Journal of Nuclear Medicine, published on December 11, 2014 as doi:10.2967/jnumed.114.144055

FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort

Mathieu Hatt[1], Mohamed Majdoub[1*], Martin Vallières[2*], Florent Tixier[1,3], Catherine Cheze Le Rest[3,1], David Groheux[4], Elif Hindié[4], Antoine Martineau[4], Olivier Pradier[1,5], Roland Hustinx[6], Remy Perdrisot[3], Remy Guillevin[7], Issam El Naqa[2], Dimitris Visvikis[1]

* Equally contributed

[1] INSERM, UMR 1101 LaTIM, Brest, France.

[2] Department of Oncology, McGill University, Montreal, Canada.

[3] Nuclear Medicine, CHU Milétrie, Poitiers, France

[4] Nuclear Medicine, CHU Saint Louis, Paris, France.

[5] Radiotherapy, CHRU Morvan, Brest, France.

[6] Nuclear Medicine, CHU Liège, Belgium.

[7] Radiology, CHU Milétrie, Poitiers, France.

Running title: FDG-PET heterogeneity and volume

Corresponding author: M. Hatt,

INSERM, UMR 1101, LaTIM

CHRU Morvan, 2 avenue Foch

29609, Brest, France

Tel: +33(0)2.98.01.81.11

Fax: +33(0)2.98.01.81.24

E-mail: hatt@univ-brest.fr

Wordcount: 4925

Disclosure of Conflicts of Interest: No conflicts of interest.

Funding: This work has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference ANR-10-LABX-07-01.

ABSTRACT

Intra-tumor uptake heterogeneity in ^{18}F-FDG PET has been associated with patient treatment outcomes in several cancer types. Textural features (TF) analysis is a promising method for its quantification. An open issue associated with the use of TF for the quantification of intratumoral heterogeneity concerns its added contribution and dependence on the metabolically active tumor volume (MATV), which has already been shown as a significant predictive and prognostic parameter. Our objective was to address this question using a larger cohort of patients covering different cancer types.

Methods: A single database of 555 pre-treatment ^{18}F-FDG PET images (breast, cervix, esophageal, head & neck and lung cancer tumors) was assembled. Four robust and reproducible TF-derived parameters were considered. The issues associated with the calculation of TF using co-occurrence matrices (such as the quantization and spatial directionality relationships) were also investigated. The relationship between these features and MATV, as well as among the features themselves was investigated using Spearman rank coefficients, for different volume ranges. The complementary prognostic value of MATV and TF was assessed through multivariate Cox analysis in the esophageal and NSCLC cohorts.

Results: A large range of MATVs was included in the population considered (3-415 cm^3, mean=35, median=19, SD=50). The correlation between MATV and TF varied greatly depending on the MATVs, with reduced correlation for increasing volumes. These findings were reproducible across the different cancer types. The quantization and the calculation method both had an impact on the correlation. Volume and heterogeneity were independent prognostic factors (P=0.0053 and 0.0093 respectively) along with stage (P=0.002) in NSCLC, but in the esophageal tumors, volume and heterogeneity had less complementary value due to smaller overall volumes.

Conclusion: Our results suggest that heterogeneity quantification and volume may provide valuable complementary information for volumes above 10cm^3, although the complementary information increases substantially with larger volumes.

Keywords: ^{18}FDG-PET/CT, heterogeneity, textural features, metabolically active tumor volume, prognosis.

^{18}F-FDG Positron Emission Tomography/Computed Tomography (PET/CT) is a powerful tool for diagnosis and staging in oncology (1). Its use in therapy assessment (2,3) is increasing. Within this context more and more emphasis is being given to image-derived indices (4). On the one hand, features extracted from PET images, including metabolically active tumor volume (MATV), mean standardized uptake value (SUV_{mean}) and total lesion glycolysis (TLG) have provided potentially higher prognostic value than standard maximum SUV (SUV_{max}) in various cancer types (5). On the other hand, more recently the heterogeneity of ^{18}F-FDG uptakes within tumors has been associated with treatment failure (4,6–8). Proposed approaches for the assessment of intra-tumor activity distribution heterogeneity include visual evaluation (9), SUV coefficient of variation (SUV_{COV}) (10), area under the curve of the cumulative histogram (CH_{AUC}) (11), fractals (12) or textural features (TF) analysis (10,13). The latter can provide a number of parameters quantifying tumor heterogeneity at the scales of voxels or groups of voxels. A recent study, based on the use of one of these parameters (local entropy calculated from co-occurrence matrices) has suggested that a minimum MATV of 45cm^3 is required to provide an estimate of heterogeneity independent of MATV confounding effects (14). However, this study investigated a single heterogeneity parameter, in a single cancer type, and used an image quantization scheme with >150 grey levels. Another recent study investigating the relationship between MATV and TF (15) has included a relatively small number of tumors (including patients who already had distant metastases) and cancer types, without explicitly reporting on the lesion sizes considered or making conclusions regarding the minimum tumor volume that should be considered. In addition, this study used Pearson correlation to test linear relationship between MATV and TF, which would miss non-linear trends that may exist between these parameters.

To our knowledge, the potential interaction between MATV and TF has not been previously considered within the context of patient outcome prognosis.

This study was therefore designed to investigate in detail, and across a large number of primary tumors and cancer types, the relationship between tumor MATV and derived heterogeneity measurements using TF, in order (a) to determine whether a minimum MATV should be considered in such analyses and (b) if tumor heterogeneity quantified through TF could provide complementary prognostic value relative to MATV.

MATERIALS AND METHODS

FDG PET images

Several patient cohorts were retrospectively collected in a dataset of 555 ^{18}F-FDG PET baseline images of different primary locally advanced tumors, excluding patients with distant metastases because they usually have a very different prognosis and treatment management. In all cohorts, patients were selected as consecutive patients with an ^{18}F-FDG PET/CT scan at diagnosis prior to any treatment.

The resulting dataset consisted of 158 breast tumors with three different subtypes (luminal, her2+/ER-, triple negative), 45 cervix tumors, 112 esophageal tumors, 139 head and neck (H&N) tumors, and 101 non-small cell lung cancer (NSCLC) tumors (Fig. 1). These tumor entities were chosen because they have often been considered in the literature for studying FDG uptake heterogeneity. They have a wide range of tumor size,

significant FDG uptake, and a high rate of treatment failures. Each cohort for a given cancer site came from one clinical center except H&N tumors that involved two different University Hospitals (Supplemental Table 1). Within each cohort, all acquisitions followed corresponding institutional protocols. All cohorts except the 66 H&N tumors from McGill were acquired using the same acquisition settings and protocols: a Philips GEMINI PET/CT scanner, CT-based attenuation correction, reconstruction with 3D Row-Action Maximum Likelihood Algorithm with $4\times4\times4mm^3$ voxels and 5mm full-width-at-half-maximum Gaussian post-filtering, 6-hour fasting period, 3D whole-body acquisition performed 60 min after injection of ^{18}F-FDG, and SUV normalized using body weight. The acquisitions for the 66 H&N tumors from McGill shared the same settings expect for the scanner model (a GE Discovery ST) and Ordered-Subset Expectation Maximization reconstruction ($3.52\times3.52\times3.27mm^3$ voxels) without post-filtering.

Only primary tumors (not lymph nodes) with MATVs $>3cm^3$ (which, assuming a spherical shape, corresponds to ~1.8cm diameter) were included due to the limitations of PET imaging to characterize tracer distribution within smaller tumors because of its limited spatial resolution and resulting partial volume effects (PVE).

The institutional review boards of each involved group in this work approved this retrospective study and the requirement to obtain informed consent was waived.

Image Analysis

MATVs were first delineated using the Fuzzy Locally Adaptive Bayesian (FLAB) algorithm (16), which has been previously validated for accuracy and robustness using simulated and clinical datasets, including homogeneous and heterogeneous MATVs (17,18). FLAB was used with two or three classes in order to adequately cover the entire MATV, including low uptake areas. The H&N tumors from McGill were manually delineated by an experienced radiation oncologist on fused PET/CT images as part of their radiotherapy treatment planning.

The differences in scanner, reconstruction and delineation between the H&N McGill dataset and the other cohorts were taken into account by only considering features that were shown to be the most robust with respect to PVE and segmentation (19) or reconstruction settings (20), as well as test-retest reproducibility (21).

Consequently, the present analysis included four TFs, which have been previously shown to have a predictive and prognostic value in different cancer types. The two local TFs calculated using co-occurrence matrices (13) were entropy (E) and dissimilarity (D). D and E were calculated according to two different methods: 1) using 13 matrices, one for each spatial direction, followed by averaging the values calculated separately in each matrix, and 2) using only one matrix taking into account all 13 directions simultaneously without an averaging step. From a conceptual point of view, the second method is more accurate, as it fully describes the 3D co-occurrence properties within the volume. The first method results in taking an average from fewer co-occurrence measurements, thus information could be lost and the complexity of the distribution of grey levels may not be

optimally captured. On the other hand, averaging could artificially reduce the effect of the residual noise from the quantization process. The parameters related to the first method (averaging 13 matrices) will be denoted as E^{13} and D^{13}, whereas the parameters from the second method (1 matrix) will be denoted as E^{1} and D^{1}.

Regional TFs calculated using size-zone matrices (13) were high intensity large area emphasis (HILAE) and zone percentage (ZP).

Other TFs were not included in the present analysis because of high correlation with D and/or E (correlation>0.8), or previously shown poor reproducibility/robustness (19–21) *(supplemental Table 2)*.

Although an optimal quantization of 64 grey levels was identified in our previous work concerning the reproducibility (19), we also investigated the impact of the quantization pre-processing step by considering values from 4 to 256.

Finally, SUV_{max}, SUV_{mean}, and SUV_{COV} were included for comparison purposes.

Statistical Analysis

Statistical analyses were performed using Medcalc™ (MedCalc Software, Belgium). Spearman rank correlation (*rs*) was used to study relationships between parameters, since such relationships are non-linear and all these parameters are frequently not normally distributed. In order to assess the potential complementarity and/or additional clinical value of MATV and derived heterogeneity parameters, a survival analysis was performed in the esophageal and NSCLC cohorts for which overall survival (OS), follow-up and other clinical data were available (Supplemental Tables 3 and 4), which was not the case for the other cohorts. Association with OS was assessed using univariate Cox proportional hazard regression including features as continuous variables (*i.e.* no dichotomization). Correction for multiple testing was performed using the false discovery rate Benjamini-Hochberg step-up procedure. It consists in declaring positive discoveries at level α (here α=0.05), among the $k=1\ldots K$ tested variables ordered according to their p-values p in increasing order, those ranked above the one satisfying the condition $p(k) \leq \frac{k}{K} \times \alpha$ (22). After univariate analysis, a stepwise multivariate Cox analysis (significant variables are entered sequentially, then removed if they become non-significant) was performed to identify independent prognostic factors. Kaplan-Meier survival curves (with optimal cut-off values determined through ROC curves analysis) were subsequently used to assess the prognostic value of previously identified independent factors. Median survival, percentage of deaths in each group and hazard ratios (HR) were reported for each risk factor separately and for their combination, in order to quantify any improved patient stratification.

RESULTS

Correlation between parameters

To avoid confusion, absolute *rs* values are reported, although correlation direction results can be found in figure 2. Three important observations can be emphasized.

First, significant details regarding the grey-levels distribution are lost when using a quantization <32 (Fig. 3), and the quantization had an important impact on the correlation between volume and E^1, decreasing from almost 1 for a quantization of 256, to <0.6 for quantization<64. Correlation with MATV was much lower for E^1 compared to E^{13}, except for quantization <16. In contrast, dissimilarity (either D^1 or D^{13}) was insensitive to the quantization value (*rs*~0.8) and there was no difference between D^1 and D^{13}. The correlation of HILAE and ZP with MATV was very sensitive to the quantization, although contrary to entropy, *rs* increased from <0.5 to >0.8 with decreasing quantization (Fig. 3). For quantization=32, the correlation with volume was >0.75 and <0.85 for all TFs except E^1 (0.3), whereas with quantization=64 there was a wider range (*rs*<0.2 for HILAE to *rs*≈1 for E^{13}), and several TFs had a correlation <0.7, suggesting a higher potential of complementary information with respect to MATV for 64 than 32. Quantization=64 was also previously shown to provide the highest TFs' reproducibility (21) and robustness (19). A quantization into 64 grey-levels was thus considered for the rest of the analysis, as it represents the best compromise between sufficient sampling of voxel SUVs, preservation of original intensities information, and potential complementary information with respect to MATV.

Secondly, significant correlations were found amongst almost all features considered. In addition, MATV correlated with those features, as well as with SUV_{max} and SUV_{mean} (Fig. 2). The correlation between the co-occurrence matrix-derived features themselves and their respective correlation with MATV were sensitive to the use of a single matrix compared to averaging 13 directional matrices. This was particularly true for entropy. The correlation between D^{13} and E^{13} was 0.76, whereas the correlation between D^1 and E^1 was 0.18. Correlation of D^{13} and E^{13} with MATV was 0.80 and 0.96, whereas it was 0.82 and 0.56 for D^1 and E^1 respectively (Fig. 2).

Thirdly, correlation between D^1, E^1, ZP, and HILAE with MATV ranged from 0.17 to 0.96, suggesting that a substantial amount of complementary information with respect to MATV may be found in some of these heterogeneity quantification features, similarly to SUV measurements which exhibited correlations of 0.31 to 0.42.

Figures 4 and 5 provide visual representations of the distributions of TFs with respect to MATV. Supplemental figures 1-4 provide similar graphs for each cancer type, for other TFs and quantization values. The difference between E^{13} and E^1 was important (Fig. 4A-B), the calculation using only one matrix leading to a much tighter distribution with a smaller range (5.6-8 for E^1 *vs.* 3.9-7.3 E^{13}) of higher values (7.1±0.4 *vs.* 5.7±0.8), and with substantially lower correlation with volume (0.56 *vs.* 0.96). The difference between D^{13} and D^1 (Fig. 4C-D) was less important but nonetheless led also to a tighter distribution for D^1 with a smaller range (3.5-24.1 *vs.* 0.9-30.5) of higher values (12.6±3.8 *vs.* 10.7±3.7) but with a similar correlation with MATV (0.82 *vs.* 0.80).

By restricting the analysis to larger tumor volumes (from ≥10cm^3 to ≥60cm^3 using 5cm^3 steps), it was found that the correlation between TFs and MATV tended to decrease substantially with ranges of increasing volumes (Fig. 6). This was observed for all TFs but was especially true for entropy. The only exception to this observation was HILAE since its correlation with MATV was low even when considering the entire dataset. E^{13} correlation with MATV dropped from 0.96 when considering the entire range of volumes, to <0.3 when considering only tumors >60cm^3. E^1 dropped to <0.25 for volumes >10cm^3. Similar observations were made for D^{13} with rs>0.8 when considering all tumors, dropping <0.6 in tumors >15cm^3. Although the overall correlation of D^1 with MATV was slightly higher than for D^{13}, it was also more rapidly reduced with increasing MATV (Fig. 6). The same analysis for ZP led to similar observations, with a reduced correlation with increasing MATV when considering larger tumors, from 0.68 for all tumors to 0.5 for those >15cm^3.

The relationships between MATV and TFs were similar across the different tumor types, although measured correlations varied: cervix tumors included mostly large tumors (MATV>20cm^3), whereas the other tumor types included a larger number of tumors with MATV<10cm^3 (Fig. 1). The resulting correlation between E^1 and MATV was not significant for cervix cancer tumors (*rs*<0.003, *P*>0.9), whereas it was for esophageal, NSCLC, H&N, and breast tumors (*rs*=0.80, 0.39, 0.35, and 0.74 respectively). For dissimilarity (D^1) and ZP, smaller differences were observed, with *rs* from 0.73 and 0.63 (cervix) to 0.93 and 0.83 (NSCLC) for D1 and ZP respectively. HILAE correlations with MATV were <0.2 for all tumor types.

Survival prognosis in esophageal and NSCLC cohorts

In the esophageal cohort, median OS was 17.0 months (range 1.0-71.0, mean 22.0), with 44 patients still alive at last follow-up. The trend for MATV did not reach statistical significance (p=0.0315) and none of the variables were prognostic factors except D^1 (p=0.0016) (Table 1), therefore no multivariate analysis was performed. However, if dichotomized with optimal cut-off values in the K-M analysis, both D^1 (HR 1.92, p=0.0052) and MATV (HR 1.66, p=0.0375) could differentiate survival curves (Fig. 7A). Adding volume and heterogeneity increased HR to 2.02 (p=0.0024, 95% CI 1.22-3.34) with 23months (*N*=64) *vs.* 10months (*N*=48) median OS (supplemental table 5, Fig. 7A).

In the NSLC cohort, median OS was 18.4 months (range 1.1-57.4, mean 27.4), with 39 patients still alive at last follow-up. In the univariate analysis, all variables were significantly associated with OS, expect age, histology, smoking history and SUV_{max} (Table 1). The multivariate analysis retained clinical stage (*P*=0.0018), MATV (*P*=0.0053) and heterogeneity (E^1, *P*=0.0093) as independent prognostic factors. MATV allowed for the highest differentiation (HR=2.8), whereas stage and heterogeneity led to lower HRs (2.3 and 2.1 respectively). Adding stage to MATV did not increase stratifying power (HR=2.84), whereas adding heterogeneity increased HR to 3.55, with 49months

vs. 9.1months median OS. The addition of the three led to the highest HR of 3.81 (supplemental table 6, Fig. 7B). Survival curves were also evaluated according to a 3-valued score (MATV<35cm^3 AND E^1<7.35, MATV>35cm^3 OR E^1<7.35, and MATV>35cm^3 AND E^1>7.35), leading to survival curves with median OS of 49, 20 and 9 months ($p<0.0001$, HRs of 1.8 and 4.3) (supplemental Fig. 5B, Fig. 7C), highlighting the higher complementary value of heterogeneity and volume in this cohort.

DISCUSSION

There is an increasing interest in the use of PET image textural features for quantification of intra-tumor heterogeneity (4,6). Few studies investigated the relationships between tumor volume and TFs (15,19,23).

Most studies using textural features considered volumes>3-5cm^3, assuming that PET could not characterize heterogeneity on smaller volumes due to its limited spatial resolution. A recent theoretical analysis suggested that volumes >45cm^3 should be considered to avoid volume related confounding effects (14). However, this analysis considered a single parameter (entropy), calculated on 2D co-occurrence matrices over 2 spatial directions followed by averaging, and using a quantization value >150 (14).

In this work, we addressed the question of the minimum functional volume that could be considered and investigated the potential complementary prognostic value between volume and heterogeneity. We investigated the influence of the quantization pre-processing and of the textural features calculation methodology, investigating more thoroughly the relationships between heterogeneity and functional volume in a substantially larger tumors' database, covering large tumor volume ranges and different cancer types. Considering a patient cohort with variable cancer types and complementary volume ranges allowed providing a better picture of the relationship between the measured feature and its corresponding volume.

Our results partly confirm those of recent studies. Indeed, several textural features were found to be highly correlated with the volume from which they were calculated. In addition, high correlations were found between most of these parameters. We observed differences in the heterogeneity-volume distributions as a function of tumor type, although these differences can be explained by differences in volume ranges for each cancer patient cohort, rather than the histology or heterogeneity specific to each tumor type. For instance, the distribution of textural features with respect to volume in the cervix tumors was significantly different than the others because it included only large tumors (>20cm^3) relative to the other patient cohorts considered that also included smaller tumor volumes.

Our results emphasize that (i) the relationship and the level of correlation is not the same for all features, (ii) the level of correlation tends to decrease substantially when considering larger tumor volumes, (iii) the calculation method and the quantization step both have an impact on the relationship and level of correlation between volume and the feature, and (iv) volume and heterogeneity can have complementary prognostic value: in

the NSCLC cohort, heterogeneity and volume were identified as independent prognostic factors and hazard ratios were shown to increase from <2.9 to >3.8 when adding these risk factors for patient stratification. Although multivariate analysis could not be performed in the esophageal cohort, combining heterogeneity with volume led to an increased although non-statistically significant different stratification. NSCLC tumor volumes were much larger than esophageal ones (58±77cm^3, median 34, range 3-415 *vs.* 25±27cm^3, median 15, range 3-140), which is likely why heterogeneity and volume had higher complementary prognostic values in NSCLC.

When considered together, these results point to the potential added prognostic value of tumor heterogeneity quantified with textural feature, although regarding Kaplan-Meier curves, cut-off values found with ROC analysis are probably over-fitted and specific to the data, and thus validation in independent cohorts will be required.

The quasi-linear relationship between entropy (either E^1 or E^{13}) and volume in the range 3-10cm^3 (*rs*>0.9, Fig. 4) suggests that entropy for volumes<10cm^3 cannot provide complementary information. However, when choosing appropriate quantization and co-occurrence matrix calculation, the correlation between volume and entropy for volumes in the range 10-50cm^3 was much lower. The minimal volume to consider may therefore be closer to 10cm^3 than 45cm^3, a value previously suggested based on calculations carried out after a quantization into >150 grey-levels and using several co-occurrence matrices followed by averaging (14). In addition, for other TFs not considered in that last study, the correlation was weaker, even for volumes between 3 and 10cm^3. It was only 0.4 and 0.6 for D^{13} and D^1 respectively, and 0.3 and 0.2 for ZP and HILAE. Therefore, instead of excluding patients with volumes below the proposed 10cm^3 threshold from such analysis, we rather recommend to report the correlation between volume and heterogeneity and highlight their complementary value as tumor volumes increase. Indeed, the correlation with volume decreased substantially for all textural features when considering larger volumes. Larger tumors are known to exhibit higher hypoxia, necrosis or anatomical and physiological complexity at the microscopic and macroscopic scales, which logically translates to higher FDG uptake spatial distribution complexity and consequently associated heterogeneity quantification. A significant correlation between quantified heterogeneity (via any method) and its corresponding volume is therefore to be expected in a standard cohort of patients covering a large range of tumor sizes. Prospective clinical studies with optimized pre-defined image acquisition settings and heterogeneity analysis protocols, as well as experimental studies demonstrating that tumor properties on a scale comparable with the PET resolution can actually lead to identifiable image textural features could benefit from the results of the present study and should now be conducted.

One important aspect of this study is that the 555 PET images in our database came from six cohorts acquired in different centers, albeit with certain homogeneity in acquisition parameters (same scanner, reconstruction algorithm, voxel size), except for one. This variability was handled by restricting the analysis to robust features only, as to minimize the associated impact. Indeed, excluding the H&N cohort with a different imaging protocol did not change the results. On the other hand, the inclusion of a small number of heterogeneity quantification metrics may be considered as a limitation.

However, the four textural features included have been previously shown to be the most reproducible and robust amongst those shown to have a predictive and prognostic value in different cancer types. We also restricted the prognosis analysis to the esophageal and NSCLC cohorts because clinical and survival data for the other cohorts were not available for enough patients to allow for multivariate analysis.

CONCLUSIONS

Most of textural features considered to quantify intra-tumor heterogeneity were found to be significantly correlated with tumor volume. However, our detailed analysis also suggests that heterogeneity quantification through textural features may potentially provide valuable clinical complementary information in addition to functional volume, especially for tumors above $10cm^3$, with increasing complementary prognostic value for larger volumes. In 112 esophageal patients, heterogeneity was found to have some prognostic value that was not significantly improved when combined with volume, whereas in a cohort of 101 NSCLC patients, heterogeneity, volume and stage were independent prognostic factors that allowed increased stratification of patients when combined.

References

1. Krause BJ, Schwarzenbock S, Souvatzoglou M. FDG PET and PET/CT. *Recent Results Cancer Res*. 2013;187:351-69.

2. 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.

3. Herrmann K, Benz MR, Krause BJ, Pomykala KL, Buck AK, Czernin J. (18)F-FDG-PET/CT in evaluating response to therapy in solid tumors: where we are and where we can go. *Q J Nucl Med Mol Imaging*. 2011;55:620-32.

4. Visvikis D, Hatt M, Tixier F, Cheze Le Rest C. The age of reason for FDG PET image-derived indices. *Eur J Nucl Med Mol Imaging*. 2012;39:1670-2.

5. Van de Wiele C, Kruse V, Smeets P, Sathekge M, Maes A. Predictive and prognostic value of metabolic tumour volume and total lesion glycolysis in solid tumours. *Eur J Nucl Med Mol Imaging*. 2013;40:290-301.

6. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441-6.

7. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging*. 2013;40:133-40.

8. Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging*. 2012;3:573-89.

9. Miller TR, Pinkus E, Dehdashti F, Grigsby PW. Improved prognostic value of 18F-FDG PET using a simple visual analysis of tumor characteristics in patients with cervical cancer. *J Nucl Med*. 2003;44:192-7.

10. 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.

11. Van Velden FH, 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;38:1636-47.

12. Miwa K, Inubushi M, Wagatsuma K, et al. FDG uptake heterogeneity evaluated by fractal analysis improves the differential diagnosis of pulmonary nodules. *Eur J Radiol*. 2014;83:715-719.

13. Tixier F, Le Rest CC, Hatt M, 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;52:369-78.

14. 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.

15. Orlhac F, Soussan M, Maisonobe J-A, Garcia CA, Vanderlinden B, Buvat I.

Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med*. 2014;55:414-422.

16. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. *IEEE Trans Med Imaging*. 2009;28:881-93.

17. Hatt M, Cheze le Rest C, Descourt P, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys*. 2010;77:301-8.

18. Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D. Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation. *J Nucl Med*. 2011;52:1690-7.

19. 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.

20. 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;49:1012-6.

21. Tixier F, Hatt M, Le Rest CC, 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.

22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289-300.

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

Table 1. Prognostic value for OS in the NSCLC and esophageal cohorts assessed using Cox regression for univariate and multivariate analysis

Parameters		Esophageal p-value Univariate	NSCLC p-value Univariate	Multivariate
Clinical	Surgery	0.3654	0.0084	NRM
	Age	0.1861	0.0441¥	-
	Sex	0.1795	0.0227	NRM
	Smoker	0.4511	0.8997	-
	Histology	0.2154	0.3041	-
	Stage	0.0391¥	0.0003	0.0018
Volume and SUV	MATV	0.0315¥	0.0008	0.0053
	SUV_{max}	0.2781	0.0599	-
	SUV_{mean}	0.6008	0.0256	NRM
Heterogeneity (TF)	D^{13}	0.0405¥	0.0046	NRM
	D^{1}	0.0016	0.0027	NRM
	E^{13}	0.1087	0.0002	NRM
	E^{1}	0.3922	0.0287	0.0093
	HILAE	0.2596	0.0132	NRM
	ZP	0.4391	0.0005	NRM

NRM: not retained in the model

¥: Not significant after correction for multiple testing

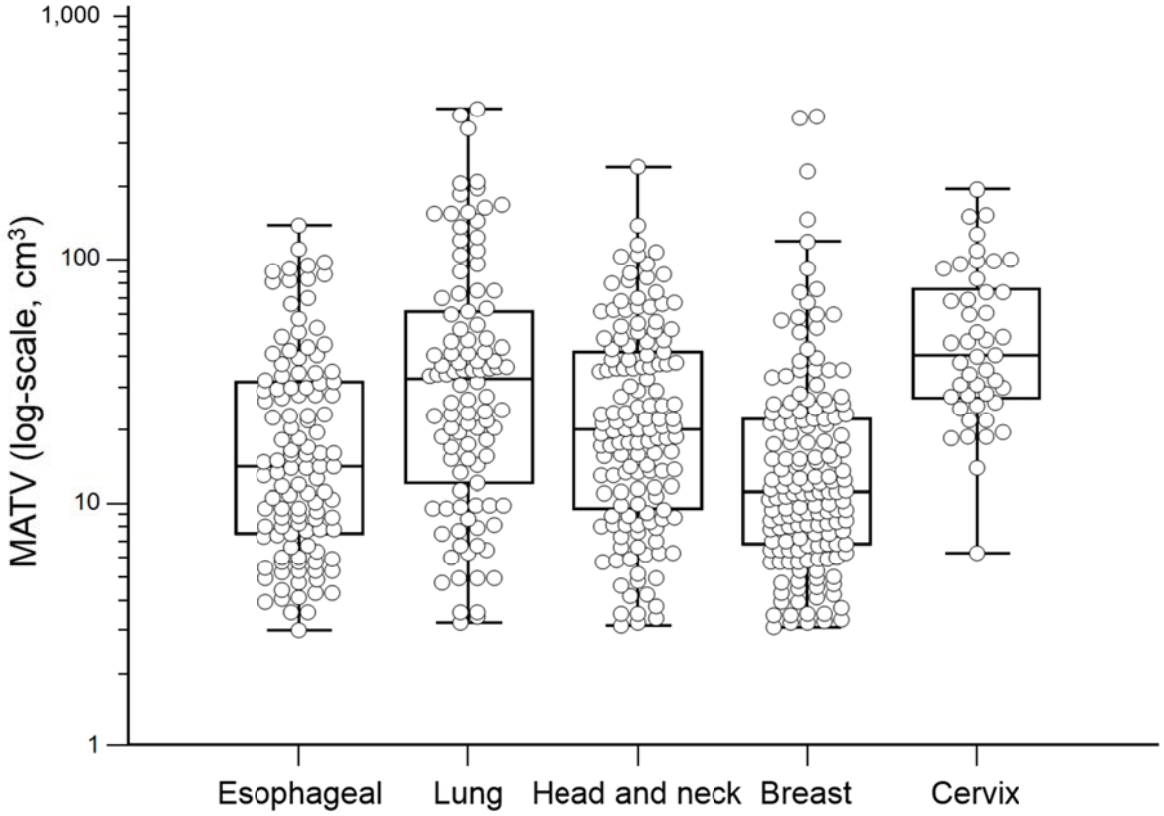
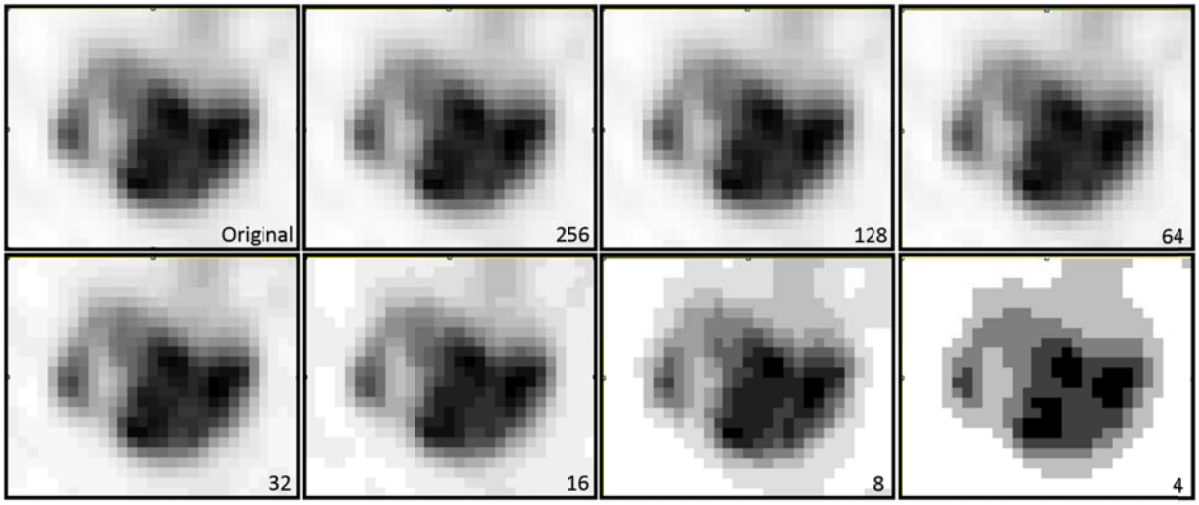


Figure 1. MATV distribution in each cancer site.

	D^{13}	D^{1}	E^{13}	E^{1}	HILAE	ZP	MATV	SUV_{max}	SUV_{mean}	SUV_{COV}
D^{13}	1									
D^{1}	0,92 <0,0001	1								
E^{13}	-0,76 <0,0001	-0,81 <0,0001	1							
E^{1}	-0,40 <0,0001	-0,18 <0,0001	0,82 <0,0001	1						
HILAE	0,05 0,4	0,12 0,001	0,23 <0,0001	0,44 <0,0001	1					
ZP	0,72 <0,0001	0,84 <0,0001	-0,64 <0,0001	-0,06 0,2	-0,10 0,02	1				
MATV	-0,80 <0,0001	-0,82 <0,0001	0,96 <0,0001	0,56 <0,0001	0,17 <0,0001	-0,68 <0,0001	1			
SUV_{max}	-0,29 <0,0001	-0,35 <0,0001	0,38 <0,0001	0,31 <0,0001	0,28 <0,0001	-0,32 <0,0001	0,42 <0,0001	1		
SUV_{mean}	-0,11 0,027	-0,13 0,0033	0,38 <0,0001	0,47 <0,0001	0,40 <0,0001	-0,13 0,0086	0,31 <0,0001	0,89 <0,0001	1	
SUV_{COV}	0,16 0,0006	-0,19 0,0001	-0,05 0,26	-0,19 0,0001	0,07 0,15	-0,30 <0,0001	0,08 0,084	0,30 <0,0001	0,07 0,17	1

Figure 2. Illustration of quantization (A) and impact on the correlation between TF and MATV (B).

A



B

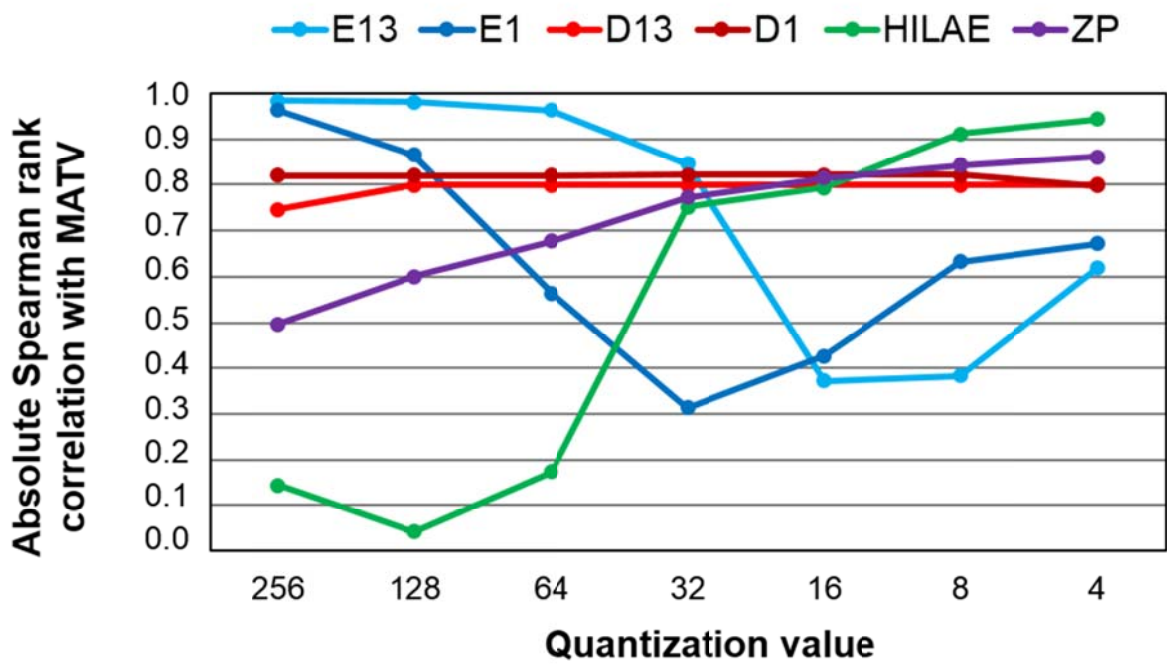
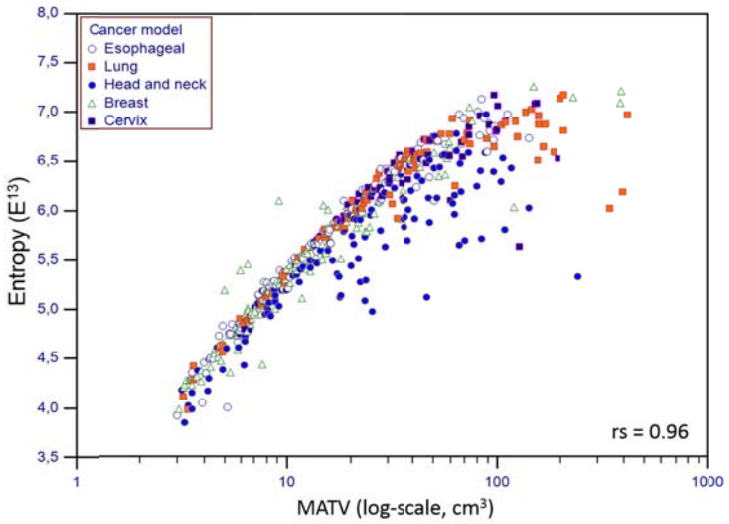
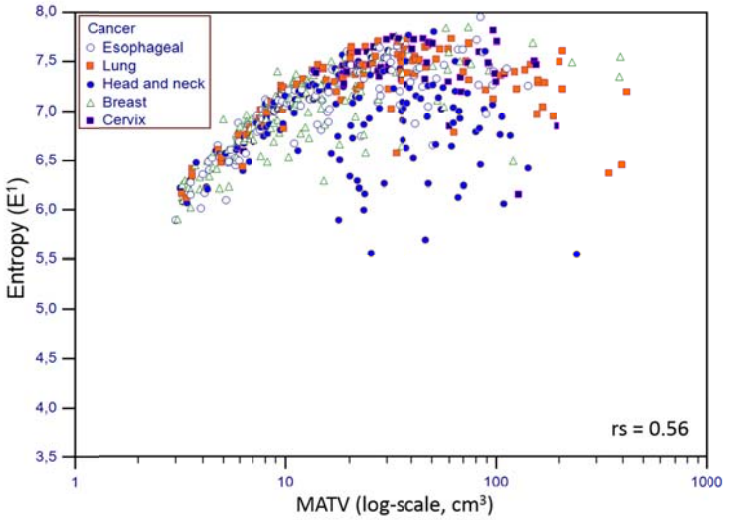


Figure 3. Spearman rank correlations between parameters (555 tumors). Red: [0.8,1.0] orange: [0.6,0.8[, green: [0.3,0.6[, violet: [0.1,0.3[and grey: [0.0,0.1].

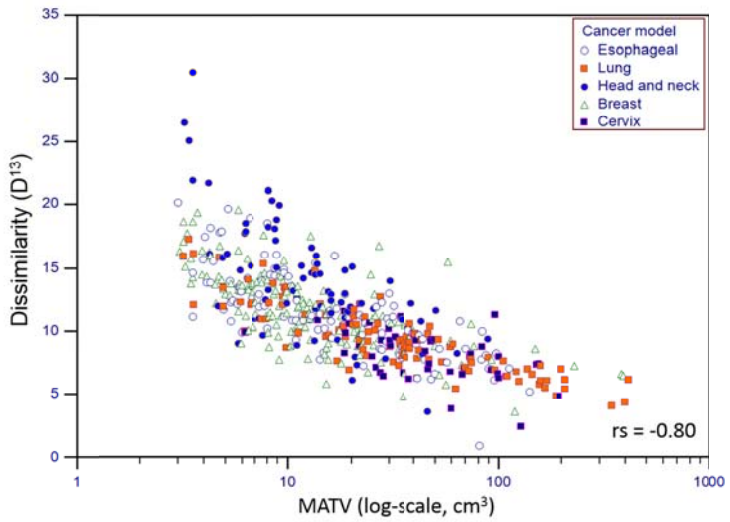
A


Cancer model
Esophageal
Lung
Head and neck
Breast
Cervix
Entropy (E13)
MATV (log-scale, cm3)
rs = 0.96

B


Cancer
Esophageal
Lung
Head and neck
Breast
Cervix
Entropy (E1)
MATV (log-scale, cm3)
rs = 0.56

C



D

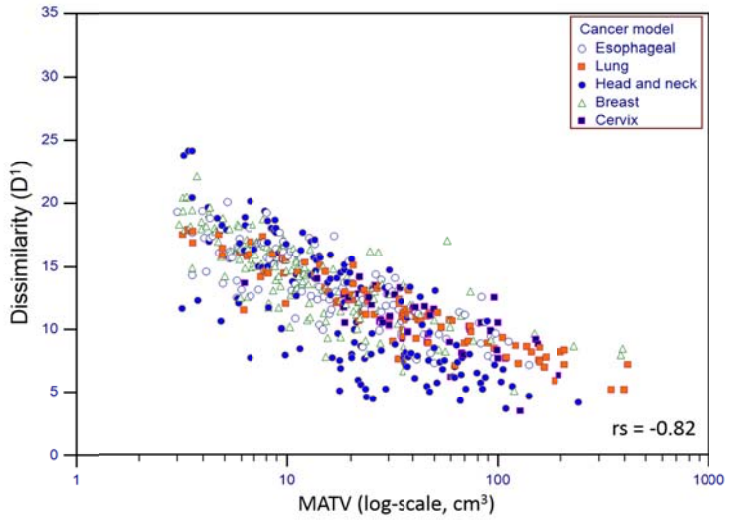
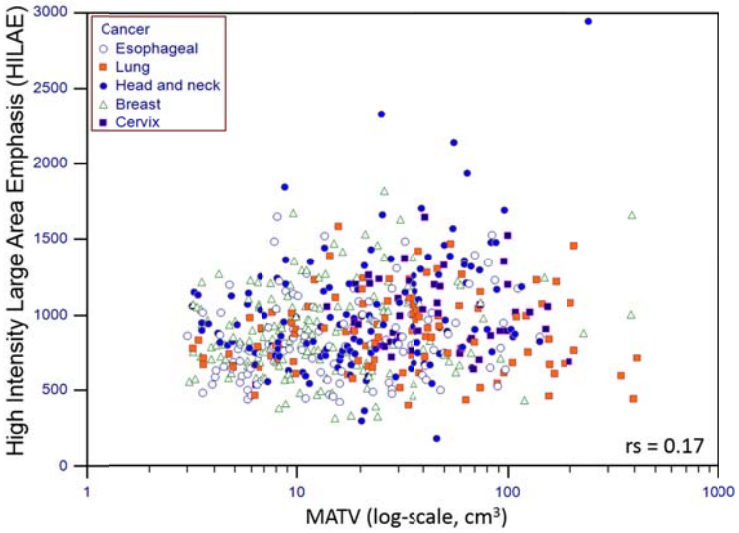


Figure 4. Distributions of E^1 (A), E^{13} (B), D^1 (C) and D^{13} (D) (quantization=64) with respect to MATV.

A



B

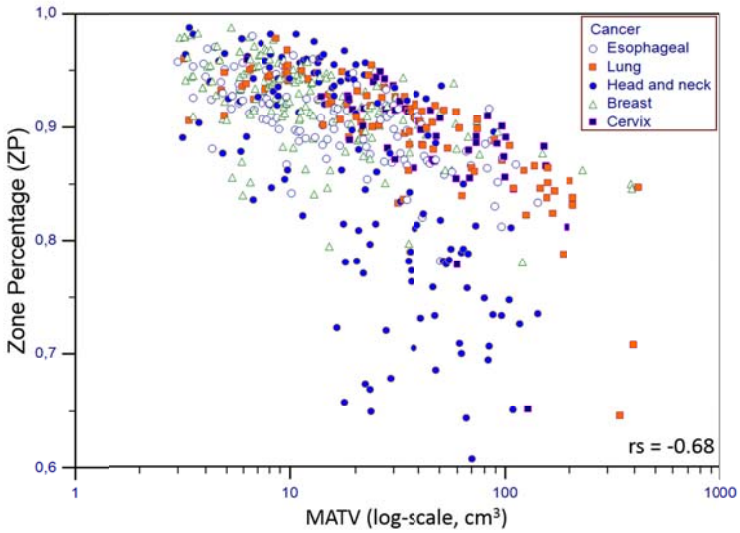


Figure 5. Distributions of HILAE (A) and ZP (B) (quantization=64) with respect to MATV.

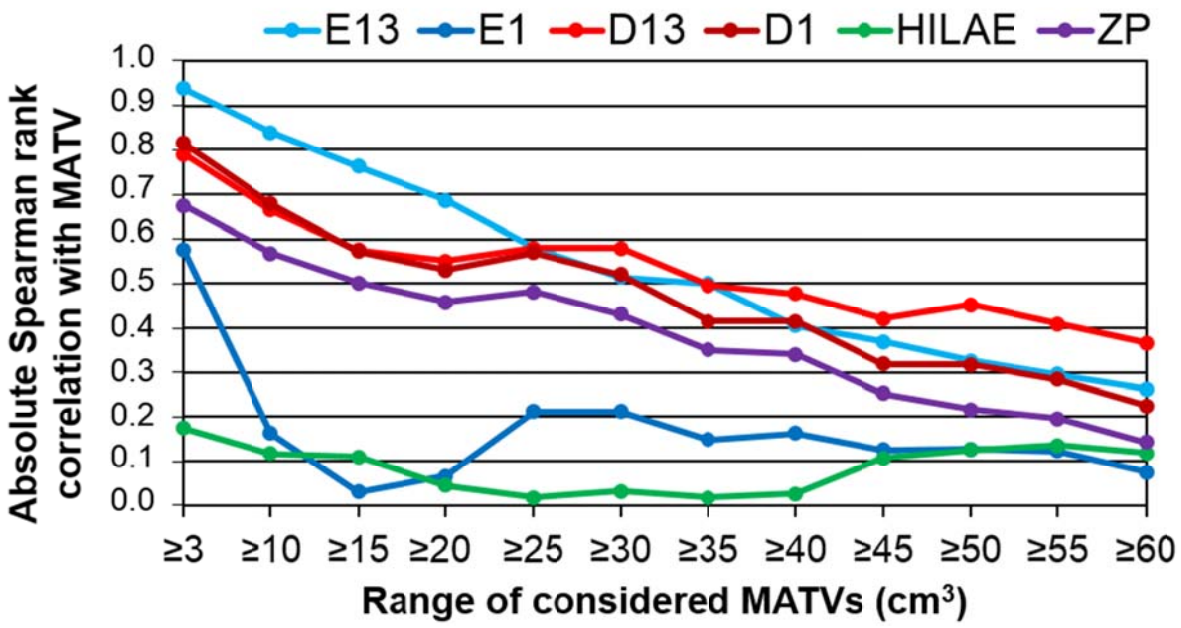
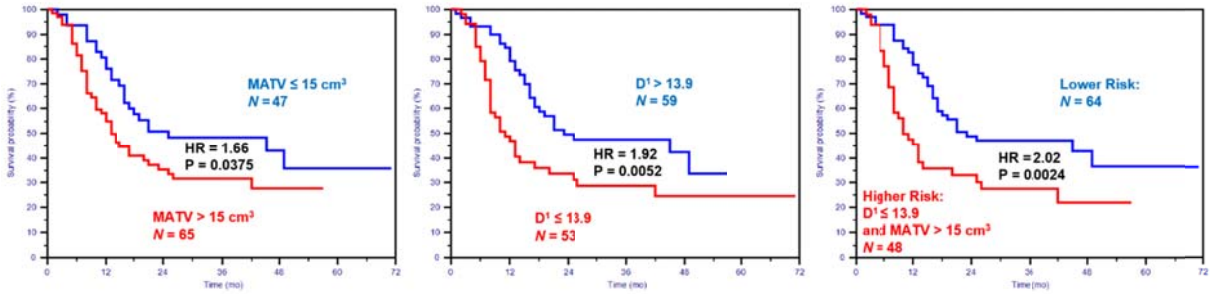
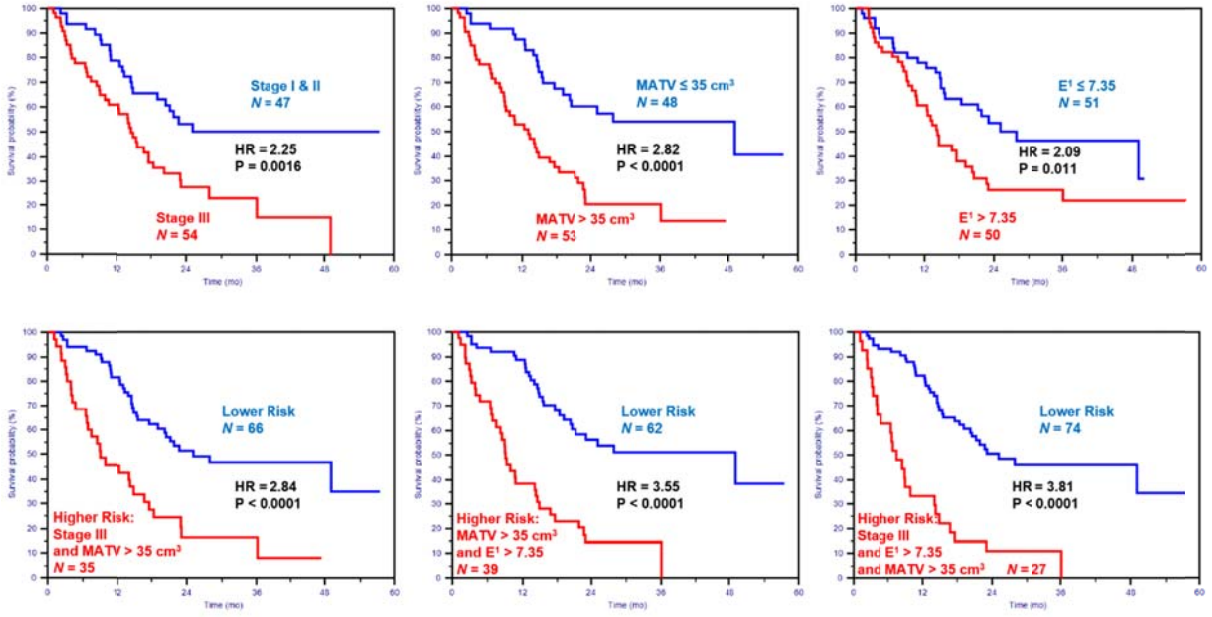


Figure 6. Absolute Spearman rank correlation with MATV for each TF (quantization=64), considering different ranges of increasing MATVs.

A



B



C

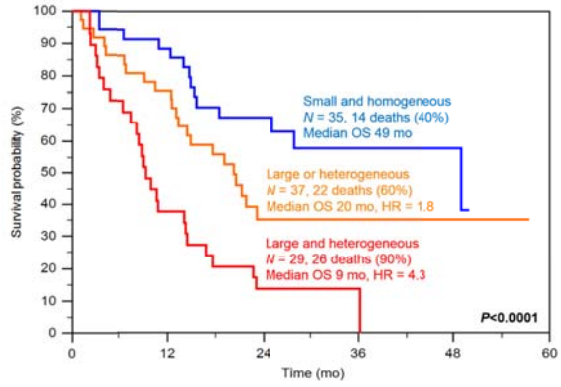


Figure 7. Kaplan-Meier curves using volume and heterogeneity for (A) 112 esophageal patients, (B-C) 101 NSCLC patients.