ORIGINAL ARTICLE

Functional 4-D clustering for characterizing intratumor heterogeneity in dynamic imaging: evaluation in FDG PET as a prognostic biomarker for breast cancer

Rhea Chitalia [1,2] · Varsha Viswanath [1,2] · Austin R. Pantel [2] · Lanell M. Peterson [3] · Aimilia Gastounioti [2] · Eric A. Cohen [2] · Mark Muzi [3] · Joel Karp [2] · David A. Mankoff [2] · Despina Kontos [2]

Abstract

Purpose Probe-based dynamic (4-D) imaging modalities capture breast intratumor heterogeneity both spatially and kinetically. Characterizing heterogeneity through tumor sub-populations with distinct functional behavior may elucidate tumor biology to improve targeted therapy specificity and enable precision clinical decision making.

Methods We propose an unsupervised clustering algorithm for 4-D imaging that integrates Markov-Random Field (MRF) image segmentation with time-series analysis to characterize kinetic intratumor heterogeneity. We applied this to dynamic FDG PET scans by identifying distinct time-activity curve (TAC) profiles with spatial proximity constraints. We first evaluated algorithm performance using simulated dynamic data. We then applied our algorithm to a dataset of 50 women with locally advanced breast cancer imaged by dynamic FDG PET prior to treatment and followed to monitor for disease recurrence. A functional tumor heterogeneity (FTH) signature was then extracted from functionally distinct sub-regions within each tumor. Cross-validated time-to-event analysis was performed to assess the prognostic value of FTH signatures compared to established histopathological and kinetic prognostic markers.

Results Adding FTH signatures to a baseline model of known predictors of disease recurrence and established FDG PET uptake and kinetic markers improved the concordance statistic (C-statistic) from 0.59 to 0.74 ($p = 0.005$). Unsupervised hierarchical clustering of the FTH signatures identified two significant ($p < 0.001$) phenotypes of tumor heterogeneity corresponding to high and low FTH. Distributions of FDG flux, or Ki, were significantly different ($p = 0.04$) across the two phenotypes.

Conclusions Our findings suggest that imaging markers of FTH add independent value beyond standard PET imaging metrics in predicting recurrence-free survival in breast cancer and thus merit further study.

Keywords Intratumor heterogeneity · Breast cancer · Dynamic PET · Imaging markers

Introduction

Cancer heterogeneity is well-established, with inter- and intratumor manifestations recognized as key prognostic and

---

✉ Despina Kontos
Despina.Kontos@pennmedicine.upenn.edu

1 Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

2 Department of Radiology, University of Pennsylvania, Rm. D702 Richards Bldg. 3700 Hamilton Walk, Philadelphia, PA 19104, USA

3 Department of Radiology, University of Washington, Seattle, WA, USA

predictive factors [1–5]. Increased intratumor heterogeneity is associated with adverse outcomes [6]. Tumor progression driven by aggressive cell subpopulations has been shown to be a mechanism for recurrence and therapy resistance [5]. Quantitative characterization of intratumor heterogeneity could allow for novel precision prognostic and predictive indicators.

Molecular and functional imaging modalities permit 4-D sampling of disease burden, capturing both spatial and temporal information that could illuminate various physiologic behaviors. Dynamic positron emission tomography (PET) imaging can quantify specific facets of tumor molecular biology [4, 7, 8] and can provide information beyond that of static imaging [9, 10]. Dynamic PET imaging of the glucose analog, ^{18}F-fluorodeoxyglucose (FDG), can provide simultaneous information on substrate delivery and metabolism [9]. Current

clinical characterization of malignant lesions using PET imaging largely utilizes qualitative descriptors [5] and quantitative measures based on static radiotracer uptake (e.g., SUVmax) [11].

The emerging field of radiomics has introduced multiparametric imaging features extracted with high-throughput computational analysis [12–15]. Previous work by Eary et al. quantified spatial heterogeneity of radiotracer uptake in static PET imaging and demonstrated improved prognostic performance over established clinical markers [16]. Stoyanova et al. identified sub-regions within pre-clinical dynamic contrast enhanced (DCE)-MRI images of prostate tumors [17]. Similarly, Cherezov et al. identified tumor habitats using established radiomic texture features [18]. While such studies demonstrate the prognostic potential of characterizing 3-D spatial heterogeneity, and are in line with studies showing differential physiologic functionality across the whole tumor [19, 20], these studies do not fully utilize the combined spatial and kinetic (e.g., 4-D) heterogeneity information available using imaging probe kinetics from modalities with high temporal imaging resolution.

The advantages of utilizing kinetic information from dynamic PET imaging have been demonstrated in breast cancer. Previous studies have demonstrated predictive improvement when FDG delivery (K_1) and FDG flux (Ki), in combination with [^{15}O]-water imaging, were utilized with static SUV measures as markers for neoadjuvant chemotherapy response in patients with locally advanced breast cancer [9, 21, 22]. However, these conventional kinetic parameters derived from FDG PET imaging at baseline alone were unable to show association to disease-free survival, likely limited by being derived from the most metabolically active portion of the tumor, and therefore not fully capturing intratumor heterogeneity [9].

We have developed a method to characterize 4-D functional tumor heterogeneity (FTH) by capturing aspects of both spatial and kinetic tumor heterogeneity seen in dynamic imaging. The improved dynamic sampling and molecular specificity available in dynamic PET as compared to other imaging modalities may allow for non-invasive, novel prognostic, and predictive markers to characterize tumor molecular biology. The developed method is agnostic to the specific radiotracer utilized and does not depend on complex kinetic modeling assumptions. Instead, the approach is data driven in terms of identifying intrinsic 4-D patterns of molecular tumor heterogeneity. We present proof-of-principle by applying our algorithm on dynamic FDG PET imaging scans of primary locally advanced breast cancer. We also investigate the role of imaging signatures as a prognostic biomarker for locally advanced breast cancer and the improved predictive value of FTH characterization compared to standard dynamic and static analytic methods for FDG PET.

Materials and methods

Radiomic functional intratumor (Rad-FIT) clustering: development and validation

Dynamic PET simulated image phantoms

Algorithm development and training were carried out using simulated 4-D PET data. With the goal of developing a method with broad applicability, we chose to perform these algorithm trainings using a different tracer and tumor type than FDG and breast cancer, but with broadly similar tracer kinetics. Simulated dynamic PET images based on data from fluorothymidine (^{18}F-FLT) (FLT) PET [23] were utilized for method development and validation. FLT was chosen as a representative tracer with kinetics similar to other trapped cancer-relevant radiotracers to assess the generalizability of our method. All simulations were done using Geant4 Application for Tomographic Emission (GATE) [24]. The scanner simulations were based on the PennPET Explorer [25] with 70-cm axial field of view. Based on prior human studies of FLT data from patients with lung cancer [23], kinetic parameters were selected to emulate low, medium, and high K_{FLT} lesions (ml/cm^3/s), and the blood input curve was derived from an FLT PET patient dataset and fit to a triexponential model. Details on simulated image generation have been previously described [26]. The simulated images were cropped to a region-of-interest (ROI) comprising the simulated regions and surround background area to a total size of $64 \times 69 \times 9$ voxels $\times 45$ frames. The simulated images consisted of two regions modeling low tracer uptake (10-mm and 13-mm sphere diameter), two regions modeling medium tracer uptake (10-mm and 13-mm sphere diameter), two regions modeling high tracer uptake (10-mm and 13-mm sphere diameter), a blood curve region, and a background region (Fig. 1).

Radiomic functional intratumor (Rad-FIT) clustering algorithm

To accurately assess both the spatial and temporal behavior seen in breast tumors, radiomic functional intratumor (Rad-FIT) clustering was developed as a 4-D-segmentation method to account for voxel similarity in the time domain as well as weight spatial contiguity. The temporal signal of an ROI over dynamic PET scans is first summarized using functional principal component analysis (FPCA), with each voxel represented using functional principal components (FPC) capturing greater than 85% of the variance seen in its dynamic behavior. Use of FPCA allows for the dimensionality reduction of time series data for each voxel while retaining the variance of its temporal information.
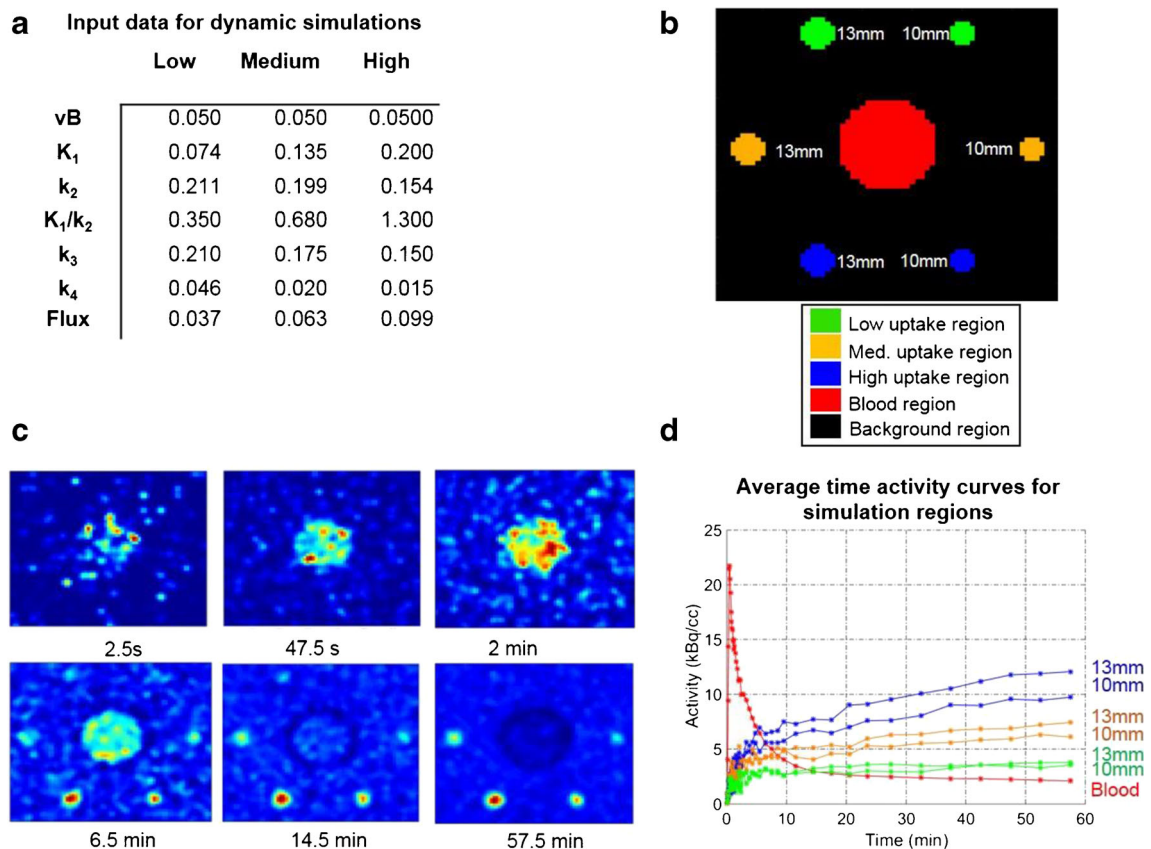
a Input data for dynamic simulations

	Low	Medium	High
vB	0.050	0.050	0.0500
K_1	0.074	0.135	0.200
k_2	0.211	0.199	0.154
K_1/k_2	0.350	0.680	1.300
k_3	0.210	0.175	0.150
k_4	0.046	0.020	0.015
Flux	0.037	0.063	0.099

Fig. 1 Dynamic FLT-PET simulation images used for method development and validation. **a** Input data used to generate dynamic simulations. **b** Labels for all regions in simulated image. **c** Center slice images from representative simulation frames. **d** Average TACs for simulation sphere and blood curve regions

A Markov Random Field [27] (MRF) segmentation paradigm is then applied to the voxel data for a learning-based method which incorporates spatial interactions between voxels allowing for a $3 \times 3 \times 3$ voxel neighborhood surrounding each voxel to influence segmentation. Initialization of mean and standard deviation values for each label k is performed using K-means clustering with a predetermined number of k labels. The final segmentations are then determined using the expectation maximization algorithm [28]. Details regarding this methodology can be found in the Supplementary materials (Supplementary information, Supplemental Fig. 1, Supplemental Table 1).

Rad- FIT evaluation and comparison to existing unsupervised clustering methods

To evaluate the performance of segmenting spatially contiguous regions of heterogeneous tracer activity with Rad-FIT, the following segmentation assessments were performed.

First, the improved value of summarizing temporal information from dynamic voxel behavior using FPCs was evaluated. The segmentation performance of Rad-FIT when simulation voxels were represented by their original time activity curves was compared to the performance when simulation voxels were represented by their FPC scores.

The segmentation performance of Rad-FIT was then evaluated for its ability to segment the high, medium, and low uptake 13-mm spheres from each sphere's surrounding background region, respectively. All segmentation performances were compared against the performance of established unsupervised segmentation algorithms including K-means clustering, hierarchical clustering, and spectral clustering [29]. Lastly, Rad-FIT segmentation performance was compared to the highest performing unsupervised clustering algorithm as outlined above when segmenting the simulated image into five distinct regions: background, blood, and low, medium, and high sphere regions. All segmentation performances were evaluated using the Dice score [30] and Jaccard index, averaged over ten replicates. Both the Dice score and Jaccard index are established statistical metrics used to determine the degree of overlap between the true regions and resulting regions from the segmentation algorithm. Both Dice scores and Jaccard indices include values ranging from 0 to 1, with a value of 1 indicating perfect similarity between true and segmented regions.

Functional tumor heterogeneity signature as a prognostic biomarker for dynamic FDG PET imaging of breast cancer

Study cohort

To investigate the role of intratumor segmentation when characterizing functional heterogeneity; the prognostic value of functional tumor heterogeneity imaging signatures was explored on a previously published data set where serial dynamic FDG PET was shown to be predictive of response and recurrence using standard static uptake and kinetic analysis [9, 21, 22]. The goal was to test functional tumor heterogeneity imaging signatures extracted from dynamic FDG PET scans of women with locally advanced breast cancer imaging prior to treatment and compare their predictive value to standard approaches.

We used an anonymized data set consisting of women presenting at the University of Washington Breast Cancer Specialty Center with histologically confirmed breast carcinoma who underwent dynamic FDG PET imaging prior to neoadjuvant chemotherapy and were followed for disease recurrence. The research protocol was approved by the institutional review board and patients studied provided informed consent prior to imaging and follow-up. The data set for this analysis was taken from a study first reported for 35 patients [21]. An additional 30 patients were later studied and added to a follow-up report of the data [9, 22]. From this pooled data set of 65 women with complete baseline dynamic FDG PET scans who also completed neo-adjuvant chemotherapy and post-therapy surgery, two women were excluded for electing not to receive chemotherapy, three women were excluded for electing for medical care elsewhere, four patients were excluded for being unwilling to undergo mid-therapy imaging, two patients were excluded due to distant disease, and one patient was excluded due to little or no tracer uptake upon pre-therapy examination resulting in a total of 53 women. Of these, two women were excluded due to image artifacts and one woman was excluded due to incomplete survival information, resulting in a total of 50 women included in this study. Dynamic FDG PET images from these 50 women comprised our study sample reported here. Details of the patient population have been previously described [22].

Each woman had undergone 60-min dynamic FDG PET centered over the breast prior to neoadjuvant chemotherapy and breast surgery. All women were imaged in the supine position and no positioning devices for immobilization were utilized. Women were infused with 218–396 MBq of FDG over 2 min in a 7–10 mL volume, with an intended injected dose of 370 MBq. Images for all women were acquired on an Advanced Tomograph (General Electric Medical Systems, Waukesha, WI) using the same image acquisition protocol. Dynamic images were acquired (25 image frames: 1-min pre-injection frame, 4×20 s, 4×40 s, 4×40 s, 4×1 min, 4×3 min, 8×5 min). Images were reconstructed into $35 \times 128 \times 128$ voxel matrices with a spatial resolution of 10–12 mm [21]. Clinical information collected as part of the study included hormone receptor (HR) status consisting of estrogen receptor (ER) and progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), clinical stage, tumor size, proliferation (Ki67), pathologic complete response (pCR), axillary lymph node (ALN) positivity, and age at diagnosis (Supplemental Table 2). In the study, recurrence free survival (RFS) was tracked for each patient, defined as date of known recurrence, date of death, date of most recent clinical follow-up with no evidence of disease, following the patient's date of surgery. Patients received standard of care follow-up including routine period imaging of CT scans, blood marker analysis (CA2729), and follow-up visits to check for symptoms.

Established ROI-based measures of uptake and kinetics for dynamic FDG PET—summed imaged standardized uptake value (SUV), and the kinetic parameters of FDG blood-to-tissue transport (K_1) and FDG trapping flux (Ki)—were calculated based on kinetic modeling of dynamic data for each woman and have been previously reported [9, 21]. These kinetic parameters were measured for each tumor from a 1.5-cm-diameter circle VOI surrounding the area of maximal tumor FDG uptake seen on the 30–60-m summed image.

Functional tumor heterogeneity (FTH) signature extraction

A 3-D bounding region surrounding each unifocal lesion was manually identified by a radiologist blinded to the outcome of each patient using the final of the summed FDG images for the 25 imaging frames (30–60 min post-injection) and guided by ROIs previously used for extraction of SUV, K_1, and Ki for consistency. An established segmentation approach was applied to the TACs generated from the 25 imaging frames of each voxel within the bounding region to segment the tumor from its surrounding background [31].

Within the segmented 3-D tumor region, Rad-FIT clustering was applied to segment each tumor region into three, spatially constrained sub-regions with distinct functional behavior. Three sub-regions were selected based on the rationale that there are currently three major subtypes of breast tumors broadly recognized: hormone receptor positive, HER2 positive, and triple negative [2]. The three sub-regions within each tumor were ranked in the order of descending mean value of the first FPC to allow for consistent comparisons across tumors.

The resulting Rad-FIT clustering within each tumor was summarized using metrics describing sub-region compactness and separation. These metrics were chosen to summarize how well the functional behavior of each tumor will cluster into three groups and to allow for comparisons of intratumor

heterogeneity across women. Compactness was measured using the between cluster sum of squares (BCSS) scaled by the total sum of squares (TSS).:

$$\frac{\sum_{k=1}^{K}\left(\overline{x_k}-\overline{X}\right)^2}{\sum_{i=1}^{N}\left(x_i-\overline{X}\right)^2} \tag{1}$$

where K represents the 3 sub-regions and N is the total number of voxels within each tumor. The separation between sub-regions was determined using the Bhattacharya distance [32] to calculate the distance between FPC distributions of two sub-regions (ϕ). Use of this distance allows for a similarity measure between the distributions of FPC values within two sub-regions.

Based on the definitions above, a total of four features summarizing intratumor heterogeneity from Rad-FIT clustering results were extracted to form an FTH signature (Fig. 2): (1) BCSS/TSS, (2) distance between sub-region 1 and 2 ($\phi(1,2)$), (3) distance between sub-region 2 and 3 ($\phi(2,3)$), and (4) distance between sub-region 1 and 3 ($\phi(1,3)$).

BCSS/TSS provides a measure of how compact the resulting clusters are; the more compact each cluster is, the greater heterogeneity between the identified sub-regions. Calculating the distance between the distributions of FPC values of two sub-regions provides a metric for how separated the clusters are; a greater distance between sub-regions indicates greater heterogeneity within the whole tumor region. This FTH signature can be used to interpret how distinct the three identified sub-regions are within each tumor. As such, the average value of the FTH signature, or FTH signature index, can provide a metric for intratumor heterogeneity across tumors.

Statistical analysis: evaluation of FTH signatures as a prognostic biomarker

Our goals in this proof of principal study were to test the prognostic value of FTH signatures from dynamic breast cancer FDG PET and to assess for incremental value compared to standard clinical parameters and conventional FDG PET static and dynamic analysis measures used in prior published analyses. FTH signatures were first z score normalized across all women. Time-to-event analysis was then used to assess the prognostic value of the FTH signatures in predicting recurrence-free survival (RFS). To this end, a threefold cross validated (CV) Cox proportional hazards model was used to compare improved prognostic discriminatory capacity over baseline models of established prognostic factors consisting of ER status, PR status, tumor size, pCR, and ALN positivity and kinetic parameters consisting of the SUV, K_1, and Ki. These prognostic factors were chosen based on the available data as well as the intent to compare analysis results to prior published data [9, 21, 22].

Model performance was evaluated using an averaged C-statistic over the test sets for all three folds and the log-likelihood statistical test.

The prognostic value of the FTH signature was evaluated via Kaplan-Meier survival analysis using each patient's risk core, dichotomizing patients into high- and low-risk groups. The risk score for each patient was defined as the patient's FTH signature weighted by the corresponding coefficients from each of the three test sets from a threefold cross validated model for each covariate in the FTH signature [12, 33]. Risk scores generated from baseline features of ER status, PR status, tumor size, pCR, and ALN positivity and from baseline and kinetic features were also assessed. Statistical significance of Kaplan-Meier stratification was evaluated using the Log Rank Test.

Lastly, an exploratory unsupervised hierarchical clustering was performed on the extracted FTH signatures from each woman. The resulting c clusters obtained from the hierarchical clustering algorithm were interpreted as c intrinsic FTH phenotypes seen in this study population. The optimal number of stable FTH phenotypes was determined using consensus clustering [34]. Statistical significance of the identified, stable FTH phenotypes was evaluated using the SigClust method [35]. The distribution of histopathologic and kinetic
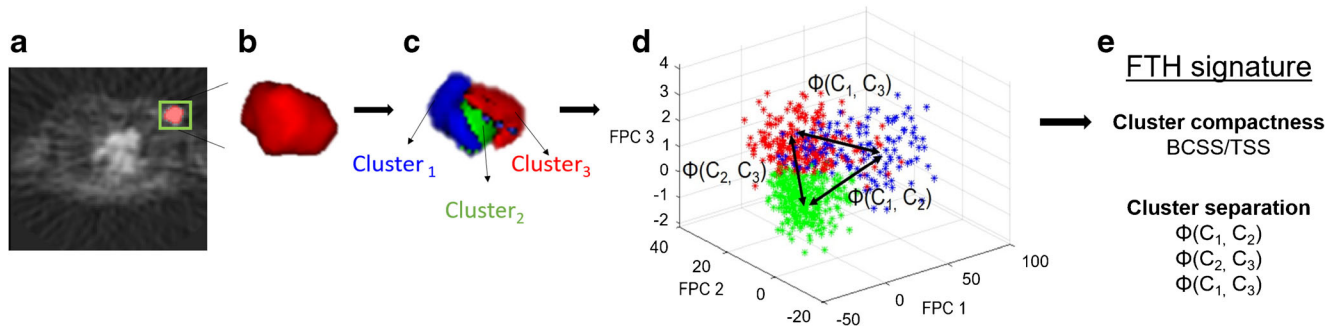

Fig. 2 FTH signature extraction. **a** 3-D tumor region identified by a radiologist shown in green. **b** Automated 3-D segmentation of tumor from background with pixels represented using time activity curves. **c** Rad-FIT clustering performed, identifying three, spatially contiguous sub-regions. **d** Cluster compactness and cluster separation distances are calculated to form features in FTH signature. **e** Intratumor heterogeneity summarized using the FTH signature

prognostic covariate values across women assigned to each of the FTH phenotypes was assessed using chi-square tests for categorical biomarkers and one-way analysis of variance test for continuous biomarkers.

Results

Rad-FIT evaluation and comparison to existing unsupervised clustering methods

Average signal to background ratios for the low, medium, and high uptake spheres were 1.78, 6.95, and 11.89, respectively. The segmentation performance of Rad-FIT clustering was compared against existing unsupervised segmentation methods, K-means clustering, hierarchical clustering, and spectral clustering, when applied to the simulated images with varying regions of known uptake value and location.

When segmenting low, medium, and high uptake spheres from its surrounding background region, segmentation performance for all segmentation methods improved when simulation voxels were represented by the FPCs capturing greater than 85% of TAC variability as opposed to using the TACs. Additionally, Rad-FIT clustering demonstrated the highest segmentation performance when segmenting low, medium, and high uptake sphere regions from its surrounding backgrounds when evaluated using the Dice score (Table 1) and Jaccard index (Supplemental Table 3).

The K-means clustering algorithm had the second highest segmentation performance across the evaluated unsupervised clustering algorithms. This algorithm was subsequently used to compare performance against the Rad-FIT clustering algorithm to segment the simulated image into five regions: background, blood, low uptake spheres, medium uptake spheres, and high uptake spheres. The Rad-FIT clustering algorithm was able to segment the low uptake spheres from its surrounding background region and identify the five regions of distinct tracer uptake (Fig. 3). In comparison, the K-means clustering algorithm was unable to identify and segment the low uptake

spheres as a distinct region from its surrounding background. Overall, the Rad-FIT clustering algorithm demonstrated improved segmentation over the K-means clustering algorithm as demonstrated by decreased mean percent error and increased Dice score and Jaccard index for each individual region's segmentation as well as segmenting the overall simulated image. Due to this improved segmentation performance, the Rad-FIT clustering algorithm was utilized as the unsupervised clustering method towards characterizing functional tumor heterogeneity as a prognostic biomarker for women diagnosed with locally advanced breast cancer.

Evaluation of the FTH signature as a prognostic biomarker

Of the 50 women included in the data set selected from prior studies for evaluation of FTH signatures as a prognostic biomarker, 17 women (34%) had recurrence events. A total of 47 (94%) women were diagnosed with infiltrating ductal carcinoma, and 3 (6%) women were diagnosed with infiltrating lobular carcinoma.

Of the non-recurrent cases, 58% were ER positive, 52% were PR positive, and 18% were HER2 positive. Of the recurrent cases, 59% were ER positive, 59% were PR positive, and 35% were HER2 positive (Supplemental Table 2).
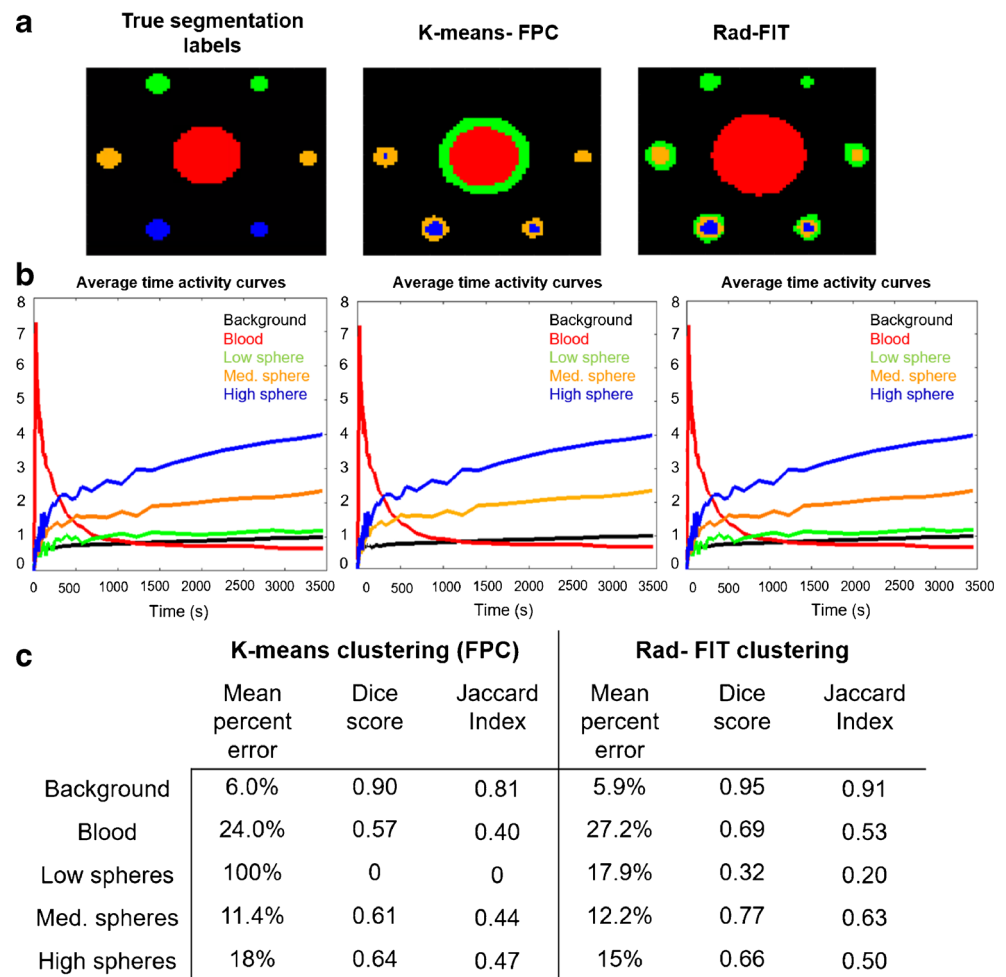
Representative tumor images after Rad-FIT clustering demonstrate intratumor heterogeneity within breast tumors (Fig. 4). Tumors with increased intratumor heterogeneity can be identified as having sub-regions with distinct time activity curve behaviors, while tumors with decreased intratumor heterogeneity display little distinction between the time activity curve behavior of the identified sub-regions.

As expected, in a full multivariate Cox proportional hazards model after adjusting for ER status, PR status, tumor size, pCR, and ALN positivity (Supplemental Table 4), $\phi(1,2)$ and $\phi(2,3)$ were associated with disease-free survival (Supplemental Table 5). A baseline, threefold CV Cox proportional hazards model consisting of ER status, PR status, tumor size, pCR, and ALN positivity resulted in a mean C-

Table 1 Average segmentation performance over ten replicates evaluated using the Dice scores when segmenting low, medium, and high uptake simulated sphere regions from surrounding backgrounds. Standard deviation in parentheses

Segmentation region	Voxel representation	Dice scores			
		Hierarchical clustering	Spectral clustering	K-means clustering	Rad-FIT clustering
Low uptake sphere	Time activity curves	0.15 (0)	0.04 (0.001)	0.13 (0.06)	0.24 (0.07)
	FPC	0.67 (0)	0.16 (0.03)	0.65 (0.02)	0.70 (0.01)
Medium uptake sphere	Time activity curves	0.77 (< 0.001)	0.06 (0.01)	0.44 (0.4)	0.73 (< .001)
	FPC	0.78 (0)	0.20 (0.18)	0.84 (0.03)	0.85 (0)
High uptake sphere	Time activity curves	0.72 (0)	0.18 (0.07)	0.52 (0.40)	0.66 (0)
	FPC	0.83 (0)	0.17 (0.03)	0.84 (0)	0.86 (0)

Fig. 3 **a** Representative center slice image of true segmentation labels (left), segmentation results using *K*-means clustering (middle), and segmentation results using Rad-FIT clustering (right) when segmenting simulation images into five classes: background, blood, and low, medium, and high uptake spheres. **b** Average time activity curves for true simulation regions (left), *K*-means clustering identified regions (middle), and Rad-FIT clustering (right) identified regions. **c** Mean percent error, generalized dice score, and Jaccard indices for segmentation results. The *K*-means clustering algorithm fails to segment the low uptake spheres from the background region, and instead segments the blood region to two separate classes

	K-means clustering (FPC)			Rad- FIT clustering		
	Mean percent error	Dice score	Jaccard Index	Mean percent error	Dice score	Jaccard Index
Background	6.0%	0.90	0.81	5.9%	0.95	0.91
Blood	24.0%	0.57	0.40	27.2%	0.69	0.53
Low spheres	100%	0	0	17.9%	0.32	0.20
Med. spheres	11.4%	0.61	0.44	12.2%	0.77	0.63
High spheres	18%	0.64	0.47	15%	0.66	0.50

statistic of 0.51 when predicting RFS. Adding SUV, K_1, and K_i parameters to the baseline model resulted in a mean CV C-statistic of 0.54. Adding the FTH signature to the baseline model improved the mean CV C-statistic to 0.74 ($p < 0.01$) (Fig. 5a).

Dichotomizing patients into low- and high-risk groups based on the baseline model risk scores (Fig. 5b) and baseline plus the kinetic model risk scores (Fig. 5c) demonstrated no statistically significant separation between Kaplan-Meier curves. Patient dichotomization into low- and high-risk groups based on the baseline plus FTH signature risk scores (Fig. 5d) resulted in a statistically significant separation between Kaplan- Meier curves ($p < 0.05$) for RFS probability.

Unsupervised hierarchical clustering of women based on the extracted FTH signatures from each tumor identified two clusters which were interpreted as FTH phenotypes seen in the study population and found to be statistically significant via the SigClust method ($p = 0.04$) (Fig. 6a).

As a higher FTH signature index suggests greater separation between the three sub-regions' FPC values and can therefore be interpreted as greater intratumor heterogeneity, the identified FTH phenotypes were ranked based on the mean FTH signature index found across all women assigned to each phenotype. The resulting FTH phenotypes 1 and 2 were interpreted as a low FTH versus high FTH phenotypes, respectively, with tumors in phenotype 1 having lower mean FTH signature indices (blue color in Fig. 6a), versus tumors in phenotype 2 which had on average higher FTH signature indices (red color in Fig. 6a).

Tumor PR status was found to be statistically significantly different across the two FTH phenotypes ($p < 0.05$), with tumors in the low FTH phenotype having a higher proportion of PR-positive tumors. Other clinical covariates including ER status, HER2 status, Ki67 status, pCR, ALN positivity, and tumor grade were not statistically significant across identified phenotypes. From the FDG PET covariates, K_i was found to be statistically significantly different across the two phenotypes ($p < 0.05$), with tumors in the high FTH phenotype having a greater interquartile range (0.025) and greater variance (0.0003) than tumors in the low FTH phenotype (interquartile range: 0.004, variance: 3.27e-5) (Fig. 6b). K_1 and SUV values were not found to be statistically significant across the identified phenotypes.
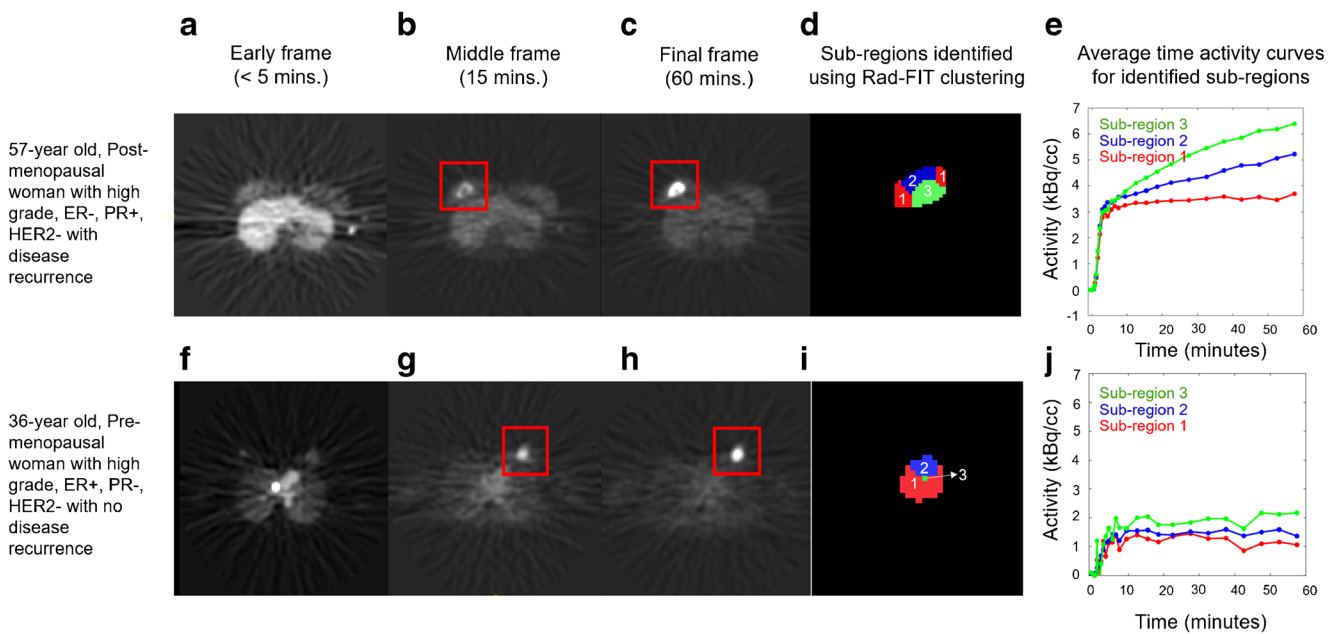
Fig. 4 Representative images for the primary tumor of two women diagnosed with locally advanced breast cancer with future tumor recurrence (top) or tumor non-recurrence (bottom). **a** Representative slice from an early frame at less than 5 min after tracer injection, (**b**) representative slice from middle frame at 15 min after tracer injection, and **c** representative slice from final frame taken at 60 min after tracer injection of a 57-year old, post-menopausal woman with a high grade-, ER−, PR+, HER2− tumor who had disease recurrence upon follow-up (top). **d** Three sub-regions identified using Rad-FIT clustering labeled as region 1 (red), 2 (blue), and 3 (green), and **e** average TACs for each identified sub-region. **f** A representative slice from an early frame at less than 5 min after tracer injection, **g** slice from middle frame at 15 min after tracer injection, and **h** representative slice from final frame taken at 60 min after tracer injection of a 36-year old, pre-menopausal woman with a high-grade, ER+, PR−, HER2− tumor with no disease recurrence (bottom). **i** Three sub-regions with distinct 4-D behavior identified using Rad-FIT clustering labeled as region 1 (red), 2 (blue), and 3 (green), and **j** average time activity curves for each identified sub-region
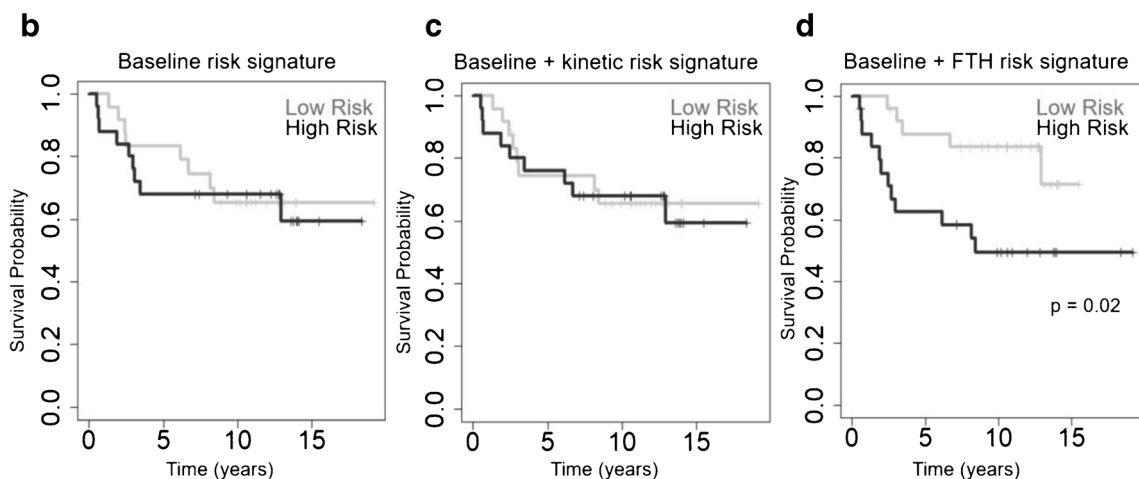
Fig. 5 **a** Cross-validated c scores and Kaplan-Meier survival curves for baseline (ER status, PR status, tumor size at baseline, pCR, and ALN positivity), baseline plus kinetic (ER status, PR status, tumor size at baseline, pCR, ALN positivity, SUV, K_1, Ki), and baseline plus FTH signature models. Kaplan-Meier curves generated when patients are stratified by risk scores generated from **b** the baseline model, **c** baseline plus kinetic features model, and **d** baseline plus FTH signature model
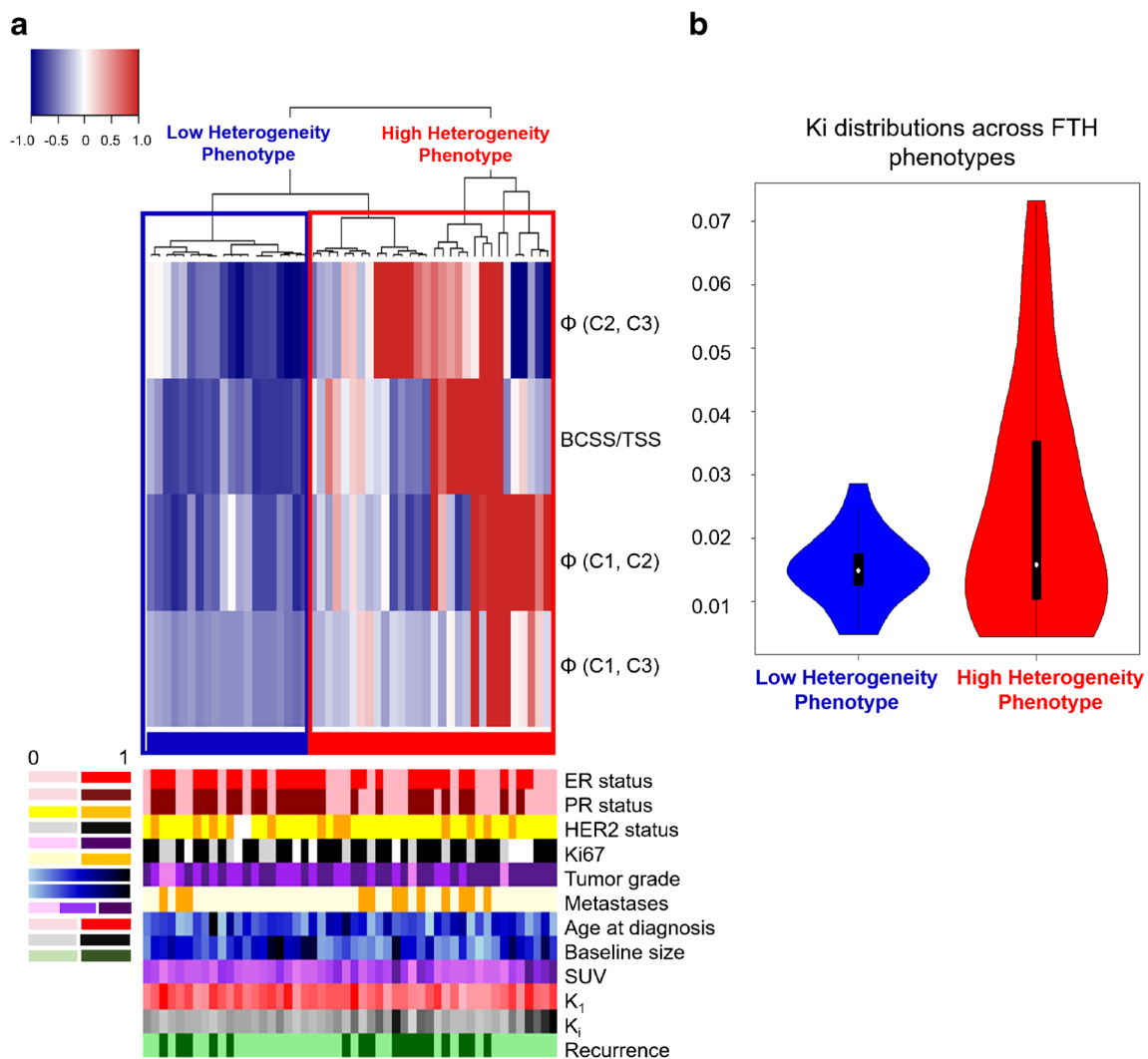
a

b

Fig. 6 **a** Unsupervised hierarchical clustering of FTH signatures identifies 2 significant phenotypes of FTH, with clinical covariate distribution across identified phenotypes displayed in the bottom legend. The resulting cluster dendrogram can be seen above a heatmap in which each row represents a feature within the FTH signature, and each column represents a tumor. **b** Distributions of Ki across the identified phenotypes were found to be statistically significant ($p < 0.05$)

Discussion

Our results suggest that incorporation of both spatial and kinetic information in a 4-D dynamic activity curve clustering paradigm allows for improved segmentation of dynamic PET imaging data over established unsupervised clustering techniques utilizing kinetic information alone. Established unsupervised voxel parcellation techniques largely assume voxel independence [5], and as such, may be inadequate for identifying spatially constrained, functionally similar sub-regions under the hypothesis that subclonal populations can occupy spatially contiguous regions with common biologic properties [36]. Along these lines, partial volume effects seen in imaging modalities suggest that neighboring voxels may share information regarding underlying tissue structure due to the spatial limitations of the imaging device [37]. Therefore, analyzing

imaging presentations of intratumor heterogeneity requires a fully 4-D approach.

Use of simulation images of distinct regions with known spatial locations and tracer kinetics allowed for method development and evaluation and allowed for the comparison across established techniques in this exploratory analysis. Comparing the segmentation performance of Rad-FIT clustering against established unsupervised segmentation algorithms demonstrated improved Dice scores of 0.70, 0.85, and 0.86 across low, medium, and high uptake sphere segmentation, respectively, when using simulated data with pre-defined ground truth. *K*-means clustering performed the second highest when all segmentation algorithms were evaluated, resulting in Dice scores of 0.65, 0.84, and 0.84 across low, medium, and high uptake sphere segmentation, respectively (Table 1). Additionally, when segmenting the simulated image into five regions of distinct tracer uptake, the Rad-

FIT clustering algorithm outperformed *K*-means clustering as demonstrated by mean percent error in average TACs from each region as well as the corresponding Dice scores and Jaccard indices. The better segmentation performance of Rad-FIT clustering emphasizes the added benefit of incorporating both spatial and kinetic information to allow for more accurate identification of functionally distinct sub-regions. As the Rad-FIT clustering algorithm outperformed the established unsupervised clustering algorithms in the simulation analysis, we chose to use this algorithm to characterize FTH as a prognostic biomarker.

We had the goal of developing methodology broadly applicable to PET tracers with similar kinetic features. With this in mind, dynamic simulations utilized for Rad-FIT development and validation were based on FLT simulated data and used to select an approach which was then applied to a previously collected FDG patient dataset to examine the role of FTH as a prognostic biomarker. The rationale was that while simulation curves were generated using kinetic parameters specific to FLT (flux between 0.03 and 0.1 mL/min/g), these kinetics curves have parameters similar to FDG PET curves (flux between 0.02 and 0.09 mL/min/g), and they can be generalized to all tracers fit to a two-compartment model, including FDG, as long as the model and range of parameters has overlap with these tracers. Additionally, and as our algorithm is in principle agnostic to the type of tracer used or any related kinetic modeling parameters, utilizing simulated images of a different two-compartment radiotracer during Rad-FIT development allowed for a more generalizable algorithm that was not biased towards a single specific tracer in subsequent analyses.

Extending the Rad-FIT clustering algorithm to characterize intratumor heterogeneity has the potential to identify intratumor sub-regions with discrete functional behavior. This is supported by the average TACs from the identified sub-regions in representative tumors (Fig. 4), where the tumor with disease recurrence clustered into three sub-regions with distinct curve patterns. The tumor with no disease recurrence and characterized as ER+, demonstrated mostly low uptake and non-rising curves in the identified sub-regions.

Quantifying intratumor heterogeneity using the FTH imaging signature demonstrates prognostic value when predicting RFS. Cox-regression models incorporating FTH signatures added to a baseline model demonstrated a statistically significant improvement in *C*-statistic. While the dichotomization of baseline risk scores based on known prognostic features did not demonstrate significant Kaplan-Meier curve separation in this relatively small sample size, consistent with previous analyses of this study cohort [9, 22], improvement over models combining baseline and kinetic features emphasizes the added prognostic value of utilizing quantitative features summarizing dynamic tumor behavior over the entire volume. Additionally, risk scores generated using the baseline features and FTH signature resulted in statistically significant patient dichotomization into low- and high-risk groups for RFS using Kaplan-Meier survival analysis as

compared to risk signatures generated from the baseline model and baseline plus kinetic model.

While prior studies have shown prognostic value for measures obtained from serial dynamic FDG PET using standard kinetic analysis methods (10), pre-therapy FDG dynamic data were not significantly predictive. Similarly, prior studies demonstrated predictive value pre-therapy measures of FDG flux and tumor blood flow obtained from combined ^{15}O-water PET and FDG PET studies [21, 22], but pre-therapy FDG kinetic measures alone were not significantly predictive of RFS. In this preliminary analysis, use of the Rad-FIT clustering algorithm extracted significantly prognostic 4-D signatures from pre-therapy dynamic FDG PET data that did predict RFS, a notable incremental improvement on standard approaches to dynamic PET analysis of considerable potential significance.

Additionally, our results suggest that intrinsic imaging phenotypes may exist within locally advanced breast tumors corresponding to FTH. In particular, statistically significant differences in the FDG flux constant, K_i, were seen across the two phenotypes with tumors corresponding to higher degrees of FTH having higher values of K_i. This finding suggests that the tumor characteristic of increased metabolic rate may be captured within the FTH imaging signature generated from the 4-D clustering performed using Rad-FIT and may have prognostic significance when expanded to a larger study cohort. Interestingly, compared to tumor clinical and histopathologic features, we found significant differences across the FTH phenotypes in PR expression, a marker shown to be an indicator of tumor ER functionality and a more differentiated breast cancer biologic phenotype [38].

Limitations of our study should be noted. First the Rad-FIT clustering algorithm utilizes *K*-means clustering as an initialization to the method, which can allow for sensitivity to cluster initialization due to *K*-means clustering identifying local optima. Future studies will be conducted to evaluate segmentation performance when random cluster initializations are selected. Additionally, our study utilized a relatively small sample size of patients. To account for potential model overfitting, we utilized threefold CV in our time-to-event analysis, to ensure model robustness. The identification of FTH phenotypes within the study population is limited by a lack of independent validation and instead was conducted as an exploratory analysis. Future work will include expanding this initial, exploratory analysis to a larger cohort as well as validating the identified FTH phenotypes. While the Rad-FIT clustering paradigm identified three clusters within each tumor, the optimal number of functionally discrete sub-regions may vary across tumors. In this exploratory study, the selection of three for the number of subtypes was chosen empirically, guided by the three major subtypes of breast cancer (ER+/PR+, Her2+, Triple negative). Future work will also include optimization of the Rad-FIT clustering algorithm such that an optimal number of clusters is identified within each tumor. In addition, we have evaluated only the pre-therapy time

point in this initial analysis. All women included in our study underwent neoadjuvant chemotherapy and repeat mid-therapy imaging. To account for the effect of treatment on intratumor heterogeneity and corresponding FTH signatures, we plan to expand our analysis to dynamic FDG PET images taken also during the midpoint of each woman's therapy in a future study. Lastly, alternative approaches exploring a linear analysis of 4-D dynamic PET using a mixture-based approach have been previously reported [39, 40]. Future work will include expanding our analysis to compare methodologies and potentially include a mixture-based component. Lastly, we developed and applied this method on simulated dynamic images and clinical dynamic scans of breast cancer patients, with a larger goal of extending this method towards analyzing other solid tumors and different PET tracers in future work.

Conclusion

In conclusion, we have developed a 4-D clustering and segmentation algorithm to identify functionally discrete, spatially constrained sub-regions within breast tumors that was able to generate prognostic measures from pre-therapy dynamic FDG PET of locally advanced breast cancer not previously identified by ROI-based kinetic analysis. Our results demonstrate that quantifying functional tumor heterogeneity can provide independent and additional prognostic value and may provide a non-invasive- 4-D characterization of breast tumors towards personalized decision making.

Declarations

References

1. Polyak K. Breast cancer: origins and evolution. J Clin Invest. 2007;117:3155–63.
2. Polyak K. Heterogeneity in breast cancer. J Clin Invest. 2011;121: 3786–8.
3. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochimica et Biophysica Acta (BBA)-reviews on. Cancer. 2010;1805:105–17.
4. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton J, Snyder A, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. Clin Radiol. 2017;72:3–10.
5. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res. 2015;21: 249–57.
6. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. Nat Med. 2016;22:105.
7. Lebron L, Greenspan D, Pandit-Taskar N. PET imaging of breast cancer: role in patient management. PET Clin. 2015;10:159–95.
8. McDonald ES, Mankoff DA, Mach RH. Novel strategies for breast cancer imaging: new imaging agents to guide treatment. J Nucl Med. 2016;57:69S–74S.
9. Dunnwald LK, Doot RK, Specht JM, Gralow JR, Ellis GK, Livingston RB, et al. PET tumor metabolism in locally advanced breast cancer patients undergoing neoadjuvant chemotherapy: value of static versus kinetic measures of fluorodeoxyglucose uptake. Clin Cancer Res. 2011;17:2400–9.
10. Humbert O, Lasserre M, Bertaut A, Fumoleau P, Coutant C, Brunotte F, et al. Breast cancer blood flow and metabolism on dual-acquisition 18F-FDG PET: correlation with tumor phenotype and neoadjuvant chemotherapy response. J Nucl Med. 2018;59: 1035–41.

11. Groheux D, Cochet A, Humbert O, Alberini J-L, Hindié E, Mankoff D. 18F-FDG PET/CT for staging and restaging of breast cancer. J Nucl Med. 2016;57:17S–26S.

12. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

13. Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. Radiology. 2016;281: 382–91.

14. Chitalia R, Rowland J, McDonald ES, Pantalone L, Cohen EA, Gastounioti A, et al. Imaging phenotypes of breast cancer heterogeneity in pre-operative breast Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) scans predict 10-year recurrence. Clin Cancer Res. 2019:clincanres.4067.2018. https://doi.org/10.1158/1078-0432.ccr-18-4067.

15. Chitalia RD, Kontos D. Role of texture analysis in breast MRI as a cancer biomarker: a review. J Magn Reson Imaging. 2019;49:927–38.

16. Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. J Nucl Med. 2008;49:1973–9.

17. Stoyanova R, Huang K, Sandler K, Cho H, Carlin S, Zanzonico PB, et al. Mapping tumor hypoxia in vivo using pattern recognition of dynamic contrast-enhanced MRI data. Transl Oncol. 2012;5:437–IN2.

18. Cherezov D, Goldgof D, Hall L, Gillies R, Schabath M, Müller H, et al. Revealing tumor habitats from texture heterogeneity analysis for classification of lung cancer malignancy and aggressiveness. Sci Rep. 2019;9:1–9.

19. Chang Y-CC, Ackerstaff E, Tschudi Y, Jimenez B, Foltz W, Fisher C, et al. Delineation of tumor habitats based on dynamic contrast enhanced MRI. Sci Rep. 2017;7:9746.

20. Jardim-Perassi BV, Martinez G, Gillies R. Habitat imaging of tumor evolution by magnetic resonance imaging (MRI). Radiomics and Radiogenomics: Technical Basis and Clinical Applications 2019:115.

21. Mankoff DA, Dunnwald LK, Gralow JR, Ellis GK, Schubert EK, Tseng J, et al. Changes in blood flow and metabolism in locally advanced breast cancer treated with neoadjuvant chemotherapy. J Nucl Med. 2003;44:1806–14.

22. Dunnwald LK, Gralow JR, Ellis GK, Livingston RB, Linden HM, Specht JM, et al. Tumor metabolism and blood flow changes by positron emission tomography: relation to survival in patients treated with neoadjuvant chemotherapy for locally advanced breast cancer. J Clin Oncol. 2008;26:4449.

23. Muzi M, Vesselle H, Grierson JR, Mankoff DA, Schmidt RA, Peterson L, et al. Kinetic analysis of 3′-deoxy-3′-fluorothymidine PET studies: validation studies in patients with lung cancer. J Nucl Med. 2005;46:274–82.

24. Strydhorst J, Buvat I. Redesign of the GATE PET coincidence sorter. Phys Med Biol. 2016;61:N522.

25. Karp JS, Viswanath V, Geagan MJ, Muehllehner G, Pantel AR, Parma MJ, Perkins AE, Schmall JP, Werner ME, Daube-Witherspoon ME. PennPET Explorer: design and preliminary performance of a whole-body imager. J Nucl Med. 2020;61(1):136–43.

26. Viswanath V, Pantel AR, Daube-Witherspoon ME, Doot R, Muzi M, Mankoff DA, Karp JS. Quantifying bias and precision of kinetic parameter estimation on the PennPET Explorer, a long axial field-of-view scanner. IEEE Transactions on Radiation and Plasma Medical Sciences. 2020;4(6):35–749.

27. Li SZ. Markov random field models in computer vision. European conference on computer vision: Springer, Berlin, Heidelberg. 1994;361–70.

28. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B Methodol. 1977;39:1–22.

29. Von Luxburg U. A tutorial on spectral clustering. Stat Comput. 2007;17:395–416.

30. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945;26:297–302.

31. Chen JL, Gunn SR, Nixon MS, Gunn RN. Markov random field models for segmentation of PET images. Biennial International Conference on Information Processing in Medical Imaging: Springer. 2001;468–74.

32. Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. Bull Calcutta Math Soc. 1943;35:99–109.

33. Jahani N, Cohen E, Hsieh M-K, Weinstein SP, Pantalone L, Hylton N, et al. Prediction of treatment response to neoadjuvant chemotherapy for breast cancer via early changes in tumor heterogeneity captured by Dce-MRi registration. Sci Rep. 2019;9:1–12.

34. Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach Learn. 2003;52:91–118.

35. Liu Y, Hayes DN, Nobel A, Marron J. Statistical significance of clustering for high-dimension, low–sample size data. J Am Stat Assoc. 2008;103:1281–93.

36. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nature reviews Clinical oncology. 2018;15(2):81.

37. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. J Nucl Med. 2007;48:932–45.

38. Chan SR, Fowler AM, Allen JA, Zhou D, Dence CS, Sharp TL, et al. Longitudinal noninvasive imaging of progesterone receptor as a predictive biomarker of tumor responsiveness to estrogen deprivation therapy. Clin Cancer Res. 2015;21:1063–70.

39. O'Sullivan F. Locally constrained mixture representation of dynamic imaging data from PET and MR studies. Biostatistics. 2006;7: 318–38.

40. O'Sullivan F, Muzi M, Mankoff DA, Eary JF, Spence AM, Krohn KA. Voxel-level mapping of tracer kinetics in PET studies: a statistical approach emphasizing tissue life tables. Ann Appl Stat. 2014;8:1065.