# Data-Driven Mapping of Hypoxia-Related Tumor Heterogeneity Using DCE-MRI and OE-MRI

Adam K. Featherstone [1,2]* James P.B. O'Connor,[2,3,4] Ross A. Little,[1] Yvonne Watson,[1] Sue Cheung,[1] Muhammad Babur,[5] Kaye J. Williams,[2,5] Julian C. Matthews,[1,2] and Geoff J.M. Parker [1,2,6]

**Purpose:** Previous work has shown that combining dynamic contrast-enhanced (DCE)-MRI and oxygen-enhanced (OE)-MRI binary enhancement maps can identify tumor hypoxia. The current work proposes a novel, data-driven method for mapping tissue oxygenation and perfusion heterogeneity, based on clustering DCE/OE-MRI data.

**Methods:** DCE-MRI and OE-MRI were performed on nine U87 (glioblastoma) and seven Calu6 (non-small cell lung cancer) murine xenograft tumors. Area under the curve and principal component analysis features were calculated and clustered separately using Gaussian mixture modelling. Evaluation metrics were calculated to determine the optimum feature set and cluster number. Outputs were quantitatively compared with a previous non data-driven approach.

**Results:** The optimum method located six robustly identifiable clusters in the data, yielding tumor region maps with spatially contiguous regions in a rim-core structure, suggesting a biological basis. Mean within-cluster enhancement curves showed physiologically distinct, intuitive kinetics of enhancement. Regions of DCE/OE-MRI enhancement mismatch were located, and voxel categorization agreed well with the previous non data-driven approach (Cohen's kappa = 0.61, proportional agreement = 0.75).

**Conclusion:** The proposed method locates similar regions to the previous published method of binarization of DCE/OE-MRI enhancement, but renders a finer segmentation of intratumoral oxygenation and perfusion. This could aid in understanding the tumor microenvironment and its heterogeneity.
Magn Reson Med 79:2236–2245, 2018. © 2017 The Authors Magnetic Resonance in Medicine published by Wiley Periodicals, Inc. on behalf of International Society for

[1]Division of Informatics, Imaging & Data Sciences, The University of Manchester, Manchester, UK.

[2]CRUK & EPSRC Cancer Imaging Centre in Cambridge and Manchester, Cambridge and Manchester, UK.

[3]Division of Cancer Studies, The University of Manchester, Manchester, UK.

[4]Department of Radiology, Christie NHS Foundation Trust, Manchester, UK.

[5]Division of Pharmacy & Optometry, The University of Manchester, Manchester, UK.

[6]Bioxydyn Ltd, Manchester, UK.

## INTRODUCTION

Alongside the biological differences between tumors of different patients and within the same patient, there exists considerable functional and structural variation within individual lesions (1). This intra-tumoral heterogeneity has been shown to be a marker of aggressive disease and poor patient prognosis (2,3). Furthermore, the presence of certain phenotypical features in a tumor, such as hypoxia, is believed to contribute to therapeutic resistance (4,5). Hypoxia can limit radiotherapeutic efficacy by reducing the radiosensitivity of tumor tissue and has also been shown to lessen the effect of chemotherapy (6). Drugs are being developed and tested that target hypoxic regions in tumors (7) with optimization and monitoring of treatment potentially benefiting from measuring intra-tumoral hypoxia.

Clinically, characterization of the tumor microenvironment is carried out taking into account only limited spatial information. For hypoxia, voxel-wise heterogeneity imaging is important to allow the detection of hypoxia in a background of normoxia. Oxygen tension probes (8,9) and biopsies followed by histological staining (10,11) are often treated as the "gold standard" for assessing oxygen tension in tissues. However, the sampled portion of tissue may not be representative of the entire tumor, biopsies of the same piece of tumor are non-repeatable, and there may also be issues with accessing the tumor in some locations, such as in many lung cancers. These considerations necessitate a non-invasive method for characterizing the tumor microenvironment and its heterogeneity, particularly one that is sensitive to hypoxia and has the potential to reliably predict and monitor treatment response.

Non-invasive methods currently available for assessing inter- and intra-tumoral oxygenation are largely positron emission tomography-based or MRI-based (5,6). Positron emission tomography has a variety of radiotracers available such as [18F]MISO (12), [18F]FAZA (13), and [64Cu]ATSM (14), which have a high specificity for measuring hypoxia (15), although these more novel radiotracers are restricted to specialist centers with appropriately equipped
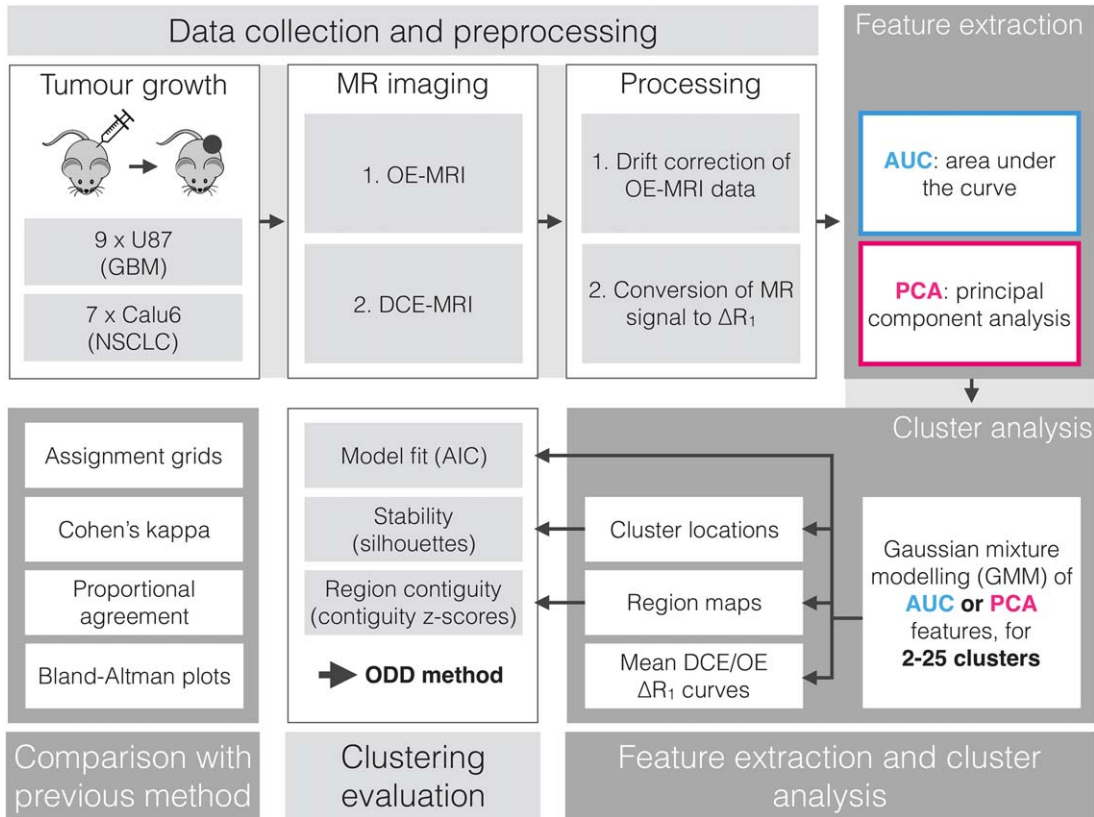
FIG. 1. A flowchart depicting the sequence of acquisition, analysis, and evaluations conducted.

radiochemistry departments. Clinically available methods of using MRI to assess tissue oxygenation (5) fall into those monitoring changes in $T_2^*$ relaxation using blood oxygenation level dependent (BOLD) contrast imaging (16) and those measuring changes in $T_1$ relaxation using tissue oxygenation level-dependent/oxygen-enhanced (TOLD/OE)-MRI (17). Their value in monitoring oxygenation in tumors has been demonstrated independently and in combination with each other (18–26). There are also studies that use dynamic contrast-enhanced (DCE)-MRI derived perfusion biomarkers, such as $K^{trans}$ and $AUC_{90}$, as surrogate measures of hypoxia (27–30).

We have previously carried out DCE-MRI (reflecting tissue perfusion) and OE-MRI (reflecting tissue oxygen delivery) on tumor xenografts, creating three classes of tumor sub-regions based on enhancing/non-enhancing voxels for both imaging techniques (31,32). The class that enhanced with DCE-MRI but not with OE-MRI was termed the "perfused Oxy-R fraction" (refractory to an oxygen challenge) and was postulated to represent hypoxic regions. Important steps have been taken toward providing biological and technical validation (33) of this biomarker for assessing the degree of tumor hypoxia.

Other workers have focused on objective, clustering-based methods for locating distinct tumor (and other tissue) sub-regions in general, without a specific focus on investigating hypoxia. This has been done in preclinical tumor models (34–37), healthy kidneys (38), head and neck cancer (39), cervical cancer (40), bone metastases (41), and breast cancer (42). These studies demonstrate the use of data-clustering techniques in segmenting

tumor and healthy tissue, with varying levels of validation, but none present a systematic optimization of the parameters used in their methods.

The work we present in this paper investigates the application and optimization of data clustering methods such as those described above, but tailored and evaluated specifically on the combination of imaging data that we showed previously to map hypoxia: DCE-MRI and OE-MRI (21,31). We present three ways of evaluating the performance of the clustering method and use these to select recommended parameters and options for the data sets in this study. We then compare our results with our previously published method to assess consistency of results and potential benefits of automated and objective tissue segmentation. We move away from a priori assumptions leading to heuristically chosen thresholds in data and present a method for mapping oxygenation and perfusion heterogeneity that is instead data-driven and demonstrated to be robustly supported by the available signals.

## METHODS

All analyses carried out in this work were carried out in MATLAB R2014a (MathWorks, Natick, MA), apart from the calculation of native tissue $T_1$ values, which was carried out using in-house software written in C. Key MATLAB functions are written in italics (e.g., *function.m*). As an overview of the current work, Figure 1 shows graphically the key steps carried out in our analysis stream, organized into the four main sections of work: (1) data collection and preprocessing, (2) feature

extraction and cluster analysis, (3) clustering evaluation, and (4) comparison with previous method.

## Data Collection and Preprocessing

Studies were performed in compliance with the National Cancer Research Institute "Guidelines for the welfare and use of animals in cancer research" (43) and with licenses issued under the UK Animals (Scientific Procedures) Act 1986 (PPL 40/3212) following local Ethical Committee review.

Experiments were carried out in two murine xenograft models of human cancer, generated by intra-dermal injection of cells on the midline lower back of nude mice. Nine U87 glioblastoma multiforme tumor models were propagated by injecting 0.1 ml of cells ($5 \times 10^6$ cells/ml), and seven Calu6 non-small cell lung carcinoma tumor models were propagated by injecting 0.1 ml of cells ($2 \times 10^7$ cells/ml). When tumors reached $> 200\,mm^3$ by caliper measurement, mice were anaesthetized using 2% isoflurane carried in medical air (21% oxygen). Core temperature was controlled at 37°C while anatomical, OE, and DCE MR imaging was carried out on a 7T Magnex instrument interfaced to a Bruker Advance III console and gradient system using a volume transceiver coil. For each image acquisition, localizer scans were carried out and imaging volumes selected to position tumors centrally within the volume.

### Anatomical Imaging

A $T_2$-weighted scan was carried out to enable tumor identification and localization. 3D rapid acquisition with relaxation enhancement sequence (44): $T_R = 2200.00$ ms; $T_E = 32.00$ ms; $\alpha = 135°$; matrix $= 96 \times 128$; field of view $= 32\,mm \times 32\,mm$; and $16 \times 1$ mm thick coronal slices.

### OE-MRI

A variable flip angle method was carried out to measure tissue baseline $T_1$ (45) using 3D modified driven equilibrium Fourier transform sequences (46): $T_R = 30.00$ ms; $T_E = 1.44$ ms; $\alpha = 5°$, $10°$, $20°$; five averages; matrix $= 64 \times 64$; field of view $= 32\,mm \times 32\,mm$; $16 \times 1$ mm thick coronal slices. This was followed by 42 dynamic acquisitions (acquisition details as above but $\alpha = 20°$ only) at a temporal resolution of one image volume every 28.80 s. The gas supply to the mouse was delivered via a nose cone and switched from air to 100% oxygen at the beginning of the 19th acquisition. The duration of the entire dynamic series was 20 min 10 s.

### DCE-MRI

A variable flip angle method was again carried out to measure tissue baseline $T_1$ using 3D modified driven equilibrium Fourier transform sequences; $T_R = 6.02$ ms; $T_E = 1.46$ ms; $\alpha = 2°$, $5°$, $10°$; five averages; matrix $= 64 \times 64$; field of view $= 32\,mm \times 32\,mm$; $16 \times 1$ mm thick coronal slices. This was followed by 96 dynamic acquisitions (acquisition details as above but $\alpha = 10°$ only) at a temporal resolution of one image volume every 5.78 s. Gd-DOTA was injected into a tail vein (0.25 mmol/kg) at the beginning of the 25th acquisition. The duration of the entire dynamic series was 9 min 15 s.

### Processing

Native tissue $T_1$ ($T_{10}$) values were estimated for both DCE and OE via fitting the spoiled gradient recalled echo signal equation to variable flip angle data (47).

Baseline signal drift was observed in OE data, which was corrected by fitting the spoiled gradient recalled echo signal equation with an empirically determined exponentially time-varying flip angle. Supporting Figure S1 shows an illustration of the fitted baseline signals.

Dynamic MR signals were then converted from raw signal units to voxel-wise $\Delta R_1$ ($R_1 = 1/T_1$) vectors, by manipulating the spoiled gradient recalled echo signal equation (48) and using $T_{10}$, $S_0(t)$, and $\alpha(t)$ values for OE, and uncorrected $T_{10}$, $S_0$, and $\alpha$ values for DCE. Supporting Figure S2 shows the effect that drift correction of OE data has on $\Delta R_1$ values.

Regions of interest were drawn around tumor volumes, and only tumor voxels were used in the following analysis. Erratically enhancing voxels that would impede feature calculation were present in most data sets. These existed likely because of movement at a boundary, and perhaps the inclusion of large vessels, in the region of interest. The area-under-the-curve (AUC) of the modulus of each DCE and OE $\Delta R_1(t)$ curve was calculated, and the voxels with the highest 1% from each imaging technique were removed from the data set to improve the suitability of the data for feature calculation.

## Feature Extraction and Cluster Analysis

Two sets of features were calculated for each tumor voxel: an AUC feature set and a principal component analysis (PCA) feature set.

For the AUC feature set, the area under the first 90 s (post-Gd) of DCE $\Delta R_1(t)$ curves and under all post oxygen-switch time points of OE $\Delta R_1(t)$ curves was calculated using trapezoidal integration using *cumtrapz.m*.

For the PCA feature set, the DCE and OE $\Delta R_1(t)$ enhancement curves were concatenated after scaling the OE data so that the mean standard deviation across all voxels for DCE and OE were equal. This created a composite DCE-OE curve for each voxel. The principal components that describe the directions of greatest variance in DCE-OE $\Delta R_1(t)$ enhancement curves, together with voxel weightings for each component, were then calculated using *pca.m*.

Gaussian mixture modeling (GMM) was applied to the two-dimensional AUC feature set and separately to the four-dimensional PCA feature set, consisting of the first four principal components, with the number of clusters, $N_C$, varying from 2 to 25. GMM was implemented using *fitgmdist.m*. Each GMM fit was repeated ten times with randomly initialized cluster locations to avoid local minima. To ensure that clusters in feature space were consistently labeled, every cluster assignment in feature space was reordered so that clusters 1 to $N_C$ increased from lowest to highest DCE AUC values. For each clustering result (choice of $N_C$ value and choice of feature set), cluster assignments were transferred into image space, creating a 3D region map for each tumor. Additionally, mean within-cluster DCE and OE $\Delta R_1(t)$ curves were calculated.
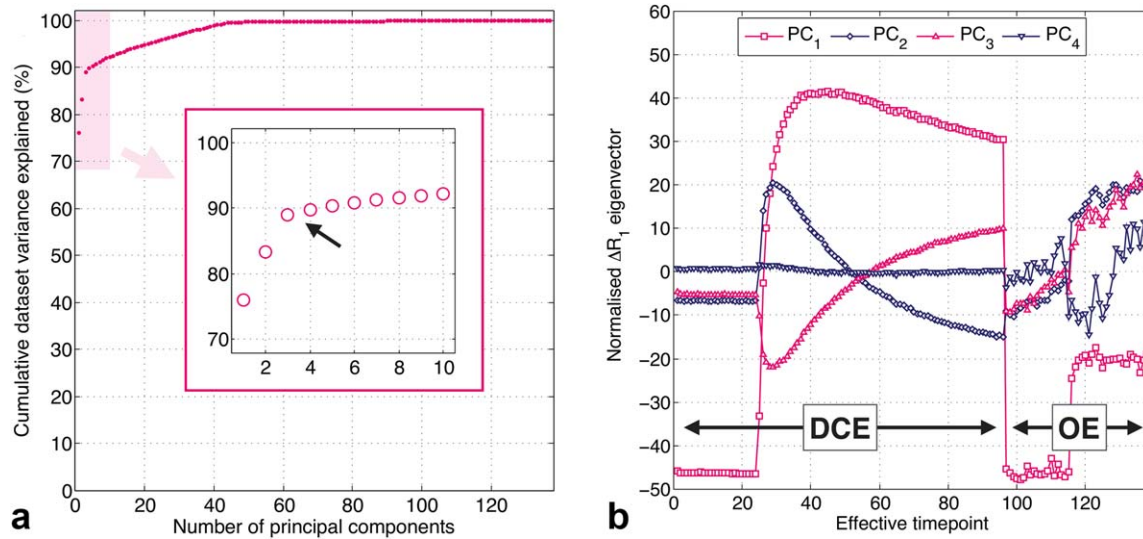
FIG. 2. Results from principal component analysis of combined DCE-OE data (OE $\Delta R_1(t)$ curves were scaled and concatenated with DCE $\Delta R_1(t)$ curves). (**a**) Cumulative variance explained with increasing number of principal components, with the arrow and inset highlighting the first shoulder in the curve ($\sim$2–5 components). (**b**) The first four principal components, showing distinct kinetics of combined DCE-MRI and OE-MRI enhancement.

## Clustering Evaluation

To investigate the suitability of the model fit using GMM clustering, the Akaike information criterion (*AIC*) (49) was used with values derived from the log-likelihood of GMM fits, for each clustering result.

To assess whether region maps identified connected regions in tumors, a contiguity metric was created. The number of connected regions ($3 \times 3 \times 3$ voxel connection kernel) for each tumor region map was calculated for each clustering result using *bwlabeln.m*. The null distribution of this metric was created by bootstrap resampling each region map and calculating the contiguity metric for 100 bootstrap realizations. From these data, contiguity z-scores were calculated for all tumors for each clustering result.

To evaluate the stability of cluster locations within feature space, both feature sets underwent bootstrap resampling over voxels 100 times. GMM for all $N_C$ values was then rerun on every bootstrapped feature set and the cluster centers saved. The set of cluster centers from each bootstrap realization were matched to the original cluster centers using the Hungarian algorithm (50), and silhouettes (51) were calculated using *silhouette.m* to quantitatively assess the dispersion of each cluster center for each realization of bootstrap resampling.

The three evaluation metrics (*AIC*, contiguity z-scores, and stability scores) were compared alongside each other and used to select (1) the feature set and (2) the $N_C$ value for GMM clustering of this data set. This is the optimized, data-driven (ODD) method we present.

## Comparison with Previous Method

Alongside the ODD method, the previously published threshold-based method (TBM) (31) that defined three tumor region classes was also applied to these data. Cluster assignments from ODD methods were compared with the three TBM classes and subsequently concatenated into three similar classes to enable a direct comparison between methods. For each tumor, the proportion of voxels assigned to the same class ($\varphi$) and Cohen's kappa ($\kappa$) were calculated to quantify the agreement between methods. Bland-Altman plots of the number of voxels in each of the three classes were also created as a method of assessing methodological agreement.

## RESULTS

### Feature Extraction and Cluster Analysis

Supporting Figure S3 shows $T_2$-weighted images of four representative tumors. Figure 2 shows results from PCA applied to all voxels' DCE-OE $\Delta R_1(t)$ time series. Figure 2a shows the cumulative variance explained with increasing number of principal components (PCs). The shoulder at 40–45 PCs most likely reflects the different noise levels within the OE and DCE measurements, as there are 42 OE time points. There is an initial shoulder in the curve at two to five PCs, at which point $\sim$90% of the variance in the data set is explained (Fig. 2a, inset). The first four PCs were chosen to create the PCA feature set based on the location of this shoulder and visual interpretation of the PCs. Figure 2b shows these first four PCs, with the first part of the curves representing the DCE and the later part the OE $\Delta R_1(t)$ signal changes. Smooth temporal changes that might be expected because of the administered Gd and oxygen are observed.

Figure 3 shows feature maps for the central slice of the same tumor in Supporting Figure S3b. AUC maps show a strong rim-core structure for both features, and regions of high $AUC_{DCE}$ with low or negative $AUC_{OE}$ are observed (see arrows). $PC_1$ and $PC_2$ feature maps show structures with similarity to those of the $AUC_{DCE}$ and $AUC_{OE}$ features, respectively, whereas $PC_3$ and $PC_4$ feature maps appear to show spatial structure distinct from
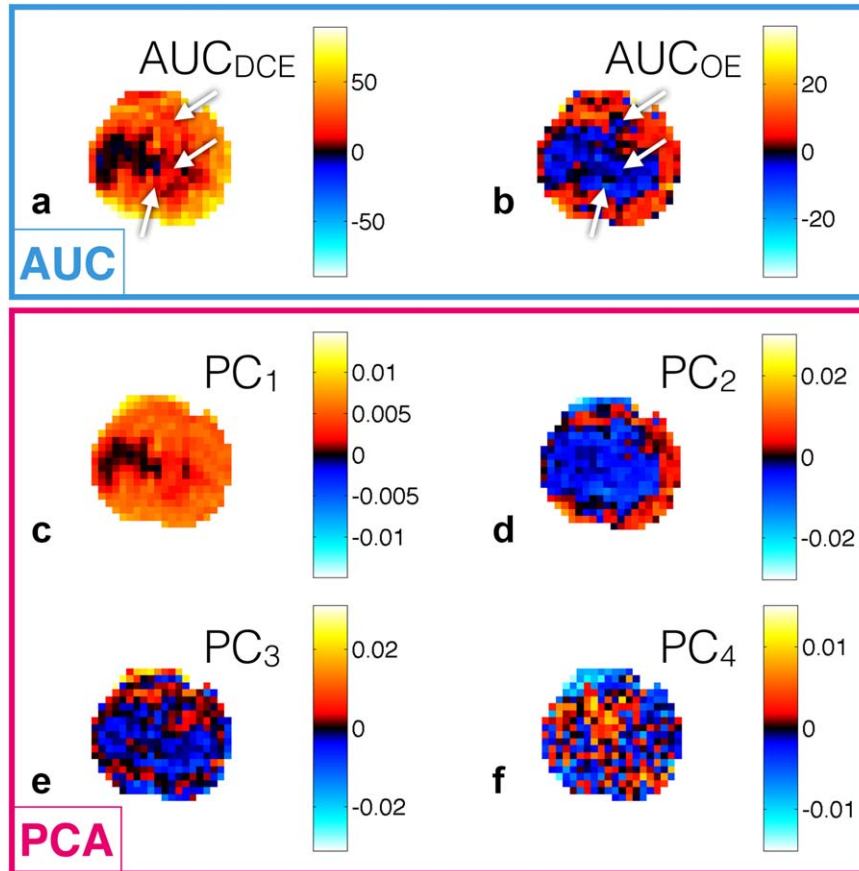
FIG. 3. Feature maps for the representative U87 tumor shown in Supporting Figure S3b. (**a** and **b**) Feature maps for the AUC feature set. (**c–f**) Feature maps of PCA weightings for the first four principal components. Arrows highlight regions of $AUC_{DCE}/AUC_{OE}$ mismatch. There is structural similarity of (**a**) and (**b**) with (**c**) and (**d**), respectively.

that of $PC_1$ and $PC_2$. Feature maps for all tumors are shown in Supporting Figures S4–S6.

Supporting Figure S7 shows two-dimensional histograms of the AUC and PCA feature sets.

## Clustering Evaluation

Figure 4 shows *AIC* values (Figs. 4a, b), contiguity z-scores (Figs. 4c, d), and stability scores (Figs. 4e, f) from GMM fits with varying $N_C$ for the AUC and PCA feature sets. For *AIC* values (Figs. 4a, b), neither feature set shows a clear minimum (optimal) value, but both show a steep decrease in *AIC* up to $N_C = 6$. After this point, there is a much shallower decrease in *AIC* with increasing $N_C$, suggesting that only modest improvement in describing the feature set distribution is achieved with more than six clusters, particularly for the AUC feature set. For contiguity z-scores (Figs. 4c, d), most values are greater than three for all $N_C$, indicating a probability of <0.3% that the contiguity of regions located in tumor region maps is because of chance. For both feature sets, there is a general trend of increasing contiguity z-scores in tumors with increasing $N_C$. For stability scores (Figs. 4e, f), values of $+1$ indicate perfect stability of cluster results. Clustering using the AUC feature set results in highly stable cluster centers when $N_C \leq 4$, whereas when using the PCA feature set the cluster centers remain highly stable when $N_C \leq 6$, although less stable cluster centers are observed when $N_C = 3$. When $N_C > 6$, a steady deterioration in stability is seen for both feature sets.

## ODD Method

*AIC* values suggest that $N_C > 6$ gives only marginal increases in GMM fit quality (for both feature sets), and stability scores remain high when $N_C \leq 6$ for the PCA feature set, whereas for the AUC feature set stability only remains high when $N_C \leq 4$. The optimized feature set and number of clusters for this data set was therefore determined to be clustering the PCA feature set with $N_C = 6$.

Results from the ODD method are shown in Figures 5 and 6 and Supporting Figures S8 and S9.

Figure 5 shows tumor region maps from the ODD method for the same four representative tumors shown in Supporting Figure S3. Contiguous regions are located, with the exception of a few voxels on the tumor periphery, and tumors show a rough rim-core structure. Heterogeneous structures are located in approximately half of the tumors, exemplified by Figures 5b, c. Region maps have different cluster compositions depending on tumor line: voxels in clusters 1 and 2 predominantly belong to Calu6 tumors (694 U87 voxels, 8050 Calu6 voxels, 1:11.6 ratio) whereas voxels in clusters 4 and 6 are predominantly in U87 tumors (7180 U87 voxels, 1467 Calu6 voxels, 4.89:1 ratio).

Figure 6 shows mean within-cluster DCE and OE $\Delta R_1(t)$ enhancement curves defined using the ODD method, with error bars of the standard error of the mean plotted. Curves show distinct kinetics of enhancement following Gd administration (Fig. 6a) and the oxygen-switch (Fig. 6b). Cluster 1 enhances in neither DCE nor
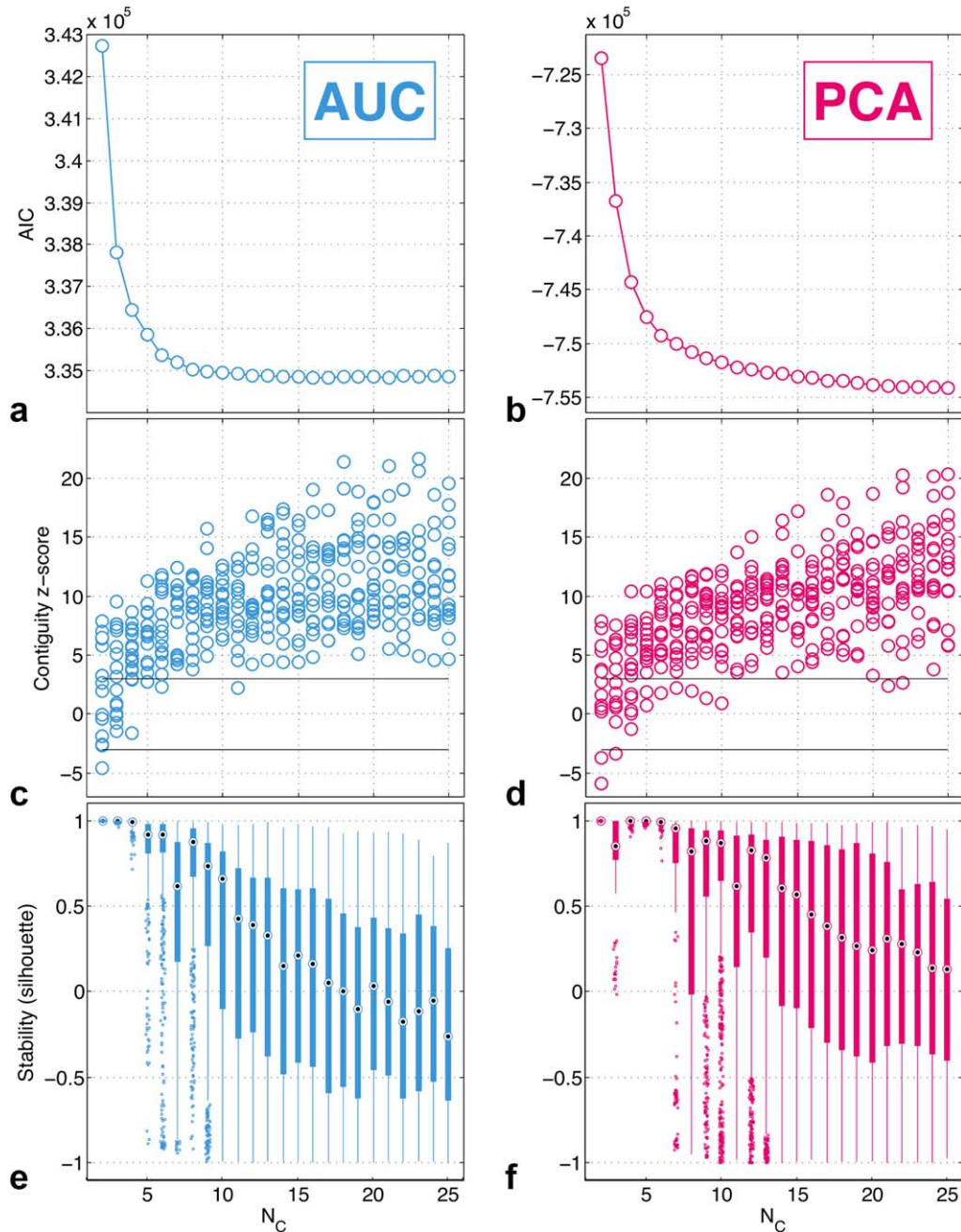
FIG. 4. Evaluation metrics from GMM fits of the AUC and PCA feature sets with varying number of clusters ($N_C$). (**a** and **b**) Akaike information criterion (*AIC*). Neither curve shows a clear minimum, with lower values observed for higher numbers of clusters. (**c** and **d**) Contiguity z-scores. There are 16 z-scores for each $N_C$ value, one for each tumor. Most values lie above 3, indicating statistically significantly greater contiguity in region maps than what would appear because of chance. (**e** and **f**) Stability scores. Each box contains $N_C$ times 100 cluster centers with a silhouette value calculated for each cluster center. Cluster center locations remain stable (located in similar areas of the feature space with stability scores of close to +1) for up to 4 clusters for AUC, and for up to 6 clusters for PCA.

OE, clusters 2 and 3 enhance only modestly in DCE, and clusters 4–6 enhance strongly in both DCE and OE. Regions with these characteristics have been previously shown to correspond to necrotic, hypoxic, and well-perfused/well-oxygenated tumor tissue, respectively, in multiple xenograft models (31). For our work, correlative histology was not available, and caution must be taken in assigning biological interpretations to clusters.

Comparison with Previous Method

Supporting Figure S10 shows assignment grids showing how tumor voxels are classified using the ODD method and TBM. Inspection of mean enhancement curves provided justification for concatenating clusters 2 and 3 into a single group, and 4, 5, and 6 into another group. This enabled further comparison of ODD methods with TBM. Supporting Figure S11 shows a side-by-side comparison
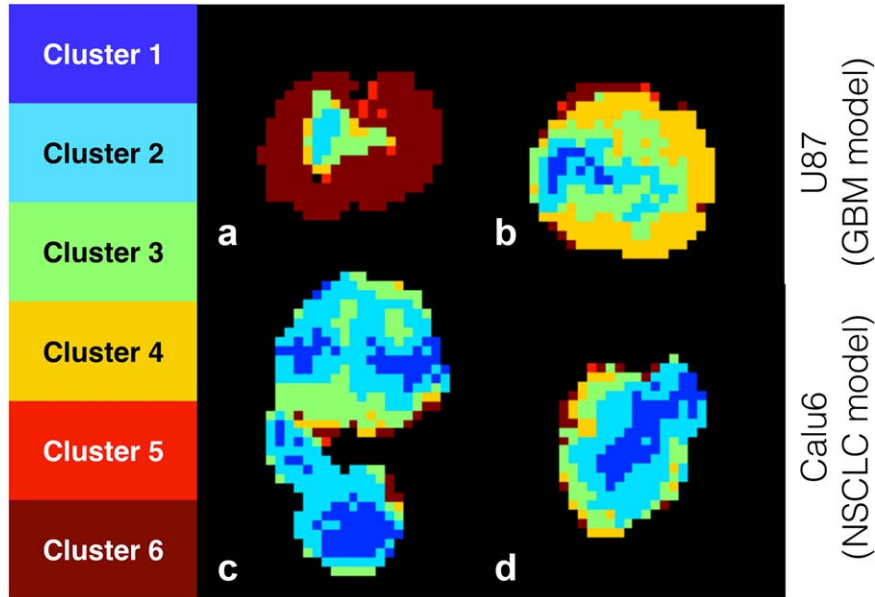
FIG. 5. Results from the ODD method. Tumor region maps for the two representative U87 tumors and two representative Calu6 tumors shown in Supporting Figure S3. Largely contiguous regions are located, with rim-core structures present in most tumors.

of tumor region maps created using both methods. The majority of tumors show a strong visual concordance between methods. Supporting Figure S12 shows proportional agreement ($\varphi$) and Cohen's kappa ($\kappa$) statistics, quantifying the agreement between methods of categorizing tumor voxels. Looking at the group values in Supporting Figure S12, both statistics show good agreement between methods: $\varphi = 0.75$ and $\kappa = 0.61$. The low values in tumors 3, 5, 14, and 15 correspond to cases where the majority of voxels are classified into a single class using one of the methods and so $\kappa$ gives artificially low values. Supporting Figure S13 shows Bland-Altman plots for the number of voxels in each of the three categories determined by both methods.

## DISCUSSION

The current work presents a thorough investigation into the application of data-led clustering methods to MRI data designed to identify putatively hypoxic tumor tissue, putting forward a combination of evaluation metrics that enable objective methodological optimization. Our results agree well with the previously published work (31,32), however we demonstrate the ability of the data to robustly support more than the three tissue classes previously located, indicating that there may be further information to exploit in the combination of DCE and OE data if methods such as those we present here are adopted.

Soft clustering methods (where clusters may overlap) generally perform more robustly than hard clustering methods (where clusters cannot overlap) under high levels of noise and when data distributions are smoothly varying. GMM was chosen as it is a well-known soft clustering technique that gives a probabilistic assignment of each voxel to each cluster, though the most probable cluster for each voxel was used as a hard classification. In this way, the final output is clear-cut and can be



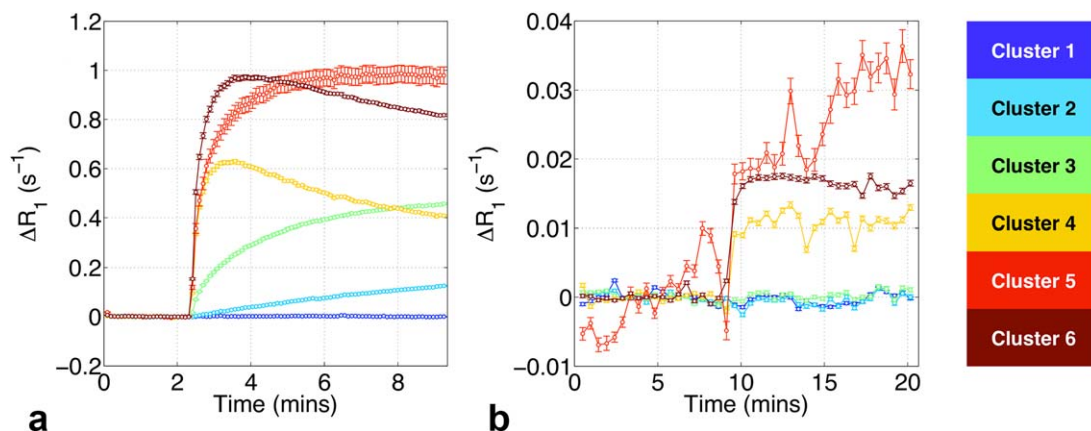FIG. 6. Results from the ODD method. Mean within-cluster $\Delta R_1(t)$ enhancement curves for DCE-MRI (**a**) and OE-MRI (**b**). Error bars show standard error of the mean. Curves show distinct, intuitive kinetics of enhancement, with lack of overlap in the post-contrast regions. Clusters 2 and 3 (light blue and green) show DCE-MRI enhancement with no OE-MRI enhancement, possibly linked with hypoxia.

easily associated with a physiological interpretation. However, this can lead to classification errors in borderline data lying between two clusters, but maintaining high cluster assignment robustness is intrinsic to the ODD method and should mitigate errors of this sort. The probabilistic cluster assignment from GMM could be maintained or an alternative soft clustering method (e.g., fuzzy c-means) used, which would avoid these errors. This might provide insight into the overlap, or lack thereof, of biological habitats in tumors, or at least allow the identification of regions with poorly categorized features.

It is far from trivial to select the most appropriate clustering method, lending importance to our combination of evaluation metrics. Contiguity z-scores (Figs. 4c, d) inform us that transferring cluster assignments into image space gives genuine structure, as opposed to the random categorization of voxels. This suggests that located tumor regions are based on biology, but does not provide a strong basis for selecting the most appropriate feature set or number of clusters. The absence of clear minima in *AIC* values (Figs. 4a, b) is likely because of the lack of distinct voxel groupings in feature space. However, the shoulders in *AIC* curves suggest that we do need at least six clusters to adequately fit the data. The stability analysis (Figs. 4e, f) provides an objective measure of the number of clusters that can be reliably supported by these data. From these metrics, running GMM clustering on the AUC feature set with four or fewer clusters and the PCA feature set with six or fewer clusters appeared to be optimal. Based on this, we selected the PCA feature set with the most clusters that the data could robustly support (and therefore more tissue classes), potentially providing more insight into the tumor microenvironment and its heterogeneity. PCA also has the advantage of requiring fewer a priori assumptions in the analysis. We desire a high number of clusters to identify features of potential importance and adequately characterize tumor heterogeneity, while simultaneously desiring high repeatability and reliability of our results, which typically deteriorates as the number of clusters increases. If greater than the optimum number of clusters is used, the clustering is increasingly subject to random errors in the data, and if fewer are used, the clustering fails to account for potentially important information contained in the data. The combination of evaluation methods used in this study provides an objective measure of where to place this trade-off—selecting the highest number of stable clusters the data can support and ensuring that significant regional contiguity exists in tumor region maps.

A potential validation of this work that was not carried out was a comparison of region maps with histologically stained tumor sections. The tumors investigated were part of a separate longitudinal radiotherapy study; all imaging data in the current study is from the baseline, pre-radiotherapy visit. Histological staining was acquired, but treatment effects and significant tumor growth before excision (up to 14 days later) prevent any valid comparisons being made with the imaging data used in the current study. However, TBM has itself been histologically validated (pimonidazole staining correlating strongly with perfused Oxy-R fraction (31)), and we demonstrate strong agreement between the ODD method and TBM, cross-validating both techniques. ODD results were necessarily simplified (six classes combined into three) to enable comparisons with TBM. A significant strength of this work therefore lies in the finer distinction the ODD method makes between tissue classes, but, in the absence of ground truth histological data, we are unable to definitively interpret the biological relevance of this distinction.

## CONCLUSIONS

The development of a non-invasive, objective, and robust tool for characterizing the tumor microenvironment is necessitated by the heterogeneous nature of tumor tissue and the push toward personalized medicine. Here, we present a data-led methodology for classifying multi-spectral tumor imaging data, alongside methods for optimizing parameter choices. We apply our methods to a cohort of DCE-MRI and OE-MRI preclinical tumor data, demonstrating the ability of our methods to extract physiologically distinct regions within tumors. Oxygenation and perfusion related characteristics are successfully spatially mapped in the tumors studied, with varying degrees of intra-tumoral heterogeneity across the cohort. We show that our methodology agrees well with a previously published method for locating hypoxia in tumors, itself validated with correlative histology, and our methods identify a greater number of distinct tumor regions than the previous method. Further work is required to ascertain the biological significance of this result, and future work will include the application of our methods in human tumors, assessing the potential of combined DCE/OE-MRI clustering in a clinical setting.

## REFERENCES

1. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta 2010;1805:105–117.
2. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature 2013;501:346–354.
3. 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.
4. Vaupel P, Mayer A. Hypoxia in cancer: significance and impact on clinical outcome. Cancer Metastasis Rev 2007;26:225–239.
5. Padhani AR, Krohn KA, Lewis JS, Alber M. Imaging oxygenation of human tumours. Eur Radiol 2007;17:861–872.
6. Hammond EM, Asselin MC, Forster D, O'Connor JPB, Senra JM, Williams KJ. The meaning, measurement and modification of hypoxia in the laboratory and the clinic. Clin Oncol (R Coll Radiol) 2014;26: 277–288.
7. Hunter FW, Wouters BG, Wilson WR. Hypoxia-activated prodrugs: paths forward in the era of personalised medicine. Br J Cancer 2016; 114:1071–1077.
8. Kallinowski F, Zander R, Hoeckel M, Vaupel P. Tumor tissue oxygenation as evaluated by computerized-po2-histography. Int J Radiat Oncol Biol Phys 1990;19:953–961.
9. Vaupel P, Schlenger K, Knoop C, Hockel M. Oxygenation of human tumors: evaluation of tissue oxygen distribution in breast cancers by computerized O2 tension measurements. Cancer Res 1991;51:3316–3322.
10. Williams KJ, Parker CA, Stratford IJ. Exogenous and endogenous markers of tumour oxygenation status: definitive markers of tumour hypoxia? Adv Exp Med Biol 2005;566:285–94.
11. Bussink J, Kaanders JHAM, Van Der Kogel AJ. Tumor hypoxia at the micro-regional level: clinical relevance and predictive value of

exogenous and endogenous hypoxic cell markers. Radiother Oncol 2003;67:3–15.

12. Rasey JS, Koh WJ, Evans ML, Peterson LM, Lewellen TK, Graham MM, Krohn KA. Quantifying regional hypoxia in human tumors with positron emission tomography of [18F]fluoromisonidazole: a pretherapy study of 37 patients. Int J Radiat Oncol Biol Phys 1996;36:417–428.

13. Piert M, Machulla H, Picchio M, et al. Hypoxia-specific tumor imaging with 18f-fluoroazomycin arabinoside. J Nucl Med 2005;46:106–113.

14. Lewis JS, McCarthy DW, McCarthy TJ, Fujibayashi Y, Welch MJ. Evaluation of 64Cu-ATSM in vitro and in vivo in a hypoxic tumor model. J Nucl Med 1999;40:177–183.

15. Lopci E, Grassi I, Chiti A, Nanni C, Cicoria G, Toschi L, Fonti C, Lodi F, Mattioli S, Fanti S. PET radiopharmaceuticals for imaging of tumor hypoxia: a review of the evidence. Am J Nucl Med Mol Imaging 2014;4:365–384.

16. Ogawa S, Lee TM, Nayak AS, Glynn P. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson Med 1990;14:68–78.

17. Ohno Y, Hatabu H. Basics concepts and clinical applications of oxygen-enhanced MR imaging. Eur J Radiol 2007;64:320–328.

18. Burrell JS, Walker-Samuel S, Baker LCJ, Boult JKR, Jamin Y, Halliday J, Waterton JC, Robinson SP. Exploring $\Delta$R2* and $\Delta$R1 as imaging biomarkers of tumor oxygenation. J Magn Reson Imaging 2013;38:429–434.

19. O'Connor JPB, Naish JH, Jackson A, et al. Comparison of normal tissue R1 and R*2 modulation by oxygen and carbogen. Magn Reson Med 2009;61:75–83.

20. Ganesh T, Estrada M, Duffin J, Cheng H-LM. T2* and T1 assessment of abdominal tissue response to graded hypoxia and hypercapnia using a controlled gas mixing circuit for small animals. J Magn Reson Imaging 2016;44:305–316.

21. Linnik IV, Scott MLJ, Holliday KF, et al. Noninvasive tumor hypoxia measurement using magnetic resonance imaging in murine U87 glioma xenografts and in patients with glioblastoma. Magn Reson Med 2014;71:1854–1862.

22. O'Connor JPB, Naish JH, Parker GJM, et al. Preliminary study of oxygen-enhanced longitudinal relaxation in MRI: a potential novel biomarker of oxygenation changes in solid tumors. Int J Radiat Oncol Biol Phys 2009;75:1209–1215.

23. Remmele S, Sprinkart AM, Müller A, et al. Dynamic and simultaneous MR measurement of R1 and R 2* changes during respiratory challenges for the assessment of blood and tissue oxygenation. Magn Reson Med 2013;70:136–146.

24. Hallac RR, Zhou H, Pidikiti R, Song K, Stojadinovic S, Zhao D, Solberg T, Peschke P, Mason RP. Correlations of noninvasive BOLD and TOLD MRI with pO2 and relevance to tumor radiation response. Magn Reson Med 2014;71:1863–1873.

25. Matsumoto KI, Bernardo M, Subramanian S, Choyke P, Mitchell JB, Krishna MC, Lizak MJ. MR assessment of changes of tumor in response to hyperbaric oxygen treatment. Magn Reson Med 2006;56:240–246.

26. Beeman SC, Shui Y-B, Perez-Torres CJ, Engelbach JA, Ackerman JJH, Garbow JR. O2-sensitive MRI distinguishes brain tumor versus radiation necrosis in murine models. Magn Reson Med 2016;75:2442–2447.

27. Egeland TAM, Gulliksrud K, Gaustad JV, Mathiesen B, Rofstad EK. Dynamic contrast-enhanced-MRI of tumor hypoxia. Magn Reson Med 2012;67:519–530.

28. Yopp AC, Schwartz LH, Kemeny N, et al. Antiangiogenic therapy for primary liver cancer: correlation of changes in dynamic contrast-enhanced magnetic resonance imaging with tissue hypoxia markers and clinical response. Ann Surg Oncol 2011;18:2192–2199.

29. Ellingsen C, Hompland T, Galappathi K, Mathiesen B, Rofstad EK. DCE-MRI of the hypoxic fraction, radioresponsiveness, and metastatic propensity of cervical carcinoma xenografts. Radiother Oncol 2014;110:335–341.

30. Halle C, Andersen E, Lando M, et al. Hypoxia-induced gene expression in chemoradioresistant cervical cancer revealed by dynamic contrast-enhanced MRI. Cancer Res 2012;72:5285–5295.

31. O'Connor JPB, Boult JKR, Jamin Y, et al. Oxygen-enhanced MRI accurately identifies, quantifies, and maps tumor hypoxia in preclinical cancer models. Cancer Res 2016;76:787–795.

32. Dewhirst MW, Birer SR. Oxygen-enhanced MRI is a major advance in tumor hypoxia imaging. Cancer Res 2016;76:769–772.

33. O'Connor JPB, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14:169–186.

34. Berry LR, Barck KH, Go MA, et al. Quantification of viable tumor microvascular characteristics by multispectral analysis. Magn Reson Med 2008;60:64–72.

35. Henning EC, Azuma C, Sotak CH, Helmer KG. Multispectral quantification of tissue types in a RIF-1 tumor model with histological validation. Part I. Magn Reson Med 2007;57:501–512.

36. Katiyar P, Divine MR, Kohlhofer U, Quintanilla-Martinez L, Scholkopf B, Pichler BJ, Disselhorst JA. A novel unsupervised segmentation approach quantifies tumor tissue populations using multiparametric MRI: first results with histological validation. Mol Imaging Biol 2017;19:391–397.

37. Zöllner FG, Heilmann M, Walczak C, Volk A, Schad LR. Tumor tissue analysis by self organizing maps from combined DCE-/DSC-MRI data. In Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis, Salzburg, Austria, 2009. p. 562–568.

38. Zöllner FG, Sance R, Rogelj P, Ledesma-Carbayo MJ, Rørvik J, Santos A, Lundervold A. Assessment of 3D DCE-MRI of the kidneys using non-rigid image registration and segmentation of voxel time courses. Comput Med Imaging Graph 2009;33:171–181.

39. Wang P, Popovtzer A, Eisbruch A, Cao Y. An approach to identify, from DCE MRI, significant subvolumes of tumors related to outcomes in advanced head-and-neck cancer. Med Phys 2012;39:5277–5285.

40. Torheim T, Groendahl AR, Andersen EKF, Lyng H, Malinen E, Kvaal K, Futsaether CM. Cluster analysis of dynamic contrast enhanced MRI reveals tumor subregions related to locoregional relapse for cervical cancer patients. Acta Oncol 2016;55:1294–1298.

41. Blackledge MD, Rata M, Tunariu N, Koh D-M, George A, Zivi A, Lorente D, Attard G, de Bono JS, Leach MO, et al. Visualizing whole-body treatment response heterogeneity using multi-parametric magnetic resonance imaging. J Algorithm Comput Technol 2016;10:290–301.

42. Wu J, Gong G, Cui Y, Li R. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. J Magn Reson Imaging 2016;44:1107–1115.

43. Workman P, Aboagye EO, Balkwill F, Balmain A, Bruder G, Chaplin DJ, Double JA, Everitt J. Guidelines for the welfare and use of animals in cancer research. Br J Cancer 2010;102:1555–1577.

44. Hennig J, Nauerth A, Friedburg H. RARE imaging: a fast imaging method for clinical MR. Magn Reson Med 1986;3:823–833.

45. Christensen KA, Grant DM, Schulman EM, Walling C. Optimal determination of relaxation times of fourier transform nuclear magnetic resonance. Determination of spin-lattice relaxation times in chemically polarized species. J Phys Chem 1974;78:1971–1977.

46. Ugurbil K, Garwood M, Ellermann J, et al. Imaging at high magnetic fields: initial experiences at 4 T. Magn Reson Q 1993;9:259–277.

47. Fram EK, Herfkens RJ, Johnson GA, Glover GH, Karis JP, Shimakawa A, Perkins TG, Pelc NJ. Rapid calculation of T1 using variable flip angle gradient refocused imaging. Magn Reson Imaging. 1987;5:201–208.

48. Yankeelov T, Gore J. Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples. Curr Med Imaging Rev 2009;3:91–107.

49. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control 1974;19:716–723.

50. Kuhn HW. The Hungarian method for the assignment problem. Nav Res Logist Q 1955;2:83–97.

51. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Fig. S1.** Mean OE-MRI signal values from each tumor (blue circles), with the fit of an exponentially time-varying baseline to pre-oxygen enhancement time points shown, and extrapolated to the post contrast time points (pink line).

**Fig. S2.** Mean OE-MRI $\Delta$R$_1$(t) values for each tumor, calculated without any form of drift correction (blue circles) and with our drift correction (pink circles).

**Fig. S3.** Representative $T_2$-weighted images of central slices through two U87 tumors (**a** and **b**) and two Calu6 tumors (**c** and **d**), demonstrating tumor anatomy and acquisition field-of-view. (**a**) An irregularly shaped tumor with a hyper-intense region, possibly corresponding to edema. (**c**) A tumor with an even more irregular, bi-lobular shape, but with comparatively homogeneous signal intensities. (**b**) A largely homogeneous tumor with a circular cross-section. (**d**) A tumor with a similar level of homogeneity but with an elliptical cross-section.

**Fig. S4.** $AUC_{DCE}$ and PC1 (first principal component) feature maps for a central slice through all 16 tumors. Note the similarity in structure in most tumors between $AUC_{DCE}$ and PC1.

**Fig. S5.** $AUC_{OE}$ and PC2 (second principal component) feature maps for a central slice through all 16 tumors. Note the similarity in structure in most tumors between $AUC_{OE}$ and PC2.

**Fig. S6.** PC3 and PC4 (third and fourth principal component) feature maps for a central slice through all 16 tumors.

**Fig. S7.** Two-dimensional histograms of the AUC and PCA feature sets alongside Spearman's $\rho$ values, with the four-dimensional PCA feature set split into its two-dimensional projections. The AUC feature set shows moderate correlation between the inputs, whereas the PCA feature set shows no strong correlation, indicating good separation of information between the four components. Neither feature set shows clear, distinct separations between voxel groupings, but we observe smooth changes in density of the feature space distributions. All plots show a dense occupation of feature space around the origin (yellow histogram bins).

**Fig. S8.** Results from the ODD method. Cluster assignments in feature space. Plots show two-dimensional projections of the PCA feature set, with color-coded cluster assignments to voxels. Clusters 1 and 2 (dark and light blue) show a large overlap, which describes the dense region around the origin in feature space, with clusters 3, 4, and 6 (green, yellow, and brown) showing much less overlap and characterizing the less dense regions of feature space. Cluster 5 (red) represents a large, diffuse Gaussian distribution of voxels that do not appear to belong to any of the other clusters.

**Fig. S9.** Results from the ODD method. Tumor region maps for central slices of all tumors, with color-coded cluster assignments to voxels. Rough rim-core structures are present in most tumors.

**Fig. S10.** Assignments grids showing how the previously published TBM and our ODD method compare at assigning voxels to categories (top row). After concatenating the six categories from ODD into three classes, the bottom row shows assignment grids of TBM versus ODD (cat), with proportional agreement ($\varphi$) and Cohen's kappa ($\kappa$) statistics calculated.

**Fig. S11.** Side-by-side comparison of region maps from the previously published TBM and from the optimized, data-driven method concatenated into three classes (ODD [cat]).

**Fig. S12.** Proportional agreement ($\varphi$) and Cohen's kappa ($\kappa$), calculated to rate agreement between the previously published, TBM and the ODD method. Values were calculated for individual tumors, the minimum, median, and max value highlighted, and group statistics were calculated for each tumor line and for the whole cohort.

**Fig. S13.** Bland-Altman plots for the number of voxels in each of the three categories determined using the previously published TBM and using the ODD method. For categories 1, 2, and 3, the bias in voxel categorization (ODD $-$ TBM) is $-19.7$, $+115.6$, and $-95.9$ voxels, corresponding to $(-)1.4\%$, $8.1\%$, and $(-)6.8\%$ of the mean tumor size (1419.2 voxels).