Title
Image Registration for Quantitative Parametric Response Mapping of Cancer Treatment Response

Journal
Translational Oncology, 7(1)

ISSN
1944-7124

Authors
Boes, Jennifer L
Hoff, Benjamin A
Hylton, Nola
et al.

Publication Date
2014-02-01

DOI
10.1593/tlo.14121

Peer reviewed

Translational Oncology
www.transonc.com

Volume 7 Number 1 February 2014 pp. 101–110 **101**

Image Registration for Quantitative Parametric Response Mapping of Cancer Treatment Response[1]

Jennifer L. Boes*, Benjamin A. Hoff*, Nola Hylton†, Martin D. Pickles‡, Lindsay W. Turnbull‡, Anne F. Schott§, Alnawaz Rehemtulla¶, Ryan Chamberlain#, Benjamin Lemasson, Thomas L. Chenevert*, Craig J. Galbán*, Charles R. Meyer* and Brian D. Ross***

*Department of Radiology, Center for Molecular Imaging, University of Michigan, Ann Arbor, MI; †Departments of Radiology and Biomedical Engineering, University of California at San Francisco, San Francisco, CA; ‡Centre for MR Investigations, Hull York Medical School, University of Hull, Hull, UK; §Department of Internal Medicine, Center for Molecular Imaging, University of Michigan, Ann Arbor, MI; ¶Department of Radiation Oncology, Center for Molecular Imaging, University of Michigan, Ann Arbor, MI; #Imbio, Minneapolis, MN; **INSERM U836, F-38042, Université Joseph Fourier, Institut des Neurosciences, UMR-S836, Grenoble, France

Abstract

Imaging biomarkers capable of early quantification of tumor response to therapy would provide an opportunity to individualize patient care. Image registration of longitudinal scans provides a method of detecting treatment-associated changes within heterogeneous tumors by monitoring alterations in the quantitative value of individual voxels over time, which is unattainable by traditional volumetric-based histogram methods. The concepts involved in the use of image registration for tracking and quantifying breast cancer treatment response using parametric response mapping (PRM), a voxel-based analysis of diffusion-weighted magnetic resonance imaging (DW-MRI) scans, are presented. Application of PRM to breast tumor response detection is described, wherein robust registration solutions for tracking small changes in water diffusivity in breast tumors during therapy are required. Methodologies that employ simulations are presented for measuring expected statistical accuracy of PRM for response assessment. Test-retest clinical scans are used to yield estimates of system noise to indicate significant changes in voxel-based changes in water diffusivity. Overall, registration-based PRM image analysis provides significant opportunities for voxel-based image analysis to provide the required accuracy for early assessment of response to treatment in breast cancer patients receiving neoadjuvant chemotherapy.

Translational Oncology (2014) 7, 101–110

Address all correspondence to: Brian D. Ross, PhD, Department of Radiology, Center for Molecular Imaging, University of Michigan, 109 Zina Pitcher Place, Ann Arbor, MI 48109-2200. E-mail: bdross@umich.edu

[1]Funding was provided by US National Institutes of Health research grants U01CA166104, P01CA087634, P01CA085878, U01CA032102, U01CA151235, and R01CA136892. J.L.B. is a recipient of support from the US National Institutes of Health training grant T32EB005172. Support was also received from Yorkshire Cancer Research to establish and fund the Hull MRI Centre. B.D.R., T.L.C., A.F.S., A.R., and C.J.G. have patents and disclosures to the University of Michigan related to the underlying technologies described in this report that have been licensed to Imbio, LLC, a company in which B.D.R. and A.R. have a financial interest.
Received 10 January 2014; Revised 17 February 2014; Accepted 17 February 2014

DOI 10.1593/tlo.14121

Introduction

Advances in functional imaging along with molecularly targeted probes and contrast agents are providing feature-rich data sets from multiple sources before and following treatment that allow for greater detail in the characterization and interrogation of the cancer disease process [1]. However, in spite of the increased toolbox of techniques for evaluating a patient's tumor response, the clinical standard assessment tool for measuring tumor treatment response is the Response Evaluation Criteria in Solid Tumors (RECIST) [2]. The RECIST clinical response metric is extensively used on anatomic tumor images acquired by computed tomography (CT) and magnetic resonance imaging (MRI) scans wherein linear measurements of the longest diameter of a tumor (or sum from multiple tumors) are obtained to compute an objective endpoint. Improvements in patient-specific imaging-based response metrics that allow for earlier detection of overall cancer response will require sensitive readouts that can be quantified and ultimately correlated with RECIST clinical trial outcome measures [3].

While the use of imaging to anatomically characterize tumor dimensions has been fundamental to the integration of imaging within clinical trial design and routine clinical practice, alternative MR image–derived functional tumor response metrics that may provide for more sensitive and thus earlier response readouts have emerged. These metrics include, for example, hemodynamic [dynamic susceptibility contrast (DSC)-MRI and dynamic contrast enhanced (DCE)-MRI] and water diffusion-weighted magnetic resonance imaging (DW-MRI) readouts of overall tumor response [4–6]. Response assessment using these scan modalities is commonly done through obtaining image-derived summary metrics that are extracted from whole tumor volume-of-interest (VOI) contours and histogram analysis of the quantitative voxel values contained within the target lesion (Figure 1) [7]. Histogram analysis allows for quantification of the image VOI using, for example, the mean value where tumor response is quantified by the magnitude of change from baseline following treatment. While whole-tumor metrics can be robust measures when tumor changes are relatively homogeneous, these VOI-based measures may be less sensitive for detecting treatment-associated changes in longitudinally acquired scans in the presence of intratumor heterogeneity where local changes are spatially varying within the mass. As shown in Figure 1, changes in tumor histogram MR metrics can be minimal (Figure 1*A*) and uniformly shifted (Figure 1*B*) with a corresponding net shift in the mean histogram values or have similar numbers of voxels with increased and decreased values, thus nullifying the overall histogram shift (Figure 1*C*). However, spatial alignment of interval images using image registration algorithms provides an opportunity to use the spatial information contained within the images to track quantitative changes in individual voxels. Individual voxels with paired, temporally resolved, quantitative values generate a joint density histogram where the axes are quantitative values for the baseline scan (*x*-axis) *versus* the follow-up scan (*y*-axis). Application of a statistically derived cutoff of significant change [95% confidence interval (CI)] provides for classification of voxels allowing for a voxel-based method capable of tracking changes in the underlying imaging metric over time (Figure 1, *D*–*F*). Figure 1 shows simulated histograms of water diffusivity, quantified as the apparent diffusion coefficient (ADC), within the tumor VOI along with corresponding

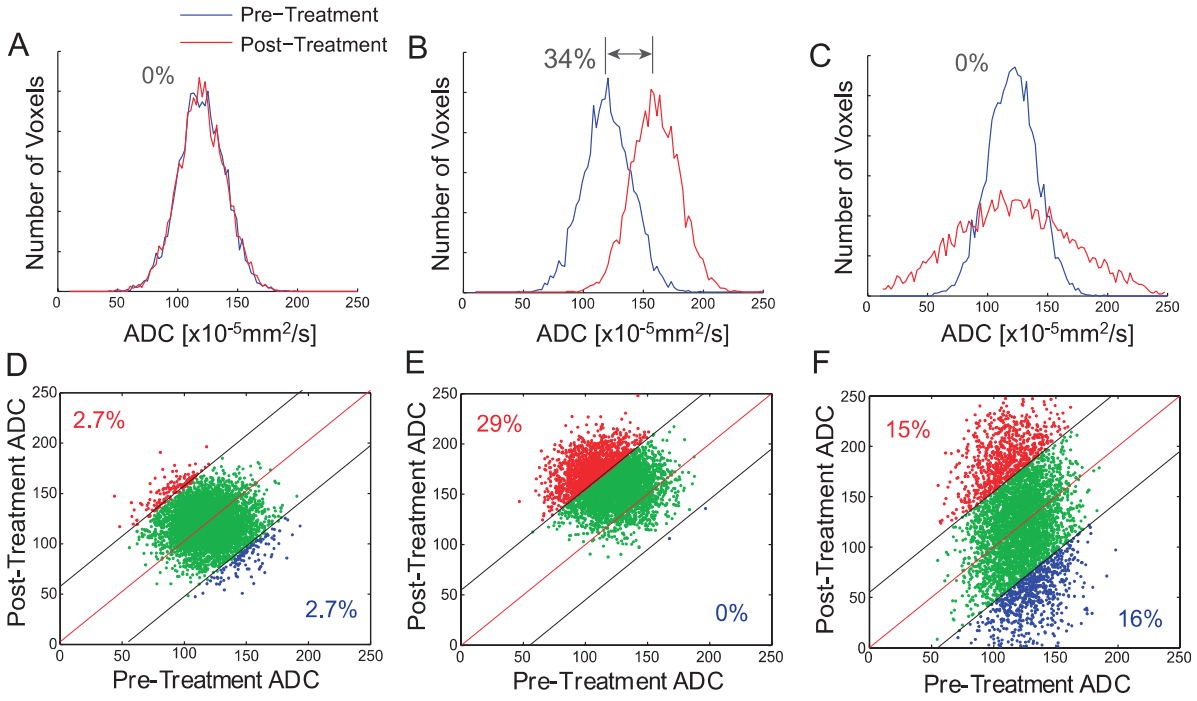

Figure 1. A simulated comparison of whole-tumor histogram analysis (top row; the blue line is the pretreatment tumor data and the red line is the posttreatment tumor data) *versus* the corresponding voxel-based analysis using a joint density histogram (bottom row). Histograms from tumors with (A) no major change, (B) significant uniform shift to higher ADC values with a 34% net mean change, and (C) increased and decreased ADC values resulting in no net detectable histogram shift. Parametric response maps from the corresponding histograms above where (D) the CI for detection of change was set to 95%, thus no significant change in red voxels (increased values) or blue voxels (decreased values) was detected. (E) An increase in the number of red voxels was detected at 29% of the total tumor voxels. (F) Both an increase and a decrease in tumor voxels of approximately 15% were detected, whereas no major shift was detected using a histogram analysis of the same data (C).

simulated joint density plots for the tumors in which pretreatment images have been registered with their corresponding posttreatment images such that each measured ADC voxel at baseline has a paired value at follow-up. The capability of applying a voxel-based analysis to "unmix" the complex biologic responses of individual tumors to treatment will improve statistical accuracy in detecting tumor changes and furthermore will provide the ability to spatially assess treatment response regionally within the lesion.

The initial application of voxel-by-voxel analysis was reported as the functional diffusion map (fDM) that was developed as a statistical approach for segmenting brain tumor response based on a defined threshold of ADC change following therapy [8]. More recently, the voxel-based method for processing longitudinally acquired MRI scans was further generalized to multimodal applications including DSC-MRI wherein it was shown to provide enhanced sensitivity for early cancer treatment response over VOI-based metrics [9]. While early application of the voxel-based methods was termed fDM as it was applied to DW-MRI scans [8,10–12], more recently it has been referred to as parametric response mapping (PRM) as it can be generalized and applied to multiple imaging modalities [MRI, positron emission tomography (PET), CT, etc.] [9,13–15].

DW-MRI can be applied to patients with breast cancer in an effort to derive an early treatment response biomarker. For voxel-based PRM analysis of ADC maps (PRM_{ADC}) to be applied, registration of breast MR images and ADC maps must be accomplished. The motivation for this research effort lies in the fact that patients who elect to undergo neoadjuvant chemotherapy (NAC) do so with the intent purpose of down-staging their disease before surgical resection [16–18]. However, patients and physicians can be hesitant to use NAC as it is not 100% effective, and since treatment response is determined late or on completion of therapy, patients may not receive timely effective intervention [19–21]. Therefore, immediate surgical resection remains the standard approach, even in patients who are known to be candidates for chemotherapy postoperatively. Having an early imaging treatment response biomarker available would make NAC more compelling and would improve the overall management of this patient population. In an effort to advance this concept, analytical algorithms for routine execution of PRM_{ADC} (e.g., image registration) must ultimately be developed and tested in a multicenter setting to validate PRM_{ADC} as an early quantitative imaging biomarker for breast cancer patients undergoing NAC.

The overarching goal of the current report is to advance the PRM imaging biomarker through consideration of interrelated issues and procedures required to develop, validate, and implement voxel-based imaging software applications for integrating the PRM biomarker within the clinical environment. Considerations will address the types of digital image registration available, the overall concepts of how image registration works, the accuracy of image registration, the procedures for proper alignment of images, and finally determination of thresholds for detection of significant change along with visualization of results and quantification of summary response metrics.

Methods

Image Registration for Response Detection

Voxel-based analysis is predicated on the ability to align tumor volumes acquired between interval exams such that homologous voxels of the different temporal instances contain approximately the same partial volumes of tissue. In the past 25 years, image registration in biomedical applications has become increasingly used [22–25]. While perhaps the most common use of registration has been in the field of radiation treatment planning [26], registration methods have been slower to move into general clinical practice. In our National Institutes of Health (NIH)-sponsored quantitative imaging network (QIN) research effort, we are developing the use of deformable registration solutions for providing early imaging biomarker readouts of breast cancer treatment response [27].

Selecting and Implementing Registration Algorithms

Due to the inherent complexity of the problems encountered in the field of image registration, registration methods are selected to suit the body part and the types of images to be registered, requiring a cost (or objective) function, deformation function, optimization scheme, and in many cases a smoothness penalty to control for too many degrees of freedom in the chosen deformation function such as cubic B-splines. A wealth of registration options, cost metrics, and deformation functions are available (Table 1) and are summarized in the literature [24,28]. Transforms can be grouped into those derived from physical models, interpolation theory, or knowledge-based. Cost/objective metrics can be described as iconic, based on voxel values, or geometric using other image features. For each transformation solution, there is the option of applying additional task-specific constraints to insure that the physical domain is modeled and the registration solution can accommodate anticipated situations, such as smoothness of the deformation, tumor growth, tumor infiltration, or possibly slip boundaries. The specific selection from

Table 1. Components of the Registration Process.

Warping Deformation Transforms for Registration	
Physical transform models	Elastic body transformations
	Viscous flow models
	Diffusion models include demon algorithms
	Curvature registration
	Flows of diffeomorphisms
Geometric transform from interpolation theory	Radial basis functions: TPSs, Gaussian, Wendland, Wu
	Elastic body splines
	Free-form deformations include cubic B-splines
	Basis functions from signal representation include Fourier basis function, wavelets, Riesz
	Locally affine
Knowledge-based	Statistically constrained geometric transforms include deformation atlases
	Geometric transforms inspired by biomechanical/biophysical models
Criteria Driving Registration (Metric or Cost Function)	
Iconic metrics	Difference type metrics (fast, assumes identical relationship between voxels or measures), sum of squared difference, attribute vectors
	Information theoretic models (single or multimodality) include mutual information, entropy
	Correlation coefficient type metrics (linear or reduction to linear relationship between corresponding voxels)
Geometric methods	Detecting points/regions of interest and mapping: Laplacian of Gaussian, Lowe's scale invariant feature, exact/inexact landmark mapping
	Feature mapping using other methods: transform invariant features matched using thresholds or graph matching
Hybrid methods	Geometric information as initialization
	Geometric information as constraint
	Coupled approaches (unified iconic and geometric constraints)

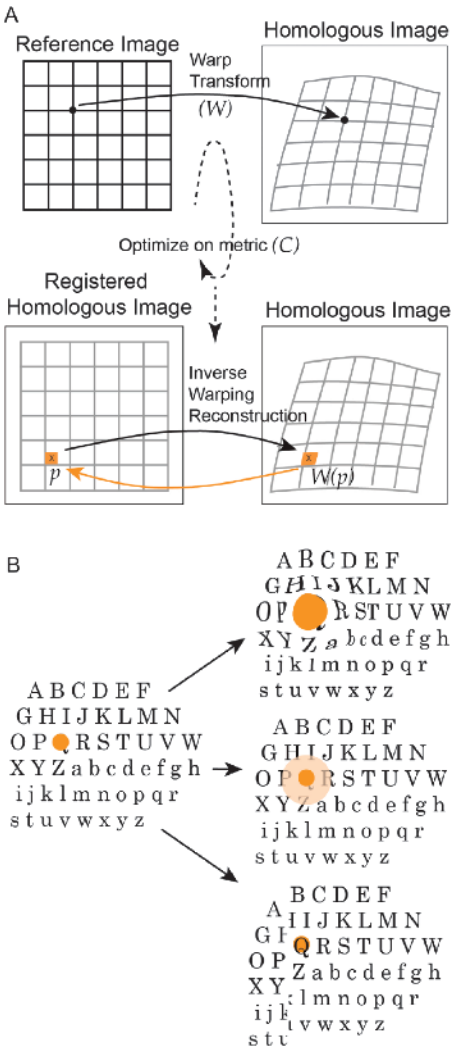

Figure 2. (A) A warp transform defines the relationship between a homologous and a reference image, where reference is the baseline image and the homologous is a later time point or different modality. To create an unwarped version of the homologous image, the location of the pixel p is found in the homologous image as $W(p)$ and its value (orange) is inversely mapped back to create a registered homologous image. (B) The warp transform W should be able accommodate expected deformations including growth with displacement of tissue (top) and infiltration into surrounding tissue (middle). Another deformation not typically found in breast tissue is a slip boundary (bottom).

the possible combinations of choices is determined by the specific task, accuracy, and computational speed required to reach the final solution.

The process is graphically shown in Figure 2. Here, two separate images (Figure 2*A*) are identified where one (homologous image) is spatially aligned to another (reference image). The extent of deformation required to spatially align the two images is presented by the grid (Figure 2*A*, *top row*). To solve the registration problem, determination of the warping transform W that maps each point in the reference to its homologue in the target space is accomplished. After finding the transform, the homologous image is reconstructed into the reference space using an inverse warping reconstruction, $W(p)$, yielding two matched images (Figure 2*A*, *bottom row*). When a voxel maps into a space between several voxels, an interpolation method is applied to find the proper value. Interpolation methods range from picking the nearest neighbor to using an interpolation method between the closest voxels, usually selected to match the type of data being mapped, and the appropriate selection of values to avoid creation of spurious values or altering a property of the data set, such as mass. Registrations optimize W based on a cost function, C (Figure 2*A*), to yield two data sets of the same dimensions, i.e., the same coordinate space, so that voxels in the same location represent the same element in space. Overall, the warping deformation transform (W) must accommodate anticipated deformations such as tumor growth with associated tissue displacement (Figure 2*B*, *top*), changes involving infiltration into surrounding tissue (Figure 2*B*, *middle*), and finally while not common in tumors, a deformation involving a slip boundary (Figure 2*B*, *bottom*).

Many different approaches can be implemented to solve breast registrations, with the caveat that assumptions should be chosen carefully and algorithm parameters tuned to the solution of the problem being solved. Rueckert's early deformable multimodal breast registration method used an affine transform followed by geometric free-form deformations interpolated with B-splines and optimized using normalized mutual information along with a smoothness penalty to register contrast-enhanced MRI of the breast [29]. Biomechanical models of the breast used finite-element modeling to corroborate Rueckert's method [30]. The algorithm was modified to add an incompressibility penalty to handle contrast changes in soft tissues [31]. More recent approaches use modality-independent feature measures such as Gabor features [32]. Optical flow with brightness shift term to account for contrast concentration has also been used and corroborated using synthetic phantoms with expected biomechanical properties [33,34]. An additional tumor-focused registration approach using a radial basis function transform with a tunable global elasticity parameter and tumor-specific stiffness constraint driven by mutual information has also been reported [35] along with a Laplacian physical model of diffusion using a finite element framework [36]. Emerging registration solutions are focused on joint segmentation and registration [37], incorporating information priors, expanded use of modality-independent metrics, and speed through parallelism or other computational efficiencies [38].

Registration Approach Developed for Breast Cancer PRM Analysis

In this paper, we have developed and implemented breast tumor registration thin-plate splines (TPSs), a geometric transform from interpolation theory that, while computationally costly, has the benefit of incorporating least-bending energy smoothness into its formulation without penalizing the cost/objective function. Furthermore, we incorporated mutual information as the cost function to register anatomic images to DW-MRI–derived ADC maps. Because of variable deformity of tumor and breast tissue, registration was achieved locally around the tumor by dilating the tumor's VOI by a few voxels (three to five) to allow the algorithm to "see" and match the border of the tumor. A Nelder-Mead simplex algorithm was used to optimize the solution. Although found to be computationally more expensive than alternatives, this algorithm has been shown to generate robust solutions [39]. Preservation of topology was assured by removal of transform control points that generate folding as measured by the presence of a negative Jacobian, i.e., folding. This approach supports the use of TPS with spatially varying degrees of freedom (DOF; using control point density) initialized to that supported by the maximum local

mutual information (MI) density but able to adapt to lower local MI that only supports lower DOFs through the automatic local removal of control points in such regions. A cubic B-spline approach would mirror this approach by spatially adapting the smoothness penalty term. Overall, this registration method provides the benefit of a maximally smooth radial basis function transform of varying local DOF optimized by a voxel-based metric.

Clinical Trial Data

These studies had institutional review board approval, and informed consent was obtained from all patients. The University of Michigan QIN research effort relies on the use of image data obtained from several multicenter prospective clinical trials including a Cancer Research UK–sponsored trial (Neo COMICE) entitled "Establishing the Efficacy of Advanced Semi-automated Functional MR Imaging in the Early Prediction of Response of Locally Advanced Breast Cancer to NAC", the USA-based I-Spy 2 trial entitled "An adaptive breast cancer trial design in the setting of NAC", and SWOG-S0800 substudy trial focused on diffusion-weighted imaging (DWI) as well as data collected at the University of Michigan. The SWOG-S0800 trial enrolled women with Her-2 negative inflammatory breast carcinoma or Her-2 negative locally advanced breast carcinoma. In addition, patients who were accrued as part of the SWOG-S0800 underwent test-retest examinations to provide image data required for evaluating the overall accuracy or variability of the PRM approach. Data presented in this current research effort were derived from the Neo COMICE and SWOG-S0800 trials.

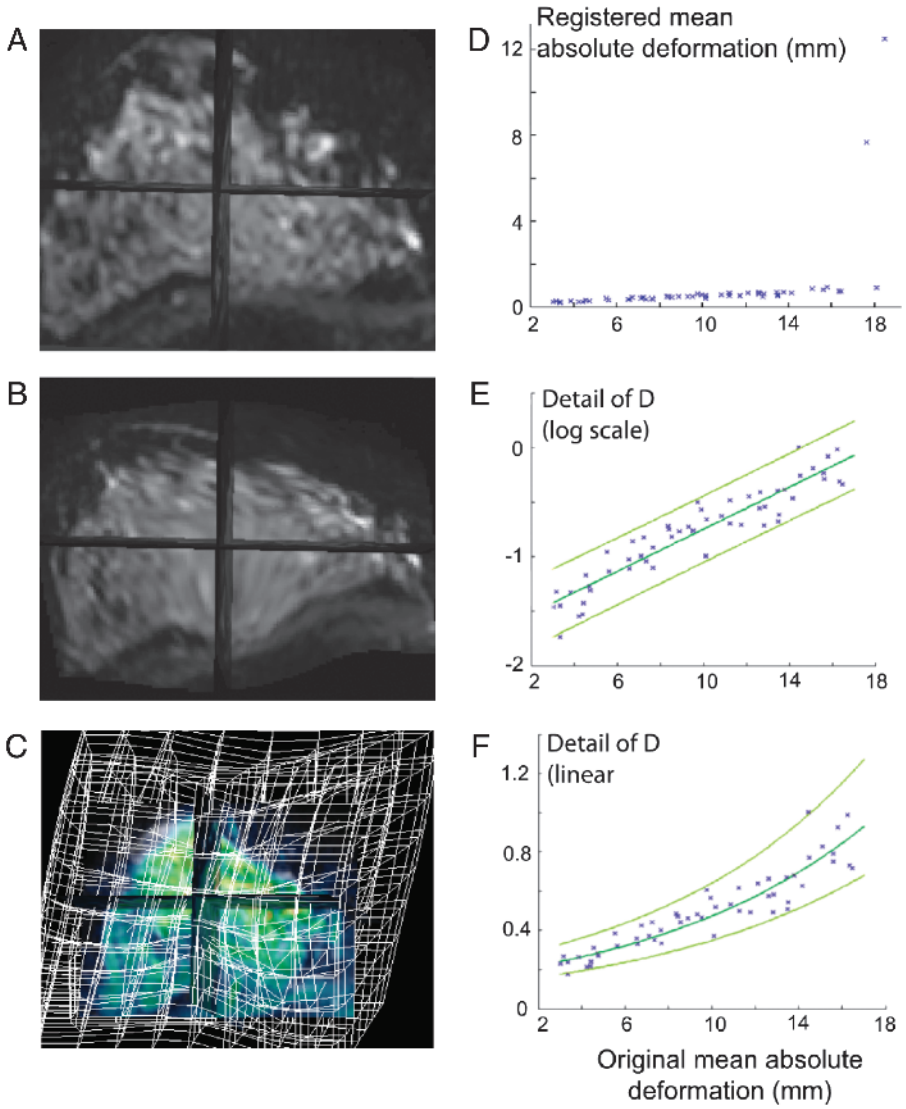

Figure 3. Image registration capture range and accuracy can be determined using MR breast images. A breast DW-MRI image (low-gradient "*b* value" image) is (A) deformed using a 3 × 3 × 3 uniform grid to form a synthetic test data set (B) where the deformation is visualized by the grid in (C) with a mean absolute deformation of 17.6 mm. The deformation was repeated with 70 different synthetic data sets with a range of mean absolute deformations. Applying our registration algorithm to these pairs with a known solution resulted in (D) that plots the original deformation and registered mean deformation (0 is perfect registration). The 95% CI in (E) is plotted in linear coordinates in (F) showing good repeatability with accuracy of <0.2 mm at 4 mm of original mean deformation and <0.5 mm for values of original mean deformation of up to 16 mm. Note that mean deformations of 4 mm indicate deformation significantly higher than that value in some regions.

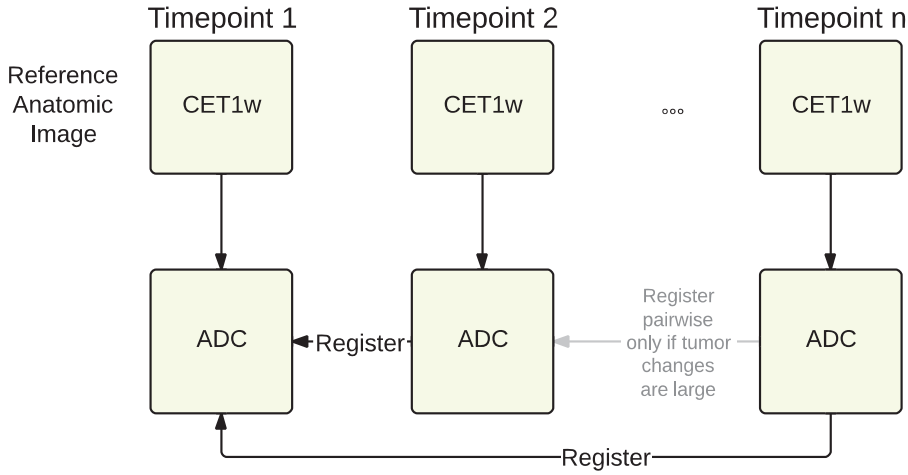

Figure 4. Registration of ADC maps over time requires a two-step process where ADC images are initially mapped to post-contrast anatomic scans within each time point to handle motion and susceptibility artifacts. Anatomic images are then registered, and the resulting transform was applied to ADC data. Registrations are computed locally in the tumor and surrounding region.

For the SWOG-S0800 trial, MR data was acquired on Achieva and Ingenia 3T Philips MRI systems using 7-channel and 16-channel dedicated breast coils. Nominal sequence parameters were as follows: three-axis, diffusion-weighted, single-shot single spin-echo, echo-planar imaging; TR = 4000 to 5000 ms; TE minimum = 55 to 80 ms; SENSE = 2 or 3; b values = 0, 100, 600 s/mm^2; bilateral axial scans with FOV of 35 to 40 cm adjusted to patient size; SPAIR fat suppression; slice thickness = 4 mm; acquisition matrix of 160 to 184; number of averages adjusted for a 4- to 5-minute scan. The diffusion maps were not corrected for gradient nonlinearity.

Results

Assessment of Registration Accuracy

An important step in validating a response assessment tool based on registration is to analyze and optimize the registration algorithm used. This step can be difficult as the same cost functions used to register images should not be used to assess the registration. A better measure of registration quality would be independent of the registration solution and relevant to the image pair under consideration. For example, difference maps might illustrate registration accuracy if the images are from the same modality but would not be appropriate to evaluate an anatomic MR image with a co-registered contrast-enhanced scan due to expected signal changes from the presence of contrast media that might be interpreted as misregistration. Evaluation of accuracy remains an ongoing research topic; however, simulations can yield valuable insights into the overall registration performance.

Accuracy, repeatability, capture range, and risk of nonconvergence can be measured using synthetic data (phantom, computer-simulated) to determine how accurately and consistently the method recovers a known solution. Such tests generate accuracy results measured at the voxel or subvoxel level, often as a mean and SD from a set of known matching locations. To measure our breast registration effectiveness, we generated a set of 70 synthetic phantoms of known, artificially deformed breast images using a 3 × 3 × 2 grid of features randomly perturbed to drive a TPS warp of the data set with no folding using SD values between 3 and 15 mm resulting in a spectrum of mean absolute deformations between 2 and 24 mm. Figure 3*A* shows the original data set (from an isotropic diffusion MR acquisition), Figure 3*B* shows one simulated phantom, and Figure 3*C* shows the resulting deformation grid. Our registration algorithm was then used to compute an optimized registration to "unwarp" each synthetic phantom to match the original image using a 4 × 4 × 2 grid, i.e., an intentionally different number and loci of control points than those used to create the synthetic phantoms. Because the original 3 × 3 × 2 grid locations are known for each phantom data set, we could back-compute the locations in each "unwarped" phantom to compare locations and measure the accuracy in mean and SD (measured in mm). The mean absolute deformation was then computed over the original and "unwarped" 3 × 3 × 2 locations across the 70 data sets. Registration capture to subvoxel accuracy was accomplished consistently for phantoms with mean absolute deformations less than 17 mm (Figure 3*D*). The detail view of Figure 3*D* using a log scale in Figure 3*E* illustrates the homoscedastic 95% CI for registration. The 95% confidence limits (definition of repeatability) were determined to be [+0.1, −0.7] or less than 0.2 mm at 4 mm original mean deformation and [+0.31, −0.2] or less than 0.5 mm for values of original mean deformations up to 16 mm. End of the robust automatic capture range is noted at the threshold of 17 mm mean absolute deformation obtained using a local maximum deformation of 46.51 mm at one of the 3 × 3 × 2 initial control points.

Voxel-Based PRM

To accomplish PRM analysis of DW-MRI breast ADC maps, multiple data sets must be co-registered across one or more time points to apply voxel-based methods for determining change in diffusion values as a response metric. PRM is based on co-registration of voxel pairs (or vectors if more than one modality is used) that can be visualized using a joint density histogram for a region of interest, in our application a tumor VOI. This distribution contains joint information about the changes. Measures of detectable change can thus be computed as well as displayed as a color overlay of the spatially varying changes over the anatomic image.

To identify response in breast tumor ADC maps that are measured at several points in a treatment cycle, co-registration of ADC maps

must be accomplished for subsequent PRM analysis (Figure 4). The approach undertakes an initial registration of the ADC map with the corresponding anatomic image to correct for any misalignment during MRI reconstruction due to susceptibility artifacts and patient motion. Successive ADC maps are registered to the baseline ADC. The result is a set of registered ADC images from which PRM metrics can be computed for analysis of treatment-induced change. When changes are minimal, as is the case here when imaging over short intervals, multiple images can be co-registered into the same space. However, when larger changes in tumor dimensions are expected, images may need to be registered pairwise.

Detection of ADC changes in the tumor requires a threshold that defines a significant change in ADC beyond system noise and misregistration. Our approach used a test-retest procedure where a patient was scanned twice within a short interval of time ($t < 30$ min), having been removed from the scanner between scans. Shown in Figure 5*A* is a representative example of a test-retest exam of a patient with breast cancer obtained from the SWOG-S0800 clinical trial where T2w images from the patient that were acquired before treatment initiation are shown. The tumor is evident within the VOI. Following registration of the two ADC maps, PRM analysis is accomplished using a joint density plot (Figure 5*B*) that provides a representation of the overall variability of the measurement. A perfect solution obtained in the absence of noise would yield a result wherein all voxel pairs would have the same value and would converge on the diagonal line of equality (dashed line in Figure 5*B*) in the joint density plot of

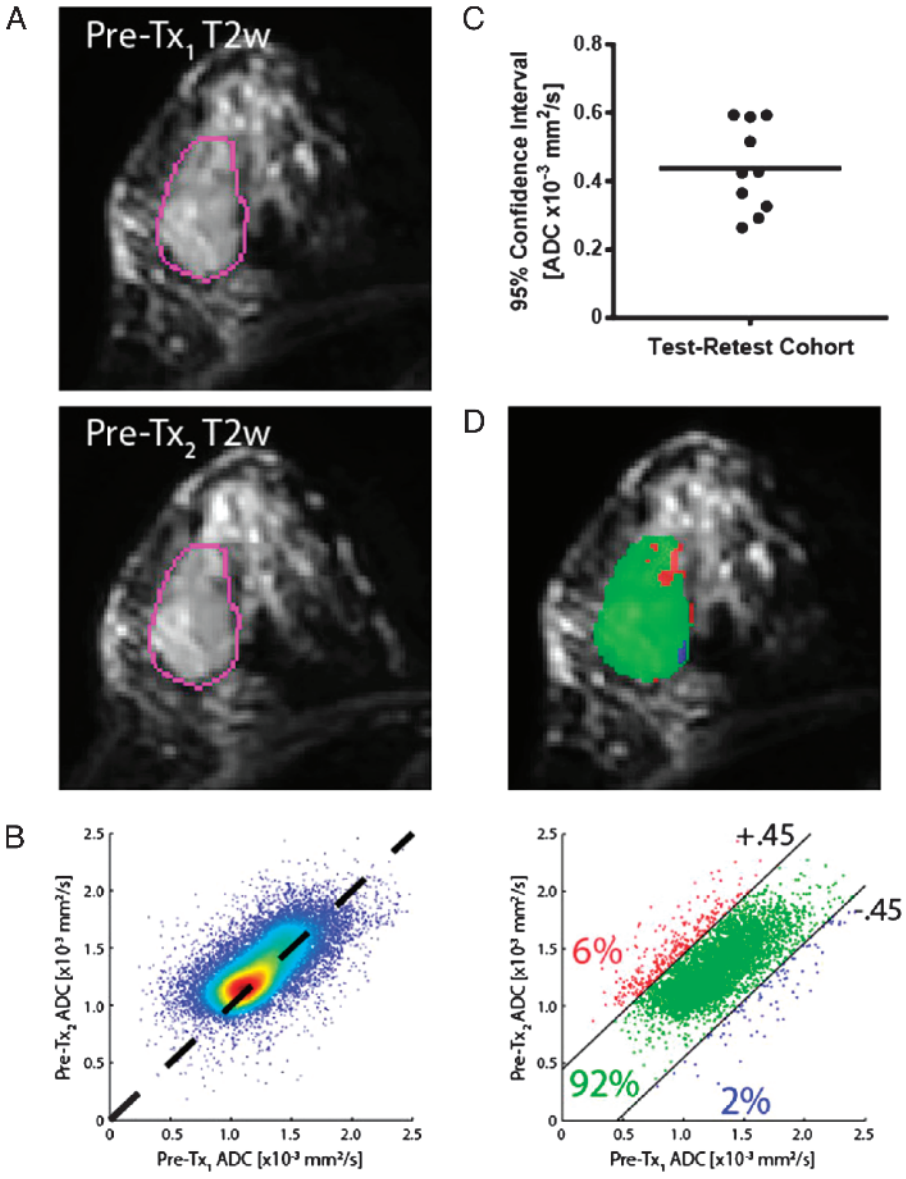

Figure 5. Joint density histogram error analysis of registered tumor ADC values using test-retest MR scans. (A) A set of test-retest T2w images of a breast cancer patient with the tumor delineated by a VOI overlay. (B) The joint density plot of the two exams reveals the distribution of differences between the exams about the dashed line of unity. (C) A cohort of 10 SWOG-S0800 patients who underwent test-retest examination followed by image registration and determination of the 95% CI is each shown as an individual point on the plot. An average CI of 0.45 $\times$ 10^{-3} mm^2/s was found and (D) applied to a joint density histogram analysis (PRM) of tumor ADC maps from the same patient shown in (A) using the pretreatment data as the baseline (abscissa) plotted against the posttreatment ADC values of the tumor (ordinate). The values for each PRM classification are shown.

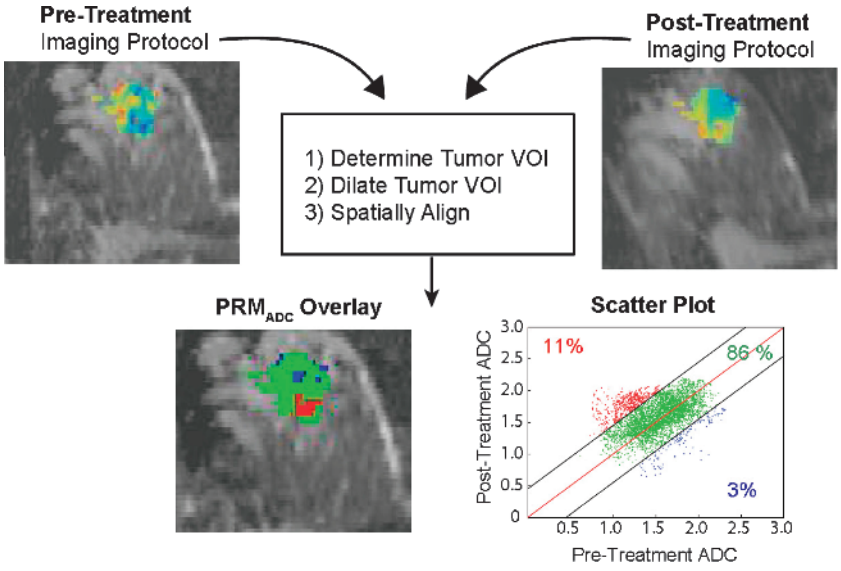

Figure 6. The overall process of quantitative PRM analysis of tumor treatment response. Two sets of image acquisitions are undertaken on an individual patient, pretreatment (baseline) and at some time interval posttreatment initiation. A VOI is drawn defining the tumor boundaries on the images that are subsequently dilated by the software to include sufficient normal surrounding parenchymal tissue to provide additional information to allow for proper image registration of the two data sets. The pretreatment data set is registered on a voxel-by-voxel basis with a posttreatment initiation data set. Application of suitable CIs for detection of significant change between the interval exams provides for quantification of ADC changes within the tumor. These can be displayed as a joint density plot with regions color coded to identify the changes when mapped back to the anatomic MR image for spatial visualization.

the two scans. In the real world, principle component analysis can be used on the distribution to identify the major axis of correspondence for the test-retest, and a CI for change can be computed using a percent measure (generally 95%). Determination of the 95% CI can be accomplished and values across the test-retest group of patients can be used to generate an average CI for future patients who do not have test-retest data available (Figure 5*C*). We determined that the average 95% CI for breast cancer patient test-retest data sets using the SWOG-S0800 data (n = 10 patients) was ±0.45 ADC units. When comparing a follow-up scan with a baseline scan from the same patient, the average CI limit can be applied to classify individual voxels based on increasing (red voxels, PRM_{ADC+}), decreasing (blue voxels, PRM_{ADC-}), and unchanged (green voxels, PRM_{ADC0}) ADC values (Figure 5*D*). PRM measures are therefore presented as relative percent volumes by summing all voxels within a class and normalizing by the total VOI volume (number of voxels). In Figure 5*D*, the color PRM overlay on the T2w image reveals the spatial distribution of PRM classifications within the individual tumor slice. In this test-retest example using the average cutoff for significant change derived from the average of 10 patients (±0.45 ADC units), values of PRM_{ADC+}, PRM_{ADC-}, and PRM_{ADC0} were 6%, 2%, and 92%, respectively.

The test-retest scheme may not be feasible within a typical clinical scanning environment (e.g., patients may not be willing to undergo the additional examination effort, time is not available for repeat scans, etc.), and using an average threshold for significant change that is derived from a pool of test-retest data may provide the most expeditious approach for determination of an overall CI for a modality. A pool of test-retest results can therefore be used to determine a globally applicable CI. Shown in Figure 6 is an example where the CI was determined using a training set of test-retest images derived from SWOG-S0800 patients and applied to Neo COMICE–derived registered patient scans. In this example, it is clearly evident that a region of large increase in ADC values could be identified within the tumor mass indicating a positive treatment response in that region (PRM_{ADC+} = 11%). As shown in Figure 6, the voxel-based PRM methods provide a value for increased ADC within the tumor (PRM_{ADC+}) that can be used as a metric for assessing treatment response but also provide for spatially identifying the regions of change within the three-dimensional tumor mass.

Discussion

We have applied fDM and PRM in multiple research domains, developing the method in animal tumor models [40–42] and translating it into human applications including brain cancer, head and neck cancer, metastatic cancer to the bone, and primary breast tumors [8–10,15,39,43,44]. The challenge in each case is to validate all components of the method: imaging, registration, and response mapping for change detection. The test of this or any other alternative measure of tumor response is its applicability across modalities and imaging centers in a repeatable and robust manner. Consistent imaging protocols that provide repeatable, quantitative readouts are important for applying PRM across clinical settings. As part of the QIN effort, we have developed MRI phantoms and gradient nonlinearity corrections aimed at improving the quantification of DW-MRI [45,46] to optimize quantitative PRM_{ADC} metrics for accuracy in detection of treatment-associated changes. Such quality control measures along with careful consideration of possible artifacts associated with acquisition protocols and post-processing algorithms will provide a solid foundation for advancing voxel-based imaging procedures at multiple institutions [7]. Calibration must occur regularly and imaging protocols should be rigorously adhered to for optimization of the PRM imaging biomarker.

PRM tumor response analysis requires robust image registration in order for quantification of change to be made on a voxel-by-voxel

basis. One of the challenges is the difficulty in obtaining test-retest data for imaging metric validation due to a more complicated study design and overall reduction in study compliance between clinical sites. Additional methods for analyzing registration algorithms using real data are possible wherein correspondence points have been identified manually. The challenge with this scenario is that user-identified points also have a noise spectrum, sometimes large. However, such analysis can still yield valuable, quantifiable insight into the power of registration solutions [47]. Shared human image data sets are also becoming more readily available with projects like the QIN. The emerging paradigm of shared research image databases provides a valuable paradigm for making test data accessible and sharing new computational approaches to problems. Overall, routine clinical use of the PRM biomarker response metric will likely require solid correlation with clinical outcome metrics such as histopathology for residual disease assessment.

Change detection using PRM is quantitative and the level of variation that indicates change must be verified for each scenario, particularly when measuring interval change in the absence of personalized test-retest data. Establishment of statistical cutoff values that indicate a detectable change in quantitative value can be achieved using two different approaches that evaluate either contralateral tissue pretreatment and posttreatment initiation or tumor in a test-retest approach that consists of acquiring duplicate scans over a very short time interval. Both approaches allow for the variation of quantitative values in the absence of changes in tissue/tumor anatomy and physiology to be evaluated, which arise from the combined noise of the imaging system and registration algorithm. Although the test-retest provides direct assessment of noise from the tumor, it is more challenging for imaging modalities that require contrast injection and/or radiation [e.g., CT, PET, single-photon emission computerized tomography (SPECT), and DCE-MRI]. The use of normal contralateral tissue as a normalization standard on an individual patient can be applied to assess the variability of the imaging metric under investigation [8]. However, care would be needed to ensure that any contralateral or comparison tissue was not changed physically, anatomically, or therapeutically, which may result in an overestimation of the threshold, thus reducing PRM's accuracy in detecting therapeutic change in tumors. From an end-point perspective, scanner variation could yield different distributions of values for the same region (a serious challenge in CT and also MR) that needs to be accounted for in any change detection thresholds. To go one step further and incorporate multimodality fusion for breast imaging and facilitate the use of, for example, PET imaging will require continued attention to each of the elements of biomarker development to ensure similar robustness [48].

The PRM cutoff has an effect on the sensitivity of the technique and has been evaluated either by testing the predictive potential of PRM using various cutoffs or deriving a more complex Bayesian model for defining the cutoff [8,9]. Although the predictive potential of PRM was improved for these cases, the improvement gained was minor compared to the simple derived cutoff from the 95% CI. As the cutoff is integral for PRM methodology, it will have to be determined for each unique case that may include tumor type, tumor location, and imaging modality. From our experience, calculation of the 95% CI provides a simplified approach to determine cutoffs for changes in quantitative values beyond system and registration noise.

Finally, to test any alternative response marker requires the availability of a near production quality method for processing data sets. The University of Michigan QIN research team has partnered with Imbio, LLC, a company that is focused on development and migration of PRM to a web-based platform to provide ease of access to automatic processing and report generation. The Imbio cloud–based approach will implement robust software tools with minimal user interaction and automated quality control to ensure fidelity of analysis. Development of this tool would make PRM accessible across multiple sites, thus providing a robust image registration algorithm for PRM analysis as a shareable tool to the overall research community.

Conclusion

Voxel-based image analysis required for following heterogeneous tumor response to therapy can be accomplished through access to high-quality MR data and use of appropriate image registration algorithms with careful selection of thresholds for determination of significant treatment-induced change along with appropriate quality controls. This paper presented approaches that can be used to develop and validate PRM-based methods as early response metrics for prediction and stratification of breast cancer responders from nonresponders using DW-MRI scans. The overarching goal of providing quantitative imaging biomarkers that will allow oncologists to remove patients from ineffective therapies quickly will, in the long term, likely improve significantly the overall clinical management of patients with breast cancer through individualized care.

Acknowledgments

The authors thank Bing Ma and Hyunjin Park for their assistance with the software and evaluation of the image registration algorithms.

References

[1] Abramson RG, Li X, Hoyt TL, Su PF, Arlinghaus LR, Wilson KJ, Abramson VG, Chakravarthy AB, and Yankeelov TE (2013). Early assessment of breast cancer response to neoadjuvant chemotherapy by semi-quantitative analysis of high-temporal resolution DCE-MRI: preliminary results. *Magn Reson Imaging* **31**, 1457–1464.

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

[3] Yankeelov TE, Atuegwu N, Hormuth D, Weis JA, Barnes SL, Miga MI, Rericha EC, and Quaranta V (2013). Clinically relevant modeling of tumor growth and treatment response. *Sci Transl Med* **5**, 187ps189.

[4] Ingrisch M and Sourbron S (2013). Tracer-kinetic modeling of dynamic contrast-enhanced MRI and CT: a primer. *J Pharmacokinet Pharmacodyn* **40**, 281–300.

[5] Sourbron SP and Buckley DL (2013). Classic models for dynamic contrast-enhanced MRI. *NMR Biomed* **26**, 1004–1027.

[6] Galban S, Brisset JC, Rehemtulla A, Chenevert TL, Ross BD, and Galbán CJ (2010). Diffusion-weighted MRI for assessment of early cancer treatment response. *Curr Pharm Biotechnol* **11**, 701–708.

[7] Jones DK and Cercignani M (2010). Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR Biomed* **23**, 803–820.

[8] Moffat BA, Chenevert TL, Lawrence TS, Meyer CR, Johnson TD, Dong Q, Tsien C, Mukherji S, Quint DJ, Gebarski SS, et al. (2005). Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proc Natl Acad Sci USA* **102**, 5524–5529.

[9] Galban CJ, Chenevert TL, Meyer CR, Tsien C, Lawrence TS, Hamstra DA, Junck L, Sundgren PC, Johnson TD, Ross DJ, et al. (2009). The parametric response map is an imaging biomarker for early cancer treatment outcome. *Nat Med* **15**, 572–576.

[10] Hamstra DA, Galbán CJ, Meyer CR, Johnson TD, Sundgren PC, Tsien C, Lawrence TS, Junck L, Ross DJ, Rehemtulla A, et al. (2008). Functional diffusion map as an early imaging biomarker for high-grade glioma: correlation with conventional radiologic response and overall survival. *J Clin Oncol* **26**, 3387–3394.

[11] Ellingson BM, Cloughesy TF, Zaw T, Lai A, Nghiemphu PL, Harris R, Lalezari S, Wagle N, Naeini KM, Carrillo J, et al. (2012). Functional diffusion maps

(fDMs) evaluated before and after radiochemotherapy predict progression-free and overall survival in newly diagnosed glioblastoma. *Neuro Oncol* **14**, 333–343.
[12] Ellingson BM, Cloughesy TF, Lai A, Nghiemphu PL, and Pope WB (2012). Nonlinear registration of diffusion-weighted images improves clinical sensitivity of functional diffusion maps in recurrent glioblastoma treated with bevacizumab. *Magn Reson Med* **67**, 237–245.
[13] Chiba Y, Kinoshita M, Okita Y, Tsuboi A, Isohashi K, Kagawa N, Fujimoto Y, Oji Y, Oka Y, Shimosegawa E, et al. (2012). Use of ^{11}C-methionine PET parametric response map for monitoring WT1 immunotherapy response in recurrent malignant glioma. *J Neurosurg* **116**, 835–842.
[14] Harris RJ, Cloughesy TF, Pope WB, Nghiemphu PL, Lai A, Zaw T, Czernin J, Phelps ME, Chen W, and Ellingson BM (2012). ^{18}F-FDOPA and ^{18}F-FLT positron emission tomography parametric response maps predict response in recurrent malignant gliomas treated with bevacizumab. *Neuro Oncol* **14**, 1079–1089.
[15] Hoff BA, Kozloff KM, Boes JL, Brisset JC, Galbán S, Van Poznak CH, Jacobson JA, Johnson TD, Meyer CR, Rehemtulla A, et al. (2012). Parametric response mapping of CT images provides early detection of local bone loss in a rat model of osteoporosis. *Bone* **51**, 78–84.
[16] Mamounas EP, Bryant J, Lembersky B, Fehrenbacher L, Sedlacek SM, Fisher B, Wickerham DL, Yothers G, Soran A, and Wolmark N (2005). Paclitaxel after doxorubicin plus cyclophosphamide as adjuvant chemotherapy for node-positive breast cancer: results from NSABP B-28. *J Clin Oncol* **23**, 3686–3696.
[17] van der Hage JA, van de Velde CJ, Julien JP, Tubiana-Hulin M, Vandervelden C, and Duchateau L (2001). Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer Trial 10902. *J Clin Oncol* **19**, 4224–4237.
[18] Mauri D, Pavlidis N, and Ioannidis JP (2005). Neoadjuvant versus adjuvant systemic treatment in breast cancer: a meta-analysis. *J Natl Cancer Inst* **97**, 188–194.
[19] Onitilo AA, Onesti JK, Single RM, Engel JM, James TA, Aiello Bowles EJ, Feigelson HS, Barney T, and McCahill LE (2013). Utilization of neoadjuvant chemotherapy varies in the treatment of women with invasive breast cancer. *PLoS One* **8**, e84535.
[20] Rapoport BL, Demetriou GS, Moodley SD, and Benn CA (2013). When and how do I use neoadjuvant chemotherapy for breast cancer? *Curr Treat Options Oncol* **15**, 86–98.
[21] Kuroi K, Toi M, Ohno S, Nakamura S, Iwata H, Masuda N, Sato N, Tsuda H, Kurosumi M, and Akiyama F (2013). Prognostic significance of subtype and pathologic response in operable breast cancer; a pooled analysis of prospective neoadjuvant studies of JBCRG. *Breast Cancer*, PMID: 24338638, E-pub ahead of print.
[22] Brown LG (1992). A survey of image registration techniques. *Computing Surveys* **24**, 325–376.
[23] Maintz JB and Viergever MA (1998). A survey of medical image registration. *Med Image Anal* **2**, 1–36.
[24] Sotiras A, Davatzikos C, and Paragios N (2013). Deformable medical image registration: a survey. *IEEE Trans Med Imaging* **32**, 1153–1190.
[25] Zitová B and Flusser J (2003). Image registration methods: a survey. *Image Vis Comput* **21**, 977–1000.
[26] Kessler ML (2006). Image registration and data fusion in radiation therapy. *Br J Radiol* **79 Spec. No. 1**, S99–S108.
[27] Guo Y, Sivaramakrishna R, Lu CC, Suri JS, and Laxminarayan S (2006). Breast image registration techniques: a survey. *Med Biol Eng Comput* **44**, 15–26.
[28] Holden M (2008). A review of geometric transformations for nonrigid body registration. *IEEE Trans Med Imaging* **27**, 111–128.
[29] Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, and Hawkes DJ (1999). Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* **18**, 712–721.
[30] Schnabel JA, Tanner C, Castellano-Smith AD, Degenhard A, Leach MO, Hose DR, Hill DL, and Hawkes DJ (2003). Validation of nonrigid image registration using finite-element methods: application to breast MR images. *IEEE Trans Med Imaging* **22**, 238–247.
[31] Rohlfing T, Maurer CR Jr, Bluemke DA, and Jacobs MA (2003). Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE Trans Med Imaging* **22**, 730–741.
[32] Heinrich MP, Jenkinson M, Bhushan M, Matin T, Gleeson FV, Brady SM, and Schnabel JA (2012). MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. *Med Image Anal* **16**, 1423–1435.
[33] Barber D and Hose D (2005). Automatic segmentation of medical images using image registration: diagnostic and simulation applications. *J Med Eng Technol* **29**, 53–63.
[34] Martel A, Froh M, Brock K, Plewes D, and Barber D (2007). Evaluating an optical-flow-based registration algorithm for contrast-enhanced magnetic resonance imaging of the breast. *Phys Med Biol* **52**, 3803.
[35] Li X, Dawant BM, Welch EB, Chakravarthy AB, Freehardt D, Mayer I, Kelley M, Meszoely I, Gore JC, and Yankeelov TE (2009). A nonrigid registration algorithm for longitudinal breast MR images and the analysis of breast tumor response. *Magn Reson Imaging* **27**, 1258–1270.
[36] Ong RE, Ou JJ, and Miga MI (2010). Non-rigid registration of breast surfaces using the laplace and diffusion equations. *Biomed Eng Online* **9**, 8.
[37] Sarkar S, Johnson TD, Ma B, Chenevert TL, Bland PH, Park H, Schott AF, Ross BD, and Meyer CR (2012). Evaluation of an automatic registration-based algorithm for direct measurement of volume change in tumors. *Int J Radiat Oncol Biol Phys* **83**, 1038–1046.
[38] Fluck O, Vetter C, Wein W, Kamen A, Preim B, and Westermann R (2011). A survey of medical image registration on graphics hardware. *Comput Methods Programs Biomed* **104**, e45–e57.
[39] Ma B, Meyer CR, Pickles MD, Chenevert TL, Bland PH, Galbán CJ, Rehemtulla A, Turnbull LW, and Ross BD (2009). Voxel-by-voxel functional diffusion mapping for early evaluation of breast cancer treatment. *Inf Process Med Imaging* **21**, 276–287.
[40] Lee KC, Moffat BA, Schott AF, Layman R, Ellingworth S, Juliar R, Khan AP, Helvie M, Meyer CR, Chenevert TL, et al. (2007). Prospective early response imaging biomarker for neoadjuvant breast cancer chemotherapy. *Clin Cancer Res* **13**, 443–450.
[41] Lee KC, Sud S, Meyer CR, Moffat BA, Chenevert TL, Rehemtulla A, Pienta KJ, and Ross BD (2007). An imaging biomarker of early treatment response in prostate cancer that has metastasized to the bone. *Cancer Res* **67**, 3524–3528.
[42] Moffat BA, Chenevert TL, Meyer CR, McKeever PE, Hall DE, Hoff BA, Johnson TD, Rehemtulla A, and Ross BD (2006). The functional diffusion map: an imaging biomarker for the early prediction of cancer treatment outcome. *Neoplasia* **8**, 259–267.
[43] Galban CJ, Mukherji SK, Chenevert TL, Meyer CR, Hamstra DA, Bland PH, Johnson TD, Moffat BA, Rehemtulla A, Eisbruch A, et al. (2009). A feasibility study of parametric response map analysis of diffusion-weighted magnetic resonance imaging scans of head and neck cancer patients for providing early detection of therapeutic efficacy. *Transl Oncol* **2**, 184–190.
[44] Lemasson B, Galbán CJ, Boes JL, Li Y, Zhu Y, Heist KA, Johnson TD, Chenevert TL, Galbán S, and Rehemtulla A (2013). Diffusion-weighted MRI as a biomarker of tumor radiation treatment response heterogeneity: a comparative study of whole-volume histogram analysis versus voxel-based functional diffusion map analysis. *Transl Oncol* **6**, 554.
[45] Malyarenko D, Galbán CJ, Londy FJ, Meyer CR, Johnson TD, Rehemtulla A, Ross BD, and Chenevert TL (2012). Multi-system repeatability and reproducibility of apparent diffusion coefficient measurement using an ice-water phantom. *J Magn Reson Imaging* **37**, 1238–1246.
[46] Malyarenko DI, Ross BD, and Chenevert TL (2013). Analysis and correction of gradient nonlinearity bias in apparent diffusion coefficient measurements. *Magn Reson Med*, PMID: 23794533, E-pub ahead of print.
[47] Murphy K, van Ginneken B, Klein S, Staring M, de Hoop B, Viergever M, and Pluim J (2011). Semi-automatic construction of reference standards for evaluation of image registration. *Med Image Anal* **15**, 71.
[48] Giger ML, Karssemeijer N, and Schnabel JA (2013). Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. *Annu Rev Biomed Eng* **15**, 327–357.