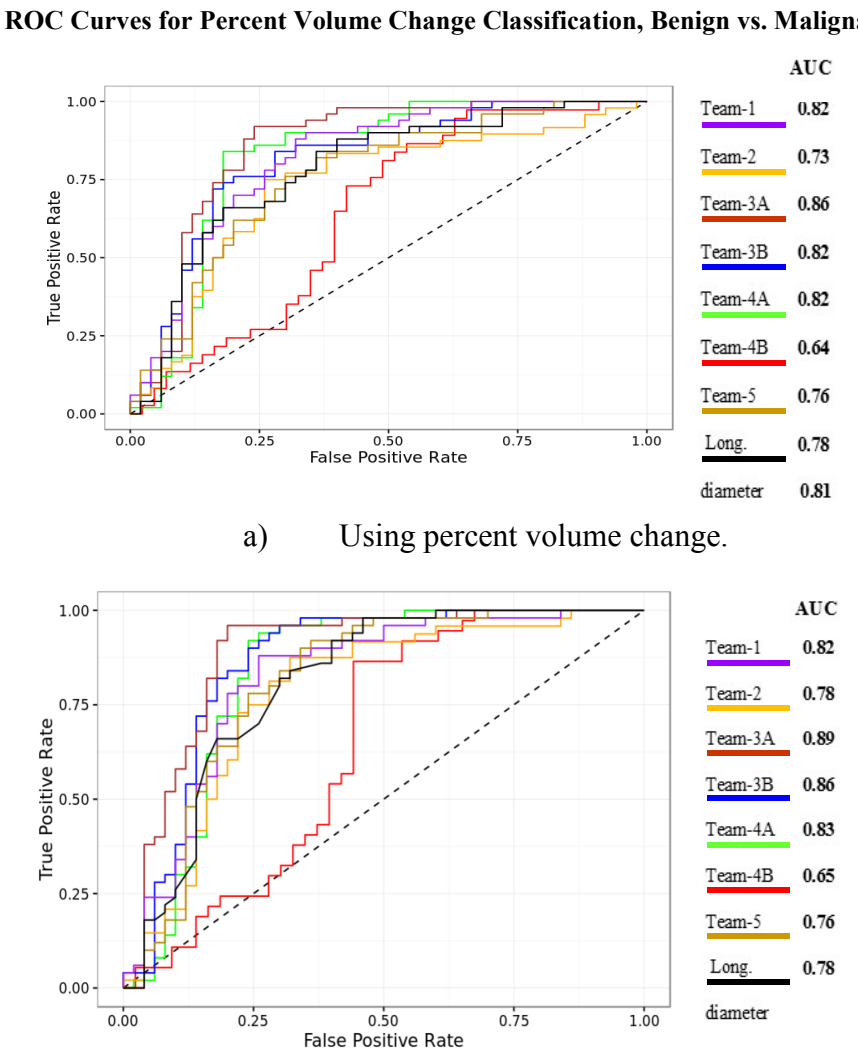


Figure 2: Receiver operator curves (ROC) across teams to predict nodules malignancy using, a) percent volume change and b) absolute volume change.

Conclusions: We find there is a fairly high concordance in the size measurements for larger nodules (≥8mm) than the lower sizes (<8mm) across algorithms. We find the change in nodule volume (absolute and percent change) were consistent predictors of malignancy across institutions, despite using different segmentation algorithms. Using volume change estimates without corrections shows slightly lower predictability (for two teams).

This challenge is successfully completed and a manuscript is published in the Medical Physics Journal, January 2018, [2].

§ Multi-center survey of PET/CT protocols for quantitative imaging in clinical trials (Hosted by Darrin Byrd/UW)

The goal of Multi-center survey of PET/CT protocols for quantitative imaging in clinical trials is to record key sources of bias and variability in PET imaging. An observational study was conducted using two surveys. The first round of surveys was designed and distributed by the American College of Radiology's Centers of Quantitative Imaging Excellence program in 2011. The second survey expanded on the first and was completed by the National Cancer Institute's Quantitative Imaging Network. Sixty-three sites responded to the first survey and 36 to the second.

Methods: Participating cancer centers underwent an initial qualification assessment that included a survey, phantom scans, assessment of clinical images, and a standardized set of quality control procedures. The CQIE qualification program included PET, CT, and MRI scanners at National Cancer Institute Designated Cancer Centers. Sites were required to submit the data reported here in order to be accredited by the CQIE.

The CQIE survey consisted of eight questions, some with sub-questions, and was repeated for both body- and brain-imaging protocols used in clinical trials at each site.

The QIN PET survey was initiated in 2013 and was designed based on the results of the CQIE survey. The survey consisted of 22 questions that were repeated for both body and brain-imaging protocols used in clinical trials at each site. Sites were encouraged to submit data on multiple PET scanners. PET scanner manufacturers and years of installation were also recorded. Both surveys asked about clinical trial protocols and did not record differences, if any existed, between clinical trial imaging and routine protocols.

Key imaging parameters varied across participating sites. The range of reported methods for image acquisition and reconstruction suggests that signal biases are not matched between sites. Patient preparation was also inconsistent, potentially contributing additional variability. For multicenter clinical trials, efforts to control biases through standardization of imaging procedures should precede patient measurements.

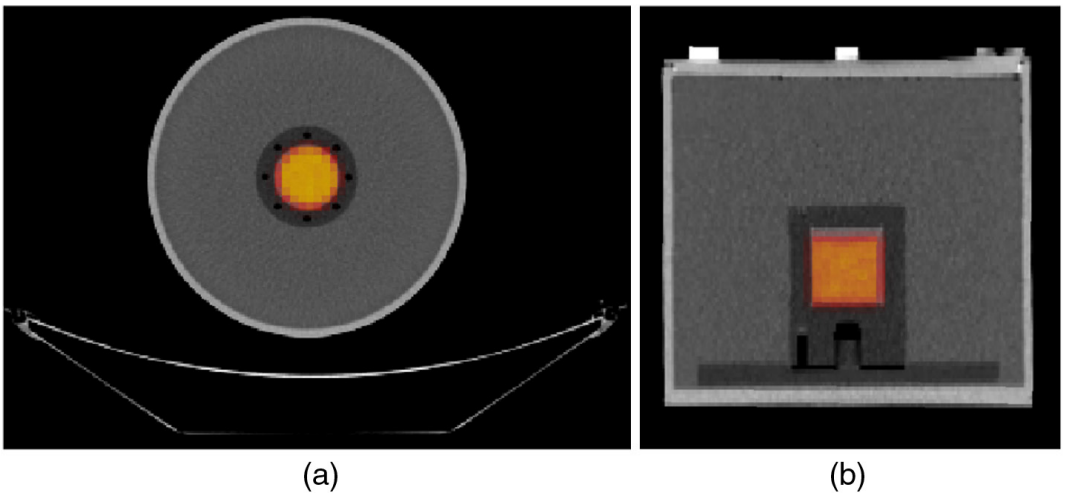


Figure 3: Fused PET/CT images of an X-Cal phantom in a water-filled 20-cm diameter cylindrical phantom. PET signal is shown in orange in (a) transaxial and (b) coronal views.

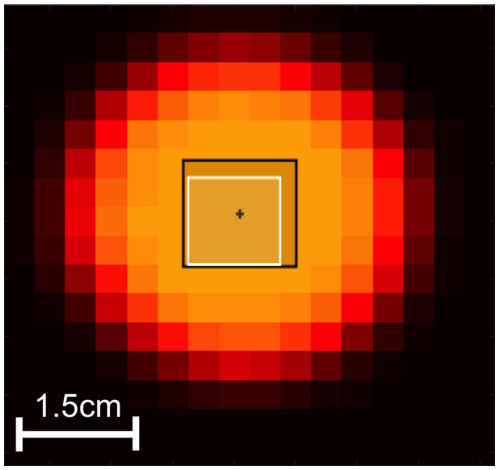


Figure 4: PET image with XCaliper ROI marked with a gray square. The black bounding box is 15 mm in each dimension (including axially) and is centered on the phantom signal's centroid, which is marked by the plus sign. The ROI includes all voxels that are completely contained in the bounding box.

Results: The CQIE survey was sent to 55 hospitals and a total of 63 unique PET/CT scanner responses were analyzed. Thirty-six hospitals responded, and a total of 44 unique PET/CT scanner responses were received.

Of the CQIE-survey scanners, 33 were General Electric, 23 were Siemens, and 7 were Philips. Of the 44 scanners included in the QIN survey, 24 were General Electric, 18 were Siemens, and 2 were Philips.

For scanners in the QIN survey, the reported trans-axial field-of view diameter was 63 ± 11 cm (range 30 to 81 cm) for body imaging and 32 ± 10 cm (range 25 to 70 cm) for brain imaging. For body imaging, the average trans-axial voxel size was 4.2 ± 1.2 mm. For brain, it was 2.0 ± 1.1 mm. Slice thicknesses had distributions of 3.57 ± 0.89 mm and 2.98 ± 0.71 mm for body and brain, respectively. Reconstruction methods varied across scanners in the QIN survey.

Of the QIN-survey sites, 24 provided a fixed value of injected radiotracer for body imaging of adults, 11 provided a range, and 15 reported a weight-based injection. For sites reporting weight-based injections for body scans, injections were computed with coefficients between 0.14 and 0.20 mCi/kg (average of 0.15 mCi/kg). Three sites did not report their method of calculating weight-based injections for body scans and one site reported using (BMI)/3 mCi, where BMI is the body mass index. The injection for brain scans was on average less than that for body scans.

All QIN-survey sites required prescan fasting for both body and brain imaging. For brain imaging, the majority of QIN-survey sites followed the same fasting period as body imaging with 17 sites requiring a 4 hours fasting period and 18 sites requiring 6 hours. For body and brain scans, carbohydrate restriction was the same at each QIN-survey site.

Twenty-nine QIN-survey sites reported 199 or 200 mg/dL as the maximum blood glucose level allowable in order to proceed with injection and scanning for body imaging. For body imaging, the most common uptake time was 60 min, and nearly half of the sites reported no difference between targeted and actual uptake times. For brain imaging, the most common time was 30 min and again approximately half of the sites reported that targeted and actual uptake times were equal.

Some QIN-survey sites provided discrete values for bed position timing and others provided a range. Fixed durations ranged from 1 to 10 min, and the distribution was 3.5 ± 1.6 min.

Conclusion: Based on the range of reported imaging parameters, we expect that large differences in biases may exist among the sites that participated in these surveys. This was an observational study, and we cannot directly estimate the combined impact of the parameter variations on the net variations in PET SUVs. One reasonable interpretation of our results is that standardization of the parameters above should precede any multicenter trial that uses PET SUVs quantitatively. This should be a high priority for future multicenter trials using quantitative imaging.

This challenge is successfully completed and a manuscript is published in the Journal of Medical Imaging, December 2017, [3].

§ Measuring temporal stability of PET SUV bias using long-lived sources in a multicenter network (Hosted by Darrin Byrd/UW)

The goal of measuring temporal stability of PET SUV bias using long-lived sources in a multicenter network challenge is to compute the bias in SUVs at multiple time points multiple sites over a 14-month period.

Methods: Calibration stability was assessed with long-lived cross-calibration kits (X-Cal kits). The kits contained two 68Ge sealed sources, one to be assayed in a dose calibrator and the other to be measured by a PET/CT scanner. The dose calibrator standard contained ~0.8 MBq (20 μCi) of 68Ge activity initially and approximated the geometry of a syringe containing a clinical dose of radiotracer.

All dose calibrator standards were calibrated to a well-characterized National Institute of Standard and Technology (NIST) 68Ge standard. The scanner source, or X-Cal phantom, was cylindrical with an active region 45 mm in diameter by 45 mm in height. The initial activity concentration was ~250 kBq/ml (6.8 μCi/ml) for a total activity of 20 MBq (0.54 mCi).

Data were collected for 19 PET/CT scanners and 16 dose calibrators at the nine sites (some dose calibrators were used for more than one PET/CT scanner). The number of scans per scanner ranged from 3 to 43 (average of 13) and the duration over which scans were performed ranged from 39 to 412 days (average of 232). The average duration between scans for the entire set of data (N ¼ 236) was 20 days. A total of 161 dose calibrator measurements were made with an average separation of 24 days between measurements.

The signal from each reconstructed image was computed as the mean of voxel values within a region of interest (ROI) in the center of the phantom. ROIs were drawn via an automated algorithm, XCaliper, which was implemented in a plug-in developed at the University of Washington for the OsiriX DICOM® viewer. For this study, the maximum ROI size was 15 mm in each dimension. This size was chosen to exclude voxels near the edge of the phantom, where signal intensity is lessened due to partial-voxel effects and finite image resolution.

For each PET/CT scan, the metadata from the DICOM® images were extracted. The metadata are saved in each image slice produced by the scanners and contain many imaging parameters.

Results: The average ρ across scanners was 0.924. The standard deviation of ρ ranged from 0.011 to 0.065. The coefficient of variation (COV) for ρ was 3.5%. Six scanners had $\Phi > 0.1$. The average Δmax was 0.057. The average r was 1.01 across all calibrators. Standard deviations of r ranged from 0.004 to 0.048. Three scanners were affected by dose calibrators with $\Phi > 0.1$. The average Φ for dose calibrators was 0.065. The average Δmax was 0.051. Average σ across scanner-calibrator pairs was 0.910. Per-pair standard deviation ranged from 0.0044 to 0.067. The standard deviation and Δmax of σ were slightly larger on average than

those of ρ and r. For many of the scanners, the scanner recovery ρ was correlated with scale factors from the DICOM® headers.

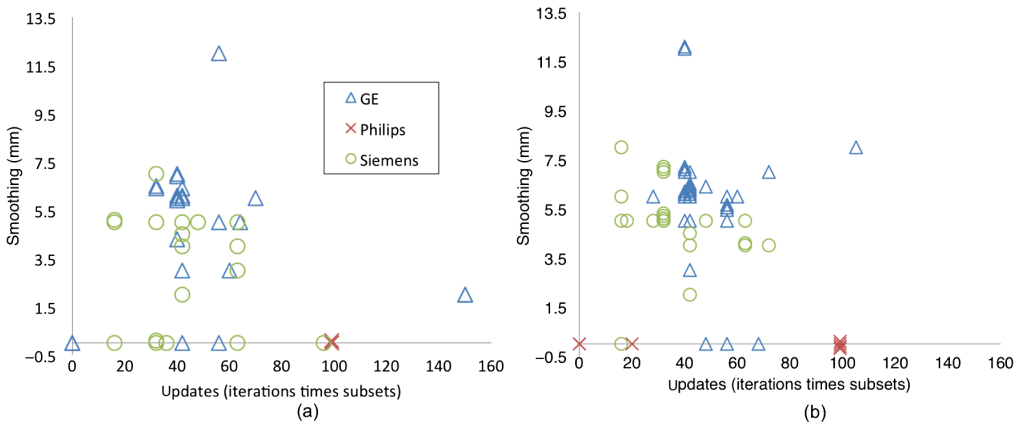


Figure 5: For body imaging, smoothing (postfilter full width at half maximum) versus iterative updates for the (a) QIN survey and (b) CQIE survey. Where multiple points lie in the same location, small shifts have been introduced to make markers visible.

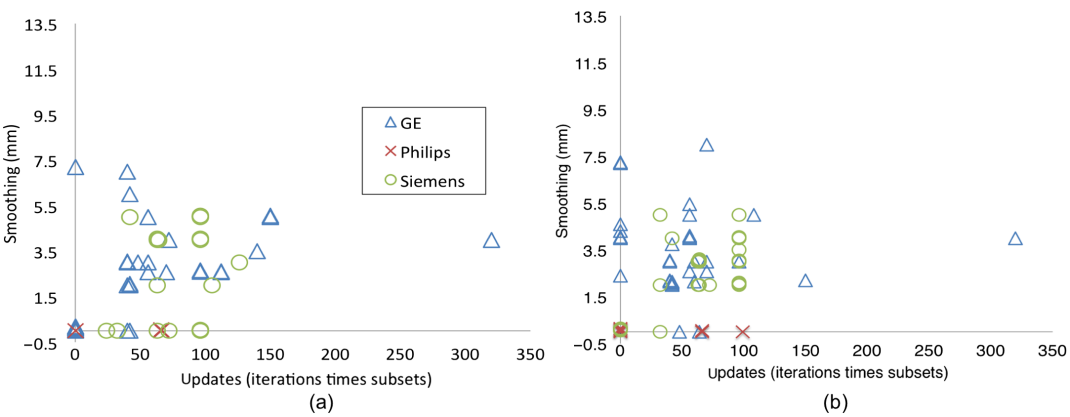


Figure 6: For brain imaging, smoothing (postfilter full width at half maximum) versus iterative updates for the (a) QIN survey and (b) CQIE survey. Where multiple points lie in the same location, small shifts have been introduced to make markers visible.

Conclusion: Our phantom images suffered some signal bias due to the epoxy used in their construction, and consequently we have focused on signal stability rather than absolute accuracy. SUV bias from instrument calibration was often stable over successive measurements, and on average had a modest COV of 3.5%. However, over the course of our 14-month study, shifts in bias were apparent for many scanners, and on average SUV recovery varied over an intra-scanner range of 11%. The biases of scanners and dose calibrators were not correlated, and estimated SUV variability was not smaller than the variability of either

instrument. Information saved in the DICOM® headers appears to show that scanner recalibration influences PET scanner bias and potentially contributes to changes in SUV bias. The variability of scale factors saved in the DICOMs® was nearly as large as that of scanner signal. This supports the conclusion that a long-lived source should be used as in independent check on the calibration process to reduce potential recalibration errors.

This challenge is successfully completed and a manuscript is published in the Journal of Medical Imaging, January 2018, [4].

§ QIN PET Segmentation Challenge (Hosted by Reinhard Beichel and Brian Smith/UIowa)

Phase I

The goal in Phase I of the multi-Institutional QIN PET segmentation challenge was to measure variability and bias in a large number of segmentations of Positron Emission Tomography (PET) scans. In addition to different lesion segmentation approaches, a high noise component related to the limited stochastic nature of the raw data, and the wide variety of reconstruction options influence segmentation performance. Understanding each element contributing to these inconsistencies in image segmentation 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 seven institutional members of the National Cancer Institute QIN. For the study, members 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 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.

Phase I of this challenge was successfully completed and a manuscript was published in Medical Physics, February 2017, [5].

Phase II

In phase II of this challenge, the goal is to assess bias and variability in imaging biomarkers derived from PET images with different segmentation methods, comparing agreement to a reference standard, studying prognostic performance, and estimating sample size for future clinical studies. For this purpose, a statistical analysis framework was developed, which is based on data obtained from phase I of the QIN PET segmentation challenge, providing tumor volumes which were measured manually and with seven different semiautomated segmentation algorithms. Estimates and comparisons of bias and variability in the resulting measurements are provided along with an R software package for the technical performance analysis and an online web application for sample size and power analysis.

Methods: The methodological approach taken aims to characterize the bias and variability that can result when estimating the risk of clinical outcomes associated with an imaging biomarker derived from different quantification methods. An underlying Bayesian model is specified for the joint distribution of method-specific biomarker measurements and sources of measurement error. Based on the model, a measurement-error-free reference biomarker is obtained and used to simulate clinical outcomes for user-specified risk associations and sample sizes. Then, based on the modeled joint distribution, risks estimated from method-specific biomarkers measured with error are compared.

Results: Statistical methods were developed and applied for the comparison of multiple quantitative methods. The Bayesian biomarker model was applied to log-transformed tumor volume measurements from the 8 QIN PET challenge methods each used to segment a common set of 47 head and neck cancer lesions. The log-transformation was needed to satisfy the model assumption of normally distributed, homoscedastic, and additive errors. From our proposed model, performance metrics commonly used in the technical validation of imaging biomarkers were derived and analyzed, including bias, operator variance, repeatability coefficient, intraclass correlation coefficient (ICC), coefficient of variation, and between-method correlation.

Conclusion: A unified Bayesian approach has been developed for the assessment, comparison, and clinical study design of quantitative imaging biomarkers. The importance of reporting more than one performance metric was discussed in relation to complementary information provided by bias, concordance, and precision measures. With respect to inference, the joint posterior distribution provided by the Bayesian approach allows for probability statements to be made about measures of interest, including credible intervals that can be

interpreted as containing the true value with a specified probability. Potential downsides to the Bayesian approach include prior specification which may be criticized as being subjective and computational challenges of obtaining MCMC samples. For study design, algorithms and an online web application are provided to determine power for and to assess the effects of estimating odds ratios with different quantification methods.

The statistical analysis approach was published in the Statistical Methods in Medical Research Journal, January 2017, [6]. So far, our analysis was focused on tumor volume as the biomarker, and currently we are working on investigating relevant quantitative PET imaging biomarkers with our analytic methods and software and plan on publishing our findings in a paper.

PLANS FOR THE NEXT YEAR (ACTIVITIES IN PROGRESS AND PLANNED)

§ Lung Nodule Malignancy Prediction, Based on Sequential CT Scans (Lead Keyvan Farahani/NCI and Yoganand Balagurunathan/MCC along with QIN members)

This challenge intends to advance methods development on the current clinical impediment to assess nodules status for lung cancer screening subjects with consecutive scans.

We will provide sequential low-dose CT (LDCT) scans at two screening intervals from the National Lung Screening Trial (NLST), with matched identified nodules from the same subject. We would like the participating teams to provide estimated nodules dimensions (longest diameter, volume) in the screening interval and the probability of malignancy. The teams are open to use any radiomic descriptors for nodules across time points and or change in size measurements including doubling time (DT) toward their assessment. If teams prefer to use doubling time (DT) metric (measured in days), following formulation is stated for reference.

$$DT = (t_2 - t_1) * (\log 2 / (\log(V_2) - \log(V_1)))$$

Where V_1, and V_2 are the nodules volume (or size) measured at two screening intervals; in our study, t_1 and t_2 are baseline and diagnostic scan time respectively. Participants may use any other preferred formulation, any variant formulation, need to be described with reasoning in their respective training summary report.

Specifically, participants are asked to submit files that include nodule size (longest diameter), volume, and a probability of malignancy score (range from 0 to 1, for absence or presence of cancer, respectively) for both the train and blinded test cases. The participants will be evaluated on the test data performance (AUC).

§ Phantom study in 'hypoxic fraction/volume' measurement (Hosted by Ivan Yeung/PMCC)

Hypoxia imaging using FMISO, FAZA and EF5 has low uptake compared to FDG imaging. One way to quantify hypoxia imaging is by 'hypoxic fraction' (HF) or 'hypoxic volume (HV) which is the proportion of voxels in a tumor having uptake value higher than a threshold. There is no consensus as how to set the threshold. In addition, the quantification of HF depends on the uptake distribution and therefore the noise characteristic of the particular scanner. A collaborative project is proposed with the primary aim to measure the variation of HF/HV measurement at different PET scanners by scanning a phantom which is manufactured to have different percentages of uptake regions. The secondary aim is to investigate possibly a more robust method of quantifying HF/HV among different scanners.

§ PET/CT Multi-Institutional, multi-reader observer study (Hosted by Reinhard Beichel and Brian Smith/UIowa)

In the first analysis phase of this study, seven institutional members of the QIN applied different segmentation methods to a common set of phantom scans acquired over a range of imaging conditions as well as to a second set of head and neck cancer scans. Tumor volumes derived from the segmentations were analyzed statistically to assess variability and bias across methods. Findings were published in Medical Physics [5] as summarized in the abstract provided earlier in this report. In the second analysis phase, we are assessing the technical and clinical performance of quantitative imaging biomarkers (QIBs) derived from the segmentations. A statistical analysis framework has been developed and published [6] for this purpose. During the upcoming year, we plan to apply the analysis framework to a panel of QIBs. Technical performance will be assessed with respect to variability and bias in QIB measurements derived from the different segmentation methods. Clinical performance will be assessed with respect to the effect of the different segmentation methods on bias and variability in estimating risk associations between QIBs and clinical outcomes. A preliminary analysis of this second phase has been conducted on 22 QIBs. The full scope of the analysis is being formulated with input from the PET/CT subgroup members, who will have the opportunity to contribute to a manuscript summarizing the analysis approach, results, and conclusion for submission later this year.

§ Feature Ontology Effort (Hosted by Michael McNitt-Gray/UCLA)

Possible Goals:

1. Mission of this effort (Gap analysis. What is our scope? What are others doing and how can we interact with and/or complement existing efforts such as ISBI)

2. Region of Interest Types/Definitions (Larry Pierce suggested this and also thought we could come up with an exhaustive list; this would be a helpful "pre-requisite" for many feature definitions.)

3. Phantoms (Digital and Physical), Reference Datasets and a testing procedure (including possibly a suggested progression). This has been suggested by several folks where one can start with analytical phantoms, move to physical phantoms (or rather digital image representations of physical phantoms) and then to patient datasets (or something like this).

4. Reference implementations of features (not an exhaustive list of features, but "representative implementations" from different feature groups that could be shared publicly).

ACTIVITIES ACROSS WORKING GROUPS

- Tool interoperability demo for 2017 F2F. (PET-CT/BIDS)
- Working with Keyvan Farahani on QIN Benchmarks
- Working with Keyvan Farahani/IBSI on a lung cancer challenge
- Welcoming Image Acquisition Group into PET-CT (two activities completed)

LIST OF PRIMARY REFERENCES

[1] Nehmeh SA, Schwartz J, Grkovski M, Yeung I, Laymon CM, Muzi M, Humm JL. Inter-operator variability in compartmental kinetic analysis of 18 F-fluoromisonidazole dynamic PET. Clinical Imaging. 2018 Jan 12.

[2] Balagurunathan Y, Beers A, Kalpathy-Cramer J, McNitt-Gray M, Hadjiiski L, Zhao B, Zhu J, Yang H, Yip SS, Aerts HJ, Napel S. Semi-Automated Pulmonary Nodule Interval Segmentation using the NLST data. Medical physics. 2018 Jan 24.

[3] Byrd DW, Christopfel R, Buatti JM, Moros EG, Nehmeh S, Opanowski A, Kinahan PE. Multicenter survey of PET/CT protocol parameters that affect standardized uptake values. Journal of Medical Imaging. 2017 Dec;5(1):011012.

[4] Byrd D, Christopfel R, Arabasz G, Catana C, Karp J, Lodge MA, Laymon C, Moros EG, Budzevich M, Nehmeh S, Scheuermann J. Measuring temporal stability of positron emission tomography standardized uptake value bias using long-lived sources in a multicenter network. Journal of Medical Imaging. 2018 Jan;5(1):011016.

[5] 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 Feb 1;44(2):479-96, PMID: 28205306.

[6] Smith BJ, Beichel RR. A Bayesian framework for performance assessment and comparison of imaging biomarker quantification methods. Statistical methods in medical research. 2017 Jan 1:0962280217741334, PMID: 29271301