

**From the Department of Radiology
at the University Hospital of Tübingen
Department of Preclinical Imaging and Radiopharmacy**

**Quantification of tumor heterogeneity using PET/MRI and
machine learning**

**Dissertation
submitted for a doctoral degree in
human sciences**

**at the
Faculty of Medicine
of the Eberhard Karls Universität
Tübingen**

**submitted by
Katiyar, Prateek**

2019

Dean: Professor Dr. I. B. Autenrieth
First reviewer: Professor Dr. B. J. Pichler
Second reviewer: Professor Dr. A. Schilling
Third reviewer: Professor Dr. K. Borgwardt

Date of oral examination: June 7th 2019

To my family

Contents

List of Figures	V
List of Tables	VII
List of Acronyms	IX
1 Introduction	1
1.1 Background	1
1.2 Outline	3
1.3 PET	4
1.3.1 Brief History	4
1.3.2 Beta-plus Decay	6
1.3.3 Interaction of Gamma Photons	7
1.3.4 Radiation Detection	9
1.3.5 Coincidence Events	10
1.3.6 Scanner Sensitivity and Spatial Resolution	11
1.3.7 PET Image Reconstruction	12
1.3.8 PET in Oncology	14
1.3.9 PET Quantification	15
1.4 MRI	17
1.4.1 Brief History	17
1.4.2 MRI Physics	19
1.4.3 Magnetization Vectors	21
1.4.4 T2 Relaxation	21

1.4.5	T1 Relaxation	22
1.4.6	Acquisition Parameters	25
1.4.7	T1-weighted Sequence	25
1.4.8	Proton Density Weighting	26
1.4.9	T2-weighted Sequence	26
1.4.10	Signal Localization	27
1.4.11	K-Space	28
1.4.12	Diffusion Weighted MRI	29
1.5	Simultaneous PET/MRI	31
1.5.1	Concept	31
1.5.2	Development and Applications	32
1.6	Machine Learning	34
1.6.1	<i>K</i> -means Clustering	36
1.6.2	Gaussian Mixture Modeling	39
1.6.3	Spectral Clustering	44
1.6.4	Random Forest	46
1.6.5	Feature Selection	51
2	Results	53
2.1	Project-1	55
2.1.1	A Population-Based Gaussian Mixture Model Incorporating ¹⁸ F-FDG PET and Diffusion-Weighted MRI Quantifies Tumor Tissue Classes	55
2.2	Project-2	81
2.2.1	Spectral Clustering Predicts Tumor Tissue Heterogeneity Using Dynamic ¹⁸ F-FDG PET: A Complement to the Standard Compartmental Modeling Approach	81
2.3	Project-3	106
2.3.1	A Novel Unsupervised Segmentation Approach Quantifies Tumor Tissue Populations Using Multiparametric MRI: First Results with Histological Validation	106

2.4	Project-4	131
2.4.1	Multi-view learning of multiparametric PET/MRI data quantifies intratumor heterogeneity	131
3	Discussion and Outlook	207
	Summary	213
	Zusammenfassung	215
	Bibliography	217
	Declaration of Originality	235
	Acknowledgements	239

List of Figures

1.1	Tumor heterogeneity in the clinic	2
1.2	Atomic models	5
1.3	Anderson's cloud chamber photograph of cosmic rays	5
1.4	Positron-electron annihilation	7
1.5	Interaction of γ photons	8
1.6	Photomultiplier tube and block detector	10
1.7	Coincidence events	11
1.8	Sinogram data	13
1.9	Two-tissue compartmental model	17
1.10	The first magnetic resonance image	19
1.11	Free hydrogen nuclei in magnetic field	20
1.12	Longitudinal and transverse magnetization	21
1.13	T_2 relaxation	23
1.14	T_1 relaxation	24
1.15	T_1 -weighted, proton density and T_2 -weighted images	26
1.16	Slice-selection gradient	28
1.17	Complete spin-echo sequence	29
1.18	K-Space	29
1.19	Diffusion-weighted images	31
1.20	PET/MRI of a melanoma patient	34
1.21	K -means clustering	38
1.22	A one dimensional mixture of Gaussians	40
1.23	Singularity during Gaussian mixture modeling	41

1.24 EM iterations of a GMM	43
1.25 Identification of non-convex clusters using spectral clustering	46
1.26 Spectral clustering solutions over different scales	47
1.27 Linear model versus a decision tree	49
1.28 Decision boundaries of a random forest	50
2.1 Tumor segmentation using a GMM	56
2.2 Modeling longitudinal PET/MRI tumor data	59
2.3 Simulated PET data	84
2.4 Input function distortions	85
2.5 Kinetic parameter estimation errors — preclinical	86
2.6 Kinetic parameter estimation errors — clinical	87
2.7 <i>In vivo</i> clustering example of a colon cancer xenograft	89
2.8 Histology imaging co-registration	108
2.9 SRSC results of a subcutaneous glioblastoma tumor	110
2.10 Qualitative comparison of different clustering algorithms	111
2.11 Benefits of spatial regularization	112
2.12 Perfusion in necrotic areas	113
2.13 Multi-view learning workflow	134
2.14 Multi-view spectral clustering	135
2.15 The $rest_{Control}$ and $rest_{Th-24}$ areas in histology	136
2.16 MSC segmentation of the Training set tumors	137
2.17 Proximity matrices and feature relevance scores	138
2.18 PET, MRI and PET/MRI classifier results	140
2.19 MRI classifiers applied to the glioblastoma tumors	141

List of Tables

1.1	Commonly used PET radio-tracers	15
1.2	Gyromagnetic ratio	20
1.3	TR and TE values of different MRI images	27
2.1	Correlation between histology and clustering tissue fractions	112

List of Acronyms

ADC	Apparent diffusion coefficient
AIF	Arterial input function
APD	Avalanche photo-diode
DSC	Dice similarity coefficient
DW	Diffusion weighted
E	Expectation
EM	Expectation maximization
FCM	Fuzzy C-means
FID	Free induction decay
GLUT-1	Glucose transporter 1
GMM	Gaussian mixture model
H&E	Hematoxylin and eosin
LOR	Line of response
M	Maximization
MRI	Magnetic resonance imaging
MSC	Multi-view spectral clustering
NMR	Nuclear magnetic resonance
NMRI	Naval medical research institute
PET	Positron emission tomography
PMT	Photomultiplier tube
RBF	Radial basis function
RF	Radio frequency

List of Acronyms

SRSC	Spatially regularized spectral clustering
SUV	Standardized uptake value
TAC	Time activity curve
TE	Time of echo
TR	Time of repetition

Introduction

1.1 Background

Continued advancements in targeted therapies and immunotherapy have resulted into an increasingly precise treatment of patients with cancer [1, 2]. Whereas targeted therapies alter pre-selected molecular pathways that are known to play an important role in cancer progression (such as proliferation and angiogenesis) [1], immunotherapies recruit the body's own mechanisms of defense to eliminate the tumor cells [2]. Unlike conventional cytotoxic therapies, which directly influence cancer cells by driving them into a rapid cellular regression, the response to targeted therapies and immunotherapies can be delayed and might not even begin with a reduction in tumor size [3, 4]. Although the biological processes resulting in such responses are not fully understood, the mechanistic differences between the conventional and modern cancer treatments advocate for a more comprehensive response evaluation criterion than the widely utilized WHO and RECIST measures [4, 5]. A second factor limiting the applicability of the two guidelines [5] in the evaluation of the cancer treatments is the heterogeneity (figure 1.1) within tumors [6, 7]. Heterogeneous lesions are notoriously aggressive and have been closely associated with therapy relapses and reduced efficacy [8]. To measure

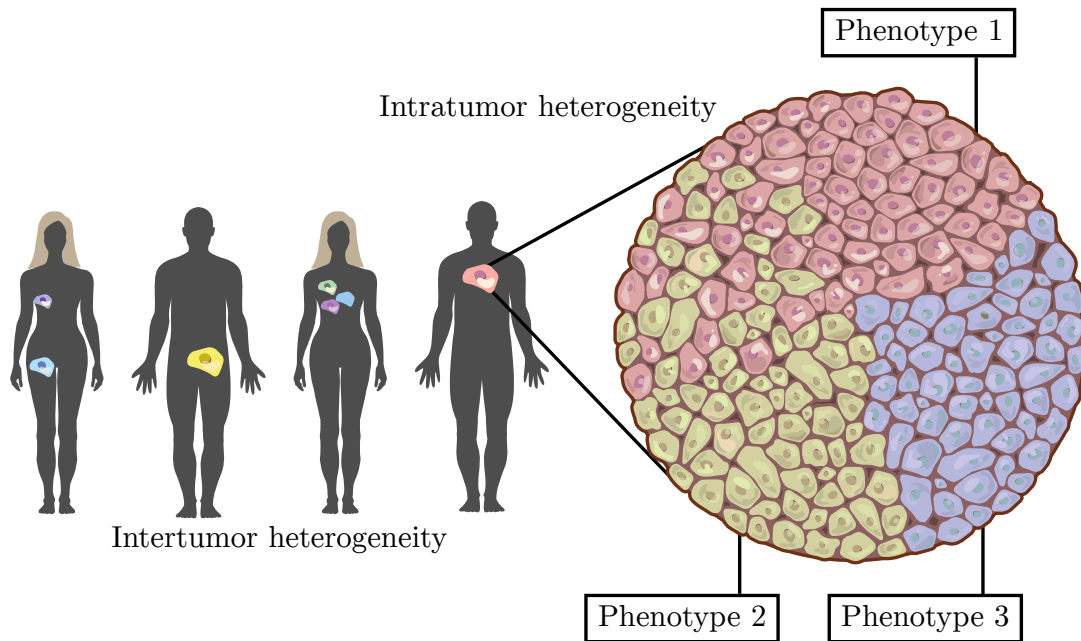


Figure 1.1: Tumor heterogeneity in the clinic. The tumor tissue characteristics can vary between different patients (intertumor heterogeneity) and also, within the same patient (intratumor heterogeneity). Moreover, the tissue types within the same tumor can either intermix with each other (phenotype 1 and 2) or can have well-described spatial boundaries (phenotype 2 and 3). Figure adapted from Burrell et al. [7].

the degree of intratumor heterogeneity, a common protocol in the clinic is to collect needle biopsies from primary tumors [6]. However, as the heterogeneity of these lesions stretches over the spatial and temporal scales, biopsies are often unable to characterize the overall phenotypic landscape of cancer.

For the purpose of tumor characterization, medical imaging is very well suited because it allows repeated non-invasive investigations of the patient. Although imaging is already widely used for the diagnosis and management of patients in oncology, a major fraction of the clinical studies rely on simple canonical features derived from the anatomical images to characterize various aspects of the tumors. To complement these investigations with additional quantitative data, more recently, a *radiomics* [9] approach has been proposed. In *radiomics*, the intratumor heterogeneity is quantified by calculating a plethora of shape- and texture-based features from the entire tumor. The hierarchy (from local to global dependencies between image voxels) of details captured by *radiomics* features provide a rich quantitative description of the underlying pathophysiology, enabling the physicians to go beyond the routine analysis of the available patient data.

While several studies have shown the prognostic benefits of such methodology [10, 11], it also has a few major limitations: the extracted features only characterize the tissue heterogeneity of tumors as a whole (commonly with a numeric score), without providing any information about the type and especially the location of molecularly distinct regions within a tumor. However, such knowledge of intratumor phenotypes has significant value in radiology, as it can provide strong indications about the failure [12] or success [13] of modern cancer treatments and may greatly assist in image-guided biopsies and radiation therapies.

The work conducted in this thesis, therefore, focused on techniques that can be applied on *in vivo* tumor data to obtain a spatial map of intratumor phenotypes. As distinct tissue types are more likely to differ in their functional characteristics than the anatomical features, we acquired multiparametric and multimodality datasets, capturing different functional processes of subcutaneous tumors in mice, using positron emission tomography (PET) and magnetic resonance imaging (MRI). Thereafter, we explored the feasibility of various machine learning techniques in analyzing the intricate multiparametric single or multimodality imaging data and in obtaining accurate maps of tumor tissue habitats. Above all, the results presented in the final study of this thesis highlight the prognostic benefits of multiparametric PET/MRI data over single modality imaging and demonstrate the important role of machine learning in augmenting the scope of simultaneous PET/MRI towards personalized medicine.

1.2 Outline

The remainder of this chapter provides the basics of the medical imaging modalities and the machine learning methods relevant for this dissertation. The entire work presented in this thesis was carried out in collaboration with several colleagues and is based on four publications. A summary of each paper and any unpublished results are provided in chapter 2. The sections in chapter 2 also include the official published copy of the respective articles and their supplemental data as attachments. Afterwards, a brief cumulative discussion on the research

findings and contributions of all papers is given in chapter 3.

1.3 PET

1.3.1 Brief History

The development of PET technology was made possible by a series of advances in many different fields. However, the inception of the PET principle largely attributes to the discovery of positrons. The existence of subatomic particles was first shown near the end of 19th century, when British physicist Joseph Thomson conducted a set of experiments, which led to the discovery of electrons. His experiments proved that cathode rays were composed of negatively charged particles (electrons), whose mass was much lower than the mass of the lightest atom [14]. Thereafter, in 1904, Thomson proposed the plum pudding model of the atom [15]. In his model, Thomson suggested that an atom was a sphere of positively charged space, which maintained an electrical neutrality due to uniformly embedded electrons (as shown in figure 1.2A). This model was later revised in a paper by Ernest Rutherford [16], where the nucleus was described as a dense positively charged core with negatively charged electrons revolving around it (figure 1.2B). Rutherford's planetary model, however, violated laws of classical physics and was insufficient in describing the spectra emitted by atoms [17]. In 1913, Niels Bohr modified the planetary model (figure 1.2C) by postulating that electrons can orbit only in certain energy levels (in stationary orbits) and that the line spectra exhibited by atoms is a result of emitted photons due to electron jumps from one level to the other [18]. More than a decade after these events, Paul Dirac coined the idea of positively charged electrons [19], which were discovered by Carl Anderson while studying cosmic radiation using a cloud chamber (figure 1.3) in 1932 [20].

In the subsequent years, the discovery of the positron was complemented by developments in radiochemistry and material physics [21], and the first clinical application of positron-emitting radioisotope was shown in 1945 [22]. The study

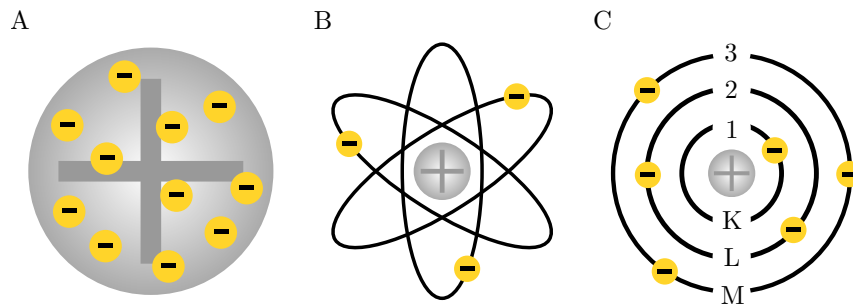


Figure 1.2: Atomic model proposed by (A) J. J. Thomson, (B) Ernest Rutherford and (C) Niels Bohr.

by Tobias et al., however, used a Geiger-Müller counter for the detection of γ photons. Within six years, two separate investigations [23, 24] laid the foundation of modern PET scanners by showing the feasibility of coincidence detection, leading to a significant improvement in overall sensitivity and tracer localization. A complete timeline of the next generation PET systems and their gradual induction in the clinic as a vital medical imaging modality is outlined by Dayton Rich [21] and Ronald Nutt [25].

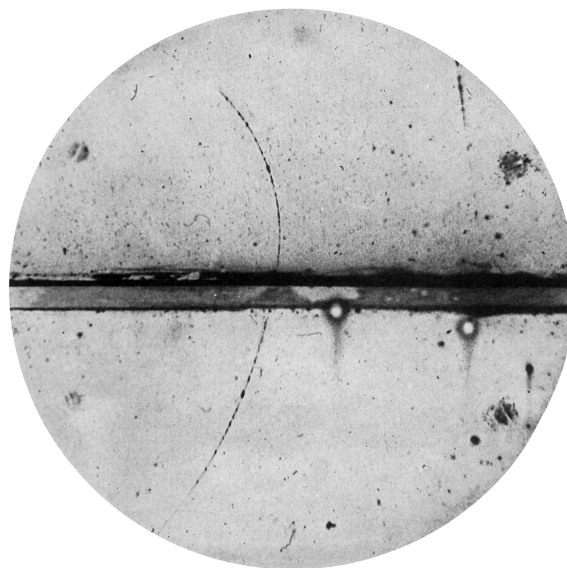
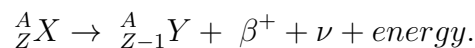


Figure 1.3: The cloud chamber photograph of cosmic rays taken by Anderson in 1932 [20]. The curved black line depicts the positron trail, which enters the chamber from the bottom, penetrates the 6 mm lead plate and loses energy while moving upwards.

1.3.2 Beta-plus Decay

Positrons can be produced in two different ways: nuclear disintegration and pair production [26]. Due to its application in radiotracer synthesis, only the former method is described in this section.

Nuclear disintegrations resulting in the emission of positrons are classified as beta-plus decay. Specifically, in beta-plus decay, a “neutron-deficient element¹” (high proton-to-neutron ratio) converts one of its excess protons into a neutron along with the emission of a positron and a neutrino. This nuclear transmutation can be summarized using the following equation:



Thus, after a beta-plus decay the daughter nucleus (Y) has the same atomic mass (A) as the parent nucleus (X), but its atomic number (Z) is reduced by one. Here β^+ and ν symbolize a positron and a neutrino, respectively. The positrons are emitted with a spectrum of energies, with the maximum energy determined by the mass difference between the daughter and parent nuclei. The average energy of the emitted positrons is roughly equal to one third of the maximum energy [27].

Post-emission, the kinetic energy of the positron is dissipated (analogous to a freely moving electron) in its interactions with the surrounding matter. The elastic and inelastic collisions [26] with atomic nuclei and electrons result in a tortuous path of the particle. After losing almost all of its energy, the positron combines with an electron to form an intermediate state, known as *positronium*. This state exists for a short fraction of time ($\sim 10^{-10}$ seconds [28]) before the sub-atomic particles annihilate and release a pair of photons. As the positron loses most of its kinetic energy prior to annihilation, the rest mass of both the particles is converted into electromagnetic radiation equal to an energy of 1022 keV. Thus, conserving the net energy and momentum, two γ photons are emitted in opposite directions,

¹ An alternative mechanism to beta-plus decay for neutron-deficient elements is electron capture decay, which results in an X -ray photon and/or an Auger electron [27].

each with an energy of 511 keV [26]. The annihilation process is shown in figure 1.4. In many cases, due to non-zero momentum of the positron, the photon pairs are not emitted strictly opposite to each other. The displacement between the positron emission and annihilation sites as well as the non-collinearity of the emitted photons limit the maximum spatial resolution of the PET imaging devices [26].

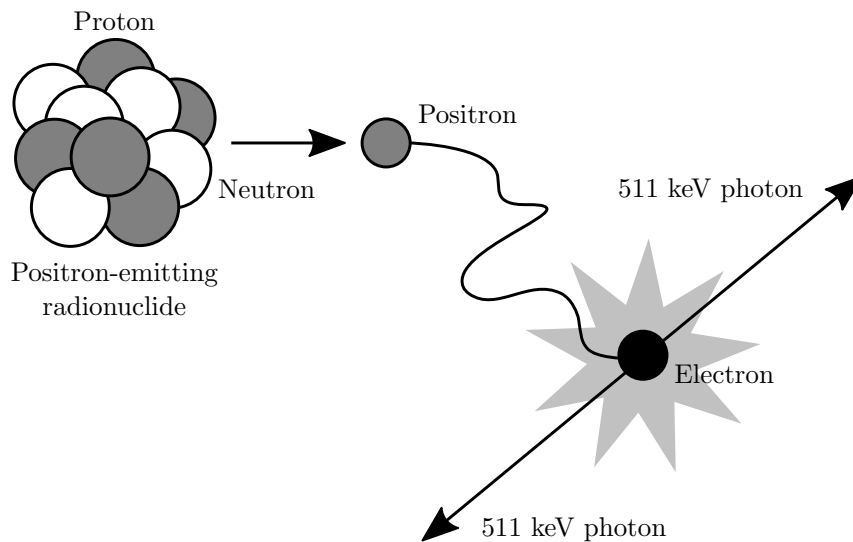


Figure 1.4: Positron emission from a radionuclide and its annihilation with an electron. The resulting photons are nearly collinear (with the annihilation site) and have an energy of 511 keV. Figure adapted from Cherry et al. [28].

1.3.3 Interaction of Gamma Photons

There are three major mechanisms through which γ photons interact with the surrounding matter: the photoelectric effect, Compton scattering and pair production. All of these processes result in either the ionization or excitation of the material in-contact.

In the photoelectric effect (figure 1.5A), an energetic γ photon transfers its entire energy to one of the electrons in the inner orbit² (figure 1.2C), resulting in the ejection of the electron from the interacting atom. The vacancy created in the inner shell of the atom is filled by an outer shell electron, with the emission of an X-ray photon carrying an energy equal to the binding-energy difference between

² A γ photon usually transfers its energy to the electrons in the K or L shell.

the two shells. In some cases, instead of transferring an electron from the outer shell, the atom may remove excess energy by ejecting a secondary electron, known as Auger electron. In living organisms, interacting photons with an energy lower than 100 keV generally result into the emission of photo-electrons because of the photoelectric effect [26].

Compton scattering takes place when a photon collides with a loosely bound electron in the outer shell of an atom. The inelastic impact transfers part of the energy from the photon to the electron, resulting in deflection of the photon and ejection of the electron (figure 1.5B). This free electron is termed as the recoil electron. In tissue, photons between the energy range of 100-2000 keV primarily lead to Compton scattering [26].

Pair production is an ionization process in which a γ photon splits into an electron and a positron (figure 1.5C). The creation of a particle and its anti-particle requires that the γ photon is in proximity to a nucleus and carries an energy equal to or

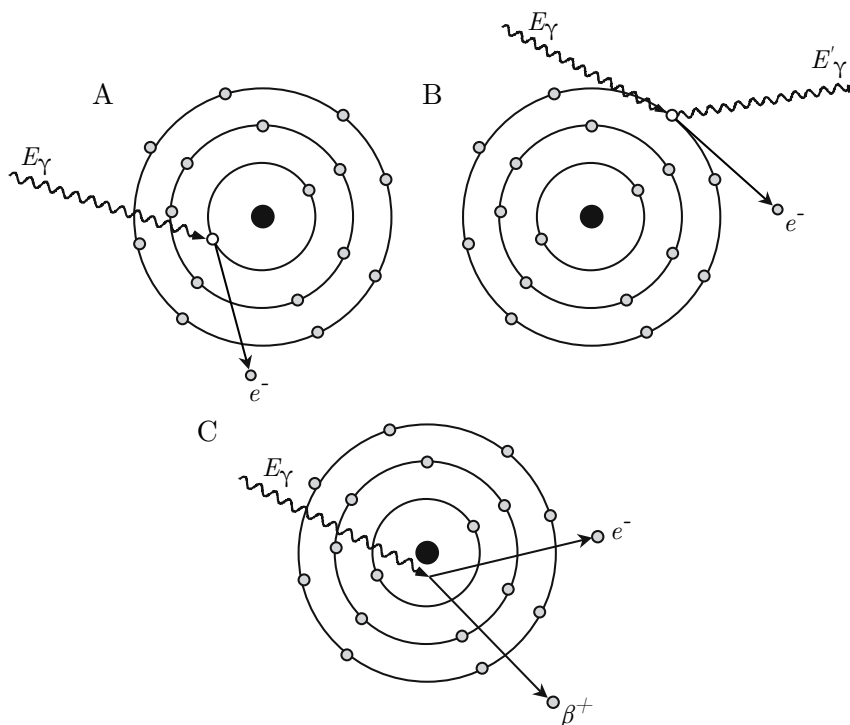


Figure 1.5: Illustrations depicting three primary processes through which γ photons interact with the surrounding matter. (A) the photoelectric effect, (B) Compton scattering and (C) pair production. Here E_γ and E'_γ denote the energy of the incident and deflected γ photons. The electron and positron are symbolized by e^- and β^+ , respectively. Figure adapted from Bailey et al. [26].

greater than the total rest mass energy of both the particles. In the case of a γ photon with higher energy, the remaining energy is distributed between the emitted particles.

1.3.4 Radiation Detection

The detection of incident radiation and its conversion into a proportional electrical signal are key processes in nuclear medicine. Based on the mechanisms of these processes, radiation detectors can be broadly classified in three categories: gas-filled, semiconductor and scintillator detectors.

A desired characteristic for PET detectors is superior stopping efficiency (their ability to stop 511 keV γ photons). Among the three detector categories mentioned previously, only scintillator detectors meet this requirement. These detectors use scintillating crystals and photo-detectors in a cascade. When a γ photon incidents upon the scintillating crystal, its energy is absorbed by the electrons in the material, causing their transition from the ground state to an excited state. As these electrons return to the low-energy state, they release the absorbed energy (equivalent to the energy gap between the ground and excited state) in the form of visible light³ [26], which is detected and converted into an electrical signal by the cascaded photo-detector.

The most routinely used photo-detector in PET scanners is a photomultiplier tube (PMT). As shown in figure 1.6A, a PMT comprises a cylindrical glass tube fitted with a photo-cathode and several dynodes, each of which is biased with a voltage supply higher than the previous dynode. The incoming light from the scintillator crystal incidents upon the photo-cathode, triggering the ejection of a photo-electron. The difference in potential between the photo-cathode and the closest dynode causes the photo-electron to accelerate towards the dynode. Upon impact, the kinetic energy of the photo-electron results in the ejection of secondary photo-electrons, which further advance towards the next dynode. This process

³ Often the radiation emitted by pure crystals is in ultra-violet range and impurities are added to shift the emitted energy in visible spectrum [26].

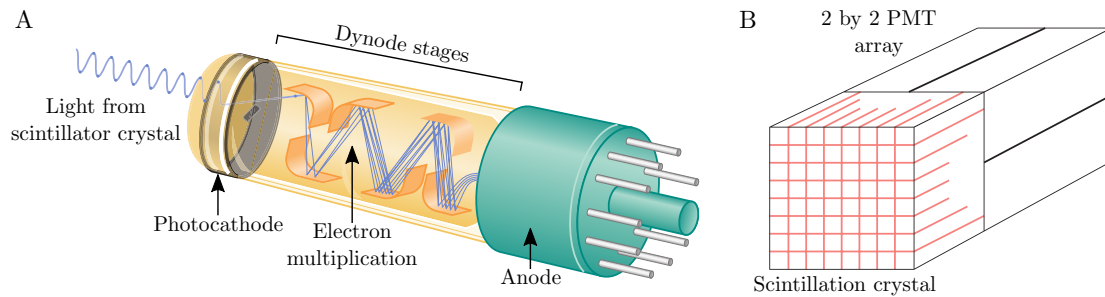


Figure 1.6: Schematic of a (A) PMT and a (B) block detector, commonly used in PET scanners for detecting γ photons. The block detector consists of a scintillator crystal cascaded with a 2 by 2 PMT array. This set-up uses a weighted combination of signals from the PMTs to calculate the exact spatial location of the incident γ photons. Slits are carved in the scintillator crystal to prevent the spreading of visible light emitted by the scintillator. Figure adapted from Bushberg et al. [27].

is repeated until the last dynode and the multiplied electrons are collected at the anode, providing an amplification factor of approximately one million [26].

1.3.5 Coincidence Events

As described in sub-section 1.3.2, positron annihilation results in the emission of two 511 keV γ photons. To detect the photons released in roughly opposite directions, PET scanners utilize a ring of detectors, as shown in figure 1.7A. The coincidences are recorded within a timing window that accounts for the time taken by the photons to reach the scintillators and the delay introduced by the front-end electronics. Afterwards, for each recorded event, a line of response (LOR) is created by joining the centers of the associated detectors. The LORs are subsequently processed by a reconstruction algorithm to create a two- or three-dimensional image indicative of the activity concentration in the PET field of view.

There are two types of coincidence events, which introduce statistical noise in the reconstructed PET image: random and scatter. A random coincidence, as the name suggests, is a phenomenon that results in a false LOR due to the detection of photons from two random annihilation events within the same timing window (figure 1.7B). The number of random coincidences between any detector pairs is directly proportional to the length of the timing window as well as the true count rate of each of the detectors (i.e. total amount of activity in the PET field of view)

[27].

A scatter coincidence (figure 1.7C), on the other hand, involves Compton scattering of either one or both the γ photons associated with the same annihilation event. Hence, a scatter coincidence is theoretically a true coincidence event, which results into a false LOR leading to increased background noise in the reconstructed PET image.

1.3.6 Scanner Sensitivity and Spatial Resolution

The sensitivity of a PET scanner is defined as the rate of coincidence events detected for each unit of the activity concentration in the PET field of view. The most important factors determining the sensitivity of a PET scanner are: geometric and detector efficiencies, energy and timing windows, and the positioning of the activity source with reference to the detectors [28]. The geometric efficiency is further dependent on the physical elements of the scanner, such as the solid angle covered by the detectors and the packing fraction. Likewise, the detector efficiency is dictated by the characteristics (scintillation decay time and stopping efficiency) of the scintillator crystal. Post acquisition, the influence of scattered coincidences (sub-section 1.3.5) is minimized by selecting the detected events only within a defined energy window [28]. As the intrinsic system sensitivity has a direct impact on the acquired image, the PET scanners are carefully designed to maximize its value.

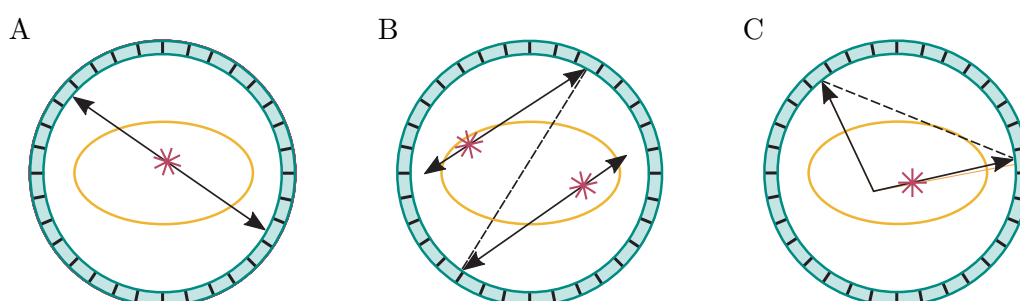


Figure 1.7: Example of (A) true, (B) random and (C) scatter coincidence events. The solid and dashed lines represent the true and false LORs, respectively. The object placed in the PET field of view is depicted in yellow and an annihilation event is shown with a red asterisk. Figure adapted from Bushberg et al. [27].

Besides the fundamental limit set by the positron range of the used radionuclide and non-collinearity of the emitted γ photons, the spatial resolution of a PET scanner is determined by the type and physical characteristics of the detector and the algorithm used for reconstructing the PET image. Common PET systems utilize block detectors in which the scintillator crystals and PMTs are arranged into an array (one such detector is shown in figure 1.6B). These detectors provide an intrinsic spatial resolution of approximately half the width of the scintillator crystal in the center of the field of view. However, because of the finite length of the crystals, the spatial resolution further degrades towards the edges⁴. In addition to system geometry, during post-processing, the choice of the image reconstruction algorithm and the corresponding parameters affect the overall system resolution. Further details of these methods are provided in the following sub-section. The finite spatial resolution of PET scanners leads to partial volume effects, which limits their ability to reliably quantify activity in small regions of interest.

1.3.7 PET Image Reconstruction

PET reconstruction algorithms are broadly divided in two categories: analytical and iterative. Both the approaches use sinogram data to reconstruct an image, reflective of the radioactivity in the PET field of view. A sinogram is a projection view of all the LORs at a certain angle (i.e. parallel to each other) from a reference coordinate system. Figure 1.8A shows a set of vertical LORs, which together form a row of the sinogram matrix, with each column representing the radial distance of the LOR from the scanner's center. Thus, each element of the sinogram matrix is the total number of coincidences accumulated by the associated pair of detectors. Panel B and C in figure 1.8 depict the axial PET image of a patient and the respective sinogram [29].

The most basic PET reconstruction algorithm is backprojection. Fundamentally, in backprojection, the PET image is calculated using a weighted summation of all the LORs. At first, the size of the target (or reconstruction matrix) image is

⁴ Moreover, the detectors not providing the depth of interaction information result in anisotropic and non-uniform spatial resolution due to parallax errors [28].

defined. Afterwards, for each pixel of the image, a sinogram coordinate is calculated based on its spatial location and a projection angle, which determines the count value to be assigned to the target pixel. If the sinogram coordinate does not fully coincide with the sinogram element, interpolation is used to calculate the exact count value. Also, a weighting factor is applied to account for the distribution of the counts from each pair of detectors to all the target pixels coinciding with its LOR. This process is repeated for all the projection angles to finally obtain the reconstructed image. Backprojection is an intuitive and simple way of reconstructing the activity distribution within an object, however, it has one major limitation. As the counts from each pair of detectors are distributed across all the pixels coinciding with its LOR, many of the events are falsely assigned outside the object, resulting into a smoothed version of the true activity distribution.

The issues encountered with backprojection algorithm are mitigated using a simple trick, based on the projection slice theorem. In particular, during reconstruction the Fourier transform is used to map the activity from the sinogram domain to the frequency domain. Thereafter, to find a balance between the high frequency components and signal to noise ratio, the Fourier transformed data is multiplied with the frequency response of a filter. This operation is analogous to convolving the sinogram data with the response of a filter in the spatial domain. Lastly,

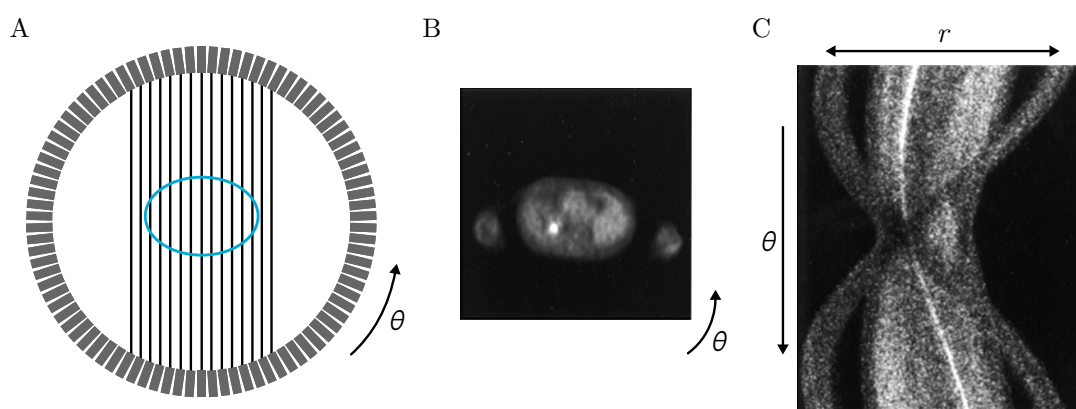


Figure 1.8: (A) A PET detector ring with vertical LORs. The object placed in the PET field of view is highlighted in blue. The projection view of these LORs will be identical to an axial image of the object captured using a collimated scanner. The direction of projection angles is indicated with θ . (B) An axial PET image of a patient and (C) the corresponding sinogram. The arms of the patient are clearly distinguishable in the sinogram image at the top and bottom projection angles. Figure adapted from Bailey et al and Turkington. [26, 29].

the filtered sinogram data is transformed back to the spatial domain using the inverse Fourier transform and backprojection is used to reconstruct the filtered PET image. Due to such post-processing, this method is termed as filtered backprojection [28]. The three most widely used filters in filtered backprojection algorithm are ramp, Shepp–Logan and Hann. The ramp filter enhances high frequency components (i.e. edges) improving the spatial resolution at the cost of signal-to-noise ratios. On the contrary, Shepp–Logan and Hann are apodizing filters, which have responses similar to a ramp filter at low frequencies but a reduced amplitude at high frequencies. These filters reduce ringing artifacts and favor high signal-to-noise ratios over the spatial resolution.

In contrast to the analytical methods, iterative reconstruction techniques take the imaging process into account by formulating a cost function followed by an optimization step to obtain the reconstructed image. These methods start with an approximated PET image (generally an image with all zeros or ones) and compute the sinogram estimate that would produce such an image. A cost function is then used to measure the discrepancy between the measured and estimated sinogram data, and to improve the initial approximation of the image. This process is repeated over several iterations until the algorithm converges to a minimum/maximum. As radioactive decay is a stochastic process, a common optimization paradigm is to obtain a maximum likelihood estimate of the count distribution under a Poisson model using the expectation maximization (EM) algorithm [30]. In order to achieve faster reconstructions, sampling can be performed while forward and backward projecting the image data (computing the sinogram estimate). One such technique is ordered subset expectation maximization that groups the projection angles prior to optimization [31].

1.3.8 PET in Oncology

Over the past decade, PET has progressed as a key imaging modality for the staging and management of patients with cancer. The developments in radiochemistry now allow physicians to quantify a wide range of physiological pro-

cesses associated with cancer, such as glucose metabolism, cell proliferation, tissue hypoxia and lipid synthesis [32, 33]. A list of tracers commonly used in clinical or preclinical PET studies is provided in table 1.1. Arguably the most widely used tracer in oncological PET examinations is ^{18}F -FDG. Due to the accumulation of glucose in many types of cancer, ^{18}F -FDG is established as a primary clinical tracer for the early detection and subsequent staging and therapy planning of various neoplasms [34]. Another factor contributing to the widespread use of ^{18}F -FDG in the clinic is the favorable half-life (109.8 min) of the radio-nuclide.

1.3.9 PET Quantification

An important aspect of PET studies is to derive quantitative measures that allow physicians to estimate the response to therapy, measure overall cancer burden and make comparisons between various patient populations examined across a group of institutions. For this purpose, static PET studies commonly rely on computing the standardized uptake value (*SUV*) [35, 36]. The *SUV* (g/mL) at time *t* can be calculated as follows:

$$SUV = \frac{C_{PET}(t)}{Dose/Weight}.$$

Where, $C_{PET}(t)$, *Dose* and *Weight* are the measured activity concentration (kBq/mL), injected dose (MBq) and body weight (kg) of the subject, respectively. Often, the quantitative reliability of the *SUV* is impaired by the simplicity of the

Radio-nuclide	$T_{1/2}$ (min)	Tracer	Application
^{15}O	2.03	^{15}O -H ₂ O	Perfusion
^{11}C	20.3	^{11}C -choline	Cell proliferation
		^{11}C -acetate	Lipid synthesis
^{18}F	109.8	^{18}F -FDG	Glucose metabolism
		^{18}F -FMISO	Tissue hypoxia
		Na ^{18}F	Bone abnormalities

Table 1.1: Commonly used positron emitting radio-nuclides, their half-lives ($T_{1/2}$), few associated tracers, and the applications of respective tracers in life-sciences.

measure. Some of the factors responsible for causing variability in the *SUV* measurements are: body size of the subject, blood glucose level, subject breathing, scanner specifications, parameters of the reconstruction algorithm and the delineated region of interest [35–37].

An alternative and more systematic way of quantifying the distribution of a tracer and its transport mechanisms is to perform tracer compartmental modeling. The objective of compartmental modeling is to find a set of parameters, which provide the most optimal fit for the observed time activity curves given the modeling constraints between the observed data and the parameters describing the physiological process. In practice, for a specific tracer, a compartmental model is a simplified approximation of a biological system that takes into account the most influential factors characterizing the functional properties of the region of interest, such as perfusion, metabolism and receptor concentration.

Figure 1.9 shows a two-tissue compartmental model for ^{18}F -FDG. The time varying tracer activity concentration in plasma and the two tissue (free and bound) compartments are denoted by $C_p(t)$, $C_1(t)$ and $C_2(t)$, respectively. As illustrated in the figure, the rate of tracer exchange between plasma and the first tissue compartment is governed by K_1 and k_2 , whereas the same is regulated between the free and bound tissue compartments by k_3 and k_4 . The parameters K_1 , k_2 , k_3 and k_4 are therefore termed as kinetic rate constants. Except for K_1 , the other three kinetic rate constants have unit min^{-1} . The parameter K_1 additionally incorporates a perfusion component and has units of $\text{mL}/\text{min}/\text{g}$. Thus, based on these modeling constraints and given the observed tissue activity concentration $C_T(t)$ and the plasma input function $C_p(t)$, the kinetic parameters of the two-tissue compartmental model can be estimated using the following equations [38]:

$$\frac{dC_1(t)}{dt} = K_1 C_p(t) - (k_2 + k_3) C_1(t) + k_4 C_2(t)$$

$$\frac{dC_2(t)}{dt} = k_3 C_1(t) - k_4 C_2(t)$$

$$C_T(t) = C_1(t) + C_2(t) + f_b C_p(t).$$

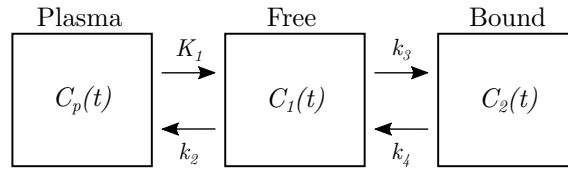


Figure 1.9: A two-tissue compartmental model for ^{18}F -FDG.

Where, f_b is the fraction of vascular space. By solving the first two ordinary differential equations, the free and bound tissue concentrations ($C_1(t)$ and $C_2(t)$) can be expressed as:

$$C_1(t) = \frac{K_1}{\alpha_2 - \alpha_1} [(k_4 - \alpha_1)e^{-\alpha_1 t} + (\alpha_2 - k_4)e^{-\alpha_2 t}] * C_p(t)$$

$$C_2(t) = \frac{K_1 k_3}{\alpha_2 - \alpha_1} [e^{-\alpha_1 t} - e^{-\alpha_2 t}] * C_p(t).$$

Where $*$ denotes the convolution operation. The variables α_1 and α_2 are algebraic functions of the kinetic parameters and have the following form:

$$\alpha_1 = \frac{k_2 + k_3 + k_4 - \sqrt{(k_2 + k_3 + k_4)^2 - 4k_2k_4}}{2}$$

$$\alpha_2 = \frac{k_2 + k_3 + k_4 + \sqrt{(k_2 + k_3 + k_4)^2 - 4k_2k_4}}{2}.$$

The kinetic parameter estimation is carried out by minimizing the sum of squared differences between the measured dynamic PET data and the data predicted by the compartmental model.

Other model-based approaches frequently applied to dynamic PET data are discussed in Bailey et al. [26].

1.4 MRI

1.4.1 Brief History

The foundations of MRI were laid by an American physicist Isidor Isaac Rabi in 1938. In his experiments, Rabi demonstrated the phenomena of nuclear magnetic resonance (NMR) by subjecting lithium chloride molecules to oscillating magnetic fields [39]. These experiments drew inspiration from the previous work of Stern

and Gerlach, which had confirmed that the intrinsic angular momentum of elementary particles is quantized [40]. Within several years, two separate investigations conducted by Edward Purcell and Felix Bloch reported the NMR characteristics of hydrogen molecules in paraffin and water, respectively [41, 42]. Purcell's approach was based on the measurement of energy absorption at resonance frequency, whereas Bloch's method relied on a crossed-coil (transmitter-receiver) arrangement for measuring the signal produced during nuclear induction. Later on, both the techniques played a significant role in formulating the theoretical basics of MRI. The next milestone in the history of MRI was marked by Erwin Hahn in 1950. In his paper titled "Spin Echoes" [43], Hahn demonstrated the application of a series of time-dependent radio-frequency (RF) pulses to produce a measurable nuclear induction signal. Afterwards, the development of Fourier transform spectroscopy in 1966 [44] greatly improved the NMR sensitivity and opened up the field of magnetic resonance to a broader community. These two contributions were particularly pivotal for the development of MRI and have been influencing to this day its application in medicine and biology.

However, it was only after 1971 that Raymond Damadian published the first NMR study [45], which distinguished between the malignant and benign tissues in a rat based on their T_1 and T_2 relaxation times. The final component for creating a magnetic resonance image (localizing NMR signals in space using magnetic field gradients) was invented by Paul Lauterbur in 1973 [46]. Lauterbur demonstrated the applicability of his approach (shown in figure 1.10) by publishing the 2-D images of test tubes filled with water. Soon after this publication, Peter Mansfield discovered the techniques of slice selection and snapshot acquisition, enabling the measurement of 2-D MRI images within a small fraction of time [47]. In the next three decades a series of technological advancements were made [48], which shaped the MRI scanners into the form that is currently being used. Lauterbur and Mansfield were jointly awarded with the 2003 Nobel Prize in Physiology or Medicine, due to their contributions that lead to the development of MRI.

1.4.2 MRI Physics

The MRI principle relies on the phenomena that in the presence of an external magnetic field, atoms of certain elements can absorb and emit energy when excited with an RF energy source. This property of the atoms is dependent on the total number of neutrons and protons present in the nucleus. If the nucleus has even number of neutrons and protons, the net nuclear magnetic moment of the atom is zero (protons and neutrons both have half integral spin, but in opposite directions). However, in case either or both the particles are odd in number, the resulting magnetic moment of the nucleus is finite. Among the other elements, hydrogen is the best MRI candidate for two main reasons: it has a large magnetic moment and it is an integral structural element of fat and water molecules, both of which are highly abundant in living organisms. In the absence of an external field, the resultant magnetic moment of a large number of loosely bound hydrogen nuclei in fat tissue and water is roughly equal to zero. However, when subjected to a static magnetic field these nuclei align either parallel or anti-parallel⁵ to the direction of the applied field, which results in an overall measurable nuclear magnetic moment (figure 1.11).

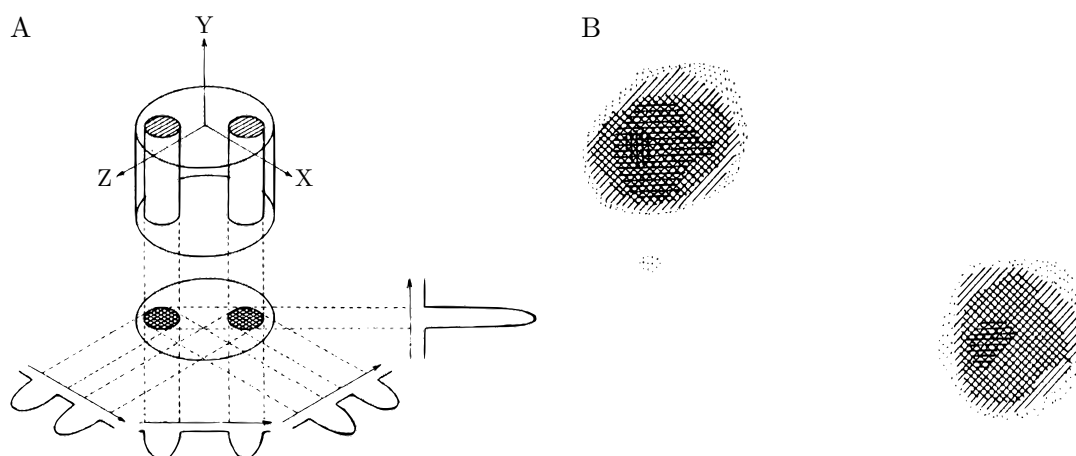


Figure 1.10: The first magnetic resonance image. (A) A schematic of the method used by Lauterbur to produce the first magnetic resonance images. (B) The acquired MRI pictures of two test tubes filled with water. The magnetic field gradients were applied in the direction of the arrows depicted in panel A. Figure adapted from Lauterbur. [46].

⁵ Slightly more parallel than anti-parallel because the former configuration results into a lower energy state than the latter.

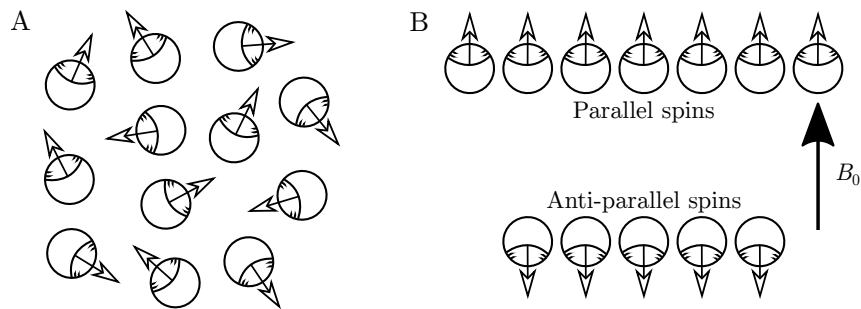


Figure 1.11: A simplified illustration depicting the behavior of hydrogen nucleus in fat tissue and water in the absence and presence of a static magnetic field (B_0). (A) In the absence of the applied field, the loosely bound protons are oriented randomly with zero net magnetic moment. (B) However, when placed in the external magnetic field, the protons align either in the direction or opposite to the direction of the applied field, resulting in a finite magnetic moment. Figure adapted from Bushberg et al. [27].

Besides the alignment in parallel and anti-parallel states, the hydrogen nuclei also precess (figure 1.12A) on their own axes due to the torque experienced by the protons orthogonal to the direction of the applied magnetic field. The angular frequency of the rotation (ω_0) is therefore termed as precession frequency and can be calculated using the Larmor equation:

$$\omega_0 = \gamma * B_0.$$

Where γ is the gyromagnetic ratio of the element and B_0 is the strength (in Tesla) of the static magnetic field. Table 1.2 details the gyromagnetic ratio (in MHz/Tesla) of some of the MRI active elements.

Nucleus	$\gamma/2\pi$ (MHz/Tesla)
^1H	42.58
^{13}C	10.7
^{17}O	5.8
^{19}F	40.0
^{23}Na	11.3
^{31}P	17.2

Table 1.2: Gyromagnetic ratio of useful MRI active elements. Data taken from Bushberg et al. [27].

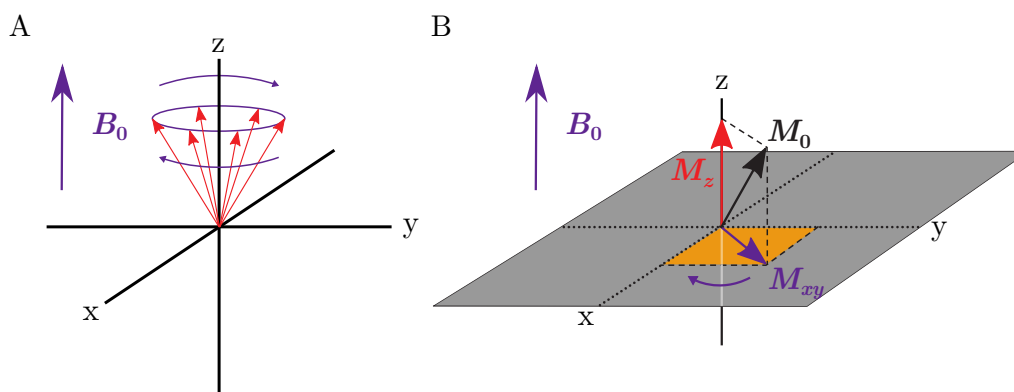


Figure 1.12: Longitudinal and transverse magnetization. (A) The precession of magnetic moment vector (indicated by red arrow) in the presence of a static magnetic field applied along the z-axis. At this point, the net magnetic moment in xy-plane is zero. (B) Once an RF pulse is applied with a flip angle of 90-degree, the magnetic moment vector is displaced to xy-plane resulting in a finite transverse magnetization vector M_{xy} . The symbol M_0 indicates the longitudinal magnetization vector M_z at equilibrium. Figure adapted from Bushberg et al. [27].

1.4.3 Magnetization Vectors

A common convention in MRI is to assume the direction of the static magnetic field along the z-axis of the Cartesian co-ordinate system. Therefore, the resultant magnetic moment in z-direction is termed as longitudinal magnetization, which can also be represented using a vectorial form (as shown in figure 1.12A). Under equilibrium conditions there is a finite longitudinal magnetization and the magnetic moment in xy-plane (known as transverse magnetization) is zero. This system can be perturbed by applying RF pulses with a specific angle and the frequency matched with the precession frequency of the hydrogen nuclei. The synchronized RF pulses cause a transfer of energy to the object placed in the magnetic field and displacement of the magnetic moment vector in the direction of the applied pulse signal. For instance, an RF pulse applied orthogonally to the z-direction flips the magnetic moment vector in xy-plane, resulting into a non-zero transverse magnetization (figure 1.12B).

1.4.4 T2 Relaxation

Upon the application of an RF pulse with a 90-degree flip angle and the Larmor frequency of protons, a spin coherence between the nuclei of hydrogen atoms is

achieved, which results in a maximum value (specific to the tissue) of the transverse magnetization. As soon as the RF pulse is ended, the individual nuclei start to dephase due to the intrinsic inhomogeneities in the magnetic properties of the sample. Thus, the amplitude of the transverse magnetization decays exponentially with time while precessing close to the Larmor frequency of protons. The decaying transverse magnetization, when converted into an electrical signal using a receiver coil, results into a damped sinusoidal wave (figure 1.13), which is termed as the free induction decay (FID). This process can be summarized using the following expression:

$$M_{xy}(t) = M_0 e^{-t/T_2}.$$

Where $M_{xy}(t)$ and M_0 are the magnitudes of the transverse magnetization at time t and at $t = 0$, respectively. The parameter T_2 is known as the T_2 relaxation time and is equal to the elapsed time until the amplitude of the transverse magnetization reaches 37% of its maximum value (i.e. M_0). The exponential decay of the transverse magnetization is also known as spin-spin relaxation, because it is caused by the intrinsic spin-to-spin interactions within the sample. Due to this fact, the molecular composition and the amount of water in various tissues greatly influence their T_2 decay values. In general, regions with less constrained structure and more water exhibit higher T_2 values, whereas tissues with densely packed molecules (such as fat) have smaller T_2 values. In addition to variations in intrinsic properties, external factors, such as inhomogeneities in the static magnetic field (B_0), and the presence of matter altering the local magnetic susceptibility of the tissue contribute in the decay of the transverse magnetization. The time constant associated with the external factors of spin dephasing is known as T_2^* decay.

1.4.5 T1 Relaxation

Concurrently to the decay of transverse magnetization, a recovery of the longitudinal magnetization takes place. As the protons start to dephase after the RF pulse is switched off, they also begin to return to their initial orientation and align

themselves either parallel or anti-parallel to the static magnetic field. This causes the longitudinal magnetization vector to recover and attain the magnitude and direction similar to those of the equilibrium state. Mathematically, the recovery process is described as follows:

$$M_z(t) = M_0 (1 - e^{-t/T1}).$$

Where $M_z(t)$ is the magnitude of the longitudinal magnetization at time t and M_0 corresponds to its fully recovered value. The parameter $T1$ is known as the $T1$ relaxation time and is equal to the elapsed time until the amplitude of the longitudinal magnetization recovers to 63% of its maximum value (i.e. M_0). Contrary to $T2$ decay, $T1$ recovery takes place due to spin-lattice interactions. During this process, the energetic hydrogen nuclei (loosely bounded in molecular structures) dissipate their energy to the surrounding lattice and return to the equilibrium state. The energy dissipation is most effective (leading to shorter $T1$ relaxation time) when the Larmor frequency of hydrogen nuclei matches with the tumbling (vibrational and random motion) frequency of the surrounding lattice and hydration layers. Thus, the dissimilarities in physical characteristics of various tissues and pathologies lead to differences in their $T1$ relaxation times. Overall, unstructured tissues and fluids have large $T1$ relaxation times in contrast to those of the structured ones (lipids and moderately sized proteins). With an increase in the strength of the static magnetic field, the Larmor frequency of hydrogen nuclei increases,

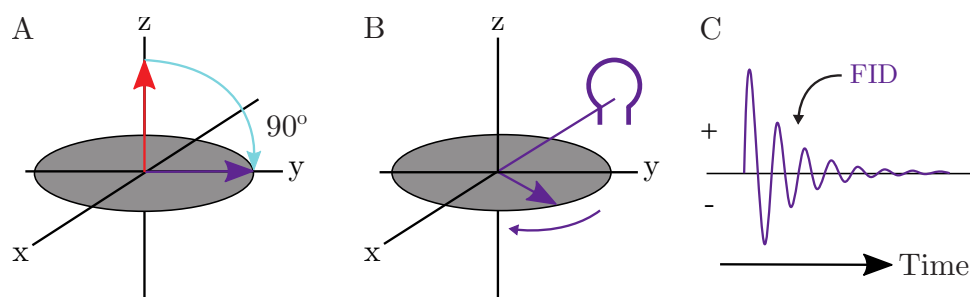


Figure 1.13: $T2$ relaxation. (A) Displacement of the longitudinal magnetization into the xy -plane upon the application of a 90-degree RF pulse. (B) The transverse magnetization vector precessing in xy -plane due to the applied magnetic field. The purple coil along the x -axis shows the receiver antenna. (C) As the transverse magnetization vector precesses, it also decays in amplitude due to spin-spin interactions within the tissue. This induces a damped oscillatory signal in the receiver coil, which is known as the free induction decay (FID). Figure adapted from Bushberg et al. [27].

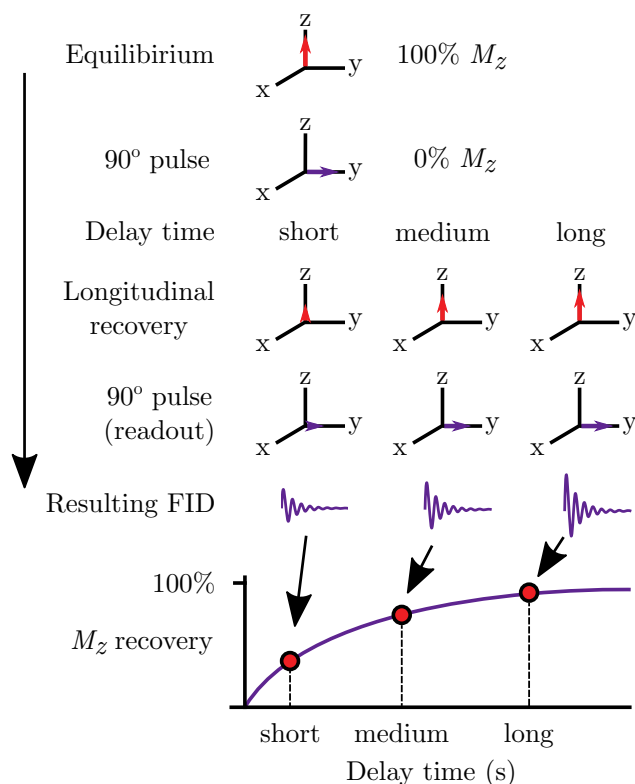


Figure 1.14: Measurement of T_1 relaxation. In order to indirectly measure the spin-lattice relaxation, a series of 90-degree RF pulses, each separated with a known delay, are applied and the corresponding peak FID amplitudes are recorded. The T_1 time is then obtained by fitting a rising exponential to the previously obtained points as a function of the delay time. The figure shows three such points obtained using short, medium and long delay times. The symbol M_z denotes the recovered value of the longitudinal magnetization (i.e. peak value of the respective FID). Figure adapted from Bushberg et al. [27].

which results into its smaller overlap with the tumbling frequency of neighboring atoms and increase in the T_1 relaxation time. One way to expedite the T_1 recovery process in tissues is to use Gadolinium chelated complex macromolecules. These contrast agents modify the local magnetic properties by forming a hydration layer, thereby reducing the effective T_1 relaxation time [27].

Although the recovery of the longitudinal magnetic moment takes place simultaneously to the decay of transverse magnetization, the former process is approximately 5–10 times longer than the latter one. Moreover, since the change in longitudinal magnetization can not be measured directly, a specific RF pulse sequence is used to determine the T_1 relaxation time (figure 1.14).

1.4.6 Acquisition Parameters

During the acquisition of MRI data, two main parameters determine the physiologic nature (T_1 -weighted, proton density and T_2 -weighted) of the recorded images: the time of repetition (TR) and the time of echo (TE).

TR is the time between the subsequent 90-degree RF pulses. During this period, both the previously mentioned processes (T_2 decay and T_1 recovery) take place in the tissues. To suppress the effects of magnetic field inhomogeneity and other external factors, after a certain delay time (TE/2) from the first RF pulse, a 180-degree inversion pulse is applied and the resultant echo signal is recorded. TE is the elapsed time between the 90-degree RF pulse and the induction of the peak echo signal.

1.4.7 T_1 -weighted Sequence

A T_1 -weighted spin-echo sequence produces an image reflective of the T_1 characteristics of different tissues by minimizing the contributions in the recorded signal from other magnetic resonance phenomena. The T_1 effects are maximized by selecting a short TR. When 90-degree RF pulses are applied within a short time interval (i.e. short TR), the longitudinal magnetization vectors of different tissues do not have enough time to recover, resulting in an enhanced T_1 contrast. Likewise, the T_2 effects are minimized by choosing short TE values. The application of successive 180-degree inversion pulses within a short period of time refocuses the transverse magnetization vector and eliminates the T_2 decay contributions. In a typical T_1 -weighted brain image (figure 1.15A), fat has the highest image intensity because of the short T_1 recovery time. The remaining tissues with a decreasing order of intensity in T_1 image are: white matter, gray matter and cerebrospinal fluid.

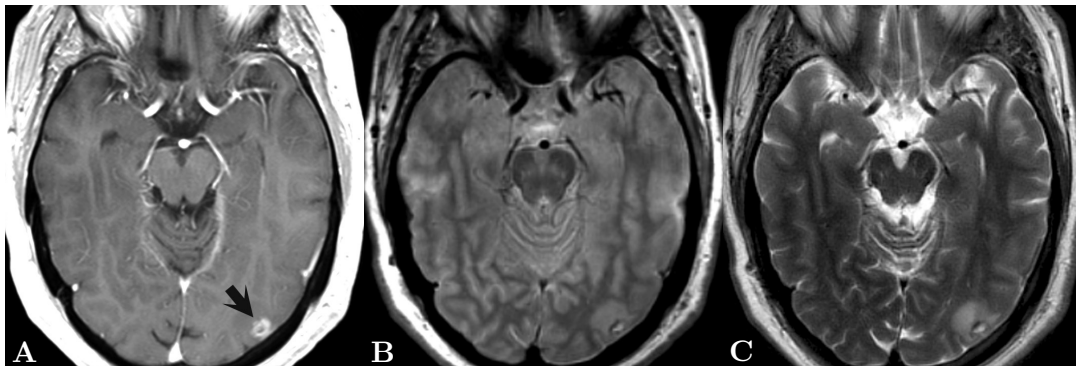


Figure 1.15: (A) T_1 -weighted, (B) proton density and (C) T_2 -weighted images of a patient. The black arrow in panel A indicates towards a metastatic lesion from lung carcinoma. Figure adapted from Runge et al. [49].

1.4.8 Proton Density Weighting

Intensity in proton density weighting is proportional to the total number of hydrogen atoms present in each voxel of the image. This is achieved by suppressing both T_1 and T_2 relaxation contributions. The T_1 recovery contrast is minimized by choosing a long TR and allowing the longitudinal magnetization component of all the tissues to recover, whereas the T_2 decay contrast is masked by selecting short TE values as mentioned in the previous subsection. The signal-to-noise ratios in proton density images are among the highest (in contrast to the T_1 and T_2 weighted images), but the overall soft-tissue contrast between different types of tissues is low (figure 1.15B).

1.4.9 T2-weighted Sequence

The T_2 -weighted image is obtained by accentuating the T_2 decay contrast and suppressing the T_1 recovery effects. Thus, the T_2 -weighted spin-echo sequence uses a long TR and a long TE. Among all the three images, T_2 -weighted image generally exhibits maximum contrast between different tissues (figure 1.15C).

Common TR and TE values used for all the three pulse sequences are detailed in table 1.3.

1.4.10 Signal Localization

Spatial localization of magnetic resonance signals is an important step in MRI and is achieved through a set of gradient coils producing linearly varying magnetic fields. The gradient magnetic fields are superimposed on the static magnetic field to induce position-dependent changes in the homogeneous magnetic field. As the precession frequency of protons depends on the strength of the external magnetic field (sub-section 1.4.2), the positional differences in the strength of the static magnetic field causes hydrogen nuclei at these locations to precess at varying frequencies. This allows to spatially select and readout magnetic resonance signals across the entire field of view by applying RF pulses pertinent to the precession frequencies of the hydrogen nuclei at different locations.

The three different gradients used for localizing magnetic resonance signals are: slice-selection, and frequency and phase encoding. The slice-selection gradient is applied along the axial direction (z-axis) of the MRI scanner. Upon activation, it causes the hydrogen nuclei along the z-axis to precess at linearly increasing or decreasing frequencies. This enables to select and perturb different axial slices by applying RF pulses of the respective frequency bands (figure 1.16). The thickness of axial planes (spatial resolution in z-direction) is determined by the slope of the slice-selection gradient and the bandwidth of the RF pulse transmitter [27]. The frequency encoding gradient is applied orthogonally to the slice-selection gradient along the x- or y-axis of the scanner's coordinate system. It is also termed as the readout gradient because it is activated during the rise and decline of the induced echo signal. Similar to the slice-selection gradient, the frequency encoding gradient alters the precession frequencies enabling a position dependent

Parameter	<i>T</i> 1-weighted	Proton density	<i>T</i> 2-weighted
TR (ms)	400-600	2,000-4,000	2,000-4,000
TE (ms)	5-30	5-30	60-150

Table 1.3: Common TR and TE values used for recording different magnetic resonance images. Data taken from Bushberg et al. [27].

dent transfer and readout of RF energy. The location of the third dimension is obtained through the phase encoding gradient. It is applied along the third dimension after the slice-selection gradient and before the frequency encoding gradient. It is called phase encoding gradient because it produces spatially encoded phase shifts across the entire field of view by incrementing the gradient strength during each TR period [27]. A typical spin-echo sequence is shown in figure 1.17, where all the three gradients are applied sequentially together with the 90-degree RF pulse to record spatially resolved magnetic resonance signals.

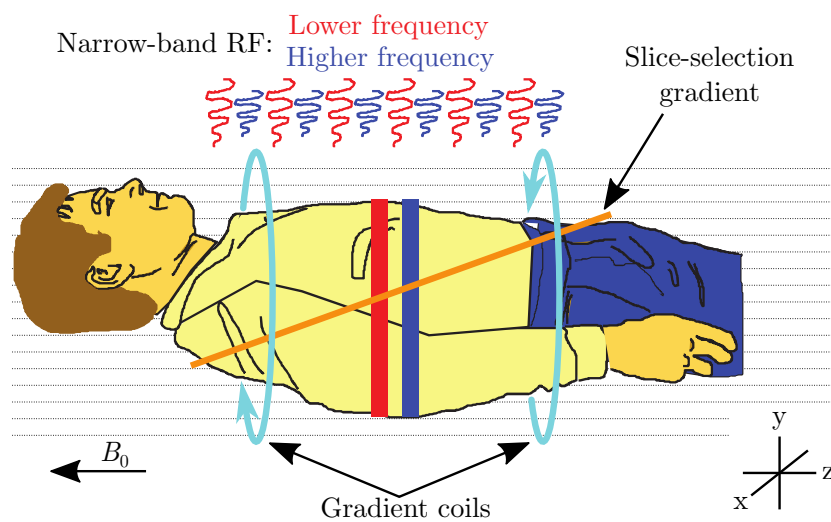


Figure 1.16: A schematic depicting the functioning of the slice-selection gradient. The magnetic field gradient is applied to alter the precession frequency of hydrogen nuclei along the z-axis. This allows to transfer and readout energy from different axial planes by applying RF pulses of the respective frequencies. In this figure two such slices are shown. The lower frequency RF pulse selects the red slice, whereas the higher frequency RF pulse selects the blue slice. Figure adapted from Bushberg et al. [27].

1.4.11 K-Space

The acquired MRI data is transformed into frequency domain and stored in a matrix as complex numbers. The elements of the so called k-space matrix (figure 1.18) represent the positive and negative spatial frequency components of the image. The data along the x- and y-axis are filled with the frequency and phase encoding gradients, respectively. Once the entire matrix is filled, the inverse Fourier transform is applied on the k-space data to reconstruct the final spatial domain im-

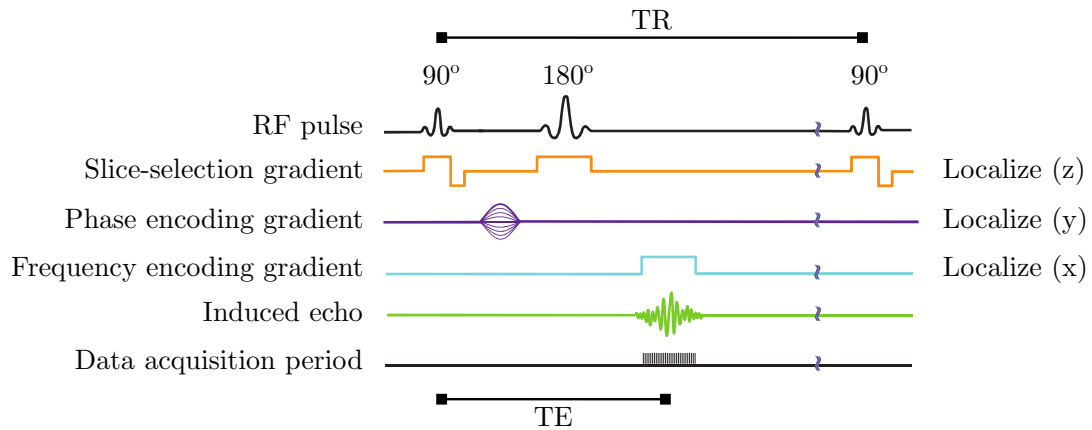


Figure 1.17: A complete spin-echo pulse sequence. The slice-selection, phase encoding and frequency encoding gradients are applied to spatially localize the induced echo signals. The illustrated TR period is repeated several times, each with a different strength of the phase encoding gradient (represented with a set of lines for the phase encoding gradient). Figure adapted from Bushberg et al. [27].

age. Due to the symmetric nature of the matrix, various schemes can be utilized to record only the partial data and expedite the acquisition process [50].

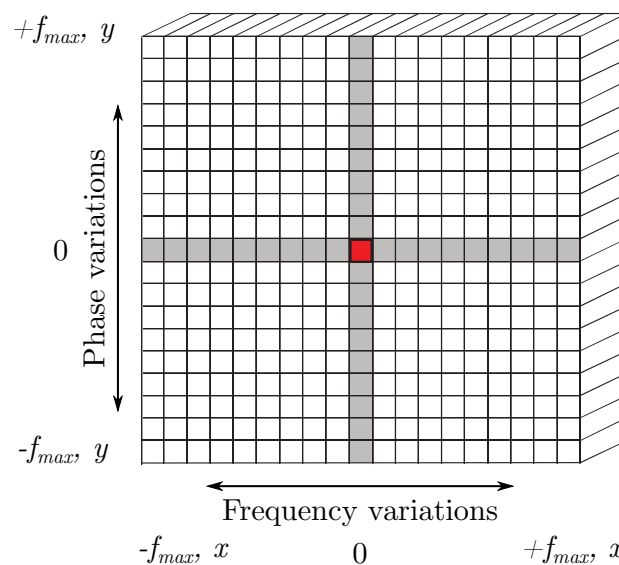


Figure 1.18: The k-space matrix used for storing the raw MRI data. The matrix is symmetric about the centers of x- and y-axis with frequency varying from $-f_{max}$ to $+f_{max}$. Phase and frequency encoding gradients are used to fill the matrix columns and rows, respectively. Figure adapted from Bushberg et al. [27].

1.4.12 Diffusion Weighted MRI

Diffusion weighted (DW) MRI is a functional imaging technique that develops image contrast based on the movement of water molecules. The magnitude of water

movement measured by DW-MRI is proportional to the average distance traveled by thermally driven protons within a specific period of time [51]. As benign and malignant tissues have differing levels of cellularity, the degree of water diffusion in these tissues is also variable. Furthermore, various cellular events, such as apoptosis, necrosis and edema result in either an increase or decrease of the extracellular space, which influences the free diffusion of water molecules specific to these phenotypes. Due to these reasons DW-MRI has become a well-established tool for the diagnosis and monitoring of many cancer types in the clinic.

One approach to quantify DW-MRI measurements is by calculating apparent diffusion coefficient (ADC) maps. The quantification is intended to remove the effects of the T_1 and T_2 relaxation times and provide scanner independent values of ADC based on the following equation [52]:

$$ADC = \frac{\ln\left(\frac{S_0}{S_1}\right)}{(b_1 - b_0)}.$$

Here S_0 and S_1 are the signal intensities measured using the b value b_0 and b_1 , respectively. A b value defines the degree to which the measured magnetic resonance signal is sensitized to diffusion effects [52]. It depends on different properties of the applied gradient pulses and is calculated as:

$$b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3).$$

Where γ is the gyromagnetic ratio of a proton. The parameters G , δ and Δ are the magnitude, duration and time interval of the applied gradient pulses. At very low b values (<200 s/mm²) the MRI signal is sensitive to perfusion contributions (figure 1.19A) from large blood vessels [53]. Above 200 s/mm², the perfusion component disappears and mainly diffusion is measured⁶(figure 1.19 B-C). In practice, MRI data acquired from multiple b values (ranging from 200 until 1000 s/mm²) and along several directions are used to calculate the ADC map.

⁶ At b values >200 s/mm² small blood vessels with less flow may still contribute to the measured signal [53].

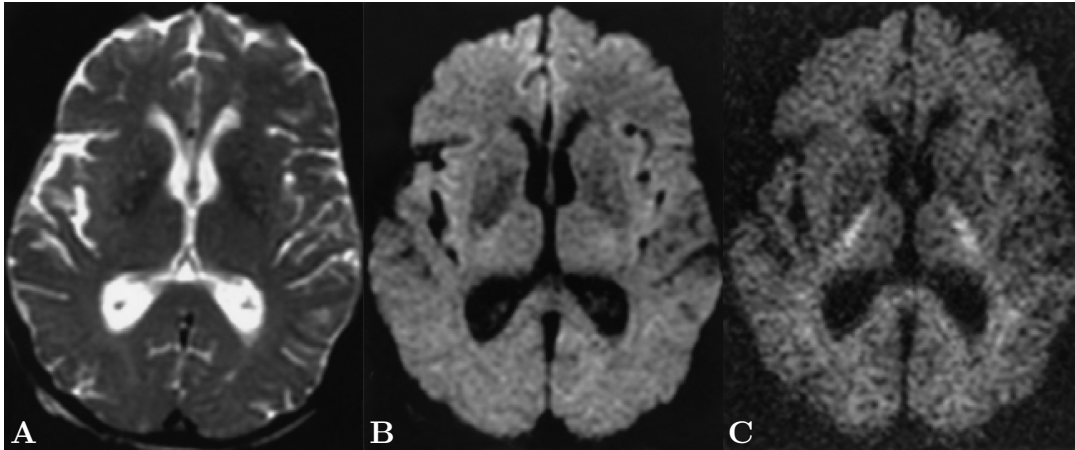


Figure 1.19: Diffusion-weighted images acquired using a b value (s/mm^2) of (A) 0, (B) 1000 and (C) 3000. Figure adapted from Burdette et al. [54].

1.5 Simultaneous PET/MRI

1.5.1 Concept

Simultaneous PET/MRI seeks to acquire multiparametric functional and anatomical information concurrently without the necessity of an additional image registration step, as performed in the sequential hybrid imaging studies [55]. The first hybrid device to acquire sequential functional and anatomical information was developed using PET and computer tomography (CT), and therefore was termed as PET/CT [56]. The anatomical references provided by CT significantly improved the localization of malignant lesions and consequently, quantification of the associated PET data, leading to a better diagnosis and staging of patients with cancer [57–59]. In addition, contrast enhanced CT can provide further evidence about perfusion and vascularization in tumors to better stratify the patients in relevant clinical categories [60].

However, a major downside of CT is the amount of radiation given to the patient [61]. Although the radiation dose does not pose a big issue in single clinical examinations, it can alter experimental conditions in longitudinal clinical and pre-clinical studies. Besides this, the limited soft tissue contrast of CT results in its low sensitivity and specificity for certain tumor types as compared to MRI [62]. Since the functioning of MRI system is based on the magnetic properties of hydrogen

nuclei, the MRI examinations do not involve exposure to radiation. Moreover, the advancements in multiparametric imaging now allow an accurate *in vivo* monitoring of various functional aspects of the tumors, such as diffusion, perfusion, vascular permeability and the metabolic profile [63]. Due to these reasons, the combination MRI and PET not only provides a precise morphological reference, but also a plethora of complementary functional information [55, 64].

1.5.2 Development and Applications

Unlike PET/CT, combining PET and MRI into one scanner in the late 1990s was a non-trivial task. In an ideal hybrid scanner enabling simultaneous data acquisition, each imaging modality should operate without influencing the performance of the other techniques. However, PET/MRI scanner carried two main sources of mutual interference [65]: inhomogeneities produced in the static magnetic field due to PET electronics and malfunctioning of PET electronics due to the static magnetic field, switching MRI gradients and RF pulses. The very first hybrid PET/MRI approach [66] therefore tried to separate the interfering components of both the modalities by placing the PMTs and other front-end electronics outside the fringe magnetic field and using optical fibers to direct the scintillation light towards the PMTs. Although this setup enabled simultaneous imaging, the use of optical fibers caused significant damping ($\sim 90\%$) of the scintillation light and a considerable reduction in the energy resolution.

The incompatibility of the PET instrumentation (in particular PMTs) was resolved with the development of avalanche photo-diode (APD) based PET detectors [67], which were unaffected by the presence of static or varying magnetic fields. However, due to small amplification factor of these detectors, a preamplifier was required in the front-end electronics. Wu et al. [68] managed to minimize the magnetic field inhomogeneities caused by the advanced front-end electronics by connecting the scintillators and APDs with small optical fibers and shifting the interfering electronics outside the scanner's field of view. The more recent MRI compatible PET systems [69] instead comprise Geiger-mode APD arrays, which

are also known as silicon PMTs. In a Geiger-mode APD, each incident photon can trigger a self sustaining avalanche breakdown (multiplication of electron–hole pairs) through the process of impact ionization, providing a gain similar to that of a conventional PMT. Moreover, like linear-mode APDs⁷, Geiger-mode APDs are also insensitive to external magnetic fields. They operate at relatively low bias voltages (30-100 V) and provide high amplification factors ranging between 10^5 and 10^6 . Besides their advantages over APDs and PMTs, there are two important characteristics of Geiger-mode APDs which can affect their normal operation: the amount of dark current generated due to increased temperature and optical crosstalk, and a highly sensitive temperature dependent gain profile. The latter attribute is particularly crucial for PET inserts intended to operate in MRI scanners, as frequent switching of gradient coils and application of RF pulses can lead to wide temperature variations inside the scanner. For this reason, a temperature stabilizing circuit is of high importance for silicon PMT based MRI compatible PET inserts [70]. To avoid electromagnetic interferences and production of eddy currents (due to variations in magnetic field), the front-end electronics are shielded using materials with specific characteristics: high density, low magnetic susceptibility and low conductivity [65].

In contrast to the adjustments required in the PET hardware, only minimal changes are needed in the MRI scanner, mainly because the majority of the prototype PET/MRI systems [65] are based on a modified PET scanner that can be fitted inside the bore of existing MRI devices. Nonetheless, efforts to optimize the data quality for both the modalities are ongoing, such as the development of PET compatible low attenuation MRI coils [71], MRI assisted attenuation correction for PET [72, 73], joint reconstruction methods for PET and MRI [74] and motion correction techniques for simultaneous imaging [75]. An example of one such method is presented in figure 1.20, where an MRI-based tissue segmentation approach [76] was used to correct the PET data for attenuation.

The complementary data obtained from PET/MRI has been shown to provide

⁷ These operate below the breakdown voltage and provide gains typically two or three orders of magnitude smaller than that of Geiger-mode APDs.

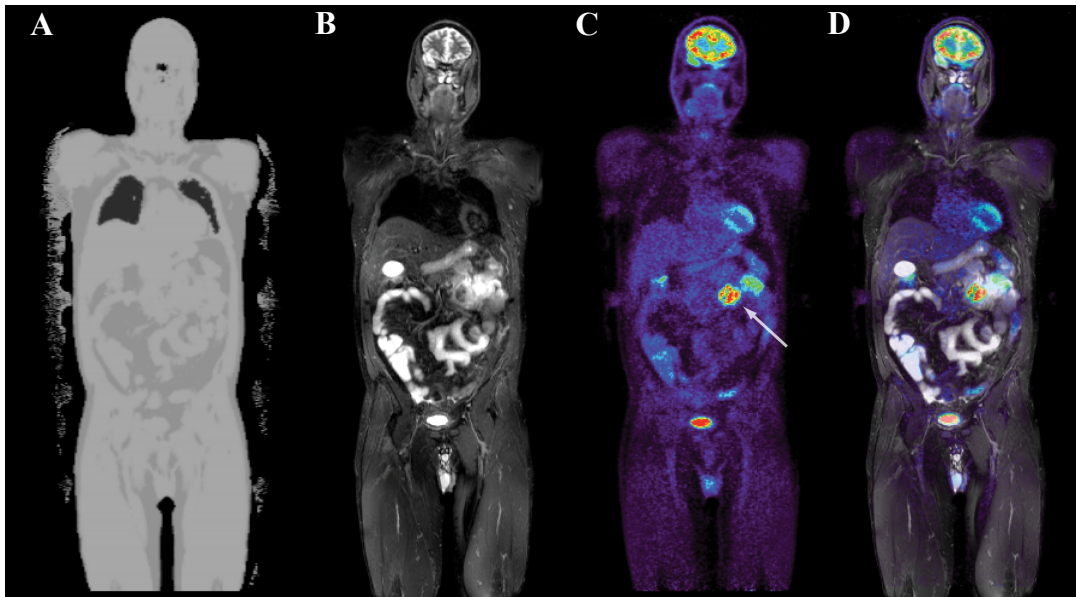


Figure 1.20: PET/MRI of a melanoma patient. (A) MRI derived attenuation map and (B) T_2 -weighted anatomy image, (C) ^{18}F -FDG PET scan, (D) and an overlay of both the previous images. The arrow in panel C indicates an enteric melanoma metastasis. Figure adapted from Disselhorst et al. [65].

a significant prognostic relevance for a wide range of applications in oncology. These include whole body staging for different types of cancer, identification of intracranial and thoracic neoplasms, localization of metastatic lesions, delineation of brain tumors and assessment of inter- and intra-tumor heterogeneity [77–79]. In the preclinical field the scope of integrated PET/MRI is broader than the clinic, as animal experiments have more flexibility in terms of scan-time and experimental protocols. A comprehensive summary of preclinical PET/MRI applications is provided by Judenhofer et al. [80].

This thesis primarily focuses on using machine learning methods to analyze multiparametric PET and MRI data, and characterize phenotypic heterogeneity in subcutaneous tumor models. The following section provides a summary of the algorithms that have been used in this work.

1.6 Machine Learning

Machine learning refers to the process of identifying structure or inferring a function from given data to perform deductive tasks, such as clustering, classification

and regression. Depending on the type of data, learning can be performed in two ways: supervised and unsupervised.

In case of supervised learning, labeled data (also known as a training set) is used to tune the parameters of an adaptive model. Each observation (or data-point) in the training set is a pair consisting of input predictors and an output variable. In the setting of medical imaging, for example, the input predictors—which are also referred to as features or measurements, can be the parameters acquired from an imaging modality and the output variable (often termed as the target or response variable) can be the survival data of patients or lesions labeled as malignant or benign. In general, both the input predictors and the output variable can be either continuous or categorical. A supervised learning algorithm determines an association between the input and output data that can be generalized for unseen observations (data-points not in the training set) to accurately predict the corresponding target variable. More formally, if the training set consists of D predictors, X_1, \dots, X_D (jointly referred to as X) and a numeric output variable Y , it is assumed that X and Y are related as following:

$$Y = f(X) + \epsilon.$$

Where f is an unknown function and ϵ is the irreducible error, which does not depend on X . The learning algorithms aim to estimate f using the observations provided in the training set [81]. The estimated function is subsequently used for either prediction (regression and classification) or inferring the relationship between each predictor and the target variable. Supervised learning is used in wide range of day to day applications, such as online recommendation systems, spam filtering, credit scoring algorithms, stock market analytics, and pattern recognition tools based on audio and video signals.

In unsupervised learning only the input predictors are available without any labels or target variable. Hence, the objective in unsupervised learning is to discover intrinsic regularities in the data, different from what would be regarded as random contributions of noise [82]. The two most intuitive examples of unsupervised learning are clustering and dimensionality reduction. In clustering, the input ob-

servations are grouped into disjoint clusters to minimize the intra-cluster variance and maximize the inter-cluster variance. Whereas, dimensionality reduction techniques aim to reduce the number of input predictors to a smaller subset that can be used either for visualization or as a pre-processing step prior to supervised learning [83].

Due to the lack of labeled data, the problem of unsupervised learning is much more challenging than supervised learning. Moreover, there are no standard procedures of parameter selection and model validation, which makes the learning process highly subjective. Despite these factors, unsupervised learning methods have witnessed a growing interest in areas where obtaining labels is either expensive or simply infeasible, such as medical imaging, genetics, anomaly detection and remote sensing.

A considerable part of the research work presented in this thesis was carried out using well known unsupervised and supervised learning methods as fundamental building blocks. A brief summary of each one of the methods is provided in the following subsections. Unless specified, all derivations in the subsequent sections are taken from Bishop [83].

1.6.1 *K*-means Clustering

K-means is perhaps the most classic of all clustering methods. Sometimes also referred to as Lloyd's algorithm, it was originally proposed by Stuart P. Lloyd at Bell Laboratories in 1957, however it was officially published 25 years later in 1982 [84]. *K*-means is an iterative partitional clustering algorithm that divides input observations into a user provided number of clusters, *K*. The clustering process iterates between two stages, until the algorithm converges and no further improvements in objective function can be achieved. This can be best explained using the formal description provided below.

Let's assume the clustering problem in a multidimensional space, where the input dataset $\{x_1, \dots, x_N\}$ comprises *N* points of a *D*-dimensional random variable *x*.

The clustering objective is to group these N observations into K clusters in such a way that within-cluster point-to-point Euclidean distances are smaller as compared to the inter-cluster point-to-point distances. Let's also introduce a set of D -dimensional vectors μ_k , where $k = 1, \dots, K$, in which μ_k can be considered as the centroid of the cluster k . Thus, during clustering the data points are assigned to different clusters based on their closest centroid. The cluster assignments can be formally represented using an indicator variable $r_{nk} \in \{0, 1\}$, which encodes whether observation x_n is included in cluster k or not. Using these notations, we can define the cost function J as follows:

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2.$$

Which is simply the sum of the squared distances between all observations and their closest centroids. In K -means clustering we minimize J by selecting the most optimal values for the variables r_{nk} and μ_k . The optimization is performed iteratively, where each iteration consists of two steps. To begin with, the variable μ_k is assigned some initial values (either random or user provided). Afterwards, in the first step, μ_k is kept fixed and the cost function J is minimized with respect to r_{nk} . Since J and r_{nk} are linearly related, this corresponds to assigning each point to its nearest centroid as described below:

$$r_{nk} = \begin{cases} 1 & \text{if } k = \arg \min_j \|x_n - \mu_j\|^2 \\ 0 & \text{otherwise.} \end{cases}$$

In the second step, the r_{nk} values determined previously are held fixed and the cost function J is minimized with respect to μ_k . Thus, setting the derivative of J with respect to μ_k equal to zero provides:

$$2 \sum_{n=1}^N r_{nk} (x_n - \mu_k) = 0$$

which is solved for μ_k to obtain:

$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}.$$

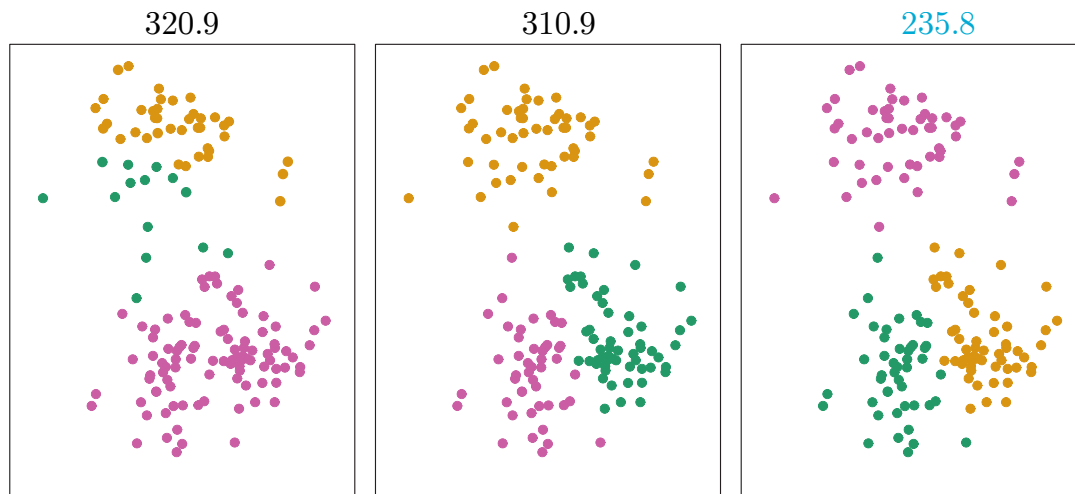


Figure 1.21: K -means clustering performed multiple times on a synthetic dataset. In each run a different random initial configuration for cluster centroids was used. The final cost function of each run is indicated above the plot. In contrast to the first and second, the third configuration resulted in a lower cost (highlighted in cyan) and well assigned cluster memberships. Figure adapted from James et al. [81].

This implies that in each iteration μ_k should simply be assigned as the mean of all observations in cluster k . The two steps of cluster assignment and centroid update are performed repeatedly until there is no (or a very small) change in the cost function J and the algorithm converges to a local or global minimum.

It should be noted that the clustering solution obtained using K -means is dependent on the initial values of the cluster centroids. A poor initialization may result in suboptimal results, because the optimization might converge to a local minimum. In practice (when there is no *a priori* information available about cluster centroids), the algorithm is run several times, each time with a different random initial configuration of cluster centroids and the solution corresponding to the lowest cost is considered as best (figure 1.21) [81].

The simplicity of Euclidean distance based K -means clustering comes at the cost of certain drawbacks. First, it assumes that the input predictors are uncorrelated and the clusters have a spherical shape [85]. This assumption provides reliable results only as long as the clusters are compact and well isolated from each other. However, this issue can be alleviated to some degree by choosing a dissimilarity measure that takes into account the covariance matrix of the data, such as the regularized Mahalanobis distance [86]. A second limitation of K -means is that in

the cluster assignment step, each observation is uniquely assigned to only one of the clusters. Such hard assignments can lead to incorrect results, especially in cases when the clusters overlap with each other. For problems like this, a soft clustering approach is more appropriate, as it would provide an uncertainty measure over the intermediate and final cluster assignments. Fuzzy C-means (FCM), an extension of K -means clustering improves upon this issue by probabilistically assigning each observation all clusters. For the details of FCM clustering the reader is referred to the original article that proposed the algorithm [87].

1.6.2 Gaussian Mixture Modeling

A Gaussian mixture model (GMM) is a collection of two or more uni-/multi-variate Gaussian distributions, whose joint probability density function is the weighted sum of the probability density functions of the independent Gaussian components [83]. A multi-variate Gaussian distribution is characterized using a mean vector μ and a covariance matrix Σ in the following manner:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) \right\},$$

where $|\Sigma|$ is the determinant of Σ and D is the dimension of the mean vector μ . As real datasets are often multimodal, a single Gaussian distribution is insufficient to capture the underlying structure and fit the data faithfully. A linear combination of several Gaussians, however, can be adjusted to generate complex probability density functions (figure 1.22) and model multimodal datasets more accurately.

A GMM with K multi-variate Gaussian distributions takes the following mathematical form:

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x|\mu_k, \Sigma_k).$$

Here, π_k is the mixture density coefficient of the normal distribution $\mathcal{N}(x|\mu_k, \Sigma_k)$, which is characterized using a mean vector μ_k and covariance matrix Σ_k . Each Gaussian density function is also termed as a component of the GMM. The mix-

ture density coefficients are constrained to fulfill the probability requirements:

$$\sum_{k=1}^K \pi_k = 1 \quad \text{and} \quad 0 \leq \pi_k \leq 1.$$

The GMM described above is characterized by three variables π_k , μ_k and Σ_k , whose optimal values are chosen during model fitting. One option to select the optimal parameters is by maximizing the log-likelihood function:

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k) \right\},$$

where, $X = \{x_1, \dots, x_N\}$. However, due to the second summation being contained in the logarithm, obtaining a closed form solution using maximum likelihood is not trivial [83, 88]. Another factor limiting the use of maximum likelihood estimation is the presence of singularities [83]. A singularity might arise during model fitting, when the mean of one of the Gaussian components matches exactly with an input observation (figure 1.23). This causes the contribution in the log-likelihood function from this Gaussian component to diverge due to a constantly increasing additive value, eventually resulting in a poorly fitted model [83].

A systematic and more reliable approach for obtaining the GMM parameters is the EM algorithm. EM is a powerful statistical technique that is widely used to find the maximum likelihood estimates of the parameters in models with hidden variables. The method alternates between expectation (E) and maximization (M) steps, which correspond to specifying a lower bound on the log-likelihood function

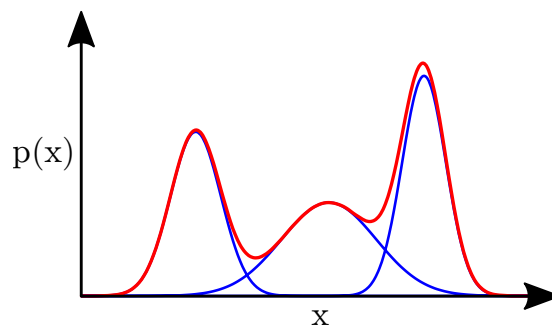


Figure 1.22: A one dimensional mixture of Gaussians. The curve in the red shows the joint probability density function obtained after linear superimposition of the three one-dimensional Gaussian distributions (shown in blue). Figure adapted from Bishop [83].

and raising this bound over each iteration [88]. The derivation of the EM algorithm for a mixture of Gaussians is provided below.

First the derivatives of the log-likelihood function with respect to the model parameters (π_k , μ_k and Σ_k) are set to zero to fulfill the maximum likelihood conditions. Starting with μ_k :

$$0 = - \sum_{n=1}^N \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\underbrace{\sum_j \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)}_{\gamma(z_{nk})}} \sum_k (x_n - \mu_k).$$

Here $\gamma(z_{nk})$ are the posterior probabilities. After organizing the terms in the equation above, we obtain:

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n.$$

Where N_k is defined as:

$$N_k = \sum_{n=1}^N \gamma(z_{nk}).$$

Likewise, equating the derivatives of the log-likelihood function with respect to Σ_k and π_k to zero, we obtain⁸:

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T$$

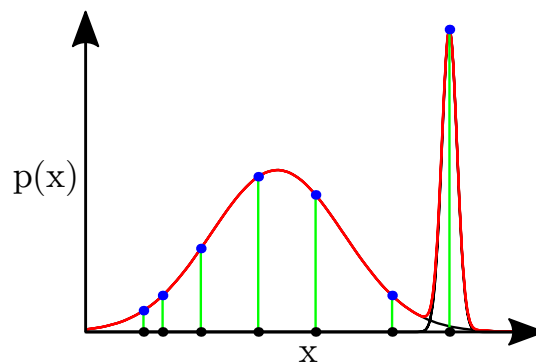


Figure 1.23: Singularities can arise during the fitting of a GMM, when the mean of one of the Gaussian components becomes equal to a data point. In such a scenario, the model parameters estimated using maximum likelihood often lead to a poor description of the data. Figure adapted from Bishop [83].

⁸ Obtaining the value of π_k is a more rigorous exercise and is beyond the scope of this thesis. Complete details can be found in Bishop [83].

and,

$$\pi_k = \frac{N_k}{N}.$$

These solutions are not closed-form, as they depend on $\gamma(z_{nk})$ in an intricate manner. Therefore, the EM algorithm uses the following steps [83] to find a set of parameters that maximize the log-likelihood function.

1. Initialize the parameters (π_k , μ_k and Σ_k) with random (or user provided) values.

2. **E step:** compute the posterior probabilities $\gamma(z_{nk})$:

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)}.$$

3. **M step:** update the parameters (π_k , μ_k and Σ_k) based on $\gamma(z_{nk})$:

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

and,

$$\pi_k^{new} = \frac{N_k}{N}.$$

Where,

$$N_k = \sum_{n=1}^N \gamma(z_{nk}).$$

4. Estimate the log-likelihood using new parameters (π_k^{new} , μ_k^{new} and Σ_k^{new}):

$$\ln p(X | \pi, \mu, \Sigma) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k^{new} \mathcal{N}(x_n | \mu_k^{new}, \Sigma_k^{new}) \right\}.$$

If the stopping criterion (for example, minimum change in the log-likelihood) is not met, repeat the next iteration from step 2.

It is interesting to point out that the cluster assignment and centroid update steps in K -means algorithm have close resemblance with the EM algorithm. While the

former step bears a close similarity with the E phase described above, the latter step is analogous to the parameter update procedure in the M phase. As a matter of fact, it can be shown that K -means algorithm is a special case of modeling a mixture of Gaussians using the EM algorithm [83]. An illustrative example of EM iterations during the fitting of a GMM is shown in figure 1.24.

It should be noted that if several local maxima are present in the log-likelihood function, the EM algorithm only guarantees convergence to one of the maxima, which may not be the global maximum. Moreover, much alike K -means, the clustering solution obtained using a GMM depends upon the initial configuration of the model parameters. Therefore, a common practice in fitting a GMM is to first run K -means clustering repetitively in order to obtain reasonable estimates of the initial parameters. This initial configuration is subsequently fed into the EM algorithm [83].

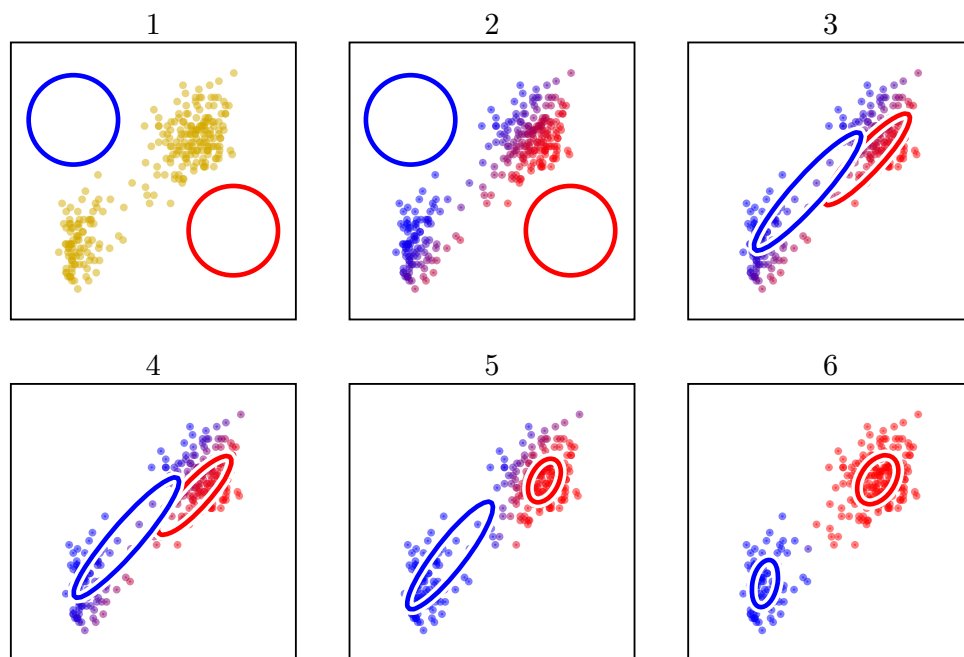


Figure 1.24: EM iterations (1-6) during the fitting of a GMM on a two dimensional dataset. Notice the probabilistic assignment (depicted by purple color) of points in both the clusters. Figure adapted from Bishop [83].

1.6.3 Spectral Clustering

Spectral clustering is a graph based clustering algorithm that represents the data points as the vertices of an undirected similarity graph $G = \langle V, E \rangle$ [89]. Here, V is the set of vertices or nodes and E is the set of edges. An edge between two vertices exists, if the similarity between them is positive. Furthermore, all the edges in graph G are weighted by the similarity between their respective vertices. Thus, the matrix containing the weights of all the edges is termed as the affinity or similarity matrix. A common approach to construct the affinity matrix for a fully-connected graph⁹ (where all the vertices are connected with each other) is by using the radial basis function (RBF) kernel. Given a set of N , D -dimensional input observations $X = \{x_1, \dots, x_N\}$, the $N \times N$ affinity matrix (W) is formed using the RBF kernel in the following manner:

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

Here, $\|x_i - x_j\|$ is the Euclidean distance between observations x_i and x_j . The RBF kernel is characterized using the scale parameter σ , which modulates the reduction in affinity with the distance between observations x_i and x_j [90].

The objective in spectral clustering is to partition the graph in such a way that the weights within a group are large and the weights of the edges connecting different groups are small. By choosing the appropriate scale of the RBF kernel, the affinity matrix is formed in a manner that preserves the local manifolds and maintains the clustering objective [91]. Once the affinity matrix is obtained, the diagonal degree matrix D with a diagonal vector d_i is defined as:

$$d_i = \sum_{j=1}^N W_{ij}.$$

Based on the affinity (W) and degree (D) matrices, various graph Laplacians can

⁹ There are several other ways to construct a similarity graph. The two other most popular graphs are: the ϵ neighborhood graph and the k -nearest neighbor graph [89].

be defined. The unnormalized Laplacian matrix is defined as:

$$L = D - W.$$

Likewise, the two normalized Laplacian matrices are defined as:

$$L_{sym} = 1 - D^{-1/2}WD^{-1/2} \quad \text{and} \quad L_{rw} = 1 - D^{-1/2}W.$$

The first Laplacian is denoted as L_{sym} because it is a symmetric matrix, whereas the second Laplacian as L_{rw} due to its close connection with a random walk. For detailed explanations about the properties of the graph Laplacians, the reader is referred to the comprehensive tutorial on spectral clustering by Ulrike von Luxburg [89].

Once the Laplacian matrix is obtained, clustering is performed by computing its m first eigenvectors (corresponding to the m smallest eigenvalues) and clustering them using K -means. A common practice in spectral clustering is to compute as many eigenvectors as the number of clusters, however this does not apply when the data is noisy and the clusters overlap with each other. In such a scenario, it is recommended to use more eigenvectors than the number of clusters. The choice of graph Laplacian also affects the clustering solution. For stability reasons, in almost all the cases, the normalized Laplacian is preferred over the unnormalized one [89].

Spectral clustering outperforms traditional clustering algorithms (K -means, FCM and Gaussian mixture modeling) primarily because it does not make any strict assumptions about the shape of the clusters. The clustering methods described previously assume the clusters to be spherical or ellipticals, which makes them unsuitable for identifying clusters with non-convex shapes. This is illustrated by the toy example in figure 1.25. Unlike K -means and Gaussian mixture modeling, spectral clustering captures nonlinear manifolds of the spiral dataset by forming an affinity matrix that preserves local distances (even in the presence of noise). Once the appropriate affinity matrix is obtained, it simply solves a linear problem without any initialization or optimization (multiple local maxima) issues and assigns the data points to their respective clusters adequately.

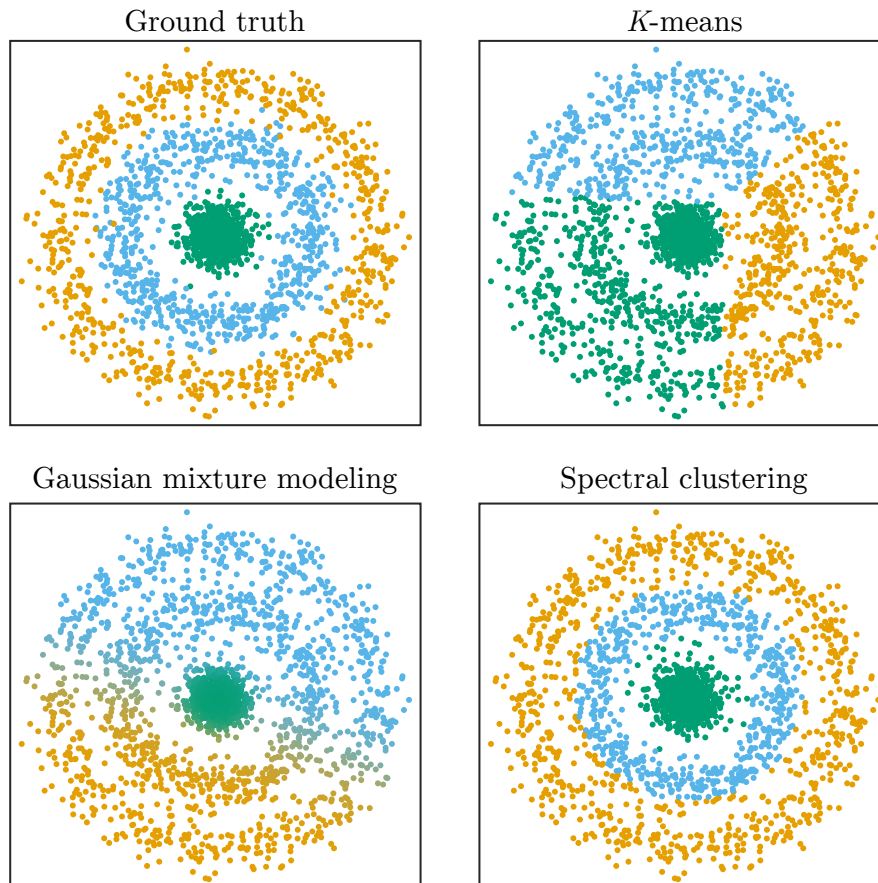


Figure 1.25: Toy example illustrating the effectiveness of spectral clustering. The synthetic data consists of points sampled from three concentric circles corrupted with additive white Gaussian noise. Due to non-convex shape of the clusters, both K -means and Gaussian mixture modeling provide poor clustering results. Contrary to these, spectral clustering retains the shape of all three clusters and results into optimal grouping of points.

An important aspect of forming a fully-connected graph is the choice of σ . Figure 1.26 shows (for the toy dataset presented in figure 1.25) how the variations in σ (too small or too big) can lead to suboptimal clustering results. In particular, one should take into account that for a specific value of σ , the total number of points with a relatively high value of affinity should not be either too many or too few [89]. This is analogous to choosing a suitable value of ϵ and k for the ϵ neighborhood and k -nearest neighbor similarity graphs, respectively.

1.6.4 Random Forest

Random forest is an ensemble based supervised learning method that is widely used for the tasks of classification or regression. Although many attempts were

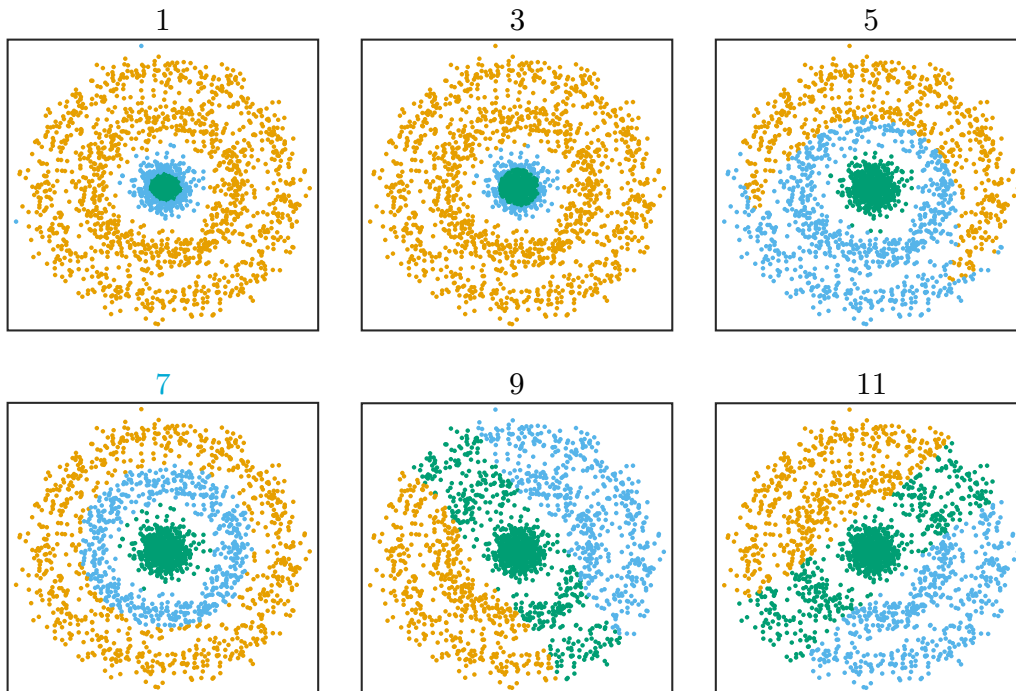


Figure 1.26: Solutions obtained using spectral clustering with variations in σ (indicated above each panel). Overall, a σ value of 7 separated the three clusters best, as matched with the ground truth plot in figure 1.25.

made in the 1990s [92–95] to develop a powerful aggregation based machine learning technique, it was only after Breiman’s paper in 2001 [96] that random forests were proven to achieve significant levels of accuracy both empirically and theoretically. In this thesis, random forests are mainly used for classification and therefore, hereinafter, we will refer to them as a random forest classifier.

A random forest classifier is an ensemble of many decision trees, where each tree plays the role of a weak learner. The trees are termed as weak learners because their classification accuracy is just marginally better than making random decisions [91]. Formally, a decision tree is non-parametric supervised learning tool that classifies observations into different categories by traversing them down the tree; from root node to one of the internal nodes and all the way to terminal nodes or leaves. Prior to classification, the tree is trained using a greedy approach that performs recursive binary splitting in the predictor space at each (except leaf) node of the tree [81]. The training paradigm is termed greedy because each binary split is carried out by choosing an optimal predictor and its corresponding threshold that maximizes a purity function (measuring the fraction

of correctly classified observations) at that particular node. This does not take into account the subsequent splits that depend on the current division and might lead to an overall higher purity. The two most preferred purity measures for growing a decision tree are the Gini index and cross-entropy.

For a dataset with K classes, the Gini index G can be defined as [81]:

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk}).$$

Where, \hat{p}_{mk} is the fraction of data points in the region created at m th node (splits at different nodes can be seen as operations that partition the predictor space into different regions) that belong to class k .

Alternatively, the cross-entropy is defined as [81]:

$$D = - \sum_{k=1}^K \hat{p}_{mk} \log \hat{p}_{mk}.$$

This implies that at the m th node, both the Gini index and cross-entropy will be small if the \hat{p}_{mk} value for all clusters is close to one or zero. Therefore, during training, binary split at each node is performed by choosing a predictor and a threshold that leads to the best separation of input observations into their respective classes and consequently, smallest possible value of the Gini index or cross-entropy.

As opposed to linear models, decision trees are better suited when the relationship between the input predictors and the output variable is non-linear (figure 1.27). But, it is often the case that decision trees do not generalize well on unseen or test datasets. The unreliability exists because decision trees suffer from high variance. This means that even a small variation in the training data can result in very different decision trees making their interpretation and predictions rather precarious. Another major weakness of decision trees is that the decision boundary learned by them during training lacks smoothness. A hard classification boundary can severely degrade the performance of a model, if the observations are expected to have probabilistic associations with different classes [91].

A random forest classifier overcomes these limitations by training a large number of uncorrelated trees and averaging their predictions to obtain probabilistic class labels. Since a deeply grown tree has a low bias and a high variance, averaging many identically distributed trees results into the same bias as that of an individual tree, but a significant reduction in the model variance [91]. The uncorrelated trees are obtained by performing a small trick in the way how each tree is grown. The training paradigm of each tree in a random forest classifier is identical to that of a single decision tree except that in the former case each split is made using only a subset of all predictors. Typically if there are p predictors, during each split $\sim \sqrt{p}$ predictors are taken into consideration. Another important operation that reduces the variance of a random forest classifier is bagging. Bagging is a common procedure in statistical learning that builds a large number training sets from an original (labeled) population by randomly sampling different observations. For a random forest classifier, it usual to perform sampling with replacement and later train each tree on one of the bagged training datasets. Thus, each trained tree in itself is considered as a weak learner because it has been trained only on a subset of all the training observations, however, their combination results in a

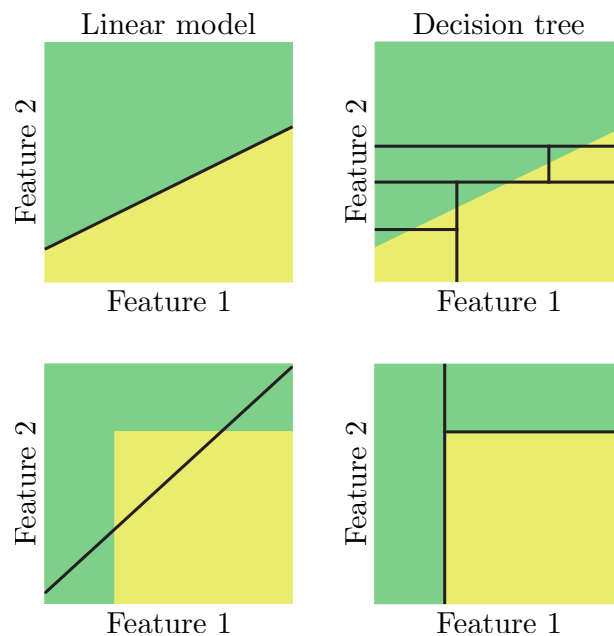


Figure 1.27: Top row: when the decision boundary between the green and the yellow class is linear, the decision tree performs poorly with multiple splits parallel to the axes. Bottom row: for non-linearly separated classes the linear model is outperformed by the decision tree. Figure adapted from [81].

very powerful statistical model that has low bias as well as low variance.

For each tree in a random forest classifier, the data points that have been used in the training are called in-bag observations, whereas the left out points are termed as out-of-bag observations. The out-of-bag observations provide a unique way to estimate the cross-validation error of the trained classifier. It has been shown [96] that bagging results in approximately two-third in-bag and one-third out-of-bag observations. Hence, each observation from the complete training dataset can be tested (by only considering the trees for which it is out-of-bag) on the trained classifier to obtain an estimate of the out-of-bag error. Interestingly, if the number of trees is large enough, the out-of-bag error is identical to leave-one-out cross-validation error [81, 91]. This provides a convenient method of training and validating a random forest classifier.

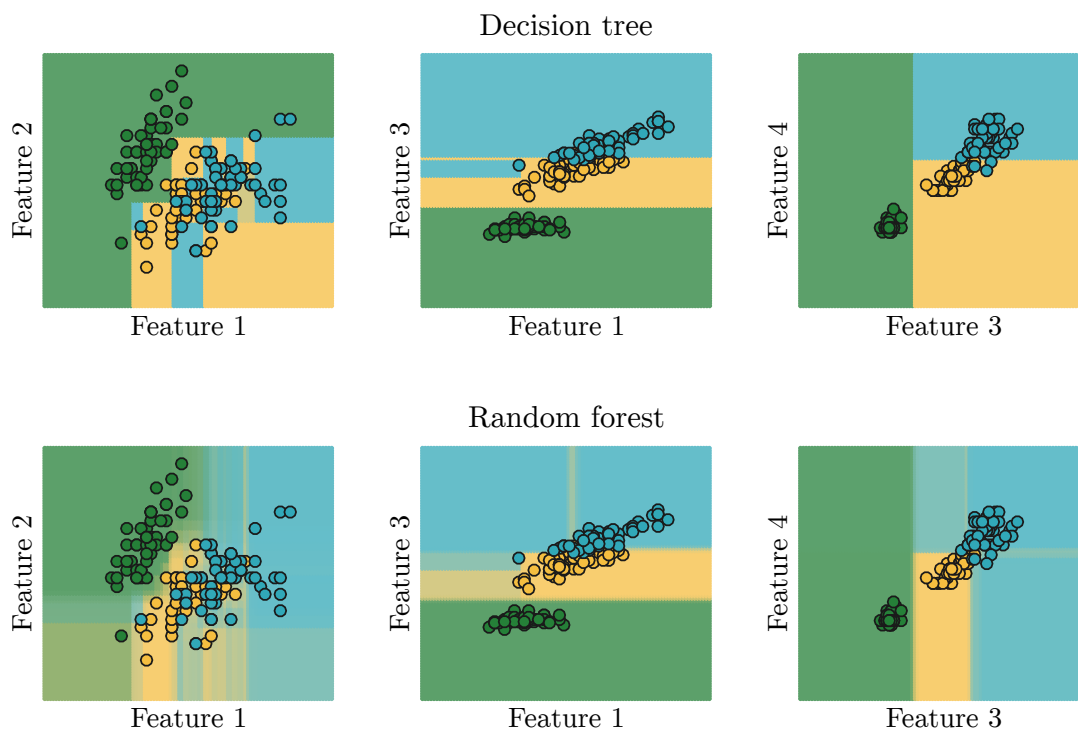


Figure 1.28: A one to one comparison between the decision boundaries learned by a decision tree (top row) and a random forest classifier (bottom row). Both the models were trained on Fisher's Iris dataset [97] that has 3 output classes, 4 predictors and 150 observations. For simplicity and visualization, both the models were trained three times on all the observations, each time with only two predictors (specified on the x- and y- axes). The random forest classifier was trained with 100 trees. It is evident from all three cases that averaging the predictions of many trees results in more accurate and smoother decision surfaces. The three colors in each panel represent the three species (setosa, virginica and versicolor) of Iris flower.

Figure 1.28 compares the prediction surfaces of a decision tree and a random

forest classifier built with 100 trees. As described previously, the random forest classifier not only has a lower model variance than the decision tree (hence more accurate decision boundaries), it also learns smoother decision surfaces that separate input observations into different output classes.

1.6.5 Feature Selection

In machine learning, feature selection refers to the process of identifying relevant predictors that improves model interpretability and generalization. In this sub-section two feature selection algorithms (Laplacian and Fisher scoring) are introduced, both of which are relevant for this thesis. For a detailed review on the topic of feature selection, the reader is referred to Guyon et al. [98].

Laplacian score is a filter based unsupervised feature selection method that selects important features based on their ability to preserve the local structure of the multidimensional data [99]. The algorithm is defined as following:

1. Given a set of N , D -dimensional input observations $X = \{x_1, \dots, x_N\}$, compute the $N \times N$ affinity matrix (W), diagonal degree matrix (D) and unnormalized Laplacian (L) as described in sub-section 1.6.3.
2. For the r th feature, compute the vector \hat{f}_r :

$$\hat{f}_r = f_r - \frac{f_r^T D \mathbf{1}}{\mathbf{1}^T D \mathbf{1}} \mathbf{1}$$

where,

$$f_r = [f_{r1}, \dots, f_{rN}]^T \quad \text{and} \quad \mathbf{1} = [1, \dots, 1]^T.$$

3. Calculate the Laplacian score (L_r) of the r th feature using the following expression:

$$L_r = \frac{\hat{f}_r^T L \hat{f}_r}{\hat{f}_r^T D \hat{f}_r}.$$

The rationale behind the Laplacian scoring algorithm is related with the applica-

tion of the graph similarity matrix, as discussed in sub-section 1.6.3. The algorithm mainly favors the features that preserve the local graph structure captured by the similarity matrix and penalizes the others [99].

Fisher score, in contrast, is a supervised feature selection algorithm, which selects a subset of features that minimize the point-to-point distances within the same class and maximize the point-to-point distances between different classes [100]. The Fisher score F_r for the r th feature is computed as:

$$F_r = \frac{\sum_{k=1}^K n_k (\mu_k^r - \mu^r)^2}{\sum_{k=1}^K n_k (\sigma_k^r)^2}.$$

Where K is the total number of classes and n_k denotes the total number of points in class k . The mean of feature r for all points is denoted as μ^r , and the mean and standard deviation of all points in class k corresponding to feature r are denoted by μ_k^r and σ_k^r .

Results

All four publications included in this cumulative thesis aimed to assess intratumor heterogeneity by applying unsupervised and supervised learning techniques on non-invasive *in-vivo* imaging data. The first paper presents a novel unsupervised learning framework based on GMMs to analyze longitudinally acquired ^{18}F -FDG PET SUV and DW-MRI data, and model the temporal and spatial phenotypic variations in subcutaneous lung cancer [101]. In the second study we simulated clinically relevant tumor tissue time activity curves (TACs) to evaluate the feasibility of spectral clustering in segmenting multi-dimensional dynamic ^{18}F -FDG PET tumor images and quantifying the population of various intratumor tissues [102]. This study also shows the benefits of analyzing PET TACs using spectral clustering over conventional dynamic and static PET quantification measures, such as compartmental modeling parametric maps and the SUV. The third publication incorporated a spatial regularization step in spectral clustering and investigated its utility in segmenting multiparametric functional MRI data of mice bearing subcutaneous colon cancer [103]. The paper further shows that as compared to the classical clustering algorithms, the suggested segmentation approach provides the most accurate estimation of phenotypic heterogeneity within tumors. Lastly, in the final publication, we extended the previously applied spectral clus-

tering techniques to a multimodality setting and incorporated them with a random forest classifier to develop phenotype specific models. Our proposed multi-view learning approach selected the most predictive features from dynamic ^{18}F -FDG PET/multiparametric MRI data and non-invasively quantified therapy-induced intratumor tissue types. Notably, in this work we demonstrated that a systematic analysis of multiparametric PET/MRI data was pivotal for a holistic voxel-wise tissue characterization of solid tumors, which could not be obtained using static or single modality imaging.

2.1 Project-1

2.1.1 A Population-Based Gaussian Mixture Model Incorporating ^{18}F -FDG PET and Diffusion-Weighted MRI Quantifies Tumor Tissue Classes

Mathew R. Divine, Prateek Katiyar, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Bernd J. Pichler, and Jonathan A. Disselhorst (2016). *The Journal of Nuclear Medicine* 57.3, pp. 473–479.

This work aimed to develop a machine learning workflow that allows a non-invasive spatial and temporal characterization of tumor tissue heterogeneity using ^{18}F -FDG PET SUV (sub-section 1.3.9) and DW-MRI (sub-section 1.4.12). Both of these parameters are acquired routinely in the clinic and have been found to be inversely correlated in various neoplasms [104–107]. Whereas an increased FDG uptake and metabolic activity of tumors is intricately related with high mitotic rates, hypoxia and the presence of a large number of viable cells [108–110], ADC values have been shown to carry an inverse relationship with tumor cellularity [111, 112]. Thus, a combination of these two parameters is believed to provide valuable information about the tissue heterogeneity in tumors. Recently, two investigations addressed the issue of intratumor tissue characterization using a GMM (sub-section 1.6.2) on combined ^{18}F -FDG PET SUV and DW-MRI data [113, 114]. However, neither of these reports utilized the two parameters to study the intratumor phenotypic changes in a longitudinal manner.

To meet this objective, four Naval Medical Research Institute (NMRI) nu/nu mice (Charles River, Sulzfeld, Germany) were inoculated with NCI-H460 non-small cell lung cancer human tumor cells (ATCC, Manassas, Virginia, USA) on the right shoulder and imaged at day 20, 22, 24 and 27. Each imaging measurement involved serial PET and MRI scans, which were later co-registered with the help of fiducial markers. After the last imaging experiment of each mouse, the tumor

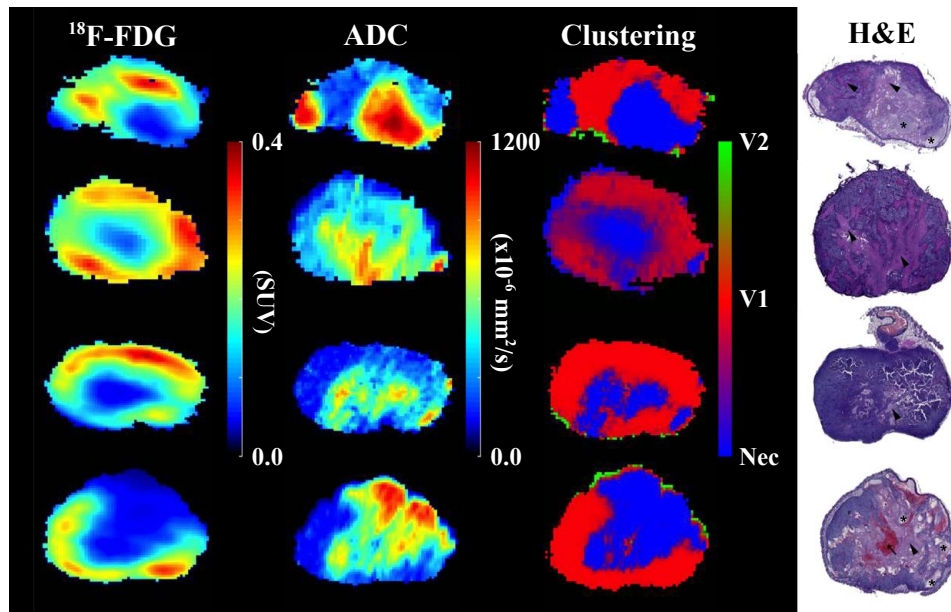


Figure 2.1: Segmentation of the last time point measurement of all four tumors (row-wise) using a GMM. Left to right: ^{18}F -FDG PET SUV and ADC images, GMM probability map and the matched H&E histology. The tumors were partitioned into three tissue clusters: viable-1 (V1, red), viable-2 (V2, green) and necrosis (Nec, blue). The arrow head and asterisk in the H&E images indicate coagulative and liquefactive necrosis, respectively. Figure adapted from Divine et al. [101].

was sectioned (along the axial plane) in two equal parts, which were further processed for hematoxylin and eosin (H&E) staining. The H&E stained sections were digitized using a NanoZoomer 2.0 HT (Hamamatsu, Hamamatsu, Japan) and different regions of interests were drawn to localize intratumor viable and necrotic tissues.

The modeling of tumor progression using PET/MRI data was carried out in two phases: initialization and segmentation. In the initialization phase, the last time point measurements of all mice were pooled into a combined dataset. This dataset was normalized to have a unit standard deviation and all voxels with an ADC value (prior to normalization) smaller than $50 (\times 10^{-6} \text{ mm}^2/\text{s})$ were labeled as noise. Subsequently, two GMMs were fitted on the combined dataset; the first model with 3 tissue clusters on the noise-free data and the second model with one cluster on the noisy data. The initialization of the first model was performed using *K*-means (sub-section 1.6.1).

In the segmentation phase, the last time point measurements of all animals were first processed independently for intratumor partitioning. This involved pre-

processing (normalization and noise removal) and clustering as described before. However, the two GMMs in this step were started using the respective model parameters obtained after the initialization phase. Post-clustering the tissue probability map of each tumor was corroborated with the associated histology. Figure 2.1 shows the imaging parameters, GMM probability maps and matching H&E histology of all four tumors. The three compartments identified in the noise-free data corresponded to a necrotic and two viable clusters. Overall, an excellent agreement was achieved between the GMM phenotypic maps and the tissue areas marked on the histology. This ground truth validation served as the basis for attributing the model parameters (fitted on the noise-free data) of different clusters from the last time point as the respective tissue class representative. Thereafter, to faithfully model the intratumor heterogeneity over time, fixed clustering was performed on the noise-free data of each tumor by applying the mixture model configuration from the last time point to the previous time points. To account for the variability in the SUV measurements [37], during fixed clustering the model configuration was kept constant and an adaptive scaling was applied on the data points (voxels) from earlier time points to obtain the most optimal fit. The data normalization was performed using the standard deviations from the last time point and the noise cluster was modeled as in the previous steps.

Figure 2.2A (density plot column) exhibits the temporal evolution of the SUV and ADC distributions for the first tumor shown in figure 2.1. On the day 20 and 22, the tumor was predominantly viable with a high ^{18}F -FDG uptake and constrained water diffusion. As the tumor grew (day 24 and 27), the necrotic areas got bigger and a new cluster emerged with a low ^{18}F -FDG uptake and high water diffusion. The proposed modeling framework (figure 2.2A, column "Divine et al.") maintained these biological characteristics of the tumor and provided tissue probability maps that depicted a faithful development of various intratumor phenotypes over time (figure 2.2B). Contrary to this, when a standard GMM was applied independently (instead of the suggested fixed clustering approach) on the previous time points, the necrotic tissue population was overestimated and the viable areas were underestimated (figure 2.2A, column "standard GMM"). The standard GMM failed,

because clustering algorithms partition data irrespective of its biological significance. Fitting a new GMM to each of the early time point measurements discarded the longitudinal nature of the data, resulting in disproportionate fractions of the tumor tissues.

In summary, this report presents a novel unsupervised learning framework that can be applied on longitudinally acquired static PET and DW-MRI data to non-invasively follow the development of intratumor phenotypes. Monitoring cancer characteristics in such a manner is of utmost relevance in the clinic, because a post-therapeutic change in the intratumor tissues can be a strong indicator of therapy resistance or efficacy [115].

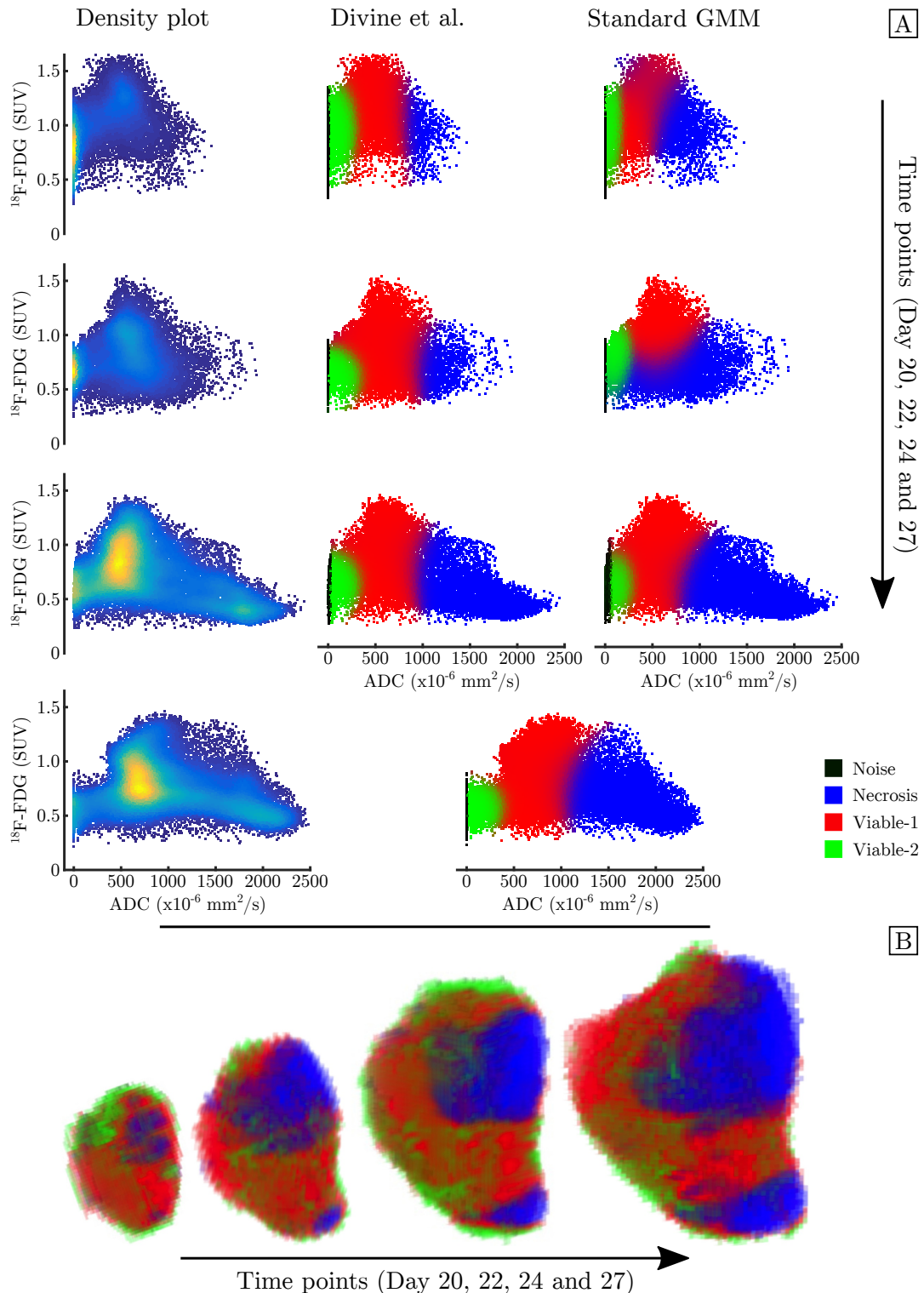


Figure 2.2: Modeling the longitudinal PET/MRI data of the first tumor shown in figure 2.1. (A) Column 1: density scatter plots depicting the changes in the tumor microenvironment from the day 20 to 27. Column 2 and 3: scatter plots obtained after applying the proposed approach and standard GMM. It is evident from the scatter plots at day 20 and 22 that as opposed to standard GMM, the fixed clustering procedure estimates the necrotic and viable tissue populations at the early time points more judiciously, providing a biologically consistent model of the tumor progression. (B) A 3-dimensional rendering of the tumor growth (based on the suggested approach and excluding noisy voxels), illustrating the spatio-temporal development of various intratumor phenotypes. Panel B adapted from Divine et al. [101].

A Population-Based Gaussian Mixture Model Incorporating ^{18}F -FDG PET and Diffusion-Weighted MRI Quantifies Tumor Tissue Classes

Mathew R. Divine¹, Prateek Katiyar^{1,2}, Ursula Kohlhofer³, Leticia Quintanilla-Martinez³, Bernd J. Pichler¹, and Jonathan A. Disselhorst¹

¹Department of Preclinical Imaging and Radiopharmacy, Werner Siemens Imaging Center, Eberhard Karls University Tuebingen, Tuebingen, Germany; ²Max Planck Institute for Intelligent Systems, Tuebingen, Germany; and ³Institute of Pathology, Eberhard Karls University Tuebingen, Tuebingen, Germany

The aim of our study was to create a novel Gaussian mixture modeling (GMM) pipeline to model the complementary information derived from ^{18}F -FDG PET and diffusion-weighted MRI (DW-MRI) to separate the tumor microenvironment into relevant tissue compartments and follow the development of these compartments longitudinally. **Methods:** Serial ^{18}F -FDG PET and apparent diffusion coefficient (ADC) maps derived from DW-MR images of NCI-H460 xenograft tumors were coregistered, and a population-based GMM was implemented on the complementary imaging data. The tumor microenvironment was segmented into 3 distinct regions and correlated with histology. ANCOVA was applied to gauge how well the total tumor volume was a predictor for the ADC and ^{18}F -FDG, or if ADC was a good predictor of ^{18}F -FDG for average values in the whole tumor or average necrotic and viable tissues. **Results:** The coregistered PET/MR images were in excellent agreement with histology, both visually and quantitatively, and allowed for validation of the last-time-point measurements. Strong correlations were found for the necrotic ($r = 0.88$) and viable fractions ($r = 0.87$) between histology and clustering. The GMM provided probabilities for each compartment with uncertainties expressed as a mixture of tissues in which the resolution of scans was inadequate to accurately separate tissues. The ANCOVA suggested that both ADC and ^{18}F -FDG in the whole tumor ($P = 0.0009$, $P = 0.02$) as well as necrotic ($P = 0.008$, $P = 0.02$) and viable ($P = 0.003$, $P = 0.01$) tissues were a positive, linear function of total tumor volume. ADC proved to be a positive predictor of ^{18}F -FDG in the whole tumor ($P = 0.001$) and necrotic ($P = 0.02$) and viable ($P = 0.0001$) tissues. **Conclusion:** The complementary information of ^{18}F -FDG and ADC longitudinal measurements in xenograft tumors allows for segmentation into distinct tissues when using the novel GMM pipeline. Leveraging the power of multiparametric PET/MRI in this manner has the potential to take the assessment of disease outcome beyond RECIST and could provide an important impact to the field of precision medicine.

Key Words: PET/MR; Gaussian mixture model; ^{18}F -FDG; DWI

J Nucl Med 2016; 57:473–479

DOI: 10.2967/jnumed.115.163972

PET and MRI have been successfully integrated into a combined imaging solution in clinics and preclinical laboratories worldwide and represent a mature imaging modality (1). However, because of the relatively short time period in which PET/MRI solutions have been available, oncologic applications that use the complementary information obtained from the separate modalities are in their infancy. Several investigations have compared ^{18}F -FDG PET/CT and diffusion-weighted (DWI) MRI measurements on the same cancer patients to determine the best approach for patient staging, prognosis, and therapy stratification, pitting PET/CT against MRI, but have not looked thoroughly into the combined utility of ^{18}F -FDG PET and MRI (2–5). In 2 of these investigations (4,5), an inverse correlation between ^{18}F -FDG uptake and DWI metrics in cancerous lesions has been shown.

DWI is a noninvasive technique to measure the self-diffusion of water in vivo and is often quantified by the apparent diffusion coefficient (ADC). Malignant tumors generally have lower ADC values than benign or necrotic tissues because of water movement restriction caused by increased cellularity (5). Measurements of tumor cellularity in non-small cell lung cancer (NSCLC) patients with DWI have been used to predict tumor invasiveness at early stages and have been shown to characterize well-differentiated adenocarcinoma from other types of lung carcinoma (6). The underlying mechanism associating changes in ADC to therapeutic response is a decreased cellularity resulting from necrosis and apoptosis (7). In some cases, an initial decrease in ADC, attributed to cellular swelling or reduced blood flow (8), is attributed to treatment response.

^{18}F -FDG uptake in lung cancers has been associated not only with glucose transporter 1 and hexokinase activity, biologic components directly related to glucose metabolism, but also with hypoxia-inducible factor 1- α , vascular endothelial growth factor, its receptor, and microvessels (9). This implicates a complex link between glucose consumption and many opposing factors involved in the multifaceted tumor microenvironment. Moreover, ^{18}F -FDG can potentially stage tumors in patients with histologically verified NSCLC (10–12), and changes in ^{18}F -FDG uptake after therapeutic intervention are correlated to patient prognosis (13–15), possibly predicting response (14). Adding biologic information measured with DWI might further improve staging and help decide on patient-individualized therapies for treating noninvasive adenocarcinomas (15).

To fully use multiparametric PET/MR imaging, assessment of disease outcome needs to go beyond RECIST (16) and move

Received Jul. 17, 2015; revision accepted Nov. 17, 2015.

For correspondence or reprints contact: Bernd J. Pichler, University of Tuebingen, Department of Preclinical Imaging and Radiopharmacy, Werner Siemens Imaging Center, Roentgenweg 13, D-72076 Tuebingen, Germany.

E-mail: bernd.pichler@med.uni-tuebingen.de

Published online Dec. 10, 2015.

COPYRIGHT © 2016 by the Society of Nuclear Medicine and Molecular Imaging, Inc.

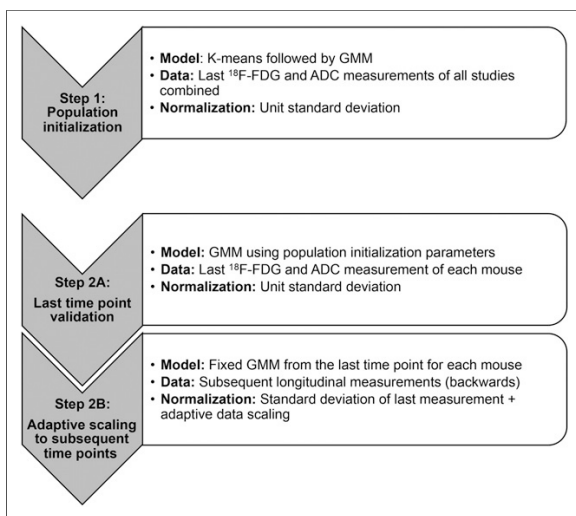


FIGURE 1. In the first step of the GMM workflow, data from the last imaging time point of all tumors are pooled into a single population and initialized. Step 2A uses initialization from step 1 to segment the last time point of each tumor. Step 2B uses adaptive scaling on all early time points for each tumor while holding the GMM parameters from step 2A constant.

toward comprehensive algorithms, which unify the complementary datasets. Recently, Schmidt et al. have reported on the correlation of ADC and ^{18}F -FDG inside of NSCLC tumors using a hybrid PET/MR (17). They applied a Gaussian mixture model (GMM) (18) and theorized that it separated the tumors into distinct and relevant tissue classes. Their investigation did not, however, include a histologic validation, track changes longitudinally, or use adaptive clustering techniques. The aim of our study was to

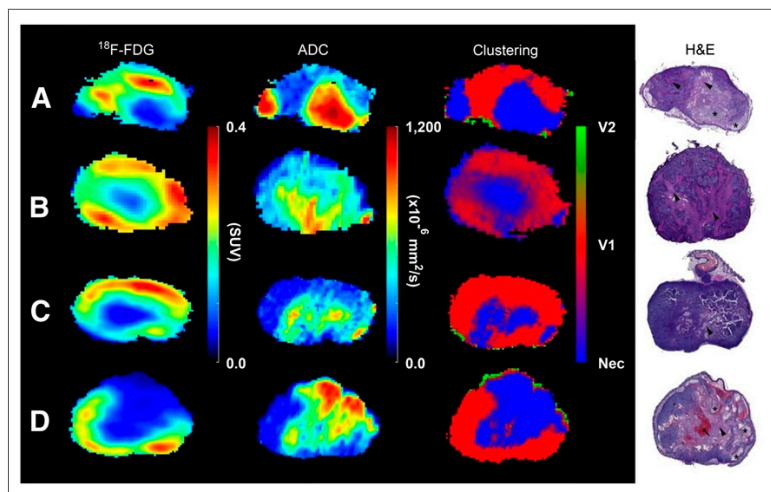


FIGURE 2. (From left to right) ^{18}F -FDG, ADC, clustering, and hematoxylin and eosin (H&E) examples are displayed for comparison. GMM was applied to the last time point of ^{18}F -FDG and DW-MR images to cluster them into viable (red, V1, and green, V2) and necrotic (blue, Nec) regions. Manually delineated H&E stainings were coregistered to tumor volumes to verify tissue classification. Liquefactive (asterisk) and coagulative necrosis (arrowhead) and blood vessel congestion (arrow) are depicted in the H&E stainings.

create a novel clustering pipeline to model the complementary information derived from ^{18}F -FDG and ADC to separate the tumor microenvironment into relevant tissue compartments and follow their development longitudinally. A correlation with imaging and histology was established to validate our model and guide a longitudinal implementation so that it was possible to observe the dynamics of individual tissue compartments over time.

MATERIALS AND METHODS

A detailed description of cell culturing, animals, tumor inoculation, ^{18}F -FDG production, PET and MR imaging, and image coregistration and statistics can be found in the supplemental data (supplemental materials are available at <http://jnm.snmjournals.org>). The rest of this section describes details of our modeling pipeline.

The clustering workflow of PET/MRI data was performed in 2 major steps: GMM on all last time points of the ^{18}F -FDG and ADC datasets and GMM on individual measurements (Fig. 1). The first step (population initialization) involved combining PET/MRI data acquired at the last time points from all tumors into one dataset. The combined dataset serves as a reference for the initialization of GMM in subsequent steps, as these voxel pairs represent the heterogeneity expected in all measurements. Before segmentation, the dataset was normalized to the unit standard deviation (SD). Voxels with ADC values lower than $50 \times 10^{-6} \text{ mm}^2/\text{s}$ were labeled as noise and excluded from initialization; a 2-dimensional (2D) Gaussian distribution was fit to the noise cluster. Cluster initialization was performed on the noise-free dataset by repeatedly running the K-means (18) algorithm with 3 randomly seeded centroids. The model parameters of the noise and noise-free clusters were used to initialize GMM on the combined dataset. The choice of 3 tissue clusters and 1 noise cluster was based on visual inspection of the data (Figs. 2 and 3).

In step 2A, the last time point of each tumor was normalized to unit SD, and the noise cluster was extracted as described above. Here, the model parameters of the noise (corresponding to last time point of respective tumor) and noise-free clusters (obtained in step 1) were used to initialize GMM on the last time points.

After termination of the expectation maximization algorithm, the clustering probability map of the last time point of each tumor was validated with respective histology (Fig. 4). After validation, the model parameters of the noise-free clusters of each last time point were used to perform hard clustering (model parameters were fixed) on the early measurements of the respective tumor (step 2B). The noise cluster was characterized as described above. Because ^{18}F -FDG measurements are vulnerable to variability, the data from early time points were adaptively scaled by a multiplicative factor to fit the given distribution. The clusters in the last step were fixed for 2 main reasons: once validated with histology, the noise-free cluster parameters of the last time point can be considered as GMM representative of the respective tumor tissue classes, and the density-based models have the tendency to partition data irrespective of their biologic significance. It is highly likely that the necrotic and viable fractions of tumors in early time points are different as compared with that in the last time point. Keeping the clusters fixed does not forcefully assign false observations to different clusters and models the tumor progression judiciously. For the sake

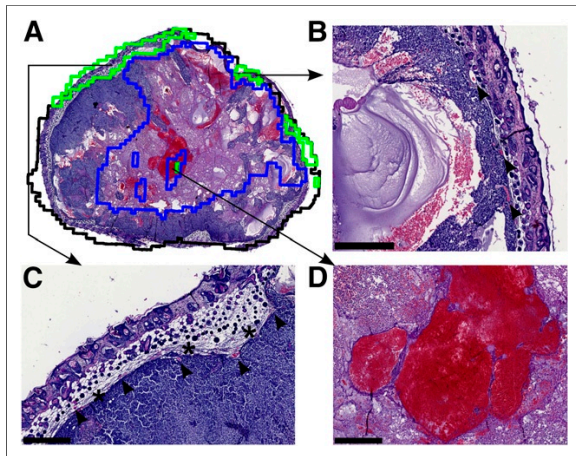


FIGURE 3. (A) Outlines of the coregistered clustering slice (black) and V2 (green) and necrosis (blue) are placed on top of the histology slice for comparison of agreement between imaging and histology. (B and C) V2 is seen at the periphery of the tumor along with blood vessels (arrows) and connective tissue (asterisk). (D) Slight misclassification of the necrotic area occurred due to artifacts produced in ADC because of the blood-pool.

of comparison, a 1-dimensional (1D) GMM was performed in the same manner with only the ADC data.

RESULTS

Hematoxylin and eosin (H&E) stainings depict tumors with large atypical cells with prominent nucleoli and abundant cytoplasm. Mitosis and apoptosis are frequently seen (Fig. 2). Liquefactive and coagulative necrosis is seen along with blood vessel congestion and dilation. Visual inspection of the manually aligned histology and delineated tumor slices revealed that a good coregistration was achieved (Fig. 3). Uncertainty in tissue classification is denoted by a weighted sum of colors and is exemplified in the clustering results (Fig. 2B), which exhibits fingerlike regions of coagulative necrosis in the histology; a similar texture is seen in the corresponding clustering probability map. There are 2 viable regions present, V2 (green) and V1 (red). V2 is seen only at the periphery and corresponds to connective tissue, dense cells, and vessels (Fig. 3), whereas V1 is bigger and represents densely packed tumors cells.

The median values and the interquartile range for the average tumor, necrotic, viable, and noise regions are summarized for both

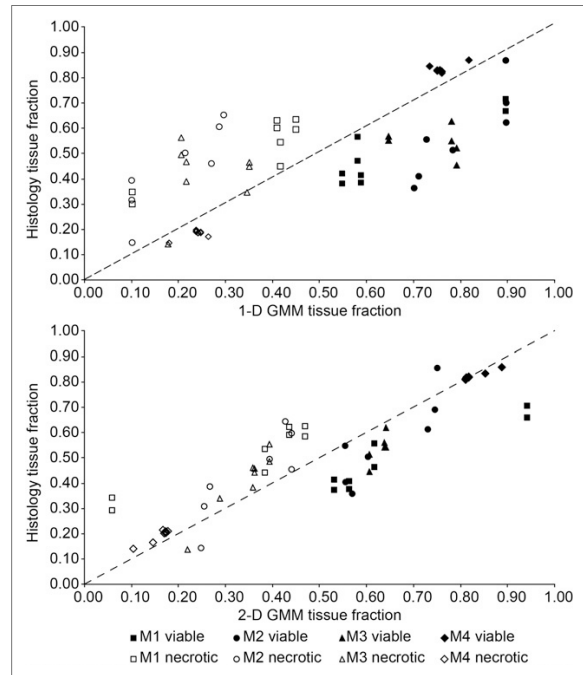


FIGURE 4. Necrotic (open symbols) and viable (filled symbols) tissue fractions taken from matching H&E, and clustering imaging slices are plotted for each mouse (M1–M4) along with the line of identity as validation of GMM tissue classification. For the sake of comparison, a 1D GMM was applied to only the ADC parameter. A 2D GMM was applied to the combined ^{18}F -FDG and ADC imaging dataset.

^{18}F -FDG and ADC (Table 1). The ^{18}F -FDG average values are highest for the viable tissue and lowest for the necrotic tissue, whereas the ADC is highest for the necrotic regions and lowest for the viable regions across all time points.

The fractions of necrosis and viable tissue in H&E stainings correlate quite well to corresponding tissue fractions in clustering slices (Table 2) and are plotted (Fig. 4) for each mouse along with the line of identity. Tissue fractions were derived from manually drawn regions on histology (Supplemental Fig. 2). In both the 1D (ADC only) and the 2D (ADC and ^{18}F -FDG) cases, the GMM had a tendency to overestimate viable regions and underestimate necrotic regions. In general, the tumor clustering results were more influenced by ADC (in the 2D case); however, the overall better agreement

TABLE 1
Median and Interquartile Range for Whole Tumor and Individual Clusters

Cluster type	^{18}F -FDG SUV		ADC ($\times 10^{-6}$ mm 2 /s)	
	Median	Interquartile range	Median	Interquartile range
Whole tumor	0.20	0.11	668	410
Viable 1 (red)	0.24	0.11	580	241
Viable 2 (green)	0.21	0.09	551	310
Necrotic (blue)	0.17	0.07	1008	394
Noise (black)	0.13	0.05	4.45×10^{-7}	6.63×10^{-5}

TABLE 2
Correlation Values for Histology and GMM Tissue Fractions

Mouse	2D GMM, r-value		1D GMM (ADC), r-value	
	Viable	Necrotic	Viable	Necrotic
Mouse 1	0.94	0.94	0.69	0.90
Mouse 2	0.85	0.85	0.86	0.86
Mouse 3	0.96	0.96	0.16	0.16
Mouse 4	0.98	0.96	0.66	0.66
All mice	0.87	0.88	0.56	0.56

r-value = Pearson correlation coefficient.

between histology in the 2D GMM serves as a validation of this methodology. The slice-by-slice correspondence between the 2D GMM and histology was calculated with the Dice coefficient (Supplemental Table 2).

In Figure 5, the histograms of ADC and ^{18}F -FDG for 4 different time points from mouse 3 are shown along with corresponding density and scatter plots. The ADC histograms become increasingly skewed toward higher values as tumors become more necrotic. The ^{18}F -FDG histograms, on the other hand, demonstrate large variability, consequently maintaining a similar average value for all time points. In general, as the tumors grow in size, so does the negative correlation between ^{18}F -FDG and ADC on a voxelwise basis (Supplemental Fig. 3).

The 3-dimensional renderings of an exemplary tumor (mouse 3) depict the spatial-temporal evolution of viable (red and green) and necrotic (blue) tissue compartments (Fig. 6). The initial necrotic portion at the first time point retains its relative position in the tumor and becomes larger over time. In the second time point, a smaller necrotic region appears in the lower portion of the tumor and also increases in size while retaining its relative position in the tumor. Moreover, purple areas on the rim of the necrotic region are due to the mixed probability of belonging to the viable or necrotic tissue class; the purple regions are possibly associated with hypoxic areas in the tumor (Supplemental Fig. 4). The green viable region is consistently located at the exterior of the tumor.

The linear regression analysis supported the hypothesis that tumor volume was positively correlated to ADC and ^{18}F -FDG for the whole tumor and for the viable and necrotic regions (Supplemental Table 1). ADC was also a highly significant predictor for ^{18}F -FDG in the whole tumor and in the segmented viable and necrotic regions. There was no significant difference between the slopes of the ADC versus volume (Fig. 7A) and ^{18}F -FDG-SUV versus volume (Fig. 7B). After the slopes of each group were constrained, the Tukey-Kramer post hoc test confirmed that a significant difference existed between all groups (Fig. 7A) and the viable and necrotic regions (Fig. 7B); the slopes of the regression lines (Fig. 7C) were significantly different from one another.

DISCUSSION

We have developed a GMM pipeline to assess tumor heterogeneity using information from both PET and MRI, showing how necrotic and viable regions develop in a longitudinal manner. We

validated the methodology both visually (Figs. 2 and 3) and quantitatively (Fig. 4) and have found a good agreement between clustering results and histology. We have shown how the intratumoral relationship between ^{18}F -FDG and ADC changes longitudinally (Fig. 5). Moreover, the manner in which the GMM segments the tumors has been shown. The visualization of an exemplary tumor reveals how tissue classes develop spatially over time (Fig. 6). Last, we have shown a positive, linear relationship between ^{18}F -FDG and ADC values in the tumor (Figs. 7A and 7B), and ADC

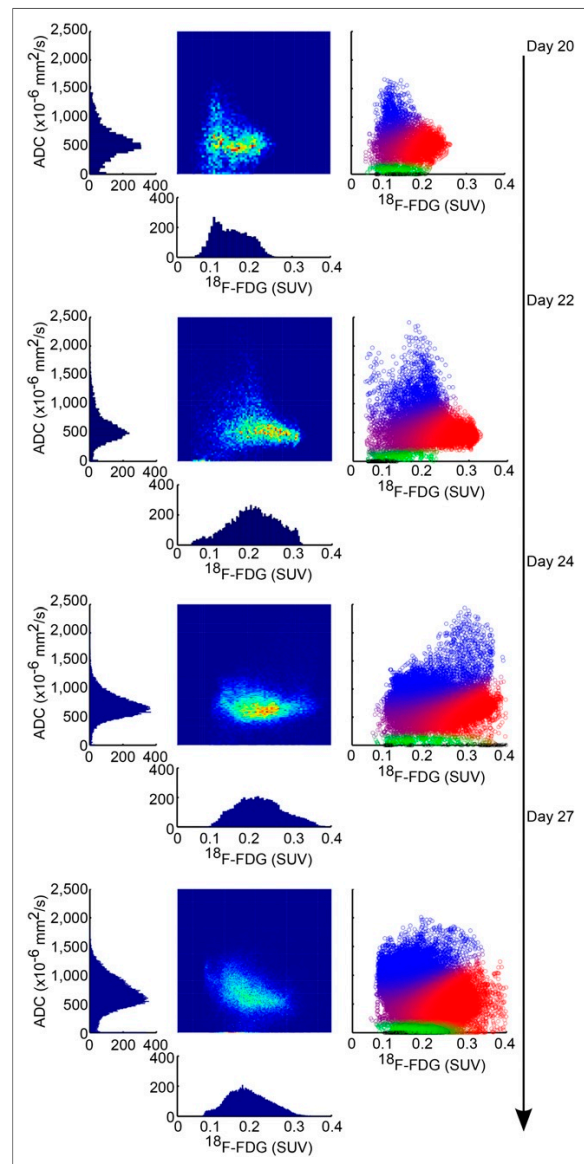


FIGURE 5. Histograms represent distributions of ADC and ^{18}F -FDG values at 4 time points for a single mouse. Density scatter plots depict the voxelwise relationship of ADC and ^{18}F -FDG values in the tumor. In the scatter plots on the right, necrotic tissue is in blue, viable is in green and red, and noise is in black. Combined probabilities of necrotic and viable classes are shown as a mixture of colors.

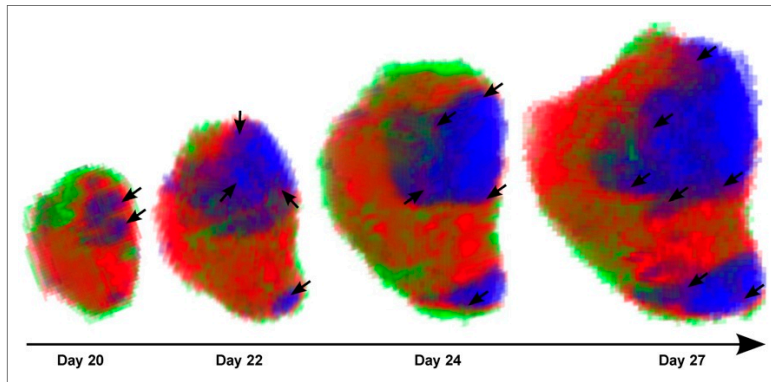


FIGURE 6. A 3-dimensional rendering depicts spatial-temporal growth of tumors and development of tissue classes. Red and green represent viable tissue, and blue represents necrosis. A purple rim (arrows) is seen around the necrosis and represents the combined probability of necrotic and viable tissue classes, possibly indicating hypoxic cells.

was shown to be an excellent predictor of ^{18}F -FDG in this tumor model (Fig. 7C).

In this study, the ADC values in the necrotic regions are low compared with results from other investigators who have also used similar segmentation techniques on purely MRI data, validated with histology, to create 2 necrotic region types with the following values: $(1,510 \pm 120$ and $1,560 \pm 240)$ (19), $(1,260 \pm 130$ and $1,610 \pm 41)$ (20), and $(2,120 \pm 50$ and $1,790 \pm 10)$ (21) $\times 10^{-6}$ mm^2/s . The b-values being used affect the ADC values, with lower b-values being heavily perfusion-dependent and responsible for increasing the overall ADC. Higher b-values (>100) are desirable to suppress perfusion-weighted components of well-perfused regions (22). In the 3 aforementioned studies, b-values below 100 were used in the calculation of the ADC maps, whereas we used b-values greater than 200.

Knowing the extent of necrosis in a tumor model can be helpful as it can be the source of confounding results in determining the efficacy of potential tumor therapies, as is the case with Berry et al. who found that necrotic fractions contributed only noise to the measurement of a therapeutic effect in an antiangiogenic drug therapy study (19). Moreover, the extent of necrosis might help to identify hypoxic tissues at the rim of necrosis and guide therapy options, because hypoxic tissues are well known to be highly resistant to radiation therapy (23). We also observed that perinecrotic tissue stains positive for hypoxia-inducible factor 1- α (Supplemental

Fig. 4), an indirect marker of tissue hypoxia. Although ^{18}F -FDG is not specific for hypoxic tissue, hypoxia has been shown to correlate to ^{18}F -FDG uptake (24).

The increase in the ADC value in the viable regions (Fig. 7A) indicates that these regions are becoming less dense, presumably because of micronecrosis as seen in the histology (Fig. 2). In the ideal case, the necrotic regions would have a slope of zero. The increase in both the parameters over time (Fig. 7C) could suggest that as the tumor becomes larger, it also becomes more aggressive because of the increased necrotic burden, which harbors an increased interstitial pressure, low oxygenation, and oxygen reactive species due to opened cell membranes (25–27). On the other hand, no partial-volume correction was performed in this study and could be a cause for the volume-dependent increase in ^{18}F -FDG values (Fig. 7B).

A positive correlation between ADC and ^{18}F -FDG SUV could seem, at first glance, inconsistent with findings from Schmidt et al., who have reported a negative correlation between SUV_{max} and ADC_{min} (17). However, in accordance with Schmidt et al., the intratumor correlation coefficient is mostly negative for all tumors and becomes more negative as the total tumor volume increases (Supplemental Fig. 3). Thus, on a voxelwise basis, an L-shaped 2D histogram is seen in the tumors of this study and in the human lung tumors from Schmidt et al. Several other authors have also reported significant negative correlations between ADC and ^{18}F -FDG in various types of malignancies. Nakajo et al. reported a correlation coefficient of $\rho = -0.56$ in head and neck squamous cell carcinomas for ADC_{mean} and ^{18}F -FDG SUV_{max} (28); Baba et al. reported $\rho = -0.36$ in breast lesions for ADC_{mean} and ^{18}F -FDG SUV_{max} (29); and Rakheja et al. reported a range of ρ values from -0.18 to -0.29 in various neoplastic lesions, for various combinations of ^{18}F -FDG SUV and ADC_{min} and ADC_{max} (30). None of these studies reported the correlation between ^{18}F -FDG SUV and ADC longitudinally, as performed in this study, making it hard to compare. Nakajo et al. did, however, correlate ADC and ^{18}F -FDG SUV to patient survival, with higher ADCs and lower ^{18}F -FDG SUVs associated with disease-free survival. The number of negative correlations observed could imply that ADC and ^{18}F -FDG SUV move along a negative slope in patients receiving treatment. In the NCI-H460 xenograft tumors of this study, the viable tissue regions are more island-shaped, with micronecrosis between dense clusters of cells, which could lead to a positive correlation in both ADC and ^{18}F -FDG SUV mean values over time.

The ability of the proposed model to accurately segment the tumor microenvironment into the proposed viable and necrotic regions is dependent on the relationship that ^{18}F -FDG and ADC voxels inside of the tumor have. There was a definite negative correlation between the 2 imaging parameters at the time point we

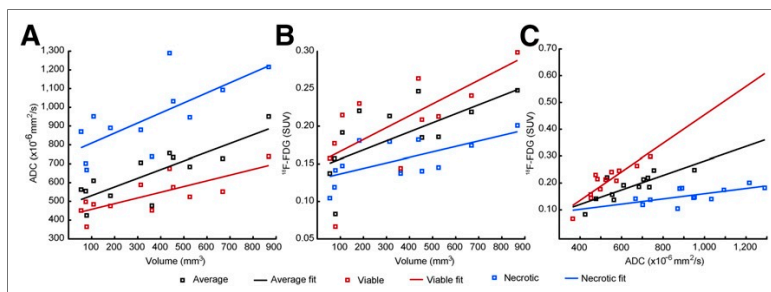


FIGURE 7. Average ADC (A) and ^{18}F -FDG (B) values for different tissue classes plotted against whole tumor volume. ^{18}F -FDG values plotted against ADC values (C) for each class.

chosed to validate our model. The tissue classes have to be present to create the shape seen in the 2D histogram of the last time point (Fig. 5); otherwise, this tissue heterogeneity will not be incorporated into the model for earlier time points.

The clear limitation of this study is that the tumors were subcutaneously inoculated in immune-compromised mice. An orthotopic or genetic mouse model could have led to greater tissue heterogeneity with an increase in the number of tissue classes to identify. In addition, serial PET/MR imaging measurements were used in this study, whereas the use of a recently developed, combined PET/MR system (31) for measurements could have decreased the error due to rigid coregistration, which might have been further increased by movement of the animal from the PET to the MRI scanner, eventually leading to nonrigid movements of the subject. Also, clustering results were not verified with histology at every time point. Future studies will focus on the automatic, nonrigid coregistration of histology with imaging data to obtain a better degree of spatial correspondence (Supplemental Table 2) than we were able to achieve in this study. Obtaining a good degree of spatial correspondence poses its own unique challenges because of nonrigid deformations of histology.

In summary, with the proposed GMM pipeline we incorporated the complementary information intrinsically associated with DW-MRI and ¹⁸F-FDG PET. One class of necrotic tissue was found, along with 2 classes of viable tissue. The green tissue class was found only at the periphery of the tumor and represents densely packed cells, vessels, and connective tissue; it has the lowest ADC values. This tissue class presumably represents the new outgrowth of the tumor and could provide a hint as to the direction of growth, as the tumor appears to progress in the direction of this tissue class at every time step (Fig. 6). Because necrosis has been indicated for poor survival outcome and has been associated with hypoxia, measuring the relative abundance of necrosis could help physicians to stratify patients accordingly and decide on the type of therapy (32). The opportunity to measure the relative size and growth patterns of different tissue types after the induction of treatment will help to gauge the overall response to tumor therapy, as well as to be useful for monitoring and optimizing the drug dose and scheduling in pre-clinical animal models.

CONCLUSION

The complementary information from ¹⁸F-FDG SUV and ADC longitudinal measurements in tumors allows for segmentation into distinct tissue classes when the proposed GMM pipeline is used. Leveraging the power of multiparametric PET/MR imaging in this manner has the potential to take the assessment of disease outcome beyond RECIST and into the realm of precision medicine.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. Research support for this study was received from the German Ministry for Education and Research (BMBF), grant 0316186E; the Swiss Werner Siemens Foundation; the Fortüne at the University of Tuebingen (Evaluation of Tumor

Heterogeneity Using Clustering of Multi-Modality Imaging Data, fortune 2131-0-0); and the European Research Council (ERC), grant 323196. No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

We acknowledge the generous technical support from Nadine Bauer, Funda Cay, Maren Harant, and Daniel Bukala.

REFERENCES

1. Wehrli HF, Sauter AW, Divine MR, Pichler BJ. Combined PET/MR: a technology becomes mature. *J Nucl Med*. 2015;56:165–168.
2. Komori T, Narabayashi I, Matsumura K, et al. 2-[fluorine-18]-fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography versus whole-body diffusion-weighted MRI for detection of malignant lesions: Initial experience. *Ann Nucl Med*. 2007;21:209–215.
3. Usuda K, Zhao XT, Sagawa M, et al. Diffusion-weighted imaging is superior to positron emission tomography in the detection and nodal assessment of lung cancers. *Ann Thorac Surg*. 2011;91:1689–1695.
4. Regier M, Derlin T, Schwarz D, et al. Diffusion weighted MRI and ¹⁸F-FDG PET/CT in non-small cell lung cancer (NSCLC): does the apparent diffusion coefficient (ADC) correlate with tracer uptake (SUV)? *Eur J Radiol*. 2012;81:2913–2918.
5. Herneth AM. Apparent diffusion coefficient: a quantitative parameter for in vivo tumor characterization. *Eur J Radiol*. 2003;45:208–213.
6. Matoba M, Tonami H, Kondou T, et al. Lung carcinoma: diffusion-weighted MR imaging—preliminary evaluation with apparent diffusion coefficient. *Radiology*. 2007;243:570–577.
7. Moffat BA, Chenevert TL, Lawrence TS, et al. Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proc Natl Acad Sci USA*. 2005;102:5524–5529.
8. Schraml C, Schwenzer NF, Martirosian P, et al. Diffusion-weighted MRI of advanced hepatocellular carcinoma during sorafenib treatment: initial results. *AJR*. 2009;193:W301–307.
9. Sauter AW, Schwenzer N, Divine MR, Pichler BJ, Pfannenber C. Image-derived biomarkers and multimodal imaging strategies for lung cancer management. *Eur J Nucl Med Mol Imaging*. 2015;42:634–643.
10. Antoch G, Stattaus J, Nemat AT, et al. Non-small cell lung cancer: dual-modality PET/CT in preoperative staging. *Radiology*. 2003;229:526–533.
11. Kligerman S, Digumarthy S. Staging of non-small cell lung cancer using integrated PET/CT. *AJR*. 2009;193:1203–1211.
12. Weber WA, Petersen V, Schmidt B, et al. Positron emission tomography in non-small-cell lung cancer: prediction of response to chemotherapy by quantitative assessment of glucose use. *J Clin Oncol*. 2003;21:2651–2657.
13. Cerfolio RJ, Bryant AS, Ohja B, Bartolucci AA. The maximum standardized uptake values on positron emission tomography of a non-small cell lung cancer predict stage, recurrence, and survival. *J Thorac Cardiovasc Surg*. 2005;130:151–159.
14. Huang W, Zhou T, Ma L, et al. Standard uptake value and metabolic tumor volume of ¹⁸F-FDG PET/CT predict short-term outcome early in the course of chemoradiotherapy in advanced non-small cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2011;38:1628–1635.
15. Tanaka R, Nakazato Y, Horikoshi H, Tsuchida S, Yoshida T. Diffusion-weighted imaging and positron emission tomography in various cytological subtypes of primary lung adenocarcinoma. *Clin Imaging*. 2013;37:876–883.
16. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228–247.
17. Schmidt H, Brendle C, Schraml C, et al. Correlation of simultaneously acquired diffusion-weighted imaging and 2-deoxy-[¹⁸F] fluoro-2-D-glucose positron emission tomography of pulmonary lesions in a dedicated whole-body magnetic resonance/positron emission tomography system. *Invest Radiol*. 2013;48:247–255.
18. Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett*. 2010;31:651–666.
19. Berry LR, Barck KH, Go MA, et al. Quantification of viable tumor microvascular characteristics by multispectral analysis. *Magn Reson Med*. 2008;60:64–72.

20. Carano RA, Ross AL, Ross J, et al. Quantification of tumor tissue populations by multispectral analysis. *Magn Reson Med.* 2004;51:542–551.
21. 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.
22. Taouli B, Koh D-M. Diffusion-weighted MR imaging of the liver. *Radiology.* 2010;254:47–66.
23. Sun X, Niu G, Chan N, Shen B, Chen X. Tumor hypoxia imaging. *Mol Imaging Biol.* 2011;13:399–410.
24. Rajendran JG. Hypoxia and glucose metabolism in malignant tumors: evaluation by [¹⁸F]fluoromisonidazole and [¹⁸F]fluorodeoxyglucose positron emission tomography imaging. *Clin Cancer Res.* 2004;10:2245–2252.
25. Koukourakis MI, Giatromanolaki A, Skarlatos J, et al. Hypoxia inducible factor (HIF-1a and HIF-2a) expression in early esophageal cancer and response to photodynamic therapy and radiotherapy. *Cancer Res.* 2001;122:1830–1832.
26. Vleugel MM, Greijer AE, Shvarts A, et al. Differential prognostic impact of hypoxia induced and diffuse HIF-1alpha expression in invasive breast cancer. *J Clin Pathol.* 2005;58:172–177.
27. Zhong H, De Marzo AM, Laughner E, et al. Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. *Cancer Res.* 1999;59:5830–5835.
28. Nakajo M, Nakajo M, Kajiya Y, et al. FDG PET/CT and diffusion-weighted imaging of head and neck squamous cell carcinoma: comparison of prognostic significance between primary tumor standardized uptake value and apparent diffusion coefficient. *Clin Nucl Med.* 2012;37:475–480.
29. Baba S, Isoda T, Maruoka Y, et al. Diagnostic and prognostic value of pre-treatment SUV in ¹⁸F-FDG/PET in breast cancer: comparison with apparent diffusion coefficient from diffusion-weighted MR imaging. *J Nucl Med.* 2014;55:736–742.
30. Rakheja R, Chandarana H, DeMello L, et al. Correlation between standardized uptake value and apparent diffusion coefficient of neoplastic lesions evaluated with whole-body simultaneous hybrid PET/MRI. *AJR.* 2013;201:1115–1119.
31. Judenhofer MS, Wehrl HF, Newport DF, et al. Simultaneous PET-MRI: a new approach for functional and morphological imaging. *Nat Med.* 2008;14:459–465.
32. Hiraoka N, Ino Y, Sekine S, et al. Tumour necrosis is a postoperative prognostic marker for pancreatic cancer patients with a high interobserver reproducibility in histological evaluation. *Br J Cancer.* 2010;103:1057–1065.

Supplemental Materials and Methods

Cell Culture

NCI-H460 non-small cell lung cancer human tumor cells (ATCC, Manassas, VA, USA) ($2 - 5 \times 10^6$) remained frozen at -80°C in Phosphate Buffered Saline (PAA Laboratories, Pasching Austria) without Ca^{2+} or Mg^{2+} at a pH of 7.5 with 10% dimethyl sulfoxide in 1.5 ml Eppendorf tube aliquots until needed. Cells were thawed by pipetting 1 ml of Roswell Park Memorial Institute cell medium (RPMI) (Biochrom, Berlin, Germany) with 10% fetal calf serum (Biochrom) and 5% penicillin streptomycin (Biochrom) into the Eppendorf tube and pipetting up and down until the cell medium was no longer frozen. Cells were then pipetted into 175 cm^2 , polystyrene, Cellstar® cell culture flasks with a red filter cap (Greiner Bio-One, Frickenhausen, Germany) with 15 ml of RPMI cell medium mixture and incubated at 37°C in a Heracell 150i cell incubator (Thermo Scientific, Karlsruhe, Germany) at 5% CO_2 . Cells were inspected every other day under the Aciovert 40 CFL 10 microscope (Zeiss, Oberkochen, Germany) for confluency and viability. Cell medium was changed every two days until cells reached 90% confluency, upon which they were split by first vacuuming excess cell medium from the flasks with an Integra Vacusafe pump (Integra Biosciences, Kassel, Germany), followed by a wash using 10 ml of PBS, followed by adding 3 ml of 0.05% trypsin (Biochrom) with 0.02% EDTA (Biochrom) and finally incubated for 5 min or until the cells visibly detached from the flask. Cells were then pipetted into a 50 ml Falcon tube (BD, Fanklink Lakes, NJ, USA) and centrifuged at 1200 rpm for 5 min in a Heraeus multifuge 35 RT centrifuge (Thermo Scientific) at 8°C . Cells were then placed back into the 10 ml of PBS and pipetted up and down until all visible cell clumps had dissolved. If cells were to be

further expanded, they were split into different cell flasks. If cells were to be injected into animals, they were first counted by taking a 10 μ l sample from the 10 ml cell suspension, mixing it thoroughly with 90 μ l of 1:2 trypan blue (Biochrom) staining, and pipetting this mixture into a Neubauer improved cell counter chamber (Assistant, Sondheim, Germany). Cells were re-centrifuged and 2×10^6 cells were placed in 200 μ l of sterile 0.9% NaCl for later subcutaneous injection. Pipetting of cells and mixing of reagents was performed in sterile conditions in a clean bench.

Animals

Six week old nmri nu/nu mice (Charles River, Sulzfeld, Germany) were ordered and allowed to acclimatize for one week in the laboratory vivarium before any procedures were performed. Briefly, mice were held in specific pathogen-free conditions in individually ventilated cages of type 2 long (Tecniplast, Buguggiate, Italy). The room climate was regulated with a room temperature of $20 \pm 1^\circ$ C, a $50 \pm 10\%$ humidity, and a 12 h night and day lighting scheme. The mice were provided with food and water *ad libitum*.

Tumor inoculation

Directly before injecting tumor cells subcutaneously, mice were anesthetized by being allowed to breathe in 1.5% isoflurane (CP Pharma, Burgdorf, Germany) dissolved in 100% O₂ (Linde AG, München, Germany) at a flow rate of 0.8 l/min until they showed loss of reflex due to deep narcosis. Twenty-two gauge micro lance needles (BD) were used to fill Injekt F 1 ml syringes (Braun, Melsungen, Germany) with cell suspension to reduce the shear force experienced by the cells while being pulled into the syringe. The injections were made subcutaneously on the right shoulder of mice with 27 gauge

micro lance needles to suppress leakage of the cells through the puncture made in the skin for the subcutaneous injection. Mice were then placed in the vivarium and tumor sizes as well as normal social activity of animals were monitored until tumors were palpable and showed signs of vascularization with a minimum length of 5 mm.

One mouse was imaged using PET and MRI at 27 days after tumor inoculation in order to establish how well *ex vivo* histology could be correlated with imaging experiments. Afterwards, three mice were measured longitudinally at days 20, 22, 24, and 27 after tumor inoculation. All experiments were performed in accordance with the German Animal Welfare Act, and protocols were approved by the Regierungspräsidium in Tuebingen.

¹⁸F-FDG production

¹⁸F was produced from the ²⁰Ne(d,α)¹⁸F reaction by bombarding a Ne(15% H₂) target with energetic protons using the 16 MeV Cyclotron. Afterwards, distilled water was used to remove no-carrier added ¹⁸F-HF from the target. Using a FDG MicroLab module (GE Healthcare, Münster, Germany), FDG was synthesized based on the protocol from Hamacher et al.(1) using mannose triflate (ABX, Radeberg, Germany) as the precursor. Radiochemical purity as determined by TLC was greater than 95%.

PET

Anesthesia was induced with 1.5% isoflurane (CP Pharma, Burgdorf, Germany) in 100% O₂ at a flow of 0.5 L/min shortly before injection of 12.0 ± 0.7 MBq of ¹⁸F-FDG in 50 µL of 0.9% NaCl in the tail vein. Mice were kept under anesthesia and warmed using heating pads for 50 minutes before PET measurements. Mice were carefully placed on an animal holder, and the temperature of the mice was maintained with a heating water pump system (Circulating Thermostat; Bruker Biospin,

Ettlingen, Germany). Three line sources were placed on the animal holder in order to co-register PET and MRI images. Tumors were placed in the middle of the field of view of the Inveon dedicated small animal PET scanner (Siemens, Knoxville, TN, USA). Scans were acquired using accompanied software (Inveon Acquisition Workplace version number 1.5.0.28) for 10 min and reconstructed using OSEM3D/FastMAP (reconstruction software version: 2.5, histogram version: 2.39, re-binning version: 2.5) with the following parameters: Ramp projection filter, image zoom = 1, image matrix size = 256 x 256 x 159 with (0.39 x 0.39 x 0.80) mm³ voxel sizes, OSEM3D Iterations = 2, MAP Iterations = 18, Beta = 0.05, Uniform set to Resolution, and FastMap setting on. Data are reported as Standard Uptake Value (SUV).

MRI

Directly after PET scans, mice were moved while lying in the same position to a dedicated small animal 7 T Clinscan MRI (Bruker Biospin) where T2-weighted anatomical and DW-MR images were acquired using accompanied software (Siemens, Syngo version number: MR B15). For better image quality, breathing was measured using a pneumatic air pad included in the Model 1030 (SA Instruments, Stony Brook, NY, USA) and used to trigger MRI sequences. The T2-weighted turbo spin echo (T2tse) sequence had the following settings: repetition time (TR) = 3,000 ms, echo time (TE) = 205 ms, averages = 1, echo train length = 161, scan time ≈ 8 – 10 min, number of slices = 112, matrix size = 160 x 256, voxel size (mm³) = 0.22 x 0.22 x 0.22. The DW images utilized a half-Fourier acquisition single-shot turbo spin echo sequence with multiple b-values (200; 400; 600; 800; 1,000) and the following settings: TR = 5,000 ms, TE = 112 ms, averages = 4, echo train length = 256, total

scan time \approx 8 – 12 min, number of slices = 7 – 12, matrix size = 148 x 192, voxel size (mm^3) = 0.21 x 0.21 x 1.00.

Histology

After the last imaging time point, mice were removed from the MRI and euthanized by means of cervical dislocation while under deep narcosis, maintaining the position of the tumor. Tumors were removed by first drawing a line parallel to the transversal imaging plane, and a scalpel was used to cut the tumor in half along this line. The tumor halves were subsequently placed in 4.7% neutral buffered formaldehyde for 48 h in preparation for histopathology. A Hematoxylin and eosin (H&E) staining was performed on 5 μm thick paraffin embedded sections, which were cut at 300 μm intervals parallel to the transverse imaging plane, and examined by an experienced mouse pathologist. To validate the GMM, histology slides were digitized with the NanoZoomer 2.0 HT (Hamamatsu, Hamamatsu, Japan) and regions of interest were manually drawn around the entire tumor and the necrotic regions using the drawing tool included in the NanoZoomer Digital Pathology software as shown in supplementary Figure 2 (Hamamatsu, version number: 2.3.1.0). Necrotic and viable portions of the histology were reported as a fraction of the total tissue slice.

Image Registration

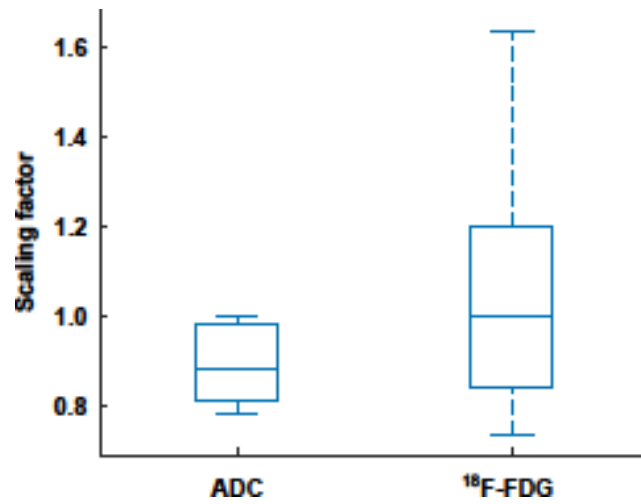
ADC maps were calculated from DWI using custom software developed in MATLAB 2013b (MathWorks, Natick, MA, USA). PET images were co-registered to MR images using a marker-based, semi-automatic co-registration tool in PMOD 3.2 (PMOD Technologies, Zurich, Switzerland), and the T2se images were an anatomical reference for drawing volumes of interest (VOIs) on each tumor. Special care was taken to exclude the skin of mice during VOI placement on the tumors of

mice in T2-weighted images. The voxel values along with voxel coordinates from each VOI for all PET/MR images were exported and further processed in MATLAB. The T2-weighted tumor voxels were reconstructed into a 3-dimensional volume and the mid-point of the tumor in the transaxial plane was used as a starting point for making visual matches to the first histology slice. Once the first histology slice was visually matched by making small translational and rotational changes to the 3-dimensional volume, subsequent image planes were deemed to correspond to subsequent histology slices based purely on the distance between subsequent histological slices; the distance between subsequent histological slices was determined at the time of cutting with the microtome. To confer spatial agreement between matched histology and the 2-D GMM slices, each was first converted into a binary mask, with ones corresponding to viable tissue and zeros to necrotic tissue. For the histology, hand drawn regions of interest were used to create the mask. For the 2-D GMM the combined probability of viable tissue classes 1 and 2 greater than 0.5 was used. A two dimensional image co-registration algorithm implemented in MATLAB matched the paired histology masks to the 2-D GMM masks. The algorithm utilized regularized gradient descent to decrease the mean square error. The co-registration algorithm was constrained to translational, rotational, and scaling transformations. The Dice coefficient was then calculated for each co-registered histology and 2-D GMM slice pair.

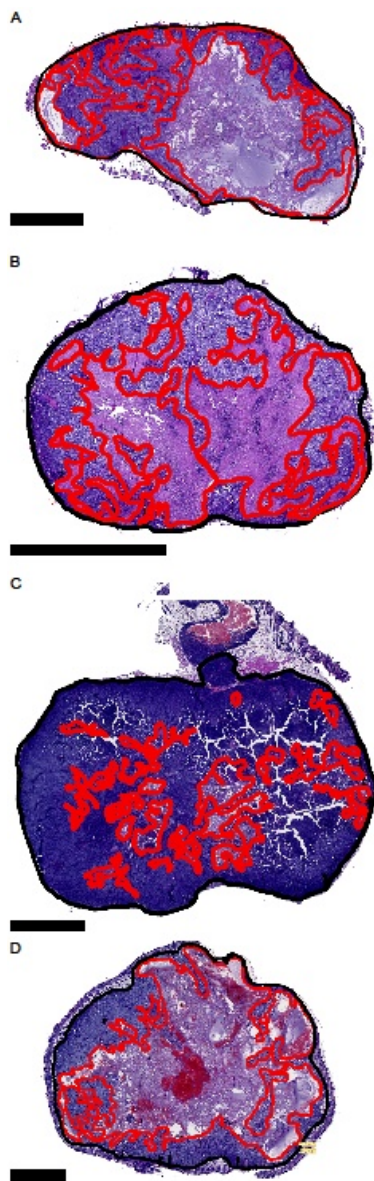
Statistics

To test whether the total tumor volume was a predictor of ADC and ^{18}F -FDG SUV, or if ADC was a good predictor of ^{18}F -FDG SUV for average values in the whole tumor or average necrotic and viable tissues, a linear regression analysis was performed to

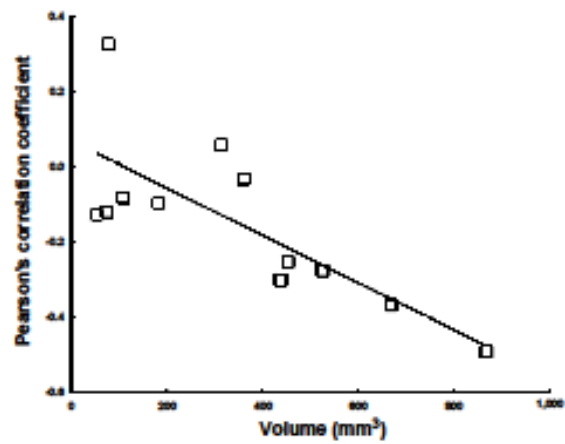
establish whether the slope of the regression line was significantly different than zero. To test the validity of the linear regression analysis, a Durbin-Watson test was performed on the residuals of the straight-line fit to check whether the residuals of the linear regression analysis are auto-correlated. The hypothesis of no auto-correlation was accepted for $p > 0.05$. Afterwards, an analysis of co-variance (ANCOVA) tested whether the groups had different slopes. If the slopes were not different ($p > 0.05$), the slopes were set to the average and the difference in groups was tested. A Tukey-Kramer post hoc test elucidated which groups had significantly different intercepts ($p < 0.05$). The degrees of freedom along with the F-statistic and p-values are reported in supplementary Table 3.



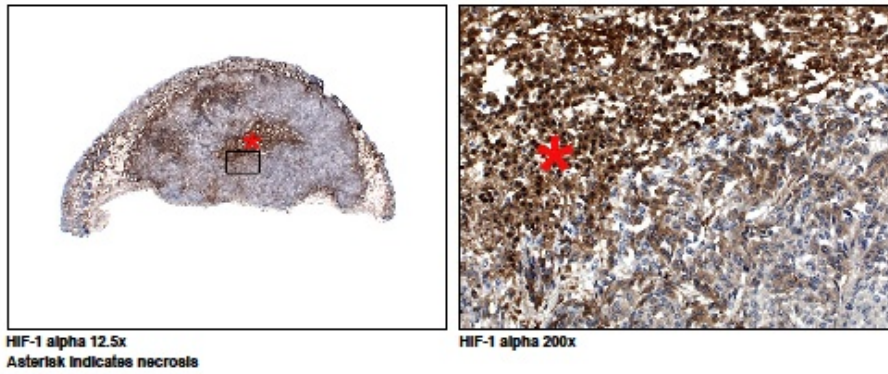
Supplemental Figure 1: An adaptive scaling factor was used in step 2B of the implemented population based GMM model. Since ¹⁸F-FDG measurements are vulnerable to variability, the data from early time points was adaptively scaled by a multiplicative factor to fit the given distribution. Keeping the clusters fixed does not forcefully assign false observations to different clusters and models the tumor progression judiciously.



Supplemental Figure 2: Black contours on H&E stained tumor tissue slices (A-D) include only tumor tissue, exclude skin, and were used to normalize the necrotic fraction. Red contours outlined the necrotic portion of the tumors. Viable fractions were calculated by subtracting the necrotic fraction from unity. The contours have been made thicker for the sake of illustration. The black bars indicate a distance of 2.5 mm.



Supplemental Figure 3: There is a high degree of correlation between the Pearson's correlation coefficient calculated on a voxel-wise basis for ¹⁸F-FDG and ADC and the tumor volume for all mice and time points ($\rho = -0.76$, $p = 0.004$). A straight line fit is shown for visualization purposes.



Supplemental Figure 4: HIF-1 α staining confers information about cells undergoing hypoxic stress. The necrotic areas (red asterisk) present a large amount of unspecific staining, but perinecrotic regions are viable cells positive for HIF1- α . The mixture of viable and necrotic tissues is represented by purple regions around blue necrotic regions in Figure 5 and could represent hypoxic regions. This example is from the same tumor cell line.

Supplemental Table 1:
Linear regression and ANCOVA

Plot	Slope different from zero	Difference in slopes	Difference in regions	Whole tumor	Viable	Necrotic
ADC vs. Whole tumor	F(1,10) = 21.43 p = 0.0009	F(2,30) = 0.98 p = 0.38	F(3,32) = 50.12 p = 1 x 10 ⁻¹⁰	-	p < 0.05	p < 0.05
Viable	F(1,10) = 14.49 p = 0.003			p < 0.05	-	p < 0.05
Necrotic	F(1,10) = 10.74 p = 0.008			p < 0.05	p < 0.05	-
FDG SUV vs. Volume tumor	F(1,10) = 6.88 p = 0.02	F(2,30) = 0.92 p = 0.41	F(3,32) = 5.34 p = 0.01	-	-	-
Viable	F(1,10) = 7.93 p = 0.01			-	-	p < 0.05
Necrotic	F(1,10) = 7.46 p = 0.02			-	p < 0.05	-
FDG SUV vs. ADC tumor	F(1,10) = 17.83 p = 0.001	F(2,30) = 11.34 p = 0.0002	-	-	-	-
Viable	F(1,10) = 36.09 p = 0.0001			-	-	-
Necrotic	F(1,10) = 7.42 p = 0.02			-	-	-

Supplemental Table 2:
Dice coefficient for corresponding histology and 2-D GMM slices

Dice coefficient for 2-D GMM		
Mouse	Median	Interquartile range
Mouse 1	0.60	0.12
Mouse 2	0.61	0.12
Mouse 3	0.69	0.01
Mouse 4	0.84	0.02
All mice	0.69	0.24

1. Hamacher K, Coenen HH, Stöcklin G. Efficient stereospecific synthesis of no-carrier-added 2-[¹⁸F]-fluoro-2-deoxy-D-glucose using aminopolyether supported nucleophilic substitution. *J Nucl Med.* 1986;27:235-238.

2.2 Project-2

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

Prateek Katiyar, Mathew R. Divine, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Bernhard Schölkopf, Bernd J. Pichler, and Jonathan A. Disselhorst (2017). *The Journal of Nuclear Medicine* 58.4, pp. 651–657.

The two acquisition methodologies to monitor and quantify the tracer distribution in PET are static and dynamic imaging. Among these, single-time-point or static PET imaging of ^{18}F -FDG is certainly the most widely used modality in nuclear medicine. It provides a spatial map of local metabolic activity, captured and averaged over a defined time interval. The static ^{18}F -FDG PET studies are commonly quantified by calculating the SUV (sub-section 1.3.9), which is subsequently used either for cancer diagnosis or for patient staging. Despite the popularity and simplicity of the SUV, its reliability for clinical and preclinical investigations is frequently questioned [35–37]. For instance, the limitations of the SUV have been shown to introduce unacceptable errors in tumor delineation and region segmentation [116, 117]. The shortcomings of the SUV pose an even bigger challenge for a reliable quantification of the tumor heterogeneity, because fundamentally, probing tumor sub-regions is a more complex problem than tumor delineation, as the former task might involve a precise multi-class classification of a highly composite environment in contrast to the binary decision of cancerous versus non-cancerous regions. In such scenarios, the spatial as well as the temporal distribution of the tracer captured by dynamic PET imaging can be highly beneficial for a voxel-wise tissue characterization. As opposed to a single static measurement, a 60 min long dynamic ^{18}F -FDG PET study measures the

full characteristics of the tracer distribution, such as the early tracer transport and the overall metabolic rate [118]. Although several investigations [119–121] have proven the benefits of dynamic PET imaging over the SUV, it has not been widely adopted in the nuclear medicine community. A major reason behind its limited application is the difficulties associated with the quantification of dynamic PET data. One of the most popular approaches to quantify dynamic PET acquisitions is tracer compartmental modeling [26] (sub-section 1.3.9). It is a deterministic approach to study the temporal distribution of a tracer in a target area/volume and is dependent on an accurate estimation of the arterial input function (AIF) and the measurement of TACs with low noise. Measuring an AIF accurately is a tedious and invasive process in humans and extremely difficult in mice that nearly prohibits its application in longitudinal studies.

An alternative approach to compartmental modeling is applying machine learning methods directly on PET TACs. Although many studies have applied unsupervised learning methods on dynamic PET data for tumor delineation and region segmentation, their utility for estimating intratumor heterogeneity is yet to be investigated. Therefore, in this work we evaluated the feasibility of spectral clustering (sub-section 1.6.3) for assessing tissue heterogeneity in cancer using simulations of dynamic ^{18}F -FDG PET tumor imaging data. We also compared the performance of PET TACs clustering with compartmental modeling parametric maps and the SUV segmentations, and justified the quantitative results obtained on the simulated data by applying the suggested methodology on an *in vivo* tumor measurement.

To perform an objective comparison between all the clustering schemes, tumor tissue TACs were simulated using the widely accepted compartmental modeling tool COMKAT [122]. An irreversible 2-tissue-compartmental model (figure 1.9, $k_4 = 0$) was implemented to simulate the TACs of three tumor tissue types based on the kinetic modeling parameters reported by Sugwara et al. [123]. Moreover, a population based curve fitting approach was used to generate various clinical AIFs [124]. The full AIF dataset was kindly provided to us by Dr. Dennis

Vriens. Afterwards, to duplicate the simulations of clinical tissue classes in pre-clinical scenario, 12 dynamic ^{18}F -FDG PET measurements were made with nu/nu NMRI mice (Charles River, Sulzfeld, Germany) bearing subcutaneous colon cancer. During these measurements, a minimally invasive blood sampling technique was used to estimate the preclinical AIFs [125]. Moreover, H&E and CD-31 (Abcam plc, 330 Cambridge Science Park, Cambridge, UK) histology was obtained from one tumor to corroborate the segmentation results on an *in vivo* measurement. Subsequently, a reasonable range of preclinical kinetic parameters was obtained by fitting an irreversible 2-tissue-compartmental model on the average TAC of the tumor for each acquisition. Thereafter, the preclinical analog of clinical TACs was simulated based on the observed kinetic parameter ranges and the relationships between the parameters of all three clinical tumor tissue classes.

For each tumor tissue type, 2000 TACs were simulated by sampling the kinetic parameters from a truncated Gaussian distribution. Both clinical and preclinical simulations were performed using the following framing protocol: $\{30 \times 2 \text{ s}, 8 \times 5 \text{ s}, 8 \times 10 \text{ s}, 6 \times 1 \text{ min}, 5 \times 2 \text{ min}, 5 \times 10 \text{ min}\}$. A clinical and preclinical example of the simulated TACs is shown in figure 2.3.

Before any further analyses, the simulated TACs were corrupted using different levels of log-normally distributed noise [126]. The noisy TACs were fitted using the implemented 2-tissue-compartmental model and their respective original AIFs to calculate the deviations in estimated kinetic parameters from their true sampled values. The kinetic parameter estimation error was defined as following:

$$\varepsilon(\%) = \begin{cases} \left(-\frac{K_{ptrue}}{K_{pestimated}} + 1 \right) \times 100 & true \leq estimated \\ \left(\frac{K_{pestimated}}{K_{ptrue}} - 1 \right) \times 100 & true > estimated. \end{cases}$$

Where K_{ptrue} and $K_{pestimated}$ are the true and estimated kinetic parameter values.

Overall for both the simulation scenarios, k_2 and k_3 were most affected by noise, with median absolute parameter estimation error ranging up to (22.93, 37.97) % and (27.32, 31.03) % for the preclinical and clinical cases, respectively.

Although not reported in the published article, in addition to noise, we also as-

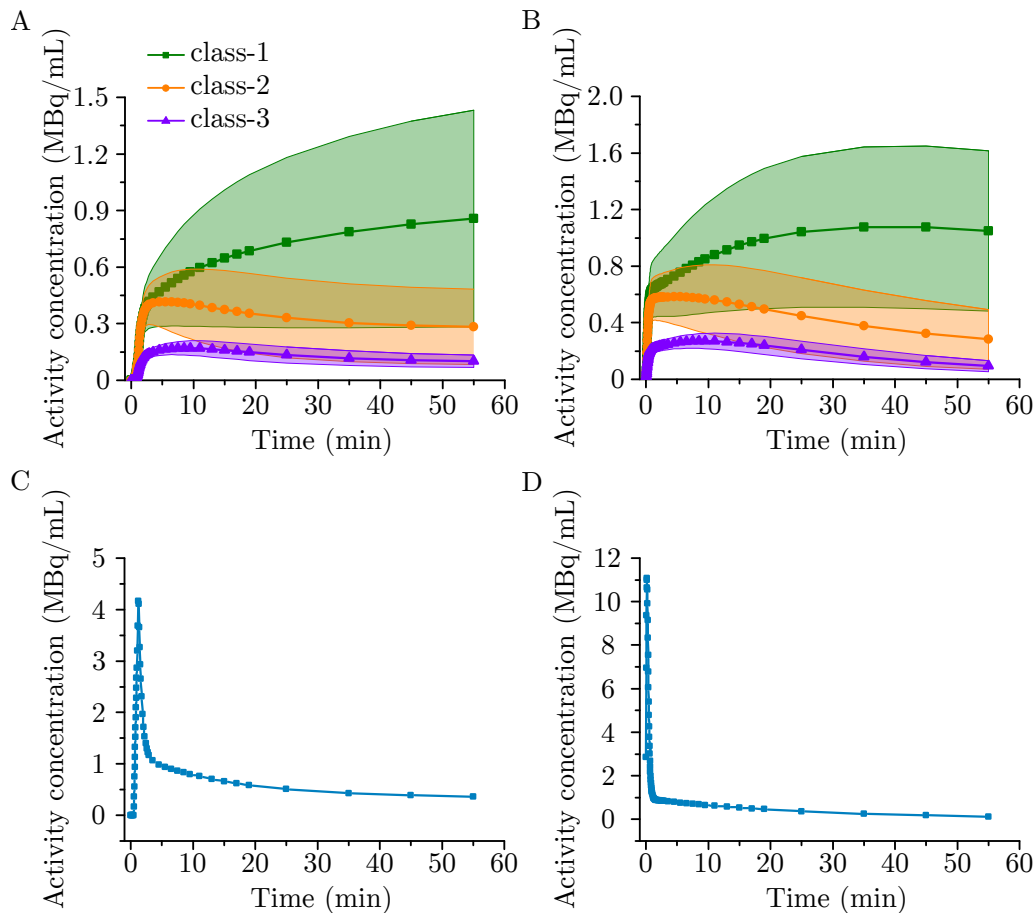


Figure 2.3: Examples of the simulated tumor TACs for (A) clinical and (B) preclinical scenarios. Solid lines and shaded areas depict the mean and unit standard deviation of the color-coded tumor tissue types. The associated clinical and preclinical AIFs are shown in panel C and D, respectively. Figure adapted from Katiyar et al. [102].

sessed the sensitivity of the kinetic rate constants to the most common distortions in the AIFs: under- or overestimation errors in the peak amplitude and delay or advance errors in the peak location of the true AIFs. Figure 2.4 depicts the aforementioned distortions in an exemplary preclinical AIF. The distorted AIFs were used to fit the noise free simulated TACs and estimate the corrupted kinetic parameters. Finally, ε was calculated to measure difference between the true and estimated kinetic parameters.

Figure 2.5 and 2.6 display the effects of changes in the peak amplitude (increase/decrease) and shifts (delay/advance) in the peak location of the true preclinical and clinical AIFs, respectively on the calculation of kinetic rate constants. We found that a decrease or increase in peak amplitude reflected in the overes-

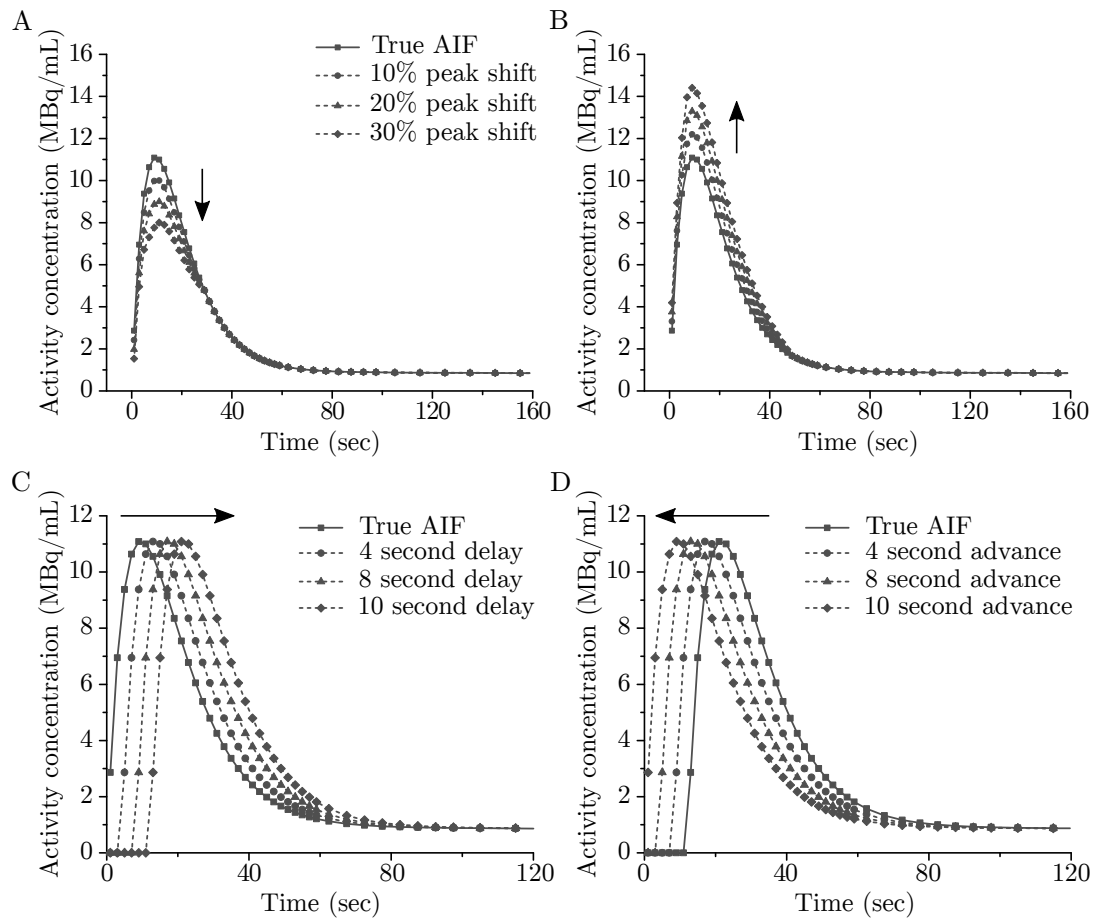


Figure 2.4: True and distorted preclinical AIFs for (A) peak amplitude decrease, (B) peak amplitude increase, (C) peak position delay and (D) peak position advance.

timination or underestimation of the parameters respectively. Similarly, a delayed input function overestimated and an advanced input function underestimated the kinetic parameters. This was also in agreement with previously reported literature [127]. In contrast to the preclinical simulation, clinical parameter estimates were more robust to the distortions in the true AIFs.

The efficacy of spectral clustering in grouping the noisy TACs into their respective tissue classes was evaluated together with two different PET frame weighting schemes. In the first scheme, the frames were weighted with the inverse of their noise variance [128], whereas in the second scheme the Laplacian scoring [99] was used to weigh the most relevant PET time points. In addition to these, to perform an objective comparison with standard PET quantification techniques, spectral clustering and K -means were applied on the estimated kinetic parameters (for noisy TACs) and just K -means was applied on the SUV. The SUV was

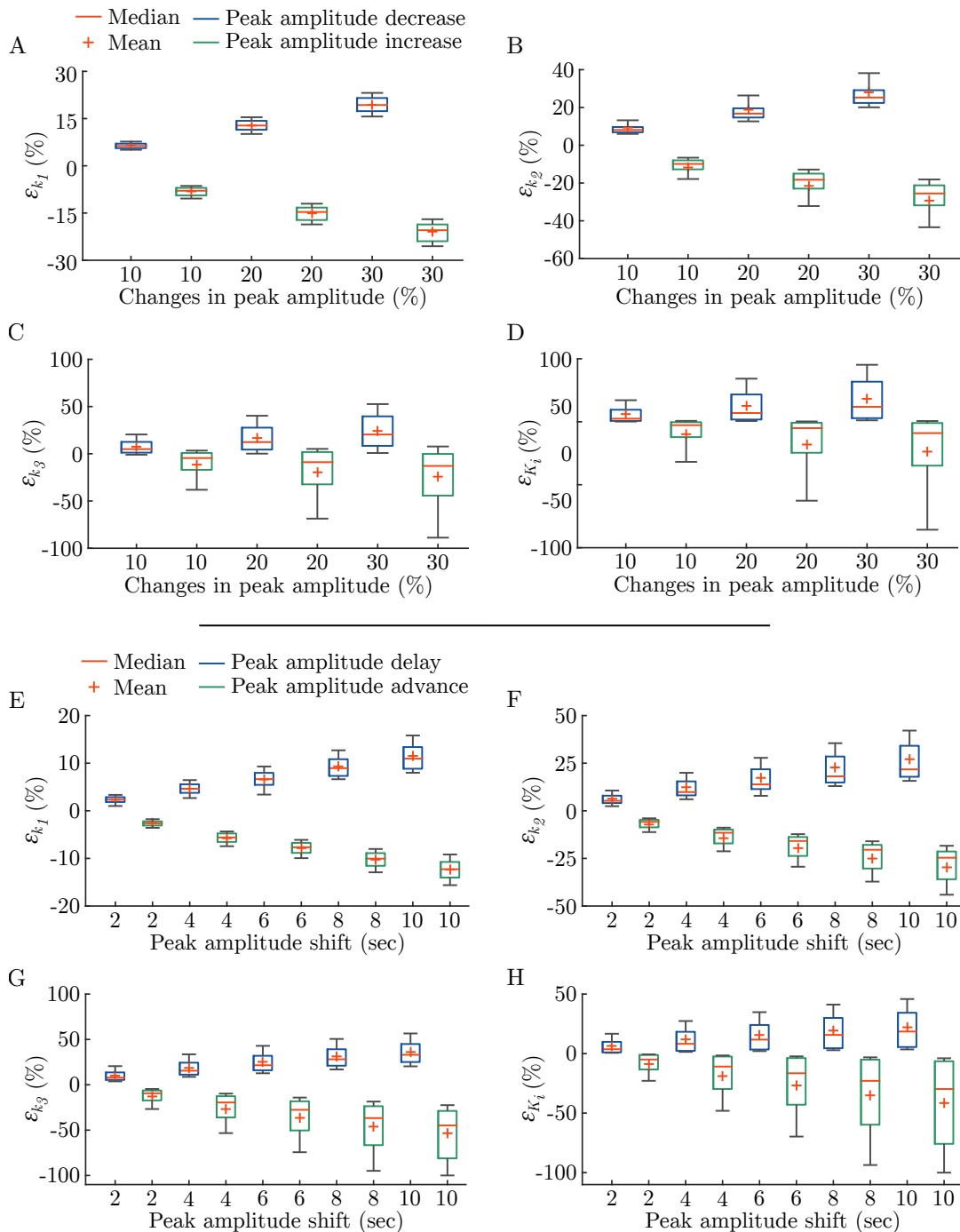


Figure 2.5: Preclinical ε (non-absolute) for K_1 , k_2 , k_3 , and K_i in panels (A)–(D) and (E)–(H) with percent change (increase/decrease) and percent shift (delay/advance) in the peak amplitude of the true AIFs, respectively. The boxes depict the interquartile range and whiskers represent the 10th and 90th percentiles of the data.

obtained by taking the mean of the last two PET time points. Throughout this exercise, the number of eigenvectors was set to 6 and an RBF kernel scale equal to 40 and 55 was used for preclinical and clinical cases, respectively. The scale and

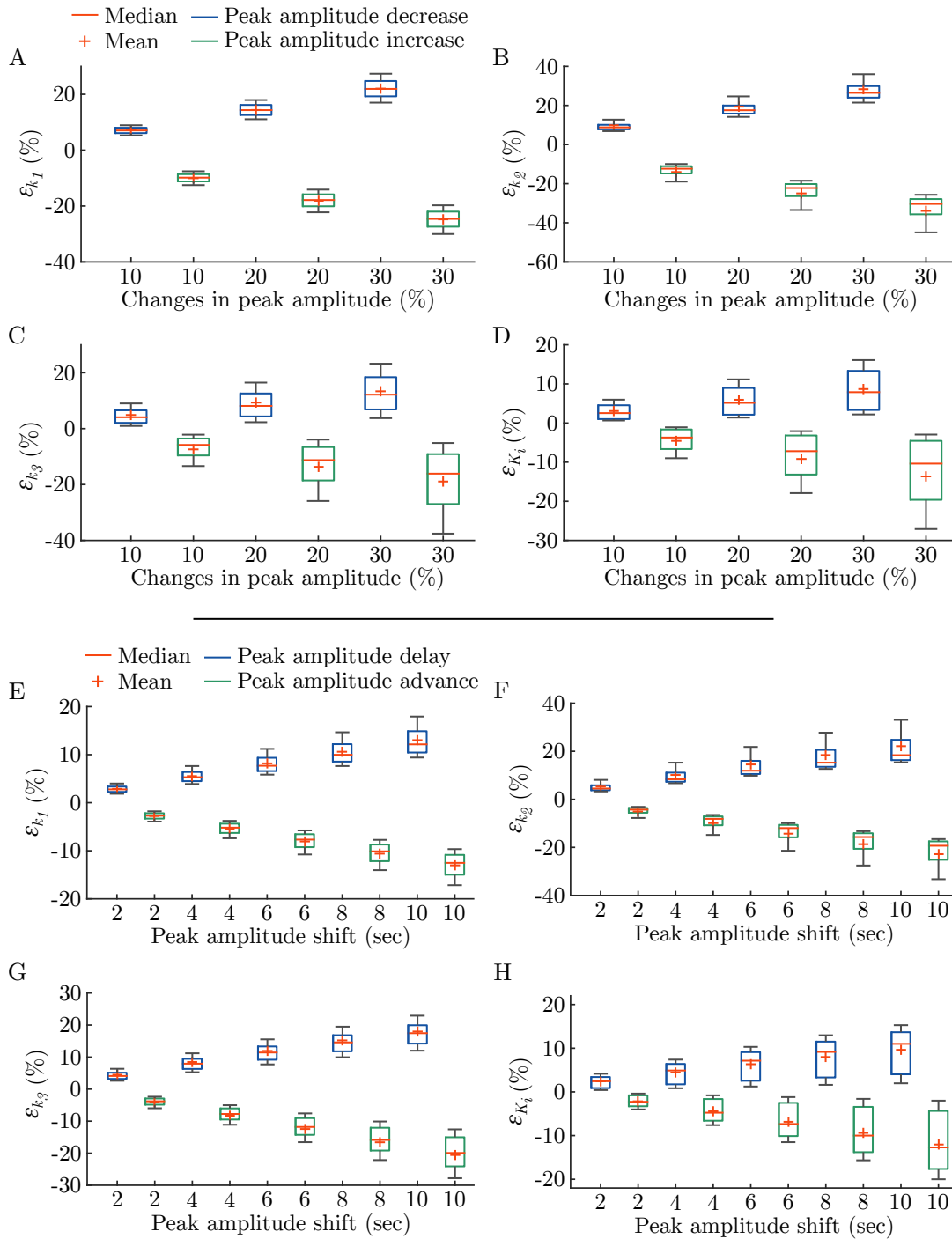


Figure 2.6: Clinical ε (non-absolute) for K_1 , k_2 , k_3 , and K_i in panels (A)–(D) and (E)–(H) with percent change (increase/decrease) and percent shift (delay/advance) in the peak amplitude of the true AIFs, respectively. The boxes depict the interquartile range and whiskers represent the 10th and 90th percentiles of the data.

number of eigenvectors were determined by applying spectral clustering on the noise free data. The performance of each clustering scenario mentioned above was assessed by calculating the tissue misclassification error.

In both preclinical and clinical settings, unweighted and weighted spectral clustering variants applied directly on noisy TACs provided the best clustering results. Although minimally affected by noise, approximately half of the TACs were misclassified by clustering the SUV with K -means. The SUV's poor clustering ability was due to the significant overlap in the last time points of all three tumor tissue classes (figure 2.3A–B). Overall, up to medium noise levels in both the simulation scenarios, the misclassification error of spectral clustering on estimated kinetic parameters was lower than that of the SUV clustering. Among both the weighted spectral clustering methods, the noise variance weighted technique was slightly better than the Laplacian weighted technique for preclinical simulations, whereas the opposite was seen in the analysis of clinical TACs.

To validate the suggested method *in vivo*, the inverse-noise-variance-weighted spectral clustering was applied on the dynamic ^{18}F -FDG measurement of a subcutaneous colon cancer xenograft that was processed to obtain histology and immunohistochemistry matching with the imaging planes. The tumor was largely viable with few focally present high vessel density areas (figure 2.7). The proposed clustering method was applied to identify three clusters in the tumor. As shown in figure 2.7, a high degree of spatial agreement was present between the tissue probability map derived from the suggested approach and the CD-31 immunohistochemistry of the tumor. The two small clusters could be clearly associated with the densely vascularized portions of the tumor. The mean TACs of these clusters also had a higher FDG uptake in early as well as late time points than that of the viable cluster.

In summary, this manuscript carries out an objective evaluation of spectral clustering for identifying intratumor tissues using realistic clinical and preclinical simulations (over varying levels of noise), and makes one to one comparisons with compartmental modeling parametric maps and SUV segmentations. It also shows the adverse effects of the most common input function distortions on the estimation of kinetic rate constants. Although the clinical and preclinical simulations did not involve modeling the respective biologic process (for instance the tumor

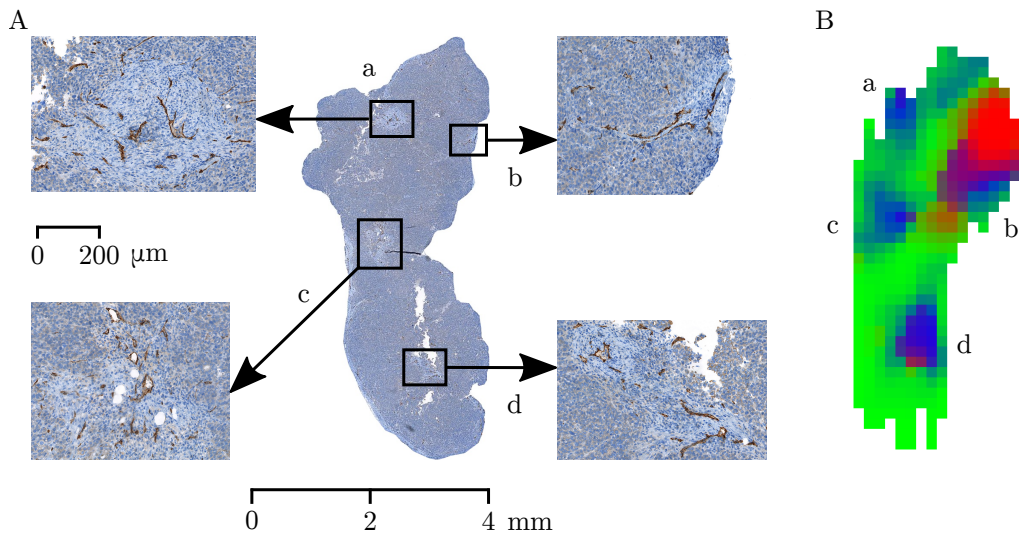


Figure 2.7: Clustering of *in vivo* dynamic PET tumor data. (A) CD31-stained immunohistochemistry of a tumor. The four insets show highly perfused areas of the tumor. (B) Tissue probability map obtained after applying the inverse-noise-variance-weighted spectral clustering on the dynamic ^{18}F -FDG PET images of the tumor. The matching regions of interest are annotated (a-d) in both the panels. Figure adapted from Katiyar et al. [102].

microenvironment), it was necessary to evaluate the efficacy of spectral clustering also in preclinical scenario due to two main reasons. First, the differences in the early kinetics of the preclinical and clinical AIFs (figure 2.3) resulted in TACs with different shapes. Since spectral clustering operates on the feature vectors built using individual TACs, it was essential to confirm that the algorithm performs comparably in preclinical settings. Second, preclinical simulations acted as a bridge between clinical simulations and the preclinical experiment. Specifically, they allowed selection of the most optimal spectral clustering scheme (the inverse-noise-variance-weighted SC) that was later applied and validated on the experimental data. The qualitative comparison to the histology presented in figure 2.7 is essentially a proof of concept, which shows how well the segmentation method works spatially on an *in vivo* measurement. An accurate quantification of these segmented intratumor compartments, however, is a challenging problem, as it requires a reliable spatial correspondence between the imaging and histology. Since a PET imaging slice is orders of magnitude thicker than the thin histology slice (800 μm versus 4~6 μm), a non-rigid co-registration between the two and a subsequent region-wise quantification can be an independent investigation on its own.

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

Prateek Katiyar^{1,2}, Mathew R. Divine¹, Ursula Kohlhofer³, Leticia Quintanilla-Martinez³, Bernhard Schölkopf², Bernd J. Pichler¹, and Jonathan A. Disselhorst¹

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

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

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

J Nucl Med 2017; 58:651–657

DOI: 10.2967/jnumed.116.181370

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

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

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

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

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

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

Published online Nov. 3, 2016.

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

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

MATERIALS AND METHODS

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

Clinical and Preclinical Tissue Class Simulation

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

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

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

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

Noisy Time-Activity Curves

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

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

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

SC

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

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

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

TABLE 1
Summary of Kinetic Parameters and Corresponding Truncation Limits Used for the Simulation of Preclinical and Clinical Tumor Tissue Classes

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

Data are mean ± SD.

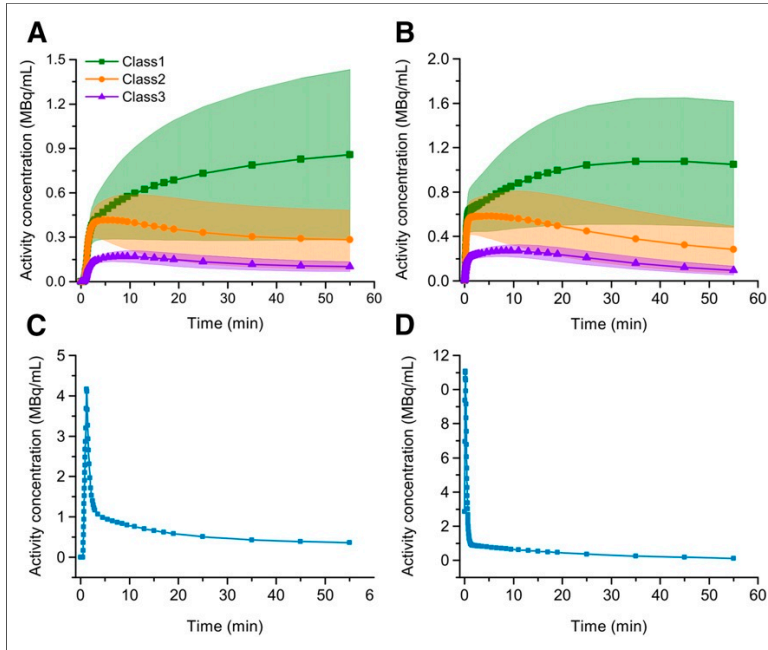


FIGURE 1. Classwise simulated time-activity curves and corresponding AIF for clinical (A and C) and preclinical (B and D) scenarios. Kinetic parameters for each class were sampled from truncated Gaussian distributions. Shaded regions depict the distribution of time-activity curves up to unit SD of the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

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

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

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

PET Frame Weighting

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

Clustering Comparisons

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

Evaluation Metrics

The percentage kinetic parameter estimation error (ϵ) was defined as:

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

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

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

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

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

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

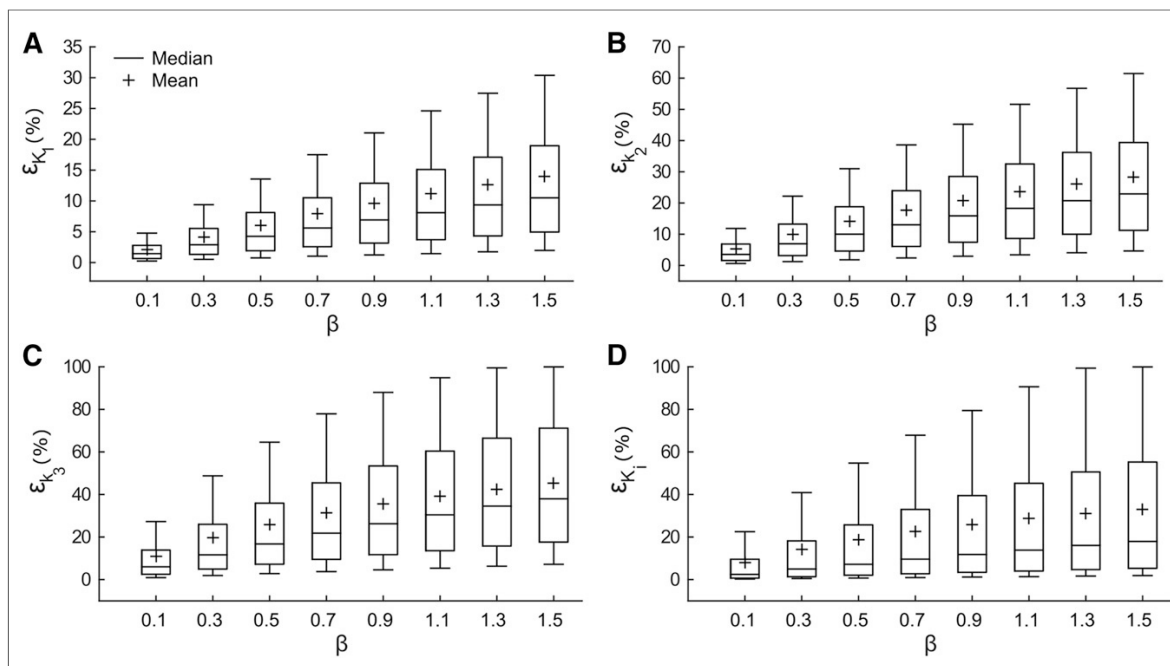


FIGURE 2. Absolute ϵ for preclinical simulations with an increase in the amount of noise (β) for K_1 (A), k_2 (B), k_3 (C), and K_i (D). The boxes depict the interquartile range, and whiskers represent the 10th and 90th percentiles of the data.

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

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

RESULTS

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

Noise Evaluation

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

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

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

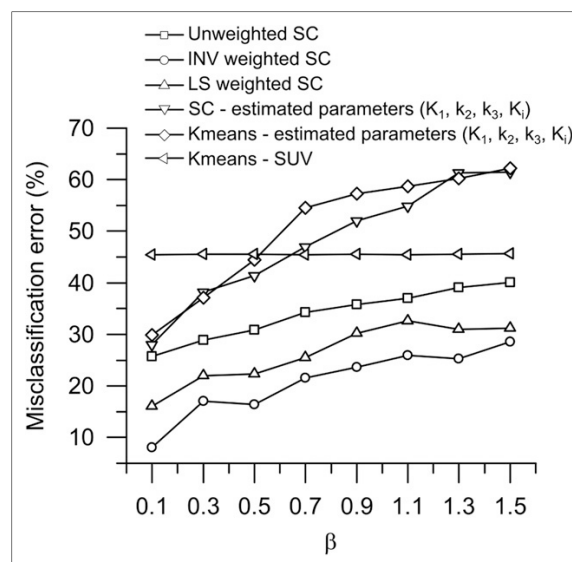


FIGURE 3. Misclassification error of various clustering schemes for preclinical simulations with increase in the amount of noise (β).

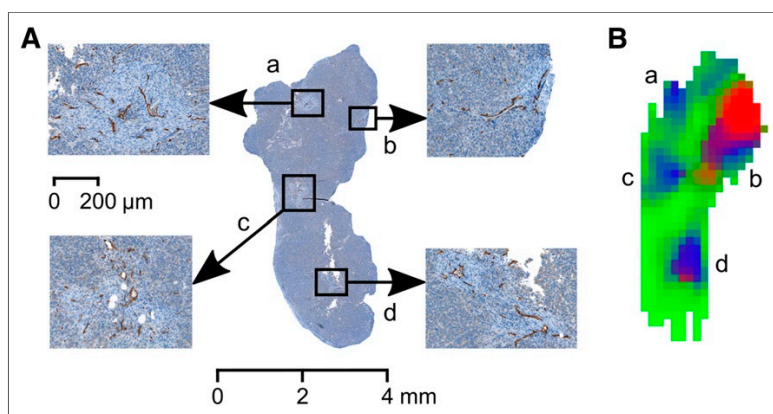


FIGURE 4. (A) CD31-stained histology of a representative tumor; the 4 insets (scale in μm) illustrate high vessel density areas. (B) Segmentation of the tumor into 3 clusters by applying SC on the dynamic ^{18}F -FDG PET data. The matched clusters are marked as a, b, c, and d in A and B respectively.

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

Example

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

DISCUSSION

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

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

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

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

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

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

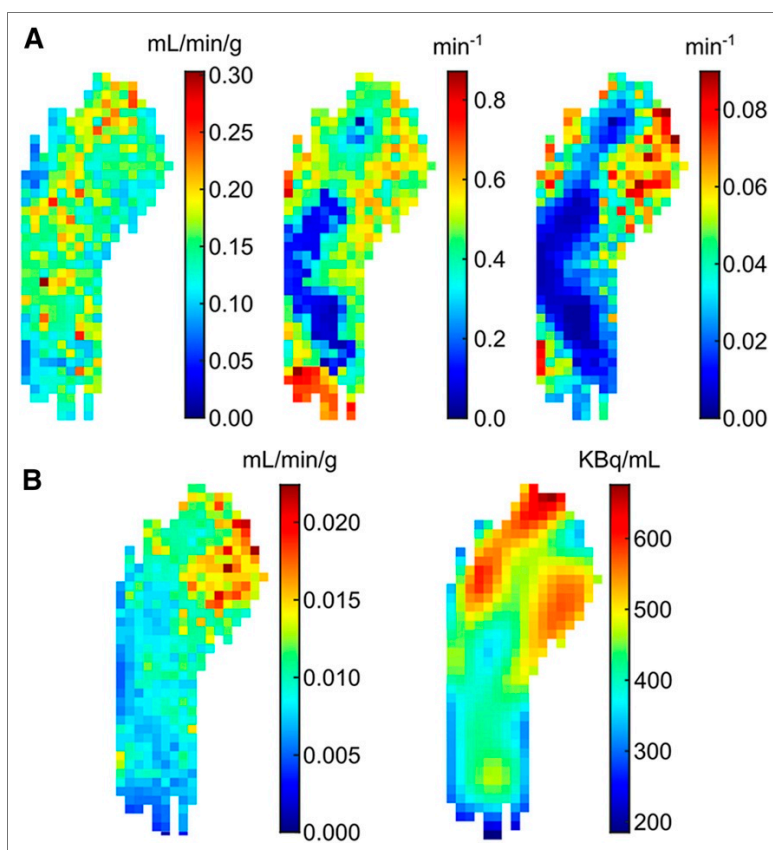


FIGURE 5. (A) Left to right: K_1 , k_2 , and k_3 maps of the tumor shown in Figure 4. (B) Left to right: K_1 map calculated using the parametric maps in A and ^{18}F -FDG uptake in the tumor in the last 20 min of the scan.

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

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

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

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

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

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

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

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

CONCLUSION

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

DISCLOSURE

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

ACKNOWLEDGMENT

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

REFERENCES

- Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013;501:346–354.
- Soussan M, Orihac F, Boubaya M, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. *PLoS One*. 2014;9:e94017.
- Zhao B, Hemann MT, Lauffenburger DA. Intratumor heterogeneity alters most effective drugs in designed combinations. *Proc Natl Acad Sci USA*. 2014;111:10773–10778.
- Jadvar H, Alavi A, Gambhir SS. ¹⁸F-FDG uptake in lung, breast, and colon cancers: molecular biology correlates and disease characterization. *J Nucl Med*. 2009;50:1820–1827.
- Gambhir SS. Molecular imaging of cancer with positron emission tomography. *Nat Rev Cancer*. 2002;2:683–693.
- Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med*. 2009;50:11S–20S.
- Dunnwald LK, Doot RK, Specht JM, et al. PET tumor metabolism in locally advanced breast cancer patients undergoing neoadjuvant chemotherapy: value of static versus kinetic measures of fluorodeoxyglucose uptake. *Clin Cancer Res*. 2011;17:2400–2409.
- McDermott GM, Welch A, Staff RT, et al. Monitoring primary breast cancer throughout chemotherapy using FDG-PET. *Breast Cancer Res Treat*. 2007;102:75–84.
- Wang W, Georgi J-C, Nehmeh SA, et al. Evaluation of a compartmental model for estimating tumor hypoxia via FMISO dynamic PET imaging. *Phys Med Biol*. 2009;54:3083–3099.
- Schmidt K, Lucignani G, Moresco RM, et al. Errors introduced by tissue heterogeneity in estimation of local cerebral glucose utilization with current kinetic models of the [¹⁸F]fluorodeoxyglucose method. *J Cereb Blood Flow Metab*. 1992;12:823–834.
- O'Connor JPB, Rose CJ, Waterton JC, Carano RAD, Parker GJM, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clin Cancer Res*. 2015;21:249–257.
- Foster B, Bageci U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. *Comput Biol Med*. 2014;50:76–96.
- Mouysset S, Zbib H, Stute S, et al. Segmentation of dynamic PET images with kinetic spectral clustering. *Phys Med Biol*. 2013;58:6931–6944.
- Muzic RF, Cornelius S. COMKAT: compartment model kinetic analysis tool. *J Nucl Med*. 2001;42:636–645.
- Sugawara Y, Zasadny KR, Grossman HB, Francis IR, Clarke MF, Wahl RL. Germ cell tumor: differentiation of viable tumor, mature teratoma, and necrotic tissue with FDG PET and kinetic modeling. *Radiology*. 1999;211:249–256.
- Vriens D, de Geus-Oei L-F, Oyen WJG, Visser EP. A curve-fitting approach to estimate the arterial plasma input function for the assessment of glucose metabolic rate and response to treatment. *J Nucl Med*. 2009;50:1933–1939.
- Feng D, Huang SC, Wang X. Models for computer simulation studies of input functions for tracer kinetic modeling with positron emission tomography. *Int J Biomed Comput*. 1993;32:95–110.
- Ferl GZ, Zhang X, Wu H-M, Kreissl MC, Huang S-C. Estimation of the ¹⁸F-FDG input function in mice by use of dynamic small-animal PET and minimal blood sample data. *J Nucl Med*. 2007;48:2037–2045.
- Logan J, Fowler JS, Volkow ND, Ding YS, Wang GJ, Alexoff DL. A strategy for removing the bias in the graphical analysis method. *J Cereb Blood Flow Metab*. 2001;21:307–320.
- Von Luxburg U. A tutorial on spectral clustering. *Stat Comput*. 2007;17:395–416.
- Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett*. 2010;31:651–666.
- He X, Cai D, Niyogi P. Laplacian score for feature selection. *Adv Neural Inf Process Syst*. 2005;18:507–514.
- Schmitz J, Schwab J, Schwenck J, et al. Decoding intratumoral heterogeneity of breast cancer by multiparametric in vivo imaging: a translational study. *Cancer Res*. 2016;76:5512–5522.
- Jovkar S, Evans AC, Diksic M, Nakai H, Yamamoto YL. Minimisation of parameter estimation errors in dynamic PET: choice of scanning schedules. *Phys Med Biol*. 1989;34:895–908.
- Divine MR, Katiyar P, Kohlhöfer U, Quintanilla-Martinez L, Disselhorst JA, Pichler BJ. A population based Gaussian mixture model incorporating ¹⁸F-FDG-PET and DW-MRI quantifies tumor tissue classes. *J Nucl Med*. 2016;57:473–479.
- Xiang T, Gong S. Spectral clustering with eigenvector selection. *Pattern Recognit*. 2008;41:1012–1029.
- Ham J, Lee DD, Mika S, Schölkopf B. A kernel view of the dimensionality reduction of manifolds. In: *ICML '04 Proceedings of the Twenty-First International Conference on Machine Learning, Banff, Alberta, Canada—July 4–8, 2004*. New York, NY: ACM; 2004:47.

Two-tissue Compartmental Modeling

A simplified two tissue compartmental model for ^{18}F -FDG is shown in Supplemental Figure 6, where $C_p(t)$, $C_1(t)$ and $C_2(t)$ are the time varying plasma, free and bound tracer activity concentrations (Bq/mL); K_1 , k_2 , k_3 and k_4 are the kinetic rate constants which control the rate of tracer exchange between compartments. In order to simulate time–activity curves, an irreversible two tissue compartmental model (i.e., $k_4 = 0$) was implemented. Additionally, the vascular fraction (F_b) was assumed to be zero. The equations for this model can be written in the following manner:

$$\frac{dC_1(t)}{dt} = K_1 C_p(t) - k_2 C_1(t) - k_3 C_1(t)$$

$$\frac{dC_2(t)}{dt} = k_3 C_1(t)$$

$$C_{model}(t) = C_1(t) + C_2(t).$$

Here, $C_{model}(t)$ is the observed activity concentration of the target tissue. The model parameters (K_1 , k_2 and k_3) are estimated by solving the ordinary differential equations and minimizing a weighted least square objective function. The net influx rate (K_i) was computed as:

$$K_i = \frac{K_1 * k_3}{k_2 + k_3}.$$

Preclinical Experiments

Six-week-old Naval Medical Research Institute nu/nu mice ($n = 4$) were ordered from Charles River, Germany and allowed to acclimatize in the on-site animal vivarium before being subcutaneously injected with 4.5×10^6 Colo-205 tumor cells on the right hind leg. The tumor size as well as normal social activity of animals was monitored during the entire study. When the tumors were palpable and showed signs of vascularization with a minimum length of 5 mm, the imaging experiments were started. Before and during tumor inoculation and imaging experiments, mice spontaneously respired 1.5% isoflurane dissolved in 100% O_2 at a flow rate of 0.8 L/min in order to maintain a deep anesthesia. All experiments were carried out in a specific-pathogen-free environment.

Three line sources were placed on the animal holder in order to co-register PET and magnetic resonance imaging (MRI) data. The mice were placed with the tumor in the middle

of the field-of-view of the Inveon dedicated small animal PET scanner (Siemens, Knoxville, TN, USA) and 12.0 MBq of ^{18}F -FDG in 50 μL of 0.9% NaCl was injected in the tail vein at a flow rate of 0.5 mL/min using an automated syringe pump system (Harvard Apparatus, Holliston, MA, USA) directly after starting the PET acquisition. Scans were acquired using manufacturer-supplied software (Inveon Acquisition Workplace, version number 1.5.0.28) for 60 min and reconstructed using OSEM3D/FastMAP (reconstruction software version: 2.5, histogram version: 2.39 and re-binning version: 2.5) with the following framing: {10 \times 2 s, 4 \times 5 s, 2 \times 10 s, 3 \times 1 min, 3 \times 2 min, 5 \times 10 min}. Other reconstruction parameters were as follows: image zoom = 1, image matrix size = 256 \times 256 \times 159 with (0.39 \times 0.39 \times 0.80) mm³ voxel sizes, OSEM3D Iterations = 2, MAP Iterations = 18, Beta = 0.05, Uniform set to Resolution and FastMAP setting on.

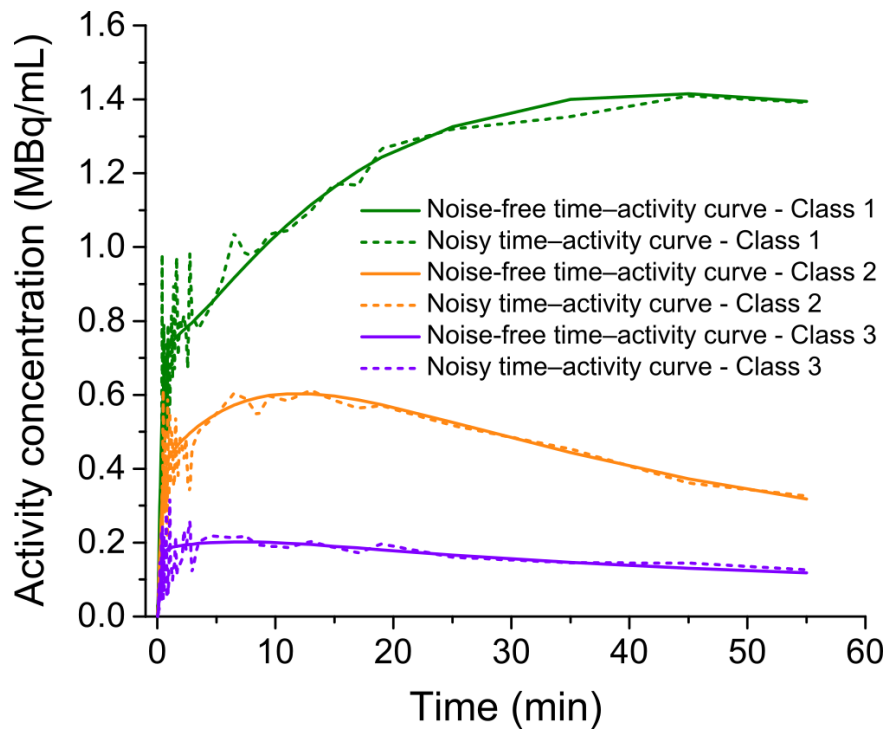
After the PET scans, the animal holder was moved to a 7 T Clinscan (Bruker BioScan, Ettlingen, Germany) small animal MRI while maintaining the position of the mouse. The following settings were used for the T2 weighted turbo spin echo sequence (T2tse): repetition time (TR) = 3000 ms, echo time (TE) = 205 ms, echo train length = 161, image size = 256 \times 160 and voxel size (mm³) = 0.22 \times 0.22 \times 0.22.

PET images were co-registered to MR images using a marker-based, semi-automatic co-registration tool in PMOD 3.2 (PMOD Technologies, Zurich, Switzerland) and the T2tse images were used as an anatomical reference for drawing volumes of interest (VOIs) on each tumor. Special care was taken to exclude the skin of the mice during the VOI placement on the tumors in T2tse images. The voxel values along with the coordinates from each VOI for all PET and MRI measurements were exported and further processed in MATLAB (Mathworks, Natick, MA, USA).

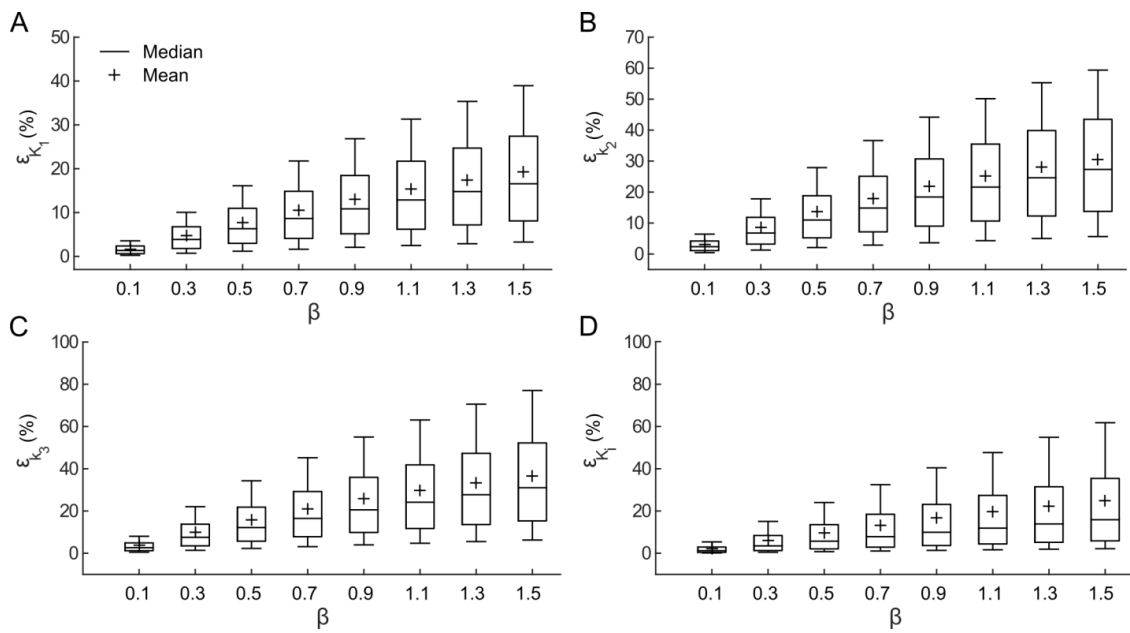
Histology

The histology and immunohistochemistry of one tumor was obtained to validate the results of SC on ^{18}F -FDG measurements. Following the dynamic PET scan, the mouse was sacrificed and a line was drawn on the tumor parallel to the transversal imaging plane. The tumor was removed using a scalpel and sectioned into two halves along the aforementioned line. The individual halves were kept in neutral buffered formaldehyde (4.7% by volume) and embedded into paraffin, before processing for staining. For histology, 3-5 μm -thick sections were cut and stained with haematoxylin and eosin. Immunohistochemical stainings with an anti-CD31 antibody (Abcam plc) were performed on an automated immunostainer (Ventana Medical Systems, Inc.) according to the company's protocols with slight modifications. Appropriate positive and negative controls were used to confirm the adequacy of the staining. Only a rigid co-registration was performed between the histology and imaging data.

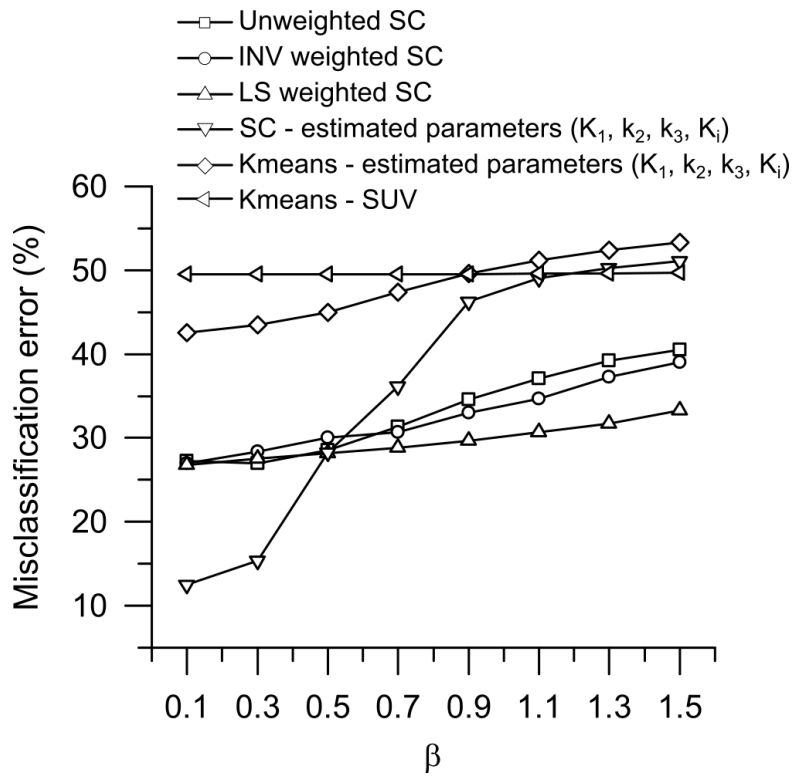
The matching imaging slice was selected based on the visual alignment of the contours of the PET and histology image.



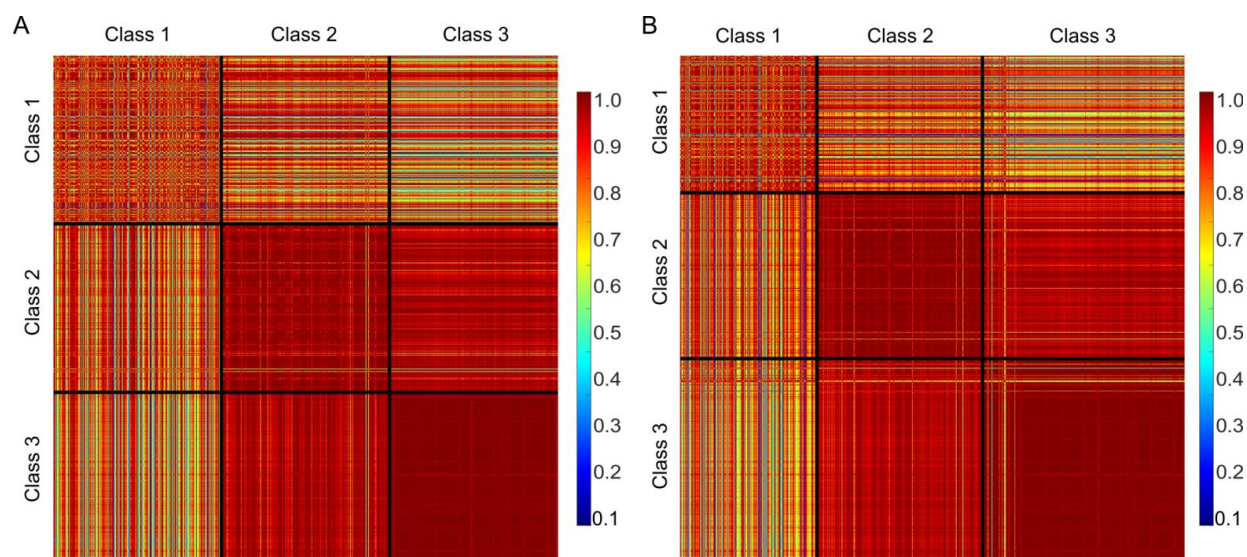
Supplemental figure 1. Preclinical noise-free and noisy (for $\beta = 0.5$) time-activity curve samples of class 1, class 2, and class 3.



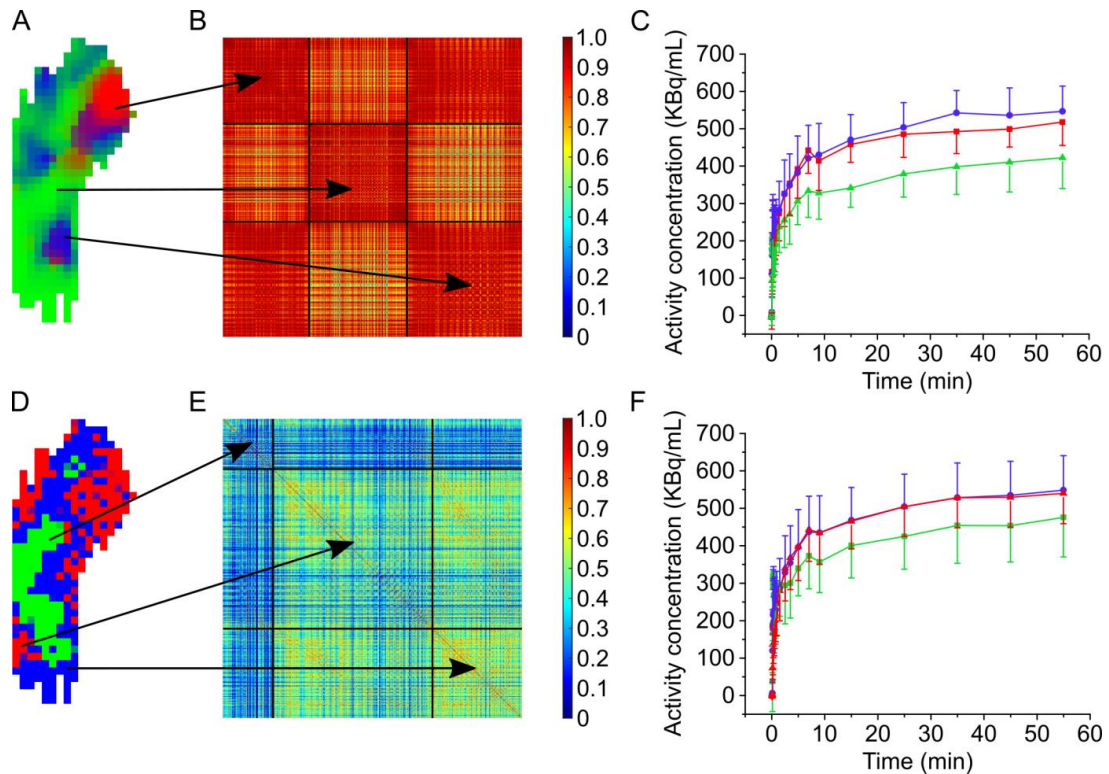
Supplemental figure 2. Absolute ε for clinical simulations with an increase in the amount of noise (β) for K_1 (A), k_2 (B), k_3 (C), and K_i (D). The boxes depict the interquartile range and whiskers represent the 10th and 90th percentiles of the data.



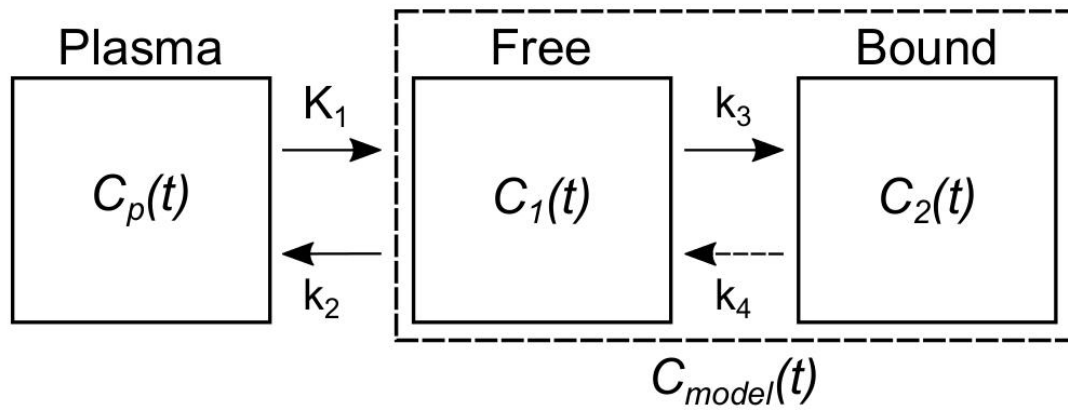
Supplemental figure 3. Misclassification error of various clustering schemes for clinical simulations with an increase in the amount of noise (β).



Supplemental figure 4. Ground truth (A) and clustering (B) affinity matrices of the preclinical noise-free time–activity curves (shown in Fig. 1B). The three block diagonal matrices (top to bottom) depict the intra-class similarity of class 1, 2, and 3. The rest of the two off-diagonal block matrices in each row display the inter-class similarity between the annotated classes. As the time–activity curves of class 2 and class 3 were relatively similar in shape (Fig. 1B), the inter-class similarity between these two clusters was also higher. On the other hand, despite significant overlap, the inter-class similarity between class 1 and class 2 was relatively lower, primarily due the differences in the shape of the simulated time–activity curves.



Supplemental figure 5. (A) SC segmented image of a representative tumor with three clusters obtained using dynamic ¹⁸F-FDG PET data. (B) Affinity matrix of the entire tumor volume computed using the clustering solution in A. The comparable cluster population of the green and blue regions is due to the fact that the end slices of the tumor are densely vascularized in contrast to the center ones. (C) Averaged time–activity curves of the respective clusters. While all three clusters depict high intra-class similarity, the red and blue clusters also contain high inter-class similarity. As the red cluster appears on the periphery of the blue cluster, both the regions also have similar average time–activity curves. (D) Segmentation of the same tumor using the parametric maps (shown in Fig. 5) and SC. (E) Affinity matrix of the entire tumor volume computed using the clustering solution in D and PET time–activity curves. (F) Averaged time–activity curves of the respective clusters.



Supplemental figure 6. A two tissue compartmental model for ^{18}F -FDG PET. The dashed box represents the observed activity concentration of the region.

2.3 Project-3

2.3.1 A Novel Unsupervised Segmentation Approach Quantifies Tumor Tissue Populations Using Multiparametric MRI: First Results with Histological Validation

Prateek Katiyar, Mathew R. Divine, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Bernhard Schölkopf, Bernd J. Pichler, and Jonathan A. Disselhorst (2017). *Molecular Imaging and Biology* 19.3, pp. 391–397.

The unmatched soft tissue contrast of MRI has resulted into its integration as a key diagnostic modality in the clinic [129, 130]. Besides the standard practice of cancer detection and management using anatomical imaging [130–132], the superior benefits of multiparametric MRI (which includes anatomical as well as functional parameters) are increasingly utilized for a better characterization of solid tumors [63, 133]. Alongside the increase in complexity of the imaging data being acquired, the development of robust and automated tools of analysis is also soaring. In this context, a large number of supervised and unsupervised learning approaches have been reported for the differentiation and detection of prostate cancer [134], and brain tumors [135], however, only a handful of techniques have been applied for the characterization of intratumor heterogeneity. Among the published reports, many have utilized variants of K -means [136–139], GMMs [140, 141] and FCM [142] to identify the intratumor viable, hypoxic and necrotic tissues. As discussed in subsection 1.6.3, all of these methods find disjoint clusters by making relatively strong assumptions about the data. In particular, for clusters with non-convex shapes these techniques are likely to fail or provide sub-optimal results (figure 1.25).

Therefore in this work, we proposed an unsupervised segmentation method that overcomes the limitations of the classical clustering techniques and provides a

reliable estimation of tumor tissue heterogeneity. Our suggested spatially regularized spectral clustering (SRSC) algorithm incorporates a penalty term in the clustering objective to penalize the voxels that have inconsistent class (tumor tissue) memberships with their neighbors and obtain spatially coherent segmentation results. We applied SRSC on mice bearing subcutaneous glioblastoma tumors and compared its performance with the three previously mentioned clustering techniques.

The imaging experiments were performed on six NMRI/nu-nu mice (Charles River, Sulzfeld, Germany) that were subcutaneously inoculated with human U87MG glioblastoma tumor cells. A total of six MRI parameters were acquired: T2-weighted anatomy image, ADC map, pre-contrast T2 and T2* maps, and post-contrast T2 and T2* maps. The post-contrast images were acquired 2 minutes after the administration of ferumoxytol (Rienso; Takeda Pharmaceuticals, Glattpark-Opfikon, Switzerland). At the end of each scan, each tumor was carefully sectioned in four approximately equal parts parallel to the axial imaging plane, which were later processed to obtain the following immunohistochemistry stains: CD-31 (Abcam plc, 330 Cambridge Science Park, Cambridge, UK), Ki-67 (Clone SP6, DCS Innovative Diagnostik-Systeme GmbH u. Co. KG, Hamburg, Germany), Glucose transporter 1 (GLUT-1, Abcam Inc., Suite B2304 Cambridge, USA) and cleaved caspase-3 (ASP 175; Cell Signaling Technology, Frankfurt am Main, Germany). Furthermore, H&E histology was obtained. The stained slides were digitized using a NanoZoomer 2.0 HT (Hamamatsu, Hamamatsu City, Japan). Afterwards, with the assistance of seasoned mouse pathologists, the high-resolution histology images were processed to estimate the fractional population of three tumor tissue types. The peri-necrotic regions were identified on GLUT-1 images, and the necrotic and viable tissues were marked on H&E histology.

Prior to clustering, the histology images were rigidly co-registered with the T2-weighted images. The co-registration was performed by selecting the adequate imaging slice and applying in-plane rotations to visually match the contours of the tumor anatomy and histology images. The qualitative registration between the

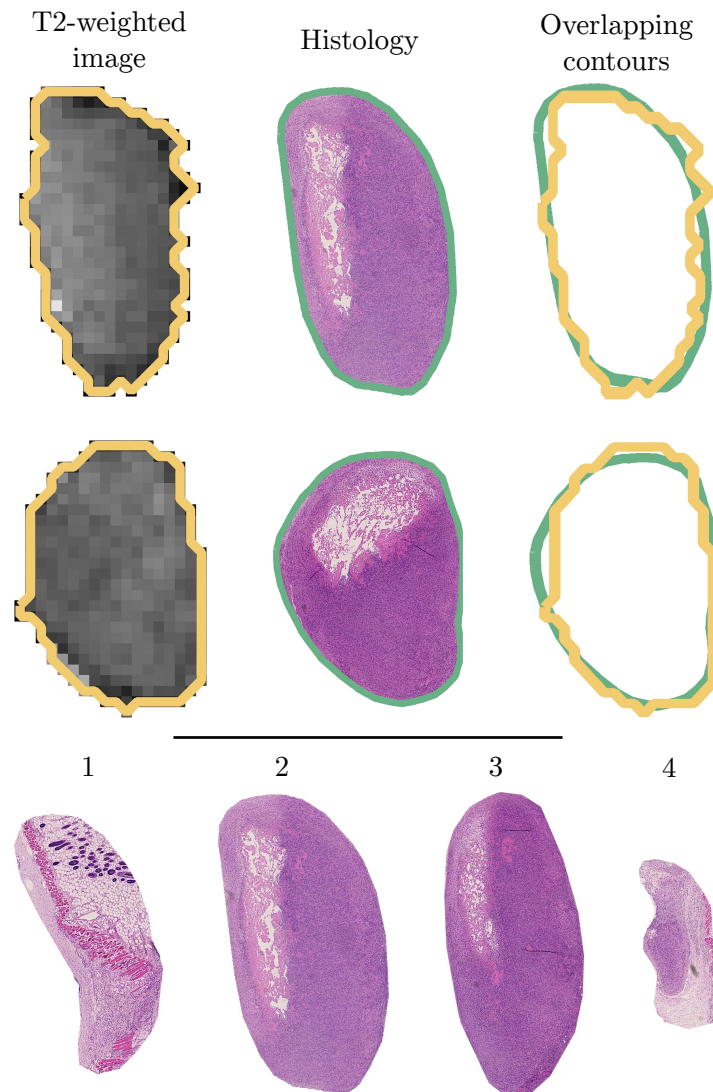


Figure 2.8: Top panel: the co-registered T2-weighted and histology images of two tumors. It is evident from their overlapping contours that the careful sectioning and handling of tumor slices provided a good spatial agreement between the *in vivo* and *ex vivo* images. Bottom panel: the four H&E stained slices (numbered as 1–4) of the first tumor shown above. Note the small amount of tumor tissue in the two end slices.

two images was solely based on the shape of the tumor because little could be inferred about the intratumor tissues using just the anatomical images (figure 2.8, top panel). Additionally, due to lack of tumor tissue or small size of the histology, only the middle histology slices were aligned with the imaging (figure 2.8, bottom panel). One tumor was excluded from further analyses due to substandard registration. For the remaining 5 tumors, a total of 7 histology slices were matched with the respective imaging planes.

Post-registration, the tumors were segmented using SRSC, *K*-means, FCM and

a standard GMM. In SRSC, the top eigenvectors of the Laplacian matrix were segmented using a constrained GMM. In particular, the constrained GMM incorporated a spatial regularization term in the EM algorithm (subsection 1.6.2), which weighted the posterior probability of each voxel with the average posterior probabilities of the 26 connected neighboring voxels. Such weighting scheme maintained a textural regularity in the segmented image by penalizing voxels which had inconsistent tissue class probabilities with their neighbors. *K*-means clustering was run several times on each tumor and the solution corresponding to the lowest cost was considered as best. The parameters of FCM were chosen by maximizing the linear correlation between the histology and clustering tissue fractions. The SRSC results were invariant to any scale value of the RBF kernel between 20 and 500.

Figure 2.9 shows the acquired imaging parameters, histology, SRSC probability map and the solution affinity matrix of an exemplary tumor. It is evident from the H&E and GLUT-1 histology that the viable, necrotic and peri-necrotic areas identified in the SRSC probability map are excellent representative of the actual tissue heterogeneity of the slice. The high similarity within the third block diagonal matrix (corresponding to the viable cluster) of the affinity matrix indicates the presence of homogeneous viable regions in the tumor. Likewise, the low similarity between the necrotic and viable clusters signifies large differences in pathology of both the tissues.

Among the remaining four tumors, two were entirely homogeneous consisting of only viable tissue, whereas the other two had characteristics similar to the tumor shown in figure 2.9. Figure 2.10 shows a qualitative comparison between the clustering results of all the tested algorithms for the three heterogeneous tumors. Although all four methods were able to differentiate between necrosis and viable tissues, only SRSC provided accurate segmentation of peri-necrotic areas.

The Pearson's correlation coefficients between the ground truth (histology) and clustering tissue fractions for all the methods are detailed in table 2.1. The correlation analysis was performed using the tissue fractions obtained from all seven

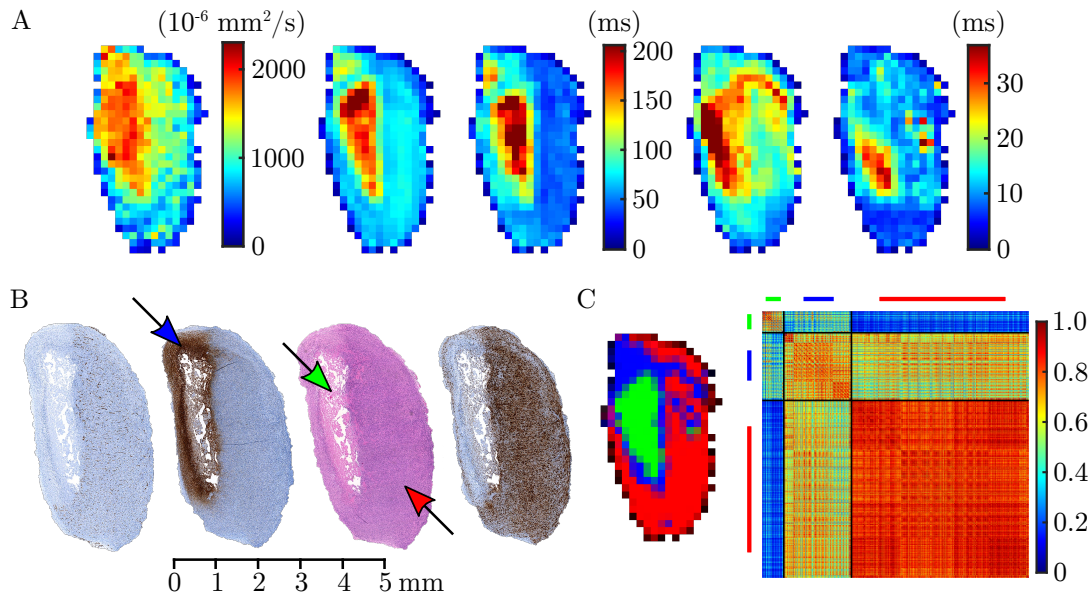


Figure 2.9: SRSC segmentation of a tumor. (A) Left to right: ADC map, pre- and post-contrast T2 maps, and pre- and post-contrast T2* maps of a tumor. (B) Left to right: CD-31, GLUT-1, H&E and Ki-67 histology of the tumor in panel A. (C) Left to right: SRSC probability map and the corresponding affinity matrix organized in cluster-wise fashion. The viable, necrotic and per-necrotic tissues are depicted with red, green and blue colors, respectively. The colored arrows point towards the corresponding tissue in the histology. Figure adapted from Katiyar et al. [103].

co-registered slices. In contrast to SRSC, all the other methods underestimated viable and overestimated peri-necrotic tissue fractions.

Although the correlation coefficients for standard spectral clustering were similar to SRSC, the benefits of spatial regularization are evident from the tissue probability maps shown in figure 2.11. Due to the disturbances caused by magnetic field inhomogeneities and subject motion, the multiparametric MRI data might contain uncertainties at voxel-level. For the experiments performed in this study, only the ADC and anatomy sequences were acquired with tracking of animal breathing, leaving T2 and T2* maps vulnerable to motion artifacts. SRSC overcomes some of these issues by imposing the spatial smoothness constraints. This assumption is fairly reasonable from a biological perspective, as it is highly likely that even in a heterogeneous tumor, neighboring voxels share similar phenotypic characteristics.

Upon close inspection of imaging parameters in figure 2.9, it appears that the T2 values seem to increase in the necrotic tissue after contrast, while there is a decrease in T2* value indicative of perfusion. This observation can be explained

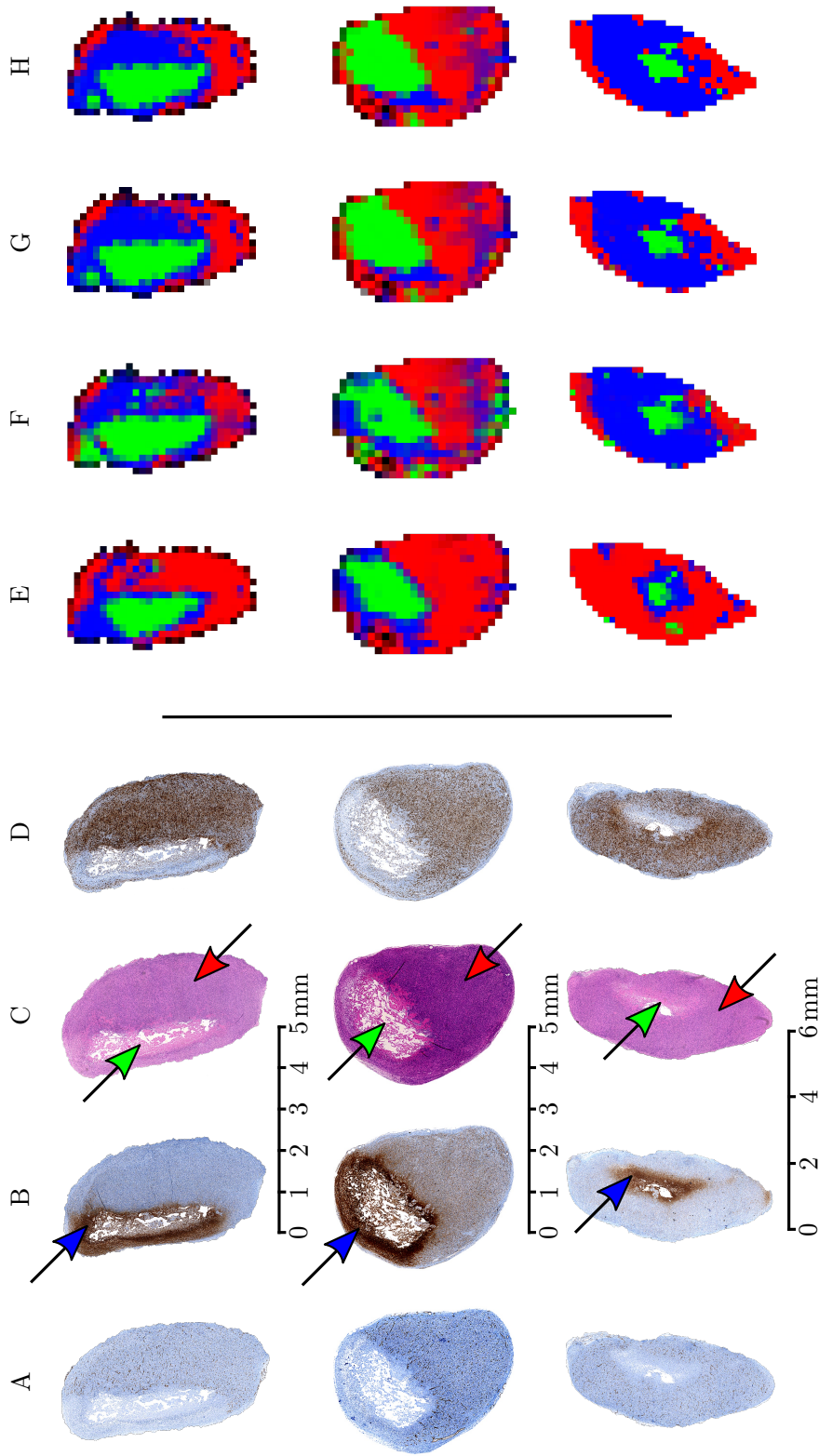


Figure 2.10: A qualitative comparison between the clustering results of all four algorithms. (A) CD-31, (B) GLUT-1, (C) H&E, and (D) Ki-67 histology of three tumors and the respective phenotypic probability maps obtained using (E) SRSC, (F) GMM, (G) FCM, and (H) K-means. The viable, necrotic and per-necrotic tissues are depicted with red, green and blue colors, respectively. The colored arrows point towards the corresponding tissue in the histology. Figure adapted from Katiyar et al. [103].

Tissue type	SRSC	K-means	FCM	GMM
Viable	0.98	0.66	0.33	0.59
Necrotic	0.92	0.76	0.79	0.88
Peri-necrotic	0.82	-0.84	-0.80	-0.72
All	0.99	0.64	0.69	0.68

Table 2.1: Pearson's correlation coefficients between the histology and clustering tissue fractions for all four algorithms. Data taken from Katiyar et al. [103].

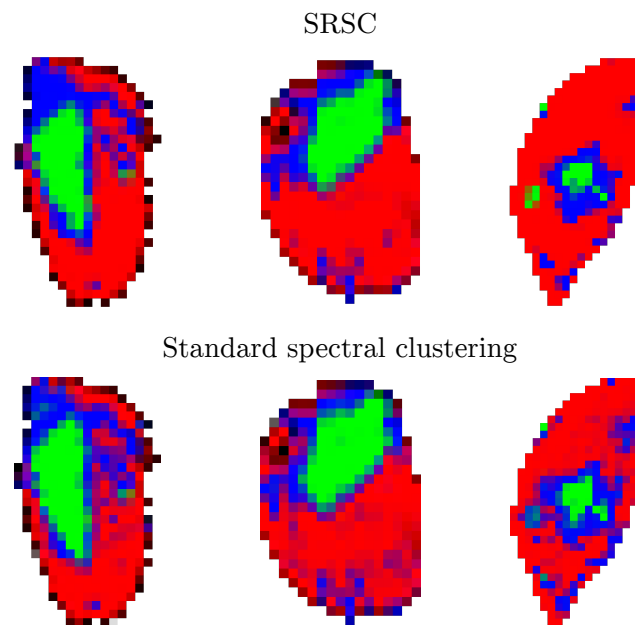


Figure 2.11: Tissue probability maps obtained using SRSC and standard spectral clustering. The viable, necrotic and peri-necrotic tissues are depicted with red, green and blue colors, respectively. For all three tumors, the SRSC tissue probability maps were more spatially consistent than those of the standard spectral clustering.

by the potential physiologic nature of the peri-necrotic tissue. The peri-necrotic areas in all three heterogeneous tumors exhibited high GLUT-1 expression, which is an indirect evidence of induced hypoxia [143]. The hypoxic and acidic regions of the tumors are well known to induce angiogenesis and modulate the vascular permeability factor, resulting in chaotic and highly permeable blood vessels [144]. The two insets of the CD-31 histology shown in figure 2.12 clearly exhibit a high vessel density in the peri-necrotic regions of the tumor. It is quite likely that these blood vessels leaked the contrast agent into the necrotic portions, causing a signal enhancement and reduction in the T2 and T2* maps, respectively. It should

also be noted that an imaging slice ($\sim 200\ \mu\text{m}$) is several orders thicker than a thin histology slice ($\sim 6\ \mu\text{m}$). Thus, the contrast variations seen in figure 2.9 might be caused by the overall vessel population of several of these histology slices.

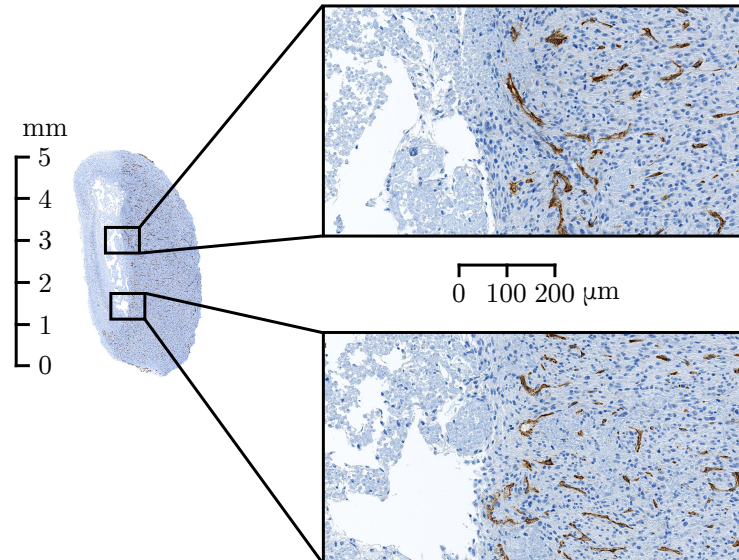



Figure 2.12: CD-31 histology image of the tumor shown in figure 2.9. Both the insets depict densely vascularized necrotic rim of the tumor.

In conclusion, this study proves the benefits of SRSC over the classical clustering methods for the assessment of intratumor heterogeneity. This work is limited in a sense that the suggested method is only tested on a single tumor model with small sample size. In future, it would be worthwhile to apply SRSC on multimodality multiparametric datasets that are acquired on a larger cohort with different cancer types and multiple therapy groups.

BRIEF ARTICLE

A Novel Unsupervised Segmentation Approach Quantifies Tumor Tissue Populations Using Multiparametric MRI: First Results with Histological Validation

Prateek Katiyar ^{1,2}, Mathew R. Divine,¹ Ursula Kohlhofer,³
Leticia Quintanilla-Martinez,³ Bernhard Schölkopf,²
Bernd J. Pichler,¹ Jonathan A. Disselhorst¹

¹Werner Siemens Imaging Center, Department of Preclinical Imaging and Radiopharmacy, Eberhard Karls University Tuebingen, Roentgenweg 13, 72076, Tuebingen, Germany

²Max Planck Institute for Intelligent Systems, Tuebingen, Germany

³Institute of Pathology and Neuropathology, Eberhard Karls University Tuebingen and Comprehensive Cancer Center, University Hospital Tuebingen, Tuebingen, Germany

Abstract

Purpose: We aimed to precisely estimate intra-tumoral heterogeneity using spatially regularized spectral clustering (SRSC) on multiparametric MRI data and compare the efficacy of SRSC with the previously reported segmentation techniques in MRI studies.

Procedures: Six NMRI nu/nu mice bearing subcutaneous human glioblastoma U87 MG tumors were scanned using a dedicated small animal 7T magnetic resonance imaging (MRI) scanner. The data consisted of T2 weighted images, apparent diffusion coefficient maps, and pre- and post-contrast T2 and T2* maps. Following each scan, the tumors were excised into 2–3-mm thin slices parallel to the axial field of view and processed for histological staining. The MRI data were segmented using SRSC, K-means, fuzzy C-means, and Gaussian mixture modeling to estimate the fractional population of necrotic, peri-necrotic, and viable regions and validated with the fractional population obtained from histology.

Results: While the aforementioned methods overestimated peri-necrotic and underestimated viable fractions, SRSC accurately predicted the fractional population of all three tumor tissue types and exhibited strong correlations ($r_{\text{necrotic}} = 0.92$, $r_{\text{peri-necrotic}} = 0.82$ and $r_{\text{viable}} = 0.98$) with the histology.

Conclusions: The precise identification of necrotic, peri-necrotic and viable areas using SRSC may greatly assist in cancer treatment planning and add a new dimension to MRI-guided tumor biopsy procedures.

Key words: Tumor heterogeneity, Multiparametric MRI, Spectral clustering, K-means, Fuzzy C-means, Gaussian mixture modeling

Electronic supplementary material The online version of this article (doi:10.1007/s11307-016-1009-y) contains supplementary material, which is available to authorized users.

Correspondence to: Prateek Katiyar; e-mail: prateek.katiyar@med.uni-tuebingen.de

Introduction

Targeted cancer therapies have experienced an unprecedented increase in approval over the past decade [1], with most recent approaches utilizing the immune system against tumors. However, due to their cytostatic effects, in treatment response evaluation of these therapies, volume- and size-based descriptors (WHO and RECIST criteria) need to be complemented with quantitative imaging biomarkers [2].

A plethora of studies have reported the prognostic value of the multiparametric magnetic resonance imaging (MRI) derived quantitative biomarkers in oncology [3, 4]. Nonetheless, little effort has been laid out in developing techniques to quantify the intra-tumoral heterogeneity. Several investigations have used *K*-means clustering [5] or related algorithms to distinguish necrosis from viable tissue and assess phenotypic variability [6–9]. Kazerooni et al. [10] combined fuzzy *C*-means (FCM) with a region growing algorithm to segment glioblastoma in patients. In addition to these, a recent study [11] has demonstrated the application of Gaussian mixture modeling (GMM) [5] on longitudinal positron emission tomography (PET)/MRI data to create a spatio-temporal profile of different tumor tissue populations.

All of the previously mentioned techniques make strong assumptions about the shape of the clusters and are classified as partitioning clustering algorithms [5]. These methods perform well as long as the clusters are easily separable and their underlying assumptions are met. However, due to highly composite microenvironment and voxel level perturbations, the multidimensional MRI tumor data may contain mixed and irregularly shaped clusters (in parameter space), limiting the applicability of the aforementioned algorithms.

In this paper, we propose a robust algorithm, which overcomes the limitations of the previously described techniques and accurately characterizes the tumor tissue variability. We show that spatially regularized spectral clustering (SRSC) outperforms *K*-means, FCM, and GMM. Furthermore, we quantitatively validate the segmentation results of SRSC on the MRI data (consisting of apparent diffusion coefficient (ADC) maps, normal and contrast-enhanced T2 and T2* maps) using tumor histology.

Materials and Methods

Data Acquisition

All the studies were performed in accordance with the German Animal Welfare Act and protocols were approved by the Regierungspraesidium in Tuebingen. Human U87MG glioblastoma tumor cells were subcutaneously inoculated in the right shoulder of six 11-week-old NMRI/nu-nu mice (1×10^7 in 200 μ l of 0.9 % NaCl). Once injected, the tumors were allowed to grow for 2 weeks, after which the imaging experiments were carried out.

The MRI scans were acquired using a dedicated small animal 7T ClinScan scanner (Bruker BioSpin, Ettlingen, Germany). The

details of the MRI sequences used for the acquisition of T2-weighted anatomy, ADC, T2 and T2* images are provided in the supplementary material. The pre- and post-contrast T2 and T2* images were obtained before and 2 min after the intravenous injection of 50 μ l of ferumoxytol (Rienso; Takeda Pharmaceuticals, Glattpark-Opfikon, Switzerland). To avoid motion artifacts, the animal breathing was tracked (Model 1030; SA Instruments, Stony Brook, NY, USA) and used for triggering the anatomy and ADC sequences. Inveon Research Workplace (Siemens, Knoxville, Tennessee, USA) was utilized to delineate the tumors on the anatomical images of the mice.

Although not included in this paper, during MRI scans, the mice were also injected with 2-deoxy-2- 18 F]fluoro-D-glucose PET tracer for independent investigations.

Histology

At the end of each scan, the mice were taken out from the MRI scanner and sacrificed by cervical dislocation, while maintaining their position on the bed. Prior to dissection, three equidistant lines (~2–4 mm apart) were drawn on the skin parallel to the imaging field of view and the tumors were frozen using a freezing spray. The frozen tumors were subsequently cut into four pieces along the parallel lines, and the sectioned parts were placed into the tissue biopsy baskets while keeping track of the slice orientation. Thereafter, the tissue baskets were placed in 4.7 % neutral-buffered formaldehyde for 48 h and processed for paraffin embedding and subsequent cutting in 6- μ m sections. An automated immunostainer (Ventana Medical Systems, Tucson, AZ, USA) was used to perform the immunohistochemistry with the following primary antibodies: GLUT-1 (Glucose transporter 1, Abcam Inc., Suite B2304 Cambridge, USA), Ki-67 (Clone SP6, DCS Innovative Diagnostik-Systeme GmbH u. Co. KG, Hamburg, Germany), and CD-31 (Abcam plc, 330 Cambridge Science Park, Cambridge, UK). Positive and negative controls were included for the immunohistochemical analysis of each antibody. Additionally, H&E staining was performed. The stained histology slides were scanned into high-resolution digital images using a NanoZoomer 2.0 HT (Hamamatsu, Hamamatsu City, Japan), and different tumor tissue populations were marked by a seasoned mouse pathologist. Utilizing these markings as reference, regions of interest were drawn on the histology slices using NDP.view (Hamamatsu, Hamamatsu City, Japan) and the fractional population of each tumor tissue type was calculated. The viable and necrotic tissues were delineated on H&E, while the peri-necrotic tissue was defined on GLUT-1 immunostaining. The registration (additional details are provided in the supplemental data) between histology and delineated tumor images was performed manually using MATLAB (MathWorks, Natick, MA, USA), as described by Divine et al. [11]. Due to inadequate registration, one mouse was excluded from further analyses.

Spatially Regularized Spectral Clustering

Spectral clustering [12, 13] utilizes voxel-wise MRI feature vectors to create affinity matrix for each tumor. The voxel-wise feature vectors were obtained by concatenating the co-registered MRI parameters (ADC, T2 pre-contrast, T2 post-contrast, T2* pre-

contrast, and T2* post-contrast), and the affinity matrices were constructed using a radial basis function (RBF) kernel:

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

Here, σ is the scale parameter of the RBF kernel and $\|x_i - x_j\|$ is the pairwise Euclidian distance between the MRI feature vectors of voxel i and j . Unsupervised clustering was performed using constrained GMM on the eigenvectors of the normalized Laplacian matrix. Further details about the algorithm can be found in the supplemental data.

Standard GMM probabilistically assigns observations to different clusters and characterizes them using a mean vector and covariance matrix. We included a spatial regularization (in the image space) into standard GMM which we will refer to as constrained GMM. Spatial constraints were imposed by weighing the tissue class probabilities of each voxel by the average tissue class probabilities of the 26 connected neighboring voxels during the optimization process. Thus, the likelihood of a voxel to be characterized as a certain tissue type is enhanced if the nearby voxels belong to the same tissue class and vice versa.

The results of SRSC were compared with K -means, FCM, and standard GMM. All of the methods were implemented in MATLAB.

Statistical Analysis

The Pearson's correlation coefficient was computed to evaluate the linear relationship between the histology and clustering tumor tissue fractions. The one-sample Kolmogorov-Smirnov test was used to test whether the distribution was normal. Due to non-normality, the differences between all of the groups were first checked using the Kruskal-Wallis non-parametric test. In case of a significant difference, the individual groups were compared using the Bonferroni corrected rank-sum tests (a p value less than 0.0167 was considered as statistically significant).

Results

The MRI parameters of one of the tumors are shown in Fig. 1a. The corresponding histology and segmentation results of SRSC are presented in Fig. 1b, c, respectively. The affinity matrix (Fig. 1c) depicts the intra- and inter-class similarity between the identified tissue classes of the tumor.

The clustering comparisons of SRSC with K -means, FCM, and standard GMM, together with the histological images of three tumors are shown in Fig. 2. The proposed method outperformed all three techniques and demonstrated the best visual correlation with the histology. The SRSC results of the remaining tumors are shown in Supplementary Figs. 1 and 2. Both tumors were highly homogeneous and mostly composed of viable portions. SRSC however also identified minor amounts of muscle and connective tissue.

The class-wise box plots and the histograms of all the MRI parameters for the aforementioned tumors are shown in

Fig. 3. The box plots and histograms were generated using the voxel-wise segmentation results of SRSC. For each MRI parameter, the distributions of all three tumor tissue classes significantly differed from each other (Supplementary Table 1).

Table 1 summarizes the Pearson's correlation coefficients (ρ values are given in Supplementary Table 2) between histology and clustering tumor tissue fractions for all four algorithms. Matching scatter plots are shown in Supplementary Fig. 3. While K -means, FCM, and GMM overestimated peri-necrotic and underestimated viable fractions, SRSC accurately predicted the fractional population of all the three tumor tissue types.

Discussion

We proposed a novel algorithm for the analysis of multiparametric MRI data and assessment of the intra-tumoral heterogeneity. We compared our algorithm with the previously reported segmentation methods in MRI studies [6–8, 11] and exhibited its efficacy over K -means, FCM, and standard GMM. We corroborated the segmentation results of SRSC with different histological stainings and demonstrated strong correlations between the tissue fractions derived from immunohistochemistry and SRSC. The precise identification of the necrotic, peri-necrotic, and viable tissue fractions using SRSC highlights the strengths of combining novel image analysis methods with multiparametric imaging and advocates the potential of the proposed method for clinical investigations.

Different types of cell death play an important role in tumor regression and progression. Among others, necrosis is designated as a lethal form of cell demise, which triggers inflammation [14]. Moreover, inflammation is a known regulator of the hallmarks of cancer [15], whose complex interplay promotes uncontrolled tumor growth. Thus, measuring the amount of necrosis can be pivotal for predicting the degree of tumor aggressiveness and cancer morbidity [16]. While all of the models were able to identify necrotic regions, only SRSC provided precise estimates of viable (revealed by high mitotic rate in Ki-67 staining) and peri-necrotic tissue populations. Furthermore, the average ADC in SRSC segmented necrotic (1252.68 ± 628.48), peri-necrotic (1132.2 ± 466.72), and viable (598.46 ± 344.67) regions was also consistent with previous findings [7, 11]. The moderate and relatively low interclass similarity of the peri-necrotic tissue with the viable and necrotic regions respectively, shown in Fig. 1c (affinity matrix), might be a result of metabolic stress, elucidating the gradual transformation of the peri-necrotic regions from viable to necrotic tissue. This is also indicated by the higher expression of GLUT-1 receptor in the peri-necrotic areas in Fig. 1b, possibly due to induced hypoxia [7]. Although we did not perform any hypoxia-specific staining, it is well established that hypoxia leads to an increase in glycolysis, eventually

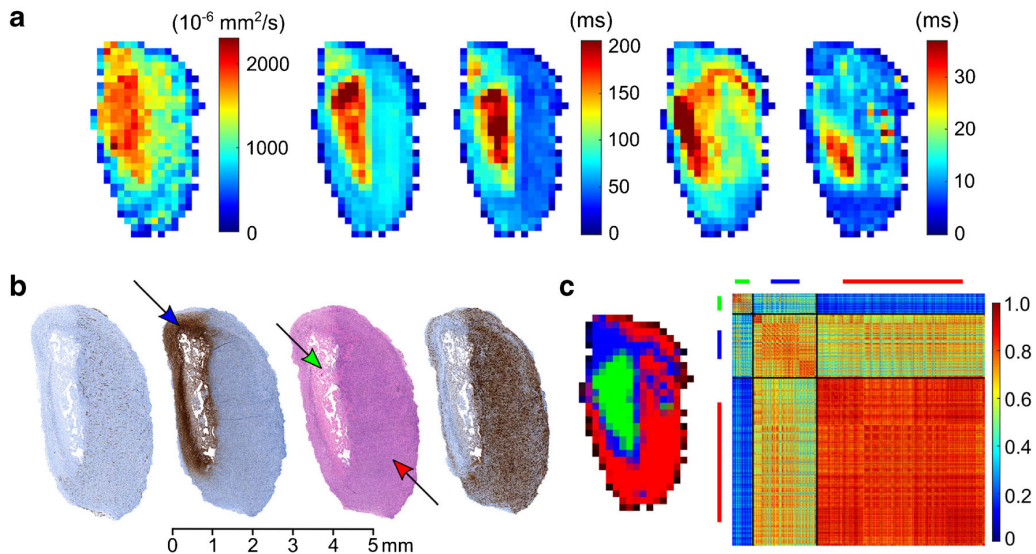


Fig. 1. **a** The acquired MRI parameters of a tumor, left to right: ADC, T2 pre, T2 post, T2* pre, and T2* post maps. **b** Left to right: CD-31, GLUT-1, H&E, and Ki-67 stained histology of the tumor from **a**. **c** Left to right: SRSC probability map and the affinity matrix of the tumor. *Green, blue, and red colors* represent the necrotic, peri-necrotic, and viable tissue, respectively. The *arrows* in the histology indicate the corresponding tissue type in the tumor. The affinity matrix was computed using the voxel-wise feature vectors from the entire tumor volume. The *diagonal* and *off-diagonal matrices* in the affinity matrix depict intra- and inter-class similarities for the labeled tissue clusters, respectively. For example, the high intra-class similarity of viable cluster indicates the presence of homogeneous viable areas in the tumor.

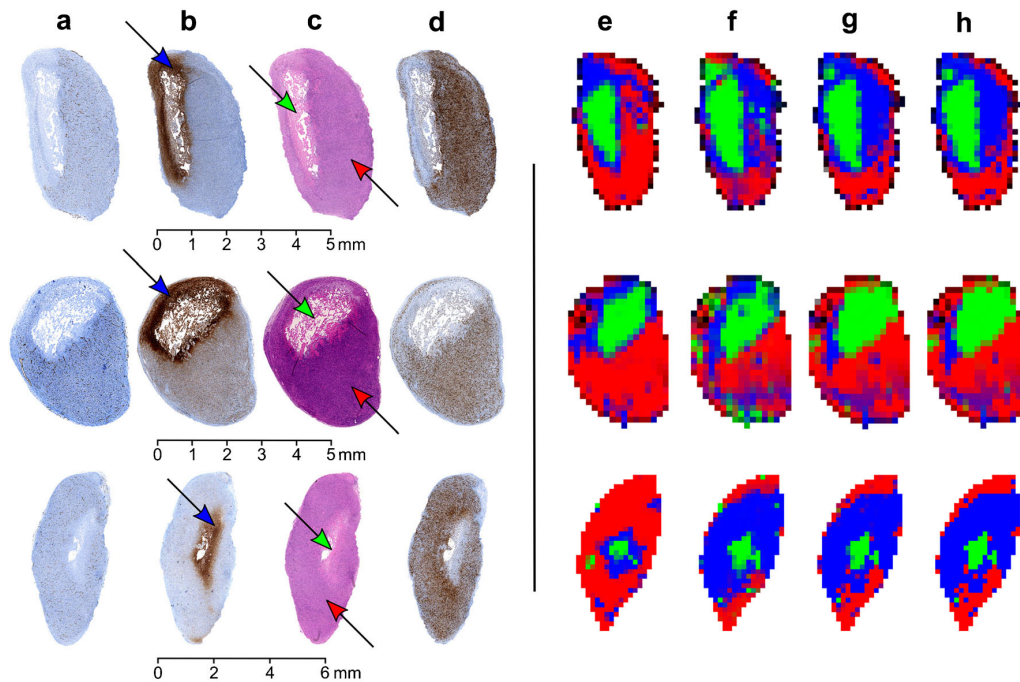


Fig. 2. **a** CD-31, **b** GLUT-1, **c** H&E, and **d** Ki-67 stained histology of three different tumors and corresponding segmentation maps obtained using **e** SRSC, **f** GMM, **g** FCM, and **h** K-means. *Green, blue, and red colors* represent the necrotic, peri-necrotic, and viable tissue, respectively. The *arrows* in the histology indicate the corresponding tissue type in the tumor.

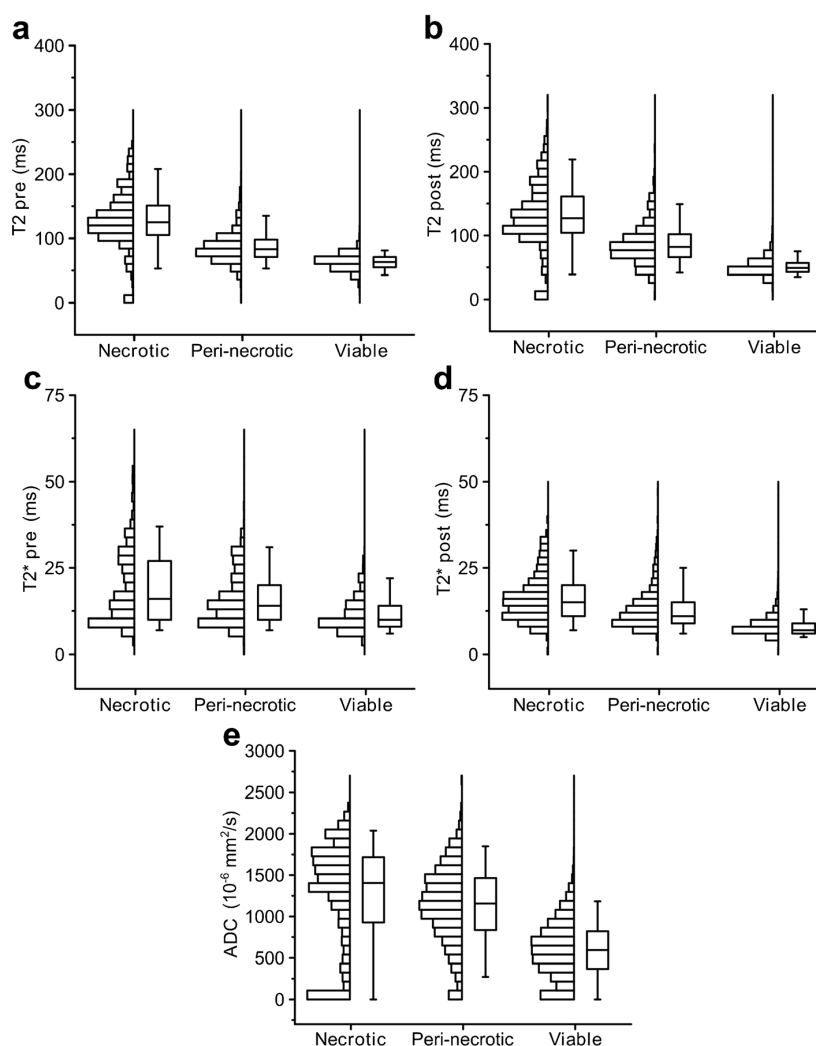


Fig. 3. The distributions of the acquired MRI parameters for each tumor tissue type. The boxes depict the interquartile range and the whiskers extend to the 5th and 95th percent value of the parameter. The line in the box shows the median of the distribution.

resulting into a higher GLUT-1 expression [17]. Similar characteristics are exemplified in Supplementary Fig. 4.

In this study, the histology was manually co-registered with the *in vivo* MRI images. The confounding challenges faced in the registration process [18] were mitigated by the

careful sectioning and fixation of the tumor. Due to the difficulties encountered in one-to-one (histology to imaging) registration, tissue labels from histology are hard to obtain, limiting the application of supervised algorithms. Recently, one investigation has improved this by using a two-step

Table 1. Pearson’s correlation coefficients for the tissue fractions obtained from histology and clustering algorithms

Tissue type	Pearson’s correlation coefficient (r)			
	SRSC	K-means	FCM	GMM
Viable	0.98	0.66	0.33	0.59
Necrotic	0.92	0.76	0.79	0.88
Peri-necrotic	0.82	-0.84	-0.80	-0.72
All	0.99	0.64	0.69	0.68

registration process, involving digital photographs of the specimen and later performing a linear discriminant analysis on the multiparametric MRI data [19]. Another way to circumvent the registration challenges is by using the unsupervised and supervised techniques in a cascaded manner. Specifically, tissue labels can be obtained by clustering the multiparametric MRI data using SRSC and labeled voxels with a probabilistic confidence score can be used to train a supervised classifier, thereby allowing the development of phenotype specific mathematical models.

Biopsies are routinely used in modern cancer diagnostics and tumor phenotyping. As tumors exhibit frequent spatial and temporal heterogeneity, the limited spatial extent of the invasive procedure can severely underestimate the disease complexity, resulting in a misleading prognosis or an unsuccessful therapy [20]. Imaging techniques on the other hand provide a complete view of the patient, allowing a comprehensive inspection of the spatio-temporal variations. Therefore, the combination of imaging diagnostics with tissue biopsy procedures could not only assist in lesion localization and selective tissue sampling, but could also deliver an extensive phenotypic and genotypic profile of the tumor, potentially uncovering the causal relationships between the two [21].

Since the acquired MRI parameters (ADC, T2 and T2* maps) in this investigation are standard protocols in the clinic, SRSC can be translated into clinical examinations. One major limitation of this study, however, is the use of a single xenograft tumor model and the small sample size. Evaluating SRSC on several tumor types along with a combination of different cancer therapies and imaging parameters could reveal the versatility of the suggested method and bring additional insights about the most robust and informative *in vivo* imaging biomarkers.

In multifaceted tumor microenvironment, it is highly probable that the neighboring cells exhibit similar functional and anatomical characteristics and there is a smooth transition from one tissue type to another. MRI measurements, however, can be corrupted by subject motion and magnetic field in-homogeneities, giving rise to voxel level uncertainties. We addressed these textural irregularities by imposing spatial constraints and achieved accurate intra-tumor segmentation results. As opposed to commonly used partitioning clustering algorithms, SRSC makes no *a priori* assumptions about cluster shapes; hence, it is likely to perform better on multidimensional data sets. Such methods of region-wise analyses are of high significance for multiparametric imaging, as they can facilitate biomarker selection and treatment planning by providing a reliable quantification of imaging measures probing inter- and intra-tumor heterogeneity [22].

Conclusion

In conclusion, through quantitative histological validation and one-to-one algorithmic comparison, we demonstrated

the efficacy of SRSC on multiparametric MRI data and delivered an accurate segmentation of the intra-tumoral heterogeneity. Multiparametric imaging in combination with image analysis tools has the ability to probe tumor heterogeneity beyond currently utilized volume- and size-based measures, which might be of great value for selective treatment planning and reliable response evaluation of personalized cancer therapies.

Acknowledgments. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under ERC grant agreement n°323196—ImageLink, German Ministry for Education and Research/Bundesministerium für Bildung und Forschung (BMBF), grant number 0316186E, and Eberhard Karls University Tuebingen (Evaluation of Tumor Heterogeneity Using Clustering of Multi-Modality Imaging Data, Fortuene 2131-0-0). We thank Dennis Thiele for the excellent histological work.

Compliance with Ethical Standards

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical Approval

All applicable institutional and/or national guidelines for the care and use of animals were followed.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Doroshow JH, Kummar S (2014) Translational research in oncology—10 years of progress and future prospects. *Nat Rev Clin Oncol* 11:649–662
- O'Connor JPB, Jackson A, Asselin M-C et al (2008) Quantitative imaging biomarkers in the clinical development of targeted therapeutics: current and future perspectives. *Lancet Oncol* 9:766–776
- Padhani AR, Miles KA (2010) Multiparametric imaging of tumor response to therapy. *Radiology* 256:348–364
- Lemaître G, Marti R, Freixenet J et al (2015) Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Comput Biol Med* 60:8–31
- Jain AK (2010) Data clustering: 50 years beyond K-means. *Pattern Recogn Lett* 31:651–666
- Carano RAD, Ross AL, Ross J et al (2004) Quantification of tumor tissue populations by multispectral analysis. *Magn Reson Med* 51:542–551
- Henning EC, Azuma C, Sotak CH, Helmer KG (2007) Multispectral quantification of tissue types in a RIF-1 tumor model with histological validation. Part I. *Magn Reson Med* 57:501–512
- Berry LR, Barck KH, Go MA et al (2008) Quantification of viable tumor microvascular characteristics by multispectral analysis. *Magn Reson Med* 60:64–72
- Hectors SJCG, Jacobs I, Strijkers GJ, Nicolay K (2014) Multiparametric MRI analysis for the identification of high intensity focused ultrasound-treated tumor tissue. *PLoS One* 9:e99936
- Kazerouni AF, Mohseni M, Rezaei S et al (2015) Multi-parametric (ADC/PWI/T2-w) image fusion approach for accurate semi-automatic segmentation of tumorous regions in glioblastoma multiforme. *Magn Reson mater physics Biol Med* 28:13–22
- Divine MR, Katiyar P, Kohlhofer U et al (2016) A population-based Gaussian mixture model incorporating 18F-FDG PET and diffusion-weighted MRI quantifies tumor tissue classes. *J Nucl Med* 57:473–479

12. von Luxburg U (2007) A tutorial on spectral clustering. *Stat Comput* 17:395–416
13. Ham J, Lee DD, Mika S, Schölkopf B (2004) A kernel view of the dimensionality reduction of manifolds. In: *Twenty-first Int. Conf. Mach. Learn.—ICML '04*. ACM Press, New York, p. 47
14. Scaffidi P, Misteli T, Bianchi ME (2002) Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature* 418:191–195
15. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144:646–674
16. Hiraoka N, Ino Y, Sekine S et al (2010) Tumour necrosis is a postoperative prognostic marker for pancreatic cancer patients with a high interobserver reproducibility in histological evaluation. *Br J Cancer* 103:1057–1065
17. Denko NC (2008) Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nat Rev Cancer* 8:705–713
18. Meyer C, Ma B, Kunju LP et al (2013) Challenges in accurate registration of 3-D medical imaging and histopathology in primary prostate cancer. *Eur J Nucl Med Mol Imaging* 40:72–78
19. Riches SF, Payne GS, Morgan VA et al (2015) Multivariate modelling of prostate cancer combining magnetic resonance derived T2, diffusion, dynamic contrast-enhanced and spectroscopic parameters. *Eur Radiol* 25:1247–1256
20. Bedard PL, Hansen AR, Ratain MJ, Siu LL (2013) Tumour heterogeneity in the clinic. *Nature* 501:355–364
21. Turkbey B, Choyke PL (2014) Decade in review—imaging: a decade in image-guided prostate biopsy. *Nat Rev Urol* 11:611–612
22. O'Connor JPB, Rose CJ, Waterton JC et al (2015) Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clin Cancer Res* 21:249–257

Electronic Supplementary Material

A novel unsupervised segmentation approach quantifies tumor tissue populations using multiparametric MRI: first results with histological validation

Journal: Molecular Imaging and Biology

Prateek Katiyar^{1,2}, Mathew R. Divine¹, Ursula Kohlhofer³, Leticia Quintanilla-Martinez³, Bernhard Schölkopf², Bernd J. Pichler¹, Jonathan A. Disselhorst¹

1: Werner Siemens Imaging Center, Department of Preclinical Imaging and Radiopharmacy, Eberhard Karls University Tuebingen, Tuebingen, Germany

2: Max Planck Institute for Intelligent Systems, Tuebingen, Germany

3: Institute of Pathology and Neuropathology, Eberhard Karls University Tuebingen and Comprehensive Cancer Center, University Hospital Tuebingen, Germany

Corresponding author:

Prateek Katiyar, M.Sc.
Eberhard Karls University Tuebingen
Department for Preclinical Imaging and Radiopharmacy
Werner Siemens Imaging Center
Roentgenweg 13, 72076
Tuebingen, Germany
Telephone: +49 7071 29 83639
Fax: +49 7071 29 4451
e-mail: prateek.katiyar@med.uni-tuebingen.de

Supplementary Materials and Methods

Animal Handling

Six-week-old NMRI nu/nu mice were ordered and allowed to acclimatize in the on-site animal vivarium before subcutaneous injection of tumor cells. Mice were kept in a specific pathogen free (SPF) environment in individually ventilated cages of type 2 long with wood chip bedding. The climate was regulated so that a temperature of $20 \pm 1^\circ \text{C}$ with a relative humidity of $50 \pm 10\%$ was maintained. The lighting was set at a 12 h night and day lighting scheme, and the mice were provided with food and water *ad libitum*. Tumor sizes as well as normal social activity of animals were monitored until tumors were palpable and showed signs of vascularization with a minimum length of 5 mm, before beginning the imaging experiments.

Before and during tumor inoculation and imaging experiments, mice spontaneously respired 1.5% isoflurane dissolved in 100% O_2 at a flow rate of 0.8 l/min in order to maintain a deep anesthesia. For injection of MRI contrast agent, catheters were placed in the tail vein of mice while they were under anesthesia using a 27 gauge micro lance needles. All experiments were carried out in a SPF environment [1].

Data Acquisition

The following settings were used for the T2 weighted turbo spin echo sequence: repetition time (TR) = 3000 ms, echo time (TE) = 205 ms, echo train length = 161, image size = 256 x 160 and voxel size (mm^3) = 0.22 x 0.22 x 0.22. The diffusion weighted images were acquired using a half-Fourier acquisition single-shot turbo spin echo (HASTE) sequence with five b-values (200, 400, 600, 800 and 1000 s/mm^2). Furthermore, the HASTE sequence had following settings: TR = 5000 ms, TE = 112 ms, echo train length = 256, number of averages = 4, image size = 192 x 120 and voxel size (mm^3) = 0.21 x 0.21 x 1.00. For T2 maps, the data were acquired using a spin echo sequence with 12 echo times (TE = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110 and 120 ms) and following parameter values: TR = 2500 ms, number of averages = 2, image size = 192 x 120 and voxel size (mm^3) = 0.21 x 0.21 x 1.00. Similarly, the data for T2* maps were acquired using a gradient echo sequence with 10 different echo times varying from 3 to 26.85 ms with an interval of 2.65 ms. The remaining parameter values were as follows: TR = 2000 ms, number of averages = 2, image size = 192 x 120 and voxel size (mm^3) = 0.21 x 0.21 x 1.00. All maps were calculated with in-house developed software in MATLAB (R2013a) using linear regression on the natural logarithm transformed imaging data. The T2 maps were calculated without the first echo (i.e., 11 echo's in total), whereas the T2* maps were calculated with all 10 echo's.

Histology

A rigid co-registration was performed between the histology and T2W images of the tumors. This was done by selecting the appropriate imaging slice and manually rotating it to match (based on visual correspondence) the corresponding histology section. The visual correspondence was based mainly on the contours of the imaging and histology slices, as little can be inferred about the tumor microenvironment using just the T2W images. Moreover, the co-registration was performed before the clustering analyses were carried out and therefore was not influenced by the segmentation results.

Furthermore, not every histology slice from each tumor could be matched adequately to imaging planes within the respective tumor. This was either due to the small size of the histology or the amount of tumor in the histology section, which corresponded to even fewer imaging voxels, reducing the ability to accurately align the contours. Thus, although clustering was performed on the imaging data from the entire tumor, the imaging to histology correlation analysis was restricted only to the tissue fractions obtained from the co-registered slices. This was also reasonable because the heterogeneity in the co-registered histology sections was not representative of the entire tumor volume.

Spatially Regularized Spectral Clustering

The normalized Laplacian was computed using the following expression:

$$L = D_{inv} * W_{ij} * D_{inv}.$$

Where, $D_{inv} = D^{-0.5}$ and D is the diagonal matrix with diagonal vector $d_i = \sum_{j=1}^n W_{ij}$. The parameter n is equal to the total number of voxels. The number of clusters was chosen based on the visual inspection of the affinity matrix constructed after clustering (i.e. number of block diagonal matrices visible in the affinity matrix). The algorithm was robust to the choice of the scale parameter σ and any value between 20 and 500 yielded the results as shown in Figure 2. The number of eigenvectors segmented using constrained GMM was always equal to the number of clusters. A window size of $3 \times 3 \times 3$ was used to perform the spatial regularization. Due to small size of the tumors, we remained conservative with the selection of window size and did not attempt to optimize it. A total number of 100 iterations were used for the expectation and maximization (EM) step in constrained GMM.

K-means, Standard GMM and FCM

The number of clusters for *K*-means, standard GMM and FCM were kept the same as chosen for SRSC. *K*-means clustering was repeated 50 times, each time with a random seed point and a solution with the lowest within-cluster sum of points-to-centroid distance was selected. Standard GMM was initialized using the clustering solution obtained from *K*-means. An exponent of 1.1 was chosen for the fuzzy partition matrix in FCM. This was obtained by performing a grid search and minimizing the difference between the histology and FCM tissue fractions.

References

1. Kilkeny C, Browne WJ, Cuthill IC, Emerson M, Altman DG (2010) Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol*, 8(6), e1000412.

Supplementary Tables

SUPPLEMENTARY TABLE 1. P-values obtained from different group comparisons made for each of the MRI parameter.

MRI parameter	Group comparison p-value*		
	Viable-Peri-necrotic	Viable-Necrotic	Peri-necrotic-Necrotic
ADC	<0.001	<0.001	<0.001
T2 pre-contrast	<0.001	<0.001	<0.001
T2 post-contrast	<0.001	<0.001	<0.001
T2* pre-contrast	<0.001	<0.001	<0.001
T2* post-contrast	<0.001	<0.001	<0.001

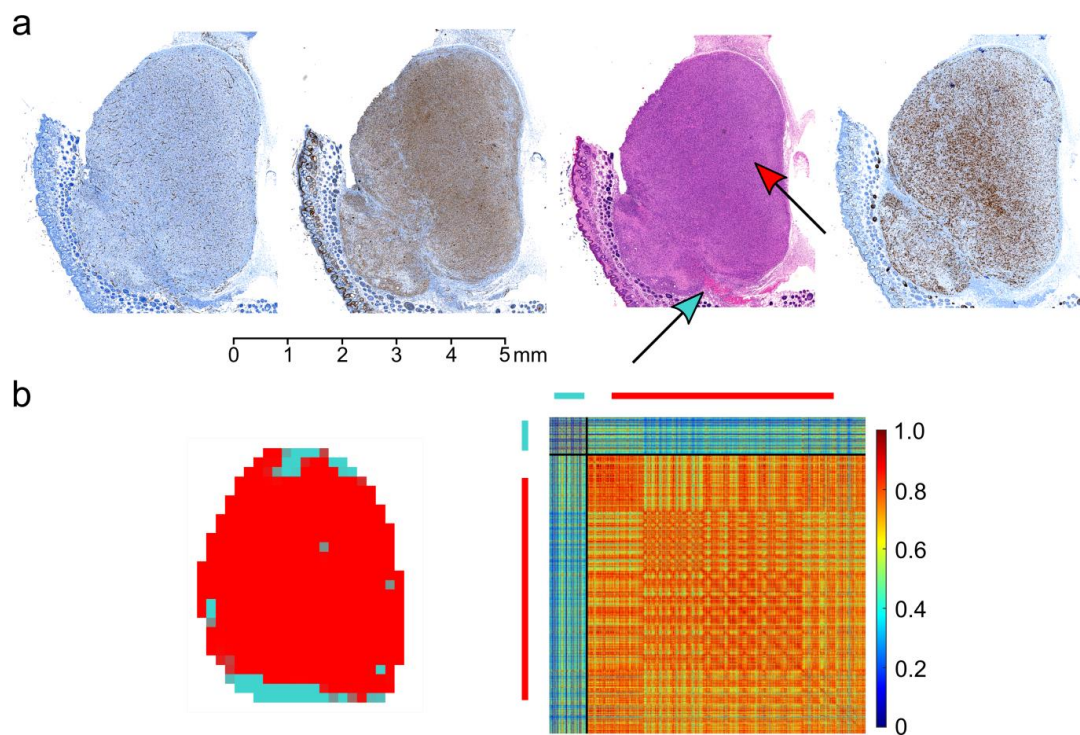
* $p < 0.0167$ was considered statistically significant

SUPPLEMENTARY TABLE 2. P-values of Pearson's correlation coefficient for the tissue fractions obtained from histology and clustering algorithms.

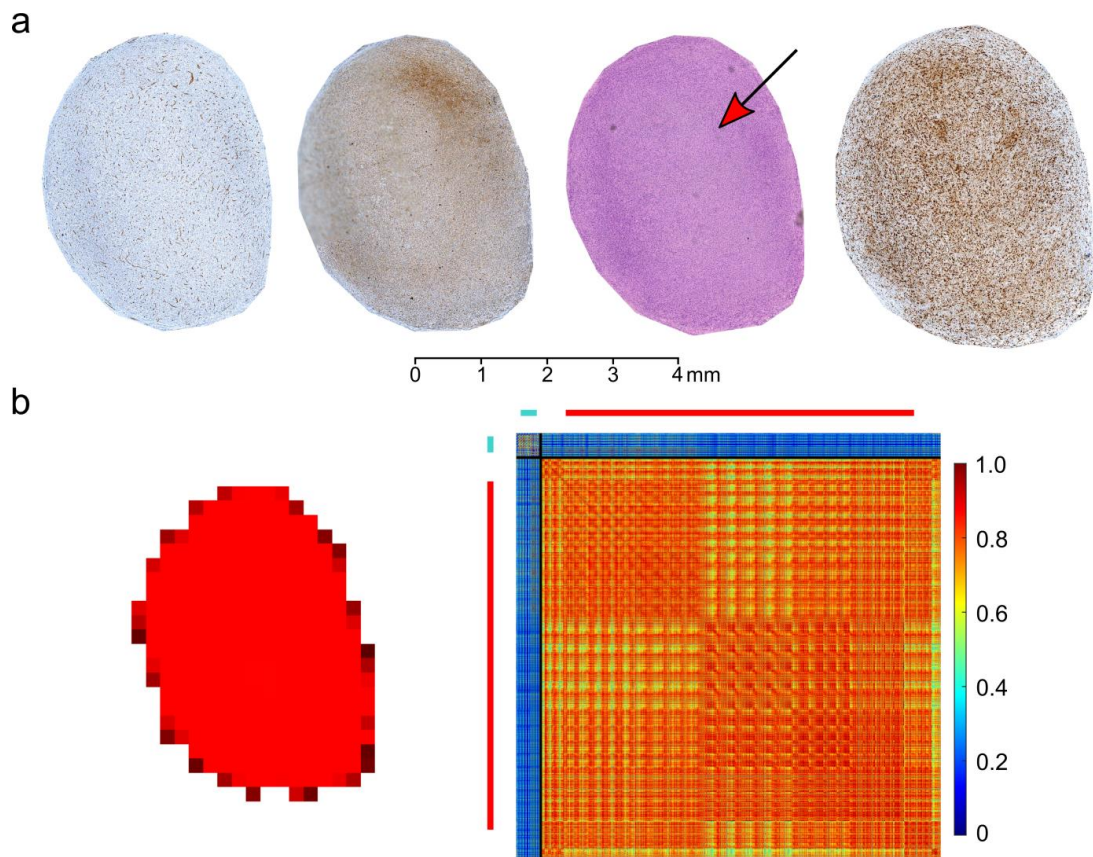
Tissue type	Pearson's correlation coefficient p-value*			
	SRSC	K-means	FCM	GMM
Viable	<0.001	0.104	0.464	0.159
Necrotic	0.026	0.135	0.108	0.048
Peri-necrotic	0.090	0.074	0.104	0.169
All	<0.001	0.006	0.002	0.003

* $p < 0.05$ was considered statistically significant

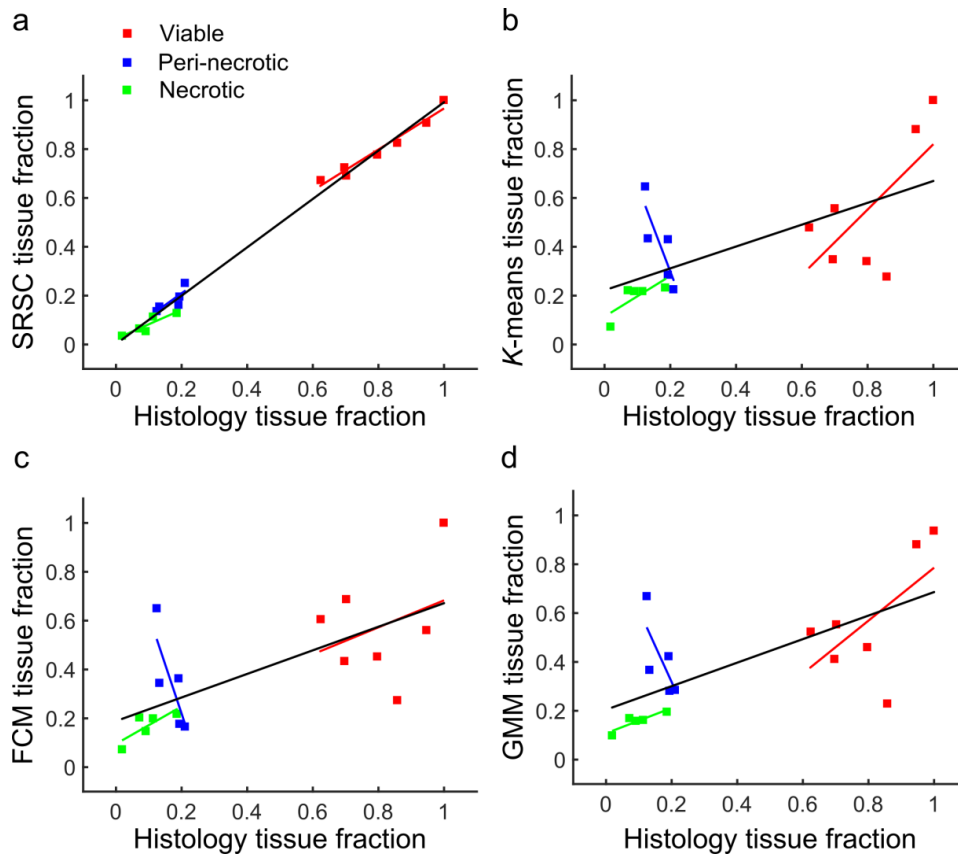
Supplementary Figures



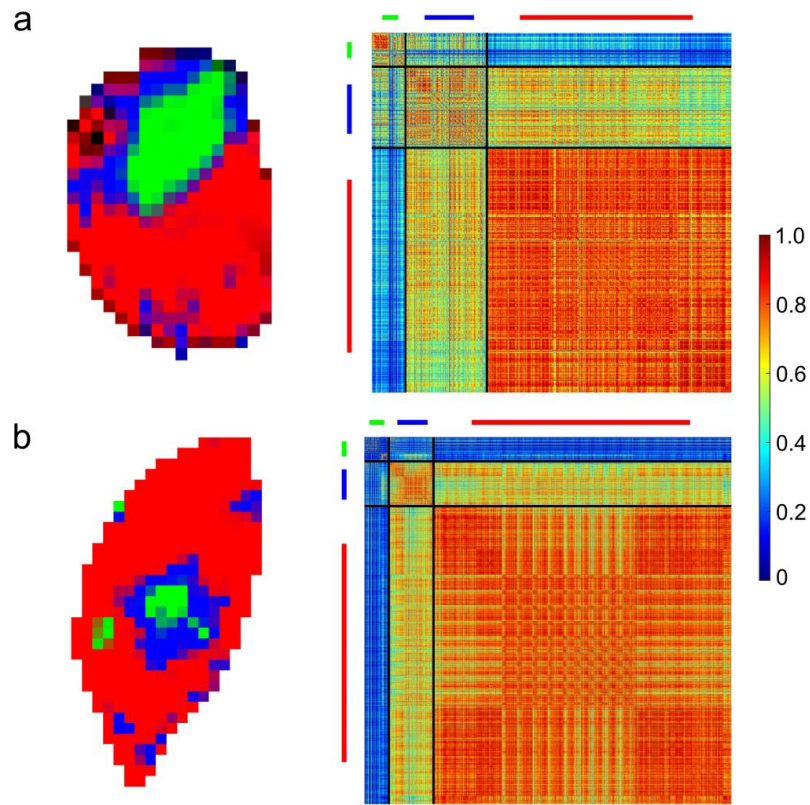
SUPPLEMENTARY FIGURE 1. (a) Left to right: CD-31, GLUT-1, H&E, and Ki-67 stained histology of tumor 4. (b) Left to right: SRSC probability map and the affinity matrix of the tumor. Red and cyan colors represent the viable, and muscle and connective tissue, respectively. The arrows in the histology indicate the corresponding tissue type in the tumor. The affinity matrix was computed using the voxel-wise feature vectors from the entire tumor volume.



SUPPLEMENTARY FIGURE 2. (a) Left to right: CD-31, GLUT-1, H&E, and Ki-67 stained histology of tumor 5. (b) Left to right: SRSC probability map and the affinity matrix of the tumor. Red color represents the viable tissue. The arrow in the histology indicates the corresponding tissue type in the tumor. The affinity matrix was computed using the voxel-wise feature vectors from the entire tumor volume.



SUPPLEMENTARY FIGURE 3. Correlation plots between the histology and clustering tumor tissue fractions obtained using (a) SRSC, (b) K-means, (c) FCM and (d) GMM. The green, blue and red lines depict the linear fit on the necrotic, peri-necrotic and viable tissue fractions. The line in black shows the linear fit on the combined points from all three tumor tissue types. The five points for each tissue type were obtained from the five imaging slices matched with the respective histology of the three tumors presented in Figure 2. The two additional points in the viable group were acquired from the tumors shown in Supplementary figure 1 and 2.



SUPPLEMENTARY FIGURE 4. Left to right: SRSC probability map and the affinity matrix of tumor (a) 2 and (b) 3. The affinity matrices were computed using the voxel-wise feature vectors from the entire tumor volume.

2.4 Project-4

2.4.1 Multi-view learning of multiparametric PET/MRI data quantifies intratumor heterogeneity

Prateek Katiyar, Mathew R. Divine, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Martin Siegemund, Roland Kontermann, Bernhard Schölkopf, Bernd J. Pichler, Jonathan A. Disselhorst

In the previous projects we have seen the utility of dynamic ^{18}F -FDG PET and multiparametric MRI in probing intratumor heterogeneity. In this paper, we went beyond the realm of single modality imaging to prove the superior benefits of simultaneously acquired dynamic ^{18}F -FDG PET/multiparametric MRI data in localizing biologically distinct intratumor tissues and providing early indications of therapeutic efficacy.

Despite the unique principles and complementary nature of PET and MRI, the medical relevance of combining these two modalities is repeatedly questioned [145]. In the field of oncology and in particular tumor heterogeneity, only few investigations have shown the potential of combined ^{18}F -FDG PET SUV and DW-MRI data [101, 113, 114]. Not only are these studies limited in the number of imaging parameters that were acquired using the PET/MRI system, their chosen imaging features also provide redundant information, as they are known to carry high correlations within tumors [101, 113]. Another factor impairing the broad applicability of these reports is the use of unsupervised learning in their analysis of intratumor heterogeneity. Applying supervised learning techniques in this context is quite challenging because of the difficulties in obtaining tissue labels, for example from tumor histology.

Therefore, in this work we advanced well beyond these studies by acquiring dynamic ^{18}F -FDG PET/multiparametric MRI data of mice with subcutaneous colon cancer that received an EGFR-specific apoptosis inducing therapy [146]. Based

on the promising results of spectral clustering (sub-section 1.6.3) on each of the two modalities [102, 103], we further extended its scope to a multimodality setting by devising an algorithm that takes into account the dimensionality and scale differences between the data obtained from the two sources. Thereafter, we combined the proposed clustering approach with an unbiased random forest classifier [147] to develop a phenotype specific mathematical model that provides an accurate *in vivo* evaluation of tumor heterogeneity and therapy efficacy.

The imaging data were acquired using 46 NMRI nu/nu mice (Charles River, Sulzfeld, Germany) that were subcutaneously inoculated with COLO-205 tumor cells (ATCC, Manassas, Virginia, USA). Once the tumors were large enough, the mice were divided in two control (C-24 and C-72) and two therapy (Th-24, Th-72) groups. The therapy animals received Db-scTRAIL [146] injections, whereas the control mice were administered with equivalent quantities of saline. An additional dose was given to the mice in C-72 and Th-72 groups 48 hours later. Imaging experiments were performed approximately 24 hours after the last injection.

The measurements were performed using a hybrid PET/MRI scanner. For each mouse, a 55 min long dynamic ^{18}F -FDG PET scan was performed during which multiparametric MRI images were also acquired. A total of six MRI parameters were recorded: T2-weighted anatomy image, ADC map, pre-contrast T2 and T2* maps, and post-contrast T2 and T2* maps. The post-contrast images were acquired 2 minutes after the administration of ferumoxytol (Rienso; Takeda Pharmaceuticals, Glattpark-Opfikon, Switzerland). The dynamic PET data was reconstructed using ordered subsets expectation maximization algorithm into 22 frames $\{1 \times \sim 30, 8 \times 30, 6 \times 60, 5 \times 300, 2 \times 600 \text{ s}\}$. Due to zero or very low activity, the first frame was discarded from any further analysis. Furthermore, because of artifacts, the pre-contrast T2* maps were also excluded. All remaining functional PET and MRI data were co-registered to the anatomy images that were subsequently used for the delineation of the tumors.

After each scan, the tumors were sectioned in four equal parts (2~3 mm thick), which were processed to obtain H&E and the following immunohistochemistry

stains: CD-31 (Abcam plc, 330 Cambridge Science Park, Cambridge, UK), cleaved caspase-3 (ASP 175; Cell Signaling Technology, Frankfurt am Main, Germany), F4-80 (SP115, Acris Antibodies GmbH, Herdford, Germany) and TK-1 (Abcam plc, 330 Cambridge Science Park, Cambridge, UK). In addition to these, Masson's trichrome staining was also obtained. The histology slides were digitized using a NanoZoomer 2.0 HT (Hamamatsu, Hamamatsu City, Japan) and rigidly co-registered with the imaging planes as described in subsection 2.3.1. The co-registration accuracy was quantified using the Dice similarity coefficient (DSC) [148] and the modified Hausdorff distance [149]. Five mice (1×C-24, 1×C-72, 1×Th-24 and 2×Th-72) were discarded from further analyses due to substandard co-registration and experimental interruptions. The control tumors from C-24 and C-72 groups were pooled to form a single Control group.

Figure 2.13 shows the suggested multi-view learning framework. The tumors from all three groups (Control, Th-24 and Th-72) were divided into a Training and two test sets (Test-1 and Test-2). The histology and imaging data of each Training set tumor was processed using multi-view spectral clustering (MSC) to obtain the histologically validated intratumor labels of viable, apoptosis, and fibrous and necrotic (*rest* cluster) tissues (figure 2.14). The MSC derived probabilistically labeled voxels from all Training set tumors were merged together to obtain a collective training dataset. To enhance the purity of the three clusters in the combined training dataset, voxels with mixed phenotype (i.e. maximum MSC probability < 0.9) were discarded and the remaining data was used to train an unbiased random forest classifier.

MSC is an extended spectral clustering approach for multimodality datasets. The algorithm can be divided in three phases. In the first phase (figure 2.14, gray area), the multiparametric MRI data of each tumor in the Training set was segmented using spectral clustering. The Control tumors were divided into two compartments, viable and $rest_{Control}$, while the Th-24 tumors were segmented into apoptosis and $rest_{Th-24}$ clusters. Figure 2.15 shows a ground truth example of the $rest_{Control}$ and $rest_{Th-24}$ clusters from a Control and a Th-24 Training

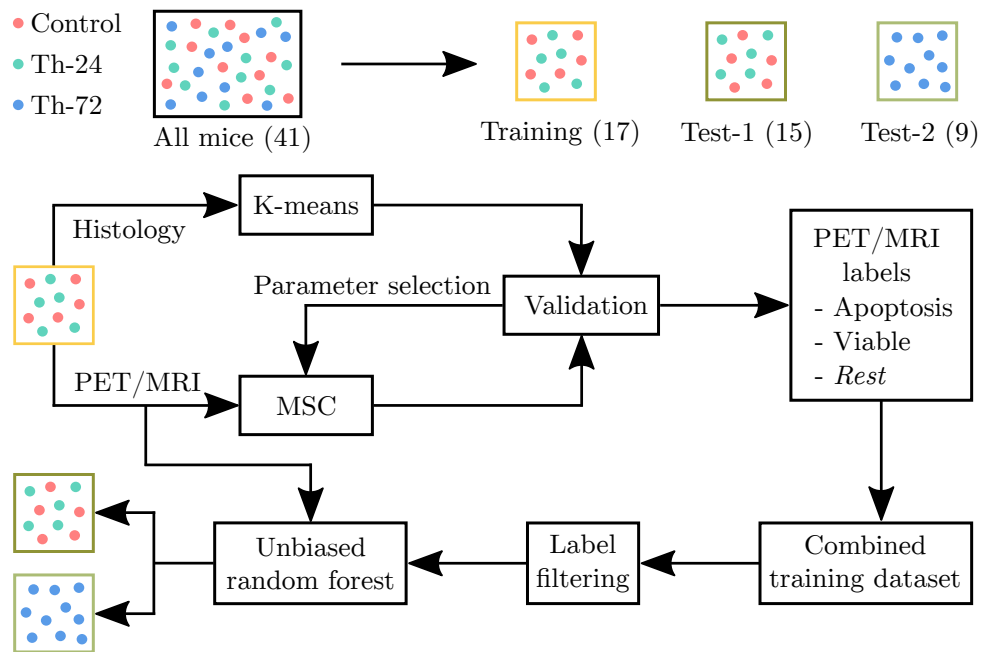


Figure 2.13: The suggested multi-view learning workflow applied to simultaneously acquired dynamic ^{18}F -FDG PET/multiparametric MRI tumor data.

set tumor, respectively. The former cluster comprised of fibrous, connective and muscle tissues, whereas the latter cluster consisted of necrotic and fibrous regions. The spectral clustering parameters (the RBF kernel scale and the number of eigenvectors) were selected by segmenting the co-registered histology slices of the respective tumors using K -means and performing a grid search to minimize the DSC between the histology and spectral clustering phenotypic maps. The matched H&E and caspase-3 histology images were segmented for the Control and Th-24 Training set tumors, respectively.

In the second phase (figure 2.14, yellow area), the dynamic ^{18}F -FDG PET data of the Training set tumors was segmented using spectral clustering. However, prior to clustering, the PET frames were weighted using the Fisher scoring algorithm [100] (subsection 1.6.5). The Fisher weights were calculated using the voxel-level histologically validated labels obtained from the clustering of the multiparametric MRI data in the previous step. The clustering on the weighted PET data was performed identically to the previous phase (not fully depicted in figure 2.14).

In the third phase, a convex combination of the multiparametric MRI (W_{MRI}) and dynamic PET (W_{PET}) affinity matrices (calculated in phase 1 and 2) was used to

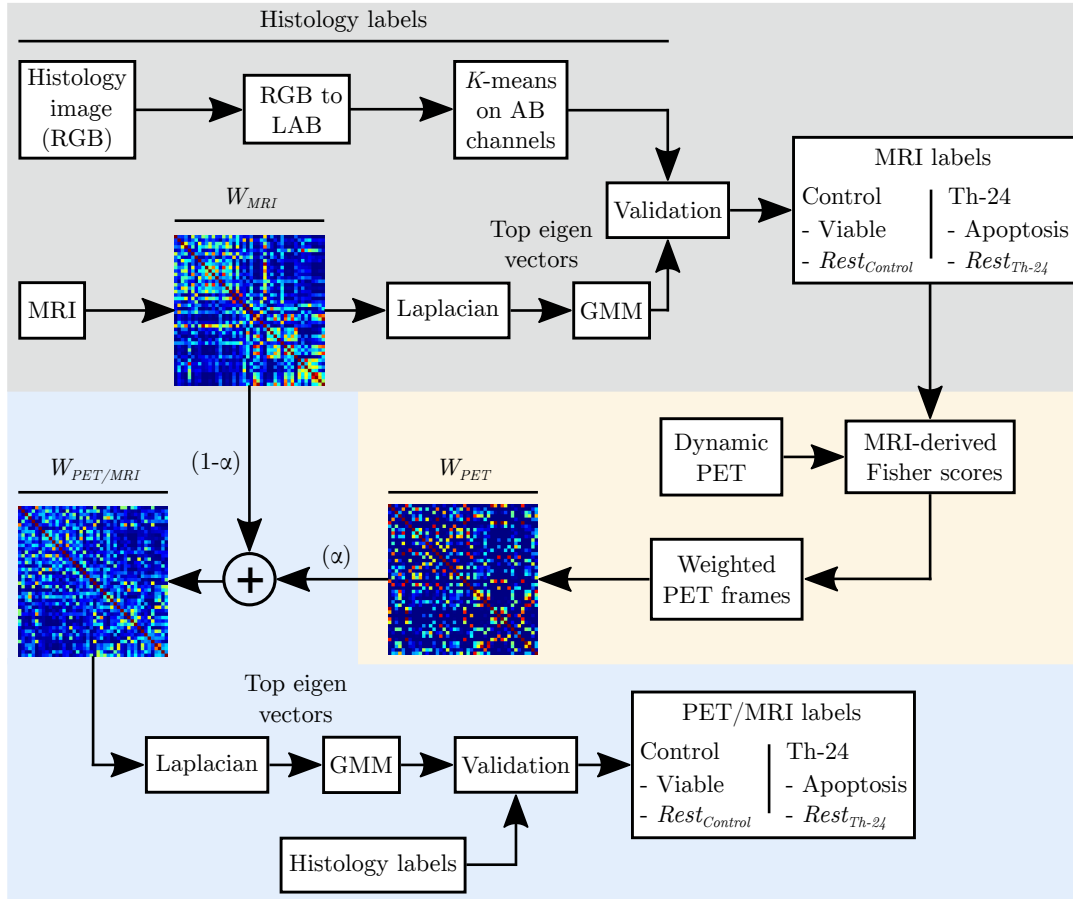


Figure 2.14: The MSC algorithm. The gray, yellow and blue regions depict the segmentation of multiparametric MRI, Fisher score weighted dynamic ^{18}F -FDG PET and multiparametric PET/MRI data using spectral clustering. The MRI, PET and PET/MRI affinity matrices are denoted as W_{MRI} , W_{PET} and $W_{PET/MRI}$, respectively and the variable α is introduced to optimize the influence of the two modalities on the final PET/MRI tissue labels.

obtain the PET/MRI affinity matrix ($W_{PET/MRI}$):

$$W_{PET/MRI} = \alpha * W_{PET} + (1 - \alpha) * W_{MRI}.$$

Here, α ($0 \leq \alpha \leq 1$) controls the contribution from each modality. The final voxel-wise tissue labels were obtained by calculating the normalized graph Laplacian and clustering its top eigenvectors using a GMM. The optimal value of α and the number of eigenvectors were selected by validating the PET/MRI phenotypic maps with the segmented histology.

In all phases above, the normalized graph Laplacian L_{sym} was calculated, as detailed in the subsection 1.6.3.

The MSC example of a Control and a Th-24 Training set tumor is shown in figure

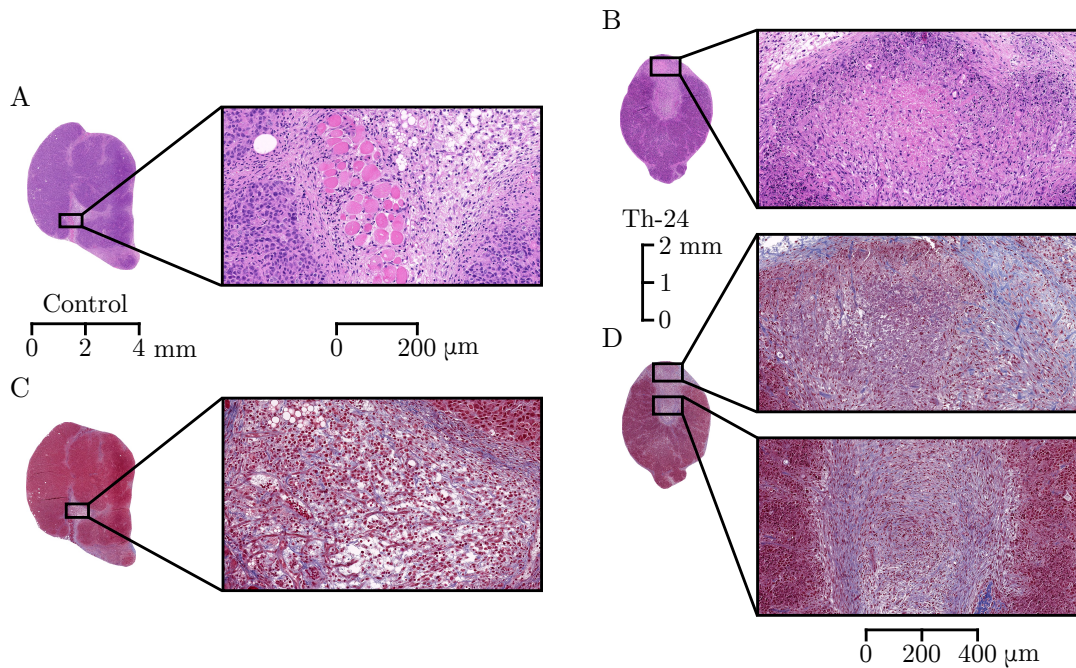


Figure 2.15: The $rest_{Control}$ and $rest_{Th-24}$ areas in histology. (A) and (B) H&E histology of a Control and Th-24 Training set tumor. The insets depict muscle and necrotic tissues, respectively. (C) and (D) Masson's trichrome staining of the tumors in A and B. The trichrome inset in C shows loose fibrous tissue. Similarly, the insets in D depict regions with necrotic and loose fibrous tissues.

2.16. Overall, for the intratumor partitioning of the Control or Th-24 Training set tumors multiparametric MRI data was more predictive than dynamic ^{18}F -FDG PET images.

Due to broad tissue heterogeneity, the $rest_{Control}$ cluster of all Control Training set tumors was excluded from the supervised learning step. Also, for simplicity the $rest_{Th-24}$ cluster is referred to as $rest$ (figure 2.13).

After the label filtering step, the redundancy in the combined training data was reduced by applying a linear fit on the last 7 PET time points of each voxel and using the fit-intercept as their substitute. Hence, the final training data consisted of 19 input predictors (PET frames 2-15, intercept, ADC maps, pre-contrast T2 and T2* maps and post-contrast T2* map) and one categorical output variable (apoptosis/viable/ $rest$). The PET/MRI classifier was trained on all input features, whereas as the PET and MRI classifiers were trained by using the features only from the respective modality. All trained models were applied on the Test-1 and Test-2 set tumors to predict the probability maps depicting intratumor heterogene-

ity.

Figures 2.17A-C show the random forest proximity matrices for all trained classifiers. Unlike the intratumor segmentation step, the prognostic relevance of the dynamic ^{18}F -FDG PET was evident from the combined training dataset. As depicted by the first two diagonal matrices of the PET proximity matrix, the trained PET classifier was able to distinguish the viable and apoptotic voxels well. Conversely, the third block diagonal matrix of the MRI proximity matrix shows that multiparametric MRI features were more predictive in identifying the voxels labeled as *rest*. The combined information from both the modalities, therefore, enabled the PET/MRI classifier to identify voxels from each one of the clusters. The relevant features for the PET/MRI classifier are shown in figure 2.17D. The six most predictive features included ADC, pre-contrast T2 and post-contrast T2* maps as well as early and late PET time points, indicating the importance of diffusion, perfusion and late tracer uptake information in the assessment of intratumor heterogeneity.

The test results of the trained classifiers are shown in figure 2.18. The Control

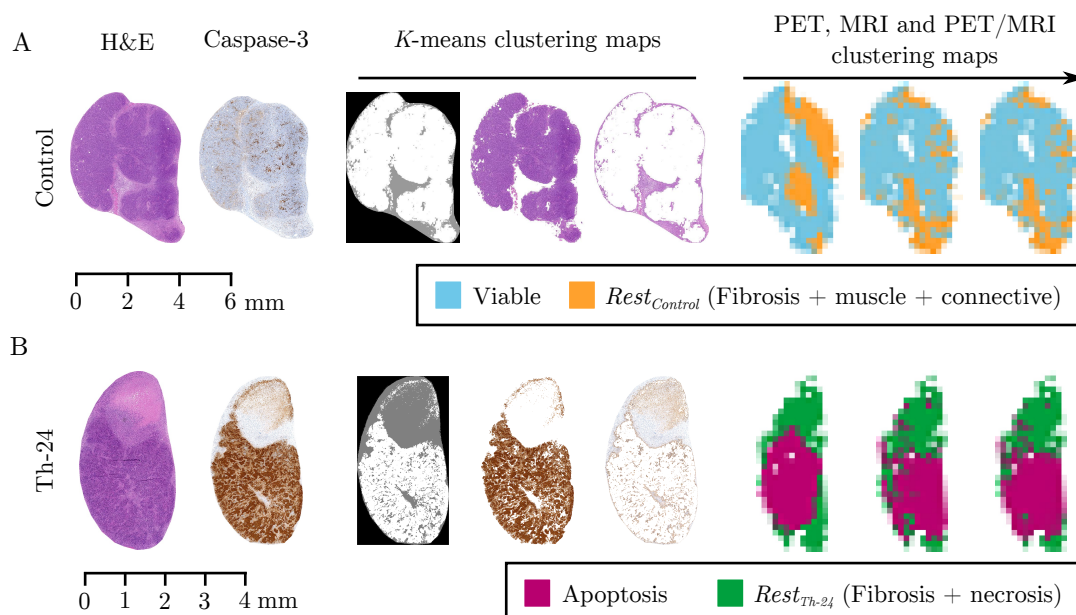


Figure 2.16: MSC segmentation of a Control and Th-24 Training set tumor. (A) Left to right: H&E and caspase-3 histology, K -means clustering indexed image, ground truth viable and $rest_{Control}$ clusters, PET, MRI and PET/MRI phenotypic maps of a Control Training set tumor. (B) H&E and caspase-3 histology, K -means clustering indexed image, ground truth apoptosis and $rest_{Th-24}$ clusters, PET, MRI and PET/MRI phenotypic maps of a Th-24 Training set tumor.

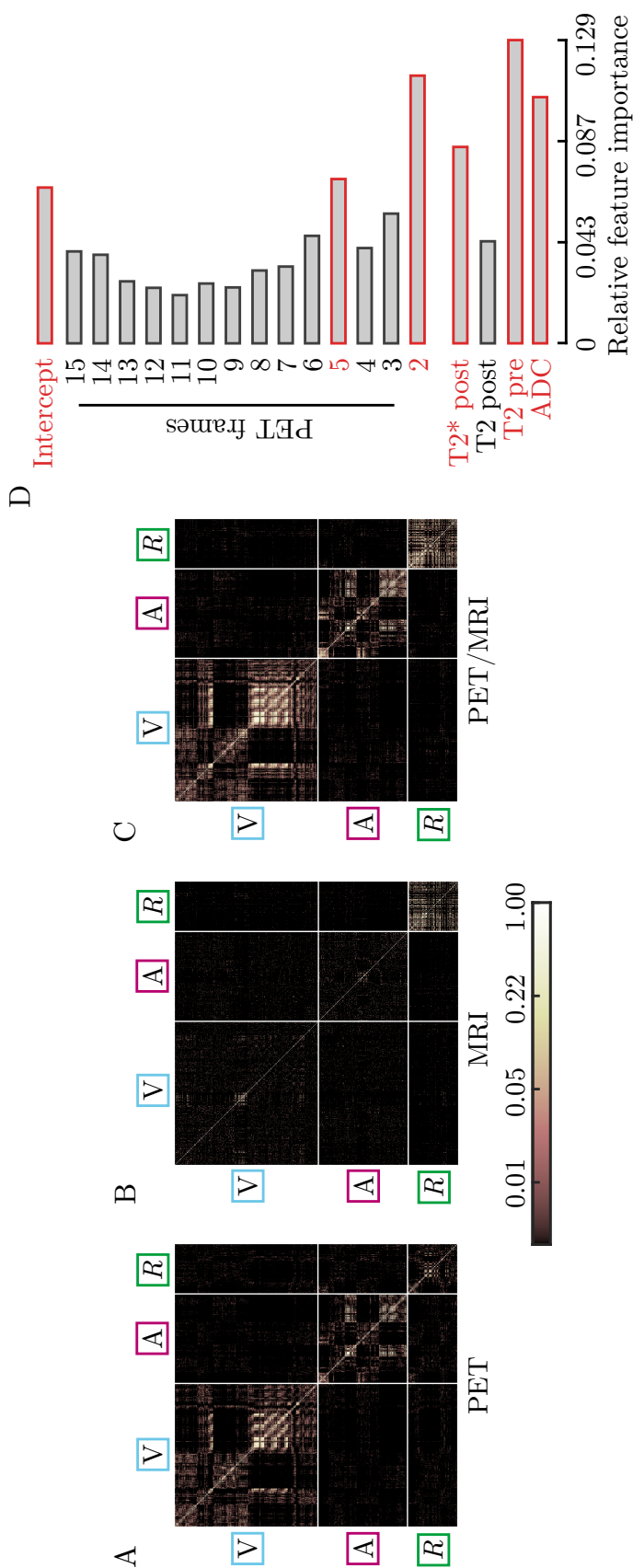


Figure 2.17: Proximity matrices and feature relevance scores. (A)-(C) PET, MRI and PET/MRI proximity matrices. The letters V, A and R denote the viable, apoptosis and rest clusters, respectively. The isolated clusters within the block diagonal matrices highlight the heterogeneity between tumors in the Training Control and the Training Th-24 groups. (D) PET/MRI classifier feature relevance scores for all input predictors. The six most relevant features are highlighted in red.

and Th-24 tumors (first four) were from the Test-1 set and the Th-72 tumors (last two) were from the Test-2 set. All three classifiers were applied on the test tumors in a voxel-wise manner to obtain the probabilistic intratumor phenotypic maps. In both the Control tumors, the PET classifier accurately distinguished between the viable and apoptotic regions, whereas the MRI classifier identified the fibrous and necrotic tissues. In the Th-24 tumors, the MRI classifier additionally recognized the apoptotic tissue, while the PET classifier was more specific for the viable areas. Lastly, in the first Th-72 tumor, the viable and apoptotic tissues were correctly identified by the PET classifier and the same was true in the second Th-72 tumor for the MRI classifier. In both the tumors, the fibrous and necrotic regions were differentiated by the MRI classifier. Due to such complementary characteristics of PET and MRI, for every test tumor, the PET/MRI classifier resulted in an accurate intratumor description of all the phenotypes.

The next step in this study was to assess the transferability of the multi-view learning concepts to a different tumor model. Therefore, we trained two additional MRI-only classifiers that were applied to the multiparametric MRI data of the subcutaneous glioblastoma tumors [103] (subsection 2.3.1). In order to account for measurement bias of the two separate studies, before training, the features were normalized to zero mean and unit standard deviation. Thereafter, only voxels from the viable and *rest* clusters were included in the training of the first model (2-class MRI classifier), whereas the entire dataset was used to train the second model (3-class MRI classifier).

Figure 2.19 shows the intratumor tissue probability maps obtained after applying the two classifiers on the multiparametric MRI data of the glioblastoma xenografts. Both the models were accurate in identifying the necrotic areas. Whereas the two-class classifier precisely characterized the viable regions, the 3-class model misclassified some of these regions as apoptosis.

To summarize, this work provides an important and novel contribution in the field of precision oncology and multimodality imaging. We have shown that the suggested multi-view learning approach selects the most predictive features from a

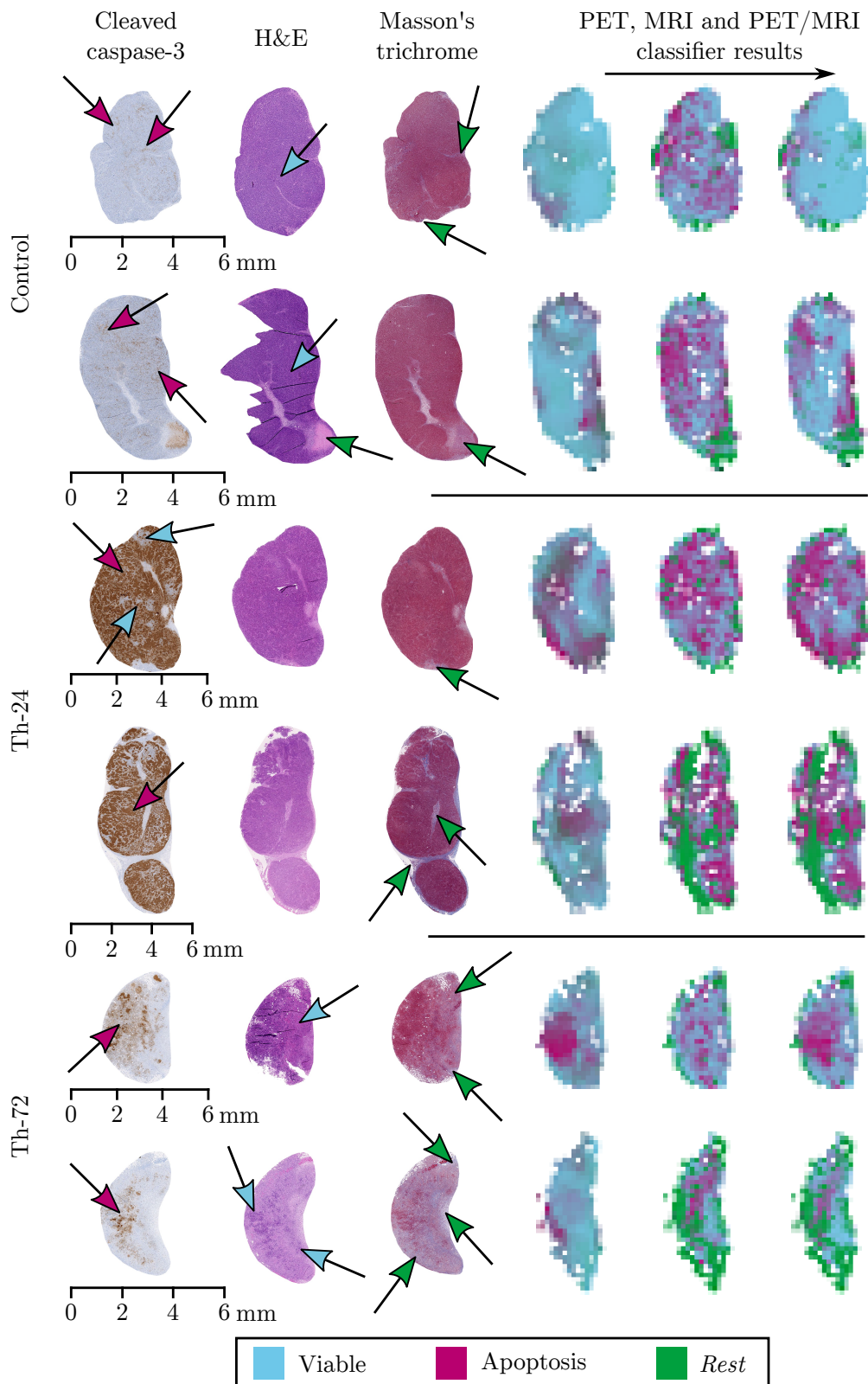


Figure 2.18: PET, MRI and PET/MRI classifier results. Column-wise, left to right: cleaved caspase-3, H&E and Masson's trichrome stainings of two exemplary test tumors from the Control, Th-24 and Th-72 groups each. The respective tissue probability maps obtained from the PET, MRI and PET/MRI models are shown in the last three columns. The arrows are color-coded according to the figure legend.

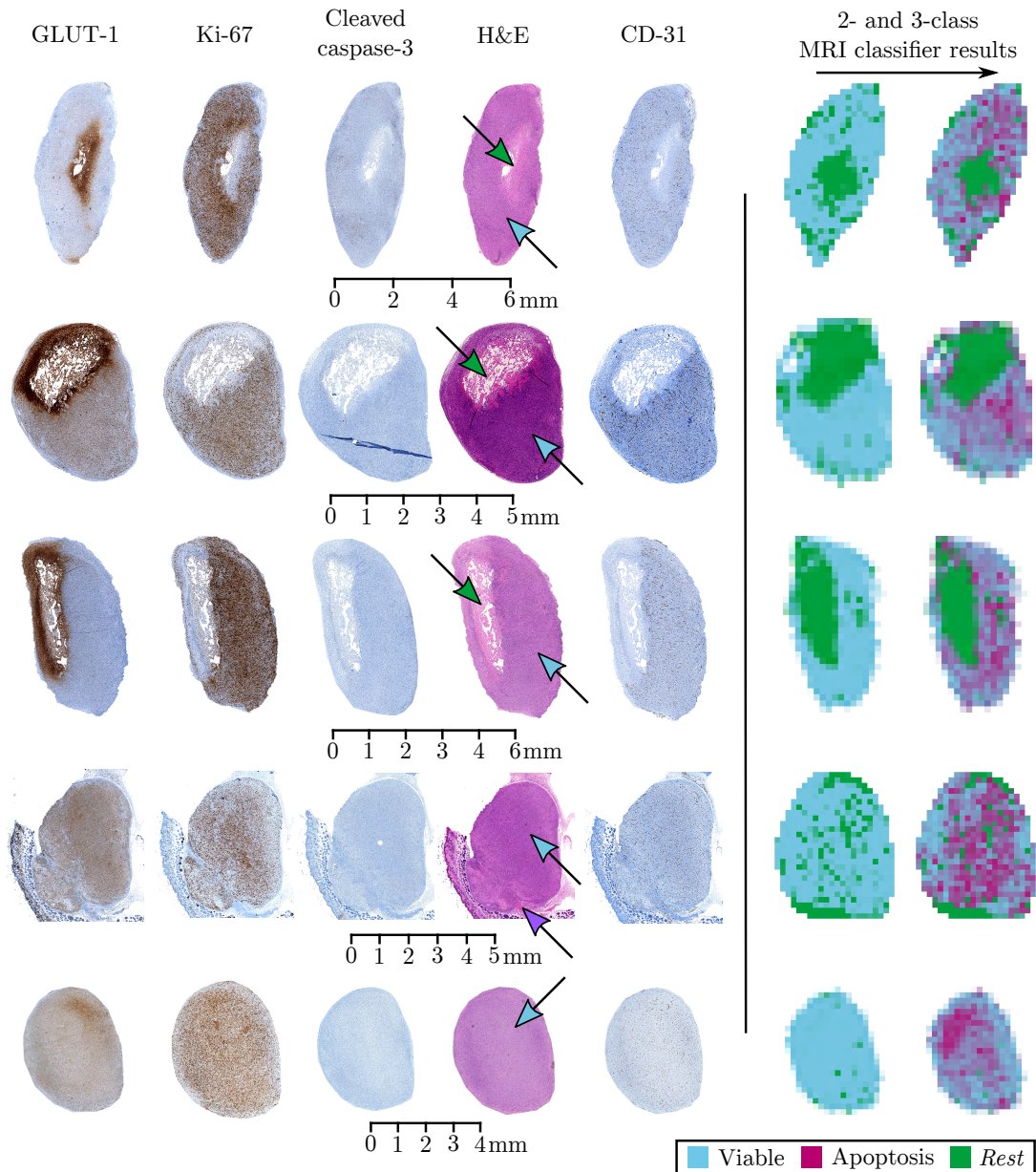


Figure 2.19: MRI classifiers applied to the glioblastoma tumors. Column-wise, left to right: GLUT-1, Ki-67, caspase-3, H&E and CD-31 histology of five glioblastoma tumors. The corresponding tissue probability maps obtained after applying the 2- and 3-class MRI classifiers. The arrows are color-coded according to the figure legend. The muscle tissue in the second last tumor is additionally indicated by the purple arrow.

plethora of imaging parameters and provides a reliable voxel-wise characterization of intratumor heterogeneity. As a proof of principle, we also showed that the classifier trained on one tumor model can be applied to another tumor model as well. Despite the fact that both tumor types were subcutaneously inoculated, approximately half of the Training set mice bearing colon cancer were undergoing an EGFR-specific apoptosis inducing therapy, and therefore were subjected to very different experimental conditions than the glioblastoma tumors. The ability of the MRI classifier to detect the viable and necrotic tissues in the latter tumor model signifies the specificity of the imaging biomarkers for the identified phenotypes. Nonetheless, the transferability of the trained classifiers on orthotopic tumor models and clinical datasets is yet to be investigated. One clear limitation of this study is the use of a supervised learning algorithm that can only consider hard tissue labels as the output variable. As MSC provided probabilistic tissue probability maps for the Training set tumors, a model utilizing the soft tissue labels [150] could have better characterized the intratumor phenotypes.

Overall, the translational prospects of this report are quite high because the PET and MRI experiments performed in this work are standard protocols in the clinic. In particular, prostate cancer investigations [151] can certainly benefit from this approach, as they occasionally include a complete surgical resection of the primary tumor [152].

Multi-view learning of multiparametric PET/MRI data quantifies intratumor heterogeneity

Prateek Katiyar^{1,2}, Mathew R. Divine¹, Ursula Kohlhofer³, Leticia Quintanilla-Martinez³,
Martin Siegemund⁴, Roland Kontermann⁴, Bernhard Schölkopf², Bernd J. Pichler¹,
Jonathan A. Disselhorst¹

1. Werner Siemens Imaging Center, Department of Preclinical Imaging and Radiopharmacy, Eberhard Karls University Tuebingen, Tuebingen 72076, Germany
2. Max Planck Institute for Intelligent Systems, Tuebingen 72076, Germany
3. Institute of Pathology and Neuropathology, Eberhard Karls University Tuebingen and Comprehensive Cancer Center, University Hospital Tuebingen, Tuebingen 72076, Germany
4. Institute of Cell Biology and Immunology, University of Stuttgart, Stuttgart 70569, Germany

Corresponding author:

Jonathan A. Disselhorst
Eberhard Karls University Tuebingen
Department for Preclinical Imaging and Radiopharmacy
Werner Siemens Imaging Center
Roentgenweg 13 D-72076, Tuebingen, Germany
Telephone: +49 7071 29 87699
Fax: +49 7071 29 4451
e-mail: Jonathan.Disselhorst@med.uni-tuebingen.de

Classification: Biological Sciences, Medical Sciences

Keywords: Intratumor heterogeneity, PET/MRI, Multi-view learning, Dynamic ¹⁸F-FDG PET, Multiparametric MRI

Abstract

Advances in immunotherapy and targeted therapy have made cancer treatment increasingly precise. However, solid tumors are well known to exhibit inter- and intratumor heterogeneity, a characteristic that is closely linked to therapy resistance and efficacy. Therefore, a reliable characterization of intratumor phenotypes is highly desirable in precision oncology. Cancer biopsies often represent only a tiny snapshot of the tumor profile, lacking the ability to comprehensively reflect spatio-temporal phenotypic changes. Recent multimodal imaging techniques could provide further valuable insights if the complementary imaging information is sufficiently analyzed. Thus, we developed a novel machine learning framework that utilizes metabolic and functional tumor profiles, captured by dynamic (~55 minutes) positron emission tomography (PET) and multiparametric magnetic resonance imaging (MRI), to provide a precise spatial characterization of intratumor heterogeneity. By implementing a unique unsupervised-supervised learning setup, which has never been applied to multimodality tumor imaging data, we investigate the feasibility of image-derived tumor tissue labels in training a phenotype specific classifier. Applying the proposed multi-view learning framework to two different subcutaneous tumor models, we show that an accurate assessment of important therapy-induced phenotypes, specifically apoptotic, fibrous and necrotic tissues was only possible using the complementary information from simultaneous PET/MRI. Moreover, feature importance scores of the PET/MRI classifier revealed that all multiparametric MRI features and both early and late PET time points were relevant in distinguishing the tissue types. Lastly, histological validation confirmed that the predicted phenotypic maps provide a definite localization of molecularly distinct regions, allowing an accurate *in vivo* evaluation of therapy efficacy.

Significance statement

Tissue heterogeneity within tumors can have a profound influence on therapeutic success and patient survival. This study presents the prognostic benefits of simultaneously acquired dynamic positron emission tomography (PET) and multiparametric magnetic resonance imaging (MRI) data in determining intratumor heterogeneity. We devised a novel machine learning approach on multiparametric oncology data to identify the most predictive PET/MRI features and obtain diagnostically relevant intratumor maps of phenotypic habitats. Our findings revealed that, as opposed to static or single modality imaging, multiparametric PET/MRI enables an early characterization of therapy-induced cellular changes in the tumor. These results demonstrate the potential of PET/MRI and machine learning in monitoring cancer patients undergoing immunotherapy or targeted therapy as well as for dose-painting in radiation treatments.

Introduction

Medical imaging plays a cardinal role in the diagnosis, treatment planning and response evaluation of patients with cancer. As opposed to omics technologies (genomics, proteomics, metabolomics and transcriptomics), imaging is non-invasive and can consequently be repeatedly applied, which may allow a complete spatial and temporal characterization of inter- and intratumor heterogeneity. Although the current clinical evaluation of solid tumors using imaging relies primarily on shape- and volume-based descriptors (1), efforts to incorporate quantitative imaging biomarkers into routine clinical practice are ongoing (2).

After computed tomography, the two most widely used imaging modalities in oncology are positron emission tomography (PET) and magnetic resonance imaging (MRI). Due to unique imaging principles, the combination of PET and MRI has been shown to provide a wealth of complementary and diagnostically relevant information (3, 4). However, despite its benefits, the ability of the PET/MRI system to offer a “key clinical application” (5) and the superiority of combining the two modalities to a single modality have been continuously debated (6, 7). Nonetheless, there is a growing interest in using machine learning for understanding the highly intricate multimodality multiparametric imaging data (6, 8).

The quantification of intratumor heterogeneity is a key issue in precision oncology. Although the PET/MRI system enables the measurement of a multitude of functional parameters, each corresponding to a distinct physiological view of the tumor, state-of-the-art PET/MRI studies investigating intratumor heterogeneity have been restricted to only two imaging features: the standardized uptake value (from static PET) and diffusion-weighted MRI (7, 9, 10). Since both of these parameters have been shown to

exhibit significant intratumor correlations (9, 10), these studies are limited in their ability to probe the multifaceted intratumor microenvironment. Furthermore, in their translational study, Schmitz et al. (7) independently modeled each parameter as a one-dimensional mixture of Gaussians, and consequently discarded the predictive value of the multidimensional PET/MRI dataset.

Although the utility of each aforementioned imaging modality has been demonstrated individually (11–13), the full potential of dynamic 2-deoxy-2-(¹⁸F)fluoro-D-glucose (¹⁸F-FDG) PET combined with multiparametric MRI in characterizing intratumor heterogeneity has never been exploited. Moreover, due to difficulties in obtaining accurate tumor tissue labels, previously published studies have not attempted to use supervised learning (14) and train a phenotype-specific classifier. Therefore, we aimed to address the following questions in this work, which together justify the important role of PET/MRI and machine learning in precision oncology: Is combined dynamic ¹⁸F-FDG PET and multiparametric MRI more valuable in decoding intratumor heterogeneity than single modality imaging? Can machine learning aid in extracting complementary information from multimodal multiparametric datasets? Is training a classifier that identifies various intratumor phenotypes feasible using *in vivo* imaging? Are these learning concepts transferrable from one tumor model to another?

We examined these questions by developing a novel multi-view learning framework, which duly combines complementary information from dynamic ¹⁸F-FDG PET/multiparametric MRI data to predict early changes in intratumor heterogeneity and monitor therapy efficacy. Specifically, we devised an unsupervised-supervised learning setup by extending our previously established tumor clustering algorithm (12, 13) to a multimodal setting and used histologically validated phenotypic maps of the multi-view clustering algorithm to train machine learning classifiers. Thereafter, we applied the

trained models to test datasets consisting of subcutaneous colon and glioblastoma tumors and corroborated the predicted phenotypic maps using the respective tumor histology. Our findings revealed that the PET/MRI data paired with multi-view learning provides a tool of significant prognostic value that can distinguish between multiple tumor tissue types and deliver reliable estimates of intratumor heterogeneity.

Materials and methods

A brief description of the materials and methods is given below; a more thorough description can be found in the supplemental materials and methods section online.

Animal model

The animal experiments performed in this study were approved by local authorities. NMRI nu/nu mice (n = 46, Charles River, Sulzfeld, Germany) were subcutaneously injected with 4.5×10^6 COLO-205 tumor cells (ATCC, Manassas, Virginia, USA) into the right flank. The tumors were allowed to grow for at least 2 weeks before being studied in one of four groups, control (C-24 = 11 and C-72 = 12) and therapy (Th-24 = 12 and Th-72 = 11). The therapy groups were intravenously (i.v.) injected with 100 μ g of Db-scTRAIL (15) approximately 3 h after an intraperitoneal injection of 5 μ g of Bortezomib (Velcade; Takeda Oncology), whereas the control groups received equal volumes of vehicle. The C-72 and Th-72 groups were administered a second dose 48 h later. The imaging experiments were performed 24 h after the first injection for the C-24 and Th-24 groups and 72 h after the first injection for the C-72 and Th-72 groups.

Imaging experiments

The dynamic ^{18}F -FDG-PET/MRI data were acquired using a PET insert inside a 7 T ClinScan MRI scanner (both Bruker BioSpin GmbH, Ettlingen, Germany). The animals were injected i.v. with 10.9 ± 1.1 MBq of ^{18}F -FDG during the 55-min long PET scan. T2-weighted (T2W) anatomical images, apparent diffusion coefficient (ADC), and T2 and T2* maps were acquired simultaneously with MRI. Contrast-enhanced T2 and T2* maps were obtained two minutes after the i.v. injection of 50 μ L of ferumoxytol (Rienso; Takeda Pharmaceuticals, Glattpark-Opfikon, Switzerland). The dynamic PET data were reconstructed into 22 PET frames (1 \times ~30, 8 \times 30, 6 \times 60, 5 \times 300, 2 \times 600 s); the first PET

frame was not used in the data analyses. The resulting dynamic PET and MRI parameter images were aligned and up-sampled to the resolution of the T2W image (0.22 mm³). The anatomical images of the mice were subjected to tumor delineation using Inveon Research Workplace (Siemens Healthcare, Oxford, UK). Two mice (1×C-24 and 1×Th-72) were excluded from further analyses due to experimental disruptions, and another three mice (1×C-72, 1×TH-24 and 1×Th-72) were discarded due to inadequate imaging-histology alignment. Moreover, all pre-contrast T2* images were excluded from the analyses due to image artifacts. Fig. 1A shows the measured parameters of one of the tumors. Due to their similar characteristics, the C-24 and C-72 tumors were combined into one Control (n = 21) group.

Histology

After each scan, the mice were sacrificed by cervical dislocation, and three equidistant lines (2~3 mm apart and parallel to the axial plane) were drawn on the skin while keeping the animals on the imaging bed. Subsequently, the tumors were superficially frozen using a freezing spray and dissected along the parallel lines, resulting in four sections (labeled as 1 – 4), as depicted in Fig. 2A. Each section was fixed in formalin, embedded in paraffin, sectioned into 2~3 µm thick slices, and stained with hematoxylin & eosin (H&E) and Masson's trichrome. Immunohistochemistry was performed on an automated immunostainer (Ventana Medical Systems, Inc.) according to the company's protocols for open procedures with slight modifications. The slides were stained with the following antibodies: cleaved caspase-3 (ASP 175; Cell Signaling Technology, Frankfurt am Main, Germany), CD-31 (Abcam plc, 330 Cambridge Science Park, Cambridge, UK), TK-1 (Abcam plc, 330 Cambridge Science Park, Cambridge, UK) and F4-80 (SP115, Acris Antibodies GmbH, Herdford, Germany). Appropriate positive and negative controls were used to confirm the adequacy of staining. The stained slides were

digitized using a NanoZoomer 2.0 HT (Hamamatsu, Hamamatsu City, Japan). T2W images of the tumors were manually co-registered to the corresponding histology in MATLAB 2015b (The MathWorks, Natick, MA, USA). The co-registration reliability was evaluated by computing the modified Hausdorff distance (MHD) (16) between the contours of the matched T2W and histology images. Moreover, the Dice similarity coefficient (DSC) (17) was calculated to measure the extent of spatial overlap among the co-registered images.

Glioblastoma tumors (12) were also stained for cleaved caspase-3.

Multi-view learning

A schematic of the utilized multi-view learning framework is presented in Fig. 2B. All tumors from the Control, Th-24 and Th-72 groups were ordered into three sets, namely Training, Test-1 and Test-2. Whereas the Training and Test-1 sets consisted of tumors from the Control and Th-24 groups, the Th-72 tumors were exclusively retained in the Test-2 set. The intratumor partitioning using multi-view spectral clustering (MSC) involved MRI, MRI-derived Fisher score-weighted PET, and PET/MRI segmentations (Fig. S1). Although MSC serves as a vital functional block of the developed multi-view learning workflow and is repeatedly mentioned in this paper, the complete details of these methods are provided in the supplementary data because they are beyond the scope of the main text.

Voxel-level image-derived tissue labels were obtained by independently segmenting the PET/MRI data of each tumor in the Training set using MSC. Both the Control and Th-24 tumors were divided into two tissue compartments. The MSC probability maps of the Control tumors were labeled as viable and $rest_{control}$, and Th-24 tumors were labeled as apoptosis and $rest_{Th-24}$. Likewise, the co-registered histology images of the Control (H&E) and Th-24 (cleaved caspase-3) Training set tumors were segmented into viable and

rest_{control} or apoptosis and *rest_{Th-24}* regions, respectively using *K*-means clustering. The ground truth histology clustering maps were utilized to improve the MSC results by maximizing the *DSC* between itself and the MSC phenotypic maps. A simple grid search was used to select the MSC parameters that yielded the highest *DSC* between the MSC and histology segmented maps. Due to extensive tissue heterogeneity in the *rest_{control}* cluster, only the viable labeled data from the Training Control tumors were utilized for the subsequent analyses. For ease, we hereafter refer to the *rest_{Th-24}* cluster as *rest*.

The voxels from the MSC labeled (as viable, apoptosis and *rest*) maps and the corresponding PET/MRI data of all the Training set tumors were pooled together into one combined training dataset. To refine the clusters in this dataset, voxels with low confidence labels (MSC probability score < 0.9) were removed, and the remaining voxels were assigned the corresponding tissue label (viable/apoptosis/*rest*) with a probability of 1.

For each voxel in the refined dataset, a linear fit was performed on the last 7 PET frames, and the fit-intercept was obtained as a surrogate feature. Finally, each voxel with 19 PET/MRI features (ADC, T2 pre, T2 post, T2* post maps, PET frames 2-15 and Intercept) and a tissue label was used to train an unbiased random forest classifier. Similarly, PET- and MRI-only classifiers were built by training the classifier on the aforementioned imaging features from each of the modalities alone. Post-training, random forest proximity matrices and feature permutation scores were obtained from all classifiers. In addition, relative feature importance scores were calculated.

The trained classifiers were applied to the tumors from both test sets in a voxel-wise manner, resulting in intratumor probability maps of tissue heterogeneity. The classifier results for the Control and Th-24 tumors from the Test-1 set were validated by calculating the *DSC* between the predicated probability maps and ground truth

phenotypic maps produced by clustering either the H&E or cleaved caspase-3 histology of the respective tumors using *K*-means (as in the training phase). The results for the 72 tumors from the Test-2 set were subjectively validated by visually corroborating the predicated probability maps with the H&E, cleaved caspase-3 and Masson's trichrome stains of the tumors.

To test the proposed framework on a different tumor model, two additional MRI classifiers were trained (as mentioned previously) on the combined MRI training data (this time normalized for zero mean and unit standard deviation) of the colon cancer xenografts. The first classifier (2-class MRI classifier) was trained only on the voxels labeled as viable and *rest*, whereas the second classifier (3-class MRI classifier) was trained on the entire dataset. Both classifiers were applied to the five glioblastoma tumors (12) in a voxel-wise manner (excluding T2* pre maps and normalized as mentioned above) to produce phenotypic maps that demonstrated the intratumor heterogeneity. These phenotypic maps were visually validated using the respective tumor histology (12).

MSC was implemented in MATLAB 2015b, and an R implementation of unbiased random forest (*cforest*, *party* package (18)) was used for supervised learning.

Statistical tests

The one-sample Kolmogorov-Smirnov test was used as a test of normality. Due to non-Gaussian distributions, the Kruskal-Wallis non-parametric test was initially applied to investigate possible differences among groups. Post-test individual comparisons were performed with the Bonferroni corrected rank sum tests, and a p-value less than 0.0167 (target p-value of 0.05 adjusted for 3 comparisons) was considered significant.

Results

Intratumor heterogeneity in the Control, Th-24 and Th-72 tumors

The tumor volumes for all three groups are shown in Fig. 1B. Although the Th-72 tumors showed significant ($p\text{-value} < 0.0167$, Table S1) volume differences from those of the Control and Th-24 tumors, no size differences were present among the latter two groups. However, the intratumor characteristics radically changed 24 h after therapy, and extensive apoptosis was visible in cleaved caspase-3 stainings (Fig. 1C), indicating the anti-tumor effects of the therapy. At this time point, macrophage infiltration was also observed at the tumor periphery (Fig. S2A, right). In all Th-24 tumors, loose fibrous tissue was present in the non-apoptotic regions (Fig. S2B) but was unrelated to the therapy because similar tissue was observed in Control tumors. Furthermore, in several ($n = 8$) Th-24 tumors, the non-apoptotic regions contained mixed areas of necrotic and fibrous tissue (*rest cluster*), as depicted in Fig. S3. In the Th-72 group, the apoptotic fraction was reduced (Fig. 1C) and the *rest cluster* portions were significantly enlarged (Fig. S4A). Unlike the tumors in both therapy groups, the Control tumors were very homogeneous and mostly consisted of viable tissue. Moreover, muscle, connective, and fibrous tissues were focally observed in some Control tumors (Fig. S5). Due to broad tissue heterogeneity, these regions were excluded from the voxel-wise analysis. The MRI and PET correlation matrices for the combined data of all 44 tumors are shown in Fig. 1D and 1E, respectively. Among all MRI parameters, T2 pre- and post-maps were moderately correlated, whereas a strong correlation was present in the last 7 frames of the dynamic ^{18}F -FDG PET data.

Co-registration of *in vivo* data with tumor histology

Fig. 2C presents two good and two poor imaging-to-histology co-registration examples. The angles of in-plane, vertical and horizontal axis rotations applied to match 80 (~2 slices for each tumor) imaging slices (along the axial plane) with the corresponding histology are detailed in Fig. 2D. The average absolute angles for in-plane and vertical axis rotations were 4.4 ± 4.9 and 1.5 ± 3.5 degrees, respectively. The resulting *MHD* and *DSC* values between all co-registered image-histology pairs are shown in Fig. 2E and 2F, respectively. The reasonably low average *MHD* (0.89 ± 0.45 mm) and high average *DSC* (0.91 ± 0.04) values highlight the excellent co-registration achieved by the careful fixation and sectioning of tumors.

MSC on multiparametric PET/MRI data provides histologically validated intratumor tissue labels

Fig. 3A illustrates examples of clustering in a Control and Th-24 tumor from the Training set. A simple color-based *K*-means clustering of the histology was adequate to distinguish the viable or apoptotic tissue from the remaining tumor. Among the *in vivo* modalities, intratumor partitioning using multiparametric MRI was more accurate in discriminating fibrous, necrotic and muscle tissues from viable or apoptotic cluster than the clustering of ^{18}F -FDG PET time activity curves. Nonetheless, tissue-labeled maps obtained by applying MSC to the PET/MRI (combined multiparametric MRI and MRI-derived Fisher score-weighted PET) data agreed best with the segmented histology. The cluster-wise pooled imaging parameters (from all Training set tumors) used to train the PET, MRI and PET/MRI classifiers are shown in Fig. 3B.

Classifier trained using multiparametric PET/MRI data and MSC tissue labels characterizes intratumor heterogeneity

Fig. 4A-C exhibit the proximity matrices formed using the trained PET, MRI and PET/MRI classifiers. In the pooled training dataset, the *rest* cluster was best distinguished from the viable and apoptosis clusters using multiparametric MRI. Conversely, metabolic information from PET was more significant in differentiating viable and apoptotic samples. The PET/MRI classifier exploited complementary information from each modality and separated all three clusters from each other. The sub-clusters in the viable and apoptosis block diagonal matrices of the PET and PET/MRI proximity matrices are a result of the intertumor heterogeneity within the Control and Th-24 Training set tumors. Fig. 4D shows the relative feature importance scores for the PET/MRI classifier. The most predictive MRI and PET features were ADC, T2 pre, and T2* post and frame 2, 5, and the Intercept, respectively.

The PET, MRI and PET/MRI classifier results for four tumors from the Test-1 set (2×Control and 2×Th-24) and two tumors from the Test-2 set are shown in Fig. 5. The tumor in the first, second and fifth rows are reasonable examples of the specificity provided by the dynamic ¹⁸F-FDG PET data, enabling a precise identification of intratumor viable and apoptotic tissue, which was not feasible using the multiparametric MRI data alone. In contrast, the second, fourth and sixth row tumors show that the vital information captured by multiparametric MRI allows the accurate localization of the *rest* cluster, which was entirely un-identified by the PET classifier. Likewise, the apoptotic regions in both Th-24 tumors (third and fourth rows) were determined by the MRI classifier. For each example, the PET/MRI classifier provided a definite characterization of all three clusters by utilizing the complementary information from both modalities.

PET/MRI classifier-predicted fractional tissue volumes are indicative of therapy efficacy

Fig. 6A details the average Training and Test-1 set *DSCs* of the Control and Th-24 tumors. MSC on the PET/MRI data resulted in the highest *DSC* for both groups in the Training set. The average *DSC* of the test Control tumors for the PET classifier was marginally higher than that of the PET/MRI classifier, and the opposite was true for the test Th-24 tumors. Fig. 6B shows the fractional volumes (only for the co-registered slices) of all three clusters as predicted by the PET/MRI classifier for the test set tumors. The viable fraction of the test Control tumors significantly differed from the viable fractions of the test Th-24 and Th-72 tumors. Similarly, the apoptotic fraction of the test Th-24 tumors significantly differed from that of the test Control and Th-72 tumors. However, the *rest* fraction of the test Th-72 tumors only differed from the *rest* fraction of the test Control tumors. The p-values for the aforementioned analyses are given in Table S2.

Multi-view learning concepts are transferable from one tumor model to another

Fig. 7 depicts the outcomes of applying the 2- and 3-class MRI classifiers to the glioblastoma tumors. The 2-class model correctly identified the viable and necrotic areas. Additionally, the muscle tissue in one of the tumors was classified as *rest*. The 3-class classifier differed from the 2-class model only in the viable portions and recognized false apoptotic areas in each tumor.

Discussion

To the best of our knowledge, this study is the first to use the synergistic value of simultaneously acquired dynamic ^{18}F -FDG PET and multiparametric MRI data to deliver a comprehensive profile of intratumor heterogeneity and, consequently, to non-invasively assess therapy efficacy. We achieved these objectives by developing a novel multi-view learning framework that judiciously combined complementary information from the two modalities and built a model that quantified phenotypic changes induced by the therapy. In this workflow, we extended our previously established single modality tumor clustering method (12, 13) to a multimodal setting and successfully showed that MSC maps of intratumor heterogeneity can be used to train a multiclass classifier. Our results revealed that an accurate intratumor classification of the viable, apoptotic, and fibrous and necrotic tissue was only feasible using combined information from the dynamic PET and multiparametric MRI data. Moreover, all multiparametric MRI features and both early and late PET time points were relevant in distinguishing the tissue types, highlighting the superior value of multiparametric PET/MRI in decoding intratumor heterogeneity. Lastly, in addition to the colon cancer xenografts, we applied the trained models to the subcutaneous glioblastoma tumors and corroborated the classification results using the corresponding histology, which demonstrated the wide applicability of the proposed approach. Our unique unsupervised-supervised learning setup offers vast opportunities in the clinic because it aptly combines the collective information of multiple imaging parameters in one single map of intratumor heterogeneity, which could serve as an accurate indicator of the therapy efficacy.

Despite the widespread use of supervised learning (14), its potential in identifying intratumoral tissues using *in vivo* imaging has not been demonstrated. The most

significant hurdle to developing a phenotype-specific classifier is the availability of reliable tumor tissue labels. Manual annotation of tumor histology is a time-consuming and tedious procedure associated with the subjective bias of the involved pathologist, and clinical studies are further limited due to difficulties in obtaining tumor histology (19). Additionally, transferring labels from histology to imaging requires voxel-wise co-registration, which is highly challenging given the resolution differences between both the sources. We mitigated some of these issues by directly extracting voxel-wise labels from the PET/MRI data of the Control and Th-24 tumors. Therefore, instead of performing a voxel-wise non-rigid imaging to histology co-registration, we rigidly matched the imaging slices with the corresponding histology sections. By ensuring the careful fixation and sectioning of tumors, the ground truth (histology-derived) and MSC phenotypic maps could be reliably matched while optimizing the MSC parameters. This approach provided histology-validated probabilistic labels of intratumor heterogeneity, which were subsequently used to train the unbiased random forest classifiers. The choice of histology segmentation algorithm largely depends on the complexity of pathology images. Our selection of *K-means* clustering was primarily due to the effectiveness and simplicity of the method.

This study goes beyond the earlier efforts of Gatidis et al. (20), who used the labels obtained from the static PET and multiparametric MRI data using a spatially constrained fuzzy c-means algorithm to train a support vector machine classifier for prostate cancer delineation. Their investigations, however, did not attempt to probe intratumor heterogeneity. Our MSC framework selectively merged relevant information from dynamic PET and multiparametric MRI data by considering the dimensionality and scale differences between both modalities. To minimize the influence of histology-to-imaging co-registration uncertainty on the tumor tissue labels, only tumors with a high *DSC* (for

MSC) were included in the Training set. Furthermore, purity within each cluster of the combined training set was maximized by discarding voxels with a low MSC probability score (voxels with mixed phenotypes) in the label-filtering step (Fig. 2B). Although random forest is a powerful machine learning tool, it is vulnerable to differences in the measurement scale and unbalanced number of observations in each category when used for variable selection. Hence, we used unbiased random forest in the supervised learning step (18), which overcomes the shortcomings of standard random forest by training unbiased classification trees, thereby ensuring a fair variable selection and reasonable predictions.

The rationale behind the specific partitioning of mice in the Training, Test-1 and Test-2 sets (Fig. 2B) stems from the characteristics of the tumors in the Control, Th-24 and Th-72 groups. The Control tumors mainly consisted of viable tissue, whereas the Th-24 tumors consisted of apoptotic and the *rest* areas. Moreover, in many Th-24 tumors, viable tissue was present in patches surrounded by large apoptotic regions (Fig. S6). The Th-72 tumors on the other hand, contained sizable populations of all three clusters. Restricting these tumors in the Test-2 set provided an independent validation set, which adequately confirmed the predictive performance of the trained PET/MRI classifier for all three labels.

Despite suitable visual agreements with the histology, the average training and test *DSCs* (Fig. 6A, corresponding to MSC and the trained classifier probability maps, respectively) for the Control and Th-24 tumors were relatively low (<0.8). Several factors account for these results. First, these *DSC* measures were affected by manual tumor delineation and co-registration inaccuracies (Fig. 2F, average *DSC* <1). Second, the indexed (probability maps thresholded to highest probability) MSC or classifier maps were used to calculate the amount of spatial overlap with the histology clustering maps.

As reflected by the histology and probability maps in Fig. 3A and 5, the tissue heterogeneity of the tumors extends beyond the spatial resolution of each voxel, causing the algorithms to characterize each voxel as a probabilistic combination of several phenotypes. In such a scenario, the most accurate assessment of the predicted spatial heterogeneity can only be obtained by calculating a fuzzy *DSC* between the histology and MSC or classifier probability maps. A multi-class probabilistic segmentation of the histology requires a pixel-wise co-registration between different stains (identifying different phenotypes) of the same section, which can be processed by an unsupervised/supervised algorithm to create a ground truth map of intratumor heterogeneity. Furthermore, to reliably quantify the *DSC* for each imaging slice, a fused 3D stack of many of these ground truth maps needs to be created to eliminate the resolution differences between the two modalities (mm vs μm). However, such an investigation is well beyond the scope of this paper. Fig. 6C shows an example of a test Control tumor, which highlights the influence of the aforementioned factors. Although the PET/MRI classifier precisely characterized the viable tissue and local apoptosis in the tumor (shown in the inset), the *DSC* of the PET/MRI phenotypic map was lower than that of the PET phenotypic map. Because the H&E histology was segmented for the Control tumors, apoptotic regions were absent in the ground truth phenotypic maps, which resulted in their better agreement with the PET classifier's output. The inclusion of texture and color information (at higher magnification) from multiple histological stains could have enabled a detailed inspection of the intratumor heterogeneity and provided more accurate *DSC* estimates of the test Control and Th-24 tumors (Fig. 6A). In the supervised learning phase, hard tumor tissue labels were used to train the classifiers; the proposed multi-view learning framework can be further improved by

selecting a classifier that can use the probabilistic MSC tissue labels (21) to learn the decision boundary between various tumor tissue classes.

Fig. 5 provides vital evidence of the sensitivity of the multiparametric MRI data for the apoptosis and *rest* clusters and dynamic ^{18}F -FDG data for viable cluster, except for the two tumors shown in the last two rows. While the PET/MRI classification map of the Th-72 tumor in the fifth row clearly benefitted more from the dynamic PET data, the multiparametric MRI data alone was sufficient to identify all three clusters in the sixth tumor. Overall, the PET/MRI classifier extracted the most relevant information from a host of imaging features by weighting the prognostic influence of different input parameters as well as each modality. The mislabeling of the viable regions as apoptosis by 3-class MRI classifier in Fig. 7 indicates the poor specificity of the multiparametric MRI data for apoptotic tissue. These results were in agreement with the MRI-only phenotypic maps of the colon cancer xenografts (Fig. 5, Control tumors), where the mislabeled regions were rectified in the PET/MRI phenotypic maps using the complementary information from the dynamic PET data.

Our approach differs from texture- and shape-based radiomics analysis (22), which extracts a plethora of synthetic features from radiology images to characterize tumor heterogeneity. Principally, in radiomics by combining the least correlated and most predictive of all features, a quantitative descriptor of the entire tumor is obtained. Although these image descriptors have proven to be indicative of intertumor heterogeneity and patient survival (23), they do not provide the spatial information of biologically distinct regions within each tumor. Moreover, the high dimensionality (>100) of the input data warrants a large number of studies (24) to identify the most robust and predictive features and avoid the curse of dimensionality. Our proposed multi-view learning workflow on the other hand, provides a spatial map of intratumor

heterogeneity by taking the voxel-level information into account. Phenotypic information at this scale has clear potential in clinical decision-making because measuring the fraction of various intratumor tissue types can provide strong clues about the success (25) or failure (26) of a specific therapy. In this case for instance, the fractional tissue volume of viable and apoptotic tissues (Fig. 6B) can be used to stratify the Control and Th-24 tumors, which would not be possible using just the volume (Fig. 1B) of respective tumors. In addition, intratumor phenotypic maps may play a key role in linking imaging with omics by serving as a reference for image-guided tumor tissue biopsies. Similarly, the exact location of these biologically distinct regions might also be of great assistance in image-guided radiation (27) and ultrasound therapies (28).

The proposed framework holds high translational relevance because the imaging parameters used in this study are routinely acquired in the clinic. Most importantly, a trade-off between the scan-duration and model accuracy can be made in the clinical examinations by acquiring only the relevant imaging parameters (Fig. 4D). As shown in Fig. S7, the classifier trained on the most predictive PET/MRI parameters ($\text{PET/MRI}_{\text{pred}}$) yielded phenotypic maps comparable to that of the PET/MRI classifier for the Control and Th-24 tumors, but it failed to fully identify the apoptotic tissue in the Th-72 tumor. Additionally, the probability maps of $\text{PET/MRI}_{\text{pred}}$ did not manifest uncertainty in heterogeneous regions, which indicates its weakness in resolving highly composite areas. In particular, the apoptotic regions in the test PET/MRI and $\text{PET/MRI}_{\text{pred}}$ phenotypic maps exhibited the least agreement (Table S3). Nevertheless, the test viable and apoptosis fractions of $\text{PET/MRI}_{\text{pred}}$ could significantly stratify the test Control and Th-24 tumors. Overall, the non-invasive application of these models is highly beneficial in the clinic because the tumor state is a continually evolving process and demands the longitudinal monitoring of intratumor characteristics (29, 30).

In conclusion, this report provides compelling evidence to substantiate the superior value of dynamic ^{18}F -FDG PET/multiparametric MRI and multi-view learning in the assessment of intratumor heterogeneity. Our results suggest that the complementary information provided by PET time activity curves and multiple MRI parameters goes beyond the limits of static or single modality imaging, enabling a robust voxel-wise characterization of the complex intratumor microenvironment. We also demonstrated the central role of machine learning in analyzing the multimodal multiparametric imaging data by devising a novel multi-view learning approach. Finally, we have shown that the full potential of PET/MRI can be exploited by using *in vivo* imaging and multi-view learning in tandem, which can significantly augment the current standards of selective treatment planning and precision oncology.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/ 2007-2013) under ERC grant agreement n°323196–ImageLink, German Ministry for Education, Research/Bundesministerium für Bildung und Forschung (BMBF), grant number 0316186E and Eberhard Karls University Tuebingen – Evaluation of Tumor Heterogeneity Using Clustering of Multi-Modality Imaging Data – Fortuene, grant number 2131-0-0.

References

1. Eisenhauer EA, et al. (2009) New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 45(2):228–247.
2. O'Connor JPB, et al. (2016) Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 14(3):169–186.
3. Judenhofer MS, et al. (2008) Simultaneous PET-MRI: a new approach for functional and morphological imaging. *Nat Med* 14(4):459–465.
4. Disselhorst JA, Bezrukov I, Kolb A, Parl C, Pichler BJ (2014) Principles of PET/MR Imaging. *J Nucl Med* 55(Supplement 2):2S–10S.
5. Bailey DL, et al. (2015) Combined PET/MR: The Real Work Has Just Started. Summary Report of the Third International Workshop on PET/MR Imaging; February 17–21, 2014, Tübingen, Germany. *Mol Imaging Biol* 17(3):297–312.
6. Gillies RJ, Beyer T (2016) PET and MRI: Is the Whole Greater than the Sum of Its Parts? *Cancer Res* 76(21):6163–6166.
7. Schmitz J, et al. (2016) Decoding Intratumoral Heterogeneity of Breast Cancer by Multiparametric In Vivo Imaging: A Translational Study. *Cancer Res* 76(18):5512–5522.
8. O'Connor JPB, et al. (2015) Imaging intratumor heterogeneity: Role in therapy response, resistance, and clinical outcome. *Clin Cancer Res* 21(2):249–257.
9. Schmidt H, et al. (2013) Correlation of simultaneously acquired diffusion-weighted imaging and 2-deoxy-[18F] fluoro-2-D-glucose positron emission tomography of pulmonary lesions in a dedicated whole-body magnetic

- resonance/positron emission tomography system. *Invest Radiol* 48(5):247–255.
10. Divine MR, et al. (2016) A population-based gaussian mixture model incorporating 18F-FDG PET and diffusion-weighted MRI quantifies tumor tissue classes. *J Nucl Med* 57(3):473–479.
 11. Stoyanova R, et al. (2016) Association of multiparametric MRI quantitative imaging features with prostate cancer gene expression in MRI-targeted prostate biopsies. *Oncotarget* 7(33):11.
 12. Katiyar P, et al. (2017) A Novel Unsupervised Segmentation Approach Quantifies Tumor Tissue Populations Using Multiparametric MRI: First Results with Histological Validation. *Mol Imaging Biol* 19(3):391–397.
 13. Katiyar P, et al. (2017) Spectral Clustering Predicts Tumor Tissue Heterogeneity Using Dynamic 18 F-FDG PET: A Complement to the Standard Compartmental Modeling Approach. *J Nucl Med* 58(4):651–657.
 14. Schölkopf B (2015) Artificial intelligence: Learning to see and act. *Nature* 518(7540):486–487.
 15. Siegemund M, et al. (2012) Superior antitumoral activity of dimerized targeted single-chain TRAIL fusion proteins under retention of tumor selectivity. *Cell Death Dis* 3(4):e295.
 16. Dubuisson M-P, Jain AK (1994) A Modified Hausdorff Distance for Object Matching. *Proceedings of 12th International Conference on Pattern Recognition* (IEEE Comput. Soc. Press), pp 566–568.
 17. Dice LR (1945) Measures of the Amount of Ecologic Association Between Species.

- Ecology* 26(3):297–302.
18. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* 8:25.
 19. Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C (2015) Translational Implications of Tumor Heterogeneity. *Clin Cancer Res* 21(6):1258–1266.
 20. Gatidis S, et al. (2015) Combined unsupervised-supervised classification of multiparametric PET/MRI data: application to prostate cancer. *NMR Biomed* 28(7):914–922.
 21. Nguyen Q, Valizadegan H, Hauskrecht M (2014) Learning classification models with soft-label information. *J Am Med Informatics Assoc* 21(3):501–508.
 22. Gillies RJ, Kinahan PE, Hricak H (2015) Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 278(2):151169.
 23. Aerts HJWL, et al. (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5:4006.
 24. Kumar V, et al. (2012) Radiomics: The process and the challenges. *Magn Reson Imaging* 30(9):1234–1248.
 25. Lee BS, et al. (2014) Induced Phenotype Targeted Therapy: Radiation-Induced Apoptosis-Targeted Chemotherapy. *JNCI J Natl Cancer Inst* 107(2):dju403-dju403.
 26. Stewart GD, et al. (2015) Sunitinib Treatment Exacerbates Intratumoral Heterogeneity in Metastatic Renal Cancer. *Clin Cancer Res* 21(18):4212–23.

27. Jaffray DA (2012) Image-guided radiotherapy: from current concept to future perspectives. *Nat Rev Clin Oncol* 9(12):688–699.
28. Hynynen K (2011) MRIgHIFU: A tool for image-guided therapeutics. *J Magn Reson Imaging* 34(3):482–493.
29. Sequist L V, et al. (2011) Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med* 3(75):75ra26.
30. Hodi FS, et al. (2016) Evaluation of Immune-Related Response Criteria and RECIST v1.1 in Patients With Advanced Melanoma Treated With Pembrolizumab. *J Clin Oncol* 34(13):1510–1517.

Author contributions

P.K., M.R.D., B.S., B.J.P., J.A.D. conceived the project. P.K., M.R.D., U.K., L.Q.M., J.A.D. collected the data. P.K., M.R.D., J.A.D. analyzed the data. M.S., R.K. provided expert knowledge and supplied the therapy. P.K., M.R.D., U.K., L.Q.M., B.S., B.J.P., J.A.D. wrote the paper. All authors edited the manuscript.

Figure legends

Fig. 1. *In vivo* parameters, volume and histology of the Control, Th-24 and Th-72 tumors and correlation matrices. (A) The acquired PET/MRI parameters of a representative tumor. Top row, left to right: ADC, T2 pre, T2 post and T2* post maps. Bottom row, left to right: dynamic ^{18}F -FDG PET frame 5, 10, 15 and 20. (B) Tumor volumes for the Control, Th-24 and Th-72 groups. The ends of the boxes represent the 25th and 75th quantiles and the whiskers extend to the 10th and 90th quantiles of the data. Asterisks indicate a significant difference between the volumes of the respective groups. (C) H&E, cleaved caspase-3 and Masson's trichrome stains representing the microenvironment of the Control, Th-24 and Th-72 group tumors. The extensive apoptosis indicated by the cleaved caspase-3 staining (dark brown areas) of the Th-24 tumor demonstrates the potent anti-tumor effects of the therapy. Likewise, the Masson's trichrome staining of the Th-72 tumor illustrates large populations of fibrous tissue (white asterisks). (D) Multiparametric MRI and (E) dynamic ^{18}F -FDG correlation matrices computed from the entire PET/MRI dataset.

Fig. 2. Tumor sectioning, the proposed multi-view learning framework and imaging-to-histology co-registration. (A) The experimental protocol used for partitioning the tumors. (B) The entire dataset was divided into a Training and two test sets. The histology and PET/MRI data of the Control and Th-24 tumors in the Training set were segmented using *K*-means and MSC, respectively. During this process, the MSC parameters were optimized by maximizing the *DSC* between the two segmentation maps. Subsequently, the labeled MSC phenotypic maps of all Training set tumors were pooled into one combined training dataset. This dataset was filtered by excluding the

voxels with a low MSC probability (<0.9) and used to train the unbiased random forest classifier. Finally, the trained classifier was applied to the imaging data of the tumors from the test sets to obtain the maps of intratumor heterogeneity. (C) Columns, left to right: T2W image, co-registered histology and the corresponding overlapping contours of four different tumors. The two top and bottom rows depict good and bad co-registration examples, respectively. The T2W images of the Control (row 1 and 4) and therapy (row 2 and 3) tumors were matched with the corresponding H&E and cleaved caspase-3 stains, respectively. (D) The in-plane, vertical and horizontal axis rotations applied to the T2W images to match the respective histology. (E) The resulting MHD and (F) DSC values between the 80 co-registered imaging-histology pairs. The ends of the boxes represent the 25th and 75th quantiles and the whiskers extend to the 10th and 90th quantiles of the data.

Fig. 3. Clustering of the Training set tumors and the cluster-wise pooled imaging parameters. (A) K -means and MSC segmentation results of the two Training set tumors. Top row, left to right: H&E and cleaved caspase-3 histology of a Control tumor, the corresponding K -means indexed image with two clusters, the ground truth viable and $rest_{control}$ clusters, and the phenotypic maps derived from the PET, MRI and PET/MRI data during MSC. Bottom row, left to right: H&E and cleaved caspase-3 histology of a Th-24 tumor, the corresponding K -means indexed image with two clusters, the ground truth apoptosis and $rest_{Th-24}$ clusters, and the phenotypic maps derived from the PET, MRI and PET/MRI data during MSC. For both tumors, the PET phenotypic map was obtained by segmenting the MRI-derived Fisher score-weighted PET data. (B) Cluster-wise pooled imaging parameters from all Training set tumors. Top row, left to right: the distributions of ADC, T2 pre and T2 post. Bottom row, left to right: the distributions of T2* post and

the intercept feature calculated by linearly fitting the last seven PET time-points of the Training set voxels and the cluster-wise average PET time activity curves. The ends of the boxes represent the 25th and 75th quantiles and the whiskers extend to the 10th and 90th quantiles of the data.

Fig. 4. Proximity matrices and relative feature importance scores. (A) PET, (B) MRI and (C) PET/MRI proximity matrices derived from the respective classifiers. The viable, apoptosis and *rest* clusters are denoted by *V*, *A* and *R*. Each element of the $N \times N$ (where, N = number of voxels in the Training set) proximity matrix represents the fraction of trees (normalized by twice the number of trees) for which the corresponding row and column voxels were classified as the same tissue type. The PET and MRI classifiers complemented each other by exhibiting relatively higher proximities in the viable and apoptosis clusters, and *rest* cluster, respectively. This feature enabled the PET/MRI classifier to adequately distinguish all three clusters. (D) Relative feature importance scores derived from the PET/MRI classifier. The six (3 PET and 3 MRI) most relevant features are highlighted in red.

Fig. 5. Phenotypic maps predicted by the trained classifiers. Column-wise, left to right: H&E, cleaved caspase-3 and Masson's trichrome histology of six representative test set tumors and the corresponding phenotypic maps predicted by the PET, MRI and PET/MRI classifiers. The top, middle and bottom pairs of the tumors belong to the test Control, Th-24 and Th-72 groups, respectively. The color-coded arrows in the histology indicate the same tissue type shown in the figure legend. For each tumor, the PET/MRI classifier extracted the most relevant information from the dynamic PET and multiparametric MRI data and provided accurate maps of the intratumor heterogeneity.

Fig. 6. DSCs, test fractional volumes and intratumor heterogeneity of a Control tumor. (A) PET, MRI and PET/MRI *DSCs* for the Training and Test-1 set tumors (Control and Th-24). The training *DSCs* were computed between the MSC and histology phenotypic maps, whereas the test *DSCs* were calculated among the classifier and histology phenotypic maps. (B) The test fractional viable, apoptosis and *rest* volumes predicted by the PET/MRI classifier. The ends of the boxes represent the 25th and 75th quantiles and the whiskers extend to the 10th and 90th quantiles of the data. Asterisks indicate a significant difference between the fractional volumes of the respective groups. (C) Top row: cleaved caspase-3 histology of an exemplary Control tumor. The inset shows natural apoptosis present in the tumor. Bottom row, left to right: the associated PET and PET/MRI phenotypic maps and H&E histology of the tumor. Among both models, the PET/MRI classifier was more sensitive in identifying apoptotic tissue in the Control tumors.

Fig. 7. Classification of the glioblastoma tumors. Column-wise, left to right: cleaved caspase-3, CD-31, glucose transporter 1, H&E, and Ki-67 histology of five glioblastoma xenografts and the associated phenotypic maps predicted by the 2- and 3-class MRI classifiers. The color-coded arrows in the histology indicate the same tissue type shown in the figure legend. Furthermore, the purple arrow in the H&E histology of the fourth tumor points towards the muscle tissue. The precise identification of the viable and necrotic areas by the 2-class MRI classifier confirms that the multi-view learning concepts are transferrable from one tumor model to another.

Figures

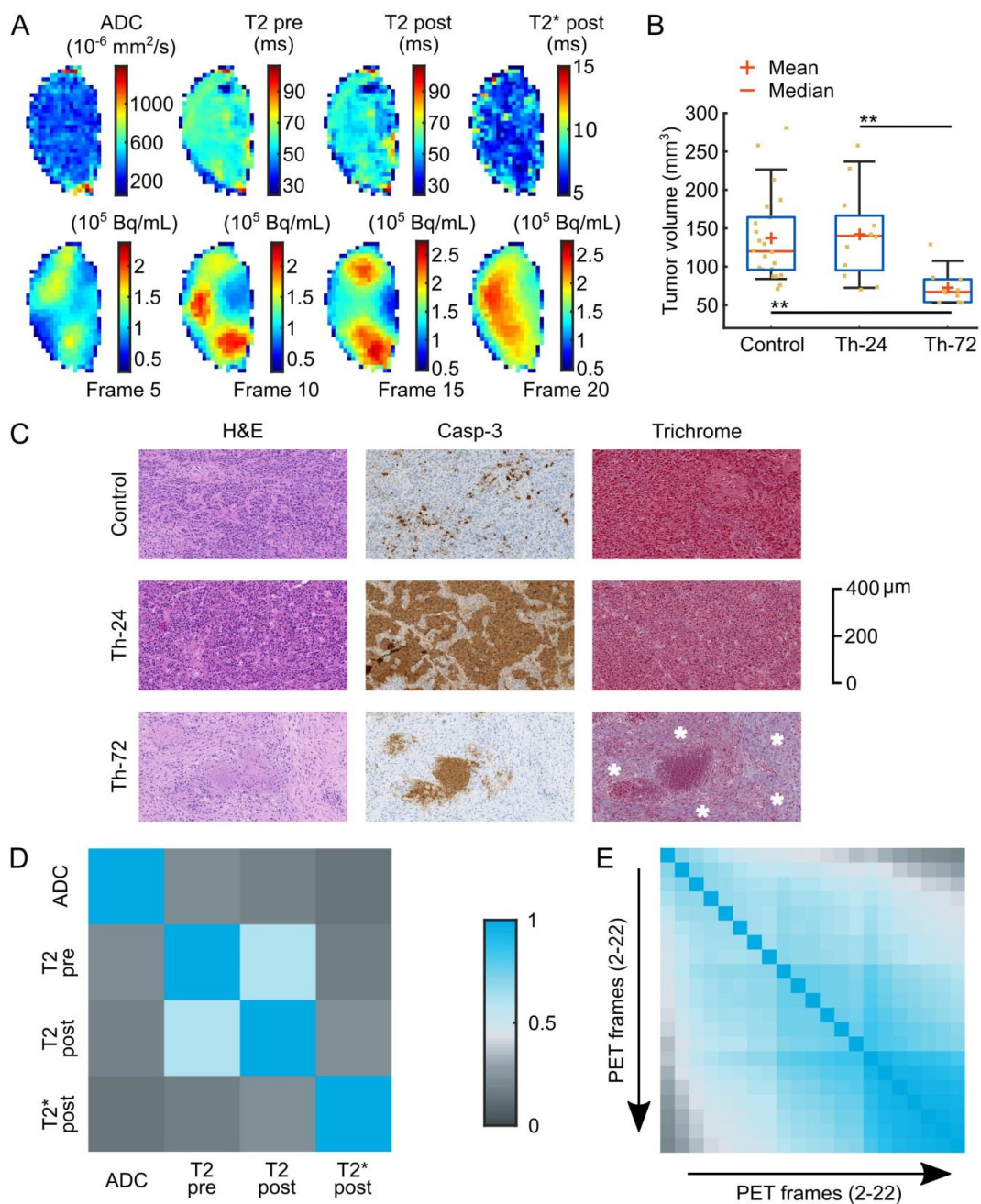


Fig. 1

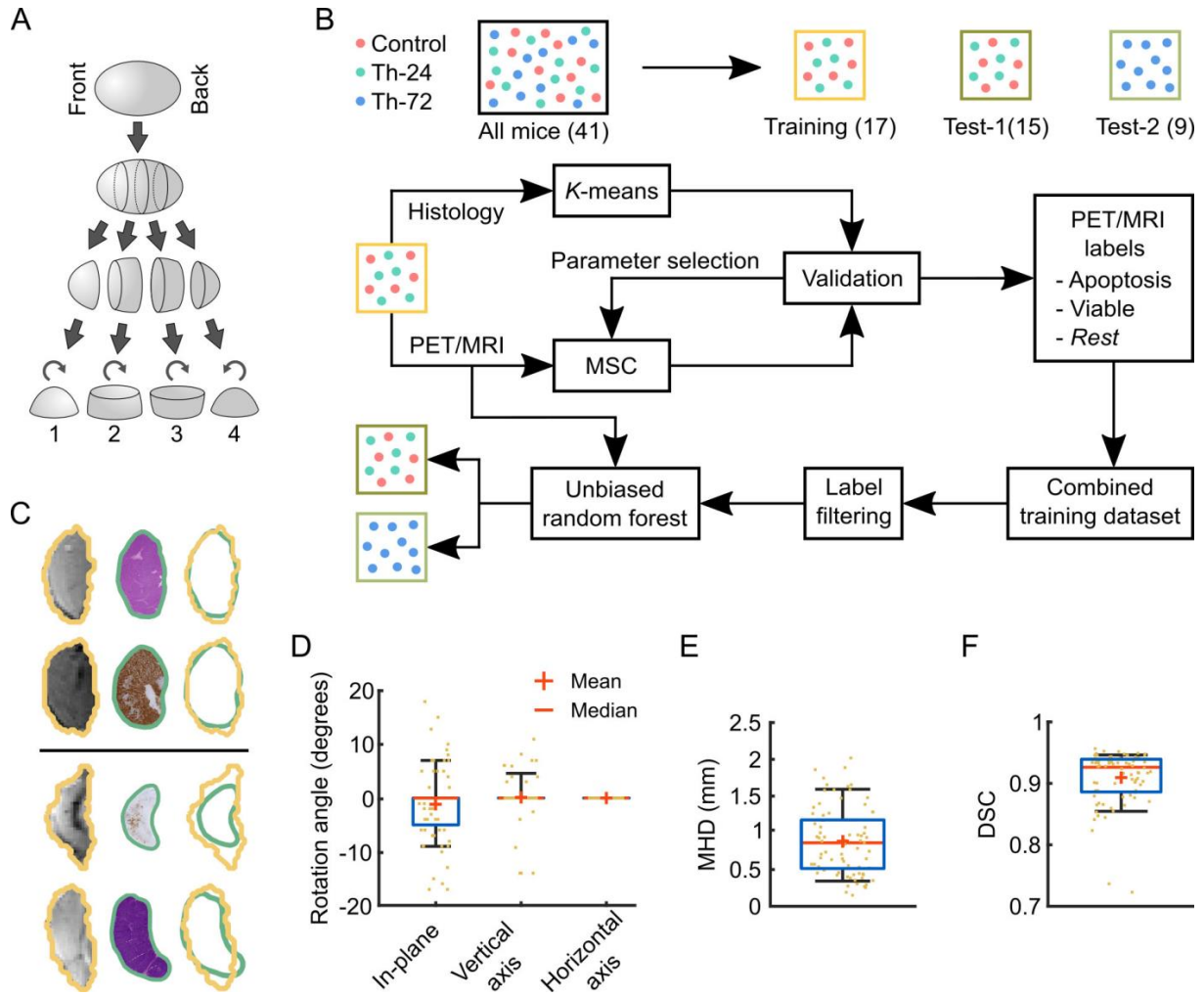


Fig. 2

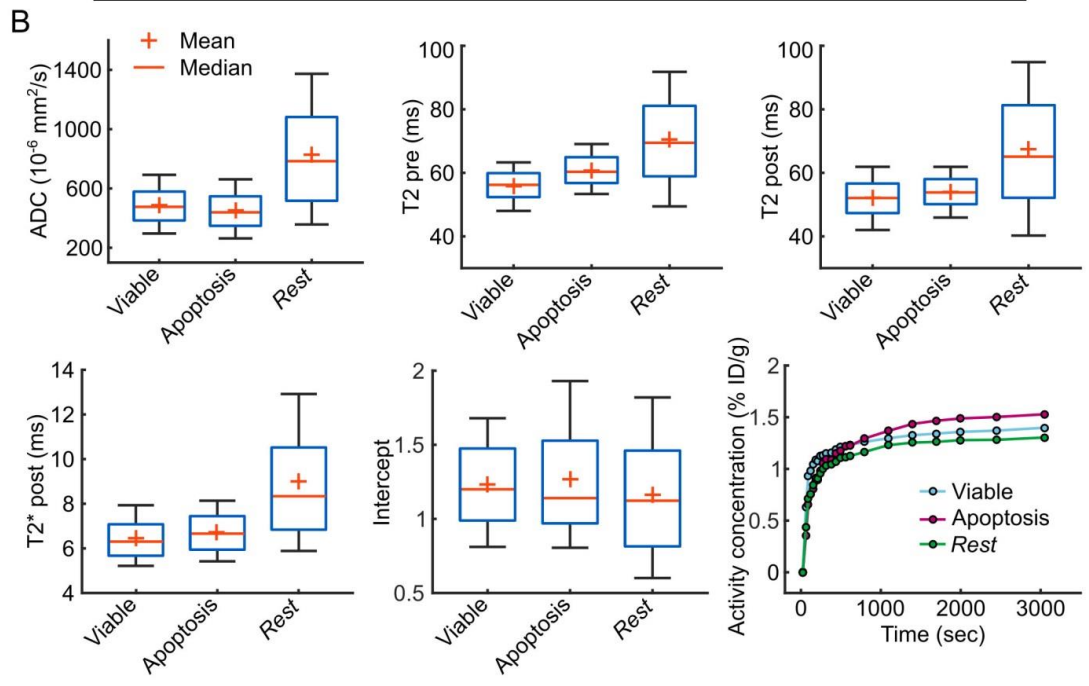
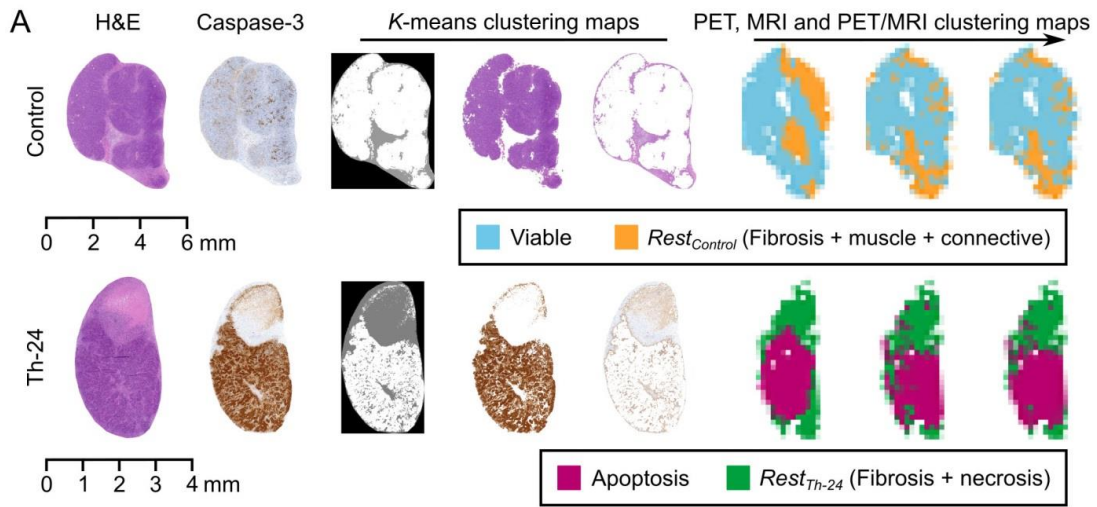


Fig. 3

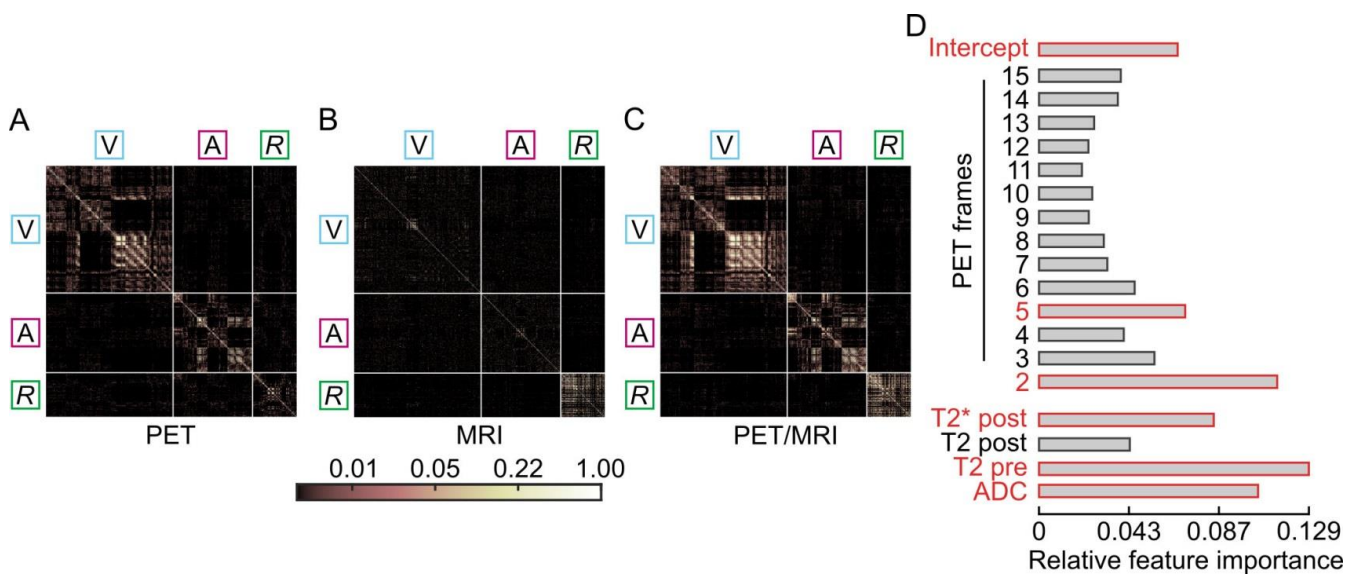


Fig. 4

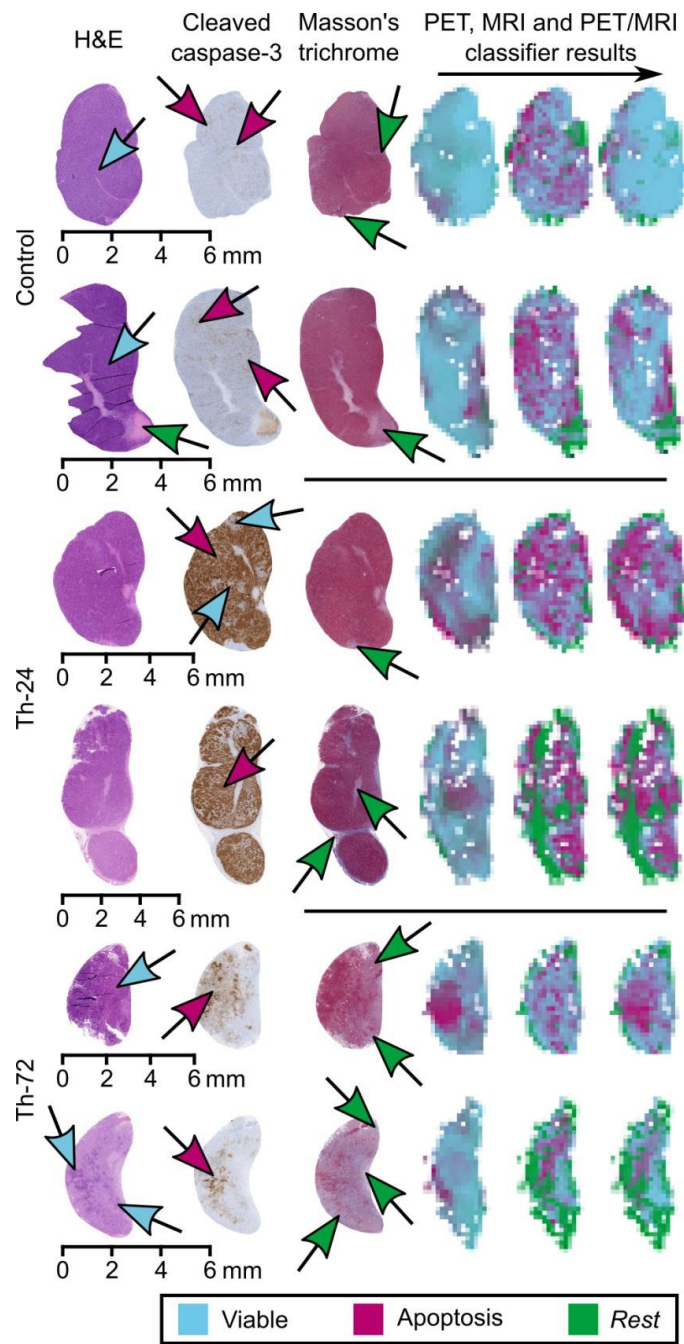


Fig. 5

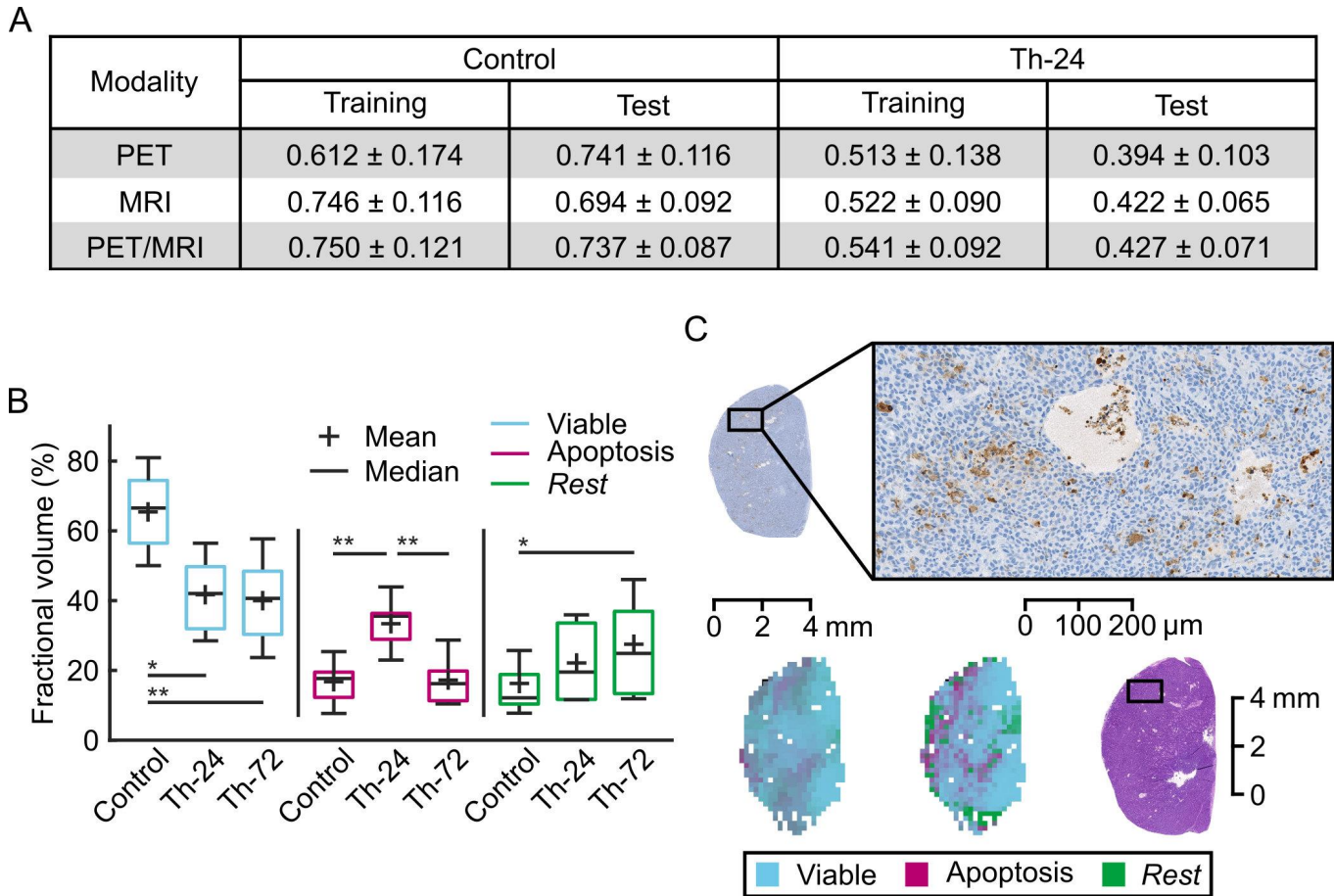


Fig. 6

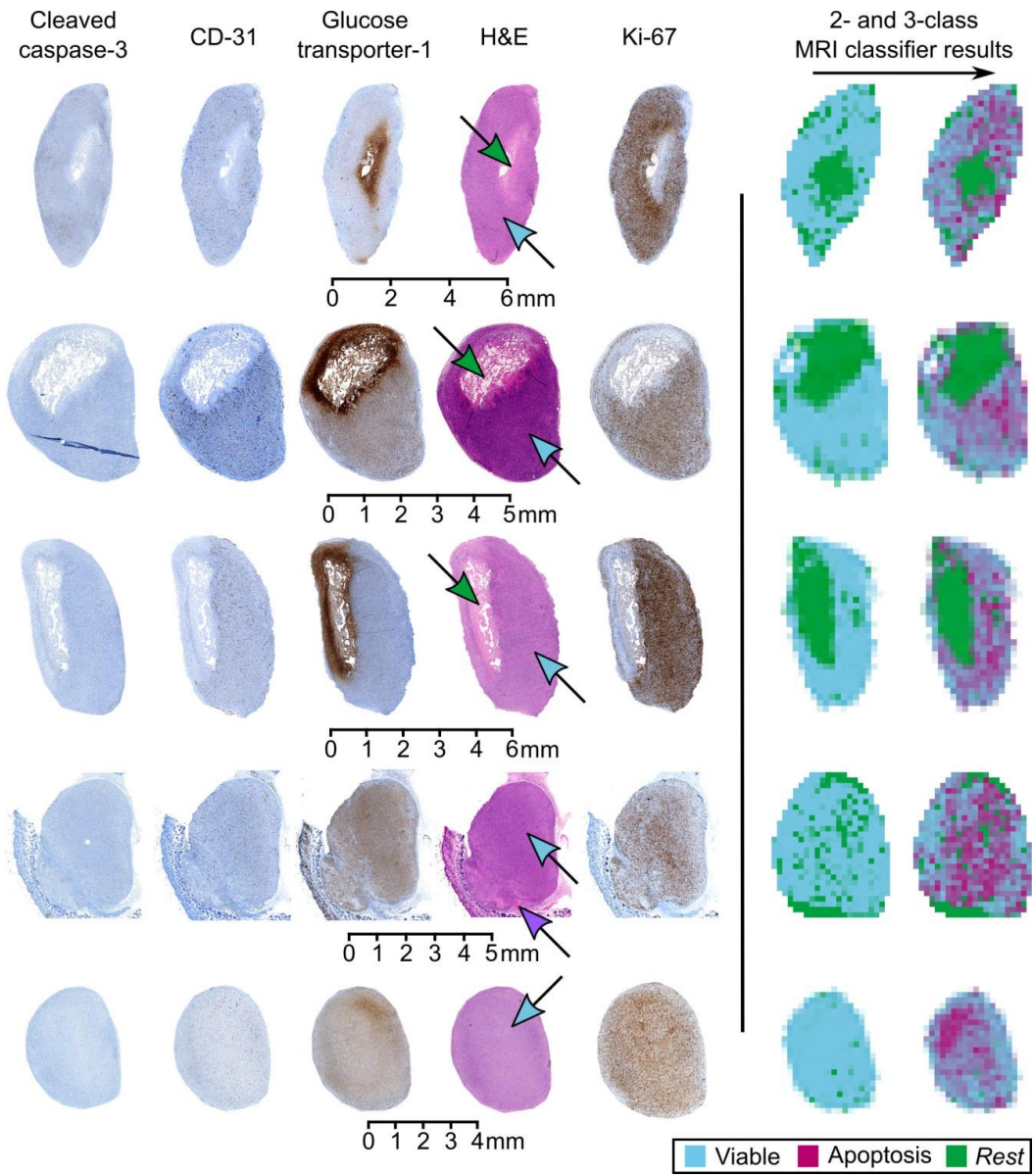


Fig. 7

Supplementary materials and methods

Animal handling

Six-week-old NMRI nu/nu female mice were ordered from Charles River, Germany and allowed to acclimate in the on-site animal vivarium prior to subcutaneous injection of COLO-205 (4.5×10^6) tumor cells (periodically tested for mycoplasma contamination). Mice were housed in a specific pathogen free environment in separate ventilated cages. The vivarium's temperature was modulated at $20 \pm 1^\circ \text{C}$ and a relative humidity of $50 \pm 10\%$ was maintained. The animals were given water and food *ad libitum* and the lighting was adjusted to a 12 h night and day scheme. The imaging experiments were performed once the tumors were palpable (with a minimum length of 5 mm) and exhibited signs of vascularization. To maintain a deep anesthesia, mice spontaneously respired 1.5% isoflurane dissolved in 100% O_2 at a flow rate of 0.8 L/min, before and during tumor inoculation and imaging experiments. FDG and ferumoxytol were injected through an 80-cm catheter in the tail vein of the mice while keeping them under anesthesia. All experiments were carried out in a specific pathogen free environment. Respiration of the mice was monitored during the imaging experiments using a pneumatic air pad connected to a monitoring and gating system (Model 1030; Small Animal Instruments, Stony Brook, NY, USA).

Data acquisition

Image acquisition

The 80 cm tail vein catheter was filled with $10.88 \pm 1.14 \text{ MBq}$ (decay-corrected to the start of the scan) of ^{18}F -FDG in 50 μL of 0.9% NaCl prior to acquisition. After the PET/MRI acquisition was started, 50 μL of ferumoxytol was used to administer the

activity to the animal. An additional 50 μL of ferumoxytol was later injected to obtain contrast-enhanced T2 and T2* maps.

MRI settings

The following MRI sequences were acquired using a mouse brain surface coil (Bruker BioSpin):

T2 weighted imaging sequence: turbo spin echo with repetition time (TR) = 3500 ms, echo time (TE) = 205 ms, echo train length = 161, image size = 256 \times 160 and voxel size (mm^3) = 0.22 \times 0.22 \times 0.22.

Diffusion weighted imaging sequence: half-Fourier acquisition single-shot turbo spin echo (HASTE). B-values: 200, 400, 600, 800 and 1000 s/mm^2 . HASTE settings: TR = 5000 ms, TE = 112 ms, echo train length = 256, number of averages = 4, image size = 192 \times 120 and voxel size (mm^3) = 0.21 \times 0.21 \times 1.00.

T2 map sequence: spin echo with 12 echo times (TE = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110 and 120 ms). Additional settings: TR = 2500 ms, number of averages = 2, image size = 192 \times 120 and voxel size (mm^3) = 0.21 \times 0.21 \times 1.00.

T2* map sequence: gradient echo with 10 different echo times varying from 3 to 26.85 ms with an interval of 2.65 ms. Additional settings: TR = 2000 ms, number of averages = 2, image size = 192 \times 120 and voxel size (mm^3) = 0.21 \times 0.21 \times 1.00.

All the maps were calculated with in-house developed software in MATLAB (R2013a) using linear regression on the natural logarithm transformed imaging data. The T2 maps were calculated without the first echo (i.e., 11 echo's in total), whereas the T2* maps were calculated with all 10 echo's. The T2 weighted and diffusion images were acquired with respiratory triggering.

PET settings

PET images were simultaneously acquired with the MR on a commercial version of the PET insert mentioned in Wehrl et al. (1). For each acquisition, 60 min of data were acquired in list-mode format. Each dataset was binned into a 22 time-frame sinogram, and reconstructed with the ordered subsets expectation maximization (OSEM2D) algorithm in 16 iterations with 4 subsets. Reconstruction resulted in a $256 \times 256 \times 89$ voxel image for each frame. The first time-frame was of variable duration (~ 30 s), from start of the acquisition until the time of injection. This frame was not used in the analyses. All subsequent time-frames (8×30 , 6×60 , 5×300 , 2×600 s, adding to 55 min) were used. Decay-correction, dead-time correction, and corrections for random were applied to the data, but no attenuation correction was applied.

For the details of glioblastoma tumors the reader is referred to Katiyar et al. (2).

Imaging to histology co-registration

A rigid co-registration between the T2W images and histology was performed by selecting the matching imaging (axial) slice, followed by in-plane (clockwise or anticlockwise), vertical and horizontal axis rotations to fit its contours with the contours of the associated histology section. The Control tumors were co-registered with H&E, whereas the Th-24 and Th-72 tumors were matched with cleaved caspase-3 histology. Special care was taken to match the tumor imaging planes in the same order (1–4) as the sectioned histology (Fig. 2A). In a few cases, the histology sections were shuffled as 4–1, and therefore, the entire tumor volume was rotated along the vertical axis by ~ 180 degrees to keep the imaging planes in the same order. In these cases, 180 degrees was subtracted from the final vertical axis rotation angle (Fig. 2D). Because of the small size of the histology or the amount of tumor in the histology section, all four histology slices

of each tumor could not be adequately matched with the corresponding imaging planes (slice 1×4, slice 2×29, slice 3×31 and slice 4×16). The co-registration was performed before implementing the multi-view learning framework.

The modified Hausdorff distance (*MHD*) (3) between the contours of the matched imaging-histology pairs was calculated in two steps. In the first step, the directed distance between two sets of points $I = \{i_1, i_2, \dots, i_{N_i}\}$ and $H = \{h_1, h_2, \dots, h_{N_h}\}$ was calculated using the following expression:

$$d(I, H) = \frac{1}{N_i} \sum_{i \in I} (d(i, H)).$$

Where, N_i and N_h are total number of points in imaging set I and histology set H , respectively and the distance metric between two points was defined as the Euclidean distance. In the second step, the directed distances $d(I, H)$ and $d(H, I)$ were combined to obtain the *MHD*:

$$MHD = \max(d(I, H), d(H, I)).$$

The Dice similarity coefficient (*DSC*) (4) between *T2W* image and *Histology* was calculated as follows:

$$DSC = \frac{2|T2W \cap Histology|}{|T2W| + |Histology|}.$$

Where, $|T2W|$ and $|Histology|$ is the area in each image.

Multi-view spectral clustering

The multi-view spectral clustering (*MSC*) workflow can be broadly divided into three major stages (Fig. S1). In the first (gray) stage, spectral clustering (*SC*) (5) was applied to

the multiparametric MRI data of the Training set tumors. In SC, the MRI features vectors (ADC, T2pre, T2post and T2*post) were used to compute the MRI affinity matrix (W_{MRI}):

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

Where σ is the scale parameter of the radial basis function kernel and $\|x_i - x_j\|$ is the pairwise Euclidian distance between the MRI feature vectors of voxel i and j . Subsequently, the normalized Laplacian (2, 5) was derived from W_{MRI} and Gaussian mixture modeling (GMM) was performed on the M first Eigenvectors of the normalized Laplacian matrix. GMM is a well-established unsupervised segmentation method. Additional details of this technique are provided in Bishop (6).

The Training Control tumors were segmented into viable and $rest_{control}$ clusters, whereas the Th-24 tumors were divided into apoptosis and $rest_{Th-24}$ clusters. Similarly, the previously co-registered tumor histology images (H&E for the Control tumors and cleaved caspase-3 for the Th-24 tumors) were segmented into the two respective classes. Histology segmentation was carried out by converting the RGB histology image into LAB color space and applying K -means clustering to the A and B channels. During clustering, the optimal SC parameters (σ and M) were obtained by minimizing the DSC between the MRI and histology segmentation maps. A simple grid search was used for the parameter selection. Due to a complex intratumor microenvironment (Fig. S4), the histology images of the Th-72 tumors could not be segmented using K -means clustering.

In the second (yellow) stage, the histologically validated MRI labels were used to calculate the feature scores (Fig. S8) of the dynamic PET data. The Fisher scoring algorithm (7) was used to weight each PET frame prior to SC. The Fisher weight $F(p)$ for frame p was calculated using the following expression:

$$F(p) = \frac{\sum_{k=1}^c n_k (\mu_k^p - \mu^p)^2}{\sum_{k=1}^c n_k (\sigma_k^p)^2},$$

where c and μ^p denote the total number of classes (in this case 2 for each tumor) and the mean of all tumor voxels for frame p . The mean and standard deviation of all the voxels (n_k) in class k corresponding to frame p are denoted by μ_k^p and σ_k^p . The voxels were grouped into different classes based on the MRI labels. For each training set tumor the Fisher scores were scaled between 0 and 1.

SC on the Fisher score-weighted PET data was performed in the same manner as on the multiparametric MRI features (not fully shown in Fig. S1). In addition, for comparison, the unweighted PET data of each Training set tumor was also segmented using SC. Overall, the average DSC of the Control and Th-24 Training set tumors for the unweighted segmentation scheme was lower than that of the weighted methodology (Table S4).

In the final (blue) stage, the PET/MRI affinity matrix ($W_{PET/MRI}$) was obtained by convex combination of the affinity matrices (corresponding to the optimum σ) of each of the modalities:

$$W_{PET/MRI} = \alpha * W_{PET} + (1 - \alpha) * W_{MRI}.$$

Here W_{PET} corresponds to the affinity matrix calculated from the Fisher score-weighted PET data and α controls the influence of each modality ($0 \leq \alpha \leq 1$). Finally, the PET/MRI labels were obtained by segmenting the M first Eigenvectors of the normalized Laplacian matrix using GMM. Again, a grid search was performed to choose the optimal parameters (α and M), while minimizing the DSC between the histology and PET/MRI segmentation maps. The resulting values of α are shown in Fig. S9.

In each step above, the normalized Laplacian matrix was calculated using the following equation:

$$L = D_{inv} * W_{ij} * D_{inv},$$

where $D_{inv} = D^{-0.5}$ and D is the diagonal matrix with diagonal vector $d_i = \sum_{j=1}^n W_{ij}$. The parameter n denotes the total number of voxels in the tumor.

Unbiased random forest

The settings used for the cforest function were as follows: random seed = 42 and number of trees = 500. The number of features randomly sampled for each split was set to $round(\sqrt{P})$, where P is the total number of input features. The remaining settings were kept to default (8).

Supplementary tables

Table S1. P-values for group-wise volume comparisons. The Th-72 tumors significantly shrunk in size (Fig. 1B) and could be separated from the Control and Th-24 tumors using just the tumor volume. The p-values were obtained using the Bonferroni corrected rank sum tests.

Volume	Groups		
	{Control, Th-24}	{Control, Th-72}	{Th-24, Th-72}
	6.14×10^{-1}	2.68×10^{-4}	5.06×10^{-4}

Table S2. P-values for group-wise fractional volume comparisons. The fractional volumes predicted by the PET/MRI classifier in the test set were indicative of the tissue heterogeneity induced by the therapy (Fig. 6B). The p-values were obtained using the Bonferroni corrected rank sum tests.

Fractional volume (cluster)	Groups		
	{Control, Th-24}	{Control, Th-72}	{Th-24, Th-72}
Viable	1.00×10^{-3}	3.79×10^{-5}	8.12×10^{-1}
Apoptosis	3.01×10^{-4}	9.62×10^{-1}	8.74×10^{-4}
<i>Rest</i>	7.64×10^{-2}	2.90×10^{-3}	2.67×10^{-1}

Table S3. DSCs between the phenotypic maps predicted by the PET/MRI and PET/MRI_{pred} classifiers. The classifier trained on the most relevant PET/MRI features provided reasonably matching results for the viable and *rest* clusters, however, it over/under-estimated apoptosis in all the test tumors, as compared to the PET/MRI model trained on all the features.

Group	Cluster		
	Viable	Apoptosis	<i>Rest</i>
Control	0.924	0.588	0.772
Th-24	0.753	0.721	0.824
Th-72	0.873	0.716	0.842

Table S4. PET frame weighting using MRI-derived Fisher score improved SC intratumor segmentation. The DSCs of the Fisher score-weighted and unweighted PET SC for the Training Control and Th-24 tumors. The former scheme improved clustering of the dynamic ^{18}F -FDG PET data by giving more weight to the predictive PET time points.

Segmentation scheme	Training set tumors	
	Control	Th-24
Fisher score-weighted PET	0.612 ± 0.174	0.513 ± 0.138
Unweighted PET	0.549 ± 0.211	0.487 ± 0.137

Supplementary figure legends

Fig. S1. MSC workflow. The MSC workflow broadly comprised three analysis stages (shaded with gray, yellow and blue). The first (gray) and second (yellow) stages utilized SC to segment multiparametric MRI and MRI-derived Fisher score-weighted dynamic ^{18}F -FDG PET data, respectively. In the third stage, complementary information from both the previous steps was combined to obtain PET/MRI maps of intratumor heterogeneity. The Control tumors in the training set were segmented into viable and *rest_{control}* clusters, whereas the Th-24 tumors were divided into apoptosis and *rest_{Th-24}* clusters.

Fig. S2. Cleaved caspase-3, F4-80 and Masson's trichrome stainings of a Th-24 tumor. (A) Left to right: cleaved caspase-3 and F4-80 staining of an exemplary Th-24 tumor. The images in the inset show macrophages present at the border and in the fibrous area of the tumor. (B) The Masson's trichrome insets in this panel show three non-apoptotic regions of the Therapy-24 tumor, which majorly consist of loose fibrous tissue.

Fig. S3. Necrosis in the Th-24 tumors. In several Th-24 tumors, the non-apoptotic regions contained mixed areas of necrotic and fibrous tissues. This is clearly depicted by the H&E inset in (A), where the necrotic region gradually mixes with the fibrous portions of the tumor (B). The two Masson's trichrome insets also highlight the complexity of the *rest* cluster, which made the separation of these two tissue classes unfeasible using imaging. The rest of the tumor was predominantly apoptotic, as shown by the cleaved caspase-3 stain in (C).

Fig. S4. Masson's trichrome and H&E stainings of a Th-72 tumor. (A) In the Th-72 tumors, the *rest* cluster portions were significantly enlarged. The white squares in the

Masson's trichrome inset exhibit large clusters of fibrous tissue surrounding small necrotic areas (shown with the white circles). (B) In addition, highly composite regions with viable and apoptotic tissues were also present, as depicted by the H&E inset. In general, due to the complex intratumor characteristics, the analysis of the Th-72 tumor histology images was beyond the scope of color based tissue segmentation and would warrant sophisticated algorithmic development.

Fig. S5. Tissue heterogeneity in the Control tumors. Focally present muscle, connective, and fibrous tissues in the Training set Control tumors were excluded from the supervised learning. The H&E (A) and Masson's trichrome (B) insets in this image present one such example, where all the three tissues can be seen.

Fig. S6. Viable tissue in the Th-24 tumors. Cleaved caspase-3 staining of two (A and B) Th-24 tumors. The small viable patches can be clearly seen in both insets. These patches were surrounded by large apoptotic regions and hence, posed significant challenges during tissue classification.

Fig. S7. Comparison of the PET/MRI and PET/MRI_{pred} classifiers. Column-wise, left to right: H&E, cleaved caspase-3 and Masson's trichrome histology of the three test set tumors, and the corresponding phenotypic maps predicted by the PET and PET/MRI_{pred} classifiers. The first, second and third row tumors belong to the Control, Th-24 and Th-72 groups from the test set, respectively. The color-coded arrows in the histology indicate the same tissue type shown in the figure legend.

Fig. S8. MRI-derived Fisher scores. The MRI-derived Fisher scores for ¹⁸F-FDG PET frames (2-22) of the Training set tumors. Due to low or zero activity, the first frames were excluded from the proposed multi-view learning analysis. The solid lines and

shaded regions depict the mean and standard deviation of the Fisher scores for all the Training Control and Th-24 tumors, respectively.

Fig. S9. Parameter α . The distribution of the parameter α . The ends of the boxes represent the 25th and 75th quantiles and the whiskers extend to the 10th and 90th quantiles of the data.

Supplementary figures

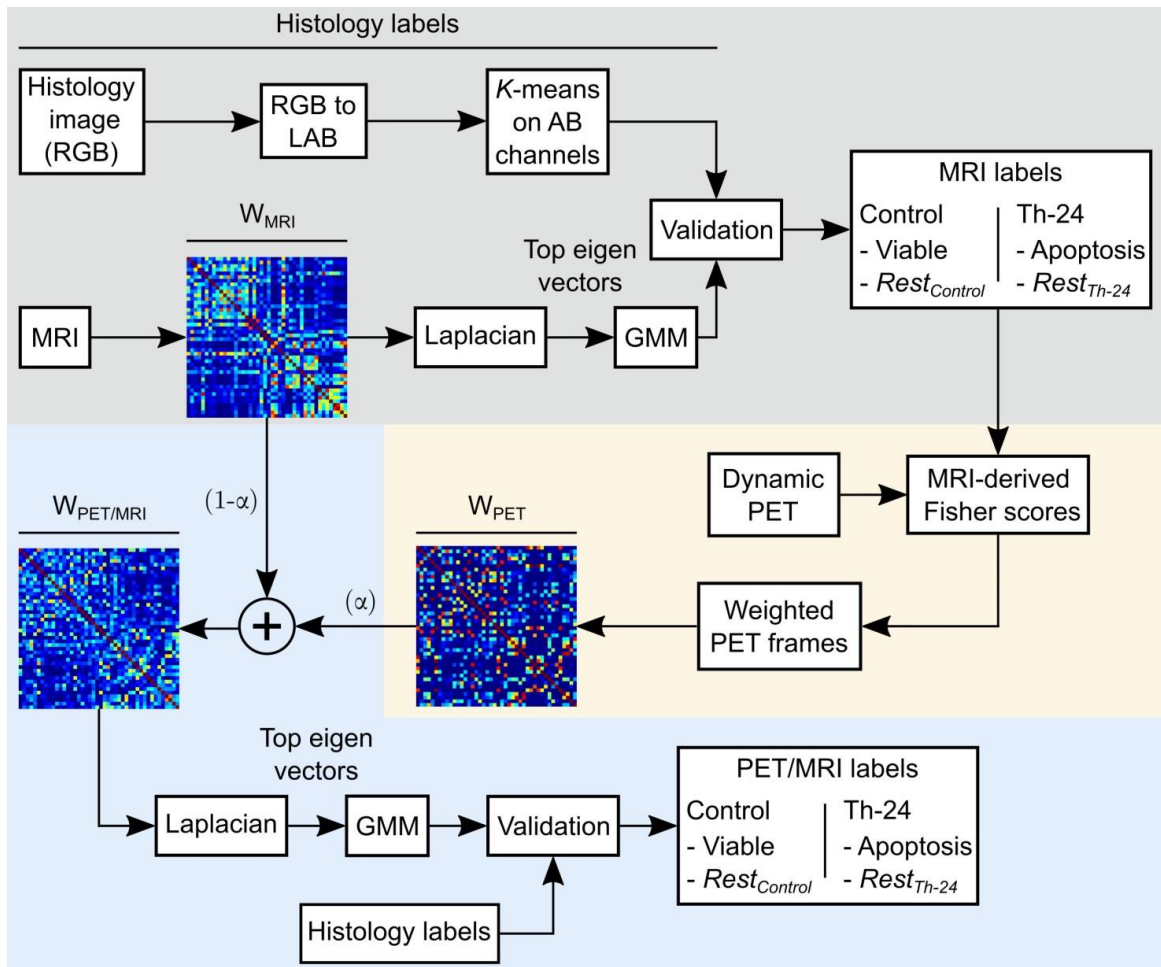


Fig. S1

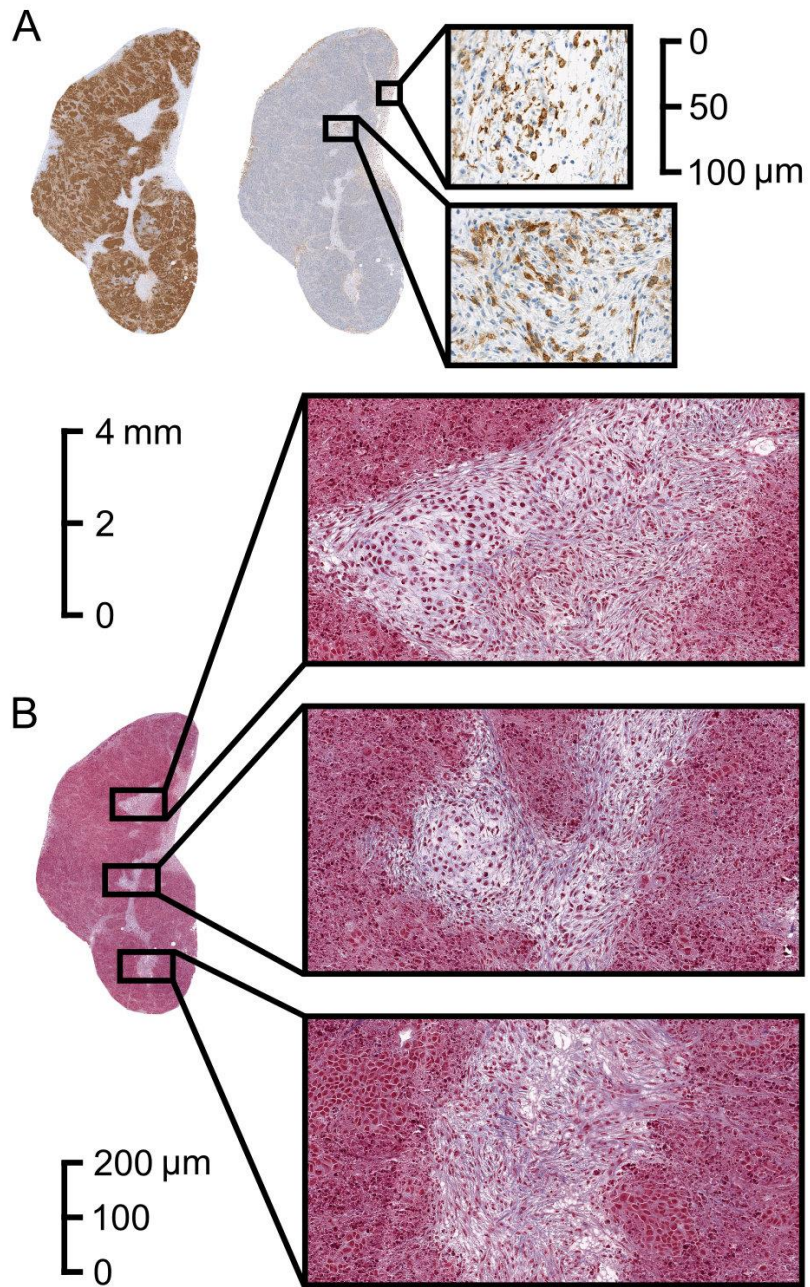


Fig. S2

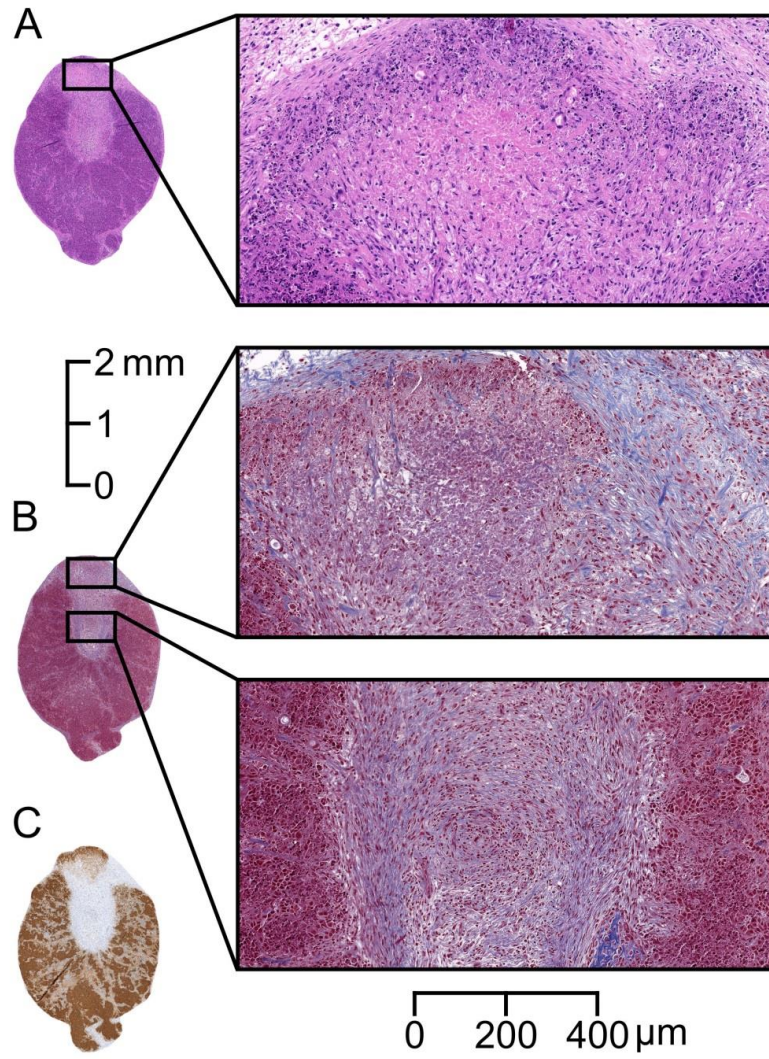


Fig. S3

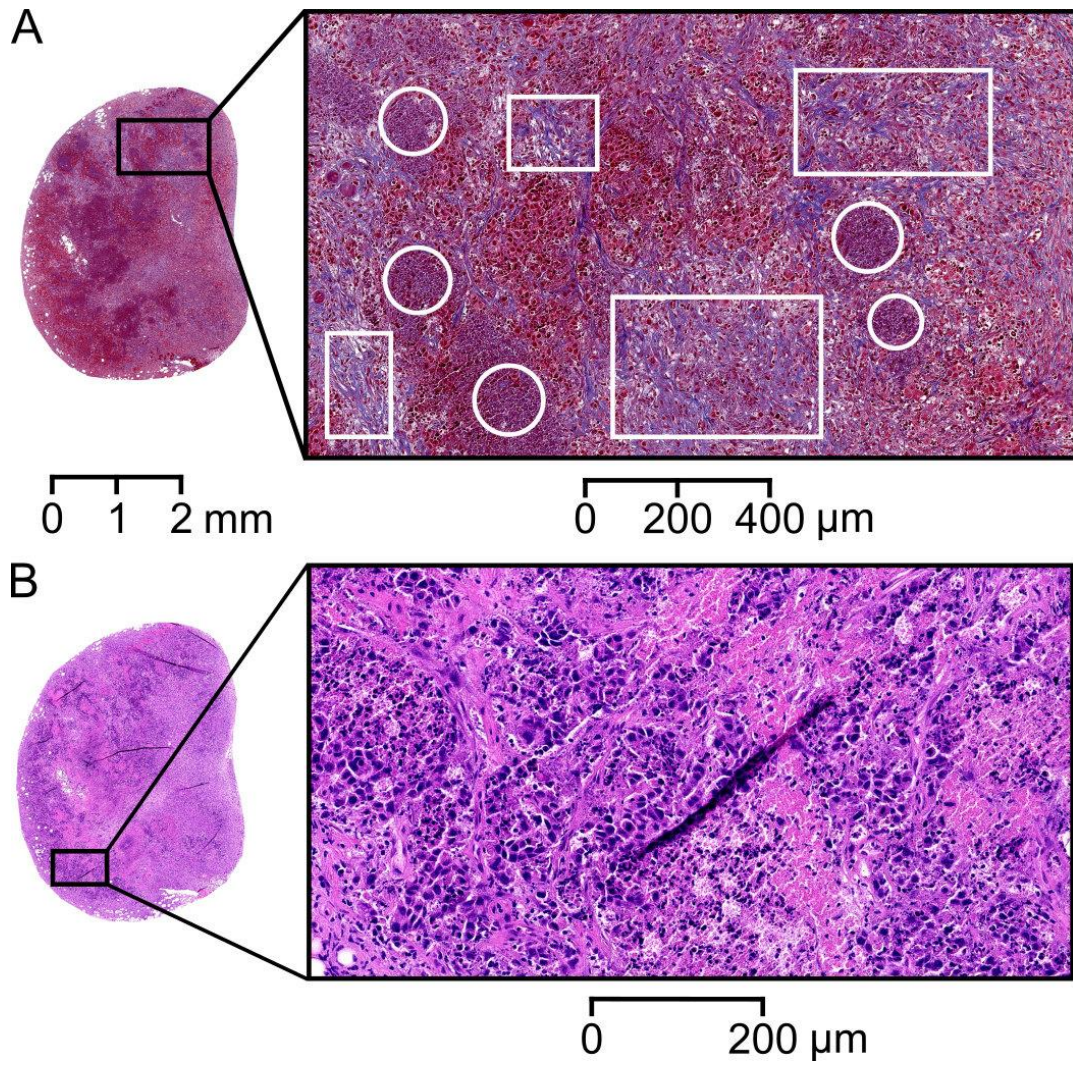


Fig. S4

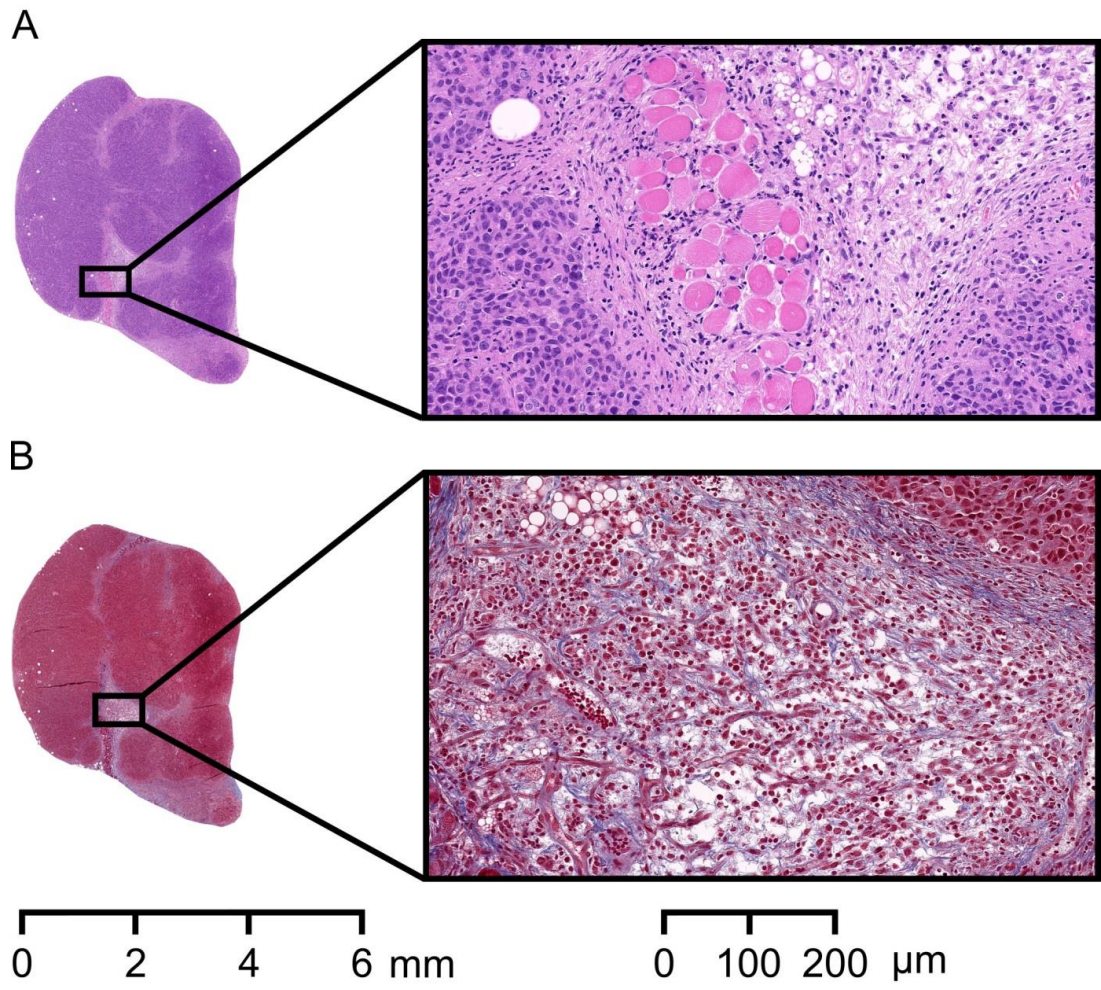


Fig. S5

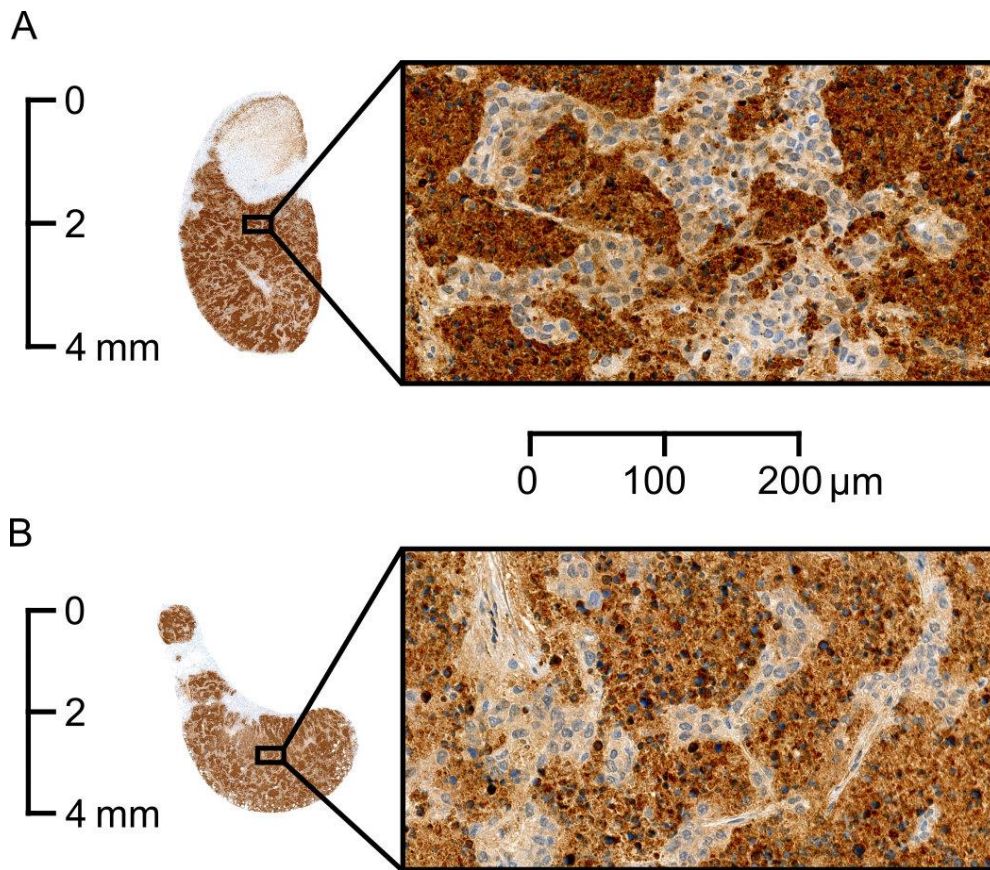


Fig. S6

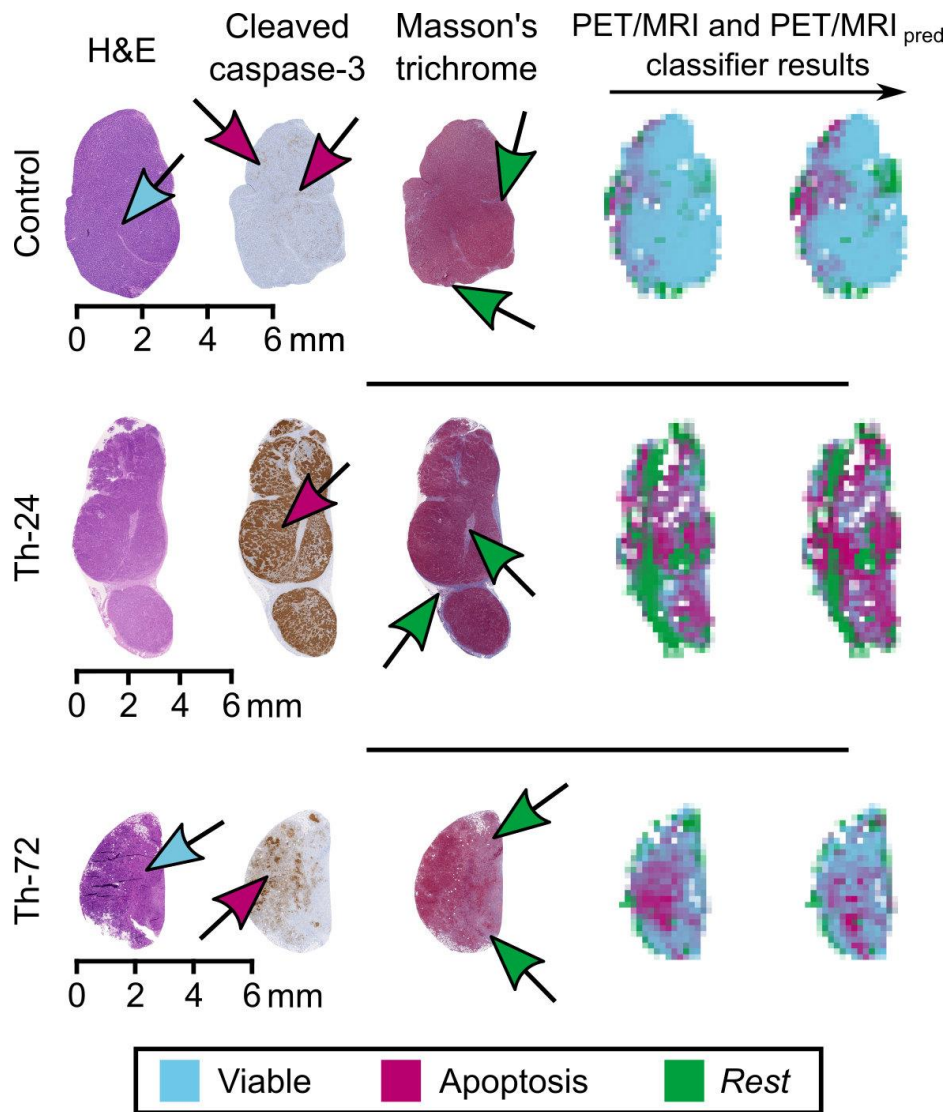


Fig. S7

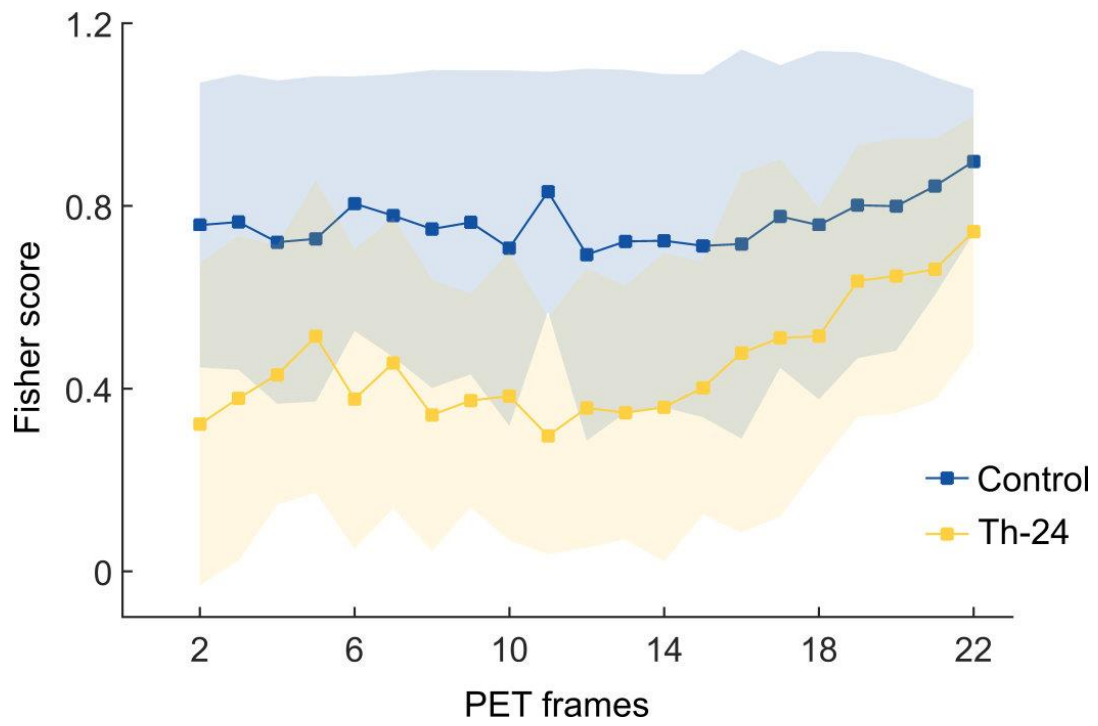


Fig. S8

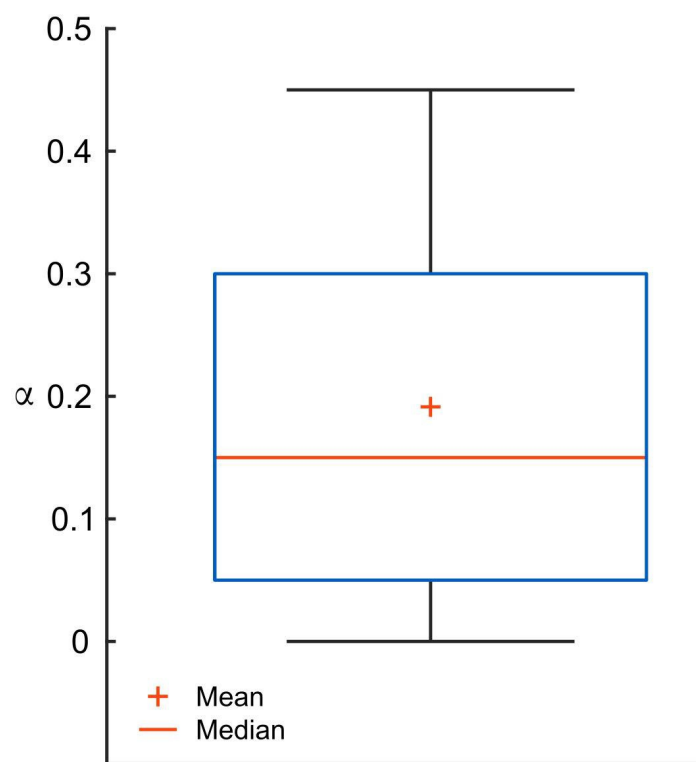


Fig. S9

Supplementary references

1. Wehrl HF, et al. (2013) Simultaneous PET-MRI reveals brain function in activated and resting state on metabolic, hemodynamic and multiple temporal scales. *Nat Med* 19(9):1184–1189.
2. Katiyar P, et al. (2017) A Novel Unsupervised Segmentation Approach Quantifies Tumor Tissue Populations Using Multiparametric MRI: First Results with Histological Validation. *Mol Imaging Biol* 19(3):391–397.
3. Dubuisson M-P, Jain AK (1994) A Modified Hausdorff Distance for Object Matching. *Proceedings of 12th International Conference on Pattern Recognition* (IEEE Comput. Soc. Press), pp 566–568.
4. Dice LR (1945) Measures of the Amount of Ecologic Association Between Species. *Ecology* 26(3):297–302.
5. Von Luxburg U (2007) A tutorial on spectral clustering. *Stat Comput* 17(4):395–416.
6. Bishop CM (2006) *Pattern Recognition and Machine Learning*
doi:10.1117/1.2819119.
7. Duda RO, Hart PE, Stork DG (2001) *Pattern Classification* doi:10.1038/npp.2011.9.
8. Strobl C, Hothorn T, Zeileis A (2009) Party on! *R J* 1(2):14–17.

Discussion and Outlook

The central goal of this doctoral thesis was to investigate the advantages of simultaneously acquired dynamic ^{18}F -FDG PET/multiparametric MRI data in decoding intratumor heterogeneity using various unsupervised and supervised machine learning techniques.

To this end, we first addressed the elementary question of modeling the spatial and temporal tumor tissue heterogeneity using longitudinally acquired two-dimensional ^{18}F -FDG PET SUV/DW-MRI data. Although the complexity of the acquired dataset for each time point was relatively low, the non-trivial aspect of this study was to model the development of various intratumor tissues in a manner that matches the temporal biological characteristics of the tumor as a whole. Therefore, we developed a novel segmentation workflow based on GMMs that begins with the last time point PET/MRI data to obtain a histologically validated mathematical description of the intratumor heterogeneity. Afterwards, for each of the previous time points, the model description was kept fixed and different tumor tissue fractions were identified by obtaining the posterior probabilities of all voxels under the specified model configuration. Moreover, to account for the inter-measurement variabilities, a scale factor was included while fitting the data from all previous time points. As discussed previously (subsection 1.6.2 and 1.6.3),

a GMM makes strong assumptions about the shape of the clusters, which might not be suitable for certain datasets. In this study, however, the assumption of Gaussian clusters was well justified due to the shape of the data distribution at all time points (figure 2.2A, first column). Also, a strong negative intratumor correlation was present between the SUV and ADC in the last two time points. A natural consequence of the high correlation between input features is the amount of redundancy. In our results, we observed that the distinction between the viable and necrotic clusters was achieved mainly based on their ADC values. Nonetheless, SUV was relevant in identifying the viable-2 cluster and achieving an overall higher spatial agreement with the histology (figure 2.2A, second column). It should be noted that the efficacy of this methodology is dependent on the choice of the validation time point. In particular, the tissue heterogeneity in the last time point should be representative of the phenotypic diversity to be modeled in all the previous time points. A clear limitation of this study is the lack of histological validation for the first three time points. Incorporating ground truth information from intermediate measurements might further improve the precision of the suggested approach.

In the next two studies we focused on the multiparametric data acquired from PET and MRI alone and evaluated the feasibility of spectral clustering in segmenting the high dimensional tumor datasets.

In PET, multiple characteristics of the tumor can be obtained by monitoring the temporal distribution of a tracer. Therefore, in the second study, we simulated the ^{18}F -FDG PET TACs of three tumor tissue classes (mature teratoma, viable and necrosis) and reported the benefits of clustering the TACs over parametric maps of kinetic parameters and SUV-based segmentation. Overall, in both clinical and preclinical scenarios, the weighted-spectral clustering applied to noisy TACs provided more accurate results than all the other clustering schemes. The misclassification error of spectral clustering on the estimated kinetic parameters for clinical TACs was initially (at lower levels of noise) the lowest, however, its performance deteriorated rapidly as the noise in the TACs was increased. These

results indicate the limitations of parametric maps based voxel-wise tumor tissue characterization, because as opposed to the TACs obtained after region averaging, the voxel-wise TACs are often corrupted by significant levels of noise. Furthermore, it should be noted that the accuracy of the clustering methods based on the estimated kinetic parameters corresponds to the most optimal case for compartmental modeling, because the rate constants were calculated by using the true AIFs. As shown in figure 2.5 and 2.6, commonly occurring peak amplitude and position distortions in AIFs either underestimate or overestimate the kinetic parameters, which would further worsen the parametric maps based estimation of intratumor heterogeneity. Moreover, aside from peak amplitude and position distortions, AIFs can generally not be acquired noise-free. Up to moderate noise levels, in contrast to the noisy TACs and parametric maps based techniques, *K*-means clustering on the SUV remained least effective in identifying the three tissue classes. These findings clearly signify the benefits of dynamic ^{18}F -FDG PET measurements over the SUV. Although this study lacked the validation of spectral clustering on clinical experimental data, a recent report by Schmitz et al. [114] has shown the feasibility of translating a GMM based approach from preclinical to clinical setting. Overall, the promising results of spectral clustering on preclinical and clinical simulations as well as on the preclinical experimental data confirm its suitability for analyzing dynamic PET tumor images and estimating the intratumor heterogeneity in a voxel-wise manner. In cases where the interpretability of kinetic parameters is more important, spectral clustering can be used in a complementary way with compartmental modeling. In particular, the impact of noise on the estimation of kinetic parameters can be reduced by using spectral clustering as a pre-processing segmentation step and performing kinetic modeling on the mean TACs of different clusters.

In the third study we applied the suggested SRSC algorithm on the multiparametric MRI data of glioblastoma xenografts and compared its ability to segment intratumor tissues with three widely utilized classical clustering algorithms (*K*-means, FCM and a GMM). To validate the phenotypic maps predicted by all four algorithms, we quantified the ground truth fractions of viable, necrotic and peri-

necrotic tissues by delineating regions of interest on the co-registered H&E and GLUT-1 histology images. A linear correlation analysis between the histology and clustering tissue fractions revealed that contrary to all other techniques, only SRSC probability maps reliably predicted the fractional population of all three tissue types. In our experiments the spatial smoothness constraints were imposed by centering a $3 \times 3 \times 3$ window around each voxel and weighting its posterior probability with the probability average of the neighboring voxels. As all five tumors were relatively small in size (the dimensions of the largest tumor being $27 \times 24 \times 29$ voxels), we did not choose a bigger window while applying SRSC. In the correlation analysis some of the data points originated from the same tumor (corresponding to different histology slices from the same tumor), which raises the question whether such analysis is actually justified, because the data points are not independent. To assess whether these points introduce bias, we took their average (for both histology and clustering tissue fractions) and recomputed the Pearson's correlation coefficients for each of the algorithm. This exercise resulted in an increase of the Pearson's correlation coefficient for all four methods indicating that the original analysis was not biased, although it was conservative. Besides the three heterogeneous tumors, the viable, connective and muscle tissues in the two homogeneous tumors were identifiable using all clustering techniques. The findings of this paper indicate that incorporating spatial constraints in a standard clustering algorithm can lead to improved segmentation results and provide more reliable intratumor quantification of imaging biomarkers. Although FCM clustering and a GMM make strong assumptions about the cluster shapes, these techniques can invariably benefit with such modifications, when the underlying assumptions are partially or fully met, for instance in studies where the acquired parameters are known to be highly correlated (figure 2.2A).

In the last paper we drew inspiration from the promising results of the two previous studies to develop a multi-view learning approach that identified therapy-induced changes in the tumor microenvironment by utilizing complementary information from dynamic ^{18}F -FDG PET/multiparametric MRI data. The proposed multi-view learning approach is essentially an unsupervised-supervised learning workflow

consisting of MSC and an unbiased random forest classifier. As Db-scTRAIL therapy induced mass apoptosis in the treated tumors, the multiparametric PET/MRI data of the Control and Th-24 Training set tumors was initially segmented using MSC to obtain the *in vivo* labels of viable, apoptosis and *rest* clusters. Thereafter, three separate models (PET, MRI and PET/MRI) were trained using the voxel-wise MSC labeled combined training set data and an unbiased random forest classifier. The classification results on the Test-1 and Test-2 set tumors confirmed that the PET/MRI classifier outperformed the PET and MRI models by selectively combining the relevant information from each of the modalities and provided an accurate intratumor localization of all three clusters. We subsequently tested the additionally trained 2- and 3-class MRI-only models on the glioblastoma tumors and found that multi-view learning concepts can be transferred from one tumor model to another. The misclassification of viable tissue by 3-class MRI classifier as apoptosis (figure 2.19) was consistent with the classification results of the MRI classifier on the Control Test-1 set tumors (figure 2.18), which demonstrates the prognostic value of the complementary information provided by dynamic PET data for the PET/MRI model. Although not described in section 2.4, we also trained a PET/MRI classifier using just the six predictive features ($\text{PET/MRI}_{\text{pred}}$), shown in figure 2.17D. In contrast to the PET/MRI model trained on the full feature set, the test results of the $\text{PET/MRI}_{\text{pred}}$ classifier were in weak agreement with the apoptotic regions identified on the caspase-3 histology of the test Th-24 and Th-72 tumors. In addition, within the heterogeneous portions of the tumors, the classification maps often lacked uncertainty. Nonetheless, the treatment and control tumors could be stratified using the respective apoptotic and viable tissue fractions for both the PET/MRI models. In training of the unbiased random forest classifiers, $\text{round}(\sqrt{D})$ predictors were sampled for each split, as suggested by Strobl et al. [153]. Here, D denotes the total number of predictors. A color-based segmentation of H&E and caspase-3 histology images using K -means was adequate to obtain the ground truth viable, apoptosis and *rest* clusters. However, this was only applicable for the Control and Th-24 tumors. The Th-72 tumor histology images could not be processed as such, due to extensive tissue heterogeneity of

the tumors. To reduce the redundancy in the dynamic PET data, prior to training the PET and PET/MRI classifiers, a linear fit was performed on the last 7 PET time points and the fit-intercept was used as a surrogate feature. The intercept of these time points reflects the combination of two biological processes (perfusion and the hexokinase driven phosphorylation rate of FDG) in a complex fashion, which are likely to work against one another in deriving its value across heterogeneous tissue regions. An alternative and possibly more generic approach to mitigate this issue would be to use the average of the correlated last PET time frames in training the PET and PET/MRI classifiers.

In conclusion, this thesis systematically presents the advantages of multiparametric and multimodality imaging for the assessment of tumor tissue heterogeneity and therapeutic efficacy. Despite the fact that some of the methods presented in this thesis can not be readily translated into the clinic, due to difficulties in obtaining tumor histology and its alignment with imaging, they might play a pivotal role in improving our understanding of cancer and discovering diagnostically relevant patterns from highly intricate PET/MRI datasets. In the same vein, the phenotypic classifiers trained using the proposed multi-view learning approach have great potential, as they can be applied on different tumor models with minimal supervision.

In future work, we will therefore acquire preclinical PET/MRI datasets using orthotopic tumor models that closely match with an ongoing or culminated clinical trial. The objectives of the planned future studies include the development of histologically validated phenotypic classifiers and their prospective translation on clinical datasets. Furthermore, to obtain reliable ground truth phenotypic maps, greater focus will be placed on the development of multi-stain automated histology analysis techniques.

Summary

Despite a broad understanding that solid tumors exhibit significant tissue heterogeneity, clinical trials have not seen a remarkable development in techniques that aid in characterizing cancer. Needle biopsies often represent only a partial view of the tumor profile, lacking the ability to comprehensively reflect spatio-temporal phenotypic changes. Recent multimodal multiparametric imaging techniques could provide further valuable insights if the complementary imaging information is sufficiently analyzed. Therefore, in this work I developed and applied machine learning methods on multiparametric positron emission tomography (PET) and magnetic resonance imaging (MRI) datasets, acquired using mice bearing subcutaneous tumors, to obtain a precise spatio-temporal characterization of intratumor heterogeneity.

In the first study, I modeled longitudinal development of viable and necrotic tissues in subcutaneous lung cancer using multiparametric PET/MRI data and Gaussian mixture models. The subsequent two studies used spectral clustering and its variants to analyze dynamic PET and multiparametric MRI data, respectively and to quantify phenotypic heterogeneity in tumors. These reports also present the benefits of the suggested algorithms over conventional dynamic PET quantification methods and clustering techniques. In the final study, I proposed a novel multi-view learning framework that selectively utilized complementary information from dynamic PET/multiparametric MRI data to predict early changes in intratumor heterogeneity and monitor therapy efficacy. The results of this investigation provide compelling evidence that the full potential of PET/MRI can be exploited by using *in vivo* imaging and machine learning in tandem, which can positively

Summary

impact the current standards of personalized treatment planning and precision oncology.

Zusammenfassung

Trotz des umfassenden Verständnisses, dass solide Tumore eine signifikante Gewebe-heterogenität aufweisen, haben klinische Studien keine bemerkenswerte Entwicklung in unterstützenden Techniken zur Charakterisierung von Krebs erfahren. Mit Nadelbiopsien können oft nur Teilansichten eines Tumorprofils dargestellt werden, welche die räumlich-zeitlichen phänotypischen Veränderungen nicht umfassend reflektieren. Neuere multimodale und multiparametrische Bildgebungstechniken können weitere wertvolle Erkenntnisse liefern, wenn die komplementären Bildgebungsinformationen ausreichend analysiert werden. Hier setzt diese vorliegende Arbeit an, mit der ich maschinelle Lernverfahren für multiparametrische Positronen-Emissions-Tomographie (PET)- und Magnetresonanztomographie (MRT)-Datensätze entwickelte und implementierte. Die Datensätze wurden von Mäusen mit subkutanen Tumoren gewonnen, um eine genaue räumlich-zeitliche Charakterisierung der intratumoralen Heterogenität zu erhalten.

Im Rahmen der ersten Studie modellierte ich die longitudinale Entwicklung von gesundem und nekrotischem Gewebe bei subkutanem Lungenkrebs mit multiparametrischen PET/MRT-Daten und Gaußschen Mischverteilungsmodellen. Für die nachfolgenden zwei Studien wurden Spektral Clustering und seine Varianten angewandt, um dynamische PET- und multiparametrische MRT-Daten zu analysieren und die phänotypische Heterogenität in Tumoren zu quantifizieren. Die Ergebnisse dieser Studien zeigen die Vorteile der vorgeschlagenen Algorithmen gegenüber konventionellen dynamischen Quantifizierungsmethoden und Clustering-Techniken. In der abschließenden Studie präsentiere ich ein neuar-

tiges Multi-View-Lernverfahren, das selektiv komplementäre Informationen aus dynamischen PET- und multiparametrischen MRI-Daten nutzt, um frühe Veränderungen der intratumoralen Heterogenität vorherzusagen und die Therapiewirksamkeit zu überwachen. Die Ergebnisse dieser Untersuchung zeigen, wie mittels kombinierter *in vivo* Bildgebung und maschinellen Lernens das volle Potenzial der PET/MRT-Bildgebung genutzt werden kann um neue Standards in der personalisierten Behandlungsplanung und Präzisionstherapie zu ermöglichen.

Bibliography

- 1 Sawyers, C. (2004). "Targeted cancer therapy". *Nature* 432.7015, pp. 294–297 (cit. on p. 1).
- 2 Mellman, I., Coukos, G., Dranoff, G. (2011). "Cancer immunotherapy comes of age". *Nature* 480.7378, pp. 480–489 (cit. on p. 1).
- 3 Tang, L., Zhang, X.-P., Sun, Y.-S., Shen, L., Li, J., Qi, L.-P., Cui, Y. (2011). "Gastrointestinal stromal tumors treated with imatinib mesylate: apparent diffusion coefficient in the evaluation of therapy response in patients". *Radiology* 258.3, pp. 729–738 (cit. on p. 1).
- 4 Chiou, V. L., Burotto, M. (2015). "Pseudoprogression and immune-related response in solid tumors". *Journal of Clinical Oncology* 33.31, pp. 3541–3543 (cit. on p. 1).
- 5 Tirkes, T., Hollar, M. A., Tann, M., Kohli, M. D., Akisik, F., Sandrasegaran, K. (2013). "Response criteria in oncologic imaging: review of traditional and new criteria". *RadioGraphics* 33.5, pp. 1323–1341 (cit. on p. 1).
- 6 Bedard, P. L., Hansen, A. R., Ratain, M. J., Siu, L. L. (2013). "Tumour heterogeneity in the clinic". *Nature* 501.7467, pp. 355–364 (cit. on pp. 1, 2).
- 7 Burrell, R. A., McGranahan, N., Bartek, J., Swanton, C. (2013). "The causes and consequences of genetic heterogeneity in cancer evolution". *Nature* 501.7467, pp. 338–345 (cit. on pp. 1, 2).
- 8 Jamal-Hanjani, M., Quezada, S. A., Larkin, J., Swanton, C. (2015). "Translational implications of tumor heterogeneity". *Clinical Cancer Research* 21.6, pp. 1258–1266 (cit. on p. 1).

- 9 Gillies, R. J., Kinahan, P. E., Hricak, H. (2015). “Radiomics: Images Are More than Pictures, They Are Data.” *Radiology* 278.2, p. 151169 (cit. on p. 2).
- 10 Aerts, H. J.W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Cavalho, S., Bussink, J., et al. (2014). “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. *Nature Communications* 5, p. 4006 (cit. on p. 3).
- 11 Itakura, H., Achrol, A. S., Mitchell, L. A., Loya, J. J., Liu, T., Westbroek, E. M., Feroze, A. H., et al. (2015). “Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities”. *Science Translational Medicine* 7.303, 303ra138–303ra138 (cit. on p. 3).
- 12 Stewart, G. D., O’Mahony, F. C., Laird, A., Eory, L., Lubbock, A. L. R., Mackay, A., Nanda, J., et al. (2015). “Sunitinib treatment exacerbates intratumoral heterogeneity in metastatic renal cancer”. *Clinical Cancer Research* 21.18, pp. 4212–4223 (cit. on p. 3).
- 13 Lee, B. S., Cho, Y. W., Kim, G. C., Lee, D. H., Kim, C. J., Kil, H. S., Chi, D. Y., et al. (2015). “Induced phenotype targeted therapy: radiation-induced apoptosis-targeted chemotherapy”. *JNCI: Journal of the National Cancer Institute* 107.2, dju403–dju403 (cit. on p. 3).
- 14 Thomson, J. J. (1897). “XL. Cathode rays”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 44, pp. 293–316 (cit. on p. 4).
- 15 Thomson, J. J. (1904). “XXIV. On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 7.39, pp. 237–265 (cit. on p. 4).
- 16 Rutherford, E. (1911). “LXXIX. The scattering of α and β particles by matter and the structure of the atom”. *The London, Edinburgh, and Dublin Philo-*

- sophical Magazine and Journal of Science* 21.125, pp. 669–688 (cit. on p. 4).
- 17 Beiser, A. (2003). *Concepts of modern physics*. Tata McGraw-Hill Education, New York (cit. on p. 4).
 - 18 Bohr, N. (1913). “LXXIII. On the constitution of atoms and molecules”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 26.155, pp. 857–875 (cit. on p. 4).
 - 19 Dirac, P. A. M. (1928). “The quantum theory of the electron”. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 117.778, pp. 610–624 (cit. on p. 4).
 - 20 Anderson, C. D. (1933). “The positive electron”. *Physical Review* 43.6, pp. 491–494 (cit. on pp. 4, 5).
 - 21 Rich, D. A. (1997). “A brief history of positron emission tomography”. *Journal of Nuclear Medicine Technology* 25.1, pp. 4–11 (cit. on pp. 4, 5).
 - 22 Tobias, C. A., Lawrence, J. H., Roughton, F. J. W., Root, W. S., Gregersen, M. I. (1945). “The elimination of carbon monoxide from the human body with reference to the possible conversion of CO to CO₂”. *American Journal of Physiology – Legacy Content* 145.2, pp. 253–263 (cit. on p. 4).
 - 23 Wrenn, F. R., Good, M. L., Handler, P. (1951). “The use of positron-emitting radioisotopes for the localization of brain tumors”. *Science* 113.2940, pp. 525–527 (cit. on p. 5).
 - 24 Sweet, W. H. (1951). “The uses of nuclear disintegration in the diagnosis and treatment of brain tumor”. *New England Journal of Medicine* 245.23, pp. 875–878 (cit. on p. 5).
 - 25 Nutt, R. (2002). “The history of positron emission tomography”. *Molecular Imaging & Biology* 4.1, pp. 11–26 (cit. on p. 5).
 - 26 Bailey, D. L., Townsend, D. W., Valk, P. E., Maisey, M. N. (2005). *Positron emission tomography*. Springer, Berlin (cit. on pp. 6–10, 13, 17, 82).
 - 27 Bushberg, J. T., Boone, J. M. (2011). *The essential physics of medical imaging*. Lippincott Williams & Wilkins, Philadelphia (cit. on pp. 6, 10, 11, 20, 21, 23, 24, 27–29).

- 28 Cherry, S. R., Dahlbom, M. (2006). *PET: physics, instrumentation, and scanners*. Springer, New York (cit. on pp. 6, 7, 11, 12, 14).
- 29 Turkington, T. G. (2001). "Introduction to PET instrumentation". *Journal of Nuclear Medicine Technology* 29.1, pp. 4–11 (cit. on pp. 12, 13).
- 30 Dempster, A., Laird, N., Rubin, D. (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1, pp. 1–38 (cit. on p. 14).
- 31 Hudson, H. M., Larkin, R. S. (1994). "Accelerated image reconstruction using ordered subsets of projection data". *IEEE Transactions on Medical Imaging* 13.4, pp. 601–609 (cit. on p. 14).
- 32 Juweid, M. E., Cheson, B. D. (2006). "Positron-emission tomography and assessment of cancer therapy". *New England Journal of Medicine* 354.5, pp. 496–507 (cit. on p. 15).
- 33 Ruth, T. J. (2009). "The uses of radiotracers in the life sciences". *Reports on Progress in Physics* 72.1, pp. 016701–016724 (cit. on p. 15).
- 34 Gambhir, S. S. (2002). "Molecular imaging of cancer with positron emission tomography". *Nature Reviews Cancer* 2.9, pp. 683–693 (cit. on p. 15).
- 35 Huang, S. C. (2000). "Anatomy of SUV". *Nuclear Medicine and Biology* 27.7, pp. 643–646 (cit. on pp. 15, 16, 81).
- 36 Thie, J. A. (2004). "Understanding the standardized uptake value, its methods, and implications for usage." *Journal of Nuclear Medicine* 45.9, pp. 1431–4 (cit. on pp. 15, 16, 81).
- 37 Adams, M. C., Turkington, T. G., Wilson, J. M., Wong, T. Z. (2010). "A systematic review of the factors affecting accuracy of SUV measurements". *American Journal of Roentgenology* 195.2, pp. 310–320 (cit. on pp. 16, 57, 81).
- 38 Carson, R., Cunningham, V., Gunn, R., Hoff, J. van den, Knudsen, G., Lamertsmas, A., Leenders, K., et al. (2003). "PET pharmacokinetic course manual". *University of Groningen, Groningen* (cit. on p. 16).

- 39 Rabi, I., Zacharias, J., Millman, S., Kusch, P. (1938). "A new method of measuring nuclear magnetic moment". *Physical Review* 53.4, p. 318 (cit. on p. 17).
- 40 Gerlach, W., Stern, O. (1922). "Der experimentelle Nachweis des magnetischen Moments des Silberatoms". *Zeitschrift für Physik* 8.1, pp. 110–111 (cit. on p. 18).
- 41 Purcell, E. M., Torrey, H. C., Pound, R. V. (1946). "Resonance absorption by nuclear magnetic moments in a solid". *Physical Review* 69.1-2, pp. 37–38 (cit. on p. 18).
- 42 Bloch, F., Hansen, W. W., Packard, M. (1946). "The nuclear induction experiment". *Physical Review* 70.7-8, pp. 474–485 (cit. on p. 18).
- 43 Hahn, E. L. (1950). "Spin echoes". *Physical Review* 80.4, pp. 580–594 (cit. on p. 18).
- 44 Ernst, R. R., Anderson, W. A. (1966). "Application of fourier transform spectroscopy to magnetic resonance". *Review of Scientific Instruments* 37.1, pp. 93–102 (cit. on p. 18).
- 45 Damadian, R. (1971). "Tumor detection by nuclear magnetic resonance". *Science* 171.3976, pp. 1151–1153 (cit. on p. 18).
- 46 Lauterbur, P. C. (1973). "Image formation by induced local interactions: examples employing nuclear magnetic resonance". *Nature* 242.5394, pp. 190–191 (cit. on pp. 18, 19).
- 47 Mansfield, P. (1977). "Multi-Planar image formation using NMR spin echoes". *Journal of Physics C: Solid State Physics* 10, pp. L55–L58 (cit. on p. 18).
- 48 Ai, T., Morelli, J. N., Hu, X., Hao, D., Goerner, F. L., Ager, B., Runge, V. M. (2012). "A historical overview of magnetic resonance imaging, focusing on technological innovations." *Investigative Radiology* 47.12, pp. 725–741 (cit. on p. 18).
- 49 Runge, V. M., Nitz, W. R., Trelles, M., Goerner, F. L. (2014). *The physics of clinical MR taught through images*. Georg Thieme Verlag, Stuttgart (cit. on p. 26).

- 50 Lustig, M., Donoho, D., Pauly, J. M. (2007). "Sparse MRI: The application of compressed sensing for rapid MR imaging". *Magnetic Resonance in Medicine* 58.6, pp. 1182–1195 (cit. on p. 29).
- 51 Charles-Edwards, E. M., Desouza, N. M. (2006). "Diffusion-weighted magnetic resonance imaging and its application to cancer". *Cancer Imaging* 6, pp. 135–143 (cit. on p. 30).
- 52 Le Bihan, D. (2013). "Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure." *Radiology* 268.2, pp. 318–22 (cit. on p. 30).
- 53 Le Bihan, D. (2012). "Diffusion, confusion and functional MRI". *NeuroImage* 62.2, pp. 1131–1136 (cit. on p. 30).
- 54 Burdette, J. H., Durden, D. D., Elster, A. D., Yen, Y.-F. (2001). "High b-Value diffusion-weighted MRI of normal brain". *Journal of Computer Assisted Tomography* 25.4, pp. 515–519 (cit. on p. 31).
- 55 Pichler, B. J., Judenhofer, M. S., Wehrl, H. F. (2008). "PET/MRI hybrid imaging: devices and initial results". *European Radiology* 18.6, pp. 1077–1086 (cit. on pp. 31, 32).
- 56 Beyer, T., Townsend, D. W., Brun, T., Kinahan, P. E., Charron, M, Roddy, R, Jerin, J, et al. (2000). "A combined PET/CT scanner for clinical oncology." *Journal of Nuclear Medicine* 41.8, pp. 1369–79 (cit. on p. 31).
- 57 Bar-Shalom, R., Yefremov, N., Guralnik, L., Gaitini, D., Frenkel, A., Kuten, A., Altman, H., et al. (2003). "Clinical performance of PET/CT in evaluation of cancer: additional value for diagnostic imaging and patient management." *Journal of Nuclear Medicine* 44.8, pp. 1200–9 (cit. on p. 31).
- 58 Czernin, J., Allen-Auerbach, M., Schelbert, H. R. (2007). "Improvements in cancer staging with PET/CT: literature-based evidence as of September 2006." *Journal of Nuclear Medicine* 48 Suppl 1.1, 78S–88S (cit. on p. 31).
- 59 Brindle, K. (2008). "New approaches for imaging tumour responses to treatment". *Nature Reviews Cancer* 8.2, pp. 94–107 (cit. on p. 31).
- 60 Pfannenber, A. C., Aschoff, P., Brechtel, K., Müller, M, Klein, M., Bares, R., Claussen, C. D., et al. (2007). "Value of contrast-enhanced multiphase CT

- in combined PET/CT protocols for oncological imaging". *The British Journal of Radiology* 80.954, pp. 437–445 (cit. on p. 31).
- 61** Brix, G., Lechel, U., Glatting, G., Ziegler, S. I., Münzing, W., Müller, S. P., Beyer, T. (2005). "Radiation exposure of patients undergoing whole-body dual-modality 18F-FDG PET/CT examinations." *Journal of Nuclear Medicine* 46.4, pp. 608–13 (cit. on p. 31).
- 62** Müller-Horvat, C., Radny, P., Eigentler, T. K., Schäfer, J., Pfannenberger, C., Horger, M., Khorchidi, S., et al. (2006). "Prospective comparison of the impact on treatment decisions of whole-body magnetic resonance imaging and computed tomography in patients with metastatic malignant melanoma". *European Journal of Cancer* 42.3, pp. 342–350 (cit. on p. 31).
- 63** Johnson, L. M., Turkbey, B., Figg, W. D., Choyke, P. L. (2014). "Multiparametric MRI in prostate cancer management". *Nature Reviews Clinical Oncology* 11.6, pp. 346–353 (cit. on pp. 32, 106).
- 64** Judenhofer, M. S., Wehrl, H. F., Newport, D. F., Catana, C., Siegel, S. B., Becker, M., Thielscher, A., et al. (2008). "Simultaneous PET-MRI: a new approach for functional and morphological imaging". *Nature Medicine* 14.4, pp. 459–465 (cit. on p. 32).
- 65** Disselhorst, J. A., Bezrukov, I., Kolb, A., Parl, C., Pichler, B. J. (2014). "Principles of PET/MR imaging". *Journal of Nuclear Medicine* 55.Supplement_2, 2S–10S (cit. on pp. 32–34).
- 66** Shao, Y., Cherry, S. R., Farahani, K., Meadors, K., Siegel, S., Silverman, R. W., Marsden, P. K. (1997). "Simultaneous PET and MR imaging". *Physics in Medicine and Biology* 42.10, pp. 1965–1970 (cit. on p. 32).
- 67** Pichler, B. J., Judenhofer, M. S., Catana, C., Walton, J. H., Kneilling, M., Nutt, R. E., Siegel, S. B., et al. (2006). "Performance test of an LSO-APD detector in a 7-T MRI scanner for simultaneous PET/MRI." *Journal of Nuclear Medicine* 47.4, pp. 639–47 (cit. on p. 32).
- 68** Wu, Y., Catana, C., Farrell, R., Dokhale, P. A., Shah, K. S., Qi, J., Cherry, S. R. (2009). "PET performance evaluation of an MR-compatible PET in-

- sert". *IEEE Transactions on Nuclear Science* 56.3, pp. 574–580 (cit. on p. 32).
- 69** Wehner, J., Weissler, B., Dueppenbecker, P., Gebhardt, P., Schug, D., Ruetten, W., Kiessling, F., et al. (2014). "PET/MRI insert using digital SiPMs: Investigation of MR-compatibility". *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 734.PART B, pp. 116–121 (cit. on p. 32).
- 70** Kolb, A., Lorenz, E., Judenhofer, M. S., Renker, D., Lankes, K., Pichler, B. J. (2010). "Evaluation of Geiger-mode APDs for PET block detector designs". *Physics in Medicine and Biology* 55.7, pp. 1815–1832 (cit. on p. 33).
- 71** Sander, C. Y., Keil, B., Chonde, D. B., Rosen, B. R., Catana, C., Wald, L. L. (2015). "A 31-channel MR brain array coil compatible with positron emission tomography". *Magnetic Resonance in Medicine* 73.6, pp. 2363–2375 (cit. on p. 33).
- 72** Bezrukov, I., Schmidt, H., Mantlik, F., Schwenzer, N., Brendle, C., Scholkopf, B., Pichler, B. J. (2013). "MR-based attenuation correction methods for improved PET quantification in lesions within bone and susceptibility artifact regions". *Journal of Nuclear Medicine* 54.10, pp. 1768–1774 (cit. on p. 33).
- 73** An, H. J., Seo, S., Kang, H., Choi, H., Cheon, G. J., Kim, H.-J., Lee, D. S., et al. (2016). "MRI-based attenuation correction for PET/MRI using multiphase level-set method". *Journal of Nuclear Medicine* 57.4, pp. 587–593 (cit. on p. 33).
- 74** Ehrhardt, M. J., Thielemans, K., Pizarro, L., Atkinson, D., Ourselin, S., Hutton, B. F., Arridge, S. R. (2015). "Joint reconstruction of PET-MRI by exploiting structural similarity". *Inverse Problems* 31.1, p. 015001 (cit. on p. 33).
- 75** Chun, S. Y., Reese, T. G., Ouyang, J., Guerin, B., Catana, C., Zhu, X., Alpert, N. M., et al. (2012). "MRI-based nonrigid motion correction in simultaneous PET/MRI". *Journal of Nuclear Medicine* 53.8, pp. 1284–1291 (cit. on p. 33).

-
- 76** Martinez-Moller, A., Souvatzoglou, M., Delso, G., Bundschuh, R. A., Chefd'hotel, C., Ziegler, S. I., Navab, N., et al. (2009). "Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data". *Journal of Nuclear Medicine* 50.4, pp. 520–526 (cit. on p. 33).
- 77** Fraum, T. J., Fowler, K. J., McConathy, J. (2016). "PET/MRI: emerging clinical applications in oncology". *Academic Radiology* 23.2, pp. 220–236 (cit. on p. 34).
- 78** Ledezma, C. J., Chen, W., Sai, V., Freitas, B., Cloughesy, T., Czernin, J., Pope, W. (2009). "18F-FDOPA PET/MRI fusion in patients with primary/recurrent gliomas: Initial experience". *European Journal of Radiology* 71.2, pp. 242–248 (cit. on p. 34).
- 79** Metz, S., Ganter, C., Lorenzen, S., Marwick, S. van, Holzapfel, K., Herrmann, K., Rummeny, E. J., et al. (2015). "Multiparametric MR and PET imaging of intratumoral biological heterogeneity in patients with metastatic lung cancer using voxel-by-voxel analysis". *PLoS ONE* 10.7, pp. 1–14 (cit. on p. 34).
- 80** Judenhofer, M. S., Cherry, S. R. (2013). "Applications for preclinical PET/MRI". *Seminars in Nuclear Medicine* 43.1, pp. 19–29 (cit. on p. 34).
- 81** James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An introduction to statistical learning*. Springer, New York (cit. on pp. 35, 38, 47–50).
- 82** Ghahramani, Z. (2004). "Unsupervised learning". *Advanced Lectures on Machine Learning*. Springer, Berlin (cit. on p. 35).
- 83** Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, New York (cit. on pp. 36, 39–43).
- 84** Lloyd, S. (1982). "Least squares quantization in PCM". *IEEE Transactions on Information Theory* 28.2, pp. 129–137 (cit. on p. 36).
- 85** Jain, A. K. (2010). "Data clustering: 50 years beyond K-means". *Pattern Recognition Letters* 31.8, pp. 651–666 (cit. on p. 38).

- 86 Mao, J., Jain, A. K. (1996). "A self-organizing network for hyperellipsoidal clustering (HEC)". *IEEE Transactions on Neural Networks* 7.1, pp. 16–29 (cit. on p. 38).
- 87 Dunn, J. C. (1973). "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters". *Journal of Cybernetics* 3.3, pp. 32–57 (cit. on p. 39).
- 88 Prince, S. J. D. (2012). *Computer vision: models learning and inference*. Cambridge University Press, New York (cit. on pp. 40, 41).
- 89 Luxburg, U. von (2007). "A tutorial on spectral clustering". *Statistics and Computing* 17.4, pp. 395–416 (cit. on pp. 44–46).
- 90 Ng, A. Y., Jordan, M. I., Weiss, Y. (2001). "On spectral clustering: Analysis and an algorithm". *Advances in Neural Information Processing Systems* 14, pp. 849–856 (cit. on p. 44).
- 91 Hastie, T., Tibshirani, R., Friedman, J. (2009). *The elements of statistical learning*. Springer, New York (cit. on pp. 44, 47–50).
- 92 Kwok, S. W., Carter, C. (1990). "Multiple decision trees". *Machine Intelligence and Pattern Recognition*. Vol. 9. C, pp. 327–335 (cit. on p. 47).
- 93 Ho, T. K. (1995). "Random decision forests". *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. IEEE, pp. 278–282 (cit. on p. 47).
- 94 Amit, Y., Geman, D., Wilder, K. (1997). "Joint induction of shape features and tree classifiers". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.11, pp. 1300–1305 (cit. on p. 47).
- 95 Ho, T. K. (1998). "The random subspace method for constructing decision forests". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.8, pp. 832–844 (cit. on p. 47).
- 96 Breiman, L. (2001). "Random forests". *Machine Learning* 45.1, pp. 5–32 (cit. on pp. 47, 50).
- 97 Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems". *Annals of Eugenics* 7.2, pp. 179–188 (cit. on p. 50).

-
- 98** Guyon, I., Elisseeff, A. (2003). “An introduction to variable and feature selection”. *Journal of Machine Learning Research* 3.3, pp. 1157–1182 (cit. on p. 51).
- 99** He, X., Cai, D., Niyogi, P. (2005). “Laplacian score for feature selection”. *Advances in Neural Information Processing Systems* 186, p. 189 (cit. on pp. 51, 52, 85).
- 100** Duda, R. O., Hart, P. E., Stork, D. G. (2001). *Pattern classification*. Wiley, New York (cit. on pp. 52, 134).
- 101** Divine, M. R., Katiyar, P., Kohlhofer, U., Quintanilla-Martinez, L., Pichler, B. J., Disselhorst, J. A. (2016). “A population-based gaussian mixture model incorporating 18F-FDG PET and diffusion-weighted MRI quantifies tumor tissue classes”. *Journal of Nuclear Medicine* 57.3, pp. 473–479 (cit. on pp. 53, 56, 59, 131).
- 102** Katiyar, P., Divine, M. R., Kohlhofer, U., Quintanilla-Martinez, L., Schölkopf, B., Pichler, B. J., Disselhorst, J. A. (2017). “Spectral clustering predicts tumor tissue heterogeneity using dynamic 18 F-FDG PET: a complement to the standard compartmental modeling approach”. *Journal of Nuclear Medicine* 58.4, pp. 651–657 (cit. on pp. 53, 84, 89, 132).
- 103** Katiyar, P., Divine, M. R., Kohlhofer, U., Quintanilla-Martinez, L., Schölkopf, B., Pichler, B. J., Disselhorst, J. A. (2017). “A novel unsupervised segmentation approach quantifies tumor tissue populations using multiparametric MRI: first results with histological validation”. *Molecular Imaging and Biology* 19.3, pp. 391–397 (cit. on pp. 53, 110–112, 132, 139).
- 104** Rakheja, R., Chandarana, H., DeMello, L., Jackson, K., Geppert, C., Faul, D., Glielmi, C., et al. (2013). “Correlation between standardized uptake value and apparent diffusion coefficient of neoplastic lesions evaluated with whole-body simultaneous hybrid PET/MRI”. *American Journal of Roentgenology* 201.5, pp. 1115–1119 (cit. on p. 55).
- 105** Grueneisen, J., Beiderwellen, K., Heusch, P., Buderath, P., Aktas, B., Gratz, M., Forsting, M., et al. (2014). “Correlation of standardized uptake value and

- apparent diffusion coefficient in integrated whole-body PET/MRI of primary and recurrent cervical cancer”. *PLoS ONE* 9.5, pp. 1–7 (cit. on p. 55).
- 106** Yu, X., Lee, E. Y. P., Lai, V., Chan, Q. (2014). “Correlation between tissue metabolism and cellularity assessed by standardized uptake value and apparent diffusion coefficient in peritoneal metastasis”. *Journal of Magnetic Resonance Imaging* 40.1, pp. 99–105 (cit. on p. 55).
- 107** Shih, I. L., Yen, R. F., Chen, C. A., Chen, B. B., Wei, S. Y., Chang, W. C., Sheu, B. C., et al. (2015). “Standardized uptake value and apparent diffusion coefficient of endometrial cancer evaluated with integrated whole-body PET/MR: Correlation with pathological prognostic factors”. *Journal of Magnetic Resonance Imaging* 42.6, pp. 1723–1732 (cit. on p. 55).
- 108** Higashi, K., Clavo, A. C., Wahl, R. L. (1993). “Does FDG uptake measure proliferative activity of human cancer cells? In vitro comparison with DNA flow cytometry and tritiated thymidine uptake.” *Journal of Nuclear Medicine* 34.3, pp. 414–419 (cit. on p. 55).
- 109** Brown, R. S., Leung, J. Y., Fisher, S. J., Frey, K. A., Ethier, S. P., Wahl, R. L. (1996). “Intratumoral distribution of tritiated-FDG in breast carcinoma: correlation between Glut-1 expression and FDG uptake.” *Journal of Nuclear Medicine* 37.6, pp. 1042–7 (cit. on p. 55).
- 110** McDonald, D. M., Choyke, P. L. (2003). “Imaging of angiogenesis: from microscope to clinic”. *Nature Medicine* 9.6, pp. 713–725 (cit. on p. 55).
- 111** Chen, L., Zhang, J., Chen, Y., Wang, W., Zhou, X., Yan, X., Wang, J. (2014). “Relationship between apparent diffusion coefficient and tumour cellularity in lung cancer”. *PLoS ONE* 9.6, pp. 1–7 (cit. on p. 55).
- 112** Eidel, O., Neumann, J. O., Burth, S., Kieslich, P. J., Jungk, C., Sahm, F., Kickingeder, P., et al. (2016). “Automatic analysis of cellularity in glioblastoma and correlation with ADC using trajectory analysis and automatic nuclei counting”. *PLoS ONE* 11.7, pp. 1–10 (cit. on p. 55).
- 113** Schmidt, H., Brendle, C., Schraml, C., Martirosian, P., Bezrukov, I., Hetzel, J., Müller, M., et al. (2013). “Correlation of simultaneously acquired diffusion-weighted imaging and 2-Deoxy-[18F] fluoro-2-D-glucose positron

- emission tomography of pulmonary lesions in a dedicated whole-body magnetic resonance/positron emission tomography system". *Investigative Radiology* 48.5, pp. 247–255 (cit. on pp. 55, 131).
- 114** Schmitz, J., Schwab, J., Schwenck, J., Chen, Q., Quintanilla-Martinez, L., Hahn, M., Wietek, B., et al. (2016). "Decoding intratumoral heterogeneity of breast cancer by multiparametric in vivo imaging: a translational study". *Cancer Research* 76.18, pp. 5512–5522 (cit. on pp. 55, 131, 209).
- 115** Sequist, L. V., Waltman, B. A., Dias-Santagata, D., Digumarthy, S., Turke, A. B., Fidias, P., Bergethon, K., et al. (2011). "Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors". *Science Translational Medicine* 3.75, 75ra26–75ra26 (cit. on p. 58).
- 116** Visser, E. P., Philippens, M. E., Kienhorst, L., Kaanders, J. H., Corstens, F. H., Geus-Oei, L.-F. de, Oyen, W. J. (2008). "Comparison of tumor volumes derived from glucose metabolic rate maps and SUV maps in dynamic 18F-FDG PET". *Journal of Nuclear Medicine* 49.6, pp. 892–898 (cit. on p. 81).
- 117** Cheebsumon, P., Velden, F. H. van, Yaqub, M., Hoekstra, C. J., Velasquez, L. M., Hayes, W., Hoekstra, O. S., et al. (2011). "Measurement of metabolic tumor volume: static versus dynamic FDG scans". *EJNMMI Research* 1.1, p. 35 (cit. on p. 81).
- 118** Muzi, M., O'Sullivan, F., Mankoff, D. A., Doot, R. K., Pierce, L. A., Kurland, B. F., Linden, H. M., et al. (2012). "Quantitative assessment of dynamic PET imaging data in cancer imaging". *Magnetic Resonance Imaging* 30.9, pp. 1203–1215 (cit. on p. 82).
- 119** Muzi, M., Vesselle, H., Grierson, J. R., Mankoff, D. A., Schmidt, R. A., Peterson, L., Wells, J. M., et al. (2005). "Kinetic analysis of 3'-deoxy-3'-fluorothymidine PET studies: validation studies in patients with lung cancer." *Journal of Nuclear Medicine* 46.2, pp. 274–82 (cit. on p. 82).
- 120** Dunnwald, L. K., Doot, R. K., Specht, J. M., Gralow, J. R., Ellis, G. K., Livingston, R. B., Linden, H. M., et al. (2011). "PET tumor metabolism in locally advanced breast cancer patients undergoing neoadjuvant chemother-

- apy: value of static versus kinetic measures of fluorodeoxyglucose uptake”. *Clinical Cancer Research* 17.8, pp. 2400–2409 (cit. on p. 82).
- 121** Kristian, A., Nilsen, L. B., Røe, K., Revheim, M.-E., Engebråten, O., Mælandsmo, G. M., Holm, R., et al. (2013). “Dynamic 18 F-FDG PET for assessment of tumor physiology in two breast carcinoma xenografts”. *Nuclear Medicine and Molecular Imaging* 47.3, pp. 173–180 (cit. on p. 82).
- 122** Muzic, R. F., Cornelius, S (2001). “COMKAT: compartment model kinetic analysis tool.” *Journal of Nuclear Medicine* 42.4, pp. 636–45 (cit. on p. 82).
- 123** Sugawara, Y., Zasadny, K. R., Grossman, H. B., Francis, I. R., Clarke, M. F., Wahl, R. L. (1999). “Germ cell tumor: differentiation of viable tumor, mature teratoma, and necrotic tissue with FDG PET and kinetic modeling”. *Radiology* 211.1, pp. 249–56 (cit. on p. 82).
- 124** Vriens, D., Geus-Oei, L.-F. de, Oyen, W. J., Visser, E. P. (2009). “A curve-fitting approach to estimate the arterial plasma input function for the assessment of glucose metabolic rate and response to treatment”. *Journal of Nuclear Medicine* 50.12, pp. 1933–1939 (cit. on p. 82).
- 125** Ferl, G. Z., Zhang, X, Wu, H.-M., Huang, S.-C. (2007). “Estimation of the 18F-FDG input function in mice by use of dynamic small-animal PET and minimal blood sample data”. *Journal of Nuclear Medicine* 48.12, pp. 2037–2045 (cit. on p. 83).
- 126** Logan, J., Fowler, J. S., Volkow, N. D., Ding, Y. S., Wang, G.-J., Alexoff, D. L. (2001). “A strategy for removing the bias in the graphical analysis method”. *Journal of Cerebral Blood Flow & Metabolism* 21.3, pp. 307–320 (cit. on p. 83).
- 127** Jovkar, S, Evans, A. C., Diksic, M, Nakai, H, Yamamoto, Y. L. (1989). “Minimisation of parameter estimation errors in dynamic PET: choice of scanning schedules”. *Physics in Medicine and Biology* 34.7, pp. 895–908 (cit. on p. 85).
- 128** Wang, W., Georgi, J.-C., Nehmeh, S. A., Narayanan, M., Paulus, T., Bal, M., O’Donoghue, J., et al. (2009). “Evaluation of a compartmental model

- for estimating tumor hypoxia via FMISO dynamic PET imaging”. *Physics in Medicine and Biology* 54.10, pp. 3083–3099 (cit. on p. 85).
- 129** Weissleder, R., Pittet, M. J. (2008). “Imaging in the era of molecular oncology”. *Nature* 452.7187, pp. 580–589 (cit. on p. 106).
- 130** Fass, L. (2008). “Imaging and cancer: a review”. *Molecular Oncology* 2.2, pp. 115–152 (cit. on p. 106).
- 131** Jaffray, D. A. (2012). “Image-guided radiotherapy: from current concept to future perspectives”. *Nature Reviews Clinical Oncology* 9.12, pp. 688–699 (cit. on p. 106).
- 132** Edmund, J. M., Nyholm, T. (2017). “A review of substitute CT generation for MRI-only radiation therapy”. *Radiation Oncology* 12.1, p. 28 (cit. on p. 106).
- 133** Yankeelov, T. E., Abramson, R. G., Quarles, C. C. (2014). “Quantitative multimodality imaging in cancer research and therapy”. *Nature Reviews Clinical Oncology* 11.11, pp. 670–680 (cit. on p. 106).
- 134** Lemaître, G., Martí, R., Freixenet, J., Vilanova, J. C., Walker, P. M., Meriaudeau, F. (2015). “Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review”. *Computers in Biology and Medicine* 60, pp. 8–31 (cit. on p. 106).
- 135** Bauer, S., Wiest, R., Nolte, L.-P., Reyes, M. (2013). “A survey of MRI-based medical image analysis for brain tumor studies”. *Physics in Medicine and Biology* 58.13, R97–R129 (cit. on p. 106).
- 136** Carano, R. A., Ross, A. L., Ross, J., Williams, S. P., Koeppen, H., Schwall, R. H., Van Bruggen, N. (2004). “Quantification of tumor tissue populations by multispectral analysis”. *Magnetic Resonance in Medicine* 51.3, pp. 542–551 (cit. on p. 106).
- 137** Henning, E. C., Azuma, C., Sotak, C. H., Helmer, K. G. (2007). “Multispectral quantification of tissue types in a RIF-1 tumor model with histological validation. Part I”. *Magnetic Resonance in Medicine* 57.3, pp. 501–512 (cit. on p. 106).
- 138** Berry, L. R., Barck, K. H., Go, M. A., Ross, J., Wu, X., Williams, S. P., Gogineni, A., et al. (2008). “Quantification of viable tumor microvascular charac-

- teristics by multispectral analysis". *Magnetic Resonance in Medicine* 60.1, pp. 64–72 (cit. on p. 106).
- 139** Hectors, S. J.C. G., Jacobs, I., Strijkers, G. J., Nicolay, K. (2014). "Multiparametric MRI analysis for the identification of high intensity focused ultrasound-treated tumor tissue". *PLoS ONE* 9.6, pp. 1–11 (cit. on p. 106).
- 140** Han, S. H., Ackerstaff, E., Stoyanova, R., Carlin, S., Huang, W., Koutcher, J. A., Kim, J. K., et al. (2013). "Gaussian mixture model-based classification of dynamic contrast enhanced MRI data for identifying diverse tumor microenvironments: preliminary results". *NMR in Biomedicine* 26.5, pp. 519–532 (cit. on p. 106).
- 141** Juan-Albarracín, J., Fuster-Garcia, E., Manjón, J. V., Robles, M., Aparici, F., Martí-Bonmatí, L., García-Gómez, J. M. (2015). "Automated glioblastoma segmentation based on a multiparametric structured unsupervised classification." *PLoS ONE* 10.5, pp. 1–20 (cit. on p. 106).
- 142** Fathi Kazerooni, A., Mohseni, M., Rezaei, S., Bakhshandehpour, G., Saligheh Rad, H. (2015). "Multi-parametric (ADC/PWI/T2-w) image fusion approach for accurate semi-automatic segmentation of tumorous regions in glioblastoma multiforme". *Magnetic Resonance Materials in Physics, Biology and Medicine* 28.1, pp. 13–22 (cit. on p. 106).
- 143** Denko, N. C. (2008). "Hypoxia, HIF1 and glucose metabolism in the solid tumour". *Nature Reviews Cancer* 8.9, pp. 705–713 (cit. on p. 112).
- 144** Gillies, R. J., Schomack, P. A., Secomb, T. W., Raghunand, N. (1999). "Causes and effects of heterogeneous perfusion in tumors". *Neoplasia* 1.3, pp. 197–207 (cit. on p. 112).
- 145** Bailey, D. L., Antoch, G., Bartenstein, P., Barthel, H., Beer, A. J., Bisdas, S., Bluemke, D. A., et al. (2015). "Combined PET/MR: the real work has just started. Summary report of the third international workshop on PET/MR imaging; February 17–21, 2014, Tübingen, Germany". *Molecular Imaging and Biology* 17.3, pp. 297–312 (cit. on p. 131).
- 146** Siegemund, M., Pollak, N., Seifert, O., Wahl, K., Hanak, K., Vogel, A., Nusler, A. K., et al. (2012). "Superior antitumoral activity of dimerized targeted

- single-chain TRAIL fusion proteins under retention of tumor selectivity". *Cell Death and Disease* 3.4, e295 (cit. on pp. 131, 132).
- 147** Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T. (2007). "Bias in random forest variable importance measures: illustrations, sources and a solution". *BMC Bioinformatics* 8.1, p. 25 (cit. on p. 132).
- 148** Dice, L. R. (1945). "Measures of the amount of ecologic association between species". *Ecology* 26.3, pp. 297–302 (cit. on p. 133).
- 149** Dubuisson, M.-P., Jain, A. (1994). "A modified Hausdorff distance for object matching". *Proceedings of 12th International Conference on Pattern Recognition*. Vol. 1, pp. 566–568 (cit. on p. 133).
- 150** Nguyen, Q., Valizadegan, H., Hauskrecht, M. (2014). "Learning classification models with soft-label information". *Journal of the American Medical Informatics Association* 21.3, pp. 501–508 (cit. on p. 142).
- 151** Gillies, R. J., Beyer, T. (2016). "PET and MRI: is the whole greater than the sum of its parts?" *Cancer Research* 76.21, pp. 6163–6166 (cit. on p. 142).
- 152** Kwak, J. T., Sankineni, S., Xu, S., Turkbey, B., Choyke, P. L., Pinto, P. A., Moreno, V., et al. (2017). "Prostate cancer: a correlative study of multiparametric MR imaging and digital histopathology". *Radiology*, p. 160906 (cit. on p. 142).
- 153** Strobl, C., Hothorn, T., Zeileis, A. (2009). "Party on!" *R Journal* 1.2, pp. 14–17 (cit. on p. 211).

Declaration of Originality

General

The work conducted in this thesis was performed at the department of Preclinical Imaging and Radiopharmacy under the supervision of Prof. Dr. Bernd Pichler.

I, Prateek Katiyar, herewith declare that this dissertation is my own original work, except where indicated through the proper use of citations and references. I used no sources or aids other than those explicitly mentioned. Support provided by other people is clearly marked and acknowledged as such in this dissertation.

Within each project included in this thesis, my work primarily focused on the development and application of mathematical and computational algorithms. My contribution in the first project and the contribution of all co-authors in the remaining projects are detailed below.

Project-1

A Population-Based Gaussian Mixture Model Incorporating 18F-FDG PET and Diffusion-Weighted MRI Quantifies Tumor Tissue Classes, *The Journal of Nuclear Medicine* (2016).

Mathew R. Divine, Prateek Katiyar, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Bernd J. Pichler, and Jonathan A. Disselhorst

Development and programming of the machine learning framework by Prateek Katiyar under the supervision of Mathew R. Divine and Jonathan A. Disselhorst. Production of the Figures 2, 5, and 6 by Prateek Katiyar in cooperation with

Mathew R. Divine and Jonathan A. Disselhorst. Writing of the clustering workflow by Prateek Katiyar and editing of the manuscript by Prateek Katiyar (and all the other authors).

Project-2

Spectral Clustering Predicts Tumor Tissue Heterogeneity Using Dynamic 18F-FDG PET: A Complement to the Standard Compartmental Modeling Approach, *The Journal of Nuclear Medicine* (2017).

Prateek Katiyar, Mathew R. Divine, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Bernhard Schölkopf, Bernd J. Pichler, and Jonathan A. Disselhorst

Study conception and design by Prateek Katiyar in cooperation with Mathew R. Divine, Bernhard Schölkopf, Bernd J. Pichler and Jonathan A. Disselhorst. Algorithm development, data analysis and presentation of results by Prateek Katiyar. Preclinical imaging experiments performed by Mathew R. Divine. Radiotracer production by the radiopharmacy team. Histopathology work performed by Ursula Kohlhofer and Leticia Quintanilla-Martinez. Preparation of the manuscript by Prateek Katiyar in cooperation with Jonathan A. Disselhorst. Manuscript editing by all the authors.

Project-3

A Novel Unsupervised Segmentation Approach Quantifies Tumor Tissue Populations Using Multiparametric MRI: First Results with Histological Validation, *Molecular Imaging and Biology* (2017).

Prateek Katiyar, Mathew R. Divine, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Bernhard Schölkopf, Bernd J. Pichler, and Jonathan A. Disselhorst

Study conception and design by Prateek Katiyar in cooperation with Mathew R. Divine, Bernhard Schölkopf, Bernd J. Pichler and Jonathan A. Disselhorst. Algorithm development, data analysis and presentation of results by Prateek Katiyar. Imaging experiments performed by Jonathan A. Disselhorst and Mathew R. Di-

vine. Histopathology work performed by Ursula Kohlhofer and Leticia Quintanilla-Martinez. Histology tumor delineation and alignment with multiparametric MRI data by Prateek Katiyar. Preparation of the manuscript by Prateek Katiyar in cooperation with Jonathan A. Disselhorst. Manuscript editing by all the authors.

Project-4

Multi-view learning of multiparametric PET/MRI data quantifies intratumor heterogeneity

Prateek Katiyar, Mathew R. Divine, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Martin Siegemund, Roland Kontermann, Bernhard Schölkopf, Bernd J. Pichler, Jonathan A. Disselhorst

Study conception and design by Prateek Katiyar in cooperation with Mathew R. Divine, Bernhard Schölkopf, Bernd J. Pichler and Jonathan A. Disselhorst. Algorithm development, data analysis and presentation of results by Prateek Katiyar. Imaging experiments performed by Jonathan A. Disselhorst and Mathew R. Divine. PET/MRI tumor delineation by Mathew R. Divine. Histology tumor delineation and alignment with PET/MRI data by Prateek Katiyar. Radiotracer production by the radiopharmacy team. Histopathology work performed by Ursula Kohlhofer and Leticia Quintanilla-Martinez. Db-scTRAIL therapy provided by Martin Siegemund and Roland Kontermann. Preparation of the manuscript by Prateek Katiyar in cooperation with Jonathan A. Disselhorst. Manuscript editing by all the authors.

Signature

Date

Acknowledgements

The past four years of my doctoral research at the Werner Siemens Imaging Center and Max Planck Institute for Intelligent Systems has bestowed me with incredibly rewarding experiences, both professionally and personally. This dissertation is therefore an outcome of not just my own work, but also the efforts of many remarkable and inspiring people, whom I wish to extend my sincere gratitude.

First of all, I am extremely thankful to my doctoral advisor Prof. Dr. Bernd Pichler, director of the Werner Siemens Imaging Center, for providing me with complete autonomy ever since I joined the laboratory. His vision, support and proactive attitude towards the importance of machine learning methods for medical imaging studies have been a constant source of encouragement in my everyday research. Thanks to his diligent mentorship that I had the opportunity to attend several national and international conferences and obtain feedback about my work from a broad community. Most importantly, his experience and meticulous research practices have greatly improved the work presented in this thesis.

I am also very thankful to Prof. Dr. Bernhard Schölkopf, director of the Empirical Inference department at the Max Planck Institute for Intelligent Systems, for kindly accepting me as a visiting PhD student and giving me a highly stimulating environment to learn directly from machine learning experts. I could not have come this far without the knowledge I acquired from the talks given at the department and during the machine learning summer schools as well as the weekly pixel meetings. I am further grateful to Prof. Schölkopf for being a member of my thesis advisory committee and guiding me in the right direction during the course of my PhD.

I was fortunate to have Prof. Dr. med. Konstantin Nikolaou and Prof. Dr. med. Christian la Fougère as the other two members of my thesis advisory committee. Their expert advices and support regarding the potential clinical translation of the methods suggested in this thesis have been extremely valuable.

Although, I was guided by experts from several different fields, the person who shaped this thesis most is my immediate supervisor, Dr. Jonathan Disselhorst. Above all, I am most thankful to Jonathan for helping me improve from an inexperienced researcher towards more of an independent thinker. Whether it was a new research idea, scientific writing exercise or any technical difficulty, my discussions with Jonathan always left me with more clarity and sensible solutions than what I had before. I owe him my deepest gratitude for investing countless hours and being so observant with my work. Also, many thanks to him for his unbiased feedbacks and useful advices in both professional and personal matters; they made me feel very confident and free in expressing my opinions. Lastly, I want to thank Jonathan for being a good friend and a wonderful collaborator over the past couple of years.

Another colleague and a good friend, who significantly influenced the research directions of this thesis is Dr. Mathew Divine. I am grateful to Mathew for actively taking part in almost all imaging experiments included in this thesis and persistently presenting me with intricate biological questions that could be investigated using machine learning. I also want to thank him for his rigorous comments on all my research papers and teaching me somewhat unorthodox ways of writing scientific communications. I thoroughly enjoyed this collaboration and our thought-provoking discussions.

I cannot thank Prof. Dr. Leticia Quintanilla-Fend and Dr. Ursula Kohlhofer enough for their indispensable support with the histopathology work. A big thanks to them for always being approachable and patiently clarifying my not-so-clever questions about tumor biology. Likewise, my sincere thanks to the radiopharmacy team for their continuous support with tracer production.

Thanks to Christoph Parl, Marie-Aline Neveu, Paul Soubiran and Francesco An-

sideri for reading parts of my thesis and giving useful suggestions. Also, many thanks to Neele Hübner and Kristin Patzwaldt for helping me with German translations.

My acknowledgements would be incomplete without extending a gratifying note to the past and present colleagues at the Werner Siemens Imaging Center. Thank you very much for all the good times and making me feel at home. I am grateful to have worked with so many clever and dedicated people throughout my doctoral period. Equally, I am happy to have made some wonderful friends, whose presence made this journey much more enjoyable for me.

A very special thanks to my previous flatmates and friends from Tübingen: Julien, Gosia, Kushal, Joana, Diana and Francesco. Thank you for always being there as my second family and sharing not only the moments of joy, but also great anxiety.

Lastly, a heartfelt thanks to my family for their constant support and an unshakable belief in my scholarly abilities.

Prateek Katiyar
Tübingen, January 2018