Translational Oncology

www.transonc.com

Volume 9 Number 2 April 2016 pp. 155–162 155

Radiomics Analysis on FLT-PET/MRI for Characterization of Early Treatment Response in Renal Cell Carcinoma: A Proof-of-Concept Study[1]

Jacob Antunes*, Satish Viswanath*, Mirabela Rusu†, Laia Valls‡, Christopher Hoimes§, Norbert Avril‡ and Anant Madabhushi*

*Department of Biomedical Engineering, Case Western Reserve University, 10900 Euclid Ave, Wickenden 525, Cleveland, OH 44106; †General Electric Global Research, 1 Research Cir, Niskayuna, NY 12309; ‡Department of Radiology, Case Western Reserve University, Case Center for Imaging Research, 11000 Euclid Ave, Cleveland, OH 44106; §Department of Medicine, University Hospitals Seidman Cancer Center at the Case Comprehensive Cancer Center, 11000 Euclid Ave, Cleveland, OH 44106

Abstract

Studying early response to cancer treatment is significant for patient treatment stratification and follow-up. Although recent advances in positron emission tomography (PET) and magnetic resonance imaging (MRI) allow for evaluation of tumor response, a quantitative objective assessment of treatment-related effects offers localization and quantification of structural and functional changes in the tumor region. Radiomics, the process of computerized extraction of features from radiographic images, is a new strategy for capturing subtle changes in the tumor region that works by quantifying subvisual patterns which might escape human identification. The goal of this study was to demonstrate feasibility for performing radiomics analysis on integrated PET/MRI to characterize early treatment response in metastatic renal cell carcinoma (RCC) undergoing sunitinib therapy. Two patients with advanced RCC were imaged using an integrated PET/MRI scanner. [18 F] fluorothymidine (FLT) was used as the PET radiotracer, which can measure the degree of cell proliferation. Image acquisitions included test/retest scans before sunitinib treatment and one scan 3 weeks into treatment using [18 F] FLT-PET, T2-weighted (T2w), and diffusion-weighted imaging (DWI) protocols, where DWI yielded an apparent diffusion coefficient (ADC) map. Our framework to quantitatively characterize treatment-related changes involved the following analytic steps: 1) intraacquisition and interacquisition registration of protocols to allow voxel-wise comparison of changes in radiomic features, 2) correction and pseudoquantification of T2w images to remove acquisition artifacts and examine tissue-specific response, 3) characterization of information captured by T2w MRI, FLT-PET, and ADC via radiomics, and 4) combining multiparametric information to create a map of integrated changes from PET/MRI radiomic features. Standardized uptake value (from FLT-PET) and ADC textures ranked highest for reproducibility in a test/retest evaluation as well as for capturing treatment response, in comparison to high variability seen in T2w

Address all correspondence to: Jacob Antunes, Department of Biomedical Engineering, Case Western Reserve University, 10900 Euclid Ave, Wickenden 525, Cleveland, OH 44106.

E-mail: jacob.antunes@case.edu

[1]This publication was made possible by the National Cancer Institute of the National Institutes of Health under award numbers R21CA167811-01, R21CA179327-01, R21CA195152-01, and U24CA199374-01; the National Institute of Diabetes and Digestive and Kidney Diseases under award number R01DK098503-02; the National Center for Advancing Translational Sciences award number UL1TR000439; the DOD Prostate Cancer Synergistic Idea Development Award (PC120857); the DOD Lung Cancer Idea Development New Investigator Award (LC130463); the DOD Prostate Cancer Idea Development Award; the Ohio Third Frontier Technology development grant; the CTSC Coulter annual pilot grant; the Case Comprehensive Cancer Center pilot grant; the VelaSano grant from the Cleveland Clinic; and the Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering at Case Western Reserve University. Philips Healthcare provided funding for clinical PET/MR imaging.
Received 2 October 2015; Revised 20 January 2016; Accepted 27 January 2016

1936-5233/16

http://dx.doi.org/10.1016/j.tranon.2016.01.008

MRI. The highest-ranked radiomic feature yielded a normalized percentage change of 63% within the RCC region and 17% in a spatially distinct normal region relative to its pretreatment value. By comparison, both the original and postprocessed T2w signal intensity appeared to be markedly less sensitive and specific to changes within the tumor. Our preliminary results thus suggest that radiomics analysis could be a powerful tool for characterizing treatment response in integrated PET/MRI.

Translational Oncology (2016) 9, 155–162

Introduction

Early characterization of *in vivo* structural and functional changes in tumors due to treatment is vital for defining individualized treatment stratification and follow-up [1]. If treatment response evaluation is performed late in the treatment regimen, this may result in a substantial overtreatment or ineffective treatment of many patients: 1) those that do not respond to therapy should be referred to alternative treatments, and 2) those that developed a maximal metastatic inhibition response could benefit from adjuvant chemotherapy [2].

Fluorothymidine (FLT) has emerged as a promising positron emission tomography (PET) radiotracer for imaging cell proliferation [3] and associated treatment effects. In Murakami et al. (2013), FLT uptake levels in human renal cell carcinoma (RCC) xenograft models decreased after only 7 days into cytostatic treatment [4]. Bao et al. (2014) used a cytostatic drug, sunitinib, in their application of treating human glioblastoma xenograft models and also saw decreased FLT uptake levels after only 7 days [5]. Thus, in cytostatic treatments which attempt to inhibit tumor proliferation (contrary to conventionally used cytotoxic drugs whose effects can be measured via tumor volume reduction), FLT uptake may be a useful imaging biomarker for visualizing early treatment effects.

Magnetic resonance imaging (MRI) provides high-contrast structural and functional information for characterizing soft tissue and has been examined extensively in both pre- and posttreatment settings. T2-weighted (T2w) MRI demonstrates pathological features arising from differences in water content within internal structures. Diffusion-weighted imaging (DWI) MRI captures changes in the cellular architecture of the tissue based on differences in movement of water protons within different tissue regions. Diffusion can be quantified by generating an apparent diffusion coefficient (ADC) map, which has previously demonstrated great promise as an imaging biomarker of treatment response [6,7]. In therapeutic options that affect the vascularization of tumors (e.g., cytostatic treatment) and thus the anatomical structure of tissue, both T2w MRI and ADC maps may be expected to capture *in vivo* changes due to treatment.

Furthermore, combining multiple imaging protocols (e.g., PET and MRI) holds great promise in clinical oncology applications, especially with the advent of newly developed hybrid scanners [8]. MRI offers anatomical localization and an attenuation correction map for quantification of PET data due to concurrent acquisition within the same frame of reference, enabling an overlay of independent *in vivo* structural, functional, and metabolic characterizations of tumor response.

Conventional assessment of *in vivo* imaging is based on tumor morphology measurements, which have demonstrated a limited ability to identify treatment response. For example, the most widely used method in clinical standard of care to evaluate response to cytostatic drugs is the Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST v1.1) [9]. The RECIST v1.1 classification is solely based on unidimensional measurements on radiologic imaging to quantify changes in parameters such as the tumor diameter on MRI. However, morphological changes may only appear months into the treatment regimen, because of which RECIST v1.1 may not be able to identify any early treatment effects [10]. Alternative characterization of tumor response using *in vivo* imaging have involved measuring changes in contrast (for MRI), metabolic activity (for PET), or functional activity (for DWI). Although there are guidelines for assessing treatment response on PET and MRI independently [9,11,12], to the best of our knowledge, there are no techniques for a joint combined assessment of PET/MRI to evaluate treatment response in oncological applications.

We hypothesize that to quantitatively characterize treatment-related changes in the tumor region on imaging modalities such as PET and MRI, we need to better capture these subtle changes in the tumor region before macroscopic changes in tumor morphology become visible. Furthermore, by quantitatively combining parameters derived from PET, ADC, and T2w MRI, we expect an improved characterization of treatment response *in vivo* compared with the individual imaging sequences.

To accurately quantify changes in the molecular and functional characterization of the tumor in the posttreatment setting, specific challenges must be overcome. First, the different imaging acquisitions (PET, T2w, ADC) must be transformed into voxel-wise correspondence to evaluate the spatial heterogeneity of associated parameters. By more rigorously evaluating treatment-related changes, we would expect to achieve greater sensitivity than region-based approaches (such as RECIST v1.1). Second, acquisition artifacts due to patient movement and image parameter drift [13,14] must be corrected for, as these would significantly affect the ability of imaging parameters to capture subtle treatment-related changes in a voxel-wise approach. Finally, visual assessment of imaging parameters may not be able to identify subtle changes in the tumor region *in vivo*, a problem further exacerbated when evaluating such changes early into treatment (a time point at which these changes may be even less visually appreciable). This issue may be overcome by deriving computer-extracted features from radiographic images, termed *radiomics*[15]. Radiomic features attempt to capture "image texture" through quantification of local changes in image intensity values in relation to their voxel-wise arrangement [16–19]. This offers the ability to quantitatively capture subvisual "textural" changes in the tumor region, changes that may characterize early treatment response but escape visual identification.

The current study examines these methods in the context of evaluating treatment response in patients with metastatic RCC, which has a poor prognosis with a 5-year survival rate of 12.3% [2].

Recently, tyrosine kinase inhibitors (TKIs) have been introduced as a cytostatic treatment for advanced RCC. For example, sunitinib is a TKI used as first-line therapy for patients with relapsed or medically unresectable RCC because it targets the vascular endothelial growth factor and its receptor [20,21], inhibition of which has been shown to prolong progression-free survival in metastatic RCC [22,23]. Preclinical studies have shown that early treatment response as a result of sunitinib is a combination of both antiangiogenesis as well as inhibition of tumor proliferation [24,25], implying that FLT-PET and MRI may be able to capture and visualize these treatment-related effects *in vivo*.

In this proof-of-concept study, we present a first attempt at performing radiomics analysis to identify imaging features and a combination thereof that are both sensitive and specific for identifying voxel-wise *in vivo* treatment-related changes on FLT-PET/MRI acquired from metastatic RCC patients. For each patient, acquisition sets acquired on different days were spatially coregistered, in addition to which imaging protocols within each acquisition were spatially aligned, to enable a voxel-wise comparison across different imaging sequences. T2w intensities were additionally processed by correcting for intensity nonuniformity (within an image) as well as for intensity drift between patient acquisitions. A number of different radiomic features were then extracted from each protocol to obtain a comprehensive set of quantitative descriptors of image structure and function to ensure that we completely characterized response to treatment despite having a limited-sized data set. Finally, a weighted multiparametric (MP)-PET/MRI signature was constructed to give an optimal assessment of early treatment response *in vivo*.

Material and Methods

Data Description

Two patient studies were obtained from a prospective study evaluating the use of FLT-PET/MRI in monitoring antiangiogenic therapy (PI: Dr. N. Avril, MD, ClinicalTrials.gov Identifier: NCT02055586) at the Case Comprehensive Cancer Center Cleveland. Before imaging, patients had been diagnosed with advanced metastatic RCC and were scheduled to begin sunitinib therapy. Before starting treatment, patients were imaged on consecutive days (termed *test/retest*). Patient 1 was additionally imaged 3 weeks into the treatment regimen (termed *mid*). Image acquisitions were acquired using the Ingenuity TF PET/MR (Philips Healthcare, Cleveland, OH). Imaging sequences considered in each acquisition set included a respiratory-triggered T2w turbo spin echo MRI, DWI spin echo, and a whole-body PET protocol. DWI yielded an ADC map (b values = 0, 400, 800). [18 F]-labeled FLT was used as the PET radiotracer. Annotations of the tumor region (RCC) as well as a homogenous region of healthy tissue (termed *norm*) were obtained on the T2w volumes via an expert radiologist (Figure 1).

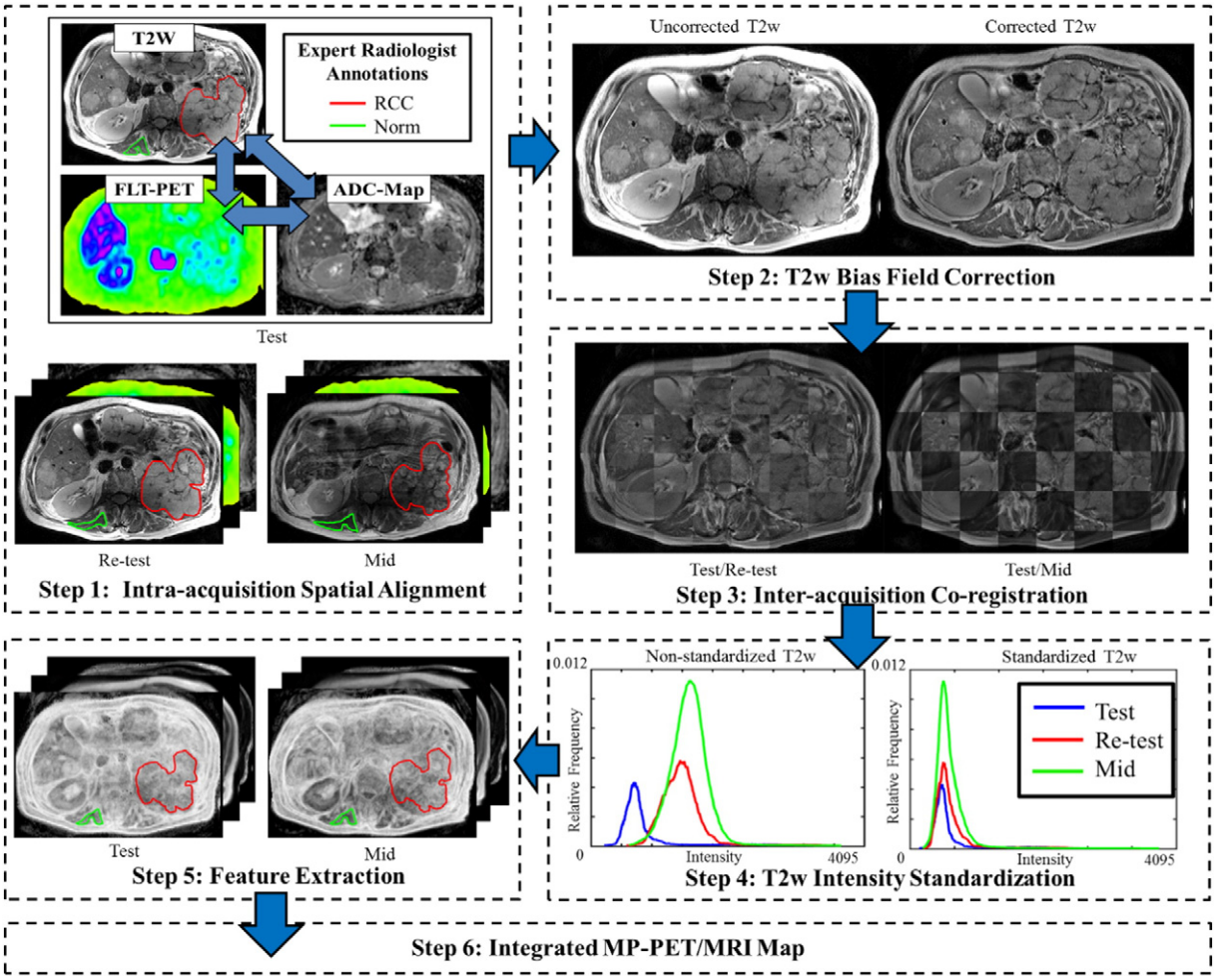

Figure 1. Overview of the methodology and overall workflow.

Table 1. Parameters for Each PET/MRI Protocol Used in Study

Parameters	T2w	DWI	PET
Sequence	RT T2w TSE	DWIse	WB PET (at ~60 min Δt)
Matrix dimensions	480 × 480	192 × 192	144 × 144
Slice thickness (mm)	5.5	5.5	4
Acquisition parameters	TR/TE = 1590/80 ms	*b* values = 0, 400, 800	Radiopharmaceutical = FLT-F^{18}

RT T2w TSE, respiratory-triggered T2w turbo spin echo; *DWIse*, DWI spin echo; *WB*, whole body; *TR*, repetition time; *TE*, echo time.

Intraacquisition Spatial Alignment

T2w, PET, and ADC map image volumes within each acquisition set were spatially aligned to one another using the T2w volume as a reference. This was done by resampling and cropping PET and ADC map volumes to the same field of view (as the T2w volume) to account for imaging differences in protocol-specific image dimensions and voxel sizes. This enabled a voxel-wise correspondence between the three PET/MRI protocols under consideration (Table 1).

Bias Field Correction

Intensity inhomogeneity is an MR image acquisition artifact which refers to the gradual intensity variation within a tissue region over the image domain [26]. This occurs due to imaging instrumentation radiofrequency nonuniformity or static field inhomogeneity and is known to significantly affect image analysis methods and radiomic feature extraction [27]. Intensity inhomogeneity is visualized as a significant variation in gray-level intensities across the MR image and was found to be present in the T2w images upon visual inspection. Thus, before registration, T2w volumes were bias field corrected for this image intensity nonuniformity by convolving the images with a Gaussian low-pass filter, resulting in uniform intensities across the volume [28].

Interacquisition Coregistration

Acquisitions acquired on each of the 3 days were found to be misaligned because of organ and patient movement. Three-dimensional (3D) affine coregistration followed by a 3D deformable coregistration was applied between T2w volumes across the different acquisition sets to enable per-voxel correspondence and comparison between test/retest and test/midtreatment acquisitions. Image registration was driven using mutual information as the similarity

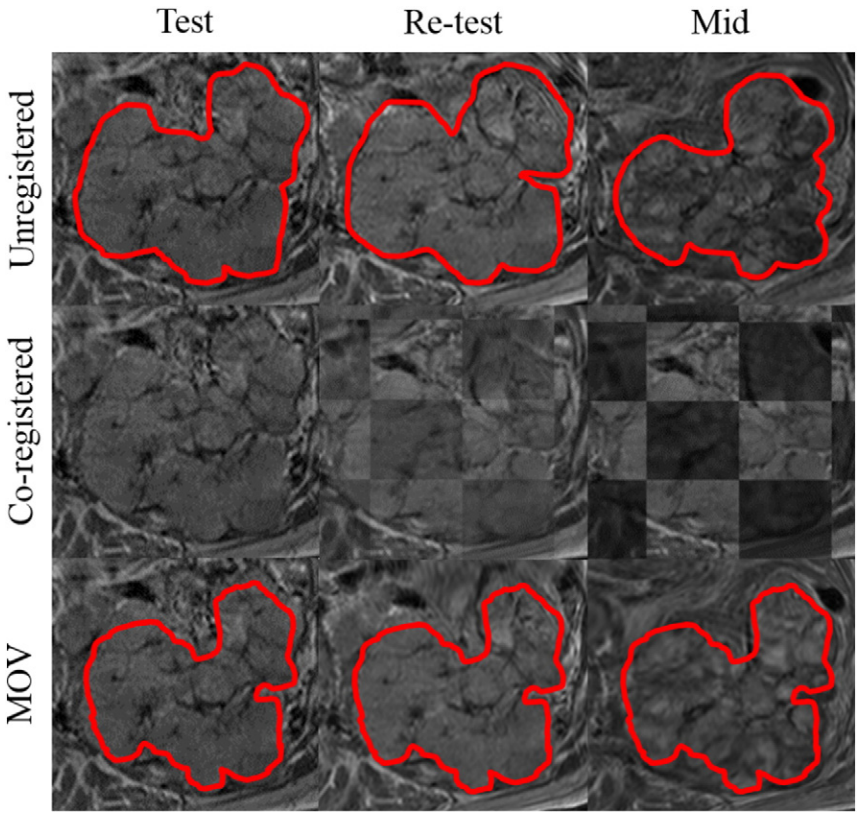

Figure 2. Transformation of RCC volumes into a voxel-wise correspondence across acquisitions. (Top row) Annotations of RCC region on original T2w volumes. Note slight differences in annotated region between each column. (Middle row) Results of 3D rigid plus deformable coregistration shown as checkerboard images. Note contiguity of structures across checkerboard, indicating accuracy of registration. (Bottom row) Creation of minimum overlapping volume mask from intersection of transformed annotations of coregistered RCC region across acquisitions to ensure voxel-wise comparison within tumor region. Note that annotated region and structures are consistent across all three columns. Dice similarity coefficients between test/retest and test/midtreatment volumes were calculated as 0.90 and 0.83, respectively (note ideal Dice value between two volumes should be 1).

measure and was performed using the ITK framework in Elastix [29]. The transformation matrices obtained from each registration T2w pairing were then applied to the corresponding PET and ADC maps (i.e., the matrix to transform retest T2w into alignment with test T2w was then applied to retest PET and retest ADC map.) Thus, all three protocols across all three acquisition sets were in the same frame of reference with the ability to perform voxel-wise correspondences.

Separate annotation volumes had been created for each of test/retest/midtreatment T2w acquisitions. The transformation matrices from the coregistration step were applied to these annotations, following which a minimum overlapping volume mask was made from the intersection of test, retest, and midtreatment acquisition annotations. This was done separately for annotations of healthy tissue and RCC regions, which were then used as regions of interest for the remainder of the analysis (Figure 2).

Intensity Standardization

T2w images visualize different *in vivo* tissue regions as having different levels of intensity contrast within the same field of view. However, the image intensities in T2w MRI lack a fixed tissue-specific meaning even when protocol, scanner, and patient parameters are held constant [30]. This artifact is termed *intensity nonstandardness*, which implies that differences in image intensities between corresponding voxels within the same tissue region, but from 2 different T2w acquisitions, cannot be assigned quantitative meaning. Thus, T2w volumes were standardized to align intensity distributions across acquisitions using the method presented by Nyul et al. (2000) [30]. Intensities of the homogenous healthy tissue regions from each T2w acquisition were used to generate standardization maps which were then applied to the entire T2w volume. The output of standardization is a pseudoquantitative T2w value at every voxel, which has tissue region-specific meaning. Intensity standardization was performed after registration because registration was found to exacerbate drift between intensity distributions (possibly because of implicit interpolation).

Radiomics Analysis

For each imaging protocol, a specific set of quantitative features (where f is the set of all features) was extracted within the annotated regions. Standard uptake value (SUV) was measured from FLT uptake in PET volumes using a decay-correction intensity computer algorithm [31]. ADC map values were automatically generated from the magnitude of diffusion of water molecules in DWI volumes (b values = 0, 400, 800).

Texture features were extracted from both the postprocessed T2w and ADC volumes. A total of 30 first- and second-order statistical features were used, including nonsteerable gradient (gray level, Kirsch, Sobel) [32] and Haralick [33] operators. Radiomic analysis on PET data has demonstrated limited robustness and reproducibility due to the low spatial resolution of PET (in comparison to T2w, ADC) [34,35]. Based on findings in a recent validation study, we decided to only extract entropy and difference average texture features from FLT-PET [36]. In total, a set of 66 radiomic features was implemented (raw T2w signal, postprocessed T2w, 30 postprocessed T2w textures, raw ADC map, 30 ADC textures, SUV, and 2 PET textures).

Statistical Analysis

To rank features, the distribution of feature intensities within the RCC and healthy tissue regions across acquisition sets was compared. Differences in intensity distributions (D) were measured using the Bhattacharyya distance [37]. Bhattacharyya values (B) were normalized from $0 \leq B \leq 1$ to represent perfectly aligned and perfectly misaligned intensity distributions, respectively.

A two-part scoring function was implemented:

$$S_1 = B\left(D_{norm}^{test}, D_{norm}^{retest}\right)$$

$$S_2 = B\left(D_{RCC}^{test}, D_{RCC}^{mid}\right)$$

where S_1 quantifies the variability of each parameter between test/retest acquisitions (i.e., a measure of specificity) and S_2 quantifies the ability of a parameter to capture early response between test/midtreatment acquisitions (a measure of sensitivity). Intensity distributions of a homogenous healthy region between test/retest acquisitions are expected to be aligned (i.e., ideally, $S_1 = 0$ implies low variability), whereas intensity distributions of the RCC region are expected to be markedly misaligned as a result of sunitinib treatment (i.e., ideally, $S_1 = 1$ implies maximal response). Features were ranked based on a combination of the feature's performance in each of the two parts of the scoring function based on a weighted ratio:

$$S_{overall} = \frac{{S_2}^{w_2}}{{S_1}^{w_1}}$$

The weighting factors were set as $w_1 = w_2 = 0.5$ to attribute equal importance to both terms of the scoring function. $S_{overall}$ was computed for each feature in f.

Multiparametric Map

Percent difference (denoted %Δ) for each feature $f_i \in f$, $i = \{1, \ldots, 66\}$ was measured on a per-voxel basis between test/midtreatment acquisitions. A weighted difference image ($\bar{I}$) was then computed by combining %Δ from a subset of features $\overline{f}_j \subset f,\ j \in \{1, \ldots, n\}$ using a weighted summation as follows:

$$\overline{I} = \sum_{j=1}^{n} a_{\overline{f}_j} * \%\Delta_{\overline{f}_j}$$

where the weighted contribution factor $a_{\overline{f}_j}$ for each feature $\overline{f}_j$ is based on how well $S_{overall}$ for each feature $\overline{f}_j$ ranked relative to other features. The number of features used in generating $\bar{I}$ was set as $n = 3$, where SUV, the top-ranked ADC feature, and the top-ranked T2w feature were combined to yield a MP-PET/MRI difference image.

Experimental Results and Discussion

Table 2 summarizes the top 25 PET/MRI radiomic features for each patient, ranked in the descending order, based on the scoring function S_1 alone. It may be observed that SUV and ADC Haralick features are relatively highly ranked in both patients, suggesting low variability in a test/retest evaluation. By comparison, ADC gradient features and the original T2w signal intensity ranked relatively high for patient 2 only and therefore may not be as reliable. Most T2w radiomic features, including the postprocessed T2w signal intensity, the raw ADC value, and both PET texture features, were among the lowest-ranked features, implying relatively high variability in a test/retest evaluation. It has previously been shown that, in a comparison of feature reliability, ADC texture features have ranked higher than T2w features in demonstrating more consistent changes in healthy tissue regions [16]. In a previous case study that presented some of the data used in this work [38], FLT-PET was shown to have an overall

Table 2. Top 25 Ranked PET/MRI Radiomic Features for Each Patient Based on Scoring Function S_1 (Quantifying Test/Retest Variability)

S_1 Rank	Patient 1	Patient 2
1	T2w gradient Y	ADC gradient X
2	ADC energy	ADC Sobel YX
3	T2w Sobel Y	ADC entropy
4	ADC difference average	ADC Sobel X
5	SUV	SUV
6	ADC gradient X	ADC sum entropy
7	ADC inverse difference moment	ADC difference entropy
8	T2w correlation	ADC inverse difference moment
9	ADC Sobel YX	ADC energy
10	T2w Sobel YX	ADC range
11	ADC difference entropy	ADC Sobel XZ
12	ADC difference variance	T2w correlation
13	ADC gradient Y	ADC gradient Y
14	ADC entropy	ADC gradient magnitude
15	ADC Sobel X	ADC difference average
16	T2w gradient X	PET difference average
17	T2w Sobel XY	Raw T2w
18	ADC inertia	ADC Sobel YZ
19	ADC sum entropy	ADC Sobel ZY
20	ADC Sobel Y	PET entropy
21	ADC Sobel XY	ADC information metric 1
22	ADC information metric 1	ADC information metric 2
23	ADC information metric 2	T2w gradient Y
24	T2w Sobel X	ADC gradient Z
25	ADC correlation	ADC Sobel Z

high repeatability (i.e., low variability) between test/retest studies. DWI-ADC was found to be less repeatable likely because of respiration artifacts affecting image quality, whereas T2w MRI and radiomic features were not evaluated.

Figure 3 visualizes box plots of %Δ between test/midtreatment within healthy tissue and RCC regions for the three top-ranked radiomic feature based on the scoring function $S_{overall}$ (in order: SUV, ADC energy, and T2w difference average) for patient 1 alone, as well as for the original signal intensities (midtreatment acquisition for patient 2 was unavailable). The three top-ranked features were combined to yield an MP-PET/MRI map, for which box plots of %Δ are also shown. SUV and both T2w and ADC texture features appear to be able to capture treatment-related changes, whereas the original T2w signal intensity does not appear to capture any significant change other than differences due to acquisition variability. The MP-PET/MRI map may be considered most reflective of a marked difference in the tumor region but may not necessarily best reflect change due to treatment alone (because of a greater change captured in the healthy tissue compared with SUV). This indicates the need for either a better weighting of the different modalities or a modified scoring function to optimize the integrated map.

Figure 4 visualizes the difference maps between test/midtreatment acquisitions for patient 1. Displayed are percent difference images for the three top-ranked radiomic features based on $S_{overall}$ and for the original signal intensities. Hotter colors represent greater percent changes as a result of sunitinib treatment. Although differences in uptake appear across the entire functional tumor volume of FLT-PET, regions of change captured by SUV appear to be rather homogeneous (minimal differences within a small neighborhood of voxels), reflecting marked changes being captured by molecular imaging due to treatment, as well as the limited spatial resolution of PET. The regions of change seen on the difference maps of the original ADC and T2w images are visually much more heterogeneous (marked differences within a small neighborhood of voxels), indicative of these modalities capturing voxel-wise changes at the anatomic or structural level. Each of the different PET/MRI protocols appears to capture unique information regarding early treatment response, and the MP-PET/MRI map appears to combine these complementary sources of information to yield a comprehensive characterization of response due to TKI treatment.

The raw T2w value and postprocessed T2w signal intensities were among the lowest-ranking features based on $S_{overall}$, reflecting high test/retest variability as well as not optimally capturing treatment-related changes. By comparison, T2w texture features appear to yield an improved performance in capturing treatment-related changes due to cytostatic treatment. Similarly, ADC texture features also appeared to capture additional information related to treatment change compared with the raw ADC value. In contrast, texture features from PET (not shown) appear to exhibit high test/retest variability and be unable to capture treatment-related changes. This however is somewhat expected because of the limited spatial resolution of PET [36]. SUV was the highest-ranking feature overall.

T2w radiomic features have previously been demonstrated to be able to differentiate tumor and healthy regions based on appearance of the prostate [17,39]. In addition, in cytotoxic applications, such as laser ablation, both T2w and ADC radiomic features were able to

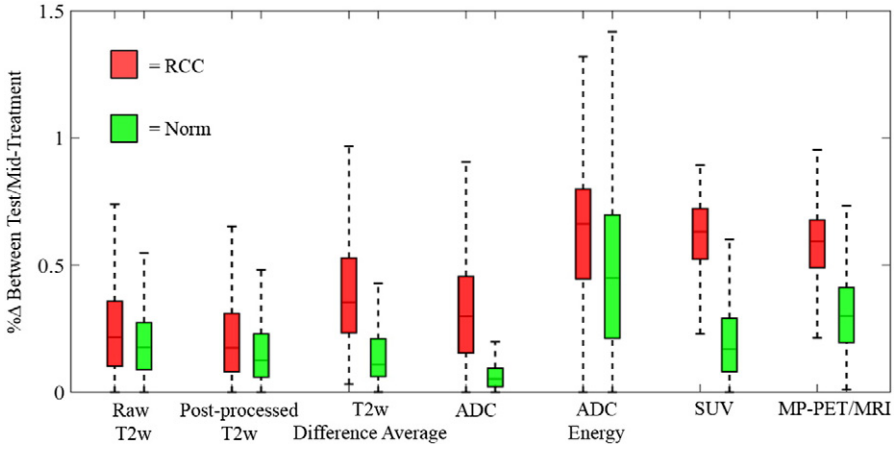

Figure 3. Box plots showing %Δ between test/midtreatment acquisitions for patient 1 in both the RCC and healthy tissue regions for top-ranked PET/MRI radiomic features (based on scoring function $S_{overall}$) as well as the original signal intensities.

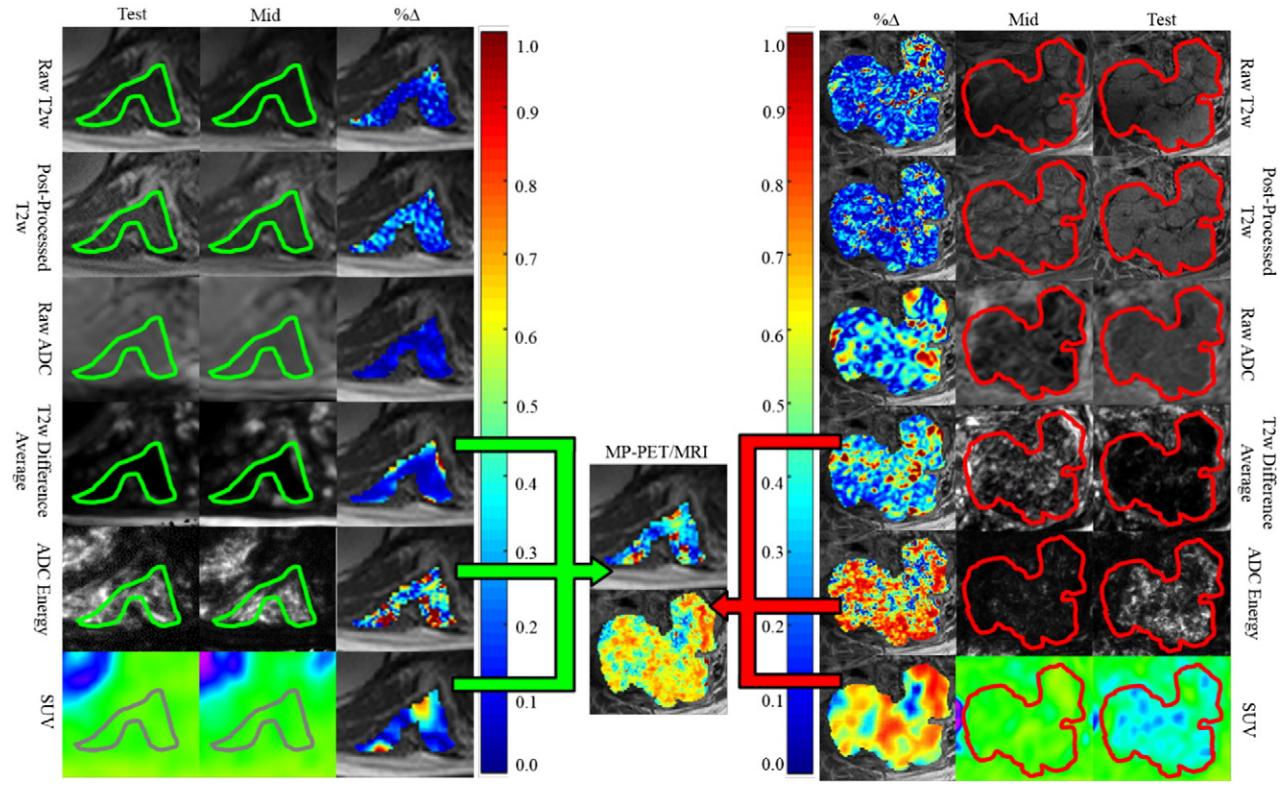

Figure 4. Percent difference maps based on treatment-related changes between corresponding voxels in test/midtreatment acquisitions for patient 1 in the (A) healthy tissue and (B) RCC regions for PET/MRI radiomic features. (C) Difference maps of SUV, ADC energy, and T2w difference average are combined based on each feature's weighted contribution factor to compute an integrated MP-PET/MRI map, reflecting a comprehensive characterization of early treatment-related changes and response.

detect treatment-related changes in the ablation zone of the prostate [16,18]. A subset of the radiomic features examined in these previous studies was used in this study to characterize treatment response in RCC based on changes in appearance before and during treatment. Because we are examining changes only 3 weeks into treatment, qualitative changes on intensity images were hard to observe. However, our initial results suggest that T2w and ADC radiomic features may be capturing subvisual changes in anatomic texture, thus quantifying changes in the microscopic structures of the tumor as a result of sunitinib treatment.

Although we have evaluated a large number of radiomic features to quantitatively characterize treatment response on FLT-PET/MRI, the goal was to determine which radiomic features best reflected treatment-related change while being resilient to acquisition-related artifacts. The results of this study could therefore help guide future studies for selecting the appropriate radiomics feature set to examine for treatment evaluation in the context of different body regions and diseases.

In summary, we have presented a first attempt of using radiomics analysis for evaluating treatment-related changes *in vivo* after early sunitinib treatment for RCC via FLT-PET/MRI. Our experimental results suggested that PET/MRI radiomic features, namely, SUV, ADC energy, and T2w difference average, might be able to identify early structural and functional response to cytostatic treatment in metastatic RCC, but our conclusions are limited because of the fact that only two studies were considered in this work. A test/retest evaluation indicated that SUV and ADC Haralick features appeared to display low variability across acquisitions for each patient. By comparison, the original T2w signal intensity and T2w radiomic features seemed inconsistent across acquisitions even after intensity standardization. T2w and ADC texture features appeared to yield an improved performance in capturing treatment-related changes compared with the original T2w and ADC signal intensities, whereas PET texture features seemed to degrade in performance. Although SUV was the highest-ranking individual feature for overall sensitivity and specificity, the integration of top-performing radiomic features yielded an even higher sensitivity by combining complementary information from the different PET/MRI protocols. As this was only a proof-of-concept study, further investigation is needed to validate our findings on a larger cohort of data. Additional studies are also needed to correlate early treatment changes with progression-free survival. Through the presented workflow, we also hope to perform radiomics analysis to identify imaging features predictive of treatment response in other cancers via PET/MRI.

Acknowledgements

This publication was made possible by the National Cancer Institute of the National Institutes of Health under award numbers R21CA167811-01, R21CA179327-01, R21CA195152-01, and U24CA199374-01; the National Institute of Diabetes and Digestive and Kidney Diseases under award number R01DK098503-02; the National Center for Advancing Translational Sciences award number UL1TR000439; the DOD Prostate Cancer Synergistic Idea Development Award (PC120857); the DOD Lung Cancer Idea Development New Investigator Award (LC130463); the DOD Prostate Cancer Idea Development Award; the Ohio Third Frontier Technology development grant; the CTSC Coulter annual pilot

grant; the Case Comprehensive Cancer Center pilot grant; the VelaSano grant from the Cleveland Clinic; and the Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering at Case Western Reserve University. Philips Healthcare provided funding for clinical PET/MR imaging. The content of this research is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] De Cecco CN, Ganeshan B, Ciolina M, Rengo M, Meinel FG, Musio D, De Felice F, Raffetto N, Tombolini V, and Laghi A (2015). Texture analysis as imaging biomarker of tumoral response to neadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. *Invest Radiol* **50**(4), 239–245.

[2] NCCN (2015). Clinical practice guidelines in oncology: kidney. *Cancer* **3**.

[3] Chen W, Cloughesy T, Kamdar N, Satyamurthy N, Bergsneider M, Liau L, Mischel P, Czernin J, Phelps ME, and Silverman DH (2005). Imaging proliferation in brain tumors with 18F-FLT PET: comparison with 18F-FDG. *J Nucl Med* **46**, 945–952.

[4] Murakami M, Zhao S, Zhao Y, Yu W, Fatema CN, Nishijima KI, Yamasaki M, Takiquchi M, Tamaki N, and Kuge Y (2013). Increased intratumoral fluorothymidine uptake levels following multikinase inhibitor sorafenib treatment in a human renal cell carcinoma xenograft model. *Oncol Lett* **6**, 667–672.

[5] Bao X, Wang M, Zhang Y, and Zhang Y (2014). Early monitoring antiangiogenesis treatment response of sunitinib in U87MG tumor xenografft by 18F-FLT MicroPET/CT Imaging. *Biomed Res Int 2014* , 1–9 218578.

[6] Padhani AR, Liu G, Mu-Koh D, Chenevert TL, Thoeny HC, Takahara T, Dzik-Jurasz A, Ross BD, Van Cauteren M, and Collins D, et al (2009). Diffusion-weighted magnetic resonance imaging as a cancer biomarker: concensus and recommendations. *Neoplasia* **11**, 102–125.

[7] Sun YS, Zhang XP, Tang L, Ji JF, Gu J, Cai Y, and Zhang XY (2010). Locally advanced rectal carcinoma treated with preoperative chemotherapy and radiation therapy: preliminary analysis of diffusion-weighted MR imaging for early detection of tumor histopathologic downstaging. *Radiology* **254**(1), 170–178.

[8] Partovi S, Dohan A, Rubbert C, Vercher-Conejero JL, Gaeta C, Yuh R, Zipp L, Herrmann KA, Robbin MR, and Lee Z, et al (2014). Clinical oncologic applications of PET/MRI: a new horizon. *Am J Nucl Med Mol Imaging* **4**, 202–212.

[9] Eisenhuuer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, and Mooney M, et al (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228–247.

[10] Kang HC, Tan K, Keefe SM, Heitjan DF, Siegelman ES, Flaherty KT, O'Dwyer PJ, and Rosen MA (2013). MRI assessment of early tumor response in metastatic renal cell carcinoma patients treated with sorafenib. *AJR Am J Roentgenol* **200**, 120–126.

[11] Wahl RL, Jacene H, Kasamon Y, and Lodge MA (2009). From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med* **50**(Suppl 1), 122S–150S.

[12] Hee K, Lee HY, Lee KS, and Kim JH (2012). Imaging-based tumor treatment response evaluation: review of conventional, new, and emerging concepts. *Korean J Radiol* **13**(4), 371–390.

[13] Zitova B and Flusser J (2003). Image registration methods: a survey. *Imag Vis Comput* **21**, 977–1000.

[14] Palumb D, Yee B, and Madabhushi A (2011). Interplay between bias field correction, intensity standardization, and noise filtering for T2-weighted MRI. *Conf Proc IEEE Eng Med Biol Soc* **2011**, 5080–5083.

[15] Lambin P, Rios-Velazquez E, Leinjenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, and Dekker A, et al (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* **48**(4), 441–446.

[16] Viswanath S, Toth R, Rusu M, Sperling D, Lepor H, Futterer J, and Madabhushi A (2014). Identifying quantitative in vivo multi-parameteric MRI features for treatment related changes after laster interstitial thermal therapy of prostate cancer. *Neurocomputing* **144**, 13–23.

[17] Viswanath SE, Bloch NB, Chappelow JC, Toth R, Rofsky NM, Genega EM, Lenkinski RE, and Madabhushi A (2012). Central gland and peripheral zone prostate tumors have significantly different quantitative signatures on 3 tesla endorectal in vivo T2-weighted MR imagery. *J Magn Reson Imaging* **36**, 213–224.

[18] Ginsburh S, Rusu M, Kurhanewicz J, and Madabhushi A (2014). Computer extracted texture features on T2w MRI to predict biochemical recurrence following radiation therapy for prostate cancer. *Proc SPIE Int Med Imag* **9035**(903509), 1–13.

[19] Litjens G, Huisman H, Elliot R, Shih NN, Feldman MD, Viswanath S, Futterer JJ, Boomers JG, and Madabhushi A (2014). Quantitative identification of magnetic resonance imaging features of prostate cancer response following laser ablation and radical prostatectomy. *J Med Imaging* **1**(3), 035001.

[20] Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, Oudard S, Negrier S, Szczylik C, and Kim ST, et al (2007). Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* **356**, 115–124.

[21] Motzer RJ, Michaelson MD, Redman BG, Hudes GR, Wilding G, Figlin RA, Ginsberg MS, Kim ST, Baum CM, and DePrimo SE (2006). Activity of SU11248, a multitargeted inhibitor of vascular endothelial growth factor receptor and platelet-derived growth factor receptor, in patients with metastatic renal cell carcinoma. *J Clin Oncol* **24**, 16–24.

[22] Ratain MJ, Eisen T, Stadler WM, Flaherty KT, Kaye SB, Rosner GL, Gore M, Desai AA, Patnaik A, and Xiong HQ, et al (2006). Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol* **24**, 2505–2512.

[23] Escudier B, Eisen T, Stadler WM, Szczylik C, Oudard S, Siebels M, Negrier S, Chevreau C, Solska E, and Desai AA, et al (2007). Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med* **356**, 125–134.

[24] Faivre S, Delbaldo C, Vera K, Robert C, Lozahic S, Lassau N, Bello C, Deprimo S, Brega N, and Massimini G, et al (2006). Safety, pharmacokinetic, and antitumor activity of SU11248, a novel oral multitarget tyrosine kinase inhibitor, in patients with cancer. *J Clin Oncol* **24**, 25–35.

[25] Chow LQ and Eckhardt SG (2007). Sunitinib: from rational design to clinical efficacy. *J Clin Oncol* **25**, 884–896.

[26] Condon BR, Patterson J, Wyper D, Jenkins A, and Hadley DM (1987). Image non-uniformity in magnetic resonance imaging: its magnitude and methods for its correction. *Br J Radiol* **60**(709), 83–87.

[27] Vovk U, Pernus F, and Likar B (2007). A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans Med Imaging* **26**(3), 405–421.

[28] Cohen M, DuBois R, and Zeineh M (2000). Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Hum Brain Mapp* **10**, 204–211.

[29] Klein S, Staring M, Murphy K, Viergever MA, and Pluim JPW (2010). elastix: a toolbox for intensity based medical image registration. *IEEE Trans Med Imaging* **29**(1), 96–205.

[30] Nyul LG, Udupa JK, and Zhang X (2000). New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging* **19**(2), 143–150.

[31] Masa-ah P and Soongsathitanon S (2010). A novel standardized uptake value (SUV) calculation of PET DICOM files using MATLAB. WSEAS Int Conf on Biomedical Electronics and Biomedical Informatics; 2010.

[32] Agner SC, Soman S, Libfeld E, McDonald M, Thomas K, Englander S, Rosen MA, Chin D, Nosher J, and Madabhushi A (2010). Textural kinetics: a novel dynamic contrast-enhanced (DCE)-MRI feature for breast lesion classification. *J Digit Imaging* **24**, 446–463.

[33] Haralick RM, Shanmugam K, and Dinstein I (1973). Textural features for image classification. *IEEE Trans Syst Man Cybern* **3**(6), 610–621.

[34] Brooks FH and Grigsby PW (2014). The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med* **55**, 37–42.

[35] Orlhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, and Buvat I (2014). Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med* **55**, 414–422.

[36] Hatt M, Majdoub M, Vallieres M, Tixier F, Le Rest CC, Groheux D, Hindie E, Martineau A, Pradier O, and Hustinx R, et al (2015). ^{18}F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi–cancer site patient cohort. *J Nucl Med* **56**(1), 38–44.

[37] Kailath T (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans Commun Technol* **15**(1), 52–60.

[38] Valls L, Hoimes C, Sher A, Hu L, Lee Z, Muzic R, and Avril N (2014). Early response monitoring of receptor tyrosine kinase inhibitor therapy in metastatic renal cell carcinoma using [F-18]fluorothymidine-positron emission tomography-magnetic resonance. *Semin Roentgenol* **49**(3), 238–241.

[39] Lopes R, Ayache A, Makni N, Puech P, Villers A, Mordon S, and Betrouni N (2010). Prostate cancer characterization on MR images using fractal features. *Med Phys* **38**, 83–95.