CrossMark

ORIGINAL ARTICLE

Texture analysis of high-resolution dedicated breast ^{18}F-FDG PET images correlates with immunohistochemical factors and subtype of breast cancer

Alexis Moscoso[1] · Álvaro Ruibal[1,2,3] · Inés Domínguez-Prado[1] ·
Anxo Fernández-Ferreiro[4] · Míchel Herranz[1] · Luis Albaina[5] · Sonia Argibay[1] ·
Jesús Silva-Rodríguez[1] · Juan Pardo-Montero[1,6] · Pablo Aguiar[1,2]

Abstract

Purpose This study aims to determine whether PET textural features measured with a new dedicated breast PET scanner reflect biological characteristics of breast tumors.

Methods One hundred and thirty-nine breast tumors from 127 consecutive patients were included in this analysis. All

Alexis Moscoso and Álvaro Ruibal equally contributed to this work

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00259-017-3830-1) contains supplementary material, which is available to authorized users.

of them underwent a ^{18}F-FDG PET scan before treatment. Well-known PET quantitative parameters such as SUV_{max}, SUV_{mean}, metabolically active tumor volume (MATV) and total lesion glycolysis (TLG) were extracted. Together with these parameters, local, regional, and global heterogeneity descriptors, which included five textural features (TF), were computed. Immunohistochemical classification of breast cancer considered five subtypes: luminal A like (LA), luminal B like/HER2− (LB−), luminal B like/HER2+ (LB+), HER2-positive-non-luminal (HER2pnl), and triple negative (TN). Associations between PET features and tumor characteristics were assessed using non-parametric hypothesis tests.

Results Along with well-established associations, new correlations were found. HER2-positive tumors had significantly higher uptake ($p < 0.001$, AUCs > 0.70) and presented different global and regional heterogeneity ($p = 0.002$, $p = 0.016$, respectively, AUCs < 0.70). Nine out of ten analyzed features were significantly associated with immunohistochemical subtype. Uptake was lower for LA tumors ($p < 0.001$) with AUCs ranging from 0.71 to 0.88 for each subgroup comparison. Heterogeneity metrics were significantly associated when comparing LA and LB− ($p < 0.01$), being regional heterogeneity metrics more discriminative than any other parameter (AUC = 0.80 compared to AUC = 0.71 for SUV). LB+ and HER2pnl tumors also showed more regional heterogeneity than LA tumors (AUCs = 0.79 and 0.84, respectively). After comparison with whole-body PET studies, we observed an overall improvement in the classification ability of both non-heterogeneity metrics and textural features.

Conclusions PET parameters extracted from high-resolution dedicated breast PET images showed new and stronger correlations with immunohistochemical factors and immunohistochemical subtype of breast cancer compared to whole-body PET.

✉ Juan Pardo-Montero
juan.pardo.montero@sergas.es

✉ Pablo Aguiar
pablo.aguiar.fernandez@sergas.es

1 Nuclear Medicine Department and Molecular Imaging Group, Complexo Hospitalario Universitario de Santiago de Compostela CHUS-IDIS, Travesía da Choupana s/n, Santiago de Compostela 15706, Spain

2 Molecular Imaging Group, Department of Radiology, Faculty of Medicine, University of Santiago de Compostela (USC), Campus Vida, Santiago de Compostela 15782, Spain

3 Fundación Tejerina, Madrid 28003, Spain

4 Pharmacy Department and Pharmacology group, Complexo Hospitalario Universitario de Santiago de Compostela CHUS-IDIS, Travesía da Choupana s/n, Santiago de Compostela 15706, Spain

5 Department of General Surgery, University Hospital A Coruña (SERGAS), A Coruña, Spain

6 Medical Physics Department, Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), Travesía Choupana s/n, Santiago de Compostela 15706, Spain

Introduction

Decades of observation and research have shown that cancer is a heterogeneous disease. The phenomenon of intratumor heterogeneity, i.e., phenotypic and functional variability among cancer cells within the same tumor, arises as a result of the genetic instability of cancer cells and tumor microenvironment. Continuous mutations lead to different cellular subpopulations within the same tumor and the spatial distribution of these populations is complex and dependent on the tumor microenvironment [1, 2]. These biological features have a strong impact on patient treatment and outcome, for instance in breast cancer [3], making its characterization essential for treatment planning.

In this context, positron emission tomography (PET) with ^{18}F-fluorodeoxyglucose (^{18}F-FDG) can reflect metabolic tumor heterogeneity by measuring the variations in metabolism of different tumor regions. Apart from well-known metrics as SUVs (standardized uptake values) and MATV (metabolically active tumor volume), texture analysis [4–6] of ^{18}F-FDG PET images has emerged as a promising tool for assessing intratumor heterogeneity [7]. The spatial distribution of SUV values within the tumor is the most important input of texture analysis since it is essential for an accurate characterization of metabolic heterogeneity. This metabolic heterogeneity is expected to reflect biological and clinical tumor properties and, therefore, most of the studies were conducted to show this correlation. Performance of texture analysis was better when predicting outcome or response to treatment in different types of tumors [8–10] than when discriminating tumor histology or immunohistochemistry (IHC) [7]. Although outcome and response are very relevant clinical variables, one would expect to find significant differences in biologically different tumors, with different aggressiveness and resistance to treatment. However, there is little evidence supporting this association [7]. Considering that the distance scales at which variations in tumor cell types are present range from cellular level (0.01-100 μm) to macroscopic level (subpopulations larger than a few mm) [11, 12], and that the spatial resolution of the most advanced PET systems is about 2–4 mm, it seems clear that we can only measure heterogeneity *directly* if it appears at the macroscopic level and if differences in the radiotracer uptake between the different subpopulations of cells can be effectively measured by PET scanners. Nevertheless, it is possible that a microscopically heterogeneous population of cells presents a macroscopic signature, for instance, a particular shape, allowing us to measure microscopic heterogeneity *indirectly* by correlating texture analysis with anatomopathological information.

Breast cancer (BC) is a good example of a heterogeneous type of tumor [12]. According to [3], BC can be classified into five IHC subtypes: luminal A like (LA), luminal B like/HER2− (LB−), luminal B like/HER2+ (LB+), HER2-positive non-luminal (HER2pnl) and triple negative (TN). Each of these is associated with a particular molecular subtype but, although both classifications are largely concordant, there is no exact overlap between IHC subtype and inherent molecular subtype [13]. Despite this non-perfect correspondence, IHC subtypes have shown prognostic and predictive power and are one of the main inputs in treatment planning [3], showing that this surrogacy encodes most of the relevant information of genetic analyses. Therefore, the study of correlations between tumor heterogeneity and IHC subtypes is valuable in the sense that it may provide the same information as IHC analyses. Furthermore, given the high overlap between IHC surrogates and genetic profiles, a positive correlation between heterogeneity descriptors and IHC subtypes is likely to imply a positive correlation between these descriptors and tumor genetic profiles. Although this can only be proven with a dedicated study, the demonstration of these correlations seems to be the necessary prerequisite before carrying out the more resource-demanding genetic studies. No previous investigations have found strong correlations between textural features and IHC subtype [14–17]. Since a good spatial characterization of tumor uptake is essential to quantify metabolic heterogeneity, an improvement of the spatial resolution of PET scanners may help to elucidate new associations between PET features and tumor biology. In the present study, we used a high-resolution dedicated breast PET (dbPET) scanner (MAMMI-PET, GEM Imaging S.A., Valencia, Spain) [18]. The aim of this work was to explore possible correlations between tumor biological characteristics and PET features measured with a high-resolution dbPET scanner.

Materials and methods

dbPET

PET studies were carried out with MAMMI-PET (GEM Imaging S.A., Valencia, Spain), a dbPET scanner designed to improve small-lesion detection while reducing received dose and study time [18]. Patients lie in prone position, allowing better tumor detection and minimizing breathing motion compared with a whole-body PET scanner. Our dbPET scanner consists of 12 continuous LYSO crystals arranged in a ring with an aperture of 186 mm that provides an axial field of view of 40 mm. The ring can be moved by a precise engine along the axial direction so acquisitions in a step-and-shoot mode can be launched to increase the axial field of view up to 170 mm. The 3D mode-acquired

images were reconstructed with a 3D maximum likelihood expectation maximization (MLEM 3D) algorithm and individual beds merged to form a single 3D image. Different settings can be used in the reconstruction, namely, voxel size of $0.5 \times 0.5 \times 0.5$ or $1 \times 1 \times 1$ mm^3; random, scatter, decay, and shape-based attenuation corrections, as well as different numbers of iterations. Spatial resolutions below 3 mm were measured in most of the field of view [18] and a comparative study between this dbPET scanner and a whole-body PET/CT reported better resolution and detectability of the former [19]. Also, an improvement in visual heterogeneity assessment was reported [20].

Patient cohort

This study was approved by the local ethics committees (CEIC-Galicia) and informed consent was obtained from all participants. The MAMMI-PET project consists of two different research lines. ARM 1 investigates the power of MAMMI-PET as a complementary imaging tool compared to MRI, so every breast cancer patient that was scheduled for MRI scan due to at least one of the following indications, (a) suspected multifocal/multicentric/bilateral disease, (b) discrepancies between ultrasonography and mammography, and (c) hidden tumor, was scanned with MAMMI-PET before any treatment. ARM 2 studies the power of MAMMI-PET in evaluating neoadjuvant treatment response. Every patient that was scheduled to receive neoadjuvant therapy was scanned with MAMMI-PET before treatment beginning. Participants in both ARMs had no previous breast cancer history. If a patient met the inclusion criteria for ARM 1 and ARM 2, then that patient was assigned to ARM 2. From July 2014 to February 2017, consecutive patients from ARM 1 and ARM 2 that met the following two imaging criteria, (1) lesions that were fully contained in the field of view and (2) lesions that were detectable, so that they could be unambiguously located by three independent nuclear physicians, were included in this study. TNM stage was established according to the 7th American Joint Committee on Cancer (AJCC).

PET protocol

Patients fasted for at least 6 h before the injection of 180 MBq of ^{18}F-FDG. To ensure correct incorporation of the radiotracer, blood glucose levels were checked and patients rested in a warm room 30 mins before administration. All PET images were reconstructed with the settings recommended by the vendor: pixel size of 0.5 mm, 12 iterations and attenuation, scatter, and randoms correction. The axial field of view covered by each bed was 40 mm, with an acquisition time of 3 min and 30 s.

Tumor histology and immunohistochemistry

Core-needle biopsy was used for diagnosis. Histological grading was performed using the modified SBR system. Tumors were considered positive for estrogen (ER) or progesterone (PR) receptors according to the Allred method, being positive those tumors with a total score (TS) equal or higher than 3 (TS $\geq$ 3). Tumors were considered HER2-positive if a 3+ score was obtained according to HercepTest criteria. In situ hybridization (FISH) was used to assess positivity of equivocal (2+) cases. IHC classification followed the recommendations of the 13th St. Gallen International Breast Cancer Conference (2013) with a Ki67 threshold of 20% [3].

Texture analysis

Each tumor was manually enclosed in a cropping box by an experienced nuclear physician. A threshold of $0.45 \times \text{SUV}_{\text{max}}$ was applied in the cropping box to segment lesions. All the possible holes of the resulting segmentation were filled with a 3D flood-fill algorithm. Quantitative analysis was carried out in the segmented tumors. In order to allow a precise comparison between whole-body PET and dbPET, we computed the same features as those used in [16], which include first-order statistics features (FOS) and textural features (TF), derived from co-occurrence matrix (CM) [4] and size-zone matrix (SZM) [21]. FOS features take the frequency histogram of tumor uptake values as the only input and therefore do not capture any information of the spatial arrangement of voxels. On the contrary, textural features quantify the presence of spatial structures at a local (CM) or regional (SZM) level. Computed FOS features were SUV_{max}, SUV_{mean}, MATV, TLG = $\text{SUV}_{\text{mean}} \times$ MATV and CH_{AUC} [22]. The latter is the only FOS feature designed to quantify global heterogeneity by means of the area under the curve of the cumulative histogram. TF were calculated after quantization with 64 grey levels [16, 17, 23–25]. CM was computed by finding a coincidence in any of the 13 possible directions, instead of averaging 13 features, each one coming from a single matrix for each direction [25]. The selected features were entropy, homogeneity, and dissimilarity from CM [4], and HILAE and ZP from SZM [21]. All these features have shown robustness with respect to partial volume effects and segmentation [24], reconstruction settings [26], and reproducibility [23]. Most of them were also tested in previous studies, so a comparison of performance between dbPET and current whole-body PET scanners can be carried out. A comprehensive definition of all of the features is provided in Supplemental Methods. Algorithms were implemented in Matlab (The MathWorks, Inc).
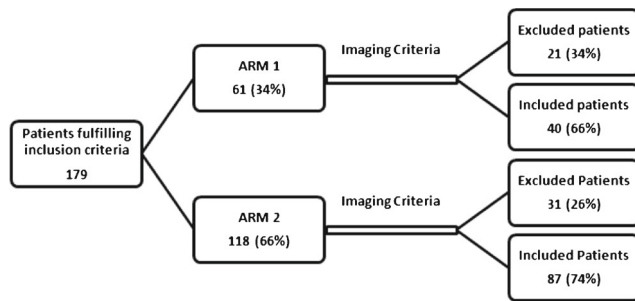
Fig. 1 Diagram showing the number of patients involved the study. Of the excluded patients from ARM 1 (ARM 2), 11 (31) did not show FDG uptake and ten (27) were not fully included in the field of view

Statistical analysis

For descriptive statistics, medians±standard deviations were used. Correlations between features were evaluated using the Spearman rho. Associations between PET features and clinical, histopathological, and IHC factors were computed using a two-sided Wilcoxon rank-sum test. Associations between PET features and molecular subtype were assessed using a Kruskal–Wallis test. All these tests were corrected for multiple testing using the Benjamini–Hochberg method [27] and a further Tukey–Kramer test was applied when evaluating the pairs of molecular groups that are statistically different. The significance level used in the test was the standard $\alpha = 0.05$. ANCOVA analysis was used to adjust for covariates. Discriminative power was quantified with the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. All of the analyses were computed with Matlab (The MathWorks, Inc).

Results

Patient population characteristics

A total of 179 patients participated in ARM 1 and ARM 2, resulting in 127 patients (median age 52±12 years) that fulfilled imaging criteria described in Section "Patient cohort" (Fig. 1). Overall, nine were diagnosed with stage I BC, 74 with stage II, 41 with stage III, and three with stage IV. Multifocal/multicentric disease with more than one biopsied lesion that fulfilled the inclusion criteria was present in ten patients. A total number of 139 tumors were analyzed. Tumor number for each molecular subtype was 36 LA, 50 LB−, 25 LB+, 16 HER2pnl, and 11 TN. Some examples of PET images from these tumors are presented in Fig 2. Patient and tumor characteristics are listed in Table 1.

Correlations between PET features

Correlations between SUV_{max} and FOS features ranged from 0.24 for MATV to 0.99 for SUV_{mean} (0.37 for CH_{AUC} and 0.63 for TLG). TF were less correlated with SUV_{max} than FOS: −0.30 for dissimilarity, 0.03 for entropy, 0.24 for homogeneity, 0.53 for HILAE and −0.34 for ZP. Correlations between MATV and FOS features were low (−0.24 for CH_{AUC}) except for TLG (0.89), as expected. Stronger correlations were found between MATV and TF: −0.86 for dissimilarity, 0.64 for entropy, 0.83 for homogeneity, 0.64 for HILAE, and −0.85 for ZP. Dissimilarity was strongly correlated with TLG (0.80), entropy (0.82), homogeneity (0.88), and ZP (0.97) (Supplementary Table 1).

Relation between PET features and clinical, histological, and immunohistochemical factors

Associations between features and tumor characteristics are summarized in Table 2. Clinical T stage was associated with MATV and TLG, as expected. CH_{AUC} and TF, with the exception of HILAE, were also correlated with T stage. No association between clinical N stage and PET features reached statistical significance after multiple tests correction.

Ductal carcinoma had significantly higher uptake than lobular carcinoma ($p = 0.0022$). None of the heterogeneity metrics was significantly associated with histology. Grade
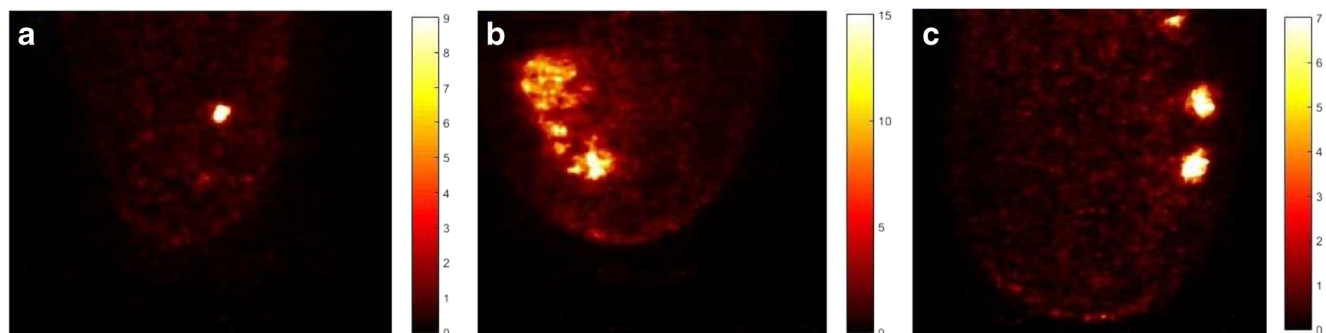

Fig. 2 Examples of some of the analyzed tumors. SUV scale is depicted on the right of each image. **a** T1, HER2pnl tumor, **b** T3, LB− tumor, **c** multifocal tumor

Table 1 Patient and tumor classification

	Number of lesions (%), N = 139
T stage (N=127)	
T1	19 (15)
T2	67 (54)
T3	21 (16)
T4	19 (15)
Not assessable	1 (<1)
N stage (N=127)	
N0	81 (64)
N1	24 (19)
N2	19 (15)
N3	3 (2)
Histology	
Invasive ductal	108 (78)
Lobular	12 (8)
Others	19 (14)
Histologic grade[a]	
1	16 (12)
2	67 (48)
3	55 (40)
Not specified	1 (<1)
Estrogen receptor status	
Positive	111 (79)
Negative	28 (21)
Progesterone receptor status	
Positive	94 (68)
Negative	45 (32)
HER2 status	
Positive	41 (29)
Negative	97 (71)
Not specified	1 (<1)
Molecular subtype	
Luminal A like	36 (26)
Luminal B like/HER2−	50 (36)
Luminal B like/HER2+	25 (18)
HER2pnl	16 (12)
Triple negative	11 (8)
Not specified	1 (<1)

[a]Histologic grade was determined using the modified Scarff–Bloom–Richardson method.

3 lesions were associated with higher uptake compared to grade 1 and 2 lesions and higher TLG. HILAE was also associated with histologic grade ($p < 0.001$). CH_{AUC} and ZP did not reach statistical significance after multiple tests correction.

Estrogen positivity was only associated with SUV_{max} and SUV_{mean}. Progesterone receptor status was significantly associated with SUV_{max} and SUV_{mean}. TLG, CH_{AUC} and HILAE were also significantly associated. HER2 status was significantly correlated with SUV_{max} and SUV_{mean} ($p < 0.001$). This result remained significant after correcting by histologic grade ($p = 0.011$ and $p = 0.006$, respectively). CH_{AUC} and HILAE were also associated with HER2 status (Fig. 3) ($p = 0.002$ and $p = 0.016$, respectively).

All features were significantly associated with molecular subtype after multiple tests correction, although entropy did not reach significance after a further Tukey–Kramer test to determine which pairs of groups were different. Box plots of each feature are presented in Fig. 4. Most of the significant associations involved LA tumors. SUV_{max} and SUV_{mean} were lower for LA tumors ($p < 0.001$) than for the other subtypes. The discriminative power ranged from AUC = 0.717 for LB− to AUC = 0.880 for HER2pnl (Supplementary Figure 1). MATV showed relation $p = 0.023$ between LA and LB− with an AUC = 0.712 (Supplemental Figure 1b). TLG was significantly different ($p < 0.001$) for LA when compared to LB−, HER2pnl and TN. CH_{AUC} was discriminative when comparing LA vs. HER2pnl and LB− vs. HER2pnl ($p = 0.022$). Among TF, dissimilarity, homogeneity, HILAE, and ZP were significantly different when comparing LA and LB− and AUCs were 0.768, 0.712, 0.797 (Supplemental Figure 1b) and 0.789, respectively. HILAE was also significantly correlated for LA vs. LB+ and LA vs. HER2pnl (Supplemental Figure 1a).

Comparison with previous studies

In order to evaluate the performance of dbPET with respect to whole-body PET, we performed a subanalysis to compare results with [16] and [17].

In [16], patients were classified into three IHC groups: TN, ER+/HER2−, and HER2+. Associations between PET features and IHC classification from [16] are summarized in Supplementary Table 2. Only SUV_{max}, SUV_{mean}, and CH_{AUC} were significantly associated with the IHC groups ($p < 0.001$ for SUV, $p = 0.001$ for CH_{AUC}). AUCs were computed using the criteria of [16], comparing each group against the other two together. Discriminative power was better for ER+/HER2− and HER2+ groups than for TN.

In [17], only ER+ tumors were included in the analysis. Significant correlations after multiple tests correction were found for SUV_{max} and SUV_{mean} when discriminating tumor grade ($p < 0.001$, AUCs of 0.758 and 0.760, respectively) and HER2 status ($p = 0.001$ and $p < 0.001$, respectively). HER2 status remained significant after correcting by histologic grade ($p = 0.017$) and $p = 0.014$, respectively. TLG and HILAE remained significant ($p = 0.009$ and $p = 0.012$) for tumor grade, although AUCs were modest (<0.7). T stage associations had low p values (0.006) but were not enough to pass multiple tests correction. These results are listed in Table 3.

Table 2 Associations (p values) between PET features and clinical, histological, and molecular factors. When significant, AUCs comparing each group are presented in parentheses

Feature	T stage 1-2 vs. 3	N stage 0 vs. 1-3	Histology ductal vs. Lobular	Histologic grade 1-2 vs. 3	Estrogen receptor	Progesterone receptor	HER2 status	Molecular subtype
SUV_{max}	0.56	0.56	0.0026 (0.765)	<0.001 (0.750)	<0.001 (0.710)	<0.001 (0.717)	<0.001 (0.705)	<0.001 (LA vs. LB− 0.717, LA vs. LB+ 0.834, LA vs. HER2 0.880, LA vs. TN 0.816)
SUV_{mean}	0.86	0.47	0.0022 (0.770)	<0.001 (0.750)	<0.001 (0.710)	<0.001 (0.720)	<0.001 (0.712)	<0.001 (LA vs. LB− 0.714, LA vs. LB+ 0.832, LA vs. HER2 0.885, LA vs. TN 0.813)
MATV	0.003 (0.707)	0.62	0.90	0.18	0.51	0.69	0.33	0.023 (LA vs. LB− 0.712)
TLG	0.017 (0.668)	0.89	0.14	<0.001 (0.676)	0.017*	0.016 (0.626)	0.25	<0.001 (LA vs. LB− 0.769, LA vs. HER2 0.764, LA vs. TN 0.770)
CH_{AUC}	0.001 (0.733)	0.008*	0.023*	0.036*	0.042*	0.0070 (0.642)	0.0024 (0.664)	0.022 (LA vs. HER2 0.743, LB− vs. HER2 0.735)
Dissimilarity	<0.001 (0.737)	0.15	0.61	0.077	0.52	0.56	0.99	0.0012 (LA vs. LB− 0.768)
Entropy	<0.001 (0.780)	0.011*	0.32	0.83	0.34	0.28	0.089	0.047**
Homogeneity	<0.001 (0.758)	0.15	0.89	0.17	0.78	0.76	0.68	0.0035 (LA vs. LB− 0.712)
HILAE	0.24	0.75	0.071	0.0014 (0.660)	0.015*	0.0041 (0.651)	0.016 (0.630)	<0.001 (LA vs. LB− 0.797, LA vs. LB+ 0.791, LA vs. HER2 0.837)
ZP	0.001 (0.726)	0.30	0.48	0.036*	0.47	0.46	0.99	<0.001 (LA vs. LB− 0.789)

*Not significant after multiple tests correction
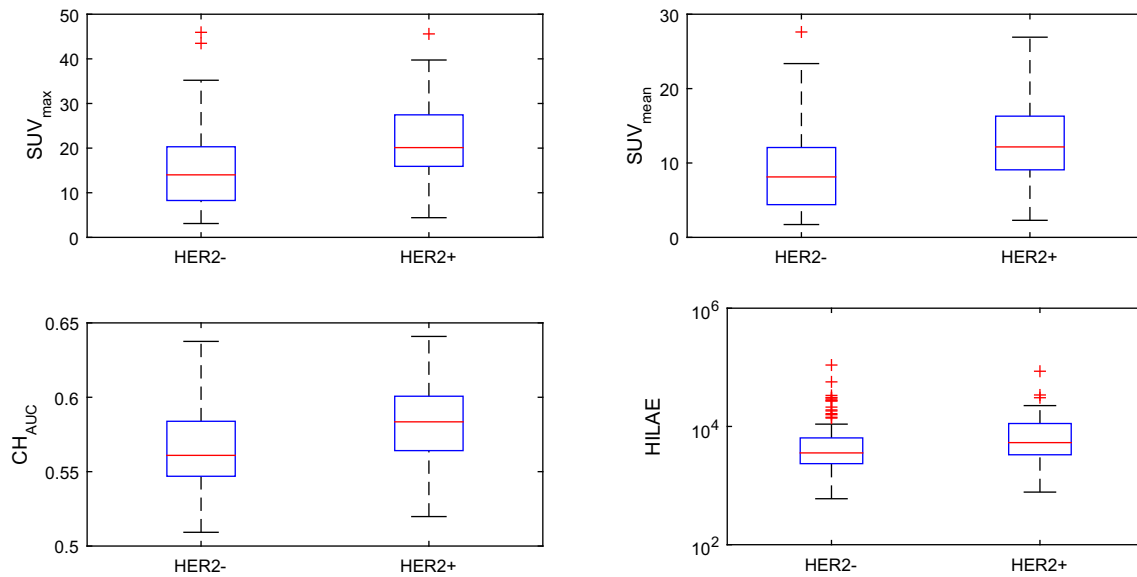
***Not significant after Tukey–Kramer test

Fig. 3 Box plots showing significantly associated features with HER2 status. The *center mark* indicates the median, and the *bottom and top edges of the box* indicate the 25th and 75th percentiles, respectively.

The *whiskers* extend to the most extreme data points not considered outliers, and the *outliers* are plotted individually using the + symbol

Discussion

As a heterogeneous disease, differences in histology and clinical outcome of breast cancer have served as the basis for tumor classification but, recently, this classification has been refined and at times replaced by molecular classifications, which have the ability to combine biological features with clinical outcome. In this context, the primary aim of texture analysis is to non-invasively characterize tumors by

medical imaging. Although correlations with response and outcome were found in other tumors and imaging modalities [8–10, 28], poor associations were found with biological features of breast cancer [14–17]. A possible reason for this failure was pointed out in [29], where it was found that texture analysis with current PET resolutions in tumor volumes below 45 cm^3 may lead to inaccurate heterogeneity quantification. This issue is specially problematic in BC, since most of the tumors are smaller than this limit [14–17]. In this
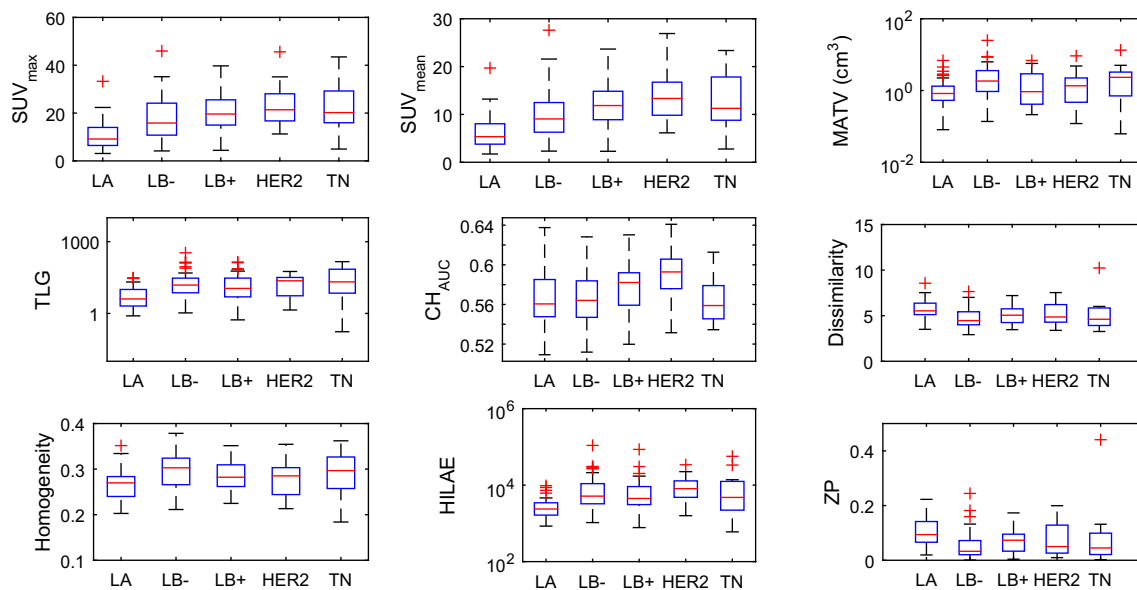

Fig. 4 Box plots showing significantly associated features with IHC subtype. The *center mark* indicates the median, and the *bottom and top edges of the box* indicate the 25th and 75th percentiles, respectively.

The *whiskers* extend to the most extreme data points not considered outliers, and the *outliers* are plotted individually using the + symbol

Table 3 Associations (p values) between PET features and IHC factors with a subsample of 111 ER+ tumors, as studied in [17]. When significant, AUCs comparing each group are presented in parentheses

Feature	T stage 1-2 vs. 3	Histology ductal vs. Lobular	Histologic grade 1–2 vs. 3	Progesterone receptor	HER2 status
SUV_{max}	0.18	0.008*	<0.001 (0.758)	0.006*	0.001 (0.717)
SUV_{mean}	0.30	0.007*	<0.001 (0.760)	0.005*	<0.001 (0.718)
MATV	0.040*	0.99	0.39	0.77	0.31
TLG	0.056	0.25	0.009 (0.657)	0.16	0.43
CH_{AUC}	0.033*	0.037*	0.068	0.038*	0.085
Dissimilarity	0.015*	0.67	0.11	0.56	0.77
Entropy	0.006*	0.37	0.98	0.64	0.50
Homogeneity	0.006*	0.99	0.22	0.62	0.98
HILAE	0.51	0.13	0.012 (0.652)	0.042*	0.090
ZP	0.010*	0.58	0.059	0.50	0.88

*Not significant after multiple tests correction

study, we have measured five FOS and five TF in a cohort of 139 different breast tumors using a dbPET scanner. All these features were previously studied in [16] and most of them in [15, 17].

Our results indicate that HER2-positive tumors show higher metabolism than negative ones, contrary to what was found before by several studies [16, 17, 30–32]. This result remains when considering only ER+ tumors so no influence of ER status is altering our measurement. A possible reason to explain the inability of whole-body PET to detect this association might be its lower spatial resolution, thus yielding a less accurate SUV_{max} measurement for a given HER2-positive tumor than the one measured by our dbPET scanner. Although a recent study has found this association using whole-body PET [33], the authors claim that their finding is only based on the histopathological nature of their patient sample. In any case, further investigation is needed to clarify the origin of the discrepancies among studies.

HILAE, a regional heterogeneity measure, was the only TF that showed correlation with IHC factors. Significant associations with tumor grade, progesterone status, and HER2 status were found. Furthermore, HILAE was the most discriminative feature of the four that were significantly associated with IHC subtype and it was moderately correlated with SUV_{max} (rho=0.53) and MATV (rho=0.64), being TLG its most correlated feature (rho=0.76). Its better discrimination power with respect to SUV when comparing LA to LB−, the two most similar IHC groups in terms of uptake, and the fact that the rest of TF also showed better discrimination than SUVs when comparing these two groups (Fig. 4), may indicate that the role of tumor uptake is minimized while heterogeneity information captured by dbPET becomes more relevant in characterizing LA and LB- tumors. Figure 5 depicts the differences in

uptake distribution for these two types of tumors, having both very similar SUV and MATV. HILAE also showed good discrimination power (AUC >0.75) differentiating LA vs. LB+ and LA vs. HER2pnl. This connects with the above discussion concerning HER2-positive tumors: if such tumors present metabolic heterogeneity at small distance scales, HILAE might be reflecting this characteristic, since LB+ and HER2pnl define our population of HER2-positive tumors.

To compare our results with [16], we used the same IHC classification as that study. We found that SUVs and CH_{AUC} were associated with IHC subtype. In [16], SUVs and TLG were significantly associated. The discriminative power of SUV for TN tumors was lower in our study, but an increase (>0.05) in discriminative power was observed for ER+/HER2− tumors, and a substantial increase (>0.15) for HER2+ tumors. Due to differences in patient population (54 TN in [16] vs. 11 TN in our study) it is not possible to establish a direct comparison of performance between dbPET and whole-body PET. However, we have shown that regrouping the initial five groups into three can spoil any heterogeneity characterization. We must bear in mind that two molecular subtypes are biologically different entities, with completely different gene expressions even when they are differentiated just by a single IHC factor (for instance LA and LB−, which basically differ in Ki67 index) [3, 34]. In this sense, IHC factors act as surrogates to classify different genetic entities, so any molecular classification based only on few of these factors is going to contain a heterogeneous mixture of different biological profiles in each group. Our results suggest that heterogeneity can be captured with dbPET if the more modern five-group IHC classification is used but, as pointed out in [16], this is still pending work in whole-body PET.
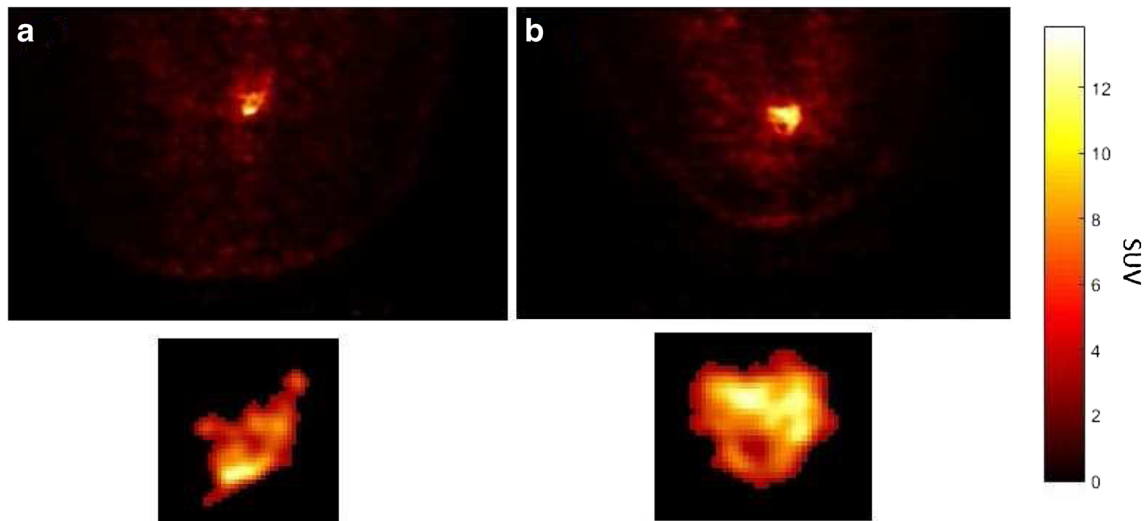
Fig. 5 Example of uptake distributions of LA and LB− tumors with very similar SUV and MATV values but different HILAE, along with resulting segmentations. **a** LA with SUV_{max} = 13.8, MATV = 1.1 cm^3 and HILAE = 2951. **b** LB− with SUV_{max} = 13.8 and MATV = 1.2 cm^3 and HILAE = 6716

Restricting the analysis to ER+ tumors, as in [17], our main results remained unchanged. HER2-positive tumors showed, again, higher uptake values. Apart from correlations of TLG and HILAE with histologic grade, we obtained very low p values for SUV associations with histology and PR status. Taking into account that our patient sample was very similar to the one used in [17], this comparison suggest that MAMMI-PET significantly improves the performance of whole-body PET in quantifying tumor characteristics.

Our study had some limitations. Although we found a positive relation between TF and IHC subtypes, the relation between TF and molecular subtype by means of genomic analyses must be explored, which are the most accurate descriptors of tumor biology and aggressiveness. The results of this work suggest that the existence of this relation is probable. We included stage I and IV patients, making our sample slightly different from the samples studied in [16, 17]. Our segmentation algorithm was chosen due to its simplicity but was different from the algorithms used in [16, 17]. This should not represent a serious problem since the chosen TF demonstrated robustness against the segmentation algorithm [24]. Ki-67 index is a measure with high degree of variation between laboratories and therefore it is important to follow the recommendations from [35] to achieve fully reproducible results.

The results of this work suggest that the direction that must be followed to achieve a complete heterogeneity characterization is two-fold. On one hand, and even though this study may indicate that TF provide additional information, it is essential to have a better understanding of TF, its correlations, and how they behave when measuring complex uptake distributions to determine the biological mechanisms behind different heterogeneity scores. This will allow to select the set of features that contain complementary and relevant information, a crucial step in order to design a machine-learning algorithm that will reliably classify tumor molecular subtypes in a non-invasive way. It should be noted that this methodology can be extended to the prediction of other relevant variables such as clinical outcome or response to treatment, providing that enough data are available and a good selection of input features was performed. On the other hand, an improvement in the quality of PET images, in terms of resolution and noise, may make emerge new and strong associations that are currently disguised, as we have shown in this study. Bearing in mind these strategies, it would be desirable, from a theoretical point of view, to collect more evidence linking heterogeneity in PET images with the expected pathological manifestations such as axillary nodal involvement. In our opinion, these are the guides that should be followed in future studies.

Conclusions

Our study suggests that dedicated breast PET resolution improvement enables a more precise heterogeneity characterization. Thus, TFs were able to discriminate IHC subgroups, showing its best performance when comparing LA and LB−, the two subgroups with most similar uptake levels. Due to the low correlation of TF with SUV, this result suggests that there is an actual metabolic difference in the uptake distribution of both types of tumors. Additionally, HER2+ tumors showed increased SUV, a result that was not reported before.

References

1. Heppner GH. Tumor heterogeneity. Cancer Res. 1984;44:2259–65.
2. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;194:23–28.
3. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thrlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol. 2013;24(9):2206–23.
4. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern. 1973;SMC-3:610–21.
5. Amadasun M, King R. Textural features corresponding to textural properties. IEEE Trans Syst Man Cybern. 1989;19:1264–74.
6. Alic L, Niessen WJ, Veenland JF. Quantification of heterogeneity as a biomarker in tumor imaging: a systematic review. PLoS ONE. 2014;9:1–1.
7. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJR. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging. 2013;40:133–40.
8. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med. 2011;52:369378.
9. Lovinfosse P, Janvary ZL, Coucke P, Jodogne S, Bernard C, Hatt M, et al. FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy. Eur J Nucl Med Mol Imaging. 2016;43:1453–60.
10. Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy?. J Nucl Med. 2013;54:1926.
11. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. Engl J Med. 2012;366:883892.
12. Zardavas D, Irrthum A, Swanton C, Piccart M. Clinical management of breast cancer heterogeneity. Nat Rev Clin Oncol. 2015;12(7):381–94.
13. Bastien RR, Rodrguez-Lescure Á, Ebbert MT, Prat A, Munárriz B, Rowe L, et al. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. BMC Med Genomics. 2012;5:44.
14. Son SH, Kim D-H, Hong CM, Kim C-Y, Jeong SY, Lee S-W, et al. Prognostic implication of intratumoral metabolic heterogeneity in invasive ductal carcinoma of the breast. BMC Cancer. 2014;14:585.
15. Soussan M, Orlhac F, Boubaya M, Zelek L, Ziol M, Vronique E, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. PLoS One. 2014;9(4):e94017.
16. Groheux D, Majdoub M, Tixier F, Le Rest CC, Martineau A, Merlet P, et al. Do clinical, histological or immunohistochemical primary tumour characteristics translate into different (18)F-FDG PET/CT volumetric and heterogeneity features in stage II/III breast cancer? Eur J Nucl Med Mol Imaging. 2015;42(11):1682–91.
17. Lemarignier C, Martineau A, Teixeira L, Vercellino L, Espi M, Merlet P, et al. Correlation between tumour characteristics, SUV measurements, metabolic tumour volume, TLG and textural features assessed with 18F-FDG PET in a large cohort of oestrogen receptor-positive breast cancer patients, Eur J Nucl Med Mol Imaging 2017. https://doi.org/10.1007/s00259-017-3641-4.
18. Moliner L, González AJ, Soriano A, Sánchez F, Correcher C, Orero A, et al. Design and evaluation of the MAMMI dedicated breast PET. Med Phys. 2012;39:5393–5404.
19. García Hernández T, Vicedo González A, Ferrer Rebolleda J, Sánchez Jurado R, Roselló Ferrando J, Brualla González L, et al. Performance evaluation of a high-resolution dedicated breast PET scanner. Med Phys. 2016;43:2261–72.
20. Koolen BB, Vidal-Sicart S, Benlloch Baviera JM, Valdés Olmos RA. Evaluating heterogeneity of primary tumor (18)F-FDG uptake in breast cancer with a dedicated breast PET (MAMMI): a feasibility study based on correlation with PET/CT. Nucl Med Commun. 2014;35(5):446–52.
21. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Texture indexes and gray level size zone matrix: application to cell nuclei classification. In: 10th international conference on pattern recognition and information processing, PRIP. Minsk, Belarus; 2009. p. 140–145.
22. Van Velden FHP, Cheebsumon P, Yaqub M, et al. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imaging. 2011;38:1636.
23. Tixier F, Hatt M, Cheze Le Rest C, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. J Nucl Med. 2012;53:693–700.
24. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur J Nucl Med Mol Imaging. 2013;40:1662–71.
25. Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multicancer site patient cohort. J Nucl Med. 2015;56:38–44.
26. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST. Impact of image reconstruction settings on texture features in 18F-FDG PET. J Nucl Med. 2015;56:1667–73.

27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B Methodol. 1995;57(1):289–300.

28. 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.

29. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. J Nucl Med. 2014;55(1):37–42.

30. Jadvar H, Alavi A. Gambhir SS, 18F-FDG uptake in lung, breast, and colon cancers: molecular biology correlates and disease characterization. J Nucl Med. 2009;50(11):1820–7.

31. Osborne JR, Port E, Gonen M, Doane A, Yeung H. Gerald W, others. 18F-FDG PET of locally invasive breast cancer and association of estrogen receptor status with standardized uptake value: microarray and immunohistochemical analysis. J Nucl Med. 2010;51(4):543–50.

32. Koolen BB, Vrancken Peeters MJ, Wesseling J, Lips EH, Vogel WV, Aukema TS. Association of primary tumour FDG uptake with clinical, histopathological and molecular characteristics in breast cancer patients scheduled for neoadjuvant chemotherapy. Eur J Nucl Med Mol Imaging. 2012;39(12):1830–8.

33. Lee SS, Bae SK, Park YS, Park JS, Kim TH, Yoon HK. Correlation of molecular subtypes of invasive ductal carcinoma of breast with glucose metabolism in FDG PET/CT: Based on the Recommendations of the St. Gallen Consensus Meeting 2013. Nucl Med Mol Imaging. 2017;51(1):79–85.

34. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009;101(10):736–50.

35. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst. 2011;103(22):1656–64.