Tumor Texture Analysis in PET: Where Do We Stand?

Irène Buvat[1], Fanny Orlhac[1], and Michaël Soussan[1,2]

[1]*Imagerie Moléculaire In Vivo, UMR 1023 Inserm/CEA/Université Paris Sud—ERL 9218 CNRS, CEA Service Hospitalier Frédéric Joliot, Orsay, France; and* [2]*Department of Nuclear Medicine, AP-HP, Avicenne Hospital, Bobigny, France*

The coexistence of different tumor cells that show distinct morphologic and phenotypic features either within a tumor or between tumors defines tumor heterogeneity. The identification, characterization, understanding, and, possibly, treatment of tumor heterogeneity are key challenges in oncology and should help design effective therapeutic and monitoring strategies (*1*). Because biopsies probe only parts of the tumors, they do not necessarily reflect tumor heterogeneity (*2*). Additional techniques are needed, among which imaging is an appealing approach to comprehensively detect, depict, and quantify local variations in tumor morphology and function. Several hundred published articles have investigated the beneficial information that can be extracted from the analysis of tumor heterogeneity using imaging since the beginning of the nineties, mostly involving MR and ultrasonography (>70% of the articles) (*3*), with a significant increase in publications since 2008. The interest in exploring tumor heterogeneity using PET dates back to 2009 (*4*) and is conceptually quite appealing given that PET reflects the biology of the tumor. Yet, a synthetic synopsis of the published results and subsequent

conclusions is extremely difficult because of major methodologic variety in the conducted studies, as described below.

Various image analysis approaches can be used to characterize tumor heterogeneity (*5*). In PET, the 2 most frequent approaches are the methods based on the analysis of the histogram of the voxel values within the tumor and the methods accounting for the spatial arrangement of voxel values. In the histogram-based methods, the heterogeneity descriptors (HDs) disregard the inherent spatial relationship between voxel values, only reflecting the voxel-value frequency distribution. They include the mean, SD, median, skewness, kurtosis, percentile values, range of standardized uptake value (SUV), entropy, and energy and are called first-order statistics (FOS). The second approach accounts for the spatial arrangement of the voxel values within the tumor using

higher-order statistics by first calculating a 2-dimensional matrix describing this spatial organization. This matrix is often the gray-level cooccurrence matrix (GLCM), giving the probability of observing a pair of voxel values at a given distance in a given direction (*6*). Several other matrices are also used, among which are the neighborhood gray-tone difference matrix, which provides information regarding how each voxel value differs from the neighbor voxel values; the gray-level run length matrix, which stores the number of voxels with identical values in each direction; and the gray-level size zone matrix, which stores the size of the 3-dimensional (3D) region that includes a given voxel value. All these matrices capture some spatial relationship between voxel values, and each matrix enables the calculation of several HDs, therefore yielding several dozen HDs. This large number of descriptors complicates the overview of published results because not all authors use the same HD. A publication bias is likely as it is often unclear whether HDs other than the presented ones were studied and did not perform well. In addition, the same descriptor name is used for descriptors calculated from different definitions or different matrices. For instance, homogeneity does not obey the same definition in the study by El Naqa et al. (*4*) and in Tixier et al. (*7*). Entropy is not defined identically in 2 different articles by Tixier et al. (*8,9*). A contrast feature can be calculated from the GLCM or from the neighborhood gray-tone difference matrix, whereas entropy can be calculated from the gray-level histogram or from the GLCM. Moreover, different descriptor names sometimes obey the same definition. The short zone high gray level emphasis HD is called Szonehigl (*10*), high-intensity short-zone emphasis (*7*), or high-intensity small-area emphasis (*8*). This lack of standardization in the names and precise definition of the various HDs creates confusion. Last, for each HD following a precise definition, several calculation options are possible. These options include the way the tumor voxel values are rescaled between a minimum and maximum SUV (SUV_{max}) before subsequent HD calculation or the 3D extension of the 2-dimensional descriptor definition to accommodate the 3D nature of the tumor. When speaking about an HD, it is therefore essential to provide its precise name and definition, including the matrix it is derived from and its 3D calculation method.

An additional source of confusion is that PET images have limited spatial resolution and (smoothed) noise, which themselves introduce some textural pattern and local signal correlation. It is of foremost importance to understand how the heterogeneity descriptors are affected by these resolution and noise components, which do not relate to the underlying biologic signal. In this issue of *The Journal of Nuclear Medicine,* Yan et al. (*11*) contribute to that understanding by reporting a thorough analysis of the robustness of 55 textural features derived from matrices and of 6 FOS

with respect to the changes in the iterative reconstruction scheme (changes in the algorithm options, in postfiltering, in iteration number, and in reconstructed grid size) that directly affect the spatial resolution and noise in the PET images. A similar analysis had previously been reported (12) but the novelty of the Yan et al. study is that up-to-date iterative reconstructions taking advantage of time-of-flight information and modeling of the point-spread function of the imaging systems have been used, and a more detailed analysis of the HD variation as a function of different parameters is presented. Among all parameters included in the analysis, they found that the voxel size (or grid size) affects the HD value the most, followed by the full width at half maximum of the gaussian postprocessing filter applied to the reconstructed images. Neither the number of iterations nor the actual reconstruction scheme (with or without time-of-flight information, with or without point-spread function modeling) affected the HD values much. As SUV_{max} and peak SUV (SUV_{peak}) have been proven useful to characterize tumors, comparing the robustness of HD with respect to that of SUV_{max} and SUV_{peak} gives insight into how useful HD might be. Considering all sources of variability investigated in their study (grid size, postprocessing filter full width at half maximum, reconstruction scheme, and iteration number), 7 HDs were as robust as or even more robust than SUV_{peak} and SUV_{max}: entropy FOS, difference entropy (DE), inverse difference (ID), inverse difference moment (IDM), and inverse difference moment normalized (IDMN) from the GLCM; low grey-level run emphasis (LGRE) and high grey-level run emphasis (HGRE) from the gray-level run length matrix; and low grey-level zone emphasis (LGZE) from the gray-level size zone matrix. Six other indices were also robust although not as much as the 7 previous ones, including the largely investigated entropy from the GLCM and high grey-level zone emphasis (HGZE) index. These findings are important because they prompt us to focus on 13 indices instead of considering the large number of possible HDs. The number of HDs of interest could actually be further reduced by accounting for the large correlation existing between some indices. In particular, Orlhac et al. (13) have shown that LGZE and LGRE were highly correlated and similarly for HGZE and HGRE. The contribution of Yan et al. (11) combined with results regarding the correlation existing among HDs (13) and with findings about the robustness of HD with respect to different segmentation approaches (13–15) and to test–retest scans (8,15) definitely help to select HDs of major interest in PET. When these results are analyzed, HGRE (or HGZE) is a good candidate (LGRE and LGZE are not robust enough to segmentation (13), as well as entropy, DE, IDM, IDMN, sum average (SA) and sum entropy (SE) from the GLCM, and FOS entropy. The ID and small number emphasis (SNE) indices identified as robust by Yan et al. have not been investigated in detail by others and might be worth additional investigation, as well as the redundancy between all these HDs.

Even focusing on a small number of robust HDs, many investigations still have to be performed before HD can be soundly used to assist in the assessment of tumor biologic heterogeneity. Indeed, some HDs are highly correlated with the metabolic volume (MV) of the tumor (13,14,16). This high correlation explains the results published in Tixier et al. (7), in which the authors concluded that some HDs predicted the tumor response in esophageal cancer patients whereas the MV only was already highly predictive of tumor response in the very same patients (17). For instance, the intensity variability (identical to grey-level nonuniformity for zone, noted GNLUz or GNLz)

shown as highly predictive of tumor response by Tixier et al. (7), is actually mostly a surrogate of the MV (13). Such misleading data interpretation could be avoided by systematically performing an adequate multivariate analysis to demonstrate the real added value of HDs with respect to conventional index (especially SUV and MV). Incidentally, it was recently demonstrated that the high correlation between HD and MV is introduced in part by the tumor-dependent SUV rescaling step involved in HD calculation (18) and that a tumor-independent SUV rescaling similar to that proposed by Leijenaar et al. (15) removes most of that correlation, while introducing a correlation between HD and SUV. This raises the question of which rescaling is the most relevant. One hint might come from the interpretation of the HD. In Tixier et al. (9), when tumor-dependent SUV rescaling was used, tumors visually rated as the most heterogeneous by doctors had the highest homogeneity index, which is extremely counterintuitive. This prompts some in-depth analysis of the actual meaning of HDs and of how they relate to the visual assessment of tumor heterogeneity. Also, a practical use of HDs will require some guidance regarding which HD values reflect heterogeneous tumor uptake. Because most studies performed so far were retrospective, interpretation rules have not yet been established. For deriving such rules, comparing tumor heterogeneity measured from PET images with that observed in histologic specimens might be useful, to determine which type and level of heterogeneity can be captured and quantified through PET and to bridge the gap between in vivo and ex vivo tumor characterization.

On the basis of all PET studies that have focused on tumor heterogeneity so far, what have we learned regarding its potential clinical value? A recent extensive literature analysis (19) could not find enough evidence to support a relationship between PET HD and patient outcome in cancer patients, mostly due to an inappropriate control of type I error in studies that often investigate many HDs on a single dataset. A definite answer regarding the usefulness of HD in PET for enhancing tumor characterization now requires some standardization of HD calculation, precise reporting on HD selection, and thorough investigation of the many open methodologic questions previously mentioned. These efforts are absolutely needed to make the most of all that PET images can offer for characterizing tumor biologic heterogeneity.

DISCLOSURE

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta.* 2010;1805:105–117.
2. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med.* 2012;366:883–892.
3. Alic L, Niessen WJ, Veenland JF. Quantification of heterogeneity as a biomarker in tumor imaging: a systematic review. *PLoS One.* 2014;9:e110300.
4. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009;42:1162–1171.
5. O'Connor JPB, Rose CJ, Waterton JC, Carano RAD, Parker GJM, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clin Cancer Res.* 2015;21:249–257.
6. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern.* 1973;3:610–621.

7. Tixier F, Cheze Le Rest C, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline ^{18}F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011;52:369–378.

8. Tixier F, Hatt M, 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.

9. Tixier F, Groves AM, Goh V, et al. Correlation of intra-tumor 18f-FDG uptake heterogeneity indices with perfusion CT derived parameters in colorectal cancer. *PLoS One.* 2014;9:e99567.

10. Willaime JMY, Turkheimer FE, Kenny LM, Aboagye EO. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of ^{18}F fluorothymidine-positron emission tomography. *Phys Med Biol.* 2013;58:187–203.

11. Yan J, Chu-Shern JL, Loi HY, et al. Impact of image reconstruction settings on texture features in 18F-FDG PET? *J Nucl Med.* 2015;56:1667–1673.

12. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol.* 2010;49:1012–1016.

13. Orlhac F, Soussan M, Maisonobe J-A, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in ^{18}F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med.* 2014;55:414–422.

14. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging.* 2013;40:1662–1671.

15. Leijenaar RTH, Carvalho S, Velazquez ER, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol.* 2013;52:1391–1397.

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

17. Hatt M, Visvikis D, Pradier O, Cheze-le Rest C. Baseline ^{18}F-FDG PET image-derived parameters for therapy response prediction in oesophageal cancer. *Eur J Nucl Med Mol Imaging.* 2011;38:1595–1606.

18. Orlhac F, Soussan M, Chouahnia K, Martinod E, Buvat I. Enhanced textural indices distinguish tissue types in FDG PET. *J Nucl Med.* 2015;56(suppl 3):430.

19. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS One.* 2015;10: e0124165.