Co-clinical FDG-PET Radiomic Signature in Predicting Response to Neoadjuvant Chemotherapy in Triple Negative Breast Cancer

Sudipta Roy[1], Timothy D. Whitehead[1], Shunqiang Li[2], Foluso O. Ademuyiwa[2], Richard L. Wahl[1,3], Farrokh Dehdashti[1], Kooresh I. Shoghi[1,4]

[1]Department of Radiology, [2]Department of Medicine, Division of Oncology, and [3]Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO, USA;
[4]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA

Corresponding author:
Kooresh Isaac Shoghi, Ph.D.
Department of Radiology
510 S. Kingshighway Blvd.
Campus Box 8225
St. Louis, MO 63110
Tel.: 314-362-8990
Email: shoghik@wustl.edu

ABSTRACT

Purpose. We sought to exploit the heterogeneity afforded by patient-derived tumor xenografts (PDX) to optimize robust radiomic features associated with response to therapy in the context of a co-clinical trial and implement PDX-optimized image features in the corresponding clinical study to predict and assess response to therapy using machine-learning (ML) algorithms.

Methods. TNBC patients and subtype-matched PDX were recruited into a co-clinical FDG-PET imaging study to predict response to therapy. One hundred thirty-one imaging features were extracted from PDX and human segmented tumors. Robust image features were identified based on reproducibility, cross-correlation, and volume independence. A rank importance of predictors using ReliefF was used to identify predictive radiomic features in the preclinical PDX trial in conjunction with ML algorithms: classification and regression tree (CART), Naïve Bayes (NB), and support vector machines (SVM). The top four PDX-optimized image features, defined as radiomic signatures (RadSig), from each task were then used to predict or assess response to therapy. Performance of RadSig in predicting/assessing response was compared to SUV_{mean}, SUV_{max}, and lean body mass normalized SUL_{peak} measures.

Results. Sixty-four out of 131 preclinical imaging features were identified as robust. NB-RadSig performed highest in predicting and assessing response to therapy in the preclinical PDX trial. In the clinical study, the performance of SVM-RadSig and NB-RadSig to predict and assess response was practically identical and superior to SUV_{mean}, SUV_{max}, and SUL_{peak}, measures.

Conclusions. We optimized robust FDG-PET radiomic signatures (RadSig) to predict and assess response to therapy in a context of a co-clinical imaging trial.

KEYWORDS: triple negative breast cancer (TNBC); FDG-PET; radiomics; co-clinical imaging; quantitative imaging; machine learning

DECLARATIONS

Funding. This work was supported by NCI grants U24CA209837, U24CA253531, and U54CA224083; U2CCA233303, and K12CA167540; Siteman Cancer Center (SCC) Support Grant P30CA091842; and Internal funds provided by Mallinckrodt Institute of Radiology.

Conflicts of interest/Competing interests. None.

Availability of data and material. All the co-clinical data will be available for download through the Washington University School of Medicine Co-Clinical Imaging Research Resource web portal at https://c2ir2.wustl.edu/, co-clinical database (CCDB).

Code availability. Not applicable.

Authors' contributions. Conceptualization: SR, FOA, KIS; Methodology: SR, TDW, SL, KIS; Formal analysis and investigation: SR, KIS; Writing - original draft preparation: SR; Writing - review and editing: RLW, FD, KIS; Funding acquisition: RWL, FOA, SL, KIS; Resources: SL; Supervision: FD, KIS. All authors read and approved the final manuscript.

Ethics approval. All studies were performed with approval from the Washington University Humans subjects research committee and animal studies committee.

Consent to participate. Informed consent to participate in the study was obtained from all participants.

Consent for publication. Not applicable.

INTRODUCTION

Triple Negative Breast Cancer (TNBC) is a highly heterogeneous and aggressive cancer characterized by poor outcome and higher relapse rates compared to other subtypes of breast cancer. Pathological complete response (pCR) is often used as a critical endpoint in the treatment of TNBC following NAC as it is often associated with favorable long-term outcome. Therefore, it is critical to identify patients who will respond to NAC therapy to avoid the use of ineffective treatments. Intratumoral heterogeneity is regarded as a major factor in tumor progression and resistance to NAC [1]. Towards that end, advanced quantitative imaging (QI) strategies, including extraction of image features, or radiomics, have been employed to characterize tumor heterogeneity and to predict/assess response to therapy [2, 3].

We designed a co-clinical trial to assess the efficacy of docetaxel/carboplatin therapy in patients with TNBC and patient-derived tumor xenografts (PDX) generated from TNBC patient biopsies. Co-clinical trials are an emerging area of investigation in which a clinical trial is coupled with a corresponding preclinical trial to inform the corresponding clinical trial [4-10]. The emergence of PDXs as co-clinical platforms is largely motivated by the realization that established cell-lines do not recapitulate the heterogeneity of human tumors and the diversity of tumor phenotypes [11]. Indeed, numerous investigations have demonstrated that PDX accurately reflect patients' tumors in terms of the histomorphology, gene expression profiles, and gene copy number alterations [12-16], as well as ability to predict therapeutic response in patients, especially when a clinically relevant drug dosage is used [17-19]. To that end, the National Cancer Institute's (NCI) Patient-Derived Models Repository (https://pdmr.cancer.gov), EuroPDX (https://www.europdx.eu), academic institutions, and numerous commercial entities have launched wide-ranging PDX and repositories to advance the use of PDX in oncologic applications.

One of the objectives of the co-clinical trial, which is still underway, is to predict response to therapy using [^{18}F]fluorodeoxyglucose (FDG) with positron emission tomography (PET). We previously identified six TNBC subtypes including 2 basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), a mesenchymal stem–like (MSL), and a luminal androgen receptor (LAR) subtype through molecular signatures of TNBC subtypes [20]. The use of PDX in preclinical imaging offers numerous advantages in translational imaging research, chief among them is retention of human tumor heterogeneity [12, 16, 21], which can be exploited to develop image metrics of response to therapy. Thus, the primary objective of this work was to utilize PDX to optimize robust radiomic features of tumor heterogeneity indicative of response to therapy in preclinical PDX trials.

The scheme outlined in Figure 1 highlights the paradigm we undertook in this effort. We used the co-clinical imaging trial to define, for the first-time, parallels in radiomic features between preclinical and clinical imaging. To address the primary objective, we characterized the reproducibility, cross correlation (auto-correlation), and volume dependency of FDG-PET radiomic features in PDX. Optimal radiomic features were then used in ML algorithms to define radiomic signatures (RadSig) of response to therapy in the preclinical PDX trial. With the RadSig at hand, we performed an interim analysis to implement radiomic signatures optimized in the preclinical PDX trial to predict response to therapy in the clinical arm. Our findings suggest that RadSig performed significantly better than standardized uptake value (SUV) measures to predict (using baseline metrics) and assess (difference in image metrics) response to therapy.

METHODS

Co-clinical protocol

The co-clinical design is outlined in the scheme of Figure 2A and described below. TNBC PDX subtypes were identified as described previously [22] based on molecular signature analysis. TNBC subtypes include: basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), and a luminal androgen receptor (LAR) subtype.

Preclinical imaging. Small animal PET/CT was performed on the Inveon microPET/CT scanner as described previously [22]. Briefly, four hours prior to imaging session, food was removed from metabolism cages while water was given ad libitum. Mice were anesthetized with 2-2.5% isoflurane by inhalation via an induction chamber. Anesthesia was maintained throughout the imaging session by delivering 1%–1.5% isoflurane via a custom-designed nose cone. A heat lamp was used to maintain body temperature. TNBC PDX were injected with ^{18}FDG (6.66 – 8.14 MBq) by tail vein immediately before a 0-60 min dynamic small animal PET acquisition. Images were reconstructed with a 3D OSEM algorithm with a Ramp filter at 0.5 cutoff and voxel size of ~0.08mm. In therapeutic studies, TNBC PDX (N=29) were imaged at baseline (BL) and four days (4D) following start of therapy (Figure 2A). Docetaxel (20mg/kg IP)/carboplatin (50mg/kg IP) was administered following BL imaging and weekly for a period of four weeks. Tumor volumes were measured bi-weekly. All animal experiments were conducted in compliance with the Guidelines for the Care and Use of Research Animals established by Washington University's Animal Studies Committee.

Clinical imaging. Twenty stage II or III TNBC patients were recruited into an ongoing co-clinical trial (ClinicalTrial.gov ID # NCT02124902). A secondary goal of the co-clinical trial was to assess the performance of FDG-PET in predicting/assess response to therapy. Simultaneous FDG-PET and MR imaging protocols were implemented on the Siemens Biograph mMR. Subjects were imaged at baseline (BL) prior to therapy and between the first cycle (C1) and second cycle of docetaxel/carboplatin for a total of 6 cycles (21 days per cycle). At each imaging time point, patients fasted for ~4hrs prior to injection of ~10mCi of FDG. After an uptake period, patients were positioned prone on the PET/MR scanner. FDG-PET imaging was performed starting at 30min to 70min post FDG administration. Default Dixon sequence was used for attenuation correction. Images were reconstructed to produce four 10min frames. In parallel with FDG-PET acquisition, T1-weighted (T1w) and T2-weighted (T2w) MR acquisitions were performed. The co-clinical trial is ongoing; however, we performed an interim analysis to assess the performance of PDX-optimized FDG-PET image features to predict/assess response to therapy in the clinical arm. Quantification of clinical MR data is not included at this stage since preclinical imaging did not include MR acquisition.

Image analysis and extraction of radiomic features

Preclinical imaging: Static 10min PET/CT images obtained 50min post-administration of FDG (representative image in Figure 2B) were processed in two-steps. In the first step, co-registered PET/CT images were analyzed using the Inveon Research Workplace (IRW) software (Siemens Healthcare). Volumes of interest (VOIs) were manually drawn on co-registered PET/CT images to include tumor(s). Second, VOIs and individual voxels were normalized to SUV in MATLAB using the relation: SUV = [activity (Bq / mL)] x [animal weight (g)] / [injected dose (Bq)].

Clinical imaging: Tumor VOIs were manually drawn on 20min static PET images obtained by averaging two 10min frames 50-70min post-administration of FDG (representative image in Figure 2C). To ensure harmonization of preclinical and clinical pipelines, IRW was used to segment tumors on PET/MR images. Mean SUV (SUV_{mean}) for the entire tumor was calculated as per above. Peak SUV was normalized to lean body mass (SUL_{peak}) based on positron emission tomography response criteria in solid tumors (PERCIST) [23].

Extraction of imaging features. One hundred thirty-one imaging features were extracted from preclinical and clinical tumors. These include: one hundred twenty radiomic features, tumor volume, metabolic tumor volume, and nine SUV metrics as tabulated in Supplemental Table S1.

Radiomic features were determined per the image biomarker standardization initiative (IBSI) guidelines [24, 25]. Equal-probability quantization algorithms to quantize raw data into gray level (Ng) were implemented using histeq MATLAB function. Resampling to isotropic voxel size in all three directions was applied to all higher order features. Thirty-seven first order features were extracted directly from raw data. All higher order features were extracted after applying fixed quantization of gray level Ng=64.

Robustness of radiomic features

We evaluated the robustness of radiomic features in terms of reproducibility (test-retest), cross-correlation, and the dependency on tumor volume. Robust radiomic features were then used as predictors of response to therapy.

Test-retest. A preclinical test-retest protocol was implemented to optimize the reproducibility of radiomics features. PDX (N=40) were imaged on consecutive days (Day 1 and Day 2) in identical conditions.

Cross-correlation. The cross-correlation between features was determined using Spearman correlation. A threshold Spearman correlation of $\rho \geq 0.9$ and significance value $P<0.001$ were chosen as significantly high correlation between features.

Volume-dependent radiomic features. Radiomic features were regressed against their corresponding tumor volumes. Linear or nonlinear functional forms were used to fit all significant volume-dependent features.

Prediction and assessment of response to therapy

Prediction vs. assessment of response to therapy. We make a distinction between predicting and assessing response to therapy. In predicting response to therapy, BL imaging features were used to predict response to therapy in either the preclinical or the clinical arm. In assessing response to therapy, the change (Δ) in imaging feature between on-treatment (4D in preclinical and post C1 in clinical) and BL was used to predict response to therapy in either the preclinical or the clinical arm.

Classification of response to therapy. In preclinical studies, end-point caliper volume change from start of treatment was considered as surrogate of response to therapy with response to therapy corresponding to >20% decrease in volume; partial response corresponding to ≤|20|% change in volume, and no response corresponding to >20% increase in volume. Baseline radiomic features and change in radiomic features between 4D post-treatment and baseline scans were used as the predictive criterion for ML algorithms. In clinical studies, pCR was used to determine response to therapy.

Feature selection. In preclinical studies, the Relief-based algorithm (RBA) [26] was used to select a subset of features as inputs to the ML algorithms. A relevance threshold (τ=0.05) [27] was used to select most relevant weighted features to facilitate in expansive modeling, reduce overfitting, and make the task tractable for inputs in ML algorithms. These optimal features were used to predict response or assess response to therapy using BL and difference between on-treatment and BL optimal features, respectively.

Machine learning for outcome prediction. The ML algorithms used in this study include CART [28], SVM [29], and NB [30]. In CART, Gini index was used at each partition to determine splitting criteria with a binary threshold of CART. In implementing SVM, radial Basis Function (RBF) kernel was used to make the hyperplane decision boundary between the classes. Objective function L2-norm regularization was used to overcome overfitting problem. CART, SVM, and NB work well with datasets as low as N=20 [31]. Ten-fold cross-validation was used to avoid overfitting the ML model [32].

Statistical Analyses

Robustness of features. Lin's concordance correlation coefficient (LCC) [33] was used to assess reproducibility using Stata version 12.1. LCC≥0.7 was considered as a threshold of reproducible radiomic feature [34, 35]. As indicated above, cross-correlation between features was evaluated using the Spearman correlation $\rho \geq 0.9$ at significance value $P<0.001$. To display clusters of correlations, hierarchical clustering of the Spearman correlation heatmap was performed. In evaluating volume-dependency of features, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were calculated for each functional form, and the appropriate model was selected based on the minimum value of AIC and BIC. The Spearman correlation (ρ) was used to determine the correlation between each feature and tumor volume.

Performance metrics of response to therapy prediction. Common performance metrics including accuracy, F-score, sensitivity, specificity, precision, and negative predictive value (NPV) were used to assess performance of response to therapy [20]. The performance of the radiomic features was additionally compared with SUV_{mean}, SUV_{max}, and SUL_{peak} based on PERCIST [23].

RESULTS

Reproducibility of preclinical radiomic features

Test-retest was performed to assess the reproducibility of radiomic features using LCC as a measure of reproducibility. Ninety-four out of 129 radiomic features (72.9%) were identified as reproducible with LCC≥0.7. The frequency of correlations along with the cumulative percent is displayed in Figure 3A. Approximately 22% of features were highly reproducible with LCC≥0.9. The reproducibility by class of features is depicted in Figure 3B. Figures 3C depicts the LCC values of all reproducible radiomic features. Supplemental Table S1 summarizes the reproducibility of all 131 features.

Cross-correlation between features (preclinical and clinical)

We ascertained the cross-correlation between 129 features using the Spearman correlation (ρ). Highly correlated features (ρ≥0.9) were removed and reduced to 94 features from 129 features. Hierarchical clustering of the Spearman correlation heatmap is shown Figure 4. Twenty-one clusters were identified in the preclinical heatmap (Figure 4A) and similarly 21 clusters were identified in the clinical cross-correlation heatmap (Figure 4B). Membership of features to clusters is available in Supplemental Table S2. The distribution of Spearman correlations is available in Figure 4C and 4D for preclinical and clinical cross-correlations, respectively.

Volume-dependent radiomic features (preclinical and clinical)

In total, 10 radiomic features were highly correlated to volume ($\rho>0.9$; $P<0.001$). The functional form of the volume dependency and corresponding goodness-of-fit measures for preclinical and corresponding clinical images is shown in Figure 5, which was similar for both preclinical and clinical features. Supplemental Table S3 summarizes the statistical analyses for the correlations.

Prediction and assessment of response to therapy

At the intersection of robustness analyses, 62 of the 129 (48.06%) features were found to be optimal and were passed to ReliefF feature selection followed by ML. ReliefF rank importance identified top performing 15 features for prediction (based on BL features) and assessment (based on 4D-BL features) of response to therapy (Figures 6B and 6C, respectively). The rank importance of radiomics features is given in Supplemental Table S4.

Preclinical PDX studies

The accuracy of ML in predicting/assessing response to therapy as a function of the number of radiomic features is depicted in Figures 6D and 6E for BL and 4D-BL, respectively. The number of radiomic features to maximize prediction accuracy saturated at 4 features (Figure 6D) with NB exhibiting the highest accuracy at 86.21 %, followed by SVM, and CART. In contrast, the accuracy of assessing response to therapy (4D-BL) increased with increasing number of radiomic features; the accuracy of NB is 86.9% followed by SVM and CART (Figure 6E). We opted to compare performance between prediction and assessment (i.e., BL versus 4D-BL) using the least number of robust features. For this reason, Table 1 tabulates the performance of the ML algorithms to predict/assess response prior to and following optimization for robust features using only the top 4 radiomic features for each classification (prediction versus assessment of response). The set of 4 radiomic features from each task (prediction and assessment of response) make up the Radiomic Signature (RadSig). As tabulated in Table 1, RadSig performs as well as, or marginally better than, non-optimized features (all features) in predicting response. The performance of prediction/assessment of response to therapy stratified by TNBC subtype is tabulated in Supplemental Table S5 and highlights differences in prediction by TNBC subtype.

Table 1: Accuracy of predicting (BL) and assessing (4D-BL) response to therapy using top 4 radiomic features.

	All features		RadSig	
Methods	Prediction	Assessment	Prediction	Assessment
CART	80.34	74.86	78.48	72.57
Naïve Bayes	82.62	82.76	86.21	78.26
SVM	78.48	78.45	81.14	75.13

The performance of RadSig in comparison to SUV_{mean}, SUV_{max}, and SUL_{peak} for the top two performing ML algorithms (NB and SVM) is summarized in Figure 7. NB performed marginally better than SVM in predicting/assessing response to therapy (Figure 7A) in the preclinical PDX trial. The percent increase in predicting/assessing response to therapy relative to SUV_{mean}, SUV_{max}, and SUL_{peak} is depicted in Figure 7B for NB. NB-RadSIG improved prediction of response by over 60% in all performance measures. In assessing response to therapy, RadSig performed better than SUV_{mean} in most performance criteria and marginally better than SUL_{peak} and SUV_{max}

(Figure 7B). Thus, RadSig has greater impact in predicting response to therapy than assessing response to therapy. Full performance data is available is Supplemental Table S6. We then performed an interim analysis of the ongoing clinical trial to implement PDX-optimized RadSig to predict/assess response to therapy using ML.

Table 2 summarizes patient characteristics, pathological response, SUV metrics at BL, and percent change in SUV metrics between on-treatment (post C1) and BL for the interim analyses. Of the twenty patients, ten patients exhibited pCR; however, all patients exhibited reduction in SUV. Average percent (±1SD) reduction in the non-pCR group was -46.94±21.56, -53.20±19.91, and -51.33±19.78 for SUV$_{mean}$, SUL$_{peak}$, and SUV$_{max}$, respectively; and -57.70±14.83, -60.32±16.47, and -66.16±13.74 for SUV$_{mean}$, SUL$_{peak}$, and SUV$_{max}$, respectively. Figure 7 also depicts the performance of the ML algorithms in predicting and assessing response to therapy in the clinical arm (Figure 7C). The performance of SVM and NB with RadSig as a predictor were marginally similar, although overall SVM performed better than NB when using SUV metrics as predictors (Supplemental Table S6). SVM-RadSig exhibited higher prediction rates of response to therapy relative to SUV$_{max}$, SUV$_{mean}$ and SUL$_{peak}$ in all performance measures (20-40% higher), as well as in assessing response to therapy (15-75% higher) (Figure 7D). Overall, RadSig performed better than SUV metrics in predicting and assessing response to therapy.

Table 2: Patient characteristics, pathologic response, and SUV metrics

Stage at Diagnosis	Grade at Diag.	pCR	BL SUV$_{mean}$	BL SUL$_{P}$	BL SUV$_{max}$	%Δ SUV$_{mean}$	%Δ SUL$_{peak}$	%Δ SUV$_{max}$
IIB (T2N1)	2	No	1.86	1.57	4.03	NA	NA	NA
IIB (T2N1)	3	Yes	6.59	4.58	11.27	-60.97	-68.11	-57.82
IIIA (T3N1)	3	No	10.55	12.34	26.41	NA	NA	NA
IIA (T2N0)	?	Yes	8.77	7.80	21.48	-75.49	-82.21	-86.77
IIB (T2N1)	3	No	3.18	1.64	5.31	-4.32	-13.19	-9.40
IIB (T2N1)	3	No	2.92	1.29	4.54	-38.67	-41.17	-41.77
			8.01	6.43	15.55	-40.82	-44.37	-42.90
IIB (T2N1)	3	No	8.58	6.20	17.02	-64.92	-70.33	-73.40
			8.91	6.62	20.11	-62.72	-68.17	-61.72
IIA (T2N0)	3	Yes	2.49	1.61	5.25	NA	NA	NA
IIA (T2N0)	3	Yes	4.17	3.06	12.45	-48.84	-48.53	-64.23
IIA (T2N0)	3	Yes	3.89	3.12	8.62	-57.76	-59.91	-73.72
IIB (T2N1)	3	No	2.85	2.59	6.28	-41.45	-55.68	-55.97
			2.48	2.03	5.03	-27.04	-37.81	-49.88
IIA (T2N0)	3	No	8.66	7.56	21.87	-75.82	-82.48	-81.35
IIB (T2N1)	3	Yes	10.29	8.32	28.33	-74.14	-76.39	-81.07

			3.64	2.02	5.65	-58.93	-57.11	-55.17
			8.00	5.60	16.06	-71.75	-77.22	-77.14
IIA (T2N0)	3	Yes	1.86	1.65	4.18	-30.42	-36.91	-53.02
IIA (T2N0)	3	No	9.80	7.74	17.69	-67.65	-64.44	-54.15
IIA (T2N0)	3	Yes	7.19	5.21	14.57	-46.36	-42.49	-43.23
IIA (T2N0)	3	No	7.98	7.26	18.46	-45.99	-54.32	-42.74
IIA (T2N0)	3	Yes	2.91	2.03	5.24	-40.23	-40.04	-58.66
IIB (T2N1)	3	Yes	6.19	5.53	11.28	-69.81	-74.57	-76.91
IIB (T2N1)	3	No	2.44	1.99	5.53	NA	NA	NA

DISCUSSION

The emergence of co-clinical models is largely motivated by the realization that established cell-lines do not recapitulate the heterogeneity of human tumors and the diversity of tumor phenotypes [11] and that better oncology models are needed to support high-impact translational cancer research [12, 16, 21]. An underlying premise in the co-clinical study design is that the heterogeneity of the human tumor is retained in PDX. Indeed, tumor genomic and pathological investigations have confirmed that PDX recapitulate the heterogeneity of human tumors [12-16] and that these can be used to a better inform cancer biology, therapeutic design [17-19], and therefore by extension imaging studies, albeit with some limitations [21]. With that in mind, in this this work, we exploited the heterogeneity of TNBC PDX subtypes to 1) identify robust radiomic features in preclinical TNBC PDX; 2) optimize RadSig-ML algorithms to predict response to therapy in PDX; and 3) implement PDX-optimized RadSig to predict/assess response to therapy in the corresponding clinical trial.

To our knowledge, this study represents the first such effort to optimize radiomic features in preclinical PET imaging to predict/assess response to therapy in TNBC PDX. We recently characterized the dependency of preclinical MR radiomic features on tumor volume [36]. In this work, we confirmed dependency of preclinical PET radiomic features on tumor volume with strikingly similar clinical parallels. This is particularly relevant in longitudinal studies during which tumor volumes will change with the course of the disease or following therapy. Ideally volume-independent features should be used as to not bias image features longitudinally. We further evaluated the cross-correlation of preclinical and clinical radiomic features with the goal of reducing the dimensionality of features. Finally, we evaluated the repeatability of radiomic features in preclinical PET imaging to identify robust features for inclusion in ML-based prediction of response to therapy. At the thresholds defined within to screen for volume-dependency, repeatability, and cross-correlation, we identified 62 optimal features to predict/assess response to therapy.

RBF [26] was used to rank image features using three ML algorithms as to their relevance in predicting/assessing response to therapy. Our data suggests that overall SVM performed better than NB and CART in predicting response to therapy. We used the top four ML-RBF-optimized radiomic features—referred to as radiomic signature (RadSig)—from each task (prediction vs. assessment) to either predict or assess response to therapy. RadSig performed significantly better in predicting response to therapy in the preclinical and clinical arm, as well as in assessing response to therapy in the clinical arm. Antunovic et al. [37] reported the utility of FDG-PET radiomic features to assess response to therapy using four different models in 79 patients with heterogenous breast cancer subtypes. The reported area-under the curve of an ROC analysis ranged from 0.70 to 0.73. Li et al. [38] recently assessed the utility of both PET and CT radiomic features to predict response to therapy in a retrospective study that included 100 heterogenous breast cancer patients. The PET/CT radiomic predictors achieved a prediction accuracy of 87% on the training split set and 77% on the independent validation set.

In the small, albeit homogenous, dataset of TNBC patients where PDX-optimized radiomic features were implemented in the clinical imaging arm, we observed an impressive accuracy of 72% and 71% when predicting and assessing response, respectively, compared to SUV metrics. We were unable to perform a validation test on an independent dataset; however, the primary objective was to compare the performance of all predictive metrics in the training phase. In addition, we did not report MR radiomic features in this work. However, prediction of response to therapy can be further enhanced through integration of MR imaging features [39], liquid biopsies such as circulating tumor DNA (ctDNA) analyses [40], and molecular/genomic features of tumors [41], all of which are an active area of investigation. Finally, numerous recent studies have documented that pCR rates varied with breast cancer molecular subtypes. TNBC and HER2-positive molecular subtypes have shown to have higher pCR rates after NAC [42]. Importantly, several studies have demonstrated an association between imaging features and molecular phenotypes, risk of recurrence, and prognosis [43-45]. Interestingly, our PDX studies similarly suggest that response to therapy (and prediction thereof) is a function of the TNBC subtype, however further studies are needed to support this hypothesis and the utility of radiomic features in classifying TNBC subtypes. With that in mind, one of the most critical aspects in practical implementation of radiomics is a consensus on the most effective features and their standardization.

CONCLUSIONS

We identified robust FDG-PET radiomic features from an ongoing co-clinical (PDX and human) trial to predict and assess response to therapy. The number of radiomic features to maximize accuracy was further optimized in the preclinical PDX trial to yield ML radiomic signatures (RadSig) of response to therapy. We then implemented RadSig in an interim analysis of the corresponding clinical trial. The performance of SVM-RadSig in predicting/assessing response to therapy was superior to SUV_{max}, SUV_{mean} and SUL_{peak} metrics in the clinical setting; however, given the small sample size additional studies are warranted to further validate RadSig and potentially integrate with multi-scale features to enhance prediction/assessment of response to therapy.

Acknowledgements

The authors acknowledge the staff of the Preclinical Imaging Facility and the Center for Clinical Imaging Research (CCIR) at Mallinckrodt Institute of Radiology (MIR), Washington University School of Medicine for performing imaging studies.

Conflict of interest

No potential conflicts of interest relevant to this article exist.

FIGURE LEGENDS

Figure 1: Overview of methodology in optimizing radiomic features in the co-clinical trial. TNBC PDX were generated from human tumor biopsies. Tumor were segmented following co-clinical imaging to extract radiomic features. Radiomic features were extracted per IBSI guidelines. Repeatability, cross-correlation, and volume-dependency were performed to identify the robust features. ReliefF, and then ML were used to predict/assess the response to therapy in PDX and to identify radiomic signature (RadSig). RadSig was implemented in the clinical trial to predict/assess response to therapy.

Figure 2: Co-clinical imaging. (A) Co-clinical imaging study design. Stage II or III TNBC patients were recruited into the study for 6 cycles of docetaxel/carboplatin therapy. Imaging timepoints are indicated on the timeline. PDX are generated from patient tumor biopsies to assess response to therapy with imaging at baseline and +4 days post-therapy. (B) representative preclinical PET/CT and (C) clinical PET/MR images of response to therapy. Tumor is indicated by white arrow.

Figure 3: Repeatability of preclinical radiomic features. (A) Frequency (bar plot, left Y-axis) of repeatability by LCC values and cumulative percent (solid line, right Y-axis). (B) Percent of repeatable features with LCC≥0.7 by class (left Y-axis labels). (C) Reproducible radiomic features with LCC≥0.7 (denoted by dashed vertical black line).

Figure 4: **Cross-correlation between radiomic features**. Hierarchical clustering of the Spearman cross-correlation heatmap for (A) preclinical and (B) clinical. The dendrogram shows the clustering, and each color represents a different cluster. The frequency and cumulative sum at each Spearman correlation (ρ) is displayed in (C) for preclinical and (D) for clinical.

Figure 5: Volume-dependent preclinical and clinical radiomic features. The correlation between features and tumor volume was assessed for preclinical and clinical segmented tumors. Ten common preclinical and clinical features exhibited high correlation (ρ≥0.9) with tumor volume.

Figure 6. ML-based selection of radiomic features. (A) At the thresholds defined within to screen for volume-dependency, repeatability, and cross-correlation, we identified 62 optimal features. Implementation of relief-based algorithm (RBA) to identify a subset of features as inputs to ML-based prediction (B) and assessment (C) of response to therapy. Accuracy of ML algorithms CART, SVM, and Naïve Bayes to predict (D) or assess (E) response to therapy as a function of number of radiomic features in PDX.

Figure 7: Performance of ML algorithms. (A) Performance of RadSig in predicting/assessing response to therapy in the preclinical PDX trial with NB and SVM. (B) Percent improvement in NB-RadSig prediction/assessment of response relative to SUV_{max}, SUV_{mean} and SUL_{peak}. (C) similar to (A) but for the clinical investigation. (D) similar to (B) but using SVM.

References:

1. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. Cancer Cell. 2020;37:471-84. doi:10.1016/j.ccell.2020.03.007.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278:563-77. doi:10.1148/radiol.2015151169.
3. 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. Nature communications. 2014;5:4006. doi:10.1038/ncomms5006.
4. Chen Z, Akbay E, Mikse O, Tupper T, Cheng K, Wang Y, et al. Co-clinical trials demonstrate superiority of crizotinib to chemotherapy in ALK-rearranged non-small cell lung cancer and predict strategies to overcome resistance. Clin Cancer Res. 2014;20:1204-11. doi:10.1158/1078-0432.Ccr-13-1733.
5. Kim HR, Kang HN, Shim HS, Kim EY, Kim J, Kim DJ, et al. Co-clinical trials demonstrate predictive biomarkers for dovitinib, an FGFR inhibitor, in lung squamous cell carcinoma. Ann Oncol. 2017;28:1250-9. doi:10.1093/annonc/mdx098.
6. Kwong LN, Boland GM, Frederick DT, Helms TL, Akid AT, Miller JP, et al. Co-clinical assessment identifies patterns of BRAF inhibitor resistance in melanoma. J Clin Invest. 2015;125:1459-70. doi:10.1172/jci78954.
7. Lunardi A, Ala U, Epping MT, Salmena L, Clohessy JG, Webster KA, et al. A co-clinical approach identifies mechanisms and potential therapies for androgen deprivation resistance in prostate cancer. Nature genetics. 2013;45:747-55. doi:10.1038/ng.2650.
8. Nishino M, Sacher AG, Gandhi L, Chen Z, Akbay E, Fedorov A, et al. Co-clinical quantitative tumor volume imaging in ALK-rearranged NSCLC treated with crizotinib. Eur J Radiol. 2017;88:15-20. doi:10.1016/j.ejrad.2016.12.028.
9. Owonikoko TK, Zhang G, Kim HS, Stinson RM, Bechara R, Zhang C, et al. Patient-derived xenografts faithfully replicated clinical outcome in a phase II co-clinical trial of arsenic trioxide in relapsed small cell lung cancer. Journal of translational medicine. 2016;14:111. doi:10.1186/s12967-016-0861-5.
10. Sia D, Moeini A, Labgaa I, Villanueva A. The future of patient-derived tumor xenografts in cancer treatment. Pharmacogenomics. 2015;16:1671-83. doi:10.2217/pgs.15.102.
11. Sulaiman A, Wang L. Bridging the divide: preclinical research discrepancies between triple-negative breast cancer cell lines and patient tumors. Oncotarget. 2017;8:113269-81. doi:10.18632/oncotarget.22916.
12. DeRose YS, Wang G, Lin YC, Bernard PS, Buys SS, Ebbert MT, et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. Nat Med.17:1514-20. doi:nm.2454 [pii]10.1038/nm.2454.

13. Zhao X, Liu Z, Yu L, Zhang Y, Baxter P, Voicu H, et al. Global gene expression profiling confirms the molecular fidelity of primary tumor-based orthotopic xenograft mouse models of medulloblastoma. Neuro-Oncology.14:574-83. doi:nos061 [pii] 10.1093/neuonc/nos061.

14. Morton CL, Houghton PJ. Establishment of human tumor xenografts in immunodeficient mice. Nat Protoc. 2007;2:247-50. doi:nprot.2007.25 [pii] 10.1038/nprot.2007.25.

15. Reyal F, Guyader C, Decraene C, Lucchesi C, Auger N, Assayag F, et al. Molecular profiling of patient-derived breast cancer xenografts. Breast Cancer Research.14:R11. doi:bcr3095 [pii] 10.1186/bcr3095.

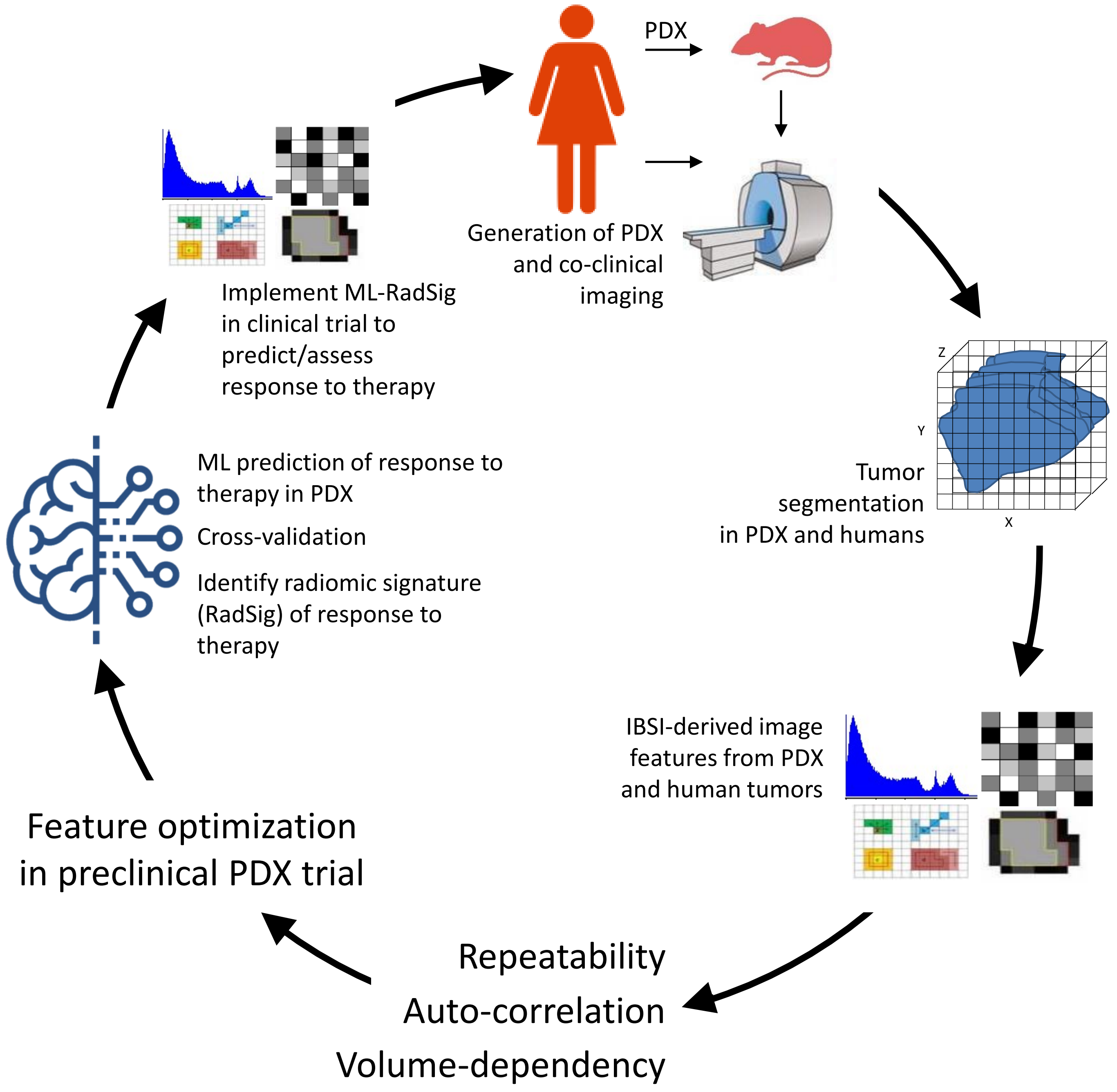
16. Krepler C, Xiao M, Spoesser K, Brafford PA, Shannan B, Beqiri M, et al. Personalized pre-clinical trials in BRAF inhibitor resistant patient derived xenograft models identify second

line combination therapies. Clinical Cancer Research. 2015. doi:10.1158/1078-0432.CCR-15-1762.

17. Kerbel RS. Human tumor xenografts as predictive preclinical models for anticancer drug activity in humans: better than commonly perceived-but they can be improved. Cancer biology & therapy. 2003;2:S134-9.
18. Johnson JI, Decker S, Zaharevitz D, Rubinstein LV, Venditti JM, Schepartz S, et al. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. Br J Cancer. 2001;84:1424-31. doi:10.1054/bjoc.2001.1796.
19. Scholz CC, Berger DP, Winterhalter BR, Henss H, Fiebig HH. Correlation of drug response in patients and in the clonogenic assay with solid human tumour xenografts. Eur J Cancer. 1990;26:901-5.
20. Savaikar MA, Whitehead T, Roy S, Strong L, Fettig N, Prmeau T, et al. Preclinical PERCIST and 25% of SUVmax Threshold: Precision Imaging of Response to Therapy in Co-clinical (18)F-FDG PET Imaging of Triple-Negative Breast Cancer Patient-Derived Tumor Xenografts. J Nucl Med. 2020;61:842-9. doi:10.2967/jnumed.119.234286.
21. Shoghi KI, Badea CT, Blocker SJ, Chenevert TL, Laforest R, Lewis MT, et al. Co-Clinical Imaging Resource Program (CIRP): Bridging the Translational Divide to Advance Precision Medicine. Tomography. 2020;6:273-87. doi:10.18383/j.tom.2020.00023.
22. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121:2750-67. doi:10.1172/JCI45014.
23. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med. 2009;50 Suppl 1:122S-50S. doi:10.2967/jnumed.108.057307.
24. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295:328-38. doi:10.1148/radiol.2020191145.
25. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. Ieee T Syst Man Cyb. 1973;Smc3:610-21. doi:Doi 10.1109/Tsmc.1973.4309314.
26. Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn. 2003;53:23-69. doi:Doi 10.1023/A:1025667309714.
27. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. J Biomed Inform. 2018;85:189-203. doi:10.1016/j.jbi.2018.07.014.
28. Cheng Z, Nakatsugawa M, Hu C, Robertson SP, Hui X, Moore JA, et al. Evaluation of classification and regression tree (CART) model in weight loss prediction following head and neck cancer radiation therapy. Adv Radiat Oncol. 2018;3:346-55. doi:10.1016/j.adro.2017.11.006.
29. Huang MW, Chen CW, Lin WC, Ke SW, Tsai CF. SVM and SVM Ensembles in Breast Cancer Prediction. Plos One. 2017;12. doi:ARTN e016150110.1371/journal.pone.0161501.
30. Gao HY, Zeng X, Yao CH. Application of improved distributed naive Bayesian algorithms in text classification. J Supercomput. 2019;75:5831-47. doi:10.1007/s11227-019-02862-1.
31. Murali N, Kucukkaya A, Petukhova A, Onofrey J, Chapiro J. Supervised Machine Learning in Oncology: A Clinician's Guide. Dig Dis Interv. 2020;4:73-81. doi:10.1055/s-0040-1705097.
32. Ha S, Choi H, Paeng JC, Cheon GJ. Radiomics in Oncological PET/CT: a Methodological Overview. Nuclear medicine and molecular imaging. 2019;53:14-29. doi:10.1007/s13139-019-00571-4.

33. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989;45:255-68.
34. Groutz A, Blaivas JG, Chaikin DC, Resnick NM, Engleman K, Anzalone D, et al. Noninvasive outcome measures of urinary incontinence and lower urinary tract symptoms: a multicenter study of micturition diary and pad tests. J Urol. 2000;164:698-701. doi:10.1097/00005392-200009010-00019.
35. Matheson GJ. We need to talk about reliability: making better use of test-retest studies for study design and interpretation. PeerJ. 2019;7:e6918. doi:10.7717/peerj.6918.
36. Roy S, Whitehead TD, Quirk JD, Salter A, Ademuyiwa FO, Li S, et al. Optimal co-clinical radiomics: Sensitivity of radiomic features to tumour volume, image noise and resolution in co-clinical T1-weighted and T2-weighted magnetic resonance imaging. EBioMedicine. 2020;59:102963. doi:10.1016/j.ebiom.2020.102963.
37. Antunovic L, De Sanctis R, Cozzi L, Kirienko M, Sagona A, Torrisi R, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. Eur J Nucl Med Mol Imaging. 2019;46:1468-77. doi:10.1007/s00259-019-04313-8.
38. Li P, Wang X, Xu C, Liu C, Zheng C, Fulham MJ, et al. (18)F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. Eur J Nucl Med Mol Imaging. 2020;47:1116-26. doi:10.1007/s00259-020-04684-3.
39. Hu Q, Whitney HM, Giger ML. Radiomics methodology for breast cancer diagnosis using multiparametric magnetic resonance imaging. J Med Imaging (Bellingham). 2020;7:044502. doi:10.1117/1.JMI.7.4.044502.
40. Magbanua MJM, Swigart LB, Wu HT, Hirst GL, Yau C, Wolf DM, et al. Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. Ann Oncol. 2021;32:229-39. doi:10.1016/j.annonc.2020.11.007.
41. Zanfardino M, Franzese M, Pane K, Cavaliere C, Monti S, Esposito G, et al. Bringing radiomics into a multi-omics framework for a comprehensive genotype-phenotype characterization of oncological diseases. Journal of translational medicine. 2019;17:337. doi:10.1186/s12967-019-2073-2.
42. Haque W, Verma V, Hatch S, Suzanne Klimberg V, Brian Butler E, Teh BS. Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. Breast Cancer Res Treat. 2018;170:559-67. doi:10.1007/s10549-018-4801-3.
43. Li W, Yu K, Feng C, Zhao D. Molecular Subtypes Recognition of Breast Cancer in Dynamic Contrast-Enhanced Breast Magnetic Resonance Imaging Phenotypes from Radiomics Data. Comput Math Methods Med. 2019;2019:6978650. doi:10.1155/2019/6978650.
44. Wang Q, Mao N, Liu M, Shi Y, Ma H, Dong J, et al. Radiomic analysis on magnetic resonance diffusion weighted image in distinguishing triple-negative breast cancer from other subtypes: a feasibility study. Clin Imaging. 2021;72:136-41. doi:10.1016/j.clinimag.2020.11.024.
45. Huang SY, Franc BL, Harnish RJ, Liu G, Mitra D, Copeland TP, et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. NPJ Breast Cancer. 2018;4:24. doi:10.1038/s41523-018-0078-2.

Fig. 1



Generation of PDX
and co-clinical
imaging

Tumor
segmentation
in PDX and humans

IBSI-derived image
features from PDX
and human tumors

Repeatability

Auto-correlation

Volume-dependency

Feature optimization
in preclinical PDX trial

ML prediction of response to
therapy in PDX

Cross-validation

Identify radiomic signature
(RadSig) of response to
therapy

Implement ML-RadSig
in clinical trial to
predict/assess
response to therapy

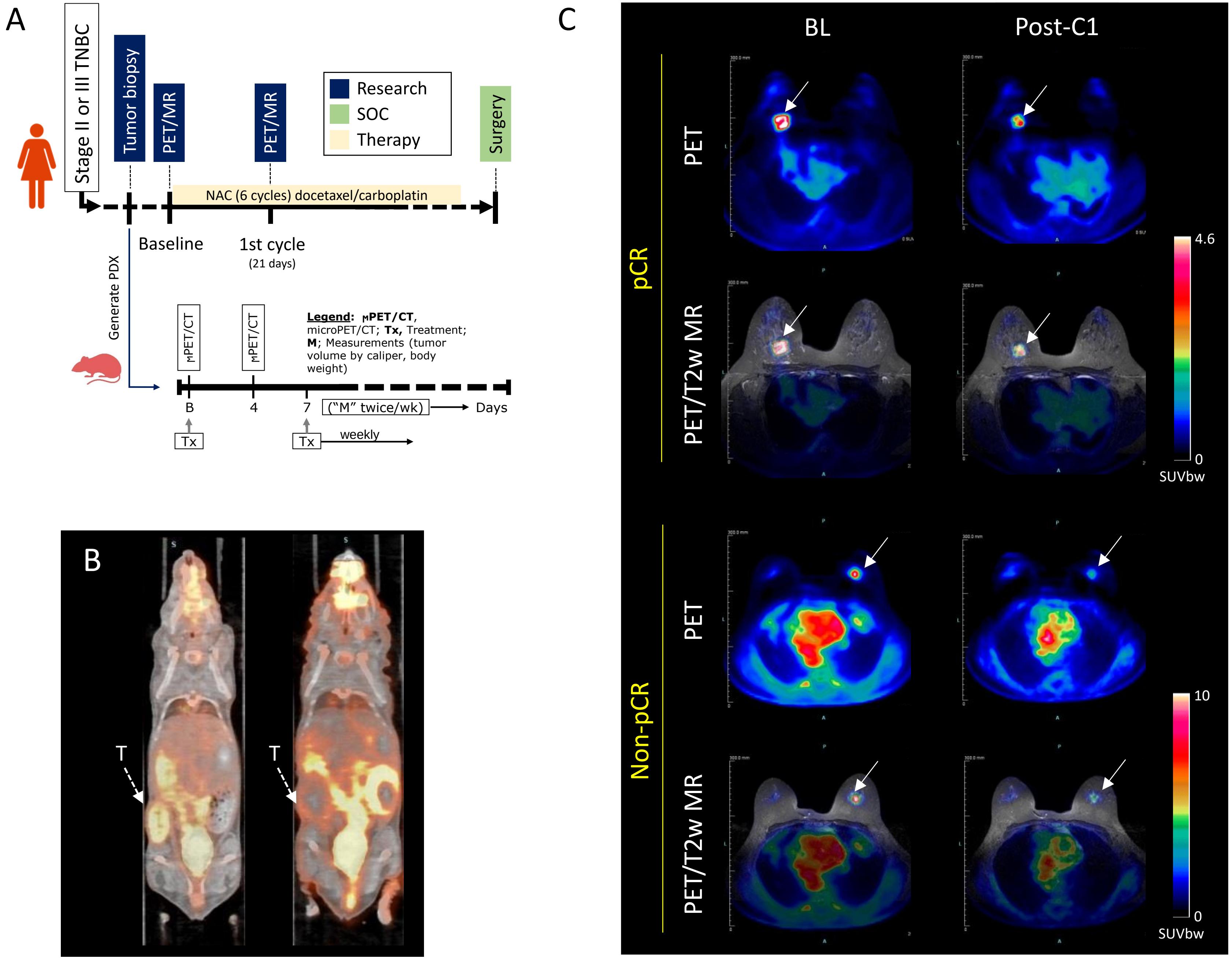

A
Stage II or III TNBC
Tumor biopsy
PET/MR
PET/MR
Research
SOC
Therapy
Surgery
NAC (6 cycles) docetaxel/carboplatin
Baseline
1st cycle
(21 days)
Generate PDX
μPET/CT
μPET/CT
Legend: μPET/CT, microPET/CT; Tx, Treatment; M; Measurements (tumor volume by caliper, body weight)
B
4
7
("M" twice/wk)
Days
Tx
Tx
weekly
B
T
T
C
BL
Post-C1
pCR
PET
PET/T2w MR
Non-pCR
PET
PET/T2w MR
4.6
0
SUVbw
10
0
SUVbw

Fig. 2

Fig. 3

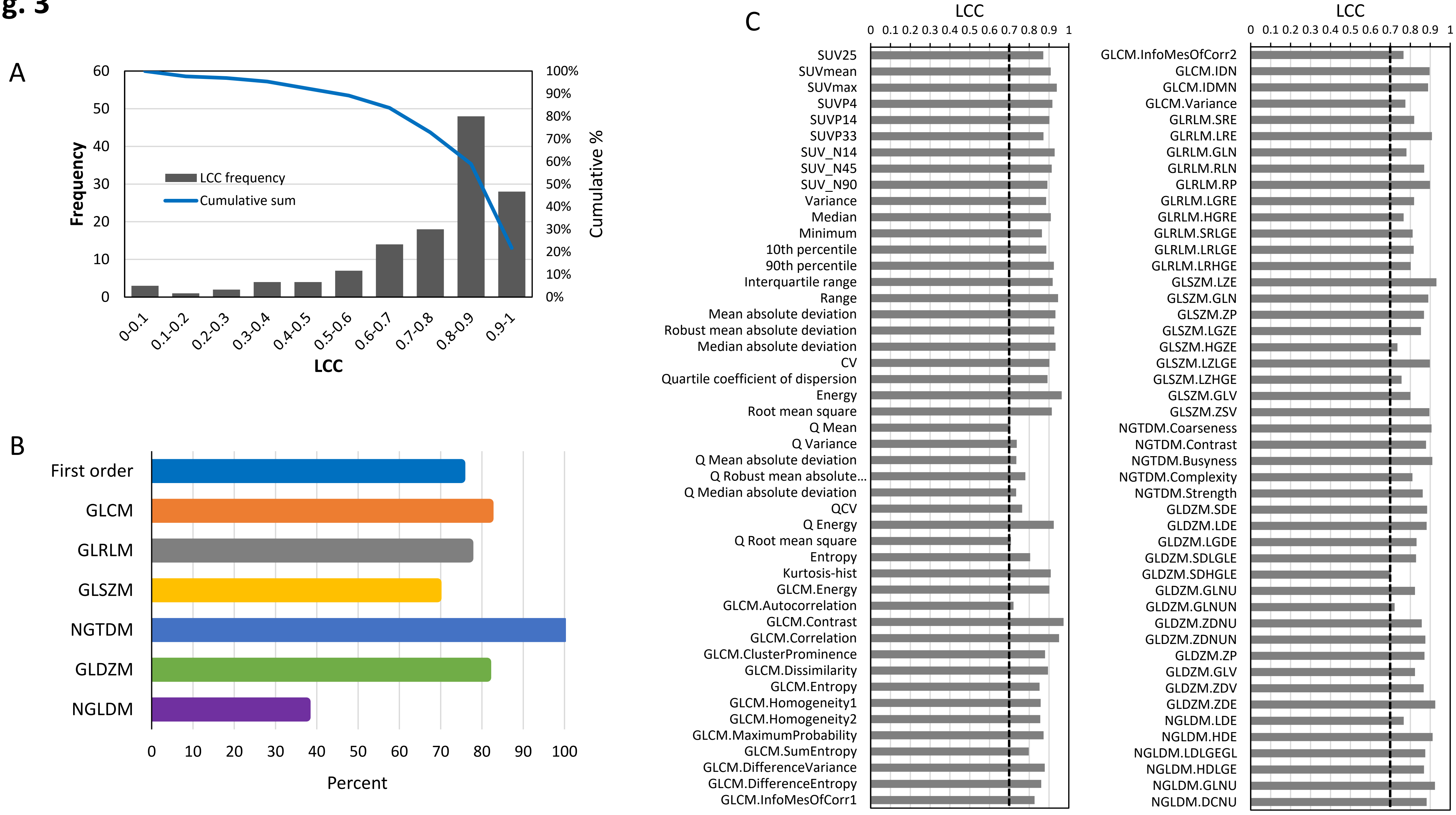
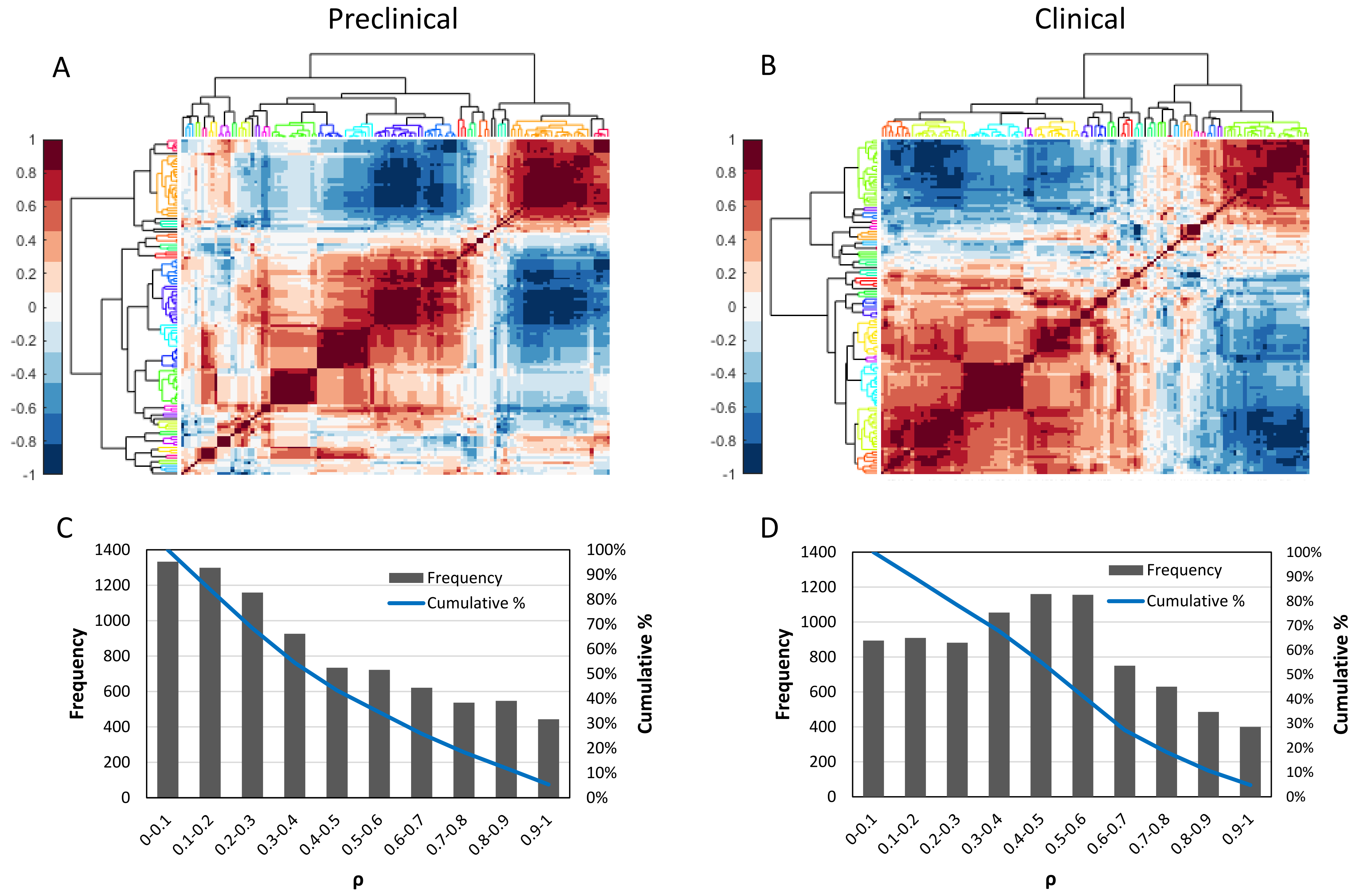
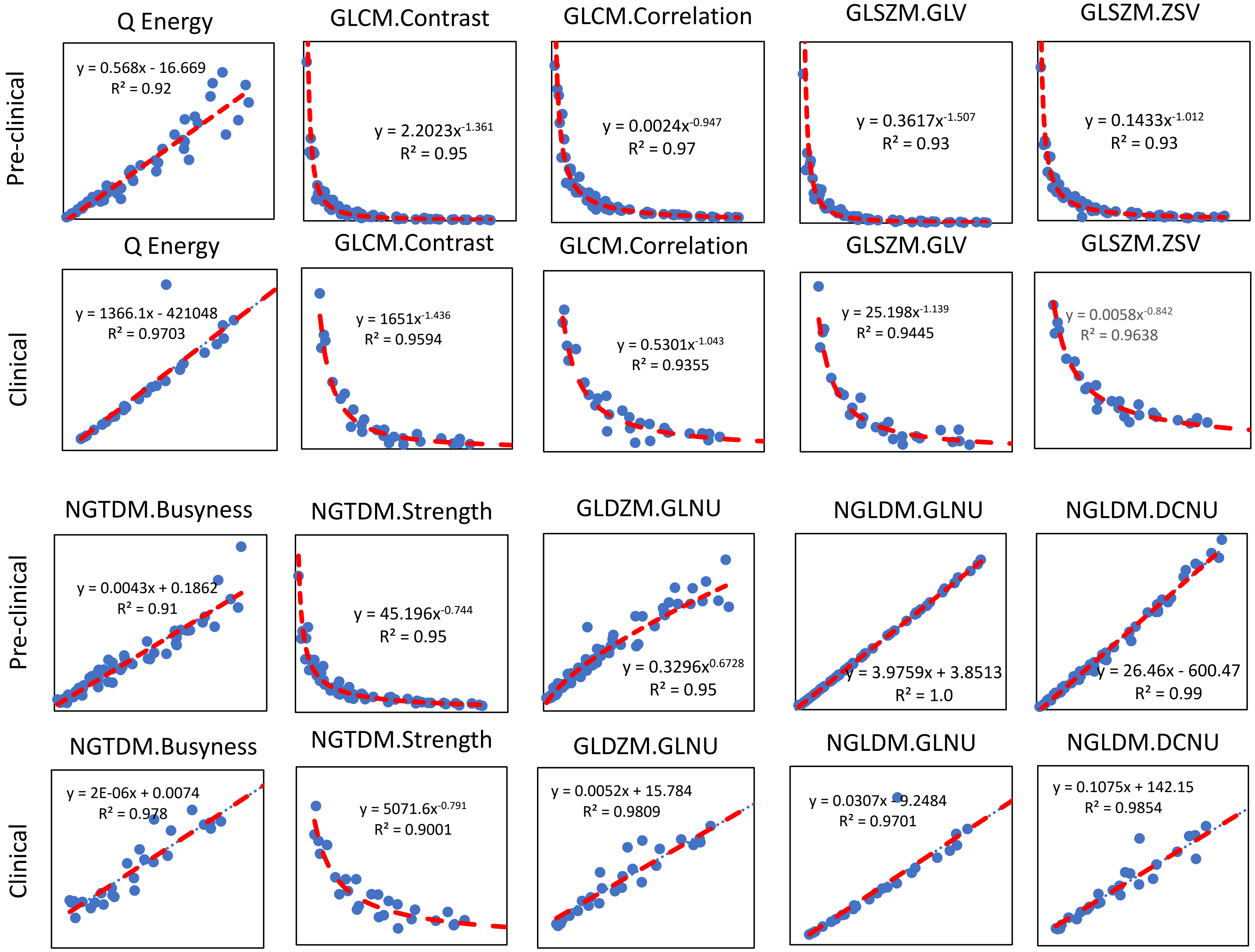

A
Frequency
60
50
40
30
20
10
0
LCC frequency
Cumulative sum
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Cumulative %
0-0.1
0.1-0.2
0.2-0.3
0.3-0.4
0.4-0.5
0.5-0.6
0.6-0.7
0.7-0.8
0.8-0.9
0.9-1
LCC
B
First order
GLCM
GLRLM
GLSZM
NGTDM
GLDZM
NGLDM
0
10
20
30
40
50
60
70
80
90
100
Percent
C
LCC
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
SUV25
SUVmean
SUVmax
SUVP4
SUVP14
SUVP33
SUV_N14
SUV_N45
SUV_N90
Variance
Median
Minimum
10th percentile
90th percentile
Interquartile range
Range
Mean absolute deviation
Robust mean absolute deviation
Median absolute deviation
CV
Quartile coefficient of dispersion
Energy
Root mean square
Q Mean
Q Variance
Q Mean absolute deviation
Q Robust mean absolute...
Q Median absolute deviation
QCV
Q Energy
Q Root mean square
Entropy
Kurtosis-hist
GLCM.Energy
GLCM.Autocorrelation
GLCM.Contrast
GLCM.Correlation
GLCM.ClusterProminence
GLCM.Dissimilarity
GLCM.Entropy
GLCM.Homogeneity1
GLCM.Homogeneity2
GLCM.MaximumProbability
GLCM.SumEntropy
GLCM.DifferenceVariance
GLCM.DifferenceEntropy
GLCM.InfoMesOfCorr1
LCC
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
GLCM.InfoMesOfCorr2
GLCM.IDN
GLCM.IDMN
GLCM.Variance
GLRLM.SRE
GLRLM.LRE
GLRLM.GLN
GLRLM.RLN
GLRLM.RP
GLRLM.LGRE
GLRLM.HGRE
GLRLM.SRLGE
GLRLM.LRLGE
GLRLM.LRHGE
GLSZM.LZE
GLSZM.GLN
GLSZM.ZP
GLSZM.LGZE
GLSZM.HGZE
GLSZM.LZLGE
GLSZM.LZHGE
GLSZM.GLV
GLSZM.ZSV
NGTDM.Coarseness
NGTDM.Contrast
NGTDM.Busyness
NGTDM.Complexity
NGTDM.Strength
GLDZM.SDE
GLDZM.LDE
GLDZM.LGDE
GLDZM.SDLGLE
GLDZM.SDHGLE
GLDZM.GLNU
GLDZM.GLNUN
GLDZM.ZDNU
GLDZM.ZDNUN
GLDZM.ZP
GLDZM.GLV
GLDZM.ZDV
GLDZM.ZDE
NGLDM.LDE
NGLDM.HDE
NGLDM.LDLGEGL
NGLDM.HDLGE
NGLDM.GLNU
NGLDM.DCNU

Fig. 4



Preclinical
Clinical
A
B
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
C
D
1400
1200
1000
800
600
400
200
0
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Frequency
Cumulative %
0-0.1
0.1-0.2
0.2-0.3
0.3-0.4
0.4-0.5
0.5-0.6
0.6-0.7
0.7-0.8
0.8-0.9
0.9-1
ρ

Fig. 5



Pre-clinical
Q Energy
$y = 0.568x - 16.669$
$R^2 = 0.92$
GLCM.Contrast
$y = 2.2023x^{-1.361}$
$R^2 = 0.95$
GLCM.Correlation
$y = 0.0024x^{-0.947}$
$R^2 = 0.97$
GLSZM.GLV
$y = 0.3617x^{-1.507}$
$R^2 = 0.93$
GLSZM.ZSV
$y = 0.1433x^{-1.012}$
$R^2 = 0.93$
Clinical
Q Energy
$y = 1366.1x - 421048$
$R^2 = 0.9703$
GLCM.Contrast
$y = 1651x^{-1.436}$
$R^2 = 0.9594$
GLCM.Correlation
$y = 0.5301x^{-1.043}$
$R^2 = 0.9355$
GLSZM.GLV
$y = 25.198x^{-1.139}$
$R^2 = 0.9445$
GLSZM.ZSV
$y = 0.0058x^{-0.842}$
$R^2 = 0.9638$
Pre-clinical
NGTDM.Busyness
$y = 0.0043x + 0.1862$
$R^2 = 0.91$
NGTDM.Strength
$y = 45.196x^{-0.744}$
$R^2 = 0.95$
GLDZM.GLNU
$y = 0.3296x^{0.6728}$
$R^2 = 0.95$
NGLDM.GLNU
$y = 3.9759x + 3.8513$
$R^2 = 1.0$
NGLDM.DCNU
$y = 26.46x - 600.47$
$R^2 = 0.99$
Clinical
NGTDM.Busyness
$y = 2E-06x + 0.0074$
$R^2 = 0.978$
NGTDM.Strength
$y = 5071.6x^{-0.791}$
$R^2 = 0.9001$
GLDZM.GLNU
$y = 0.0052x + 15.784$
$R^2 = 0.9809$
NGLDM.GLNU
$y = 0.0307x - 9.2484$
$R^2 = 0.9701$
NGLDM.DCNU
$y = 0.1075x + 142.15$
$R^2 = 0.9854$

Fig. 6

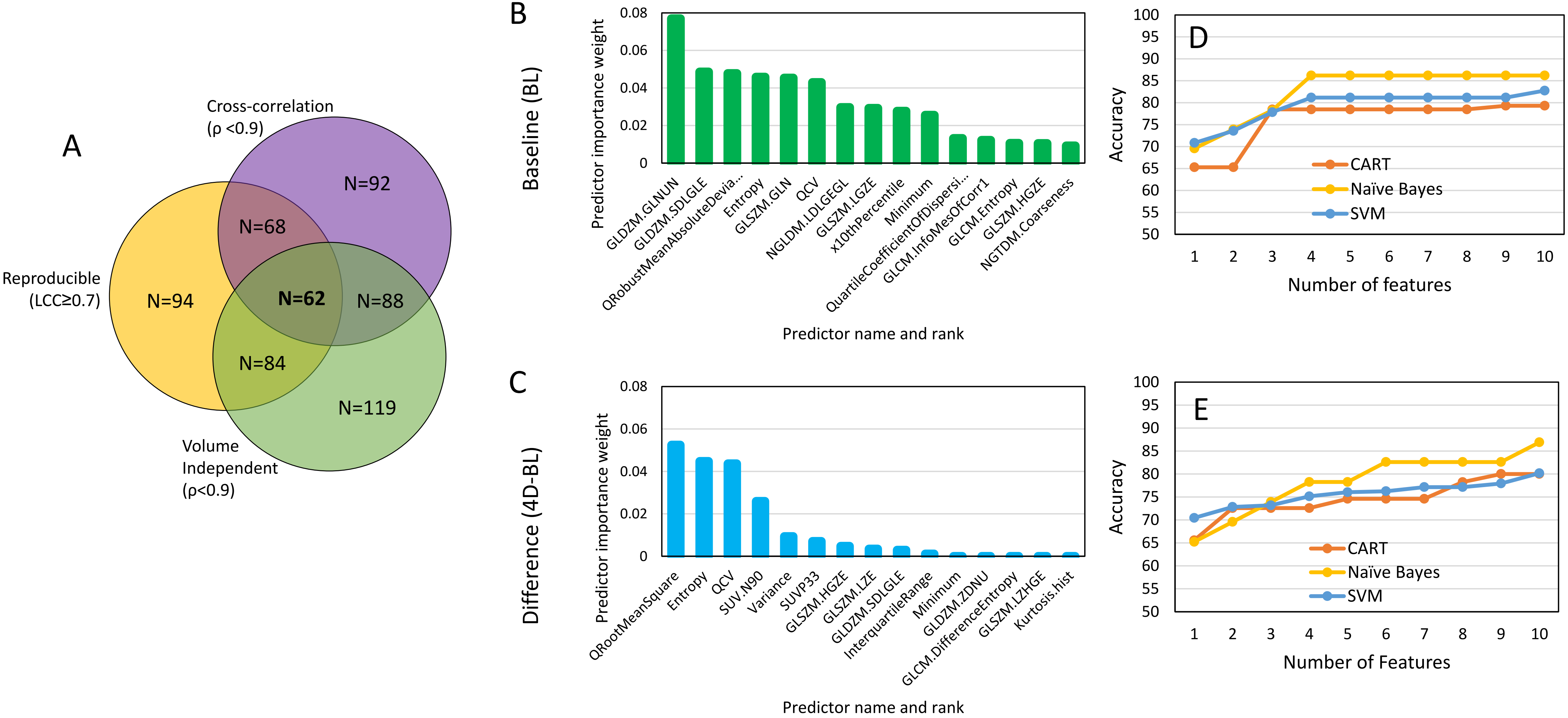
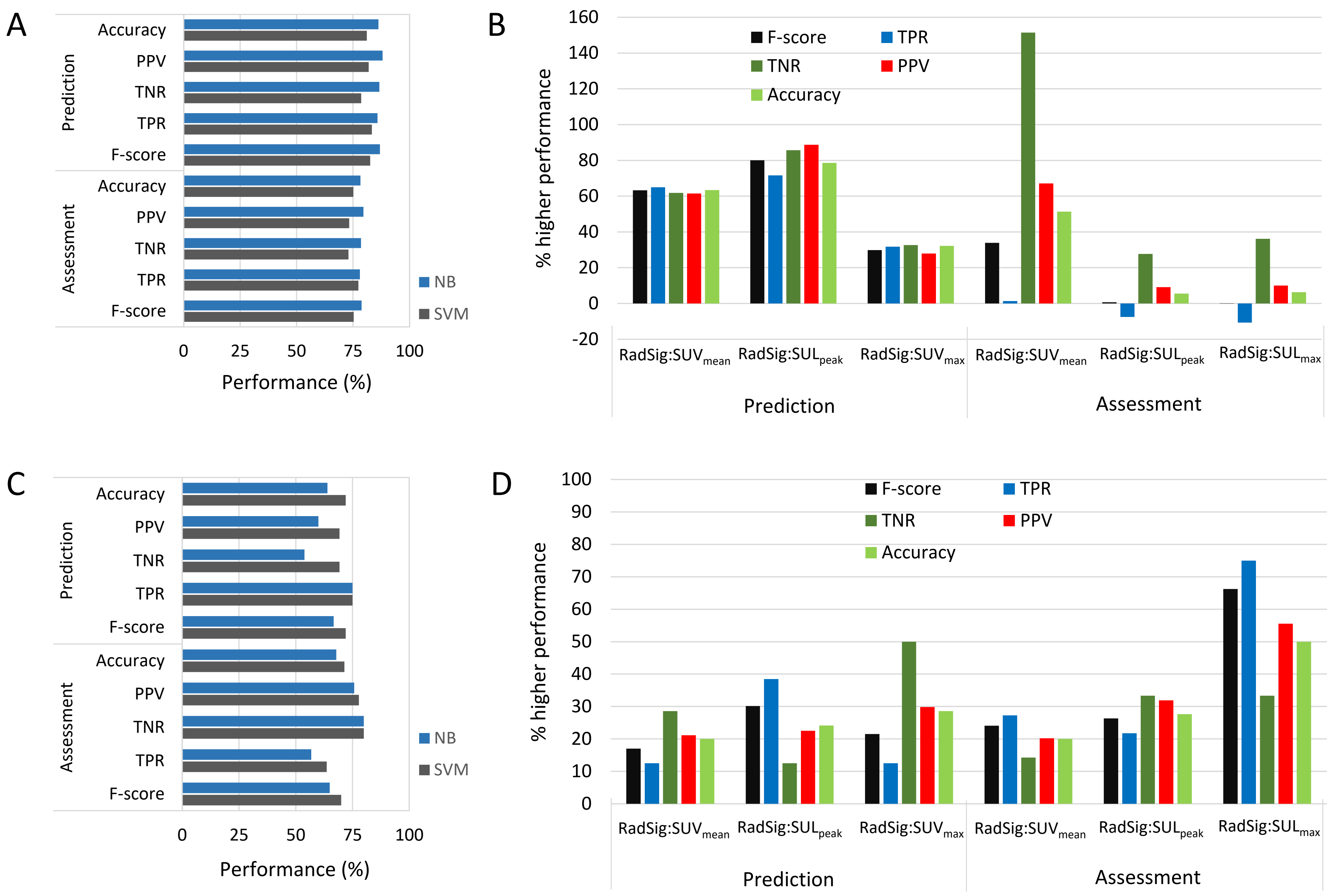

A
Cross-correlation
(ρ <0.9)
Reproducible
(LCC≥0.7)
Volume
Independent
(ρ<0.9)
N=92
N=68
N=94
N=62
N=88
N=84
N=119
B
Baseline (BL)
Predictor importance weight
0.08
0.06
0.04
0.02
0
GLDZM.GLNUN
GLDZM.SDLGLE
QRobustMeanAbsoluteDevia...
Entropy
GLSZM.GLN
QCV
NGLDM.LDLGEGL
GLSZM.LGZE
x10thPercentile
Minimum
QuartileCoefficientOfDispersi...
GLCM.InfoMesOfCorr1
GLCM.Entropy
GLSZM.HGZE
NGTDM.Coarseness
Predictor name and rank
C
Difference (4D-BL)
Predictor importance weight
0.08
0.06
0.04
0.02
0
QRootMeanSquare
Entropy
QCV
SUV.N90
Variance
SUVP33
GLSZM.HGZE
GLSZM.LZE
GLDZM.SDLGLE
InterquartileRange
Minimum
GLDZM.ZDNU
GLCM.DifferenceEntropy
GLSZM.LZHGE
Kurtosis.hist
Predictor name and rank
D
Accuracy
100
95
90
85
80
75
70
65
60
55
50
1
2
3
4
5
6
7
8
9
10
Number of features
CART
Naïve Bayes
SVM
E
Accuracy
100
95
90
85
80
75
70
65
60
55
50
1
2
3
4
5
6
7
8
9
10
Number of Features
CART
Naïve Bayes
SVM

Fig. 7


A
Prediction
Assessment
Accuracy
PPV
TNR
TPR
F-score
Accuracy
PPV
TNR
TPR
F-score
0
25
50
75
100
Performance (%)
NB
SVM
B
% higher performance
160
140
120
100
80
60
40
20
0
-20
F-score
TPR
TNR
PPV
Accuracy
RadSig:SUV$_{mean}$
RadSig:SUL$_{peak}$
RadSig:SUV$_{max}$
RadSig:SUV$_{mean}$
RadSig:SUL$_{peak}$
RadSig:SUL$_{max}$
Prediction
Assessment
C
Prediction
Assessment
Accuracy
PPV
TNR
TPR
F-score
Accuracy
PPV
TNR
TPR
F-score
0
25
50
75
100
Performance (%)
NB
SVM
D
% higher performance
100
90
80
70
60
50
40
30
20
10
0
F-score
TPR
TNR
PPV
Accuracy
RadSig:SUV$_{mean}$
RadSig:SUL$_{peak}$
RadSig:SUV$_{max}$
RadSig:SUV$_{mean}$
RadSig:SUL$_{peak}$
RadSig:SUL$_{max}$
Prediction
Assessment

Co-clinical FDG-PET Radiomic Signature in Predicting Response to Neoadjuvant Chemotherapy in Triple Negative Breast Cancer

Sudipta Roy[1], Timothy Whitehead[1], Shunqiang Li[2], Foluso O. Ademuyiwa[2], Richard L. Wahl[1,3], Farrokh Dehdashti[1], Kooresh I. Shoghi[1,4]

[1]Department of Radiology, [2]Department of Medicine, Division of Oncology, and [3]Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO, USA; [4]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA

Supplemental Table S1: *Radiomic features.* Light shaded color radiomics were non reproducible features.

Type	Method	Description	Sl. NO	Features short name	Features Full name/description	LCC value
First order (33)	Gray-level intensity, histogram, and statistics	Global distribution of intensity values in terms of spread, symmetry, flatness, uniformity and randomness	1	SUV25	Mean of Standardized Uptake Values of top 25% threshold	0.87
			2	SUVmean	Mean intensity	0.91
			3	SUVmax	Max intensity	0.94
			4	SUVP4	Mean intensity of sphere of radius 1	0.92
			5	SUVP14	Mean intensity of sphere of radius 2	0.90
			6	SUVP33	Mean intensity of sphere of radius 3	0.87
			7	SUV_N14	Mean intensity of top 14 voxels	0.93
			8	SUV_N45	Mean intensity of top 45 voxels	0.91
			9	SUV_N90	Mean intensity of top 90 voxels	0.89
			10	Variance	-	0.89
			11	Median	-	0.91
			12	Minimum	-	0.86
			13	10th percentile	-	0.89
			14	90th percentile	-	0.92
			15	Interquartile range	-	0.92
			16	Range	-	0.95
			17	Mean absolute deviation	-	0.93
			18	Robust mean absolute deviation	-	0.93
			19	Median absolute deviation	-	0.93
			20	CV	Coefficient of Variance	0.90
			21	Quartile coefficient of dispersion	-	0.89
			22	Energy	-	0.96
			23	Root mean square	-	0.91
			24	Q Mean	Quartile Mean	0.70
			25	Q Variance	Quartile Variance	0.74
			26	Q Mean absolute deviation	Quartile Mean absolute deviation	0.74
			27	Q Robust mean absolute deviation	Quartile Robust mean absolute deviation	0.78
			28	Q Median absolute deviation	Quartile Median absolute deviation	0.73
			29	QCV	Quartile CV	0.76
			30	Q Energy	Quartile Energy	0.92

			31	Q Root mean square	Quartile Root mean square	0.71
			32	Entropy	-	0.80
			33	Kurtosis-hist	-	0.91
			34	Q Median		0.68
			35	Q Minimum		0.67
			36	Q 10th percentile		0.65
			37	Q 90th percentile		0.66
			38	Q Maximum		0.68
			39	Q Interquartile range		0.62
			40	Q Range		0.63
			41	Q Quartile coefficient of dispersion		0.21
			42	Global.Variance		0.61
			43	Global.Skewness		0.67
			44	Global.Kurtosis		0.62
Second order	Gray-level co-occurrence matrix (GLCM)	Spatial relationship between pixel in a specific direction, highlighting property of uniformity, homogeneity, randomness and linear dependency of images.	45	Energy	-	0.90
			46	Autocorrelation	-	0.72
			47	Contrast	-	0.97
			48	Correlation	-	0.95
			49	Cluster Prominence	Measure of the skewness and asymmetry	0.88
			50	Dissimilarity	-	0.89
			51	Entropy	-	0.85
			52	Homogeneity1	-	0.86
			53	Homogeneity2	-	0.86
			54	MaximumProbability	Occurrences of the most predominant pair of neighboring intensity values.	0.87
			55	SumEntropy	-	0.80
			56	DifferenceVariance	-	0.88
			57	DifferenceEntropy	-	0.86
			58	InfoMesOfCorr1	Information measures of correlation 1	0.83
			59	InfoMesOfCorr2	Information measures of correlation 2	0.77
			60	IDN	Inverse Difference Normalized	0.90
			61	IDMN	Inverse Difference Moment Normalized	0.89
			62	Variance		0.78
			63	ClusterShade		0.68
			64	SumOfSqauresVariance		0.67
			65	SumAverage		0.68
			66	SumVariance		0.69
Higher order	Gray-level run-length matrix (GLRLM)	Texture in specific direction, where fine texture has more short runs whilst coarse texture presents more long runs with different intensity values.	67	SRE	Short Run Emphasis	0.82
			68	LRE	Long Run Emphasis	0.91
			69	GLN	Gray-Level Non-uniformity	0.78
			70	RLN	Run-Length Non-uniformity	0.87
			71	RP	Run Percentage	0.90
			72	LGRE	Low Gray-Level Run Emphasis	0.82
			73	HGRE	High Gray-Level Run Emphasis	0.77
			74	SRLGE	Short Run Low Gray-Level Emphasis	0.81
			75	LRLGE	Long Run Low Gray-Level Emphasis	0.82
			76	LRHGE	Long Run High Gray-Level Emphasis	0.80
			77	SRHGE		0.63

			78	GLV		0.46
			79	RLV		0.58
	Gray-level size zone matrix (GLSZM)	Regional intensity variations or the distribution of homogeneity regions.	80	LZE	Large Zone Emphasis	0.93
			81	GLN	Gray-Level Non-uniformity	0.89
			82	ZP	Zone Percentage	0.87
			83	LGZE	Low Gray-Level Zone Emphasis	0.85
			84	HGZE	High Gray-Level Zone Emphasis	0.74
			85	LZLGE	Large Zone Low Gray-Level Emphasis	0.90
			86	LZHGE	Large Zone High Gray-Level Emphasis	0.76
			87	GLV	Gray-Level Variance	0.80
			88	ZSV	Run-Length Variance	0.90
			89	SZE		0.50
			90	ZSN		0.44
			91	SZLGE		0.61
			92	SZHGE		0.35
	Neighborhood gray-tone difference matrix (NGTDM)	Spatial relationship among three or more pixels, closely approaching the human perception of image	93	Coarseness	Measure of average difference between the center voxel and its neighborhood	0.91
			94	Contrast	Measure of the spatial intensity change on the overall gray level dynamic range	0.88
			95	Busyness	Measure of the change from a pixel to its neighbor	0.91
			96	Complexity	Rapid changes within gray level intensity	0.81
			97	Strength	Measure of the primitiveness in an image	0.86
	Gray Level Distance Zone Matrix (GLDZM)	Counts the number of groups of connected voxels with a specific gray level (GL) value and distance to region of interest edge. It contains an extra-level of information compared to GLSZM: it captures the relation between the distance from the edge, zone-size, and GL.	98	SDE	Small Distance Emphasis	0.88
			99	LDE	Large Distance Emphasis	0.88
			100	LGDE	Low Gray Dependence Emphasis	0.83
			101	SDLGLE	Small Dependence Low Gray Level Emphasis	0.83
			102	SDHGLE	Small Dependence High Gray Level Emphasis	0.70
			103	GLNU	Gray Level Non-Uniformity	0.82
			104	GLNUN	Gray Level Non-Uniformity Normalized	0.72
			105	ZDNU	Size Zone Non-Uniformity	0.86
			106	ZDNUN	Size Zone Non-Uniformity Normalized	0.88
			107	ZP	Zone Percentage	0.87
			108	GLV	Gray Level Variance	0.82
			109	ZDV	Zone Distance Variance	0.87
			110	ZDE	Zone Distance Emphasis	0.93
			111	HGLE		0.68
			112	LDLGLE		0.56
			113	LDLGLEn		0.56
	Neighboring Grey Level Dependence Matrix (NGLDM)	The coarseness of the overall texture.	114	LDE	Low Dependence Emphasis	0.77
			115	HDE	High Dependence Emphasis	0.91
			116	LDLGEGL	Low Dependence Low Gray Emphasis Gray Level	0.88
			117	HDLGE	High Dependence Low Gray Emphasis	0.87
			118	GLNU	Gray Level Non-Uniformity	0.92

			119	DCNU	Dependence counts non-Uniformity	0.88
			120	LGLCE		0.10
			121	HGLCE		-0.19
			122	LDHGE		0.54
			123	HDHGE		0.49
			124	GLNUN		0.25
			125	DCNUN		0.55
			126	GLV		-0.03
			127	DCV		0.51
			128	DCE		0.38
			129	DCEn		0.33
			130	Volume		0.92
			131	Metabolic Tumor Volume		0.80

Supplemental Table S2: *Hierarchical clustering on cross correlation for preclinical and clinical features (dendrogram with distances 3 unit)*

Sl. NO	Features short name	Preclinical Cluster	Clinical Cluster	Sl. NO	Features short name	Preclinical Cluster	Clinical Cluster
1	SUV25	12	3	66	SumVariance	6	7
2	SUVmean	5	3	67	SRE	20	21
3	SUVmax	12	3	68	LRE	15	2
4	SUVP4	12	3	69	GLN	19	21
5	SUVP14	12	3	70	RLN	20	21
6	SUVP33	12	3	71	RP	20	21
7	SUV_N14	12	3	72	LGRE	19	19
8	SUV_N45	12	3	73	HGRE	8	7
9	SUV_N90	12	2	74	SRLGE	19	19
10	Variance	13	3	75	LRLGE	15	7
11	Median	5	3	76	LRHGE	10	13
12	Minimum	20	9	77	SRHGE	7	2
13	10th percentile	1	9	78	GLV	20	21
14	90th percentile	4	3	79	RLV	20	21
15	Interquartile range	13	3	80	LZE	14	21
16	Range	13	3	81	GLN	19	2
17	Mean absolute deviation	13	3	82	ZP	20	20
18	Robust mean absolute deviation	13	3	83	LGZE	20	21
19	Median absolute deviation	13	3	84	HGZE	1	21
20	CV	15	6	85	LZLGE	14	18
21	Quartile coefficient of dispersion	15	6	86	LZHGE	14	7
22	Energy	13	2	87	GLV	20	18
23	Root mean square	4	3	88	ZSV	20	16
24	Q Mean	11	5	89	SZE	1	10
25	Q Variance	11	5	90	ZSN	2	2
26	Q Mean absolute deviation	11	5	91	SZLGE	20	21
27	Q Robust mean absolute deviation	11	5	92	SZHGE	17	21
28	Q Median absolute deviation	11	5	93	Coarseness	20	21
29	QCV	11	8	94	Contrast	21	17
30	Q Energy	14	2	95	Busyness	14	2
31	Q Root mean square	11	5	96	Complexity	21	16
32	Entropy	13	15	97	Strength	20	21
33	Kurtosis-hist	18	14	98	SDE	20	21
34	Q Median	15	6	99	LDE	14	4
35	Q Minimum	11	5	100	LGDE	20	10
36	Q 10th percentile	11	12	101	SDLGLE	20	19
37	Q 90th percentile	5	1	102	SDHGLE	17	10
38	Q Maximum	11	5	103	GLNU	14	21

39	Q Interquartile range	11	5	104	GLNUN	3	10
40	Q Range	11	5	105	ZDNU	14	4
41	Q Quartile coefficient of dispersion	11	5	106	ZDNUN	20	2
42	Global.Variance	11	8	107	ZP	20	6
43	Global.Skewness	16	15	108	GLV	20	1
44	Global.Kurtosis	18	14	109	ZDV	14	21
45	Energy	6	1	110	ZDE	14	21
46	Autocorrelation	20	21	111	HGLE	17	13
47	Contrast	20	21	112	LDLGLE	9	4
48	Correlation	10	1	113	LDLGLEn	14	8
49	Cluster Prominence	21	20	114	LDE	20	21
50	Dissimilarity	8	17	115	HDE	15	2
51	Entropy	15	21	116	LDLGEGL	20	11
52	Homogeneity1	15	2	117	HDLGE	14	12
53	Homogeneity2	14	1	118	GLNU	14	21
54	MaximumProbability	8	1	119	DCNU	14	21
55	SumEntropy	21	1	120	LGLCE	2	2
56	DifferenceVariance	21	5	121	HGLCE	18	2
57	DifferenceEntropy	8	5	122	LDHGE	1	2
58	InfoMesOfCorr1	3	5	123	HDHGE	13	13
59	InfoMesOfCorr2	15	1	124	GLNUN	20	2
60	IDN	15	16	125	DCNUN	16	21
61	IDMN	10	16	126	GLV	18	12
62	Variance	16	2	127	DCV	9	2
63	ClusterShade	1	21	128	DCE	7	1
64	SumOfSqauresVariance	6	11	129	DCEn	2	21
65	SumAverage	6	11				

Supplemental Table S3: AIC and BIC stats for volume dependent radiomic features

Features name	rho(ρ)	Stats	Linear	Exponential	Log	Power	Polynomial
Q Energy	0.98	AIC	**459.3**	577.0	531.2	461.9	460.5
		BIC	**477.6**	595.3	549.6	480.3	485.0
GLCM.Correlation	-0.98	AIC	-1281.3	-1299.1	-1193.8	**-1304.3**	-1229.8
		BIC	-1262.9	-1280.8	-1175.4	**-1285.9**	-1205.3
GLCM.Contrast	-0.97	AIC	-642.2	-648.3	-681.5	**-686.7**	-658.3
		BIC	-623.8	-629.9	-663.1	**-668.3**	-633.8
GLSZM.GLV	-0.98	AIC	-945.1	-954.8	-993.0	**-905.8**	-811.3
		BIC	-926.7	-936.4	-974.7	**-887.5**	-786.8
GLSZM.ZSV	-0.96	AIC	-822.1	-832.7	-868.6	**-925.9**	-830.2
		BIC	-803.7	-814.3	-850.2	**-907.5**	-805.7
NGTDM.Busyness	0.97	AIC	**-68.5**	-16.4	-21.3	-60.0	-60.8
		BIC	**-86.9**	-34.8	-39.6	-78.4	-85.3
NGTDM.Strength	-0.97	AIC	-53.6	-63.2	-118.8	**-156.6**	-86.1
		BIC	-35.3	-44.8	-100.5	**-138.2**	-61.6
GLDZM.GLNU	0.97	AIC	129.7	218.9	171.9	**121.5**	137.7
		BIC	148.0	237.3	190.3	**139.9**	162.2
NGLDM.GLNU	0.99	AIC	**340.7**	772.6	728.3	477.2	334.9
		BIC	**359.1**	791.0	746.7	495.6	359.4
NGLDM.DCNU	0.99	AIC	**772.3**	993.0	958.3	1079.6	752.6
		BIC	**790.7**	1011.3	976.7	1098.0	777.1

* Selected model is highlighted in bold

Supplemental Table S4: *Rank importance (Using ReliefF) of radiomic features based on treatment response prediction (94 reproducible features, tumor volume and metabolic tumor volume)*

Rank	Features name for baseline (BL)	Features name for difference(4D-BL)
1	'GLDZM_GLNUN'	QRootMeanSquare'
2	'GLDZM_SDLGLE'	'QMean'
3	'GLDZM_LGDE'	'QMedianAbsoluteDeviation'
4	'QRobustMeanAbsoluteDeviation'	'QMeanAbsoluteDeviation'
5	'QMedianAbsoluteDeviation'	'QRobustMeanAbsoluteDeviation'
6	'QMeanAbsoluteDeviation'	'Entropy'
7	'QMean'	'QCV'
8	'Entropy'	'QVariance'
9	'QRootMeanSquare'	'SUV_N90'
10	'QVariance'	'SUV25'
11	'GLSZM_GLN'	'SUVP4'
12	'GLSZM_GLV'	'SUV_N45'
13	'GLSZM_ZSV'	'SUVmax'
14	'QCV'	'SUV_N14'
15	'GLCM_Contrast'	'GLCM_Correlation'
16	'NGLDM_LDLGEGL'	'NGLDM_GLNU'
17	'GLSZM_LGZE'	'Variance'
18	'GLSZM_LZLGE'	'SUVP33'
19	'x10thPercentile'	'SUVP14'
20	'Minimum'	'GLDZM_GLNU'
21	'QuartileCoefficientOfDispersion'	'GLSZM_HGZE'
22	'GLCM_InfoMesOfCorr1'	'NGLDM_DCNU'
23	'Median'	'NGTDM_Busyness'
24	'GLSZM_HGZE'	'GLSZM_LZE'
25	'QEnergy'	'GLDZM_SDLGLE'
26	'GLCM_Entropy'	'InterquartileRange'
27	'GLCM_MaximumProbability'	'Minimum'
28	‘GLCM_Energy’	'GLDZM_ZDNU'
29	'SUVmean'	'GLCM_DifferenceEntropy'
30	'GLSZM_LZE'	'RobustMeanAbsoluteDeviation'
31	'GLCM_SumEntropy'	'GLSZM_LZHGE'
32	'RootMeanSquare'	'GLDZM_LGDE’
33	'GLDZM_GLV'	'Range'
34	'Variance'	'GLCM_MaximumProbability'
35	'GLCM_Autocorrelation'	'Kurtosis_hist'
36	'NGTDM_Coarseness'	'Energy'
37	'GLCM_Correlation'	'MedianAbsoluteDeviation'
38	'GLRLM_GLN'	'MeanAbsoluteDeviation'
39	'GLCM_Variance'	'NGTDM_Complexity'
40	'NGTDM_Strength'	'QEnergy'
41	'GLDZM_ZDNU'	'GLDZM_ZDV'
42	'GLSZM_LZHGE'	'GLDZM_LDE'
43	'InterquartileRange'	'GLCM_Entropy'
44	'GLCM_InfoMesOfCorr2'	'NGLDM_HDE'
45	'GLDZM_ZP'	'NGTDM_Contrast'
46	'GLRLM_HGRE'	'GLRLM_LRLGE'
47	'GLDZM_GLNU'	'GLCM_DifferenceVariance'
48	'x90thPercentile'	'x90thPercentile'
49	'GLDZM_SDHGLE'	'GLRLM_LRE'
50	'RobustMeanAbsoluteDeviation'	'GLDZM_SDHGLE'
51	'Energy'	'GLSZM_LZLGE'
52	'CoefficientOfVariation'	'GLDZM_ZDNUN'
53	'GLCM_ClusterProminence'	'GLCM_Dissimilarity'
54	'GLSZM_ZP'	'GLCM_SumEntropy'

55	'MedianAbsoluteDeviation'	'RootMeanSquare'
56	'GLRLM_LRHGE'	'GLDZM_GLV'
57	'NGLDM_GLNU'	'SUVmean'
58	'GLRLM_LRE'	'x10thPercentile'
59	'MeanAbsoluteDeviation'	'NGLDM_HDLGE'
60	'GLDZM_ZDE'	‘Volume’
61	'SUV25'	'GLSZM_ZP'
62	'SUV_N14'	'GLSZM_LGZE'
63	'SUV_N45'	'GLCM_ClusterProminence'
64	'SUV_N90'	'CoefficientOfVariation'
65	'NGLDM_DCNU'	'GLSZM_ZSV'
66	'NGLDM_LDE'	'Median'
67	'GLRLM_LRLGE'	‘Metabolic tumor volume’
68	'GLCM_Homogeneity1'	'GLCM_InfoMesOfCorr1'
69	'SUVmax'	'GLDZM_SDE'
70	'Range'	'NGLDM_LDLGEGL'
71	'GLCM_Homogeneity2'	'GLCM_Autocorrelation'
72	'GLRLM_RP'	'NGTDM_Strength'
73	'SUVP4'	'GLDZM_ZDE'
74	'NGLDM_HDE'	'GLCM_Homogeneity1'
75	'NGTDM_Busyness'	'GLRLM_RLN'
76	'GLRLM_RLN'	'GLRLM_SRE'
77	'Kurtosis_hist'	'GLCM_Homogeneity2'
78	'SUVP14'	‘GLCM_Energy’
79	'GLRLM_SRE'	'QuartileCoefficientOfDispersion'
80	'GLDZM_ZDNUN'	'GLCM_Variance'
81	'SUVP33'	'GLRLM_HGRE'
82	'GLDZM_SDE'	'GLCM_Contrast'
83	'GLDZM_ZDV'	'GLDZM_ZP'
84	'GLDZM_LDE'	'GLSZM_GLV'
85	‘Volume’	'NGLDM_LDE'
86	'NGLDM_HDLGE'	'GLRLM_LGRE'
87	'GLCM_DifferenceEntropy'	'GLDZM_GLNUN'
88	'GLRLM_SRLGE'	'GLRLM_SRLGE'
89	'GLCM_IDN'	'GLRLM_LRHGE'
90	'GLCM_Dissimilarity'	'GLRLM_RP'
91	'GLCM_IDMN'	'GLSZM_GLN'
92	'GLRLM_LGRE'	'GLCM_IDN'
93	‘Metabolic tumor volume’	'GLCM_InfoMesOfCorr2'
94	'GLCM_DifferenceVariance'	'GLCM_IDMN'
95	'NGTDM_Contrast'	'NGTDM_Coarseness'
96	'NGTDM_Complexity'	'GLRLM_GLN'

Supplemental Table S5: *Accuracy by TNBC Subtyp* ***(%)***

	CART		Naïve Bayes		SVM	
Subtype	Baseline	Difference	Baseline	Difference	Baseline	Difference
IM	84.37	77.75	91.25	78.5	86.5	76.25
BL2	72.72	70.4	83.5	73.25	75.55	71.72
M	81.25	71.25	81.25	82.5	83.75	75
BL1	87.7	68.25	96.25	85.25	91.25	82.5
LAR	62.5	75.00	72.5	82.5	65	75

Supplemental Table S6: Performance of ML algorithms

		Preclinical				Clinical			
		Prediction		Assessment		Prediction		Assessment	
		NB	SVM	NB	SVM	NB	SVM	NB	SVM
RadSig	F-score	86.93	82.62	78.78	75.30	66.67	72.00	64.94	70.00
	TPR	85.81	83.33	78.00	77.38	75.00	75.00	56.82	63.64
	TNR	86.67	78.59	78.57	72.94	53.85	69.23	80.00	80.00
	PPV	88.08	81.92	79.59	73.33	60.00	69.23	75.76	77.78
	Accuracy	86.21	81.14	78.26	75.13	64.00	72.00	67.86	71.43
SUV_{mean}	F-score	53.24	62.30	58.82	76.47	51.06	61.54	50.00	56.41
	Sensitivity (TPR)	52.00	63.33	76.92	86.67	50.00	66.67	45.45	50.00
	Specificity (TNR)	53.57	57.14	31.25	57.14	57.69	53.85	60.00	70.00
	Precision (PPV)	54.55	61.29	47.62	68.42	52.17	57.14	55.56	64.71
	Accuracy	52.76	60.34	51.72	72.41	54.00	60.00	52.38	59.52
SUL_{peak}	F-score	48.28	64.41	78.26	76.54	52.17	55.32	53.66	55.42
	Sensitivity (TPR)	50.00	65.52	84.38	83.22	50.00	54.17	50.00	52.27
	Specificity (TNR)	46.67	62.07	61.54	63.83	61.54	61.54	60.00	60.00
	Precision (PPV)	46.67	63.33	72.97	70.86	54.55	56.52	57.89	58.97
	Accuracy	48.28	63.79	74.14	73.79	56	58	54.76	55.95
SUV_{max}	F-score	60.00	63.64	77.43	75.36	48.00	59.26	40.00	42.11
	Sensitivity (TPR)	60.00	63.23	86.67	86.67	50.00	66.67	36.36	36.36
	Specificity (TNR)	57.14	59.26	60.14	53.57	46.15	46.15	50.00	60.00
	Precision (PPV)	60.00	64.05	69.97	66.67	46.15	53.33	44.44	50.00
	Accuracy	58.62	61.38	73.86	70.69	48.00	56.00	42.86	47.62