



**HAL**  
open science

# New method for quantification of intratumoral heterogeneity: a feasibility study on Ktrans maps from preclinical DCE-MRI

Charly Girot, Andreas Volk, Christine Walczak, Nathalie Lassau, Stéphanie Pitre

## ► To cite this version:

Charly Girot, Andreas Volk, Christine Walczak, Nathalie Lassau, Stéphanie Pitre. New method for quantification of intratumoral heterogeneity: a feasibility study on Ktrans maps from preclinical DCE-MRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 2021, 34, 10.1007/s10334-021-00930-3 . hal-03328004

**HAL Id: hal-03328004**

**<https://hal.science/hal-03328004>**

Submitted on 27 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# New Method for Quantification of Intratumoral Heterogeneity: A feasibility study on $K^{\text{trans}}$ maps from Preclinical DCE-MRI.

Charly Girot<sup>1</sup> · Andreas Volk<sup>1</sup> · Christine Walczak<sup>3</sup> · Nathalie Lassau<sup>1,2</sup> · Stéphanie Pitre-Champagnat<sup>1</sup>



Corresponding author : [Charly.girot@universite-paris-saclay.fr](mailto:Charly.girot@universite-paris-saclay.fr)

<sup>1</sup> Université Paris-Saclay, CEA, CNRS, Inserm, BioMaps, Gustave Roussy, 114 Rue Edouard Vaillant, 94805 Villejuif, France

<sup>2</sup> Département de Radiologie, Gustave Roussy, 94805 Villejuif, France

<sup>3</sup> Institut Curie, Inserm, Université Paris-Saclay, CNRS, 91405 Orsay, France

Accepted at Magnetic Resonance Materials in Physics, Biology and Medicine: May 2021

<https://doi.org/10.1007/s10334-021-00930-3>

## **Corresponding author**

Charly Girot, Gustave Roussy, BioMaps, PR2, 114 Rue Edouard Vaillant, 94805, Villejuif, France

[Charly.girot@universite-paris-saclay.fr](mailto:Charly.girot@universite-paris-saclay.fr)

## **Abstract**

**Object:** To develop new imaging biomarkers of therapeutic efficacy through the quantification of intratumoral microvascular heterogeneity.

**Materials and Methods:** The described method was a combination of non-supervised clustering and extraction of intratumoral complexity features (ICF): number of non-connected objects, volume fraction. It was applied to a set of 3D DCE-MRI  $K^{\text{trans}}$  maps acquired previously on tumor bearing mice prior to and on day 4 of anti-angiogenic treatment. Evolutions of ICF were compared to conventional summary statistics (CSS) and to heterogeneity related whole tumor texture features (TF) on treated (n=9) and control (n=6) mice.

**Results:** Computed optimal number of clusters per tumor was 4. Several intratumoral features extracted from the clusters were able to monitor a therapy effect. Whereas no feature significantly changed for the control group, 6 features significantly changed for the treated group (4 ICF, 2 CSS). Among these, 5 also significantly differentiated the two groups. (3 ICF, 2 CSS). TF failed in demonstrating differences within and between the two groups.

**Discussion:** ICF are potential imaging biomarkers for anti-angiogenic therapy assessment. The presented method is expected to have advantages with respect to texture analysis based methods regarding interpretability of results and setup of standardized image analysis protocols.

## **KEYWORDS**

Tumor biomarker, Cluster Analysis, Magnetic Resonance Imaging, Angiogenesis, Treatment outcome

## **1. INTRODUCTION**

Tumor heterogeneity has been identified as an important factor in need of evaluation for clinical diagnosis, prognosis, prediction of treatment efficacy and relapse of the disease. It is well known that tumor heterogeneity can be found at different levels: at population level, between patients with the same tumor type not responding equally to the treatment, which has triggered personalized therapeutic strategies over the last years; at patient level, between different lesions, or between primary tumor and metastasis that may have different characteristics [1]; at tumor level, due to the presence of cell sub-populations with different genotypes and phenotypes that may differ in sensitivity to treatment [2], as well as due to microvascular heterogeneity [3], this latter aspect being addressed in the present article.

Heterogeneity of tumor microvasculature is promoted by the neoangiogenic process which initializes and supports the creation of new blood vessels with abnormal structure and function, possibly leading to local tumor-cell hypoxia which in turn supports malignancy [4]. Anti-angiogenic treatments modify blood vessel morphology, network, and function, therefore impacting tumor vascular distribution, which is related to treatment outcome. In a previous study, our team demonstrated that the evaluation of tumor vasculature with dynamic contrast-enhanced ultrasonography (DCE-US) provided a reliable imaging biomarker of anti-angiogenic treatments efficacy [5].

Non-invasive imaging is indeed a technique of choice to evaluate macroscopic aspects of intratumoral heterogeneity [6]. It enables monitoring of the entire tumor and may guide or even replace invasive biopsies when using validated imaging biomarkers. Today, one of the radiological challenges is to identify and qualify new potential biomarkers based on vascular heterogeneity. These new biomarkers would reinforce the reliability of those currently used in clinical practice, mainly anatomic, and possibly provide pertinent predictive information earlier after therapy onset [7].

Two basic approaches are currently used to quantify intratumoral heterogeneity as an imaging biomarker. These two approaches consist in the extraction of features, on the one hand, from the whole tumor volume and, on the other hand, from various tumor sub-regions.

When extracted from the whole tumor volume, features describing image heterogeneity can be classified into two categories: Ia) The spatial gray-level methods (SGLM), which quantify spatial voxel distribution patterns based on intensity values, and Ib) Fractal features and geometry properties, which provide information on the ability of a shape to spread and occupy a defined surface or volume and inform on its complexity. For example, as a tool to depict microvasculature heterogeneity, category Ia) was used to compute texture features (TF) on dynamic contrast-enhanced MRI (DCE-MRI) parametric maps by Thibault et al., who showed these features to be potential biomarkers for early prediction in breast cancer [8], and by Alic et al. for monitoring treatment response in limb sarcomas [9]. Category Ib) was used by Rose et al. who employed the Rényi fractal-dimensions and geometric properties to describe the complexity of DCE-MRI parametric maps. They compared these features with conventional summary statistics (CSS) and demonstrated they were providing new additional information to discriminate between low and high-grade gliomas [3].

Concerning the sub-regions approach, tumor are subdivided, in a supervised or non-supervised manner, into multiple regions of interest (ROI) also called clusters. Features extracted from these clusters can be classified into two categories: IIa) CSS features for histogram description and IIb) features describing image organization, like the clusters' voxel count or volume fraction. As examples for category IIa), Pitre-Champagnat et al. matched histological measures of the mean vessel density within multiple ROI to the mean area under the curve extracted from DCE-US derived parametric maps [10], and Longo et al. evaluated CSS from the pharmacokinetic model parameters  $K^{trans}$  and  $vp$  in multiple clusters to assess tumor response to therapy [11]. As an example for category IIb), Andersen et al. associated the volume fraction of clusters from  $K^{trans}$  and  $ve$  maps with local tumor control [12].

Here we present a new method for quantification of tumor microvasculature heterogeneity. This method is a combination of the previous approaches and is based on the analysis of tumor sub-regions with a specific focus on

the quantification of the spatial organization complexity of these sub-regions, quantified by features we called intratumoral complexity features (ICF). Our method was applied to DCE-MRI derived 3D  $K^{\text{trans}}$  parametric maps acquired on experimental tumors in mice during longitudinal follow-up with or without anti-angiogenic therapy. These parametric maps depict tissue perfusion and capillary permeability, and their heterogeneity is therefore related to microvascular heterogeneity. The dataset, considered as a test dataset for our method, had been previously acquired during a separate methodological study [13]. More specifically, the aim of the present study was threefold: I) to develop a new method for heterogeneity quantification based on ICF providing potential imaging biomarkers of therapeutic efficacy, II) to investigate if this method, applied to a set of DCE-MRI 3D  $K^{\text{trans}}$  parametric maps, can differentiate the treated from the control group, III) to compare ICF with CSS as well as with SGLM based heterogeneity features extracted from the whole tumor.

## 2. MATERIALS AND METHODS

Our method is based on the combination of two successive image processing techniques, clustering and ICF extraction, and its schematic overview is given in Fig.1. First (2.1), the clustering process was applied to parametric maps to highlight voxel clusters with similar  $K^{\text{trans}}$  values, thus corresponding to tissues with similar perfusion and capillary permeability. Second (2.2), features were extracted from each cluster after binarization to quantify its complexity and thereby the spatial heterogeneity of corresponding tumor sub-regions. These features were independent of the voxels'  $K^{\text{trans}}$  values. The method was retrospectively applied to data originating from a previous study [13]. These data correspond to  $K^{\text{trans}}$  parametric maps of manually segmented entire tumor volumes, hereafter termed ROI. Main data characteristics and the MRI acquisition technique (described previously [13, 14]) are summarized in section 2.3.

### 2.1. Image Clustering

The K-means algorithm [15] was used for the clustering process. K-means randomly selects  $k$  voxels as cluster centers, then assigns each voxel of the ROI to the cluster with lowest quadratic distance between voxel and cluster center values. After assigning all voxels within the ROI to a cluster, each cluster center is redefined as the mean value of all voxels belonging to the cluster. The process is repeated until convergence, which is reached when the center of each cluster does not change from one iteration to the next. Prior to the clustering process, the number of clusters  $k$ , which is the hyper-parameter for K-means, was optimized with the procedure mentioned in [12] and described in [16]. The optimization compares the mean intra-cluster distance (MICD) and the inter-cluster minimum distance (ICMD) after the clustering process. MICD and ICMD respectively decreases and increases when increasing the number of clusters. The optimal number of clusters is the one associated to the minimum value of the sum of MICD and ICMD normalized to their respective maximum value. This optimization process allows to select the number of clusters with the best compromise between a low intra-cluster variance and a large inter-cluster distance. It was run 30 times for each number of clusters from  $k=2$  to 8 on every  $K^{\text{trans}}$  map, and the final number of clusters was the one with the highest appearance rate among all maps. Finally, each parametric map was submitted 30 times to the clustering process, to avoid falling into local minima, and the resulting cluster centers were those with the maximum appearance rate among the 30 computations. The clustering process was performed in MATLAB (The MathWorks Inc., Natick, MA, USA).

### 2.2. Features extraction and selection

To describe for each cluster the complexity of spatial organization, heterogeneity related ICF were computed. ICF were selected based on the definition of spatial heterogeneity of a categorical map by Li and Reynolds, stating that heterogeneity can be defined as the complexity of the patches (i.e. objects) composing the categorical map, defined by their configuration and composition [17]. Each of the  $k$  clusters was converted into a binary image (categorical map with two levels) and two ICF were computed to describe its complexity: the number of objects, computed with a 6-connected component analysis (corresponding to the voxel's faces), and the volume fraction of the cluster with respect to the volume of the ROI. The number of objects describes the spatial heterogeneity of a given cluster.

A high number of objects corresponds to a cluster split into several parts within the ROI, whereas a low number of objects indicates that the cluster is spatially gathered, thus representing low spatial heterogeneity. The volume fraction defines the contribution of the cluster's related heterogeneity to the overall degree of tumor heterogeneity. Four cases can be distinguished and are summarized in Table. 1. All in all, a total of  $k$  clusters  $\times$  2 features was extracted using a MATLAB algorithm.

These ICF were compared to TF, i.e. second order statistical features, evaluated over the whole 3D tumor ROI. Amongst available features, our attention was focused on those being specifically related to spatial heterogeneity. The following 6 TF were extracted from the gray-level co-occurrence matrix (GLCM): Homogeneity, Energy, Contrast, Correlation, Entropy, Dissimilarity. LIFEx open access software was used for this analysis, and feature extraction process and mathematical definitions are described in more detail in the software documentation (<http://www.lifex.soft.org> , Orsay, France, [18]). Feature extraction was performed with the following computational parameter settings: number of gray levels 256, absolute intensity rescaling with a maximum value of  $K^{\text{trans}} = 0.8 \text{ min}^{-1}$  corresponding to the overall maximum  $K^{\text{trans}}$  value amongst all available parametric maps.

Moreover, CSS were used to describe the distributions of the  $K^{\text{trans}}$  values both for 3D tumor ROI and within each cluster. Mean and median values, as well as the volume, were computed over the 3D tumor ROI and over the clusters. Mean and median highlights global differences in perfusion and permeability distribution patterns. The whole tumor volume was used as a common biomarker for treatment.

Finally, correlation between all individual features was assessed. They were considered to be correlated for Pearson correlation coefficient  $\geq 0.95$ . In each group of correlated features only one was kept for further analysis.

### 2.3. Dataset properties

Investigated data originated from a previous study [13]. Main properties of the dataset and the data acquisition protocol are summarized hereafter. Data had been acquired on 15 male swiss nude mice carrying a human colorectal tumor xenograft (TC302, Institut Curie, France) subcutaneously implanted at the abdominal level. Animals had been treated for 10 days by single oral administration of Sunitinib® (50 mg/kg, n=9) or vehicle only solution (n=6). Mice underwent DCE-MRI immediately prior to (day 1) and at day 4 after therapy initiation. DCE-MRI experiments had been carried out on a 4.7 T small animal MRI system (Bruker, Biospin, Germany) using the previously described radial 3D DCE-MRI technique RETIA [14], with a home built quadrature birdcage probe (D=35 mm, l=50 mm). Animals had been anesthetized with isoflurane (AErrane®, Baxter, France) and the body temperature had been stabilized. For localization purpose, pre-contrast high-resolution images had been acquired with the same 3D FOV as for radial DCE-MRI covering the whole tumor (respiratory triggered multi-slice spin echo sequence TR~1s, TE = 10.7 ms, matrix =  $128 \times 128$ , slice thickness = 0.47 mm, 64 slices). An experimental medium molecular weight contrast agent P846 (3.5 kDa, Guerbet, France) had been injected at a dose of 0.089 mmol/kg (injection rate 600  $\mu\text{L}/\text{min}$ , injected volume 200  $\mu\text{L}$ ). During DCE-MRI acquisition, 15 3D datasets (tumor) and 960 2D datasets (Arterial Input Function (AIF) measured on a single slice positioned on the heart) had been acquired continuously and simultaneously including one 3D pre-contrast image (64 2D pre contrast images). Main 3D parameters were TE1 = 0.9 ms,  $\Delta\text{TE} = 1.11$  ms, 10 echoes, matrix =  $64 \times 64 \times 64$ , FOV = 30 mm  $\times$  30 mm  $\times$  30 mm, 64 readout points, 4096 projections, temporal resolution = 2 min. Main 2D parameters were TE1 = 1.9 ms,  $\Delta\text{TE} = 1.11$  ms, 10 echoes, matrix =  $64 \times 64$ , FOV = 30 mm  $\times$  30 mm, slice thickness = 2 mm, 64 readout points, 64 projections, temporal resolution = 2 s, TR = 31 ms. Images had been reconstructed using a C++ standard regridding algorithm [14, 19]. Sliding window reconstruction had been performed for 3D data providing a virtual temporal resolution of 30 seconds. CA concentrations in voxels (tumor) or AIF-ROI (heart) had been estimated from  $R2^*$  corrected  $R1(t)$  using  $r1 = 15 \text{ s}^{-1}\text{mM}^{-1}$ . AIF time constants had been measured by fitting a biexponential decay to the concentration time-curve.  $K^{\text{trans}}$  values had been computed voxel wise using the Tofts-Kermode pharmacokinetic model [20]. The process of reconstructing  $K^{\text{trans}}$  maps produced not-a-number (NaN) voxels. NaN voxels were generated when fitting results did not fall in a physical range or in the case of fit failures due to poor signal to noise ratio and/or lack of  $R1$  increase [14]. The mean fraction of NaN voxels was  $23.8 \pm 13.2 \%$ . They were excluded from further analysis. The entire tumor ROI were manually segmented on the high-resolution spin

echo images and then transferred to the  $K^{\text{trans}}$  maps. Previous CSS analyses highlighted a significant decrease of median  $K^{\text{trans}}$  between day 1 and day 4 in the treated group [13].

## 2.4. Statistical Analysis

Selected features were analyzed individually to characterize their behavior regarding three comparisons: (i) evolution within the control group between the examinations at day 1 and day 4; (ii) evolution within the treated group between the examinations pre- and post-treatment; (iii) evolution differences between the two groups. For (i) and (ii), the feature values pre- (day 1) and post-treatment (day 4) were analyzed using a Wilcoxon paired test. For (iii), feature differences were computed between post- and pre-treatment, then analyzed using a Wilcoxon unpaired test. Resulting p-values  $\leq 0.05$  were considered demonstrating statistically significant differences. A correction for multiple comparisons was applied to all selected features using the Benjamini-Hochberg procedure. The false discovery rate (FDR) was set to 0.05. Analyses were conducted with R software (R Core Team, 2019).

# 3. RESULTS

## 3.1. Clustering

The optimization process resulted in an optimal number of  $k = 4$  clusters per parametric map. Table 2 represents the normalized sums of normalized MICD and ICMD and shows that 4 clusters were the best compromise between low intra-cluster variance and high inter-cluster distance. Clusters were numbered #1 to #4 in ascending order of mean  $K^{\text{trans}}$  with mean centroids over the whole dataset of 0.8, 2.2, 4.4 and 9.9 ( $10^{-2} \text{ min}^{-1}$ ). Fig.2 shows the clustering result for a single tumor slice. It illustrates how the clusters were used, either as binary images to compute ICF, or as masks for selecting sub-populations of voxels and extract CSS.

## 3.2. Visual inspection of heterogeneity evolution

Visual inspection of 2D maps allowed qualitative evaluation of heterogeneity evolution in the images. As illustrated in Fig.3 on representative slices, tumor  $K^{\text{trans}}$  heterogeneity evolves over the 4 days follow-up period. While heterogeneity appears to decrease in the treated mouse, with an evolution from clusters with intermediate (#2 and #3) towards low  $K^{\text{trans}}$  values (#1), this is not the case for the control mouse, for which cluster organization complexity increases. Our method intended to quantify these spatial heterogeneity modifications within the clusters through ICF extraction.

## 3.3. Feature extraction and selection

A total of 29 features were computed from each parametric map: 8 ICF were computed with our method after the clustering process; 15 CSS were extracted from the clusters (4 clusters  $\times$  3 features) and from the whole tumor ROI (3 features); 6 GLCM were extracted from the whole tumor ROI.

Feature correlations were evaluated to define a limited number of independent features. None of the ICF were correlated with each other. Means and medians were correlated for each cluster and for the ROI, therefore, only the medians were kept, except the median of cluster #1 which was correlated to the ROI median. Amongst GLCM features, contrast and entropy were respectively correlated to dissimilarity and homogeneity and therefore not selected.

As a result, a total of 21 features were kept for further analysis: 8 ICF (number of objects and volume fraction of each cluster), 4 GLCM features (homogeneity, energy, correlation, and dissimilarity), and 9 CSS (median and volume of each cluster and of the ROI – except median of cluster #1). We checked that these 21 features were not correlated with each other.

## 3.4. Statistical Analysis

Wilcoxon tests were performed individually on the 21 features and results with p-value  $\leq 0.1$  are summarized in Tables 3 to 5. Tables 3 and 4 report mean and standard error values respectively for the control group at day 1 and day 4 and for the treated group for pre- and post-treatment examinations. Statistical significance of the difference between day 1 and day 4 values as well as the adjusted p-values (q-values) after Benjamini-Hochberg correction are also reported. Table 5 reports the delta features (post – pre-treatment) for control and treated group and statistical analysis between these groups. In Tables 3 to 5, features that remained significant after correction with FDR=0.05, are labelled with “\*”. Feature values were similar for the control and the treated group before treatment initiation with no significant differences. Also, within the control group, no feature demonstrated significant differences between day 1 and day 4 (Table 3).

Amongst the ICF evaluated at cluster-level, after correction for multiple comparisons, 4 demonstrated significant differences between pre- and post-treatment evaluations within the treated group (Table 4): number of objects in cluster #3 (p=0.0039) and volume fractions in clusters #1 (p=0.0039), #2 (p=0.0039), #3 (p=0.0117). 3 demonstrated significant differences in their delta-values between the two groups (Table 5): delta-number of objects in cluster #3 (p=0.0016), and delta-volume fractions in clusters #1 (p=0.0048), #2 (p=0.0008). Amongst the 4 GLCM features, after correction for multiple comparisons none presented significant differences within the treated group, neither between both groups. Amongst the 9 CSS, after correction for multiple comparisons, 3 demonstrated significant differences within the treated group (Table 4): volumes in clusters #2 (p=0.0039), #3 (p=0.0039) and median in ROI (p=0.0078). 3 demonstrated significant differences between the two groups (Table 5): delta volumes in clusters #2 (p=0.0004), #3 (p=0.012) and delta volume in ROI (p=0.0028). Features from Table 5 that were statistically discriminant after correction between both groups are represented as boxplot in Fig.4.

## 4. DISCUSSION

In oncology, a growing interest for targeted therapies and particularly for anti-angiogenic drugs led to a growing number of studies assessing tumor microvasculature and vascular heterogeneity due to the ability of these therapies to normalize and reorganize microvascular structure. Nonetheless, in clinical practice, tumors are still assessed using morphological criteria without considering the modification of heterogeneity patterns which may arise earlier during anti-angiogenic treatments. Some studies tried to exploit tumor heterogeneity using texture analysis to assess therapeutic response [21–23], but none of these approaches succeeded yet to reach use in clinical practice, mainly because of a lack of standardization and guidelines [24].

In this study, we developed a new method for assessing intratumoral heterogeneity, and we described a first application to the longitudinal follow-up of microvascular heterogeneity on an experimental tumor model. The method associates a clustering process and the extraction of features, we called ICF, to identify and quantify different levels of complexity of intratumoral heterogeneity. The development of this method was guided by the following strategy: First, processed images should be parametric maps. This allows to directly interpret heterogeneity of the parametric maps in terms of physiological heterogeneity. Second, a non-supervised clustering process was applied to these parametric maps. This procedure allows to subdivide the tumor into several functional sub-regions characterized by the measured physiological parameter levels. It has indeed been suggested that, in some clinical scenarios, measurements performed in tumor sub-regions may provide more useful biomarkers than measurements averaged over the whole tumor [25]. Third, the features used to quantify heterogeneity should provide a more intuitive spatial representation (qualitatively cross-checkable by visual inspection of 2D images) than second order statistical analysis approaches in texture analysis.

According to this strategy, the method was applied to a set of 3D DCE-MRI  $K^{\text{trans}}$  parametric maps, previously acquired in mice during a longitudinal study. Heterogeneity related features extracted from DCE-MRI parametric

maps are of interest, as they should monitor treatment effects by assessing microvasculature modifications. Information from these features is expected to be particularly useful in the absence of morphological changes. In this study, heterogeneity was assessed by ICF. They were compared to GLCM features describing heterogeneity over the whole tumor ROI and to classical non-spatial CSS features.

The ability of ICF, GLCM, and CSS features to monitor therapeutic efficacy was evaluated by the following comparisons: i) day 1 vs day 4 within the control group, ii) pre- (day 1) vs post-treatment (day 4) within the treated group, and iii) evolution differences for the control vs the treated group. It was checked that all features had similar values in the treated and control groups at day 1 confirming similar initial status of both groups.

Statistical analysis showed that several intratumoral features extracted from the clusters were able to monitor a therapy effect. Whereas no feature significantly changed for the control group (comparison (i)), 6 features significantly changed for the treated group (4 ICF, 2 CSS - comparison (ii)). Among these, 5 also significantly differentiated the two groups. (3 ICF, 2 CSS – comparison (iii)).

On the contrary, heterogeneity related GLCM features extracted from the whole tumor ROI didn't change significantly for either comparison. Therefore, the presented method seems to be more sensitive than classical texture analysis for monitoring therapy effects under our experimental conditions.

Median  $K^{trans}$  measured over the whole tumor ROI decreased for the treated group (comparison (ii)), as previously described [13]. This result is compatible with feature evolution at the cluster level: an increase of the volume fraction in cluster #1, representing the lowest  $K^{trans}$  values, associated to a decrease of the volume fractions in clusters #2 and #3, representing intermediate  $K^{trans}$  values. These findings indicate a shift of tumor microvasculature characteristics to lower perfusion and capillary permeability, as may be expected under anti-angiogenic therapy.

It has to be noted that, even though the tumor volume tends to decrease for the treated group at day 4 of treatment, this change is not significant in our study. Whole tumor volume is generally used as a reference biomarker in oncology. However, in clinical practice during anti-angiogenic treatment, morphological evaluation of tumors based on the RECIST criterion has been shown to significantly underestimate tumor response to treatment [26]. For example, some studies showed that no tumor size modification was observed during the course of treatment for respectively 50 to 75% of patients [27, 28]. This is mainly because vascular modifications often occur before any size reduction of the tumor. Our preclinical results are in line with these clinical findings.

As mentioned above, concerning heterogeneity description, in our study only ICF were demonstrating significant differences for comparisons (ii) and (iii) after BH correction. More precisely, 2 GLCM features with  $p < 0.05$  (Energy, comparison (ii); Correlation, comparison (iii)) were discarded after BH correction. In our study BH correction was applied with  $FDR = 0.05$  in order to discard false positives in the case of multi-testing. Nevertheless, the use of a correction for multi-testing is still a topic of discussion as some authors advise not to use any correction when a study is exploratory with unplanned comparisons being considered as potential ones for further investigation [29]. Further studies will be necessary to compare (or to combine) ICF and GLCM features as potential imaging biomarkers related to tumor microvascular heterogeneity.

Evolution of ICF can be relatively easily interpreted in terms of microvascular properties. Indeed, for comparison (ii) within the treated group, observed ICF evolutions can be considered indicative of  $K^{trans}$  map homogenization: Volume fraction for cluster #1, which had the smallest and also constant number of objects amongst clusters #1 to #3, increased by 37%, and the number of objects in cluster #3 was reduced by 43 %. Volume fraction of cluster #2, the most heterogeneous one, decreased by 38% ( $100 - 100 * 22.5 / 35.9$ ) with a stable number of objects. Thus, by considering clusters #1 and #3, a total of 76.1% of the analyzed tumor volume can be considered more homogenous in terms of perfusion and capillary permeability with lower  $K^{trans}$  at day 4 post-treatment, which, again, may be expected under anti-angiogenic therapy.



Furthermore, it is noteworthy that, for the control group, while no ICF demonstrated a significant evolution between day 1 and day 4, numbers of objects in all clusters tended to increase with volume fractions staying nearly identical. This observation is compatible with an overall increase of microvascular heterogeneity, which may be expected in growing tumors with neo-angiogenic activity.

Results also indicated that none of the significant features were found within the cluster #4 with highest  $K^{\text{trans}}$  values, which may be considered to be "hotspots". This is in contrast with previous studies which focused on such "hotspots" considering they were bearing more information than values averaged over the whole tumor. For example, a greater range of  $K^{\text{trans}}$  changes was identified by Hayes et al. in these "hotspots" when compared to the tumor ROI [30]. The difference with respect to our results may be explained by the definition of the "hotspots". While they were defined as the 75<sup>th</sup> percentile of the tumor histogram by Hayes et al., our non-supervised clustering approach resulted in "hotspot" clusters with few voxels and volume fractions around only 2%.

From a technical point of view, the main characteristics and advantages of the presented approach appear to be the following:

Our approach quantifies heterogeneity by the description of tumor sub-regions complexity, related, in our study, to  $K^{\text{trans}}$  map heterogeneity. This is in contrast with previous studies who assessed either the complexity of the tumor as a whole [3], or only the histogram distribution in tumor sub-regions without considering spatial distribution patterns [11].

Moreover, a potential specific advantage comes from the use of ICF extracted from binary images. As they do not depend on absolute voxel values after the non-supervised clustering process, they should be less sensitive to acquisition related variations or bias in parameter quantitation (e.g. due to poor AIF quality or pre-contrast T1 mapping in the case of DCE-MRI). This should be of advantage for the setup of multicentric studies where measurement results need to be independent of the MRI scanner and protocol. Furthermore, in this context, it is noteworthy that the only user defined computational parameter for the proposed method is the "connectivity" parameter set during computation of object numbers within the clusters. We checked, that, when varying connectivity from 6 (all voxel faces), as used in the study, to 18 (6 faces and 12 edges), results still discriminated both groups of mice with the same significant features. However, this was no longer the case for the maximal 3D connectivity of 26 (6 faces, 12 edges and 8 vertices). Indeed, in this latter case, all voxels are merged into few objects thus no longer capturing image heterogeneity. Consequently, this connectivity should be avoided using our method.

On the contrary, in the case of texture analysis, TF depend on several user-defined computational parameters and settings: number of bins for discretization, size of bins, absolute (user defined min, max values) or relative (ROI defined min, max values) intensity rescaling. These computational parameters and settings may also be handled differently by different software, making standardization a more complicated task [31].

Consequently, when analyzing data acquired with different MRI systems and protocols, our approach may be expected to be more robust in comparison to texture analysis. This aspect will be addressed in future studies.

In conclusion, in this preclinical feasibility study, we presented a new method for image heterogeneity analysis, and we showed that, when applied to DCE-MRI derived  $K^{\text{trans}}$  parametric maps, this method may provide potential imaging biomarkers for anti-angiogenic therapy assessment in agreement with the consensus on Imaging Biomarkers [7]. The method is expected to have advantages with respect to texture analysis based methods regarding interpretability of results and setup of standardized image analysis protocols. Further studies applying this method to different parametric maps across multiple imaging modalities will be needed to evaluate its usefulness in a clinical setting for treatment monitoring or, more generally, tissue characterization.

## Acknowledgements

For MRI data acquisition the tumor model had been provided by Institut Curie (Preclinical Investigation Laboratory, LIP); Sunitinib® had been provided by Pfizer; MRI contrast agent had been provided by Guerbet.

## Authors' Contributions

Charly Girot contributed to study conception and design, analysis and interpretation of data, drafting of manuscript and critical revisions. Andreas Volk contributed to data acquisition, analysis and interpretation of data and critical revision. Stéphanie Pitre-Champagnat contributed to analysis and interpretation of data and critical revision. Christine Walczak contributed to acquisition and analysis of data. Nathalie Lassau contributed to critical revision.

## Compliance with Ethical Standards

Funding: the authors did not receive support from any organization for the submitted work.

Conflict of Interest: all authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in the manuscript.

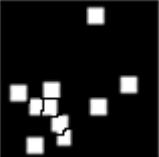



Ethical approval: Care and use of animals had been in compliance with national and institutional regulations and guidelines

## References

1. Marusyk A, Polyak K (2010) Tumor heterogeneity: Causes and consequences. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1805:105–117.

2. Melo FDSE, Vermeulen L, Fessler E, Medema JP (2013) Cancer heterogeneity—a multifaceted view. *EMBO Rep* 14:686–695.
3. Rose CJ, Mills SJ, O'Connor JPB, Buonaccorsi GA, Roberts C, Watson Y, Cheung S, Zhao S, Whitcher B, Jackson A, Parker GJM (2009) Quantifying spatial heterogeneity in dynamic contrast-enhanced MRI parameter maps. *Magn Reson Med* 62:488–499.
4. Jain RK (2014) Antiangiogenesis Strategies Revisited: From Starving Tumors to Alleviating Hypoxia. *Cancer Cell* 26:605–622.
5. Lassau N, Bonastre J, Kind M, Vilgrain V, Lacroix J, Cuinet M, Taieb S, Aziza R, Sarran A, Labbe-Devilliers C, Gallix B, Lucidarme O, Ptak Y, Rocher L, Caquot L-M, Chagnon S, Marion D, Luciani A, Feutray S, Uzan-Augui J, Coiffier B, Benastou B, Koscielny S (2014) Validation of Dynamic Contrast-Enhanced Ultrasound in Predicting Outcomes of Antiangiogenic Therapy for Solid Tumors: The French Multicenter Support for Innovative and Expensive Techniques Study. *Investigative Radiology* 49:794–800.
6. O'Connor JPB (2017) Cancer heterogeneity and imaging. *Seminars in Cell & Developmental Biology* 64:48–57.
7. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G, Buckley DL, Chenevert TL, Clarke LP, Collette S, Cook GJ, deSouza NM, Dickson JC, Dive C, Evelhoch JL, Faivre-Finn C, Gallagher FA, Gilbert FJ, Gillies RJ, Goh V, Griffiths JR, Groves AM, Halligan S, Harris AL, Hawkes DJ, Hoekstra OS, Huang EP, Hutton BF, Jackson EF, Jayson GC, Jones A, Koh D-M, Lacombe D, Lambin P, Lassau N, Leach MO, Lee T-Y, Leen EL, Lewis JS, Liu Y, Lythgoe MF, Manoharan P, Maxwell RJ, Miles KA, Morgan B, Morris S, Ng T, Padhani AR, Parker GJM, Partridge M, Pathak AP, Peet AC, Punwani S, Reynolds AR, Robinson SP, Shankar LK, Sharma RA, Soloviev D, Stroobants S, Sullivan DC, Taylor SA, Tofts PS, Tozer GM, van Herk M, Walker-Samuel S, Wason J, Williams KJ, Workman P, Yankeelov TE, Brindle KM, McShane LM, Jackson A, Waterton JC (2017) Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 14:169–186.
8. Thibault G, Tudorica A, Afzal A, Oh KY, Naik A, Troxell ML, Kemmer kathleen A, Roy N, Jafarian N, Holtorf ML, Huang W, Song X (2017) DCE-MRI Texture Features for Early Prediction of Breast Cancer Therapy Response. *Tomography* 3:23–32.
9. Alic L, van Vliet M, van Dijke CF, Eggermont AMM, Veenland JF, Niessen WJ (2011) Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response? *Phys Med Biol* 56:1601–1616.
10. Pitre-Champagnat S, Leguerney I, Bosq J, Peronneau P, Kiessling F, Calmels L, Coulot J, Lassau N (2015) Dynamic Contrast-Enhanced Ultrasound Parametric Maps to Evaluate Intratumoral Vascularization: *Investigative Radiology* 50:212–217.
11. Longo DL, Dastrù W, Consolino L, Espak M, Arigoni M, Cavallo F, Aime S (2015) Cluster analysis of quantitative parametric maps from DCE-MRI: application in evaluating heterogeneity of tumor response to antiangiogenic treatment. *Magnetic Resonance Imaging* 33:725–736.
12. Andersen EKF, Kristensen GB, Lyng H, Malinen E (2011) Pharmacokinetic analysis and k-means clustering of DCEMR images for radiotherapy outcome prediction of advanced cervical cancers. *Acta Oncologica* 50:859–865.
13. El Tannir El Tayara N, Ben Achour N, Walczak C, Nemati F, Mispelter J, Decaudin D, Vautier J, Volk A (2011) Anti-angiogenic therapy follow-up in a mouse tumor model by a novel 3D radial multi-gradient echo DCE MRI technique with individual AIF measurement. *Proc Intl Soc Mag Reson Med* 19:1056.
14. Vautier J, El Tannir El Tayara N, Walczak C, Mispelter J, Volk A (2013) Radial multigradient-echo DCE-MRI for 3D  $K^{trans}$  mapping with individual arterial input function measurement in mouse tumor models: 3D Radial DCE-MRI in Mice with Individual AIF. *Magn Reson Med* 70:823–828.
15. Macqueen J (1967) Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1:281–297.

16. Rokach L, Maimon O (2005) Clustering Methods. In: Maimon O, Rokach L (eds) *Data Mining and Knowledge Discovery Handbook*. Springer-Verlag, New York, pp 321–352
17. Li H, Reynolds JF (1995) On Definition and Quantification of Heterogeneity. *Oikos* 73:280.
18. Nioche C, Orhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, Pellot-Barakat C, Soussan M, Frouin F, Buvat I (2018) LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res* 78:4786–4789.
19. Jackson JI, Meyer CH, Nishimura DG, Macovski A (1991) Selection of a convolution function for Fourier inversion using gridding (computerised tomography application). *IEEE Trans Med Imaging* 10:473–478.
20. Tofts PS, Kermode AG (1991) Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magn Reson Med* 17:357–367.
21. Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, Followill D, Jones AK, Stingo F, Liao Z, Mohan R, Court L (2017) Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Sci Rep* 7:588.
22. Liu J, Mao Y, Li Z, Zhang D, Zhang Z, Hao S, Li B (2016) Use of texture analysis based on contrast-enhanced MRI to predict treatment response to chemoradiotherapy in nasopharyngeal carcinoma: MRI in Treatment Response Prediction. *J Magn Reson Imaging* 44:445–455.
23. Panzeri MM, Losio C, Della Corte A, Venturini E, Ambrosi A, Panizza P, De Cobelli F (2018) Prediction of Chemoresistance in Women Undergoing Neo-Adjuvant Chemotherapy for Locally Advanced Breast Cancer: Volumetric Analysis of First-Order Textural Features Extracted from Multiparametric MRI. *Contrast Media & Molecular Imaging* 2018:1–7.
24. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762.
25. O'Connor JPB, Rose CJ, Waterton JC, Carano RAD, Parker GJM, Jackson A (2015) Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. *Clinical Cancer Research* 21:249–257.
26. Smith AD, Lieber ML, Shah SN (2010) Assessing Tumor Response and Detecting Recurrence in Metastatic Renal Cell Carcinoma on Targeted Therapy: Importance of Size and Attenuation on Contrast-Enhanced CT. *American Journal of Roentgenology* 194:157–165.
27. Motzer RJ, Tannir NM, McDermott DF, Arén Frontera O, Melichar B, Choueiri TK, Plimack ER, Barthélémy P, Porta C, George S, Powles T, Donskov F, Neiman V, Kollmannsberger CK, Salman P, Gurney H, Hawkins R, Ravaud A, Grimm M-O, Bracarda S, Barrios CH, Tomita Y, Castellano D, Rini BI, Chen AC, Mekan S, McHenry MB, Wind-Rotolo M, Doan J, Sharma P, Hammers HJ, Escudier B (2018) Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *N Engl J Med* 378:1277–1290.
28. Luo Y, Chen J, Huang K, Lin Y, Chen M, Xu L, Li Z-P, Feng S-T (2017) Early evaluation of sunitinib for the treatment of advanced gastroenteropancreatic neuroendocrine neoplasms via CT imaging: RECIST 1.1 or Choi Criteria? *BMC Cancer* 17:154.
29. Armstrong RA (2014) When to use the Bonferroni correction. *Ophthalmic Physiol Opt* 34:502–508.
30. Hayes C, Padhani AR, Leach MO (2002) Assessing changes in tumour vascular function using dynamic contrast-enhanced magnetic resonance imaging. *NMR Biomed* 15:154–163.
31. Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *International Journal of Radiation Oncology\*Biography\*Physics* 102:1143–1158.

		Number of Objects	
		Low	High
Volume Fraction	Low	<p>Voxels are gathered in one or a few areas and thus are <b>homogeneous both by their value and their spatial organization</b>. With a small volume fraction, their <b>contribution to global tumor homogeneity is small</b></p> <p><i>Example: 10 objects, 10% volume fraction</i></p> 	<p><b>Voxel cluster is shattered through the tumor</b> but does not occupy an important fraction of the tumor thus its <b>impact on global tumor heterogeneity is small</b></p> <p><i>Example: 1000 objects, 10% volume fraction</i></p> 
	High	<p>Voxels are gathered in one or a few areas and thus are <b>homogeneous both by their value and their spatial organization</b>. With a large volume fraction, their <b>contribution to global tumor homogeneity is important</b></p> <p><i>Example: 10 objects, 60% volume fraction</i></p> 	<p><b>Voxel cluster is shattered through the tumor</b> and its important occupation of the tumor volume leads to an <b>important contribution to global tumor heterogeneity</b></p> <p><i>Example: 1000 objects, 60% volume fraction</i></p> 

**Table 1: Interpretation of Intratumoral Complexity Features (ICF) combinations to describe tumor heterogeneity along with illustration by corresponding simulated image examples (single cluster)**

Number of clusters	2	3	4	5	6	7	8
MICD + ICMD	1.00	0.81	<b>0.73</b>	0.76	0.81	0.89	1.00

**Table 2: Sum of normalized [0, 1] MICD and ICMD for each number of clusters using K-means algorithm. In bold, the optimal number of clusters**

	<i>Features</i>	<i>p-value</i>	<i>Pre-Treatment Mean ± se</i>	<i>Post-treatment Mean ± se</i>	<i>Adjusted P (q-value)</i>
<b>Sub-regions / clusters</b>	<b>ICF</b>				
	Num. objects #1	0.0938	65 ± 6	83 ± 13	-
	Num. objects #2	0.0625	133 ± 17	175 ± 18	-
	Num. objects #3	-	167 ± 18	182 ± 16	-
	Num. objects #4	-	54 ± 8	60 ± 13	-
	Vol. fraction #1 (%)	-	50.1 ± 2.6	51.8 ± 3.7	-
	Vol. fraction #2 (%)	-	35.9 ± 1.3	34.4 ± 1.7	-
	Vol. fraction #3 (%)	-	12.1 ± 1.8	12.0 ± 1.9	-
	Vol. fraction #4 (%)	-	1.8 ± 0.3	1.8 ± 0.5	-
	<b>CSS</b>				
	Median #2 (min <sup>-1</sup> )	-	0.024 ± 0.003	0.021 ± 0.003	-
	Median #3 (min <sup>-1</sup> )	-	0.045 ± 0.007	0.039 ± 0.004	-
	Median #4 (min <sup>-1</sup> )	-	0.095 ± 0.016	0.08 ± 0.008	-
	Volume #1 (mm <sup>3</sup> )	-	198.8 ± 15.1	245.7 ± 24.3	-
Volume #2 (mm <sup>3</sup> )	-	143.0 ± 10.0	162.5 ± 10.9	-	
Volume #3 (mm <sup>3</sup> )	-	48.3 ± 7.6	56.1 ± 8.8	-	
Volume #4 (mm <sup>3</sup> )	-	7.3 ± 1.4	8.4 ± 2.3	-	
<b>Whole tumor ROI</b>	<b>CSS</b>				
	Median (min <sup>-1</sup> )	-	0.017 ± 0.003	0.015 ± 0.003	-
	Volume (mm <sup>3</sup> )	0.0625	474.5 ± 24.9	585.2 ± 24.2	-
	<b>GLCM</b>				
	Homogeneity	-	0.466 ± 0.032	0.493 ± 0.027	-
	Energy	-	0.017 ± 0.005	0.023 ± 0.005	-
	Correlation	-	0.486 ± 0.015	0.481 ± 0.031	-
Dissimilarity	-	3.14 ± 0.537	2.668 ± 0.396	-	

**Table 3: Features mean and standard-error values for the control group at day 1 and day 4 and p-values of Wilcoxon paired tests comparing the two measurements. P-values greater than 0.1 and q-values for non-significant features are not reported**

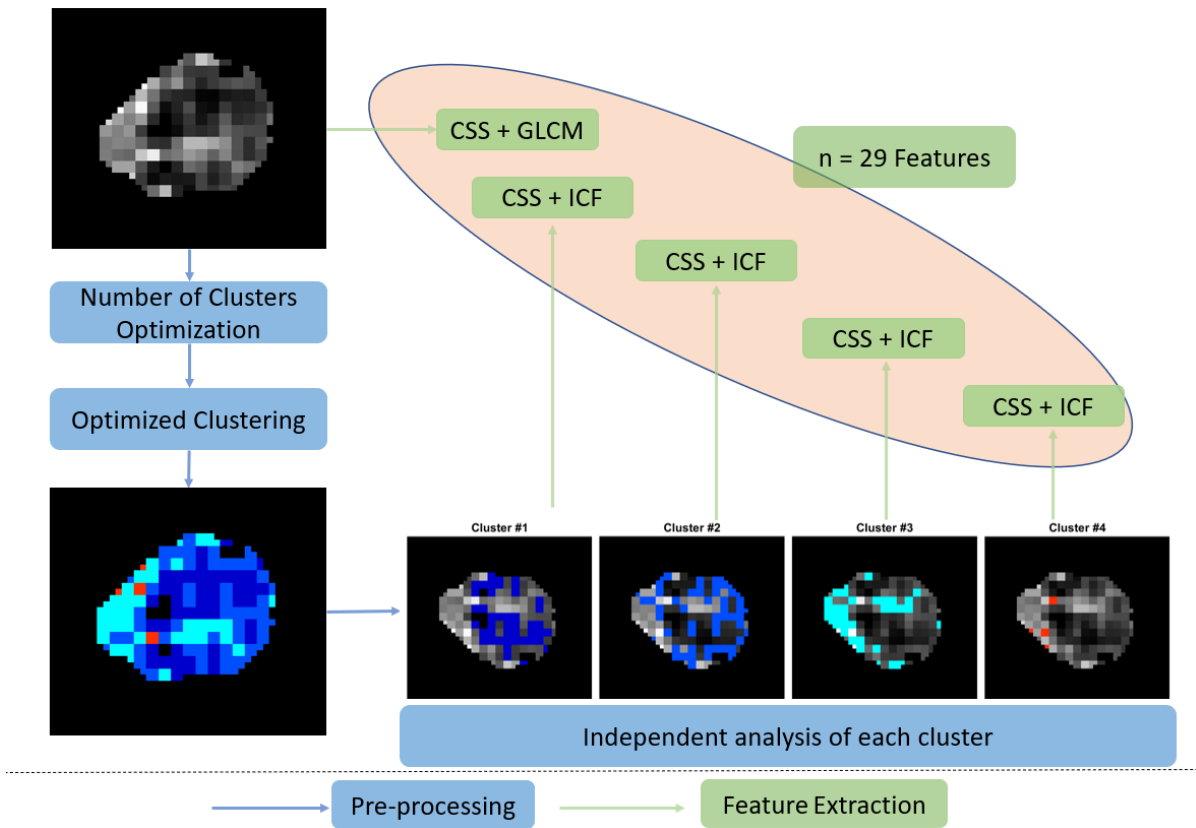
	<b>Features</b>	<b>p-value</b>	<b>Pre-Treatment Mean <math>\pm</math> se</b>	<b>Post-treatment Mean <math>\pm</math> se</b>	<b>Adjusted P (q-value)</b>
Sub-regions / clusters	<b>ICF</b>				
	Num. objects #1	-	74 $\pm$ 18	76 $\pm$ 14	-
	Num. objects #2	-	161 $\pm$ 23	186 $\pm$ 41	-
	Num. objects #3 *	0.0039	174 $\pm$ 16	100 $\pm$ 12	0.082
	Num. objects #4	0.0547	50 $\pm$ 6	31 $\pm$ 4	-
	Vol. fraction #1 * (%)	0.0039	50.4 $\pm$ 1.4	68.8 $\pm$ 2.4	0.041
	Vol. fraction #2 * (%)	0.0039	35.9 $\pm$ 0.8	22.4 $\pm$ 1.6	0.0273
	Vol. fraction #3 * (%)	0.0117	12.1 $\pm$ 0.7	7.3 $\pm$ 1.0	0.0352
	Vol. fraction #4 (%)	-	1.6 $\pm$ 0.3	1.5 $\pm$ 0.3	-
	<b>CSS</b>				
	Median #2 (min <sup>-1</sup> )	-	0.02 $\pm$ 0.003	0.02 $\pm$ 0.002	-
	Median #3 (min <sup>-1</sup> )	-	0.036 $\pm$ 0.006	0.047 $\pm$ 0.005	-
	Median #4 (min <sup>-1</sup> )	-	0.081 $\pm$ 0.014	0.098 $\pm$ 0.011	-
	Volume #1 (mm <sup>3</sup> )	-	205.9 $\pm$ 27.0	197.8 $\pm$ 32.7	-
	Volume #2 * (mm <sup>3</sup> )	0.0039	144.9 $\pm$ 17.0	64.7 $\pm$ 10.8	0.0205
Volume #3 * (mm <sup>3</sup> )	0.0039	49.0 $\pm$ 6.7	19.8 $\pm$ 2.8	0.0164	
Volume #4 (mm <sup>3</sup> )	0.0547	6.5 $\pm$ 1.4	3.9 $\pm$ 0.5	-	
Whole tumor ROI	<b>CSS</b>				
	Median * (min <sup>-1</sup> )	0.0078	0.014 $\pm$ 0.002	0.008 $\pm$ 0.001	0.0273
	Volume (mm <sup>3</sup> )	0.0547	503.7 $\pm$ 76.8	462.0 $\pm$ 73.9	-
	<b>GLCM</b>				
	Homogeneity	-	0.495 $\pm$ 0.028	0.54 $\pm$ 0.021	-
	Energy	0.0273	0.025 $\pm$ 0.006	0.044 $\pm$ 0.008	-
Correlation	0.0547	0.43 $\pm$ 0.023	0.504 $\pm$ 0.019	-	
Dissimilarity	-	2.709 $\pm$ 0.436	2.643 $\pm$ 0.336	-	

**Table 4: Features mean and standard-error values for the treated group pre-treatment (day 1) and post-treatment (day 4) and p-values of Wilcoxon paired tests comparing the two measurements. P-values greater than 0.1 are not reported. Q-values are reported for features that remained significant after Benjamini-Hochberg correction and these features are labeled with “\*”**

	<i>Features</i>	<i>p-value</i>	<i>Delta Control Mean ± se</i>	<i>Deltra Treated Mean ± se</i>	<i>Adjusted P (q-value)</i>
Sub-regions / clusters	<b>ICF</b>				
	Num. objects #1	-	19 ± 9	2 ± 11	-
	Num. objects #2	-	42 ± 18	25 ± 34	-
	Num. objects #3 *	0.0016	15 ± 19	- 73 ± 11	0.0112
	Num. objects #4	-	6 ± 12	- 19 ± 6	-
	Vol. fraction #1 * (%)	0.0048	1.7 ± 4.0	18.4 ± 2.8	0.0201
	Vol. fraction #2 * (%)	0.0008	- 1.5 ± 1.5	- 13.5 ± 1.7	0.0084
	Vol. fraction #3 (%)	0.0879	- 0.2 ± 2.1	- 4.8 ± 1.2	-
	Vol. fraction #4 (%)	-	0.0 ± 0.6	- 0.1 ± 0.3	-
	<b>CSS</b>				
	Median #2 (min <sup>-1</sup> )	-	0.004 ± 0.004	0.0 ± 0.002	-
	Median #3 (min <sup>-1</sup> )	-	- 0.006 ± 0.009	0.01 ± 0.005	-
	Median #4 (min <sup>-1</sup> )	-	- 0.015 ± 0.018	0.017 ± 0.012	-
	Volume #1 (mm <sup>3</sup> )	-	46.9 ± 28.5	- 8.1 ± 13.6	-
Volume #2 * (mm <sup>3</sup> )	0.0004	19.5 ± 10.8	- 80.1 ± 14.6	0.0084	
Volume #3 * (mm <sup>3</sup> )	0.012	7.8 ± 9.3	- 29.17 ± 6.3	0.042	
Volume #4 (mm <sup>3</sup> )	-	1.0 ± 2.7	- 2.6 ± 1.4	-	
Whole tumor ROI	<b>CSS</b>				
	Median (min <sup>-1</sup> )	-	- 0.003 ± 0.003	- 0.006 ± 0.002	-
	Volume * (mm <sup>3</sup> )	0.0028	110.7 ± 33.8	- 41.7 ± 18.2	0.0147
	<b>GLCM</b>				
	Homogeneity	-	0.026 ± 0.04	0.046 ± 0.025	-
	Energy	-	0.005 ± 0.008	0.02 ± 0.008	-
Correlation	0.0496	- 0.006 ± 0.032	0.077 ± 0.033	-	
Dissimilarity	-	- 0.473 ± 0.659	- 0.066 ± 0.395	-	

**Table 5: Delta-features (post-treatment minus pre-treatment) mean and standard-error values for the control and the treated group and p-values of Wilcoxon unpaired tests between control and treated groups. P-values greater than 0.1 are not reported. Q-values are reported for features that remained significant after Benjamini-Hochberg correction and these features are labeled with “\*\*”**





**Fig. 1: Schematic overview of the method proposed in this study to investigate intratumoral heterogeneity from parametric maps. The method is a combination of clustering and extraction of complexity features (ICF). This method was compared to heterogeneity features from the gray-level co-occurrence matrix (GLCM) and to conventional summary statistics (CSS)**

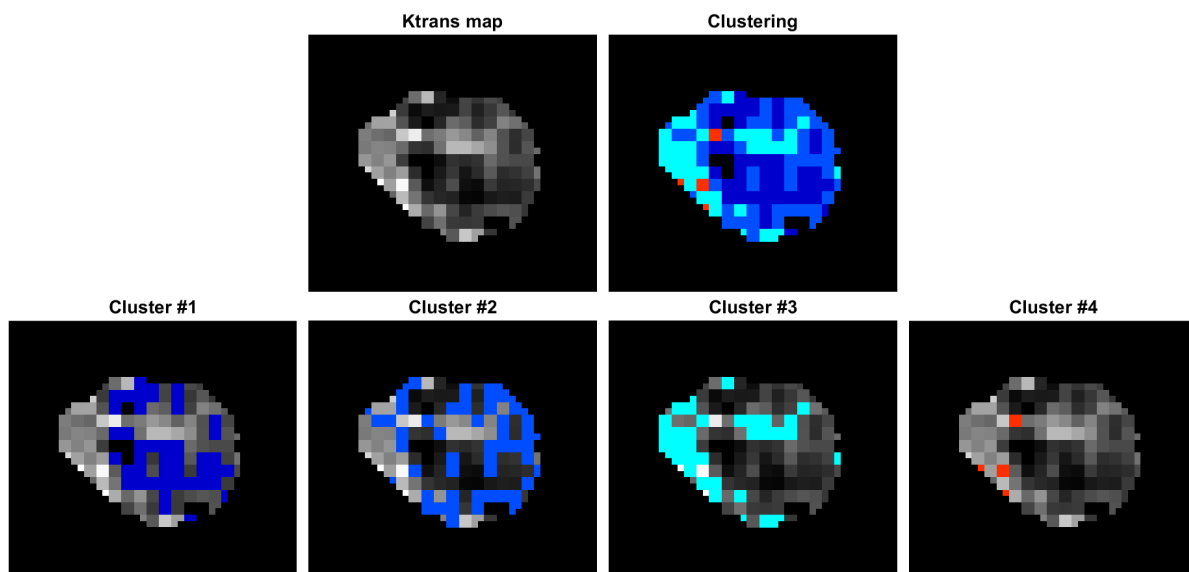


Fig. 2: Single slice illustrating the cluster extraction and masking process – Top row:  $K^{trans}$  map and resulting cluster map; Bottom row: The four clusters in ascending order of mean  $K^{trans}$  values, isolated and superimposed on the  $K^{trans}$  map

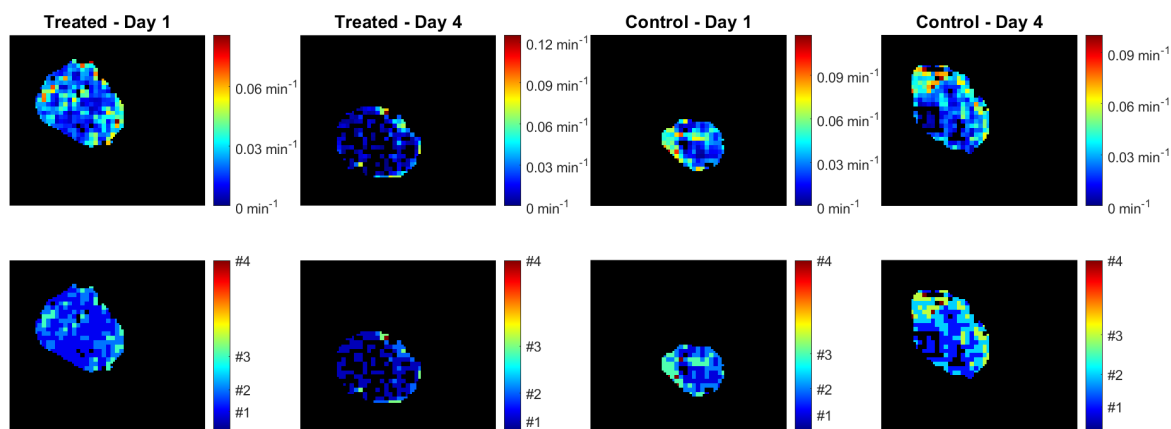
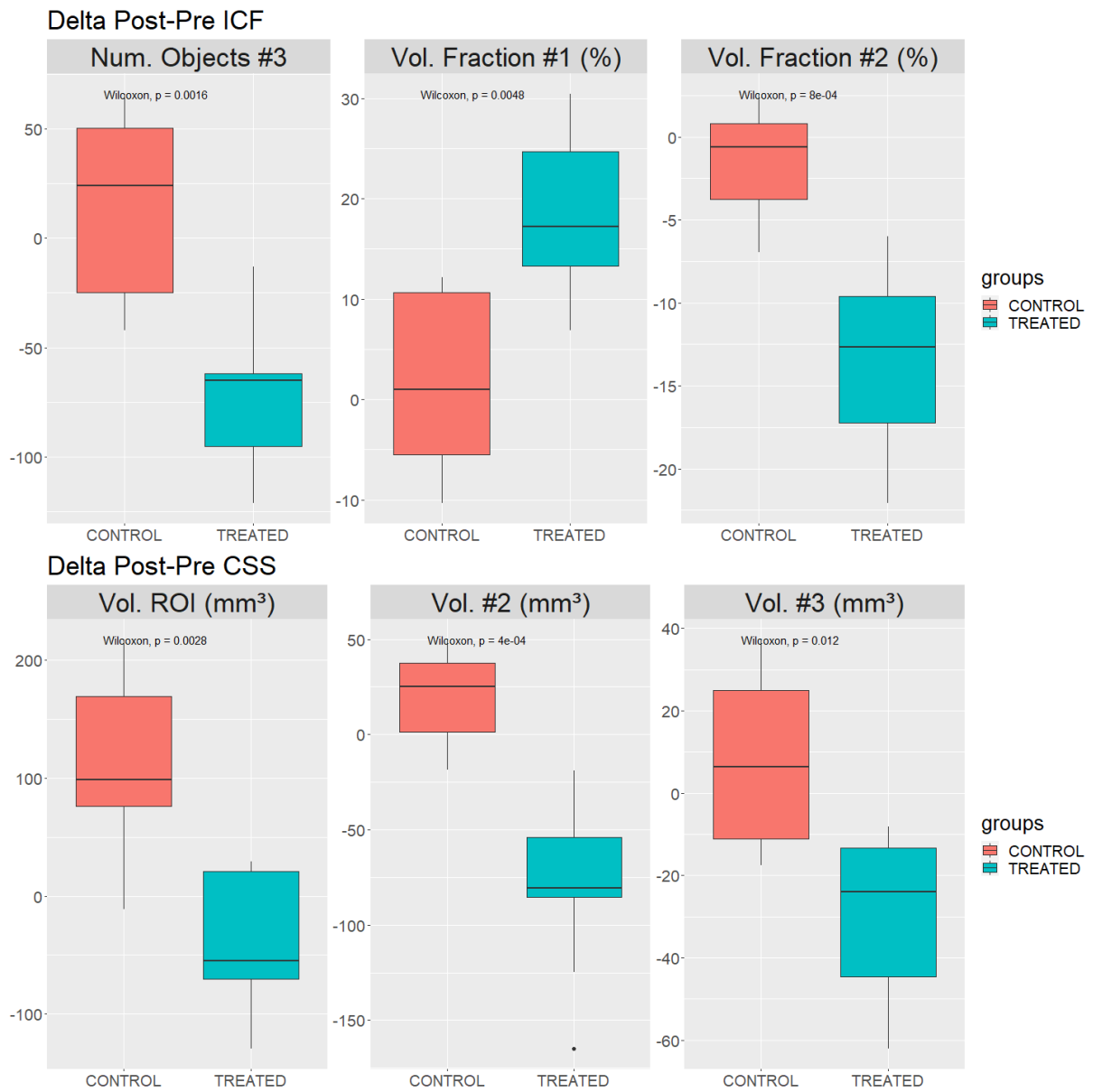


Fig. 3: Single slices illustrating  $K^{trans}$  maps (top row) and cluster maps (bottom row) evolution between day 1 and day 4 for one treated and for one control mouse. Centroids' positions for the cluster maps are indicated on colorbars



**Figure 4: Individual boxplots for features that significantly differentiated control and treated mice. Features are represented by their difference values (delta) between post-treatment (day 4) and pre-treatment (day 1). ICF = Intratumoral Complexity Feature, CSS = Conventional Summary Statistics**