

# Spectral Clustering Predicts Tumor Tissue Heterogeneity Using Dynamic $^{18}\text{F}$ -FDG PET: A Complement to the Standard Compartmental Modeling Approach

Prateek Katiyar<sup>1,2</sup>, Mathew R. Divine<sup>1</sup>, Ursula Kohlhofer<sup>3</sup>, Leticia Quintanilla-Martinez<sup>3</sup>, Bernhard Schölkopf<sup>2</sup>, Bernd J. Pichler<sup>1</sup>, and Jonathan A. Disselhorst<sup>1</sup>

<sup>1</sup>Department of Preclinical Imaging and Radiopharmacy, Werner Siemens Imaging Center, Eberhard Karls University Tuebingen, Tuebingen, Germany; <sup>2</sup>Max Planck Institute for Intelligent Systems, Tuebingen, Germany; and <sup>3</sup>Institute of Pathology and Neuropathology, Eberhard Karls University Tuebingen and Comprehensive Cancer Center, University Hospital Tuebingen, Tuebingen, Germany

In this study, we described and validated an unsupervised segmentation algorithm for the assessment of tumor heterogeneity using dynamic  $^{18}\text{F}$ -FDG PET. The aim of our study was to objectively evaluate the proposed method and make comparisons with compartmental modeling parametric maps and SUV segmentations using simulations of clinically relevant tumor tissue types. **Methods:** An irreversible 2-tissue-compartmental model was implemented to simulate clinical and preclinical  $^{18}\text{F}$ -FDG PET time-activity curves using population-based arterial input functions (80 clinical and 12 preclinical) and the kinetic parameter values of 3 tumor tissue types. The simulated time-activity curves were corrupted with different levels of noise and used to calculate the tissue-type misclassification errors of spectral clustering (SC), parametric maps, and SUV segmentation. The utility of the inverse noise variance- and Laplacian score-derived frame weighting schemes before SC was also investigated. Finally, the SC scheme with the best results was tested on a dynamic  $^{18}\text{F}$ -FDG measurement of a mouse bearing subcutaneous colon cancer and validated using histology. **Results:** In the preclinical setup, the inverse noise variance-weighted SC exhibited the lowest misclassification errors (8.09%–28.53%) at all noise levels in contrast to the Laplacian score-weighted SC (16.12%–31.23%), unweighted SC (25.73%–40.03%), parametric maps (28.02%–61.45%), and SUV (45.49%–45.63%) segmentation. The classification efficacy of both weighted SC schemes in the clinical case was comparable to the unweighted SC. When applied to the dynamic  $^{18}\text{F}$ -FDG measurement of colon cancer, the proposed algorithm accurately identified densely vascularized regions from the rest of the tumor. In addition, the segmented regions and clusterwise average time-activity curves showed excellent correlation with the tumor histology. **Conclusion:** The promising results of SC mark its position as a robust tool for quantification of tumor heterogeneity using dynamic PET studies. Because SC tumor segmentation is based on the intrinsic structure of the underlying data, it can be easily applied to other cancer types as well.

**Key Words:** spectral clustering; tumor heterogeneity; compartmental modeling;  $^{18}\text{F}$ -FDG PET; SUV

**J Nucl Med 2017; 58:651–657**

DOI: 10.2967/jnumed.116.181370

Tumors exhibit widespread genetic and phenotypic heterogeneity. The local tissue variability is known to mediate drug resistance and influence therapeutic efficacy (1). The magnitude of intratumor diversity is also linked with tumor aggressiveness and has been shown to predict cancer mortality (2). The robust characterization of the tumor heterogeneity is an urgent requirement for not only precision medicine, but also for preclinical and pharmaceutical research (3).

The sensitivity and quantitative ability of PET make it a promising prognostic tool for cancer diagnosis and in vivo monitoring of therapy response. Accumulation of  $^{18}\text{F}$ -FDG in cancerous lesions is widely associated with tumor grade and prognosis (4,5). The most common clinical assessment of  $^{18}\text{F}$ -FDG is based on visual inspection and basic quantification of the SUV. Although the SUV as a metric is practical and easy to measure, it is vulnerable to numerous sources of variability (6). Whereas static measures lack the ability to distinguish between nonphosphorylated and phosphorylated  $^{18}\text{F}$ -FDG, kinetic methods measure the complete aspects of the tracer distribution, providing vital information about glycolysis and blood flow. Kinetic modeling can play an especially essential role when evaluating the drug response of cancer patients with low pretherapy  $^{18}\text{F}$ -FDG uptake, which results in poor sensitivity of the SUV and other static measures (7,8).

Despite the quantification benefits over static measures, kinetic methods such as compartmental modeling and graphical analysis have not been widely adopted, partly due to their reliance on the acquisition of time-activity curves with low noise and a precise measurement of the arterial input function (AIF). Moreover, to improve signal-to-noise ratios, a common practice in dynamic PET studies is to perform region averaging (9) before compartmental modeling. Because compartmental modeling assumes the region of interest to be functionally homogeneous (10), user-defined delineations might lead to incorrect estimation of kinetic parameters in regions with tissue variability.

Received Jul. 21, 2016; revision accepted Oct. 19, 2016.

For correspondence or reprints contact: Jonathan A. Disselhorst, University of Tuebingen, Department for Preclinical Imaging and Radiopharmacy, Werner Siemens Imaging Center, Roentgenweg 13, D-72076 Tuebingen, Germany.

E-mail: Jonathan.Disselhorst@med.uni-tuebingen.de

Published online Nov. 3, 2016.

COPYRIGHT © 2017 by the Society of Nuclear Medicine and Molecular Imaging.

A voxel-level analysis is essential to create a holistic profile of the spatial and temporal heterogeneity of cancerous lesions (11). Over the past decades, several segmentation methods have been proposed for the region-wise analysis of PET images (12). Recently, one investigation has applied spectral clustering (SC) on dynamic PET data for brain image segmentation (13). The study by Mouysset et al., however, lacked a histologic validation. In the present study, we aimed to examine the suitability of SC in the segmentation of the tumor microenvironment. Through comprehensive simulations, we present an objective evaluation of SC and compare its robustness with the parametric maps and SUV segmentation. We also tested the proposed methodology in vivo on a mouse model of subcutaneous colon cancer with a histologic validation.

## MATERIALS AND METHODS

The widely accepted pharmacokinetic modeling tool COMKAT (14) was used to simulate  $^{18}\text{F}$ -FDG PET time-activity curves. The complete details of the implemented compartmental model, preclinical experiments, and histology are provided in the supplemental materials (available at <http://jnm.snmjournals.org>).

### Clinical and Preclinical Tissue Class Simulation

To simulate clinically relevant and comparable scenarios, the kinetic parameter values of different tissue classes were derived from Sugwara et al. (15). The authors studied 21 patients with primary germ cell tumors using  $^{18}\text{F}$ -FDG PET and reported the kinetic parameter values of 3 different tumor tissue classes, namely the viable tissue, mature teratoma, and necrosis. Because tumor tissue types were confirmed by histologic findings, we extended the average kinetic parameter values of each tissue type as corresponding class representative. Likewise, the clinical AIF was selected from a population-based AIF model (16). The study identified the parameters of the mathematic equations by fitting a 3-compartment blood-pool model (17) on the arterial blood samples taken from 80 different patients. We contacted the authors to obtain the complete dataset because the published details were insufficient for simulations.

To extrapolate the clinical scenario into the preclinical setting, twelve 60-min dynamic  $^{18}\text{F}$ -FDG PET scans (4 mice  $\times$  3 scans) were acquired from 8-wk-old Naval Medical Research Institute *nu/nu* mice bearing subcutaneous Colo-205 tumors. The AIFs of all the measurements were approximated using a minimal blood sampling scheme (18). A 2-tissue-compartmental model was fitted to the mean time-activity curve of each tumor, for each measurement. The obtained kinetic parameters from all 12 PET scans provided realistic values of kinetic parameters observable in preclinical studies, which formed

the basis to simulate the preclinical tumor tissue classes. First, the averages of these kinetic parameters were used to simulate the viable tissue. Afterward, the parameters of teratoma and necrotic tissues were obtained by scaling the viable parameters to achieve the same parameter ratios (between different tissue classes) as in the clinical settings. The SDs were chosen to match the mean-to-SD ratio of the respective clinical tissue type. All the animal experiments were performed in accordance with the German Animal Welfare Act, and local authorities approved all experimental protocols.

A total of 2,000 time-activity curves were sampled from a truncated Gaussian distribution (Table 1) for each tumor tissue class. The distributions were truncated to avoid sampling time-activity curves with an unrealistic shape. The framing protocol was kept the same for both clinical and preclinical simulations:  $\{30 \times 2 \text{ s}, 8 \times 5 \text{ s}, 8 \times 10 \text{ s}, 6 \times 1 \text{ min}, 5 \times 2 \text{ min}, 5 \times 10 \text{ min}\}$ . For simplicity, throughout this article, we refer to the simulated tumor tissue classes (viable, teratoma, and necrosis) as class 1, class 2, and class 3, respectively.

### Noisy Time-Activity Curves

The noisy realizations of the simulated time-activity curves were obtained by estimating the noise SD for each time frame and distributing it log-normally to the noise-free curve (9,19). The noise SD for each frame  $i$  can be computed as follows:

$$\text{SD} = \beta \sqrt{\frac{e^{-\lambda t_i} \times \text{ROI}(t_i)}{\Delta t_i}},$$

where  $\text{ROI}(t_i)$  is the decay-corrected activity concentration of the region of interest,  $e^{-\lambda t_i}$  is the decay uncorrection factor,  $\lambda$  refers to the ratio  $\frac{\ln(2)}{\text{half-life}}$ ,  $\Delta t_i$  is the frame duration, and  $\beta$  is a scale factor to limit the amount of noise within practical conditions. An illustrative example of noisy time-activity curves can be found in Supplemental Figure 1.

### SC

SC (20) uses the eigenstructure of the affinity matrix and one of the classic clustering methods (e.g.,  $k$ -means, fuzzy c-means, Gaussian mixture modeling) (21) to partition voxels into disjoint clusters. The affinity matrix  $W_{ij}$  of the dynamic PET data was computed as follows:

$$W_{ij} = \begin{cases} e^{-\|x_i - x_j\|^2 / 2\sigma^2} & \text{if } i \neq j \\ 0 & \text{otherwise.} \end{cases}$$

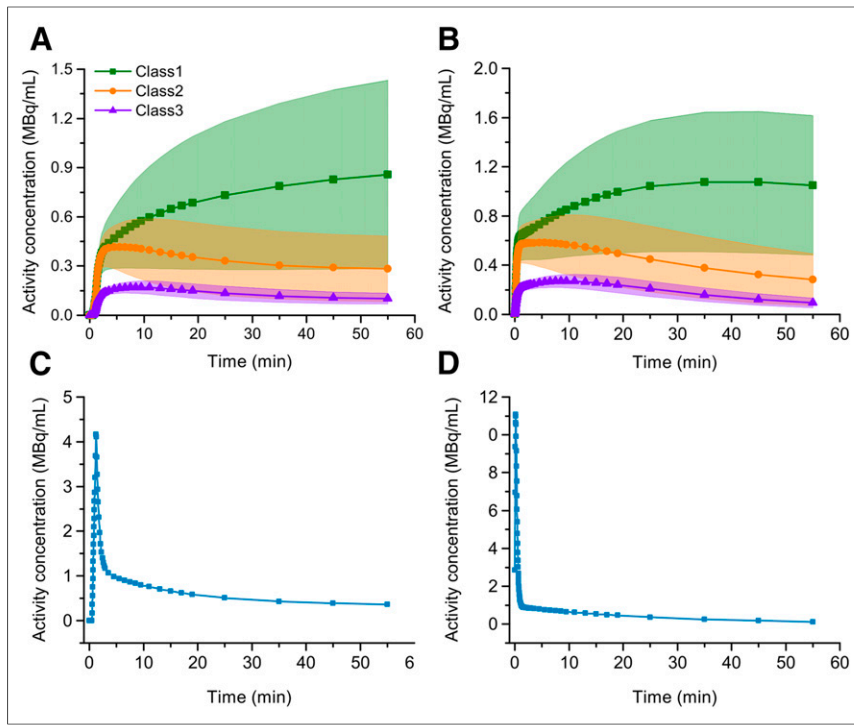
Here,  $\|x_i - x_j\|$  is the Euclidian distance between the time-activity curves  $i$  and  $j$ , and  $\sigma$  is the scale parameter of the Gaussian kernel.

**TABLE 1**

Summary of Kinetic Parameters and Corresponding Truncation Limits Used for the Simulation of Preclinical and Clinical Tumor Tissue Classes

Kinetic parameter	Class 1		Class 2		Class 3		Truncation limits
	Preclinical	Clinical	Preclinical	Clinical	Preclinical	Clinical	
$K_1$	$0.138 \pm 0.043$	$0.110 \pm 0.034$	$0.123 \pm 0.033$	$0.114 \pm 0.026$	$0.045 \pm 0.006$	$0.036 \pm 0.005$	0.01–1.0
$k_2$	$0.116 \pm 0.136$	$0.195 \pm 0.228$	$0.180 \pm 0.069$	$0.301 \pm 0.116$	$0.105 \pm 0.025$	$0.176 \pm 0.042$	0.01–1.0
$k_3$	$0.085 \pm 0.056$	$0.073 \pm 0.048$	$0.014 \pm 0.008$	$0.012 \pm 0.007$	$0.005 \pm 0.002$	$0.004 \pm 0.001$	0.001–1.0

Data are mean  $\pm$  SD.



**FIGURE 1.** Classwise simulated time-activity curves and corresponding AIF for clinical (A and C) and preclinical (B and D) scenarios. Kinetic parameters for each class were sampled from truncated Gaussian distributions. Shaded regions depict the distribution of time-activity curves up to unit SD of the respective tumor tissue type.

Subsequently, the affinity matrix  $W_{ij}$  was used to compute the normalized graph Laplacian using the following expression:

$$L = D_{inv} \times W_{ij} \times D_{inv},$$

where  $D_{inv} = D^{-0.5}$ , and  $D$  is the diagonal matrix with  $d_i = \sum_{j=1}^n W_{ij}$  as the diagonal vector. To perform unsupervised clustering, the set of first  $k$  eigenvectors (corresponding to  $k$  largest eigenvalues (20)) of the normalized graph Laplacian was fitted using Gaussian mixture modeling. Throughout the study, we used the first 6 eigenvectors ( $k = 6$ ) of the normalized Laplacian matrix and set  $\sigma$  equal to 40 and 55 for segmentation of preclinical and clinical time-activity curves, respectively. The scale was chosen experimentally, based on the misclassification error of the method on the noise-free time-activity curves. The same scale was used for segmentation of the preclinical example, but we could determine that segmentation was robust to the choice of  $\sigma$ .

## PET Frame Weighting

In this study, the performance of 2 different weighting schemes for SC was investigated. In the first case, weights for each frame were set equal to the inverse of the noise variance (INV) of the respective frame, thus, dependent on frame length and total amount of activity in that specific frame. In the second scheme, weights were derived from the Laplacian scoring (LS) algorithm (22). In the end, the weighted SC scheme with the best results (for pre-clinical simulations) was applied on the experimental data.

## Clustering Comparisons

The clustering potential of SC was tested on the simulated data over varying levels of noise. The proposed methodology was also compared with SUV and parametric map segmentation. In the former case, the average of the last 2 frames of the simulated dataset was clustered using  $k$ -means, and in the latter case the estimated kinetic parameters ( $K_1$ ,  $k_2$ ,  $k_3$ , and  $K_i$ ) were segmented into 3 tissue classes using  $k$ -means and SC.

## Evaluation Metrics

The percentage kinetic parameter estimation error ( $\epsilon$ ) was defined as:

$$\epsilon(\%) = \begin{cases} \left( -\frac{K_{p_{true}}}{K_{p_{estimated}}} + 1 \right) \times 100 & \text{true} \leq \text{estimated} \\ \left( \frac{K_{p_{estimated}}}{K_{p_{true}}} - 1 \right) \times 100 & \text{true} > \text{estimated}, \end{cases}$$

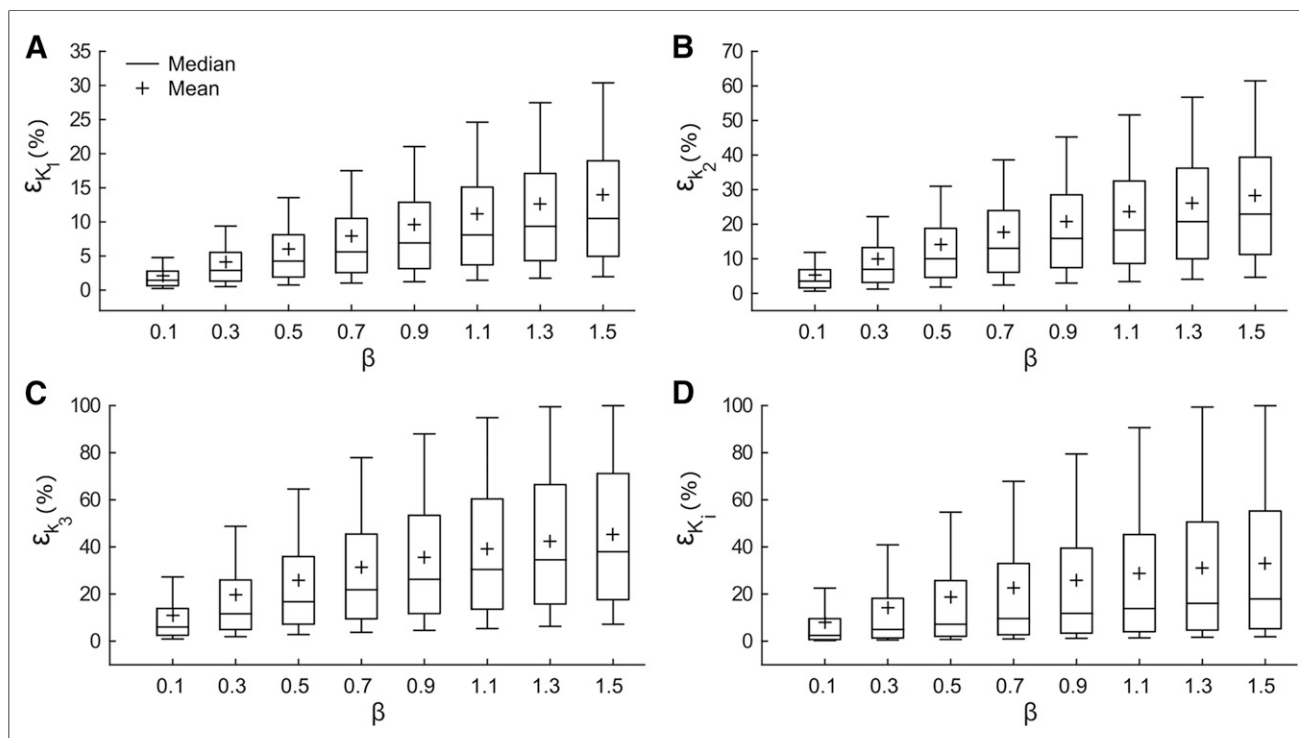
where  $K_{p_{estimated}}$  is the estimated and  $K_{p_{true}}$  is the true value of the compartmental modeling rate constant. The misclassification error was defined as follows:

$$\frac{\sum_{i=1}^{N_{TAC}} \sum_{j=1}^K I(O_{ij}, T_{ij})}{N_{TAC} \times K},$$

where  $O_{ij}$  is the output and  $T_{ij}$  is the true label of the time-activity curve  $i$  from class  $j$ ,  $N_{TAC}$  represents the total number of time-activity

**TABLE 2**  
Kinetic Parameter Estimation Errors Obtained After Fitting the Preclinical and Clinical Noise Free Time-Activity Curves Using Respective AIFs

Kinetic parameter	Kinetic parameter estimation error ( $\epsilon$ ) %			
	Preclinical		Clinical	
	Median	Interquartile range	Median	Interquartile range
$K_1$	-0.003	-0.020 to 0.012	0.012	-0.073 to 0.116
$k_2$	-0.040	-0.180 to 0.071	0.050	-0.140 to 0.287
$k_3$	-0.121	-0.649 to 0.213	0.049	-0.260 to 0.387
$K_i$	-0.033	-0.385 to 0.091	0.000	-0.143 to 0.165



**FIGURE 2.** Absolute  $\epsilon$  for preclinical simulations with an increase in the amount of noise ( $\beta$ ) for  $K_1$  (A),  $k_2$  (B),  $k_3$  (C), and  $K_i$  (D). The boxes depict the interquartile range, and whiskers represent the 10th and 90th percentiles of the data.

curves in each class, and  $K$  is equal to the number of tumor tissue types (i.e., 3). The indicator variable  $I$  is given as:

$$I = \begin{cases} 1 & \text{if } O_{ij} \neq T_{ij} \\ 0 & \text{otherwise.} \end{cases}$$

## RESULTS

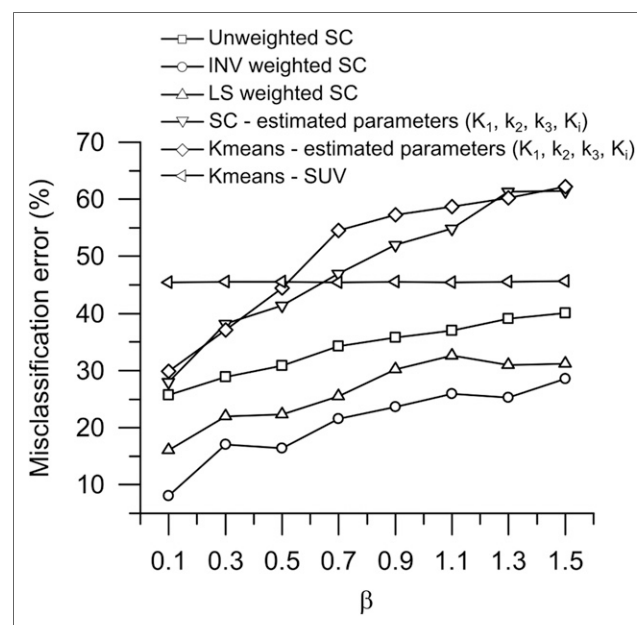
Examples of simulated time–activity curves of class 1, class 2, class 3 and corresponding AIF for clinical and preclinical scenarios are shown in Figure 1. To assess the influence of selected framing and the bias introduced by COMKAT, noise-free curves were fitted using their respective AIF. The interquartile range and median  $\epsilon$  for  $K_1$ ,  $k_2$ ,  $k_3$ , and  $K_i$  for preclinical and clinical simulations are reported in Table 2.

### Noise Evaluation

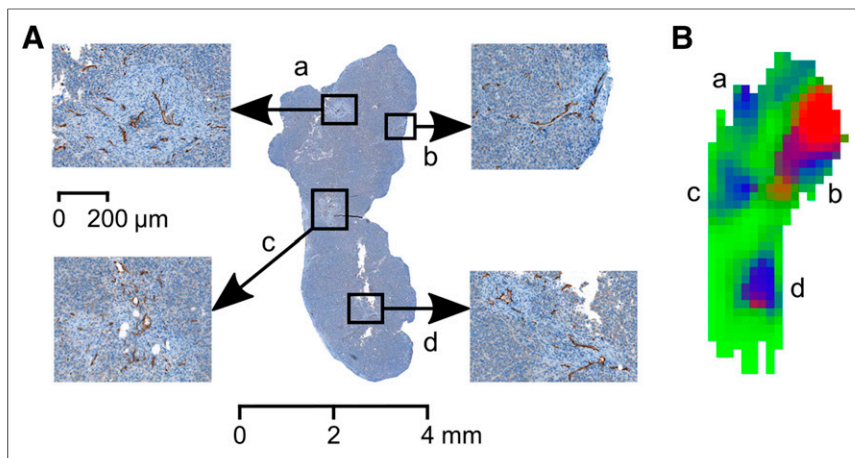
Figure 2 shows the absolute  $\epsilon$  for noisy preclinical time–activity curves with different levels of log-normally distributed noise ( $\beta = 0.1$ –1.5). Among all,  $k_2$  and  $k_3$  showed the highest deviations from the true parameter values. Moreover, the errors in  $k_2$  and  $k_3$  also propagated to  $K_i$ . A similar tendency was seen in the case of noisy clinical time–activity curves (Supplemental Fig. 2), although the  $\epsilon$  for  $k_3$  and  $K_i$  in the clinical case carried less variability than those in the preclinical settings.

The segmentation ability of different clustering methods for noisy preclinical time–activity curves is shown in Figure 3. While the INV-weighted SC exhibited the lowest misclassification error, both the weighted and the unweighted SC techniques outperformed other clustering schemes. Figure 3 also depicts the misclassification errors obtained after clustering the SUV and estimated kinetic parameters. Up to moderate noise levels ( $\beta < 0.7$ ),  $k$ -means and SC applied on the estimated kinetic parameters yielded lower errors in comparison to

clustering the SUV, signifying the efficacy of dynamic measures over the static ones. Supplemental Figure 3 shows the aforementioned clustering results for the clinical scenario. At low noise levels ( $\beta < 0.5$ ), SC on the estimated kinetic parameters displayed the highest accuracy but became worse with a gradual increase in noise. Overall for clinical simulations, the misclassification error of LS-weighted SC remained most steady at all noise levels.



**FIGURE 3.** Misclassification error of various clustering schemes for preclinical simulations with increase in the amount of noise ( $\beta$ ).



**FIGURE 4.** (A) CD31-stained histology of a representative tumor; the 4 insets (scale in  $\mu\text{m}$ ) illustrate high vessel density areas. (B) Segmentation of the tumor into 3 clusters by applying SC on the dynamic  $^{18}\text{F}$ -FDG PET data. The matched clusters are marked as a, b, c, and d in A and B respectively.

Supplemental Figure 4 shows the ground truth and clustering affinity matrices for the noise-free preclinical time-activity curves (shown in Fig. 1B). It is clearly visible that the clustering solution retains the approximate block diagonal structure of the original affinity matrix. Here, the clustering solution corresponds to the INV-weighted SC of the simulated noise-free preclinical time-activity curves. The grid lines in Supplemental Figure 4B give an impression as to the extent of overestimation of class 3 and respective underestimation in class 1.

### Example

Figure 4 shows the segmentation result of INV-weighted SC on an  $^{18}\text{F}$ -FDG measurement. The algorithm effectively identified the densely vascularized regions (depicted with blue and red) from the rest of the tumor (green cluster). The segmented regions were visually validated by CD-31 histology of the tumor section (Fig. 4A). The affinity matrix of the aforementioned clustering solution is shown in Supplemental Figure 5B. The average time-activity curves of well-perfused areas also showed a significantly higher uptake than that of the rest of the tumor (Supplemental Fig. 5C). The parametric maps of this tumor are presented in Figure 5; the figure also shows an  $^{18}\text{F}$ -FDG PET image exhibiting the tumor uptake in the last 20 min of the scan. The outcome of segmenting the tumor parametric maps using SC is shown in Supplemental Figures 5D–5F. It is evident from Supplemental Figure 5B that clustering the tumor time-activity curves yielded compartments with high intracluster similarity, whereas the uncertainties in the parametric maps resulted in poor segmentation of the tumor with low within-cluster similarity (Supplemental Figs. 5D–5F).

### DISCUSSION

This study shows the potential of spectral clustering for the assessment of tumor heterogeneity using dynamic  $^{18}\text{F}$ -FDG PET data. It also contrasts SC with the widely used 2-tissue-compartmental model and the SUV, using dynamic PET simulations of clinically relevant tumor tissue types. The clinical tissue classes were duplicated in the preclinical setting and studied for different levels of noise. A meaningful comparison of the proposed algorithm with compartmental modeling was performed by fitting the noisy time-activity curves and subsequently clustering the esti-

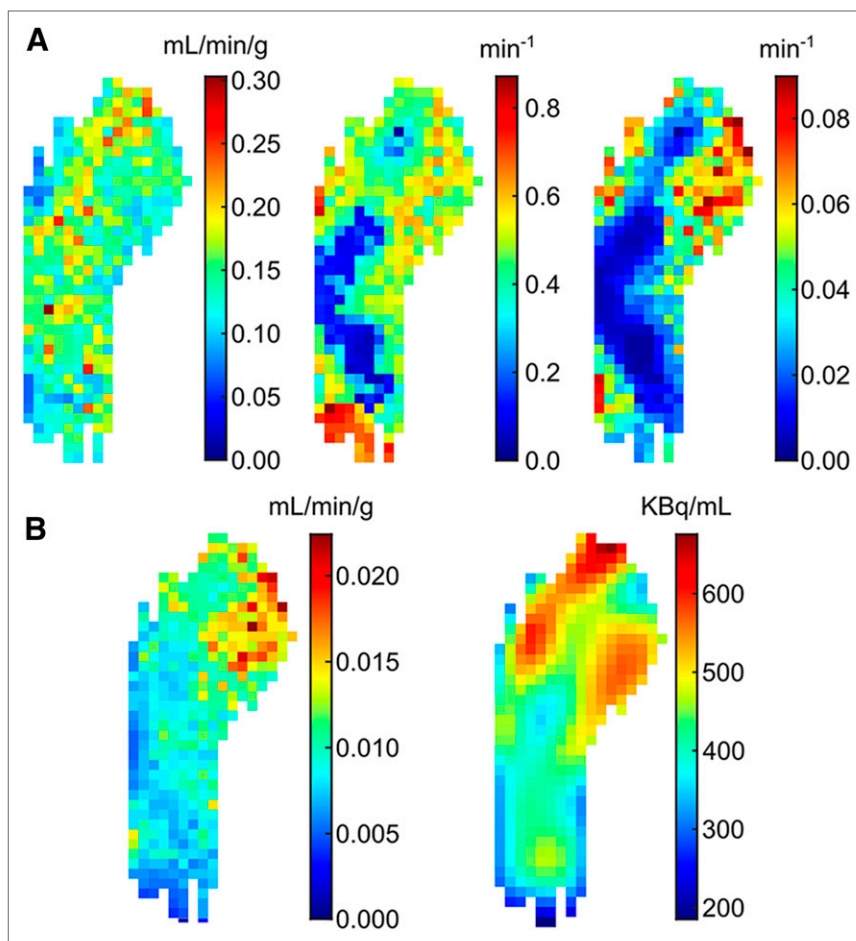
mated kinetic parameters using *k*-means and SC. Furthermore, as a proof of principle we also applied the suggested method to an in vivo mouse model of colon cancer and validated it with histology. Recently, the value of unsupervised segmentation has been shown in a translational study (23). The promising results of SC on the simulated datasets as well as on an in vivo mouse model strongly indicate its potential for dynamic  $^{18}\text{F}$ -FDG PET clinical investigations.

A precise characterization of the tumor microenvironment requires a robust voxel-level analysis. However, the variability of kinetic rate constants with the amount of noise and distortions in AIF (9,24) indicate the shortcomings of compartmental modeling for a voxel-wise analysis. Although clustering the estimated kinetic parameters in the preclinical case produced

more accurate results than clustering the SUV ( $\beta < 0.7$ ), the misclassification error of the INV-weighted SC was lower than that of any of the other schemes. In clinical simulations, SC applied on the estimated kinetic parameters seemed promising at low noise levels but failed to distinguish tumor tissue types accurately as the time-activity curves became noisier. *k*-means and SC errors on the estimated kinetic parameters reflect the best-case scenario for compartmental modeling-based tumor tissue segmentation, because the noisy time-activity curves were modeled using their respective true AIFs (without any shape distortions). Uncertainties in AIF are most likely to introduce adverse effects on kinetic parameter estimation and consequently in parametric map-based tumor tissue segmentation. The poor predictive ability of the SUV in both preclinical and clinical settings was due to the considerable overlap in the last time points of the time-activity curves of all 3 tumor tissue types. This shows that the faster static PET acquisition comes at the cost of vital physiologic information, which can play a principal role in probing intratumoral heterogeneity. The errors caused by noise in kinetic modeling on the other hand, can be minimized to a moderate extent by first using the proposed algorithm for region segmentation and later estimating the kinetic parameters from the averaged time-activity curves.

In dynamic PET imaging, early, middle, and late frames capture different kinetics of the time-activity curve. However, because of non-uniform frame durations and different activity concentration levels they are also affected by varying levels of noise. Thus, while clustering the simulated time-activity curves, we compared the efficacy of 2 different frame-weighting schemes: INV and LS. Whereas the former scheme intuitively favors frames with a higher signal-to-noise ratio, the latter one exploits the intrinsic structure of the high dimensional dynamic PET data. In the analysis of preclinical simulations, the INV-weighted SC performed marginally better than the LS-weighted SC; the opposite was true in the case of clinical simulations.

Some of the results presented in this article may slightly vary with a different choice of frame-sampling schedule. For example, longer early frames might increase the robustness of kinetic parameter estimates at the expense of faster early kinetics. Likewise, the rebinning will also influence the misclassification errors of different clustering schemes. Because this can be an independent study on its own, we did not optimize the simulations for the best framing



**FIGURE 5.** (A) Left to right:  $K_1$ ,  $k_2$ , and  $k_3$  maps of the tumor shown in Figure 4. (B) Left to right:  $K_1$  map calculated using the parametric maps in A and  $^{18}\text{F}$ -FDG uptake in the tumor in the last 20 min of the scan.

schedule. Similar considerations apply to different tracer infusion protocols. Furthermore, to be consistent with Sugawara et al. (15), the 2-tissue-compartmental model was implemented with  $k_4$  and  $F_b$  (fractional blood volume) equal to 0. Although the tissue types identified in colon cancer were different from the simulated tissue classes (except for the viable), the synthetic time-activity curves enabled a thorough objective evaluation of the proposed technique. Moreover, because SC tumor segmentation is based on the intrinsic structure of the underlying data, it can be easily applied to other cancer types as well.

The number of clusters in the example in Figure 4 was determined on the basis of the visual inspection of the data and solution affinity matrices for different number of clusters. Significant off-diagonal similarity between the red and blue clusters was evident from the similar average time-activity curves of the respective regions (Supplemental Fig. 5C). While the blue cluster corresponds well to regions with high vessel density, the red cluster appears to the periphery of the blue regions, resulting in similar uptake patterns. The histology was rigidly registered with the imaging, and we did not perform any nonrigid registration between the two. Although the tumor was carefully partitioned into 2 parts parallel to the transversal field of view, imaging to histology registration remains nontrivial because of the substantial differences in resolution (mm vs.  $\mu\text{m}$ ). Additionally, during the dehydration process the tissue sections undergo a series of nondeterministic affine deformations, which cannot be corrected

using rigid transformations. However, by sectioning the tumor along the reference (imaging) plane and keeping a track of its orientation, the errors in the manual registration can be minimized (25).

Unlike  $k$ -means, SC does not make any assumptions about the shape of the clusters. The efficacy of SC mainly lies in the change of representation (from abstract data points to points in the feature space), which enhances the segregation tendency of the input data. The optimal SC solution depends on the number of chosen eigenvectors from the normalized graph Laplacian. In ideal scenarios, the top  $k$  eigenvectors corresponding to the  $k$  largest eigenvalues of the normalized Laplacian matrix (where  $k$  = number of biologic classes) contain the class discriminative information (20). However, because of the complex microenvironment, resolution limit, and large statistical noise, compartments in oncologic dynamic PET studies often display similar tracer uptake patterns. To a certain extent, these perturbing effects can be dealt with by choosing a larger number of eigenvectors than the potential number of clusters. Throughout our study, we used 6 eigenvectors to segment the dynamic PET data (simulated and measured) into relevant biologic compartments. It has been shown that a prior eigenvector selection can further enhance the clustering stability (26), but we did not explore any such possibility. Additionally, the choice of graph Laplacian can also affect the outcome of SC. As suggested in the literature (20), we used the normalized graph Laplacian rather than the unnormalized one. Also, we did not notice any difference in the performance of two previously established normalized graph Laplacians.

Also, we did not notice any difference in the performance of two previously established normalized graph Laplacians.

A clear limitation of this study is the lack of clinical experimental data; however, accurate alignment of histology to imaging in a clinical setting is difficult to achieve, making validation of intratumoral tissue classes challenging. In preclinical studies, this alignment can be more easily performed. Yet, Figure 4 presents only a qualitative comparison of the segmented tumor compartments with the histology. Future preclinical studies will include an automated nonrigid imaging to histology coregistration to provide reliable quantification of intratumoral heterogeneity. Because PET scanners have a finite spatial resolution, tissue inhomogeneities occurring at the cellular level cannot be observed and analyses are limited to large-scale heterogeneity. Information about variations at this scale has clear potential, for example, in radiotherapy for dose painting and as a basis in image-guided biopsy procedures.

To the best of our knowledge, this is the first study investigating the feasibility of SC for the assessment of the tumor microenvironment incorporating exhaustive dynamic PET simulations and augmented by real data with histologic validation. SC exploits the temporal characteristics of dynamic studies and uses high dimensional embedding (27) to effectively segment the tumor into distinct biologic compartments. This could play an instrumental role in in vivo cancer studies, because the tumor microenvironment stems from complex

genetic alterations and phenotypic interactions, which might not be readily discernible using the existing methods for analyzing dynamic PET measurements.

## CONCLUSION

We have shown the feasibility of SC for the segmentation of 4-dimensional dynamic PET tumor images. The proposed technique showed a performance superior to that of the SUV- and parametric map-based segmentation of tumor tissue variability. Overall, SC can be used as a potential tool for the voxel-level characterization of the tumor microenvironment.

## DISCLOSURE

Grant support was provided by the European Research Council (grant no. 323196), German Ministry for Education and Research/Bundesministerium für Bildung und Forschung (BMBF) (grant no. 0316186E), and Eberhard Karls University Tuebingen (Evaluation of Tumor Heterogeneity Using Clustering of Multi-Modality Imaging Data, Fortuene 2131-0-0). No other potential conflict of interest relevant to this article was reported.

## ACKNOWLEDGMENT

We gratefully thank Dennis Vriens for providing the clinical AIF data and many useful comments on the manuscript.

## REFERENCES

- Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013;501:346–354.
- Soussan M, Orlhac F, Boubaya M, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. *PLoS One*. 2014;9:e94017.
- Zhao B, Hemann MT, Lauffenburger DA. Intratumor heterogeneity alters most effective drugs in designed combinations. *Proc Natl Acad Sci USA*. 2014;111:10773–10778.
- Jadvar H, Alavi A, Gambhir SS. <sup>18</sup>F-FDG uptake in lung, breast, and colon cancers: molecular biology correlates and disease characterization. *J Nucl Med*. 2009;50:1820–1827.
- Gambhir SS. Molecular imaging of cancer with positron emission tomography. *Nat Rev Cancer*. 2002;2:683–693.
- Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med*. 2009;50:11S–20S.
- Dunnwald LK, Doot RK, Specht JM, et al. PET tumor metabolism in locally advanced breast cancer patients undergoing neoadjuvant chemotherapy: value of static versus kinetic measures of fluorodeoxyglucose uptake. *Clin Cancer Res*. 2011;17:2400–2409.
- McDermott GM, Welch A, Staff RT, et al. Monitoring primary breast cancer throughout chemotherapy using FDG-PET. *Breast Cancer Res Treat*. 2007;102:75–84.
- Wang W, Georgi J-C, Nehmeh SA, et al. Evaluation of a compartmental model for estimating tumor hypoxia via FMISO dynamic PET imaging. *Phys Med Biol*. 2009;54:3083–3099.
- Schmidt K, Lucignani G, Moresco RM, et al. Errors introduced by tissue heterogeneity in estimation of local cerebral glucose utilization with current kinetic models of the [<sup>18</sup>F]fluorodeoxyglucose method. *J Cereb Blood Flow Metab*. 1992;12:823–834.
- 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.
- Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. *Comput Biol Med*. 2014;50:76–96.
- Mouysset S, Zbib H, Stute S, et al. Segmentation of dynamic PET images with kinetic spectral clustering. *Phys Med Biol*. 2013;58:6931–6944.
- Muzic RF, Cornelius S. COMKAT: compartment model kinetic analysis tool. *J Nucl Med*. 2001;42:636–645.
- Sugawara Y, Zasadny KR, Grossman HB, Francis IR, Clarke MF, Wahl RL. Germ cell tumor: differentiation of viable tumor, mature teratoma, and necrotic tissue with FDG PET and kinetic modeling. *Radiology*. 1999;211:249–256.
- Vriens D, de Geus-Oei L-F, Oyen WJG, Visser EP. A curve-fitting approach to estimate the arterial plasma input function for the assessment of glucose metabolic rate and response to treatment. *J Nucl Med*. 2009;50:1933–1939.
- Feng D, Huang SC, Wang X. Models for computer simulation studies of input functions for tracer kinetic modeling with positron emission tomography. *Int J Biomed Comput*. 1993;32:95–110.
- Ferl GZ, Zhang X, Wu H-M, Kreissl MC, Huang S-C. Estimation of the <sup>18</sup>F-FDG input function in mice by use of dynamic small-animal PET and minimal blood sample data. *J Nucl Med*. 2007;48:2037–2045.
- Logan J, Fowler JS, Volkow ND, Ding YS, Wang GJ, Alexoff DL. A strategy for removing the bias in the graphical analysis method. *J Cereb Blood Flow Metab*. 2001;21:307–320.
- Von Luxburg U. A tutorial on spectral clustering. *Stat Comput*. 2007;17:395–416.
- Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett*. 2010;31:651–666.
- He X, Cai D, Niyogi P. Laplacian score for feature selection. *Adv Neural Inf Process Syst*. 2005;18:507–514.
- Schmitz J, Schwab J, Schwenck J, et al. Decoding intratumoral heterogeneity of breast cancer by multiparametric in vivo imaging: a translational study. *Cancer Res*. 2016;76:5512–5522.
- Jovkar S, Evans AC, Diksic M, Nakai H, Yamamoto YL. Minimisation of parameter estimation errors in dynamic PET: choice of scanning schedules. *Phys Med Biol*. 1989;34:895–908.
- Divine MR, Katiyar P, Kohlhofer U, Quintanilla-Martinez L, Disselhorst JA, Pichler BJ. A population based Gaussian mixture model incorporating <sup>18</sup>F-FDG-PET and DW-MRI quantifies tumor tissue classes. *J Nucl Med*. 2016;57:473–479.
- Xiang T, Gong S. Spectral clustering with eigenvector selection. *Pattern Recognit*. 2008;41:1012–1029.
- Ham J, Lee DD, Mika S, Schölkopf B. A kernel view of the dimensionality reduction of manifolds. In: *ICML '04 Proceedings of the Twenty-First International Conference on Machine Learning, Banff, Alberta, Canada—July 4–8, 2004*. New York, NY: ACM; 2004:47.