A Monte Carlo simulation framework for histology-informed diffusion MRI cancer characterisation and microstructural parameter estimation

Athanasios Grigoriou[1,2], Carlos Macarro[1,2], Marco Palombo[3,4], Anna Voronova[1,2], Kinga Bernatowicz[1], Ignasi Barba[1,5], Alba Escriche[1], Emanuela Greco[1], María Abad[1], Sara Simonetti[1], Garazi Serna[1], Richard Mast[6,7], Xavier Merino[6,7], Núria Roson[6,7], Manuel Escobar[6,7], Maria Vieito[8], Paolo Nuciforo[1], Rodrigo Toledo[1], Elena Garralda[8], Roser Sala-Llonch[9,10], Els Fieremans[11], Dmitry S. Novikov[11], Raquel Perez-Lopez[1,*,**], and Francesco Grussu[1,*,**]

[1] Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain
[2] Department of Biomedicine, Faculty of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain
[3] Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, United Kingdom
[4] School of Computer Science and Informatics, Cardiff University, Cardiff, UK
[5] University of Vic - Central University of Catalonia (UVic-UCC), Vic, Spain
[6] Institut de Diagnostic per la Imatge (IDI), Barcelona, Spain
[7] Department of Radiology, Hospital Universitari Vall d'Hebron, Barcelona, Spain
[8] Medical Oncology Service, Vall d'Hebron Barcelona Hospital Campus, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain
[9] Department of Biomedicine, Faculty of Medicine, Institute of Neurosciences, Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain
[10] Centro de Investigación Biomedica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Barcelona, Spain
[11] Bernard and Irene Schwartz Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine, New York, NY, USA
[*] *joint corresponding authors: fgrussu@vhio.net, rperez@vhio.net*
[**] *joint senior (last) authors*

Abstract

Computer simulations within substrates that mimic the complexity of biological tissues are key to the development of biophysical diffusion Magnetic Resonance Imaging (dMRI) models. Realistic simulations have enabled, for example, the non-invasive estimation of fine neuronal sub-structures, which is playing an increasingly key role in neurology and neuroscience. However, biologically-realistic, simulation-informed dMRI techniques are also needed in other applications, as for example in oncological imaging of body tumours. This article aims to fill this gap by presenting a Monte Carlo (MC) framework tailored for histology-informed simulations in body imaging applications. The framework, which combines free software with custom-written routines, is demonstrated on substrates reconstructed from hematoxylin-eosin (HE) stains of human liver biopsies, including non-cancerous liver and primary/metastatic liver cancer tissues. The article has four main contributions. Firstly, it provides practical guidelines on how to conduct realistic MC diffusion simulations informed by HE histology. Secondly, it reports reference values on cell size (CS), cell density and on other cellular properties in non-cancerous and cancerous liver — information not easily found in the literature, yet essential to inform the design of innovative dMRI techniques. Thirdly, it presents a detailed characterisation of synthetic signals generated for clinically feasible dMRI protocols, shedding light onto patterns of intra-/extra-cellular (IC/EC) water diffusion in liver. Finally, it illustrates the utility of the framework, by devising a strategy where synthetic signals inform the estimation of unexplored microstructural properties, as the EC intrinsic diffusivity and CS distribution skewness. The strategy is demonstrated on actual dMRI scans, acquired on *ex vivo* mouse tissue and in cancer patients *in vivo*.

1 Introduction

The ultimate aim of diffusion MRI (dMRI) is the estimation of statistics of the cellular environment — referred to as *tissue microstructure* — from sets of diffusion-weighted (DW) signal measurements, by solving an inverse mathematical problem [Novikov et al., 2019, Kiselev, 2017]. Multi-compartment biophysical dMRI models have gained momentum as practical approaches for this task, being capable of providing maps of biologically-meaningful properties, such as cell size (CS) indices. These have found applications in multiple organs, e.g., brain [Veraart et al., 2020], muscles [Fieremans et al., 2017], breast [Xu et al., 2020], liver [Jiang et al., 2020b], prostate [Panagiotaki et al., 2015, Lemberskiy et al., 2018] and beyond. Non-invasive CS measurement may be particularly relevant for disease characterisation and treatment response assessment in oncology, given the variety of cell types that can coexist within tumours, each featuring unique, distinctive dimensions (e.g., normal vs malignant cells, immune cell infiltration, etc) [Reynaud, 2017, Jiang et al., 2020a, Hoffmann et al., 2023, Palombo et al., 2023].

However, current biophysical models are often based on idealised representations of tissue components, such as spheres of fixed radii to describe cells [Panagiotaki et al., 2015, Jiang et al., 2020b, Hoffmann et al., 2023]. This implies that they may neglect other, relevant features of intra-voxel microstructure, e.g., the existence of distributions of CSs, intracellular (IC) kurtosis [Lee et al., 2020a, Grussu et al., 2022], or extra-cellular (EC) diffusion time dependence [Burcaw et al., 2015, Xu et al., 2023]. Neglecting such characteristics may bias parameter estimation, and may also cause clinically-relevant information to be missed.

Recently, numerical methods based on more realistic tissue representations have enabled the development of accurate dMRI signal models [Nedjati-Gilani et al., 2017, Lee et al., 2021],

increasing the biological specificity of parameter estimation [Nilsson et al., 2010, Palombo et al., 2016, Palombo et al., 2019, Buizza et al., 2021, Morelli et al., 2023]. In particular, Monte Carlo (MC) diffusion simulations within 3D meshes derived from histology have enabled the characterisation of fine, sub-cellular microstructural details, such as axonal beading/undulation [Lee et al., 2020b, Lee et al., 2024], or neural process complexity [Palombo et al., 2016, Palombo et al., 2019]. Nevertheless, to date histology-informed dMRI has focussed heavily on neural tissue, with only a few examples outside the central nervous system [Berry et al., 2018]. More accurate biophysical models are urgently needed in a variety of other contexts beyond brain dMRI, as in oncological body imaging of solid tumours [O'connor et al., 2017]. New dMRI approaches could tackle several, unmet clinical needs, such as patient stratification for treatment selection, response assessment in immunotherapy [Pilard et al., 2021], or the determination of the malignity of lesions that cannot be biopsied [Doblas et al., 2013].

This article aims to fill this gap by introducing a histology-informed Monte Carlo (MC) dMRI simulation framework tailored for microstructural characterisation of body tissues. We present a rich database of meshed cellular environments reconstructed directly from hematoxylin and eosin (HE) stains of liver biopsies, and use these to synthesise signals for clinically feasible dMRI protocols. The database provides the community with reference values of key cellular properties in cancerous and non-cancerous tissues, and enables the synthesis of realistic dMRI signals. This set of cellular-level characteristics allowed us to gain insight into IC and EC diffusion, and to devise a strategy for the numerical estimation of microstructural properties that are not commonly accessible to analytical biophysical models. In particular, we tested the estimation of the intrinsic EC diffusivity and of CS distribution moments, which we showcase in pre-clinical scans of fixed mouse tissues and in cancer patients *in vivo*.

2 Materials and methods

2.1 Simulation framework

In our framework, illustrated in Fig. 1, we create 3D meshes of histological structures, such as cells, from segmentations drawn on histological images. These meshes can be used to generate random walks in MC simulations and, finally, dMRI signals, for any dMRI protocol of interest. We proceed as follows.

First, a histological image is opened with QuPath [Bankhead et al., 2017] and a region-of-interest (ROI) is selected and cropped, taking care to include in the image the scale of magnification. The image is then opened in Inkscape, where cells and other geometric features are manually segmented and separated into layers. We segmented cells and cell debris, luminal spaces, and vessels. Here we demonstrate the framework with careful, manual segmentation, but automatic segmentations would also be possible. Two types of files are then exported: a 3D object with all the features included included in a single SVG file, as well as an individual SVG file for each feature. The SVG format is used as it allows for further manipulation with Blender. In Blender, SVG files are then transformed into 3D ASCII PLY triangular meshes. In this article, we reconstructed 2D cellular environments from standard HE biopsies and obtained 3D meshes by simply replicating 2D contours along the trough-plane direction, thus generating cylinders with irregular sections. Nonetheless, 3D segmentations could also be used (e.g., from 3D confocal microscopy).

Meshes are then fed to the MCDC Simulator, an open-source MC engine [Rafael-Patino et al., 2020], in order to synthesise random walks. We used the large mesh including all the segmented structures to generate EC random walks, by seeding walkers outside all such structures, uniformly. Conversely, we generated IC random walks for each cell independently. For this first proof-of-concept we did not model permeability of barriers, i.e., we did not account for exchange between EC and IC spaces, with walkers experiencing elastic collisions at cell boundaries.

We used 10 linearly-spaced values in the range [0.8, 3] $\mu m^2/ms$ for both IC and EC intrinsic diffusivities (referred to as $D_{0|in}$ and $D_{0|ex}$), covering all possible combinations of the two (100 unique ($D_{0|in}$, $D_{0|ex}$) pairs for each substrate). Each simulation was run with a duration of $T = 140\ ms$ and step number of $N_{step} = 3000$. The IC simulations were performed with 1000 water molecules for each cell, while EC random walks were obtained for 10 000 walkers. As mentioned, vessel structures were included in the segmentation as they influence the patterns of EC diffusion, but they were not seeded with walkers (so they do not directly generate MRI signal). Regarding lumen spaces, spin walks were generated similarly to the IC case: spins were seeded within each lumen independently, and no crossing of the lumen membrane was possible (elastic reflection in presence of collisions; 1000 walkers per lumen).

Lastly, custom-written python code was used to synthesise dMRI signals from the random walks for a given acquisition protocol of interest. The signal s_n for the n-th structure is [Fieremans and Lee, 2018]

$$S_n = \left| \frac{1}{W_n} \sum_{w=1}^{W_n} e^{-j\,\gamma\,\Delta t\,\sum_{t=0}^{TE} \mathbf{g}(t)^{\mathrm{T}}\,\mathbf{r}_{n,w}(t)} \right|. \tag{1}$$

Above, $\mathbf{r}_{n,w}(t)$ is the w-th walker trajectory within the n-th structure; $\Delta t = T/N_{step}$ is the temporal resolution; T is the simulation duration; and $\mathbf{g}(t)$ is the diffusion-encoding gradient. The IC/EC signals S_{in}/S_{ex} were respectively obtained as

$$S_{in} = \sum_{n \in IC} \frac{V_n}{\sum_{k \in IC} V_k}\, S_n \tag{2}$$

and

$$S_{ex} = \sum_{n \in EC} \frac{V_n}{\sum_{k \in EC} V_k}\, S_n. \tag{3}$$

with V_n/V_k standing for the volume of the n-th/k-th structure. The index n loops over cells (including debris) in Eq. 2, referring to the IC signal; likewise, it loops over the set of EC structures in Eq. 3, i.e., the EC space itself and the luminal spaces. Last, the total signal is

$$S = f_{in}\, S_{in} + (1 - f_{in})\, S_{ex}, \tag{4}$$

with f_{in} being the IC volume fraction. Note that signal fractions f_{in} and $f_{ex} = 1 - f_{in}$ are T_2-weighted in principle, given that the intra-/extra-cellular spaces may feature different T_2 constants [Lemberskiy et al., 2018, Palombo et al., 2023]. Nonetheless, in this first demonstration of our MC framework, we do not account for intra-compartment relaxation properties, in order to reduce the number of tissue parameters required to characterise the signal.

A repository with step-by-step guidelines on how to implement the framework is released at https://github.com/radiomicsgroup/dMRIMC.

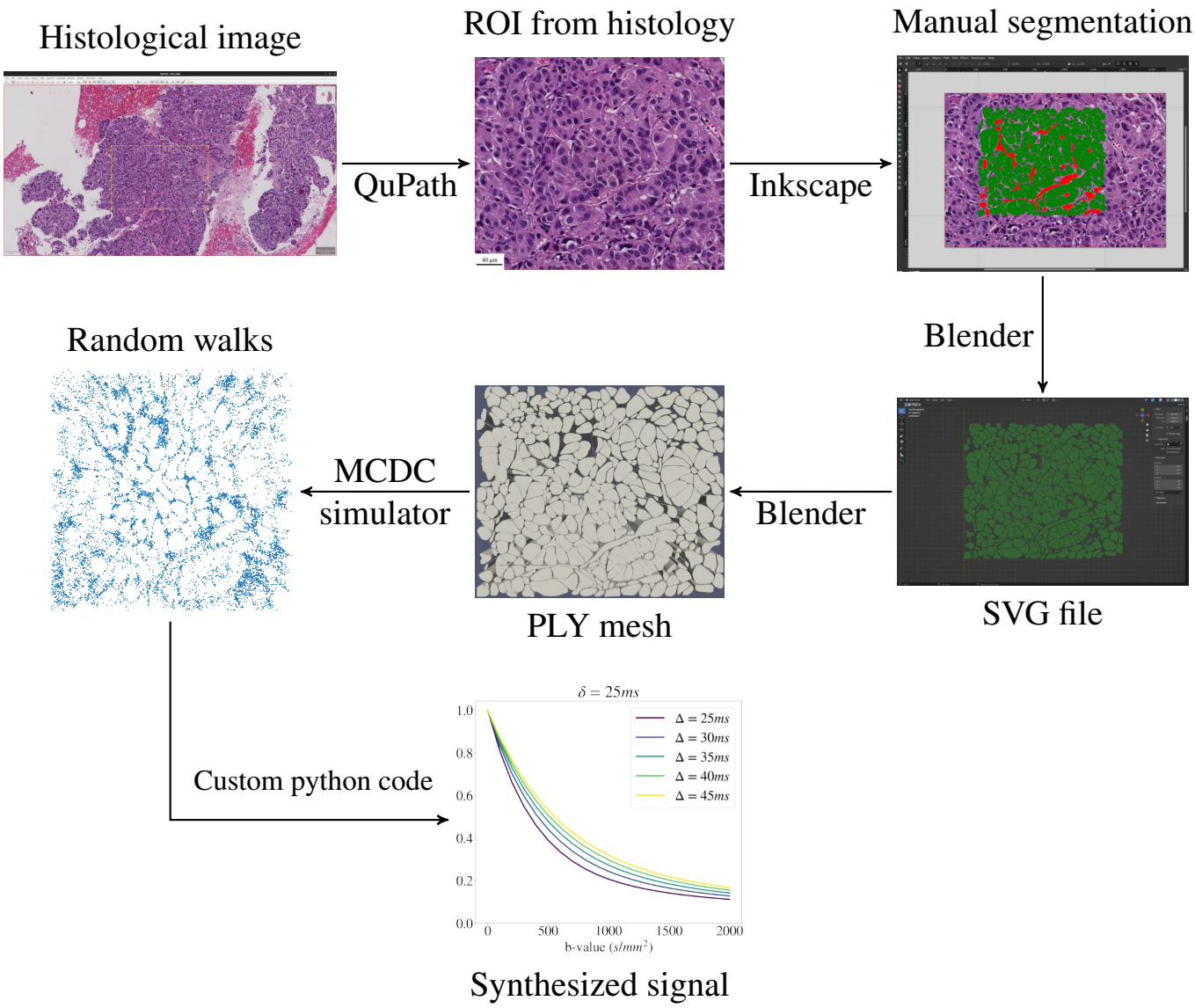


Figure 1: Illustration of our MC simulation framework generating synthetic dMRI signals from histological images. The framework relies on the following open-source software packages: QuPath, Inkscape, Blender.

2.2 Substrates

We reconstructed 18 cellular environments, referred to as *substrates* from now on. These were derived from biopsies of malignant solid tumours of the liver (primary cancer and mestastatic) of 10 different patients (1 to 3 substrates drawn per patient, see Table 1), acquired as part of ongoing imaging studies at the Vall d'Hebron Institute of Oncology (Barcelona, Spain). The substrates spanned a rich set of different cytoarchitectures, from non-cancerous liver parenchyma to cancer areas, such as dense cancer cell packings, fibrosis, necrosis, and a mix of all the above.

We characterised each substrate with the following microstructural parameters:

- ROI area and cellularity (number of cells per mm^2 of biopsied tissue);
- IC area fraction f_{in};
- lumen fraction of EC area f_l;
- lumen diameters $d_{lumen} = 2\sqrt{A_{lumen}/\pi}$, with A_{lumen} being the segmented lumen area;
- mean CS index mCS $= \langle d_{cell} \rangle$, where $d_{cell} = 2\sqrt{A_{cell}/\pi}$ is the individual cell diameter calculated from its area A_{cell}, and $\langle \dots \rangle$ is the average over the distribution in a substrate;
- CS variance index varCS $= \langle (d_{cell} - \text{mCS})^2 \rangle$;
- CS skewness index

$$\text{skewCS} = \langle (d_{cell} - \text{mCS})^3 \rangle / \text{varCS}^{3/2};$$

- volume-weighted CS (vCS) index for a system with spherical geometry [Novikov et al., 2019, Grussu et al., 2022] (vCS_{sph}), defined as

$$\text{vCS}_{sph} = \left(\frac{\langle d_{cell}^7 \rangle}{\langle d_{cell}^3 \rangle} \right)^{1/4};$$

- vCS index for a system with cylindrical geometry [Burcaw et al., 2015, Veraart et al., 2020] (vCS_{cyl}), defined as

$$\text{vCS}_{cyl} = \left(\frac{\langle d_{cell}^6 \rangle}{\langle d_{cell}^2 \rangle} \right)^{1/4};$$

- shape and scale parameters of a gamma-distribution [Assaf et al., 2008], fitted to the set of cell diameters $\{d_{cell,1}, d_{cell,2}, ...\}$ (see Supplementary Methods for more details).

2.3 Characterisation of IC and EC diffusion

We used the substrates to generate synthetic dMRI signals for routine pulsed-gradient spin echo (PGSE) acquisitions (Supplementary Figure S1). We considered diffusion times (e.g., gradient duration δ and separation Δ) that could be realistically probed in pre-clinical and clinical imaging. Signals were processed with the aim of gaining insight into the biological factors that drive the diffusion contrast in both IC and EC spaces, in order to inform biophysical signal modelling.

To this end, we studied the diffusion-time dependence of the IC and EC apparent diffusion coefficients (ADCs) ADC_{in} and ADC_{ex}. In all substrates, we estimated ADC_{in} and ADC_{ex} for all possible pairs of δ and Δ in the set $(\delta, \Delta) \in \{5, 15, 25, 35\}\ ms \times \{5, 15, 25, 35, 45, 55, 65\}\ ms$, s.t. $\Delta \geq \delta$. The estimation of ADC_{in} and ADC_{ex} was performed by regressing $\ln(S_{in/ex}) = -b\,\text{ADC}_{in/ex}$ with respect to $b = \{0, 100, 200, 300, 400, 500\}\ s/mm^2$ (with averaging of two in-plane, orthogonal gradient directions for each b). For this, we used increasing values of the intrinsic IC/EC diffusivities $D_{0|in}$ and $D_{0|ex}$ (both ranging in $\{0.8, 1.5, 2.2, 3.0\}\ \mu m^2/ms$ for any fixed (δ, Δ)).

We used the set of synthetic ADC_{in} and ADC_{ex} to characterise IC and EC diffusion, framing the observed patterns in the context of common assumptions used in biophysical dMRI models. Specifically:

1. we compared values of ADC_{in} and ADC_{ex} at varying (δ, Δ), to test the parameter constraint $\text{ADC}_{in} < \text{ADC}_{ex}$ used in body dMRI techniques such as Restriction Spectrum Imaging [White et al., 2014, Feng et al., 2021, Conlin et al., 2023];

2. we scattered ADC_{ex} against f_{in} and assessed the dependence of ADC_{ex} on Δ, in order to test whether the mean-field tortuosity limit of ADC_{ex}, i.e.,

$$\text{ADC}_{ex} = (1 - f_{in})\, D_{0|ex} \tag{5}$$

   for cylindrical geometries [Szafer et al., 1995, Zhang et al., 2012], offer a good approximation of ADC_{ex} at clinical diffusion times;

3. finally, we also assessed the dependence of ADC_{in} on salient features of the underlying CS distribution, given the urgent need for non-invasive cell profiling in diseases such as cancer (i.e., to differentiate lymphocytes from malignant cells [Jiang et al., 2020a, Hoffmann et al., 2023].). We specifically tested whether salient features of CS distribution, namely its moments mCS, varCS and skewCS, are encoded in the intra-cellular dMRI signal, by scattering ADC_{in} against each of those at varying Δ.

2.4 Simulation-informed microstructural parameter inference

We investigated the potential utility of our synthetic signals to inform microstructural parameter estimation. We used the MC random walks within our histological substrates to synthesise signals for realistic dMRI protocols, and devised a strategy enabling histological property estimation from noisy dMRI signal measurements. The strategy was tested *in silico*, and then on *ex vivo* and *in vivo* dMRI scans.

2.4.1 *In silico* data

We synthesised DW signals according to three dMRI protocols, matching those implemented for *in vivo* and *ex vivo* MRI acquisition. We simulated 100 signals per substrate (10 values of $D_{0|in} \times 10$ values of $D_{0|ex}$), for a total of 1800 signals. The protocols were:

- a PGSE protocol, matching that implemented on a 3T clinical system *in vivo* (referred to as: PGSE-in). It consisted of 3 $b = 0$ and 18 DW measurements, namely: $b = \{50, 100, 400, 900, 1200, 1500, 50, 100, 400, 900, 1200, 1500, 50, 100, 400, 900, 1200, 1500\}\ s/mm^2$, $\delta = \{3.9, 5.2, 9.2, 15.0, 18.2, 21.0, 3.9, 5.2, 9.2, 13.0, 15.8, 18.5, 3.9, 5.2, 9.2, 13.0, 15.8, 18.5\}\ ms$, $\Delta = \{27.8, 29.0, 33.0, 28.7, 31.8, 34.7, 7.8, 29.0, 33.0, 37.0, 39.6, 42.3, 7.8, 29.0, 33.0, 37.0, 39.6, 42.3\}\ ms$.

- A DW twice-refocussed spin echo (TRSE) protocol, matching that implemented on a 1.5T clinical system *in vivo* (referred to as: TRSE). It consisted of 3 $b = 0$ and 18 DW measurements, namely: $b = \{0, 50, 100, 400, 900, 1200, 1600\}\ s/mm^2$, repeated for 3 different diffusion times. The duration/separation of the gradient lobes (Supplementary Figure 1) for the 3 diffusion times were: $\delta_1 = \{8.9, 13.2, 18.9\}\ ms$, $\delta_2 = \{17.6, 19.3, 21.0\}\ ms$, $\delta_3 = \{20.4, 24.8, 30.5\}\ ms$, $\delta_4 = \{6.0, 7.7, 9.5\}\ ms$, $\Delta_{1,2} = \{17.4, 21.7, 27.5\}\ ms$, $\Delta_{1,4} = \{63.9, 74.2, 87.5\}\ ms$.

- A second PGSE protocol, matching that implemented on a pre-clinical 9.4T system for *ex vivo* imaging (referred to as: PGSE-ex). It consisted of 2 $b = 0$ and 6 DW measurements, namely: $b = \{0, 500, 2000, 4500\}\ s/mm^2$ acquired for each of $\Delta = \{16.5, 37.0\}\ ms$, with $\delta = 12\ ms$.

For all protocols, we obtained the final measurement set by averaging signals generated for two orthogonal directions, perpendicular to the substrate longitudinal axis.

We then interpolated the set of paired examples of tissue parameters $\mathbf{p}$ and dMRI signals $s(\mathbf{p})$ with a radial basis function (RBF) regressor, which implements the forward model $\mathbf{p} \mapsto s(\mathbf{p})$. This was finally embedded into routine non-linear least square (NNLS) fitting, based on maximum-likelihood estimation [Panagiotaki et al., 2012]. To test the feasibility of using simulation-informed forward models for parameter estimation, we performed a leave-one-out experiment. Briefly, for all substrates in turn, we learnt $\mathbf{p} \mapsto s(\mathbf{p})$ on noise-free signals from 17/18 substrates, and then plugged the learnt model in NNLS fitting of the noisy signals from the 18th substrate (Rician noise; $b = 0$ signal-to-noise ratio (SNR) of 50).

We performed fitting twice, considering two different forward models $\mathbf{p} \mapsto s(\mathbf{p})$:

- in *forward model 1*,

$$\mathbf{p} = \left\{ f_{in}, D_{0|in}, vCS_{cyl}, D_{0|ex} \right\},$$

  estimating a single CS statistic (vCS_{cyl}) per voxel. We chose to estimate vCS_{cyl}, rather than mCS, to enable the comparison of model 1 to fitting a well-established multi-compartment analytical signal model (see below);

- in *forward model 2* instead,

$$\mathbf{p} = \left\{ f_{in}, D_{0|in}, \text{mCS}, \text{varCS}, \text{skewCS}, D_{0|ex} \right\},$$

estimating the first three moments of the full CS distribution.

The quality of parameter estimation was assessed by scatter density plots and Pearson's correlation coefficients between estimated and ground truth parameter values. Moreover, fitting of forward model 1 was compared to a widely-used analytical model , describing the dMRI signal as the sum of IC/EC contributions, i.e.,

$$S = f_{in}\, e^{-b\,\text{ADC}_{in}(D_{0|in},R,\delta,\Delta)} + (1 - f_{in})\, e^{-b\,\text{ADC}_{ex}}, \quad (6)$$

where ADC_{in} characterises signal decay due to restriction within cells. This approach is used, for example, in popular techniques such as VERDICT and IMPULSED [Panagiotaki et al., 2015, Jiang et al., 2020b]. However, while VERDICT and IMPULSED ADC_{in} is based on a model of spherical cells, here we used the expression for diffusion within cylinders, given the cylindrical symmetry of our substrates. We used an effective radius $R = \frac{1}{2}\text{vCS}_{cyl}$. We used vCS_{cyl} rather than mCS since the former accounts for the fact that larger cells contain more water, and hence contribute more to the DW signal, than smaller cells [Veraart et al., 2020], despite being a metric prone to mesoscopic fluctuations, being highly sensitive to the tails of the cell size distribution within a voxel, with increasing sensitivity the smaller the voxel gets [Novikov et al., 2019]. In practice, ADC_{in} in Eq. 6 is written as

$$\text{ADC}_{in} = \frac{2\,R^4}{D_{0|in}\,\delta\,(\Delta - \delta/3)} \sum_{m=1}^{\infty} \frac{2\,a_m^2 + \frac{R^2}{D_{0|in}\,\delta}\,(u_m - 2)}{a_m^6(a_m^2 - 1)}, \quad (7)$$

where

$$u_m = 2\,e^{-\frac{D_{0|in}}{R^2}\delta a_m^2} + 2\,e^{-\frac{D_{0|in}}{R^2}\Delta a_m^2} - e^{-\frac{D_{0|in}}{R^2}(\Delta-\delta)a_m^2} - e^{-\frac{D_{0|in}}{R^2}(\Delta+\delta)a_m^2}. \quad (8)$$

Above, $D_{0|in}$ is the intra-cylinder diffusivity, a_m is the m-th root of $J_1'(x) = 0$, with $J_1(x)$ being the Bessel function of the first kind, order 1, and $J_1'(x) = \frac{d}{dx}J_1(x)$ [P. Vangelderen et al., 1994]. Note that the analytical model in Eq. 7, while common in dMRI literature, lacks higher-order terms in each compartment. It is therefore only a crude approximation of the signal at a given b, δ and Δ — this fact is indeed what motivates our efforts to build numerical dMRI signal models informed directly by simulations, which do not rely on approximated analytical signal expressions as the one above.

Both MC simulation-informed and analytical signal model fitting were performed with the freely-available *bodymritools* python tools (`https://github.com/fragrussu/bodymritools`; scripts *mri2micro_dictml.py* and *pgse2cylperpinex.py*).

2.4.2 *Ex vivo* data

We tested simulation-informed fitting on pre-clinical dMRI data, which were acquired on 8 formalin-fixed *ex vivo* mouse tissue specimens. These were: a non-cancerous breast and 3 breast tumours from the mouse mammary tumour virus (MMTV) polyomavirus middle T antigen (PyMT) transgenic mouse model [Guy et al., 1992, Attalla et al., 2021], obtained at weeks 9, 11 and 14; a normal spleen and a spleen suffering from splenomegaly, secondary to advanced breast cancer in one MMTV mouse; two kidneys from C57BL/6 WT male mice (9 weeks old), one normal and one pathological, following folic acid-induced injury [Yan, 2021]. Details on the animal procedures and animal models can be found in the Supplementary Methods.

Briefly, tissue was washed in phosphate-buffered saline (PBS) solution and embedded in agarose within a histological cassette. Specimens were scanned at room temperature on a 9.4T Bruker Avance system, with 200 mT/m gradient insert and a RX/TX birdcage coil. The protocol included a high resolution anatomical T2-weighted RARE scan, and dMRI (DW spin echo), with the protocol matching the PGSE-ex protocol described above (section 2.4.1; same nominal b-values, and same gradient timings). Other salient dMRI scan parameters were: resolution $0.2 \times 0.2 \times 0.57\ mm^3$, TE = 55.1 ms, TR = 2250 ms, 3 mutually-orthogonal direction for each gradient timing and b-value. After MRI, standard histological procedures provided sections at known radiographic position, with material sliced following the same orientation of the MRI slices. The material was stained with HE, and high-resolution digital images were acquired on a Hamamatsu C9600-12 scanner (resolution: 0.45 μm).

dMRI scans were denoised [Veraart et al., 2016] and Gibbs ringing mitigated [Kellner et al., 2016]. Finally, maps from forward model 2 in section 2.4.1 were computed voxel-by-voxel, via NNLS regularised maximum-likelihood fitting. Metrics were: f_{in}, $D_{0|in}$, mCS, varCS, skewCS, $D_{0|ex}$. When learning the forward signal model via RBF regression, we pooled together all 1800 signals from all substrates.

In parallel, we also processed the HE images to obtain histological counterparts of f_{in}, mCS, varCS, skewCS. We manually segmented cells on 9 ROIs across the whole sets of samples. Afterwards, we computed f_{in}, mCS, varCS, skewCS given the set of segmented cells, as illustrated in section 2.2. Fitting was not performed in areas rich of fat as seen on HE images, given that dMRI acquisitions are typically fat-suppressed. We also extracted mean and standard deviation of dMRI metrics in 10 ROIs of size of 3 voxel × 3 voxel. These were drawn in areas approximately matching the locations of the histological ROIs. Finally, we used the set of 9 paired dMRI-histological metrics to compute a Pearson's correlation coefficient.

2.4.3 *In vivo* data

Lastly, we also tested simulation-informed parameter estimation in scans from cancer patients, acquired *in vivo* for an ongoing imaging study. All participants provided informed written consent, and were scanned in imaging sessions approved by the Clinical Research Ethics Committee (CEIm) of the Vall d'Hebron University Hospital (Barcelona, Spain; code: PR(AG)29/2020)).

We studied scans from 9 patients suffering from advanced solid tumours, which were acquired either at abdominal or pelvic level (referred to as Cases A to I). 4 patients were scanned on a 1.5T Siemens Avanto system (Cases A-D), and 5 on a 3T GE SIGNA Pioneer system (Cases E-I). For the 1.5T system, the protocol included anatomical T2-weighted fast spin echo imaging and dMRI. dMRI was based on a DW TRSE Echo Planar Imaging (EPI) sequence, with the diffusion encoding pro-

tocol matching exactly the TRSE protocol used in simulations (see above for details). Other salient parameters were: resolution $1.9 \times 1.9 \times 6\ mm^3$, TE = {93, 105, 120} ms, TR = 7900 ms, trace DW imaging, NEX = 2, GRAPPA = 2, 6/8 Partial Fourier imaging, BW = 1430 Hz/pixel, acquisition of a $b = 0$ image with reversed phase encoding. For the 3T scanner instead, the protocol also included anatomical T2-weighted fast spin echo imaging and dMRI, acquired with PGSE EPI according to the "PGSE-in" protocol described in simulations above. Other salient parameters were: resolution $2.4 \times 2.4 \times 6\ mm^3$, TE = {75, 90, 105} ms, TR ≈ 3000 ms (respiratory gated), trace DW imaging, NEX = 2, ASSET = 2, BW = 1953 Hz/pixel. The dMRI protocol took approximately 15 minutes in both machines.

Scans were denoised [Veraart et al., 2016], corrected for Gibbs ringing [Kellner et al., 2016] and motion, and EPI distortions mitigated (1.5T system only) [Andersson et al., 2003]. Afterwards, each DW image was normalised to the $b = 0$ acquired at the corresponding TE [Panagiotaki et al., 2015], and *forward model 2* (see section 2.4.1) was fitted voxel-by-voxel (regularised maximum-likelihood NNLS fitting; images for $b \leq 100\ s/mm^2$ were excluded to minimising vascular contributions). This provided voxel-wise maps of f_{in}, $D_{0|in}$, mCS, varCS, skewCS, $D_{0|ex}$ within manually-outlined tumours. Mean and standard deviation of all metrics within the tumours were extracted.

In parallel, we also assessed visually per-patient HE-stained histological material (Hamamatsu C9600-12 slide scanner; resolution: 0.45 μm) from one of the imaged tumours, which was acquired through an ultrasound-guided biopsy performed approximately one week after MRI.

3 Results

3.1 Substrates

Table 1 reports the characteristics of the substrates reconstructed from histology for this study, while Supplementary Fig. S2 shows the raw histological HE-stained patches from which substrates were reconstructed. The substrates encompass a rich variety of cytoarchitectures and microenvironments. For example, they include areas featuring densely packed cells, e.g., f_{in} values as high as 0.868 in the non-cancerous liver parenchyma or 0.747 in HCC, as well as areas with f_{in} as low as 0.130 in CRC fibrosis or 0.024 (almost complete lack of cells) in necrosis. The table also highlights, for example, that the substrates encompass a variety of CSs. The largest cells are found in the non-cancerous liver (mCS around ∼16 μm), while all cancers feature the presence of smaller cells. Differences in CS are also seen within the same type of cancer, e.g., mCS of ∼13 μm and ∼6 μm in two different CRC substrates. Substrates also feature different skewnesses of the CS distribution, with positive skewCS in most cancers, and negative skewCS in the non-cancerous liver. Finally, in some substrates (e.g., CRC) the EC space features the presence of large lumina, with equivalent diameters as large as ∼90 μm. Substrates also include areas of partial volume between non-cancerous hepatocytes and cancer cells (substrates 10, 11, 12) with different proportions, a fact that is reflected in different values of skewCS.

Fig. 2 illustrates the segmentation of the different cellular structures in four representative substrates, namely: non-cancerous liver, CRC, breast cancer, and melanoma. The figure highlights again the richness of microstructural characteristics included in our substrates. For example, tightly packed cells are seen in both non-cancerous liver and in melanoma, with the former showing much larger cells than the latter (mCS of almost 16 μm in non-cancerous liver, twice as large as the approximately 8 μm seen in melanoma). A wide range of IC fraction f_{in} is also seen, ranging from 0.076 in the breast cancer substrate (containing fibrotic areas and extensive necrosis) up to 0.846 for the non-cancerous liver. Finally, large luminal spaces in CRC substrates occupy a considerable portion of EC space, with areas equivalent to the space taken by hundreds of cells.

3.2 Characterisation of IC and EC diffusion

Fig. 3 visualises ADC_{in} and ADC_{ex} as a function of diffusion gradient timing parameters δ and Δ, and for different values of the intrinsic IC and EC diffusivities ($D_{0|in}$ and $D_{0|ex}$). The figure reports ADC_{in} and ADC_{ex} for two illustrative cases, i.e., a CRC case (substrate 4) and a non-cancerous liver substrate (substrate 2). The CRC features the presence of very small cells (mCS close to 6 μm), but relatively low cell density (f_{in} of 0.44), due to the presence of stroma and luminal spaces. For any value of $D_{0|in}$ and $D_{0|ex}$, as well any possible combination of the gradient timings δ and Δ, ADC_{ex} is always considerably higher than ADC_{in} (maximum ADC_{in} of 0.19 $\mu m^2/ms$, minimum ADC_{ex} of 0.56 $\mu m^2/ms$). The non-cancerous liver substrates instead features tight packing of large hepatocytes: f_{in} is higher than in the previous CRC substrate, and mCS is almost 15 μm. In this case, when $D_{0|in} = D_{0|ex}$, then ADC_{ex} is always comparable if not much higher than ADC_{in}. However, there are cases in which ADC_{ex} is lower than ADC_{in}, provided that $D_{0|ex}$ is also lower than $D_{0|in}$ and the diffusion time is short (for example, $\mathrm{ADC}_{ex} = 0.64$ for $(\delta,\Delta) = (5\ ms,\ 15\ ms)$ and $D_{0|ex} = 1.5\ \mu m^2/ms$; $\mathrm{ADC}_{in} = 0.74$ for $(\delta,\Delta) = (5\ ms,\ 15\ ms)$ and $D_{0|in} = 2.2\ \mu m^2/ms$). Supplementary Fig. S3 shows ADC_{in} and ADC_{ex} at varying $D_{0|in}$ and $D_{0|ex}$ as a function of (δ,Δ) for an additional HCC substrate (substrate 14). Results sit in between those observed for the CRC and non-cancerous liver. ADC_{ex} is consistently larger than ADC_{in}, as in the CRC case, but the difference between the two is not as large in the CRC case, which has a lower f_{in}.

Supplementary Fig. S4 visualises the dependence of ADC_{ex} on the IC fraction f_{in}, for varying gradient separation and duration, and for an illustrative $D_{0|ex}$ of 2.5 $\mu m^2/ms$. The figure shows that ADC_{ex} decreases with increasing f_{in}, until it plateaus for f_{in} of roughly 0.75. ADC_{ex} also shows a slight decrease with increasing diffusion time (note that its long time tails are known to be $\sim \ln(\Delta)/\Delta$ [Burcaw et al., 2015]), especially at high cell density (panel b in Supplementary Fig. S4). The functional form $\mathrm{ADC}_{ex} = (1 - f_{in})\, D_{0|ex}$ for the long-time asymptotic limit agrees reasonably well with results from simulations for low and intermediate f_{in}, but departs considerably from those at higher f_{in} (f_{in} approximately equal to 0.7 or higher).

Fig. 4 shows scatter plots between ADC_{in} and CS distribution moments mCS, varCS and skewCS at varying diffusion gradient separation Δ. The figure refers to an illustrative case where δ and $D_{0|in}$ have been fixed to 15 ms and 2.2 $\mu m^2/ms$. There is a strong, positive correlation between ADC_{in} and mCS and varCS (Pearson's r of 0.98 and 0.93 for $\Delta = 65$ ms), and a strong negative correlation between ADC_{in} and skewCS (r = -0.78 for $\Delta = 65$ ms). The dependence of ADC_{in} on the CS distribution moments is approximately linear and almost perfectly monotonic as far as mCS and varCS are concerned, while it is more

complex for skewCS.

3.3 Simulation-informed microstructural parameter inference

3.3.1 *In silico* data

Fig. 5 shows scatter density plots reporting on the quality of parameter estimation *in silico*. The figure refers to the prediction of parameters from *forward model 1*, which is compared to the standard analytical model in Eq. 6. It is apparent that f_{in} and $D_{0|in}$ are, respectively, the metrics that are the most/the least accurately predicted. Correlation coefficients between ground truth and predicted values are consistently higher for MC-informed fitting compared to fitting of an analytical model. While for both models such a correlation is very strong when f_{in} is concerned, it is moderate for MC-informed fitting ($r = 0.50$) and low for analytical model fitting ($r = 0.24$) regarding vCS_{cyl}. For $D_{0|in}$ instead, it is weak for both approaches, although considerably higher for MC-informed fitting ($r = 0.17$ against 0.05). Interestingly, we also observe a moderate correlation between ground truth and predicted $D_{0|ex}$ for MC-informed fitting. Note that the analytical model in Eq. 6 enables the estimation of the EC apparent coefficient ADC_{ex}, and not of the intrinsic EC diffusivity $D_{0|ex}$.

Supplementary Fig. S5, S6 and S7 report instead scatter density plots between ground truth and predicted parameters for *forward model 2*, which enables the estimation of CS mean, variance and skewness. The figures respectively report results for the three simulated protocols (PGSE-in, TRSE, PGSE-ex). For metrics f_{in}, $D_{0|in}$ and $D_{0|ex}$, results are in line with what was seen for *forward model 1*. We observe a good agreement and a strong correlation between predicted and reference values for f_{in} in all cases (e.g., $r = 0.93$ for TRSE). For $D_{0|in}$, correlations are weak ($r = 0.36$ for PGSE-ex), while for $D_{0|ex}$ they are weak to moderate ($r = 0.40$ for PGSE-in). Regarding CS distribution moments, mCS is the metric that is most accurately predicted, with correlations varying from moderate ($r = 0.53$ for PGSE-in and 0.68 for TRSE) to strong ($r = 0.77$ for PGSE-ex). For varCS we observe moderate correlations between estimated and ground truth values (up to $r = 0.67$ for PGSE-ex), while for skewCS weaker correlations (highest correlation of $r = 0.29$ for TRSE).

3.3.2 *Ex vivo* data

Fig. 6 shows dMRI and co-localised HE images in the four breast specimens (one non-cancerous breast; three MMTV-PyM tumours, obtained at weeks 9, 11 and 14). Inspection of the HE images reveals that the specimens contain a variety of cytoarchitectural environments, with higher inter-sample and intra-sample heterogenity. For example, the non-cancerous breast specimen features areas rich in stroma. Conversely, higher cell density compared to the non-cancerous case are observed in the three MMTV-PyM tumours. At late stages (week 14 tumour), widespread necrosis is also seen. Fig. 7 shows parametric maps from forward model 2 (namely: f_{in}, $D_{0|in}$, mCS, varCS, skewCS and $D_{0|ex}$) in the same breast specimens. The variability of cellular microarchitectures seen in Fig. 6 is reflected in the parametric maps. For example, reduced f_{in} is seen in areas compatible with necrosis within the week 14 tumour (ROI 2, Fig. 7). Additionally, higher f_{in} is seen in the week 11 tumour, compared to the non-cancerous breast. On histology, this contrast corresponds to presence of areas featuring high cellularity (Fig. 7, ROI 4), compared to stroma in the non-cancerous breast (Fig. 7, ROI 5). Changes in CS moments with respect to the non-cancerous breast are also seen, e.g., reduced mCS and varCS in areas compatible with presence of cell debris in necrosis (ROI 2, Fig. 7). Local variations of intrinisic IC and EC diffusivities $D_{0|in}$ and $D_{0|ex}$ are also seen. For example, $D_{0|in}$ is lower in areas with high f_{in} (e.g., in ROI 4 in the week 14 tumour), and $D_{0|ex}$ is the highest at the interface between specimens and agarose.

Supplementary Fig. S8 shows dMRI and HE data in a normal spleen and in splenomegaly secondary to late-stage MMTV tumour growth.The spleens exhibit a patchy structure in most dMRI metrics. The same pattern is seen in HE histology, where an alternation of white and red pulps (higher T-cell density in the white pulps, and lower T-cell density in the red pulps, which are instead rich in blood and iron) is seen.

Supplementary Fig. S9 shows results from the two kidney samples: one normal, and one collected following folic-acid induced injury. On histology, the former shows normal representation of all kidney structures, while the injuried case shows proximal tubule alteration and extensive inflammation. In terms of dMRI metrics, the injured kidney shows increased f_{in} and reduced $D_{0|in}$ and $D_{0|ex}$ as compared to the normal case. Higher f_{in} is also seen in the injured kidney cortex as compared to its medulla, a finding that corresponds to higher cell density on visual inspection of histology stains.

Finally, Table 2 provides descriptive statistics (mean and standard deviation) for all dMRI metrics within the different ROIs, alongside quantitative histology indices from patches whose location matched the dMRI ROIs. Contrasts in histological metrics agree with dMRI in several cases. For example, histological mCS is lower in necrotic compared to non-necrotic areas in the week 14 breast tumour (ROI 2 vs 3), or in the spleens compared to the kidneys (ROI 6 compared to ROI 8). Histological f_{in} is higher in the week 9 breast tumour than in the non-cancerous breast (ROI 1 vs 5). In some cases, differences between dMRI and histology contrasts are also seen. For example, the low dMRI f_{in} seen in the healthy kidney underestimates considerably the corresponding f_{in} values from histology (ROI 8).

The table also reports Pearson's correlation coefficients between dMRI metrics and histology. The correlation between histological and dMRI values is moderate for all metrics: f_{in}: $r = 0.694$, $p = 0.037$; mCS: $r = 0.671$, $p = 0.047$; varCS: $r = 0.510$, $p = 0.160$; skewCS: $r = 0.460$, $p = 0.212$. The strongest correlations are seen between MRI f_{in} and mCS with their histological counterparts.

3.3.3 *In vivo* data

Examples of parametric maps from *forward model 2* obtained *in vivo* are shown in Fig. 8 in two patients (ovarian cancer liver metastases, scanned at 1.5T; endometrial cancer, scanned at 3T). Maps show intra-tumour variability. For example, in the ovarian cancer case, the largest liver metastasis features reduced f_{in} and mCS and increased $D_{0|ex}$ in the necrotic core compared to the tumour outer ring. For the endometrial cancer case, maps suggest the existence of different microstructural environments, i.e., areas with higher/lower f_{in}, matching areas with lower/higher mCS. Inspection of histological images confirms the existence of heterogeneous cellular characteristics in both cases (Fig. 8). We observe presence of active can-

cer and necrosis in the ovarian cancer case, and presence of necrotic areas with abundance of cell debris adjacent to areas with high cellularity in the endometrial tumour. Supplementary Table S1 summarises mean and standard deviation of all dMRI metrics within the segmented tumours. The table reveals inter-tumour heterogeneity: as an example, melanoma tumours show the highest f_{in}, with higher figures than endometrial and, even more, CRC tumours. Among all tumours, melanoma cases are the only ones exhibiting, on average, slightly negative skewCS. For all other tumours, a positive mean skewCS is recorded.

Substrate	Description of ROI	ROI area (μm^2)	Cellularity (cells / mm^2)	f_{in}	mCS (μm)	varCS (μm^2)	skewCS	vCS_{sph} (μm)	vCS_{cyl} (μm)	Γ distribution shape	Γ distribution scale (μm)	Lumen fraction of EC space f_l	Lumina diameters (μm)
1	non-cancerous liver, patient 1	47689	3921	0.773	15.3	14.2	-0.115	18.7	18.1	14.69	1.04	0.000	-
2	non-cancerous liver, patient 1	59248	4878	0.868	14.5	15.0	-0.156	18.1	17.4	12.32	1.18	0.000	-
3	non-cancerous liver, patient 2	59377	3907	0.827	15.9	15.9	-0.528	19.1	18.5	12.58	1.26	0.000	-
4	non-cancerous liver, patient 2	69172	4091	0.846	15.6	19.7	-0.283	19.9	19.1	9.69	1.60	0.000	-
5	CRC with lumina, patient 3	23752	8968	0.498	8.1	4.8	0.861	11.6	10.6	14.04	0.58	0.157	[27.7, 39.0]
6	CRC with fibrosis and lumina, patient 3	90130	2408	0.323	12.8	12.8	0.528	17.4	16.3	14.49	0.91	0.144	[27.8, 88.9]
7	CRC with lumina, patient 3	39305	10508	0.523	7.7	4.9	0.349	10.3	9.7	11.56	0.66	0.512	[33.9, 57.9, 34.0, 67.8]
8	CRC with lumina, patient 4	62434	13743	0.437	6.1	3.3	0.628	8.5	8.0	11.34	0.53	0.253	[42.9, 26.2, 23.1, 24.1, 57.4, 64.2]
9	CRC with fibrosis, patient 5	56891	3375	0.130	6.8	2.3	0.287	8.2	7.9	19.23	0.35	0.000	-
10	CRC and hepatocytes, patient 1	49637	6870	0.658	10.2	17.5	-0.191	14.7	14.1	4.56	2.23	0.000	-
11	CRC and hepatocytes, patient 1	43056	7873	0.730	10.4	9.7	0.230	14.0	13.2	10.20	1.02	0.000	-
12	CRC and hepatocytes, patient 1	45109	7737	0.831	11.1	11.9	0.396	15.4	14.5	10.12	1.10	0.000	-
13	HCC with necrosis, patient 6	128132	2622	0.117	7.1	5.2	0.786	10.9	9.9	9.79	0.73	0.000	-
14	HCC, patient 7	62088	5492	0.747	12.4	17.7	-0.004	17.3	16.3	7.43	1.67	0.000	-
15	HCC with necrosis, patient 7	57699	676	0.024	6.4	3.4	0.481	8.5	8.1	12.09	0.53	0.000	-
16	Melanoma, patient 8	51700	8704	0.718	9.9	5.5	-0.005	12.1	11.6	16.34	0.61	0.000	-
17	Breast cancer with fibrosis and necrosis, patient 9	63121	2139	0.076	6.4	3.2	0.614	8.6	8.2	13.20	0.49	0.000	-
18	Melanoma, patient 10	41632	11434	0.587	7.7	5.2	0.603	10.8	10.1	11.29	0.68	0.000	-

Table 1: Properties of the substrates used for MC simulations, drawn on histological tissue coming from liver tumour biopsies from 10 cancer patients. Substrates 1 to 4 contain only non-cancerous hepatocytes. Substrates 5 to 9 contain colorectal cancer (CRC) tissue, featuring large luminal spaces or presence of fibrosis, as in substrate 6 and especially substrate 9. Substrates 10 to 12 contain a mix of non-cancerous and different concentrations of CRC cells. Substrates 13 to 15 are sampled from hepatocellular carcinoma (HCC) cases, with substrate 13 in particular containing HCC cells and necrosis, and substrate 15 taken from an area of extended necrosis with scattered cells and cell debris. Substrates 16 and 18 are ROIs of a melanoma liver metastasis. Finally, substrate 17 was taken from a breast cancer liver metastasis, and shows areas of necrosis and fibrosis.

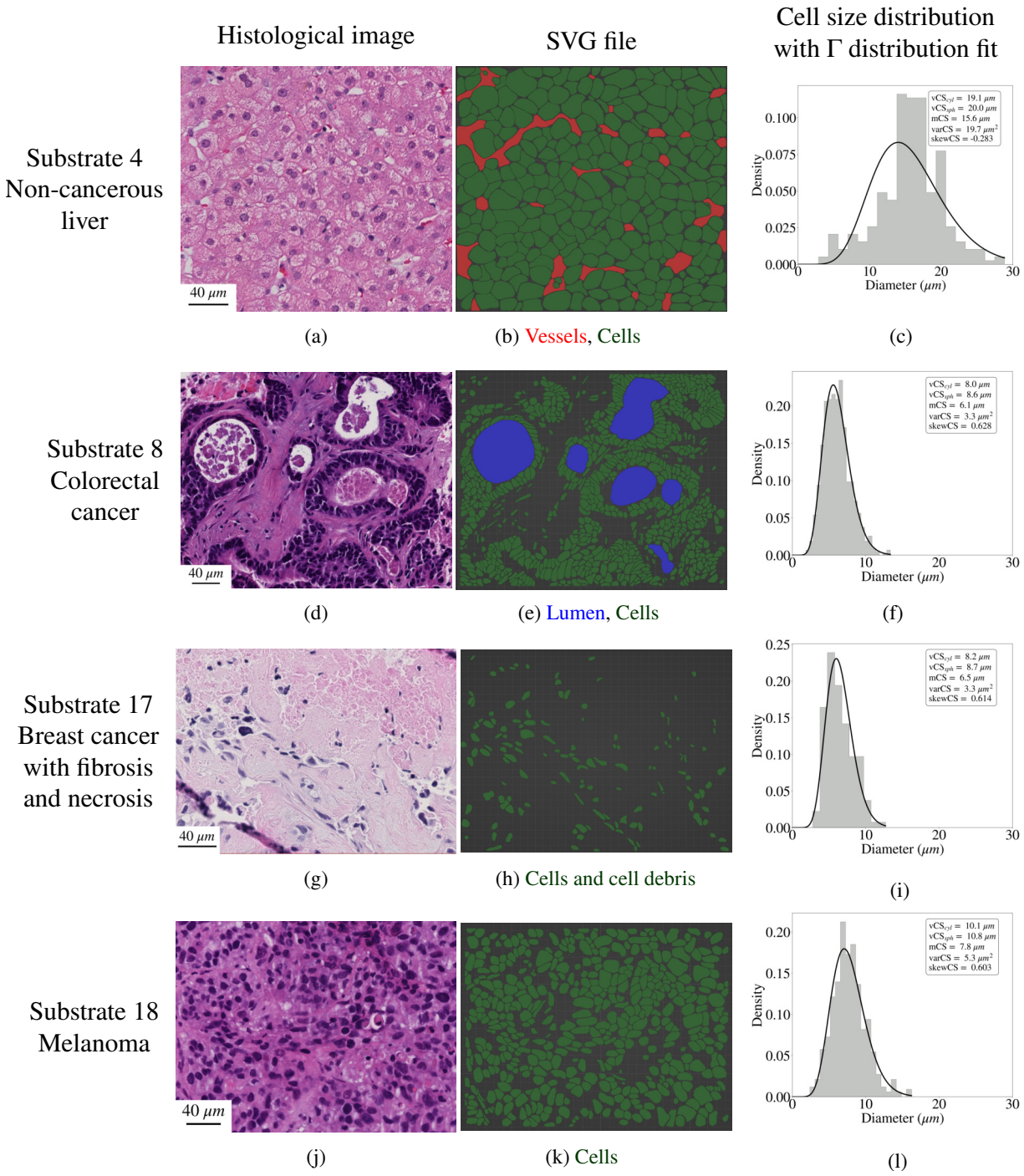


Figure 2: Visualisation of four illustrative substrates used for MC diffusion simulations. Left column (panels (a), (d), (g), (j)): HE histological images. Central columns (panels (b), (e), (h), (k)): SVG files reconstructed with the Blender software package, showing different substrate features (e.g., cells and debris in green, vessels in red, lumina in dark blue). Right column (panels (c), (f), (i), (l)): histograms depicting the CS (i.e., cell diameter) distribution for each substrate, with summary statistics and with a Gamma distribution fit superimposed onto it (black solid line). From top to bottom: non-cancerous liver (substrate 4), colorectal cancer (substrate 8), breast cancer (substrate 17), melanoma (substrate 18).

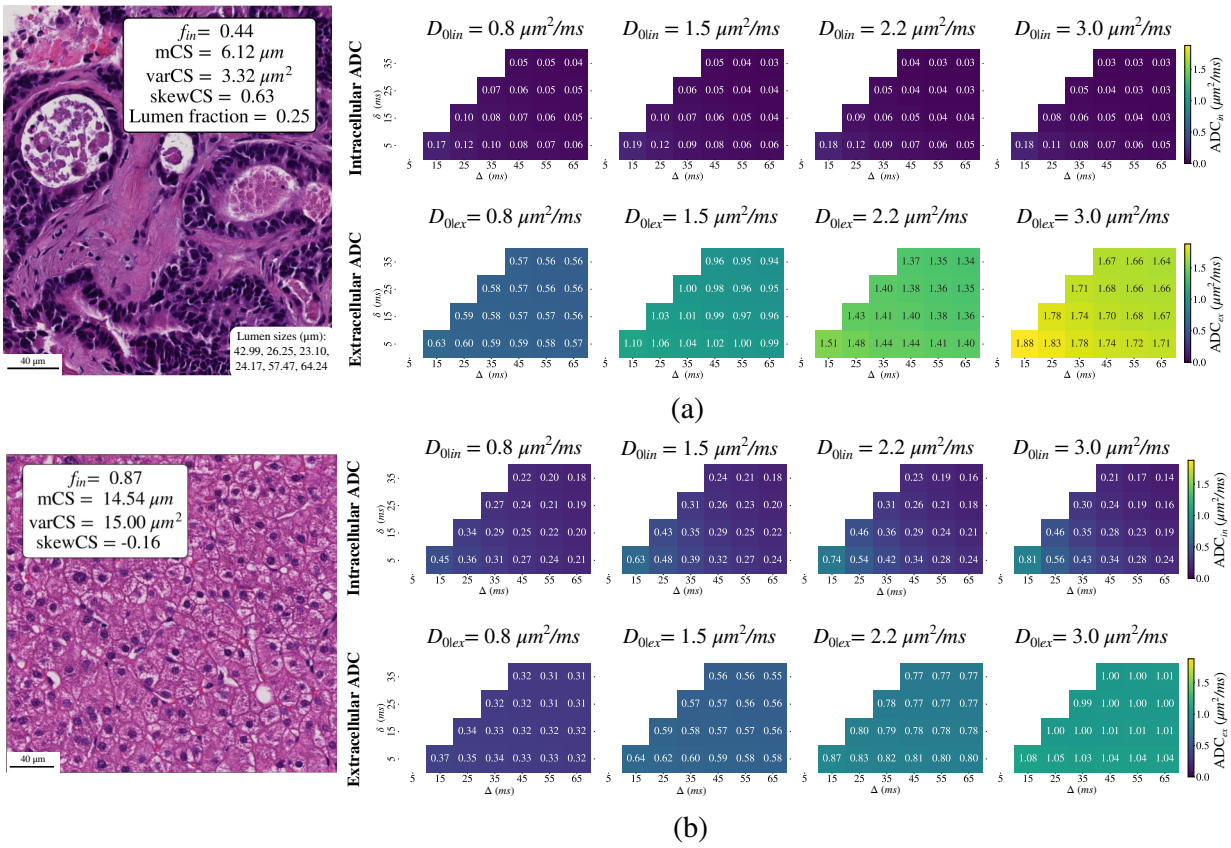


Figure 3: Plots visualising IC/EC ADC (ADC$_{in}$ and ADC$_{ex}$) as a function of the diffusion gradient duration δ and separation Δ and varying intrinsic IC/EC diffusivities $D_{0|in}$ and $D_{0|ex}$ for two representative substrates. (a) on top: CRC substrate, featuring presence of stroma and luminal spaces, as well as packing of small cells (substrate 4). (b): non-cancerous liver tissue, featuring the presence of tightly packed large hepatocytes (substrate 2). From left to right: substrate with properties, and ADC$_{in}$ and ADC$_{ex}$ for increasing values of $D_{0|in}$ and $D_{0|ex}$.

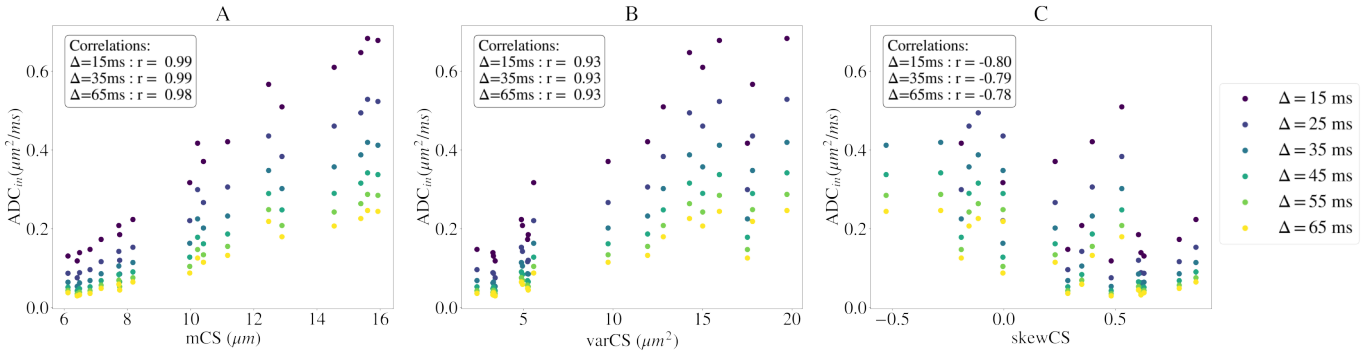
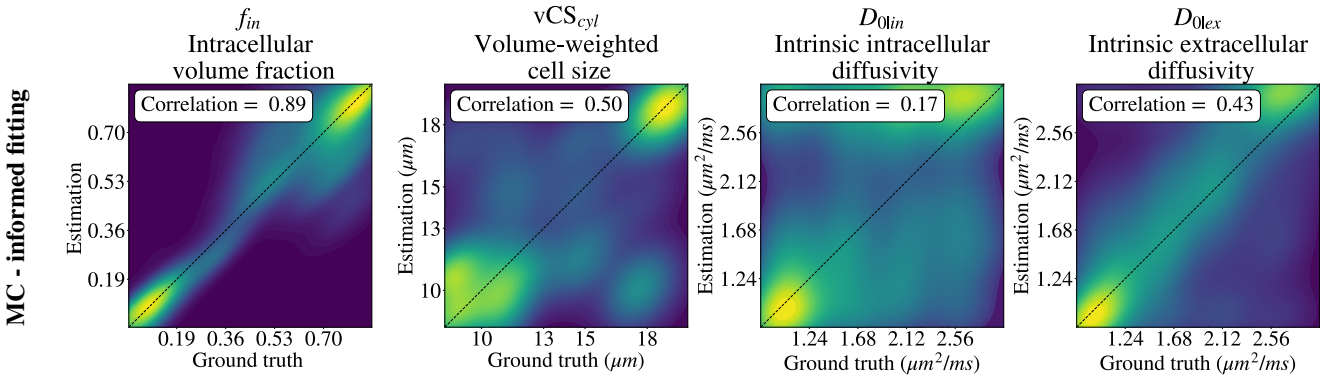
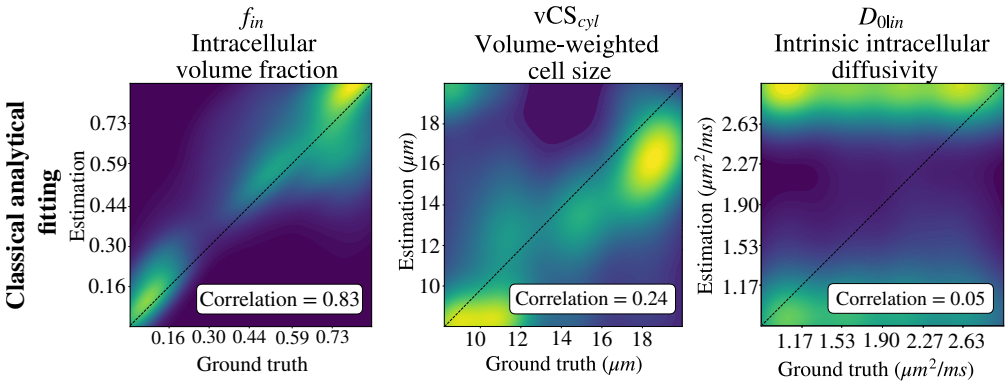


Figure 4: Relation between intracellular ADC and CS distribution moments, shown for $D_{0|in} = 2.2\mu m^2/ms$ and fixed gradient duration $\delta = 15ms$ with increasing gradient separation Δ.



(a) Fitting of a numerical signal model informed by MC simulations



(b) Fitting of a standard hindered-restricted diffusion analytical signal model

Figure 5: Scatter density plots between ground truth and estimated tissue parameters for MC-informed parameter estimation (*forward model 1*; panel (a) on top) and for standard NNLS fitting of a related hindered-restricted diffusion analytical model (panel (b), on the bottom). From left to right: IC fraction f_{in}, CS statistic vCS$_{cyl}$, intrinsic IC diffusivity $D_{0|in}$, intrinsic EC diffusivity $D_{0|ex}$ (only for MC-informed fitting on top). The plots also include the identity line for reference, and the Pearson's correlation coefficient between ground truth and estimated parameter values. Results are shown for dMRI protocol PGSE-in.

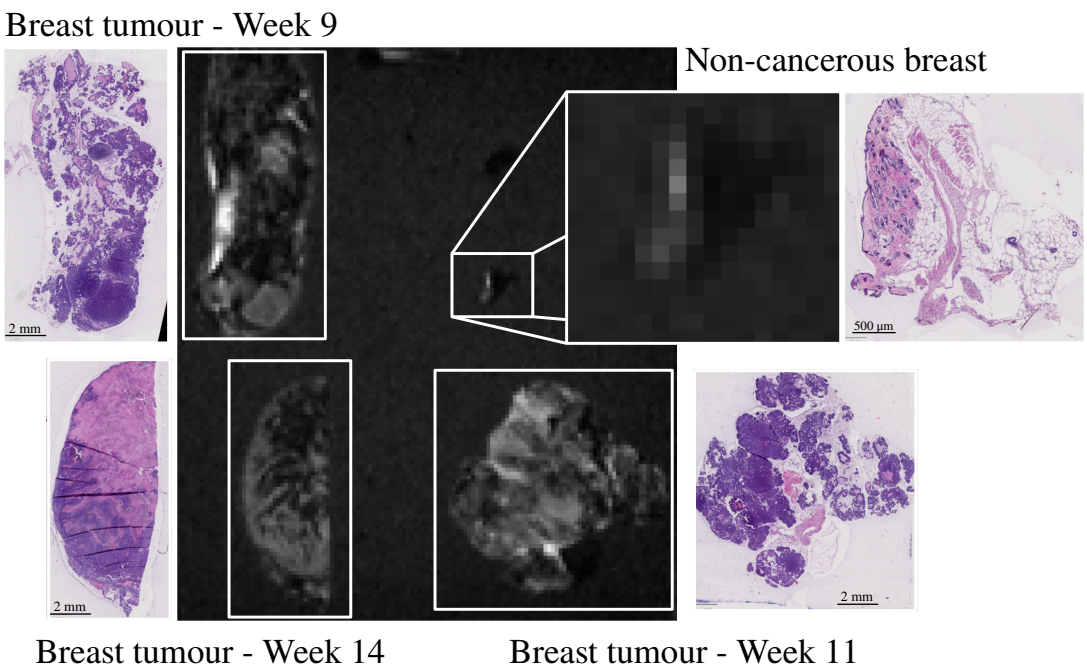


Figure 6: Images from the breast tissue samples that were scanned *ex vivo* at 9.4T. Top (panel 1): $b = 0$ dMRI image and co-localised HE-stained section. From top left, clock-wise: week 9 MMTV-PyM breast tumour, non-cancerous breast, week 11 MMTV-PyM breast tumour, week 14 MMTV-PyM breast tumour.

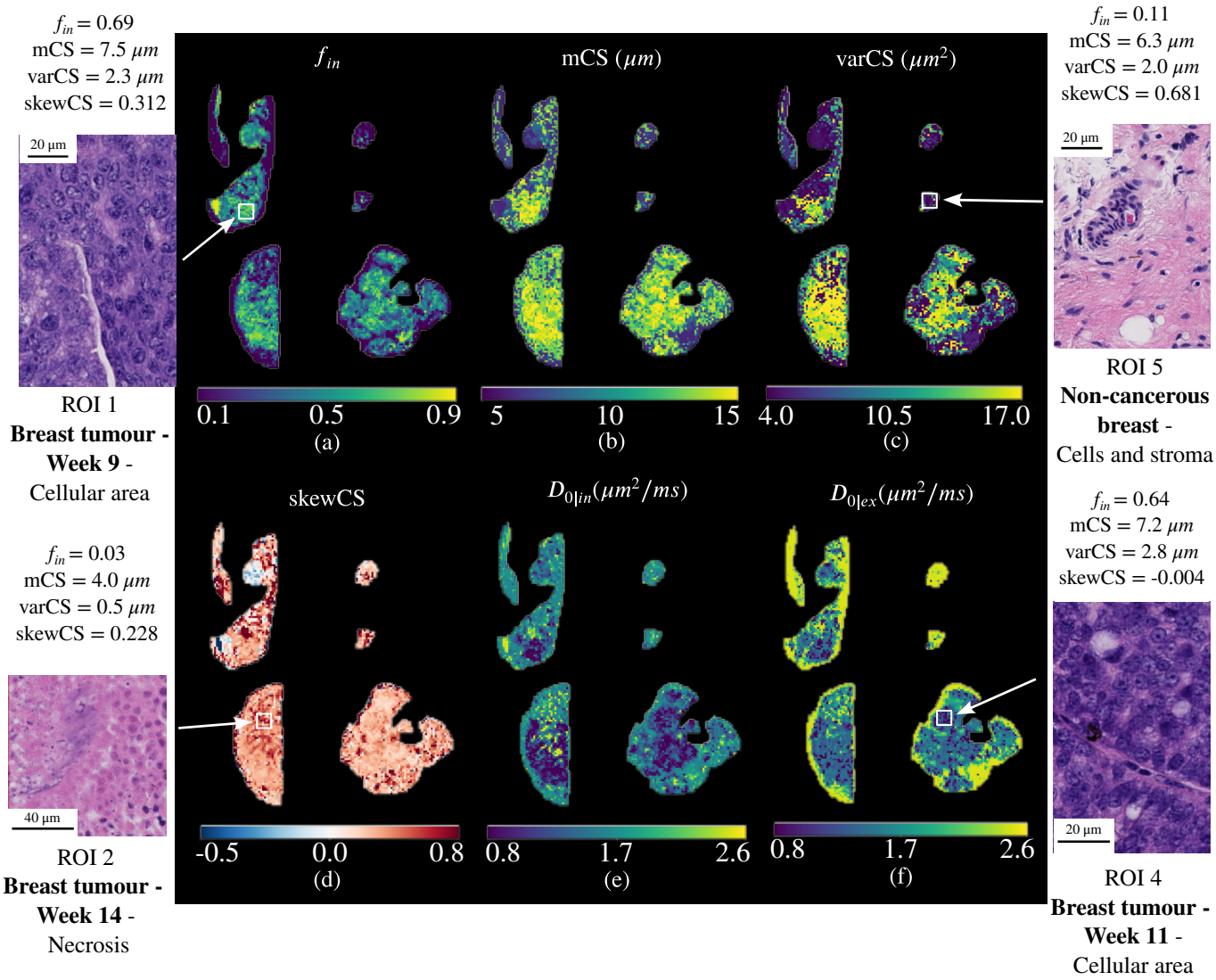


Figure 7: Parametric maps from *forward model 2* as obtained on the mouse breast specimens scanned *ex vivo* on a 9.4T system. First row: IC fraction f_{in} (a); mean CS index mCS (b); variance of CS varCS (c). Second row: skewness of the CS distribution skewCS (d); intrinsic IC diffusivity $D_{0|in}$ (e); intrinsic EC diffusivity $D_{0|ex}$ (f). For each metric, we show results on the four breast specimens. Moving clock-wise: week 9 MMTV-PyM breast tumour (top left), non-cancerous breast (top right), week 11 MMTV-PyM breast tumour (bottom right), week 14 MMTV-PyM breast tumour (bottom left).

Description	f_{in}	f_{in} - histology	mCS (μm)	mCS - histology (μm)	varCS (μm^2)	varCS - histology (μm^2)	skewCS	skewCS - histology	$D_{0\|in}$ ($\mu m^2/ms$)	$D_{0\|ex}$ ($\mu m^2/ms$)
ROI 1: Breast tumour - Week 9	0.64 [0.07]	0.69	14.9 [1.3]	7.5	16.2 [1.5]	2.3	0.310 [0.116]	0.312	0.80 [0.29]	1.40 [0.33]
ROI 2: Breast tumour - Week 14 (necrosis)	0.02 [0.11]	0.03	7.5 [2.5]	4.0	3.7 [6.1]	0.5	0.345 [0.101]	0.228	1.75 [0.35]	2.06 [0.49]
ROI 3: Breast tumour - Week 14 (cellular area)	0.70 [0.10]	0.72	15.3 [0.7]	9.0	15.9 [3.1]	3.6	0.295 [0.013]	-0.214	0.90 [0.21]	1.42 [0.19]
ROI 4: Breast tumour - Week 11	0.48 [0.06]	0.64	13.1 [1.0]	7.2	12.7 [4.6]	2.8	0.252 [0.067]	-0.004	1.01 [0.24]	0.80 [0.26]
ROI 5: Non-cancerous breast	0.02 [0.06]	0.11	6.5 [0.2]	6.3	5.7 [0.8]	2.0	0.614 [0.221]	0.681	1.63 [0.03]	2.50 [0.10]
ROI 6: Healthy spleen	0.34 [0.15]	0.75	8.1 [1.9]	4.4	5.1 [2.7]	1.0	0.126 [0.301]	0.244	1.47 [0.35]	2.26 [0.32]
ROI 7: Splenomegaly	0.40 [0.34]	0.64	6.2 [0.1]	4.0	4.6 [0.8]	0.7	0.506 [0.250]	0.131	1.80 [0.27]	2.31 [0.53]
ROI 8: Healthy kidney	0.18 [0.06]	0.82	12.3 [0.6]	10.1	12.5 [0.4]	6.6	0.246 [0.050]	-0.970	1.61 [0.05]	1.80 [0.23]
ROI 9: Injured kidney	0.62 [0.08]	0.66	13.6 [0.8]	5.0	15.7 [0.9]	1.3	0.249 [0.020]	0.085	1.18 [0.17]	1.40 [0.21]
Correlations	**0.694**, p-value=0.037		**0.671**, p-value=0.047		**0.510**, p-value=0.160		**0.460**, p-value=0.212			

Table 2: Summary statistics of metrics from *forward model 2* and from histology within different ROIs drawn on the breast, kidney and spleen tissue scanned *ex vivo* on a pre-clinical 9.4T MRI system. The table also includes Pearson's correlation coefficients between each dMRI metric and its direct histological counterpart.

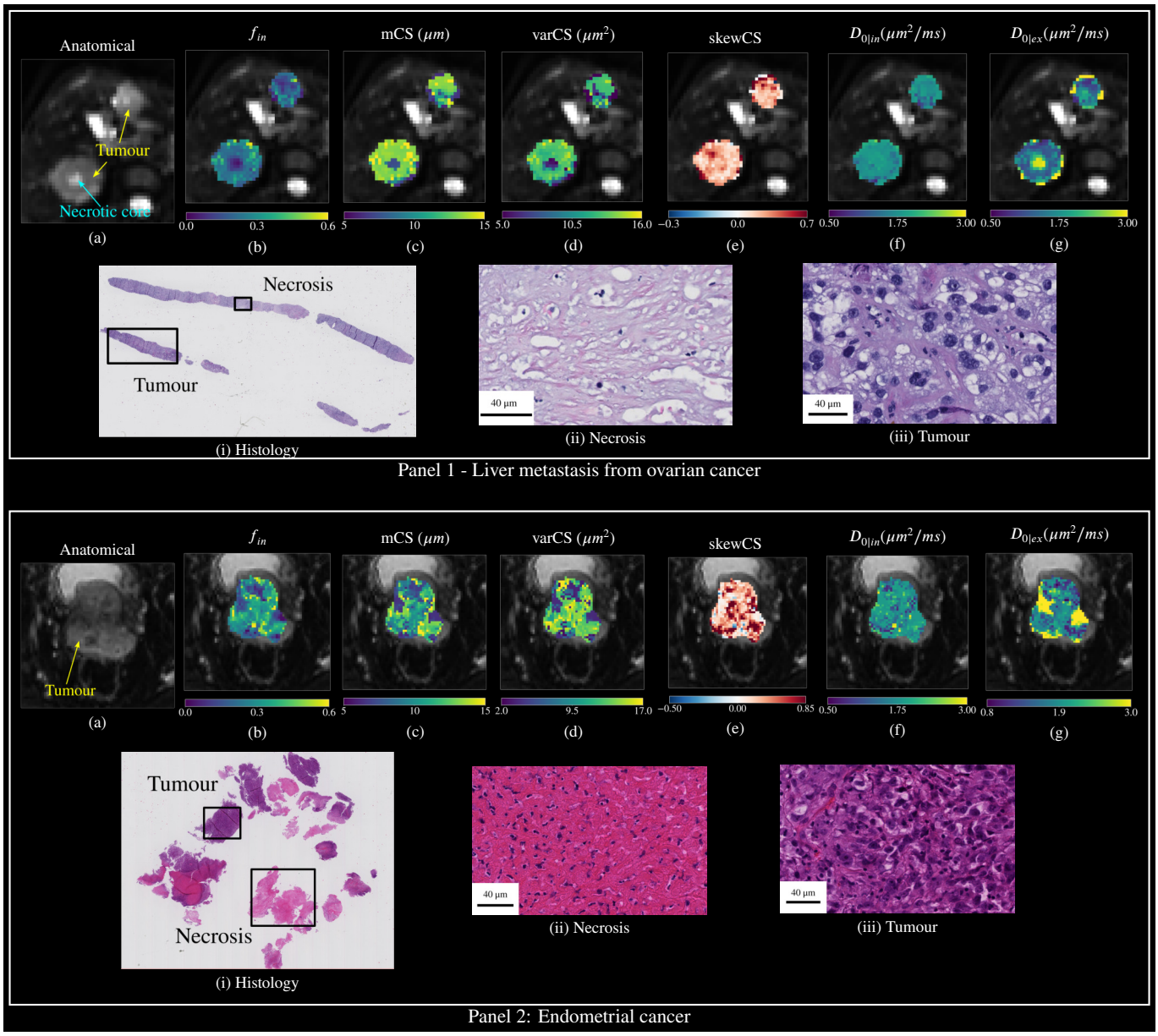


Figure 8: Parametric maps from forward model 2 as obtained on two representative patients *in vivo*, scanned on two different MRI scanners. Panel 1 on top: maps shown in Case G (liver metastases from ovarian cancer; 3T GE SIGNA Pioneer scanner). Panel 2 on the bottom: Case A (endometrial cancer; 1.5T Siemens Avanto scanner). From left to right, each panel reports a $b = 0$ image with the tumour outline (a) and then metrics f_{in} (b), mCS (c), varCS (d), skewCS (e), $D_{0|in}$ (f) and $D_{0|ex}$ (g). Below the metrics, details from a biopsy taken from one of the imaged tumours are also included (HE staining).

4 Discussion

4.1 Summary and key findings

This article presents a framework for MC diffusion simulations within cellular environments reconstructed from histology. We use it to study diffusion patterns in the liver, and to enable the non-invasive estimation of heretofore unexplored microstructural properties. Our work has four main contributions. Firstly, it describes a practical step-by-step procedure, based entirely on freely available software, to reconstruct meshed cellular environments from histological images, which can be fed to popular MC diffusion simulators. Secondly, it provides the dMRI community with unique reference values of histology-derived cell size and density in non-cancerous and cancerous human liver tissues. Information of this type, not easily found in the literature, is essential to inform the design of new dMRI signal models, or to optimise acquisition protocols for real-world imaging. Thirdly, the simulations performed within the reconstructed cellular environments shed light onto IC and EC diffusion, e.g., highlighting that for clinically realistic diffusion times, the EC ADC is higher than the IC ADC, regardless of the intrinsic IC/EC diffusivity. The measured IC and EC characteristics could be used, for example, to inform biophysically meaningful constraints in dMRI parameter estimation. Lastly, our paper showcases a numerical approach for dMRI parameter estimation informed directly by the simulated MC diffusion signals. The approach is shown to outperform classical fitting of analytical signal models, and it delivers innovative metrics, such as moments of the CS distribution or estimates of the intrinsic EC diffusivity. CS distribution moment estimates are shown to correlate moderately with histology in *ex vivo* mouse tissue, and were also obtained *in vivo* in patients within clinically acceptable scan times.

4.2 Simulation framework

Our simulation framework combines freely available software tools (i.e., QuPath [Bankhead et al., 2017], Inkscape and Blender) to reconstruct meshed cellular environments from 2D histological images. These are stored as sets of ASCII PLY files, a common file format for meshed geometrical models, being accepted by common open-source MC diffusion simulators such as MCDC [Rafael-Patino et al., 2020] or Camino [Hall and Alexander, 2009]. The procedure to convert histological data into PLY files has been described in detail in this article, and practical examples as well as tutorials for would-be users are provided in our freely accessible online repository, at the permanent address https://github.com/radiomicsgroup/dMRIMC. Our detailed guidelines will equip the community with a practical tool to increase the realism of dMRI simulations, strengthening the link between dMRI and histology. Ultimately, our framework may contribute to the development of novel dMRI techniques for body imaging, urgently needed in oncological applications.

4.3 Substrates

To demonstrate our framework, we segmented 18 cellular environments from HE-stained liver tumour biopsies, referred to as *substrates*. These included tissues of different kinds, e.g., non-cancerous liver parenchyma, as well as several liver cancers, either primary (e.g., HCC) or metastatic (e.g., CRC or melanoma), featuring areas of necrosis, fibrosis, or with tightly packed cells. The analysis of such substrates allowed us to quantify key microstructural properties in terms of cell density, IC area fraction, presence and morphology of EC luminal spaces, and CS distribution characteristics. We compiled a table reporting this information in a systematic manner, providing the dMRI community with reference histological values for cancer applications. To our knowledge, histology-derived cell morphometry literature has traditionally focussed on the study of neuroanatomy [Aboitiz et al., 1992, Duval et al., 2019], and limited quantitative data are available in body tissues or cancer. Information on the expected cell sizes and densities is essential to design and optimise acquisition protocols, e.g., in terms of b-values or diffusion times. Therefore, delivering such a data base is a major contribution of our work. The data base may be used to devise innovative dMRI acquisition protocols tailored for body imaging, for example by drawing cell diameter sets from our Gamma distribution fits, and using those for numerical dMRI optimisations.

4.4 Characterisation of IC and EC diffusion

We used the proposed MC simulation framework to synthesise dMRI signals for routine PGSE protocols that could be realistically implemented in pre-clinical and clinical settings. This allowed us to relate directly key features of the dMRI signal decay, in terms of an IC and EC ADC metric, to the histological properties of the substrates themselves. A number of key observations can be made. For example, we observed that for clinically-relevant diffusion times, the IC ADC (ADC_{in}) was smaller than the EC ADC (ADC_{ex}), regardless of the value of the intrinsic IC/EC diffusivities ($D_{0|in}$, $D_{0|ex}$). This fact held true even for the non-cancerous liver substrates, which feature the highest values of IC fraction (f_{in}), and hence the lowest ADC_{ex}. Only in very a limited number of extreme cases, e.g., for very short PGSE diffusion times and very low intrinsic diffusivities, ADC_{ex} was slightly lower than ADC_{in}. This was observed, for example, for $\Delta = 15$ ms, and when $D_{0|ex} = D_{0|in} \approx 0.8$ μm^2/ms. Considering that this scenario is likely closer to what can be expected for *ex vivo* preclinical MRI, rather than *in vivo* clinical imaging, we deduce that model fitting constraints of the type $\mathrm{ADC}_{in} < \mathrm{ADC}_{ex}$ (as in [White et al., 2014, Feng et al., 2021, Conlin et al., 2023]) appear to be biologically plausible for *in vivo* clinical application, and may thus be useful to better disentangle IC from EC diffusion signals.

The analysis of EC diffusion has also allowed us to investigate the relationship between ADC_{ex} and f_{in}. We observed a marked reduction of ADC_{ex} for increasing f_{in}, in line with well-known long-time asymptotic limits, as for example $\mathrm{ADC}_{ex}/D_{0|ex} = (1 - f_{in})$ [Reynaud, 2017]. However, our plots also revealed differences between the $(1 - f_{in})$ limit and the observed dependence of ADC_{ex} on f_{in}. First, ADC_{ex} varied with the diffusion time, albeit slightly. Second, ADC_{ex} tended to saturate to approximately $\mathrm{ADC}_{ex} \approx 0.3 D_{0|ex}$ for high cell density, diverging considerably from the monotonic decrease predicted by the asymptotic limit. In view of these results, we conclude that asymptotic limits such as $\mathrm{ADC}_{ex}/D_{0|ex} = (1 - f_{in})$ only provide a crude approximation of the actual dependence of ADC_{ex} on f_{in} for clinically achievable diffusion times, while also failing to capture the log-time tail of the diffusion-time dependence of ADC_{ex} [Burcaw et al., 2015]. Therefore, their use as parameter constraints may lead to modelling inaccuracies, especially in presence of tight cell packing.

Finally, our synthetic signals enabled the assessment of the dependence of the ADC_{in} on the characteristics of the CS distribution. Scatter plots and correlation analyses demonstrated that key moments of the CS distribution, namely, its mean, variance and skewness, are encoded in ADC_{in}, for the typical cell diameters considered here. This is apparent, for example, by noting that ADC_{in} increases with increasing mCS and varCS at fixed diffusion time (e.g., Pearson's $r = 0.98$ for $\Delta = 65$ ms for mCS), and that it shows a marked diffusion time dependence. To put this trends into context, we notice that ADC_{in} decreases from approximately 0.35 to 0.2 μm^2/ms when the gradient separation Δ increases from 35 to 65 ms, given an average mCS of roughly 12 μm, and for $D_{0|in} = 2.2$ μm^2/ms. This decrease in ADC would translate to an increase in dMRI signal attenuation $s/s(b = 0)$ from approximately 0.41 to 0.48 (17 % increase), even for a moderate $b = 1500$ s/mm^2, assuming $ADC_{ex} \approx 1.0$ μm^2/ms and $f_{in} \approx 0.5$. For comparison, note that at a conservative $b = 0$ SNR of 30 (following post-processing), the signal fluctuation is of the order of $\sigma/s(b = 0) = 1/SNR \approx 0.033$, and the noise floor the order of $\eta \approx 1.25\sigma = 0.042$ [Jones and Basser, 2004]. These observations suggest that the large size of cells found in body tissues, e.g., in liver and liver tumours, may make it feasible to resolve key characteristics of the CS distribution in clinical settings, that is, without relying on ultra-strong gradient systems as required for axon diameter measurement in the brain [Nilsson et al., 2017]. This key result motivates the need for novel microstructural models that enable body CS distribution mapping in clinical settings, as discussed in the following sections.

4.5 Simulation-informed microstructural parameter inference

We investigated whether synthetic signals from our framework can be used to devise new strategies for CS distribution mapping, urgently sought for cell population profiling in oncology [Jiang et al., 2020a, Hoffmann et al., 2023]. To this end, we interpolated the discrete dictionary of paired examples of tissue parameters and synthetic dMRI signals using RBF regressors. This provided numerical forward models that do not rely on approximated analytical functional forms for the IC/EC signal, e.g., restricted diffusion within cells of regular shape and equal size [Panagiotaki et al., 2015, Jiang et al., 2020b], or Gaussian EC diffusion. Such forward models can be easily embedded into routine NNLS fitting, based on likelihood maximisation [Panagiotaki et al., 2012].

We compared the performance of our approach in predicting a single cell diameter statistic [Veraart et al., 2020] (vCS_{cyl}) against standard analytical approaches based on restricted diffusion within cylinders. Results not only point towards the superiority of our approach in vCS_{cyl} estimation, but also show benefit in the estimation of other diffusion properties, such as the intrinsic cytosolic diffusivity or the IC fraction. We also went on and tested whether the approach can be deployed for geometries and dMRI sequences for which analytical expressions may not be readily available, provided that they can be simulated. We demonstrated this here, in forward model 2, for the estimation of the intrinsic EC diffusion coefficient and the first three CS distribution moments, without imposing any analytical functional form to the CS distribution. We observe satisfactory performances in the estimation of $D_{0|ex}$, mCS and varCS, i.e.; a moderate correlation between ground truth and estimated values for EC diffusivity $D_{0|ex}$ (r up to 0.40); strong to very strong correlations for mCS (r up to 0.77); moderate to strong for varCS (r up to 0.67). For skewCS, while performances are worse (correlations up to $r = 0.29$), they are comparable to those observed for other metrics, as the IC cytosol diffusivity $D_{0|in}$ ($r = 0.36$). Therefore, simulations suggest that our skewCS indices, while difficult to estimate, may still suffice to detect strong contrasts in CS skewness.

After demonstrating CS distribution mapping *in silico*, we tested whether it is also feasible on actual MRI scans. For this experiment, we analysed both pre-clinical *ex vivo* data from 8 mouse tissue samples, as well as *in vivo* scans acquired on cancer patients with two clinical MRI systems. Notably, the tissue scanned on the pre-clinical system was considerably different from that used to build the numerical signal models (e.g., mouse breast tumours, kidneys, and spleens, versus human liver parenchyma and liver tumours), and thus served as useful *out-of-distribution* test bed for generalisation. The parametric maps obtained *ex vivo* show a number of interesting and potentially relevant inter-sample and intra-sample contrasts. In general, several of such contrasts were confirmed by histology both qualitatively and quantitatively, such as reductions of f_{in} due to necrosis in an advanced breast tumour (14-week MMTV PyMT mouse), or presence of small cells in the spleens. Nevertheless, some differences between dMRI and histology were also seen: this was the case, for example, of discrepancies in f_{in} values in the healthy kidney (low in dMRI, high in histology). Overall, these findings led to moderate correlations between dMRI and histological properties f_{in}, mCS, varCS, and skewCS, et statistically significant for f_{in} and mCS,. These correlations are encouraging given i) the relatively small size of our sample; ii) the inherent difficulty of ensuring accurate co-localisation between dMRI and histology; and iii) given the differences between the substrates used to build the models and the tissue imaged *ex vivo*. Importantly, challenges in segmenting cell boundaries accurately over large field of views, as well as unaccounted water exchange [Jiang et al., 2022, Gardier et al., 2023] and intra-compartmental relaxation [Lemberskiy et al., 2018, Palombo et al., 2023], may have also contributed to these results (note for example that the correlations observed between dMRI and histology for metrics such as f_{in} and mCS are slightly lower than what could be expected from simulations in Supplementary Fig. S7). All in all, the *ex vivo* experiments suggest that MC-informed dMRI modelling may provide new biomarkers of tissue microstructure that are informed directly by histology, and which may shed new light onto the presence of different cell populations in a voxel, through CS distribution mapping. Metrics such as skewCS may have important applications in non-invasive cell profiling [Hoffmann et al., 2023], which we reserve to future studies.

Lastly, we demonstrated simulation-driven CS distribution mapping in a pilot cohort of patients scanned with two different MRI machines. While this demonstration was merely qualitative, it still highlights the potential utility of our approach. dMRI maps show key inter-tumour and intra-tumour contrasts, which are plausible given the high microstructural heterogenity seen on HE biopsies of the same tumours. For example, areas lying within tumour necrotic cores show reduced f_{in} and mCS, compatible with necrosis and cell debris. Moreover, melanoma tumours consistently show negative skewCS (in agreement with Table 1) as well as the highest f_{in} values, which may point towards their known malignity. Overall, the plausible metric contrasts seen in both MRI machines, the fact that maps were obtained with 15-minute scans, and without relying on custom

dMRI sequences or high-performance gradient systems, point towards the feasibility of our simulation-informed models in clinical settings.

4.6 Methodological considerations and limitations

We acknowledge some potential limitations of our approach. Firstly, while care has been taken when segmenting cells manually on histology, some segmentations may feature inaccuracies, e.g., at the cell-cell interfaces. This was due to the well-known challenge of visualising accurately cell boundaries on 2D HE stains.

Secondly, we focused here on 2D diffusion within cylindrical geometries. This was due to the availability of a large data set of HE-stained sections in human and mouse tissue, which is inherently 2D. In future, we plan to perform simulations in fully 3D substrates, reconstructed, for example, on 3D micrographs [Lee et al., 2020b] or on 3D confocal microscopy [Khan et al., 2015].

Thirdly, in this study we illustrated the benefits of relaxing some of the assumptions underlying standard analytical diffusion models through numerical simulations. However, we point out that our models are not free of assumptions, since they inherently inherit the hypotheses made to conduct the simulations themselves. In the future, we plan to increase the complexity of the simulations, and account for important aspects of tissue microstructure so far neglected, namely: variability in intrinsic diffusivity among cells or between lumina and extra-cellular space; trancytolemmal water exchange [Jiang et al., 2022, Gardier et al., 2023]; differences in intra-compartamental relaxation properties [Lemberskiy et al., 2018, Fieremans and Lee, 2018, Palombo et al., 2023].

Furthermore, we acknowledge that our signal models do not account for contributions coming from intra-voxel incoherent flow within capillaries [Le Bihan et al., 1986], which we did not simulate. For this reason, we only fitted them to $b > 100$ s/mm^2 measurements *in vivo*, where vascular signals are negligible [Cui et al., 2015]. In future, we aim to increase the realism of our simulations by including a third compartment of capillary perfusion, alongside intra-/extra-cellular diffusion.

Additionally, we remark that our *in vivo demonstration*, while promising, was merely qualitative. Therefore, it requires further validation through MRI-histology comparisons in larger cohorts, which we warrant in future work.

Lastly, in this work we did not study advanced diffusion encodings such as oscillating gradients [Jiang et al., 2020b], double diffusion [Shemesh et al., 2016] or *b*-tensor [Westin et al., 2016] encoding, since we focused on off-the-shelf, widespread clinical protocols, which are based on PGSE. In future we aim to simulate such more advanced approaches, since they can provide more detailed information on cancer microstructure than routine PGSE.

5 Conclusions

Our practical framework enables MC diffusion simulations within realistic cellular environments reconstructed directly from HE histology. The realism of the simulations offers unique opportunities to study water diffusion in the body, and to design innovative biophysical dMRI models, which we showcase through CS distribution mapping. The new CS mapping method, demonstrated on both *ex vivo* mouse data and on *in vivo* cancer patients, provides histologically-meaningful indices within clinically-acceptable scan times. It may therefore play a key role in the development of new essays for the non-invasive characterisation of solid tumors in the body, which are sought in several oncological applications.

Acknowledgements

VHIO would like to acknowledge: the State Agency for Research (Agencia Estatal de Investigación) for the financial support as a Center of Excellence Severo Ochoa (CEX2020-001024-S/AEI/10.13039/501100011033), the Cellex Foundation for providing research facilities and equipment and the CERCA Programme from the Generalitat de Catalunya for their support on this research. This research has been supported by PREdICT, sponsored by AstraZeneca. This study has been co-funded by the European Regional Development Fund/European Social Fund 'A way to make Europe' (to R.P.L.), and by the Comprehensive Program of Cancer Immunotherapy and Immunology (CAIMI), funded by the Banco Bilbao Vizcaya Argentaria Foundation Foundation (FBBVA, grant 89/2017). R.P.L is supported by the "la Caixa" Foundation CaixaResearch Advanced Oncology Research Program, the Prostate Cancer Foundation (18YOUN19), a CRIS Foundation Talent Award (TALENT19-05), the FERO Foundation through the XVIII Fero Fellowship for Oncological Research, the Instituto de Salud Carlos III-Investigación en Salud (PI18/01395 and PI21/01019), the Asociación Española Contra el Cancer (AECC) (PRYCO211023SERR) and the Generalitat de Catalunya Agency for Management of University and Research Grants of Catalonia (AGAUR) (2023PROD00178). The project that gave rise to these results received the support of a fellowship from "la Caixa" Foundation (ID 100010434). The fellowship code is "LCF/BQ/PR22/11920010" (funding F.G, A.V., and A.G.). This research has received support from the Beatriu de Pinós Postdoctoral Program from the Secretariat of Universities and Research of the Department of Business and Knowledge of the Government of Catalonia, and the support from the Marie Sklodowska-Curie COFUND program (BP3, contract number 801370; reference 2019 BP 00182) of the H2020 program (to K.B.). M.P. is supported by the UKRI Future Leaders Fellowship MR/T020296/2. A.G. is supported by a Severo Ochoa PhD fellowship (PRE2022-102586). The authors are thankful to the Vall d'Hebron Radiology department and to the ASCIRES CETIR clinical team for their assistance with MRI acquisitions, and to the GE/Siemens clinical scientists for their support with diffusion sequence characterisation.

Competing interests

This study has received support by AstraZeneca. K.B. was a researcher at VHIO (Barcelona, Spain), and is now an employee of AstraZeneca (Barcelona, Spain). AstraZeneca was not involved in the acquisition and analysis of the data, interpretation of the results, or the decision to submit this article for publication in its current form.

Data and code availability

The MC simulation framework is made freely available as a GitHub repository at the permanent address: https://github.com/radiomicsgroup/dMRIMC. The repository includes rich sets of synthetic signals that can be used to inform model fitting. The code for simulation-informed fitting is freely available as part of *BodyMRITools* at the permanent address: https://github.com/fragrussu/bodymritools (script *mri2micro_dictml.py*). The *ex vivo* mouse data will be soon released on the Radiomics Group GitHub web site (https://radiomicsgroup.github.io/data). Researchers interested in early access can contact the corresponding author Raquel Perez-Lopez (rperez@vhio.net). The *in vivo* human data cannot be made freely available at this stage due to ethical restrictions.

References

[Aboitiz et al., 1992] Aboitiz, F., Scheibel, A. B., Fisher, R. S., and Zaidel, E. (1992). Fiber composition of the human corpus callosum. *Brain research*, 598(1-2):143–153.

[Andersson et al., 2003] Andersson, J. L., Skare, S., and Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–888.

[Assaf et al., 2008] Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y., and Basser, P. J. (2008). Axcaliber: A method for measuring axon diameter distribution from diffusion MRI. *Magnetic Resonance in Medicine*, 59(6):1347–1354.

[Attalla et al., 2021] Attalla, S., Taifour, T., Bui, T., and Muller, W. (2021). Insights from transgenic mouse models of pymt-induced breast cancer: recapitulating human breast cancer progression in vivo. *Oncogene*, 40(3):475–491.

[Bankhead et al., 2017] Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., and Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1):16878.

[Berry et al., 2018] Berry, D. B., Regner, B., Galinsky, V., Ward, S. R., and Frank, L. R. (2018). Relationships between tissue microstructure and the diffusion tensor in simulated skeletal muscle. *Magnetic Resonance in Medicine*, 80(1):317–329.

[Buizza et al., 2021] Buizza, G., Paganelli, C., Ballati, F., Sacco, S., Preda, L., Iannalfi, A., Alexander, D. C., Baroni, G., and Palombo, M. (2021). Improving the characterization of meningioma microstructure in proton therapy from conventional apparent diffusion coefficient measurements using monte carlo simulations of diffusion mri. *Medical Physics*, 48(3):1250–1261.

[Burcaw et al., 2015] Burcaw, L. M., Fieremans, E., and Novikov, D. S. (2015). Mesoscopic structure of neuronal tracts from time-dependent diffusion. *NeuroImage*, 114:18–37.

[Conlin et al., 2023] Conlin, C. C., Feng, C. H., Digma, L. A., Rodríguez-Soto, A. E., Kuperman, J. M., Rakow-Penner, R., Karow, D. S., White, N. S., Seibert, T. M., Hahn, M. E., et al. (2023). A multicompartmental diffusion model for improved assessment of whole-body diffusion-weighted imaging data and evaluation of prostate cancer bone metastases. *Radiology: Imaging Cancer*, 5(1):e210115.

[Cui et al., 2015] Cui, Y., Dyvorne, H., Besa, C., Cooper, N., and Taouli, B. (2015). Ivim diffusion-weighted imaging of the liver at 3.0 t: Comparison with 1.5 t. *European journal of radiology open*, 2:123–128.

[Doblas et al., 2013] Doblas, S., Wagner, M., Leitao, H. S., Daire, J.-L., Sinkus, R., Vilgrain, V., and Van Beers, B. E. (2013). Determination of malignancy and characterization of hepatic tumor type with diffusion-weighted magnetic resonance imaging: comparison of apparent diffusion coefficient and intravoxel incoherent motion–derived measurements. *Investigative radiology*, 48(10):722–728.

[Duval et al., 2019] Duval, T., Saliani, A., Nami, H., Nanci, A., Stikov, N., Leblond, H., and Cohen-Adad, J. (2019). Axons morphometry in the human spinal cord. *Neuroimage*, 185:119–128.

[Feng et al., 2021] Feng, C. H., Conlin, C. C., Batra, K., Rodríguez-Soto, A. E., Karunamuni, R., Simon, A., Kuperman, J., Rakow-Penner, R., Hahn, M. E., Dale, A. M., et al. (2021). Voxel-level classification of prostate cancer on magnetic resonance imaging: Improving accuracy using four-compartment restriction spectrum imaging. *Journal of Magnetic Resonance Imaging*, 54(3):975–984.

[Fieremans and Lee, 2018] Fieremans, E. and Lee, H.-H. (2018). Physical and numerical phantoms for the validation of brain microstructural MRI: A cookbook. *NeuroImage*, 182:39–61.

[Fieremans et al., 2017] Fieremans, E., Lemberskiy, G., Veraart, J., Sigmund, E. E., Gyftopoulos, S., and Novikov, D. S. (2017). In vivo measurement of membrane permeability and myofiber size in human muscle using time-dependent diffusion tensor imaging and the random permeable barrier model. *NMR in Biomedicine*, 30(3):e3612.

[Gardier et al., 2023] Gardier, R., Villarreal Haro, J. L., Canales-Rodríguez, E. J., Jelescu, I. O., Girard, G., Rafael-Patiño, J., and Thiran, J.-P. (2023). Cellular exchange imaging (cexi): Evaluation of a diffusion model including water exchange in cells using numerical phantoms of permeable spheres. *Magnetic Resonance in Medicine*, 90(4):1625–1640.

[Grussu et al., 2022] Grussu, F., Bernatowicz, K., Casanova-Salas, I., Castro, N., Nuciforo, P., Mateo, J., Barba, I., and Perez-Lopez, R. (2022). Diffusion MRI signal cumulants and hepatocyte microstructure at fixed diffusion time: Insights from simulations, 9.4T imaging, and histology. *Magnetic Resonance in Medicine*, 88(1):365–379.

[Guy et al., 1992] Guy, C. T., Cardiff, R. D., and Muller, W. J. (1992). Induction of mammary tumors by expression of polyomavirus middle t oncogene: a transgenic mouse model for metastatic disease. *Molecular and cellular biology*, 12(3):954–961.

[Hall and Alexander, 2009] Hall, M. G. and Alexander, D. C. (2009). Convergence and parameter choice for monte-carlo simulations of diffusion mri. *IEEE transactions on medical imaging*, 28(9):1354–1364.

[Hoffmann et al., 2023] Hoffmann, E., Gerwing, M., Niland, S., Niehoff, R., Masthoff, M., Geyer, C., Wachsmuth, L., Wilken, E., Höltke, C., Heindel, W. L., Hoerr, V., Schinner, R., Berger, P., Vogl, T., Eble, J. A., Maus, B., Helfen, A., Wildgruber, M., and Faber, C. (2023). Profiling specific cell populations within the inflammatory tumor microenvironment by oscillating-gradient diffusion-weighted MRI. *Journal for ImmunoTherapy of Cancer*, 11(3):e006092.

[Jiang et al., 2022] Jiang, X., Devan, S. P., Xie, J., Gore, J. C., and Xu, J. (2022). Improving mr cell size imaging by inclusion of transcytolemmal water exchange. *NMR in Biomedicine*, 35(12):e4799.

[Jiang et al., 2020a] Jiang, X., Dudzinski, S., Beckermann, K. E., Young, K., McKinley, E., J McIntyre, O., Rathmell, J. C., Xu, J., and Gore, J. C. (2020a). MRI of tumor T cell infiltration in response to checkpoint inhibitor therapy. *Journal for ImmunoTherapy of Cancer*, 8(1):e000328.

[Jiang et al., 2020b] Jiang, X., Xu, J., and Gore, J. C. (2020b). Mapping hepatocyte size in vivo using temporal diffusion spectroscopy MRI. *Magnetic Resonance in Medicine*, 84(5):2671–2683.

[Jones and Basser, 2004] Jones, D. K. and Basser, P. J. (2004). "squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted mr data. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 52(5):979–993.

[Kellner et al., 2016] Kellner, E., Dhital, B., Kiselev, V. G., and Reisert, M. (2016). Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic Resonance in Medicine*, 76(5):1574–1581.

[Khan et al., 2015] Khan, A. R., Cornea, A., Leigland, L. A., Kohama, S. G., Jespersen, S. N., and Kroenke, C. D. (2015). 3d structure tensor analysis of light microscopy data for validating diffusion mri. *Neuroimage*, 111:192–203.

[Kiselev, 2017] Kiselev, V. G. (2017). Fundamentals of diffusion mri physics. *NMR in Biomedicine*, 30(3):e3602.

[Le Bihan et al., 1986] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., and Laval-Jeantet, M. (1986). Mr imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407.

[Lee et al., 2021] Lee, H.-H., Fieremans, E., and Novikov, D. S. (2021). Realistic Microstructure Simulator (RMS): Monte Carlo simulations of diffusion in three-dimensional cell segmentations of microscopy images. *Journal of Neuroscience Methods*, 350:109018.

[Lee et al., 2020a] Lee, H.-H., Jespersen, S. N., Fieremans, E., and Novikov, D. S. (2020a). The impact of realistic axonal shape on axon diameter estimation using diffusion MRI. *NeuroImage*.

[Lee et al., 2020b] Lee, H.-H., Papaioannou, A., Kim, S.-L., Novikov, D. S., and Fieremans, E. (2020b). A time-dependent diffusion mri signature of axon caliber variations and beading. *Communications biology*, 3(1):354.

[Lee et al., 2024] Lee, H.-H., Tian, Q., Sheft, M., Coronado-Leija, R., Ramos-Llorden, G., Abdollahzadeh, A., Fieremans, E., Novikov, D. S., and Huang, S. Y. (2024). The effects of axonal beading and undulation on axonal diameter estimation from diffusion mri: Insights from simulations in human axons segmented from three-dimensional electron microscopy. *NMR in Biomedicine*, page e5087.

[Lemberskiy et al., 2018] Lemberskiy, G., Fieremans, E., Veraart, J., Deng, F.-M., Rosenkrantz, A. B., and Novikov, D. S. (2018). Characterization of Prostate Microstructure Using Water Diffusion and NMR Relaxation. *Frontiers in Physics*, 6:91.

[Morelli et al., 2023] Morelli, L., Palombo, M., Buizza, G., Riva, G., Pella, A., Fontana, G., Imparato, S., Iannalfi, A., Orlandi, E., Paganelli, C., and Baroni, G. (2023). Microstructural parameters from DW-MRI for tumour characterization and local recurrence prediction in particle therapy of skull-base chordoma. *Medical Physics*, 50(5):2900–2913.

[Nedjati-Gilani et al., 2017] Nedjati-Gilani, G. L., Schneider, T., Hall, M. G., Cawley, N., Hill, I., Ciccarelli, O., Drobnjak, I., Wheeler-Kingshott, C. A. G., and Alexander, D. C. (2017). Machine learning based compartment models with permeability for white matter microstructure imaging. *NeuroImage*, 150:119–135.

[Nilsson et al., 2010] Nilsson, M., Alerstam, E., Wirestam, R., Sta˚hlberg, F., Brockstedt, S., and Lätt, J. (2010). Evaluating the accuracy and precision of a two-compartment Kärger model using Monte Carlo simulations. *Journal of Magnetic Resonance*, 206(1):59–67.

[Nilsson et al., 2017] Nilsson, M., Lasič, S., Drobnjak, I., Topgaard, D., and Westin, C.-F. (2017). Resolution limit of cylinder diameter estimation by diffusion mri: The impact of gradient waveform and orientation dispersion. *NMR in Biomedicine*, 30(7):e3711.

[Novikov et al., 2019] Novikov, D. S., Fieremans, E., Jespersen, S. N., and Kiselev, V. G. (2019). Quantifying brain microstructure with diffusion mri: Theory and parameter estimation. *NMR in Biomedicine*, 32(4):e3998.

[O'connor et al., 2017] O'connor, J. P., Aboagye, E. O., Adams, J. E., Aerts, H. J., Barrington, S. F., Beer, A. J., Boellaard, R., Bohndiek, S. E., Brady, M., Brown, G., et al. (2017). Imaging biomarker roadmap for cancer studies. *Nature reviews Clinical oncology*, 14(3):169–186.

[P. Vangelderen et al., 1994] P. Vangelderen, D. Despres, P.C.M. Vanzijl, and C.T.W. Moonen (1994). Evaluation of Restricted Diffusion in Cylinders. Phosphocreatine in Rabbit Leg Muscle. *Journal of Magnetic Resonance, Series B*, 103(3):255–260.

[Palombo et al., 2019] Palombo, M., Alexander, D. C., and Zhang, H. (2019). A generative model of realistic brain cells with application to numerical simulation of the diffusion-weighted mr signal. *NeuroImage*, 188:391–402.

[Palombo et al., 2016] Palombo, M., Ligneul, C., Najac, C., Le Douce, J., Flament, J., Escartin, C., Hantraye, P., Brouillet, E., Bonvento, G., and Valette, J. (2016). New paradigm to assess brain cell morphology by diffusion-weighted MR spectroscopy in vivo. *Proceedings of the National Academy of Sciences*, 113(24):6671–6676.

[Palombo et al., 2023] Palombo, M., Valindria, V., Singh, S., Chiou, E., Giganti, F., Pye, H., Whitaker, H. C., Atkinson, D., Punwani, S., Alexander, D. C., and Panagiotaki, E. (2023). Joint estimation of relaxation and diffusion tissue parameters for prostate cancer with relaxation-VERDICT MRI. *Scientific Reports*, 13(1):2999.

[Panagiotaki et al., 2015] Panagiotaki, E., Chan, R. W., Dikaios, N., Ahmed, H. U., O'Callaghan, J., Freeman, A., Atkinson, D., Punwani, S., Hawkes, D. J., and Alexander, D. C. (2015). Microstructural Characterization of Normal and Malignant Human Prostate Tissue With Vascular, Extracellular, and Restricted Diffusion for Cytometry in Tumours Magnetic Resonance Imaging:. *Investigative Radiology*, 50(4):218–227.

[Panagiotaki et al., 2012] Panagiotaki, E., Schneider, T., Siow, B., Hall, M. G., Lythgoe, M. F., and Alexander, D. C. (2012). Compartment models of the diffusion MR signal in brain white matter: A taxonomy and comparison. *NeuroImage*, 59(3):2241–2254.

[Pilard et al., 2021] Pilard, C., Ancion, M., Delvenne, P., Jerusalem, G., Hubert, P., and Herfs, M. (2021). Cancer immunotherapy: it's time to better predict patients' response. *British journal of cancer*, 125(7):927–938.

[Rafael-Patino et al., 2020] Rafael-Patino, J., Romascano, D., Ramirez-Manzanares, A., Canales-Rodríguez, E. J., Girard, G., and Thiran, J.-P. (2020). Robust Monte-Carlo Simulations in Diffusion-MRI: Effect of the Substrate Complexity and Parameter Choice on the Reproducibility of Results. *Frontiers in Neuroinformatics*, 14:8.

[Reynaud, 2017] Reynaud, O. (2017). Time-dependent diffusion mri in cancer: tissue modeling and applications. *Frontiers in Physics*, 5:58.

[Shemesh et al., 2016] Shemesh, N., Jespersen, S. N., Alexander, D. C., Cohen, Y., Drobnjak, I., Dyrby, T. B., Finsterbusch, J., Koch, M. A., Kuder, T., Laun, F., et al. (2016). Conventions and nomenclature for double diffusion encoding nmr and mri. *Magnetic resonance in medicine*, 75(1):82–87.

[Szafer et al., 1995] Szafer, A., Zhong, J., and Gore, J. C. (1995). Theoretical model for water diffusion in tissues. *Magnetic resonance in medicine*, 33(5):697–712.

[Veraart et al., 2016] Veraart, J., Novikov, D. S., Christiaens, D., Ades-aron, B., Sijbers, J., and Fieremans, E. (2016). Denoising of diffusion MRI using random matrix theory. *NeuroImage*, 142:394–406.

[Veraart et al., 2020] Veraart, J., Nunes, D., Rudrapatna, U., Fieremans, E., Jones, D. K., Novikov, D. S., and Shemesh, N. (2020). Noninvasive quantification of axon radii using diffusion mri. *elife*, 9:e49855.

[Westin et al., 2016] Westin, C.-F., Knutsson, H., Pasternak, O., Szczepankiewicz, F., Özarslan, E., van Westen, D., Mattisson, C., Bogren, M., O'Donnell, L. J., Kubicki, M., et al. (2016). Q-space trajectory imaging for multidimensional diffusion mri of the human brain. *Neuroimage*, 135:345–362.

[White et al., 2014] White, N. S., McDonald, C. R., Farid, N., Kuperman, J., Karow, D., Schenker-Ahmed, N. M., Bartsch, H., Rakow-Penner, R., Holland, D., Shabaik, A., et al. (2014). Diffusion-weighted imaging in cancer: physical foundations and applications of restriction spectrum imaging. *Cancer research*, 74(17):4638–4652.

[Xu et al., 2020] Xu, J., Jiang, X., Li, H., Arlinghaus, L. R., McKinley, E. T., Devan, S. P., Hardy, B. M., Xie, J., Kang, H., Chakravarthy, A. B., et al. (2020). Magnetic resonance imaging of mean cell size in human breast tumors. *Magnetic resonance in medicine*, 83(6):2002–2014.

[Xu et al., 2023] Xu, J., Xie, J., Semmineh, N. B., Devan, S. P., Jiang, X., and Gore, J. C. (2023). Diffusion time dependency of extracellular diffusion. *Magnetic Resonance in Medicine*, page mrm.29594.

[Yan, 2021] Yan, L.-J. (2021). Folic acid-induced animal model of kidney disease. *Animal models and experimental medicine*, 4(4):329–342.

[Zhang et al., 2012] Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., and Alexander, D. C. (2012). Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4):1000–1016.

Appendix

Gamma distribution fit

For each substrate we fitted a Gamma distribution to the set of individual cell diameters $\{D_1, D_2, ...\}$ obtained from manual segmentations. The distribution is defined as

$$p(D; h, c) \;=\; \frac{1}{\Gamma(h)\, c^h}\, D^{h-1}\, e^{-D/c}.$$

Above, $\int_0^\infty p(D; h, c)\, dD \;=\; 1$, D is the generic cell diameter (units: μm), c is the scale parameter (units: μm), h is the shape parameter (dimensionless), and $\Gamma(z)$ is the Gamma function

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt, \quad \Re(z) > 0.$$

Animal models and *ex vivo* MRI and histology

This section summarises the procedures followed to obtain breast,spleen and kidney tissue in mice for *ex vivo* MRI.

Mice were housed at the Specific Pathogen-Free barrier area of the Vall d'Hebron Institute of Oncology (VHIO). All animal procedures were approved by the Animal Care unit and the Ethics Committee for Animal Experimentation (CEEA) of the Vall d'Hebron Research Institute (VHIR) and the Generalitat de Catalunya, and were performed according to the European legal framework for research animal use and bioethics. Animals were monitored daily and euthanised upon signs of humane endpoints. Two mouse models were used, generating breast, spleen and kidney samples. These were processed for further histological analyses, as part of ongoing studies at VHIO. A dMRI scan of the tissue was performed at room temperature before inclusion in paraffin for histology.

Briefly, collected tissues were fixed for 24 hours in buffered 4% formaldehyde, transferred to PBS solution and embedded in 1% agarose gel dissolved in PBS. Embedded samples were kept in PBS until dMRI. After dMRI, samples were transferred to 70% ethanol for 24 hours and then embedded in paraffin. The whole samples were sectioned and were stained with HE. HE staining was performed on 3 μm paraffin sections in a Robust carousel tissue stainer (Slee Medical) according to common methods. Digital images of the HE-stained sections were acquired on a Hamamatsu C9600-12 scanner (resolution: 0.45 μm). Breast and spleen samples where obtained from the mouse mammary tumour virus (MMTV) polyomavirus middle T antigen (PyMT) transgenic mouse model. Kidney samples were obtained from a folic acid-induced kidney injury mouse model.

MMTV-PyMT transgenic mouse model. The MMTV-PyMT FVB/NJ mouse strain [Guy et al., 1992] is commonly employed to mimic human breast cancer progression [Attalla et al., 2021]. The model relies on the MMTV long terminal repeat promoter, which drives the expression of the antigen of PyMT, a potent oncogene. These transgenic mice are viable despite loss of lactational ability, which is coincident with the transgene expression. Breast tumours arise in virgin and breeder females as well as in males starting from 9 weeks of age. Splenomegaly is also observed at the latter stages of the tumour growth. For this study, we used 4 MMTV-PyMT FVB/NJ female mice, which were euthanised by CO_2 asphyxiation at different time points to collect the following samples: non-cancerous breast and non-pathological spleen (2 weeks); a breast tumour at weeks 9, 11 and 14; an enlarged spleen (splenomegaly) at late stage cancer (14 weeks).

Folic acid-induced kidney injury. The folic acid-induced kidney injury mouse model is based on the fact that high doses of folic acid are toxic, despite being the same substance beneficial at low doses [Yan, 2021]. For this study, we used two male mice (C57BL/6 WT, approximately 9 weeks old), which were intra-peritoneally injected with a single dose of vehicle (300 mM $NaHCO_3$) or with folic acid (250 mg/kg). 30 days after the injection, mice were euthanised by CO_2 asphyxiation and the kidneys were collected for downstream processing.

Supplementary Material

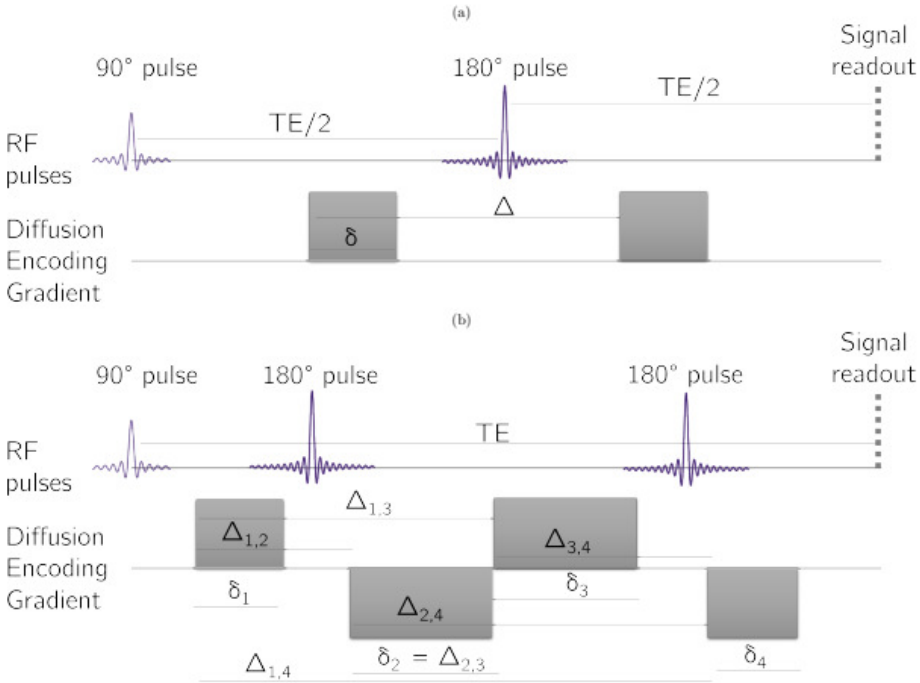


Figure S1: Illustration of the pulsed-gradient spin echo (PGSE, (a) on top) and diffusion-weighted (DW) twice-refocussed spin echo (TRSE, (b), on the bottom) sequences used in this study for both simulations and *ex vivo*/*in vivo* imaging.

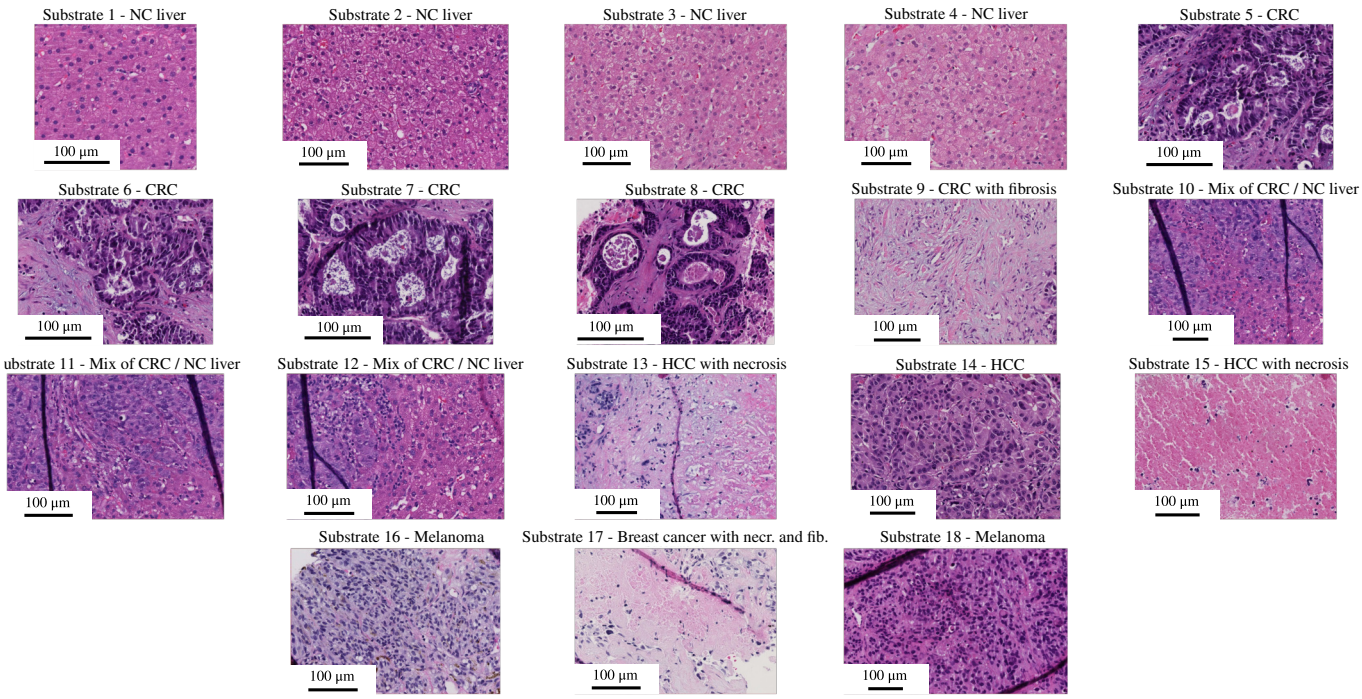


Figure S2: The dataset of histological patches from HE-stained biopsies that were manually segmented to reconstruct substrates for Monte Carlo diffusion simulations. NC: non-cancerous liver; CRC: colorectal cancer; HCC: hepatocellular carcinoma.

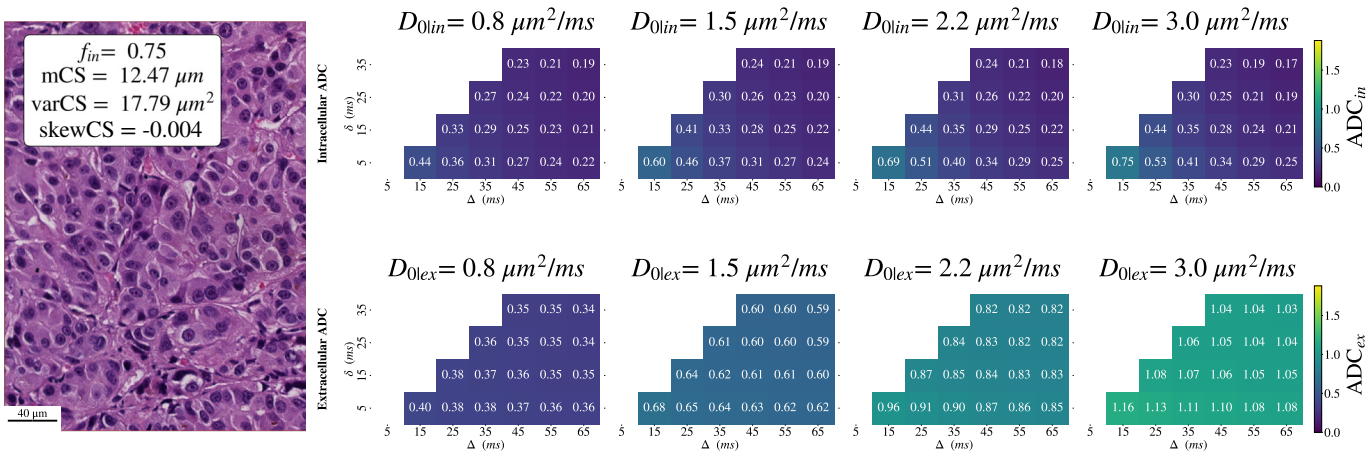


Figure S3: Plots visualising IC/EC ADC (ADC$_{in}$ and ADC$_{ex}$) as a function of the diffusion gradient duration δ and separation Δ and varying intrinsic IC/EC diffusivities $D_{0|in}$ and $D_{0|ex}$ for a HCC substrate (substrate 14). From left to right: substrate with properties, and ADC$_{in}$ and ADC$_{ex}$ for increasing values of $D_{0|in}$ and $D_{0|ex}$

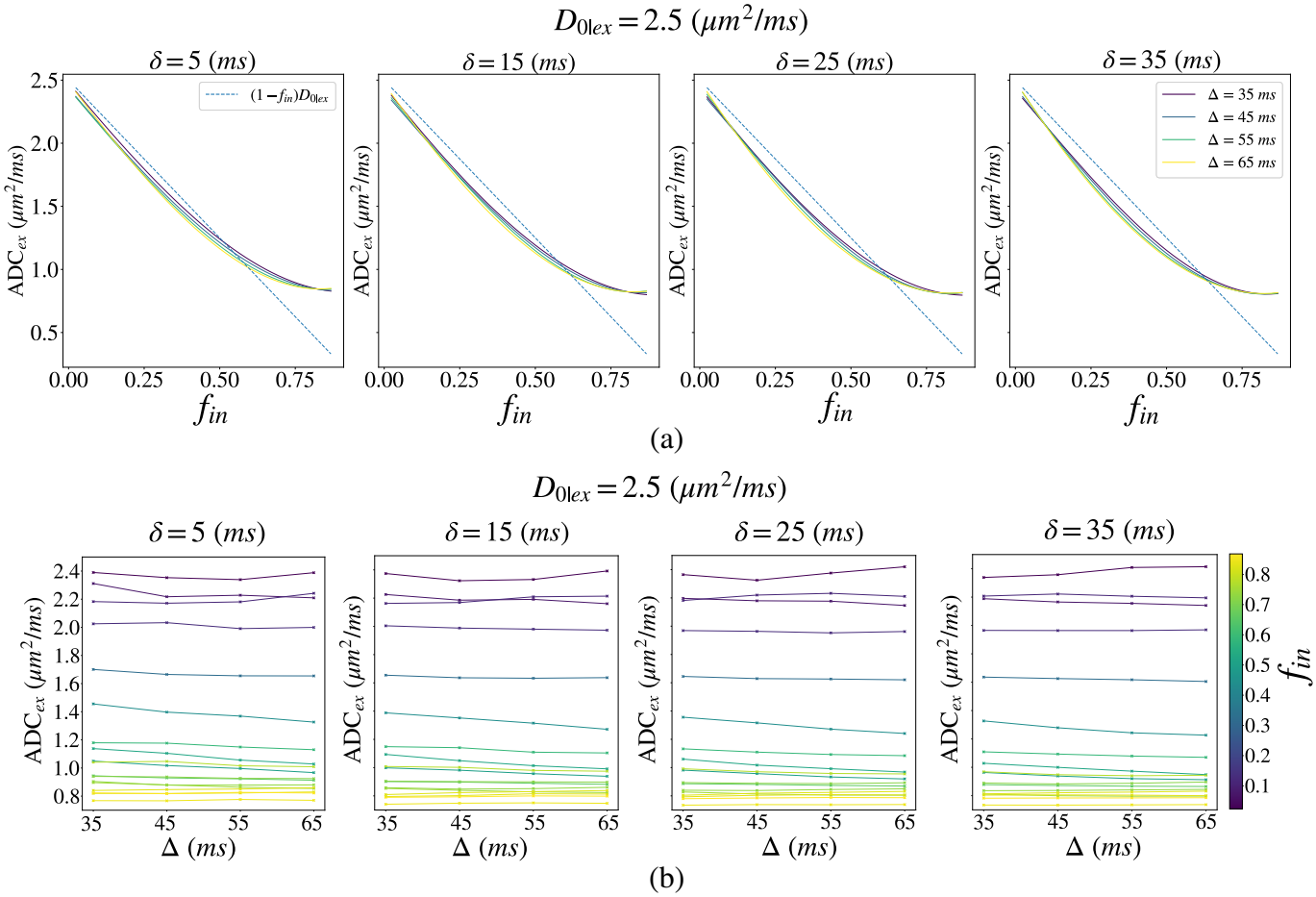


Figure S4: Plots showing the dependence of the EC ADC ADC$_{ex}$ with respect to f_{in} across substrates. Fig (a):From left to right, results from increasing values of the gradient duration δ are reported. In each plot, smooth curves are obtained by interpolating results from all substrates. These were coloured depending on the gradient separation Δ. Reference values from ADC$_{ex}/D_{0|ex} = (1 - f_{in})$ are also shown. To generate this figure, $D_{0|ex} = 2.5\ \mu m^2/ms$ was used. Fig (b): This plot shows the time dependence of $D_{0|ex}$. For $\Delta \in [35, 65]$ there is practically no t-dependence.

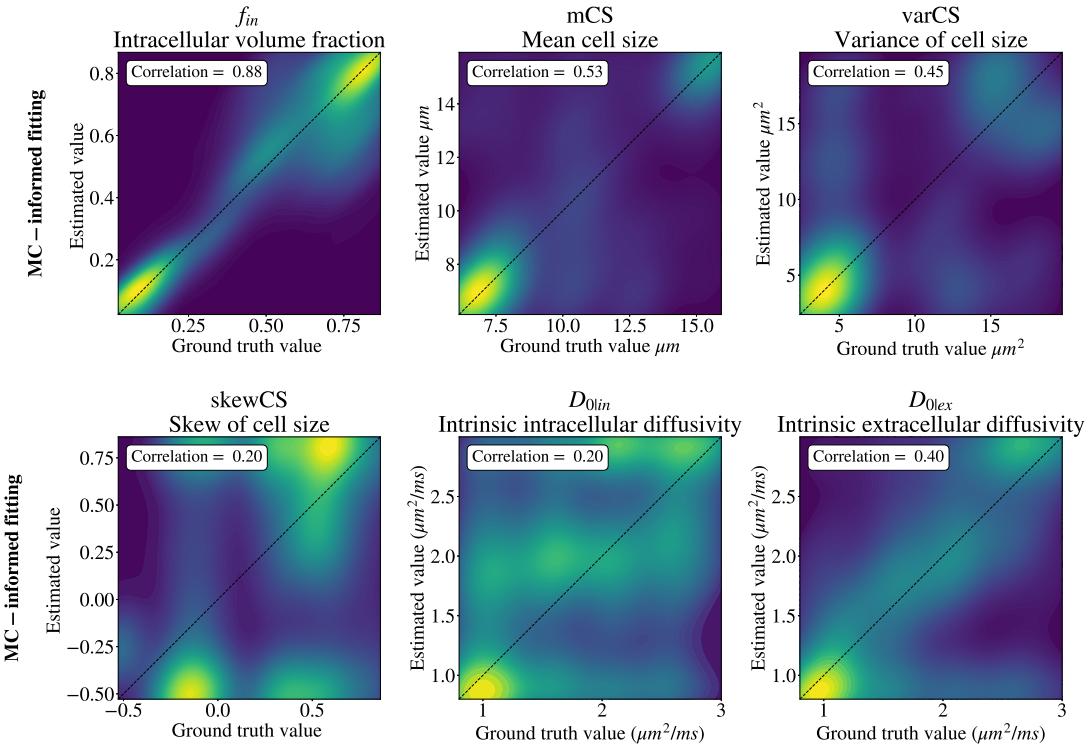


Figure S5: Scatter density plots between ground truth and estimated tissue parameters for MC-informed parameter estimation (forward model 2) and dMRI protocol PGSE-in. Each plot corresponds to a metric. From the top left corner, in clock-wise order: IC fraction f_{in}, mean CS index, variance of CS varCS, intrinsic EC diffusivity $D_{0|ex}$, intrinsic IC diffusivity $D_{0|in}$, skewness of CS distribution skewCS. The plots also include the identity line for reference, and the Pearson's correlation coefficient between ground truth and estimated parameter values. Results are shown for dMRI protocol *PGSE-in*.

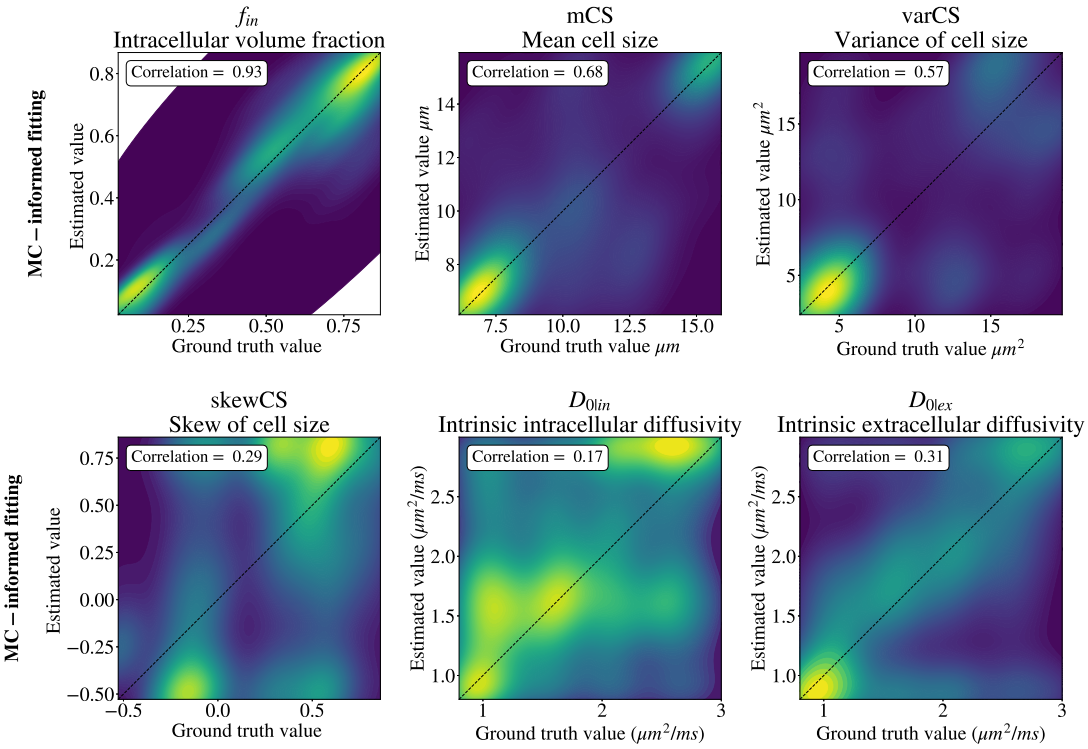


Figure S6: Scatter density plots between ground truth and estimated tissue parameters for MC-informed parameter estimation (forward model 2) and dMRI protocol TRSE. Each plot corresponds to a metric. From the top left corner, in clock-wise order: IC fraction f_{in}, mean CS index, variance of CS varCS, intrinsic EC diffusivity $D_{0|ex}$, intrinsic IC diffusivity $D_{0|in}$, skewness of CS distribution skewCS. The plots also include the identity line for reference, and the Pearson's correlation coefficient between ground truth and estimated parameter values. Results are shown for dMRI protocol *TRSE*.

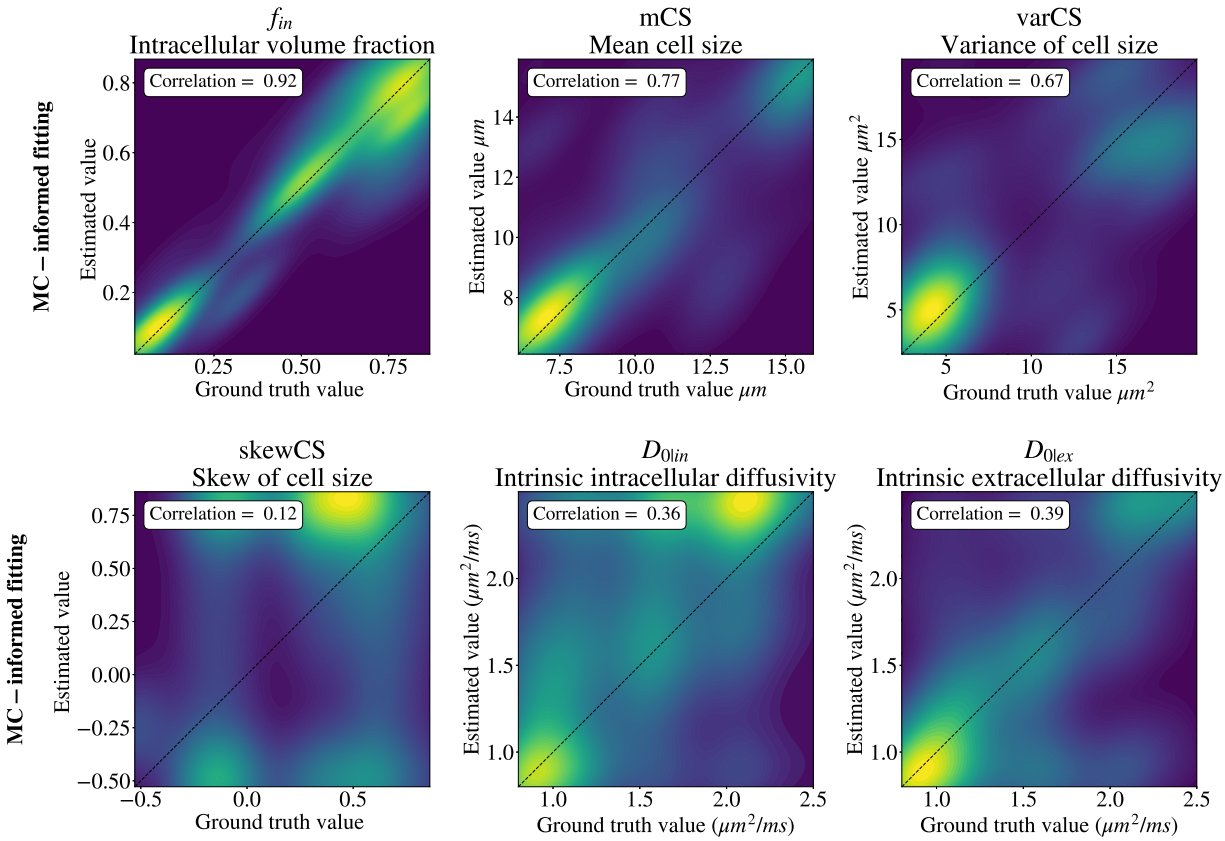


Figure S7: Scatter density plots between ground truth and estimated tissue parameters for MC-informed parameter estimation (forward model 2) and dMRI protocol PGSE-ex. Each plot corresponds to a metric. From the top left corner, in clock-wise order: IC fraction f_{in}, mean CS index, variance of CS varCS, intrinsic EC diffusivity $D_{0|ex}$, intrinsic IC diffusivity $D_{0|in}$, skewness of CS distribution skewCS. The plots also include the identity line for reference, and the Pearson's correlation coefficient between ground truth and estimated parameter values. Results are shown for dMRI protocol *PGSE-ex*.

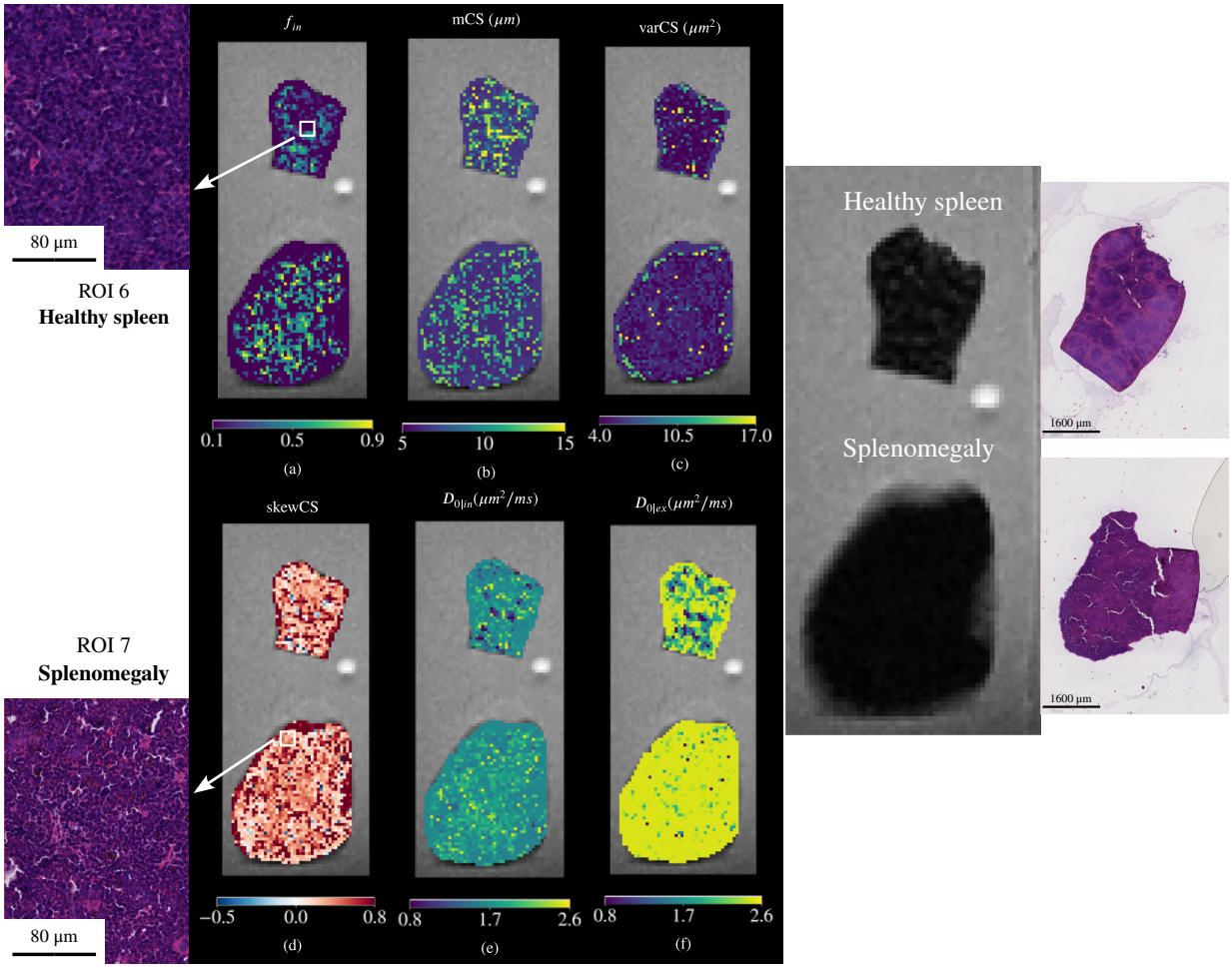


Figure S8: MRI and histological results as obtained on two mouse spleens scanned *ex vivo* on a 9.4T system. Right: $b = 0$ image, co-localised HE-stained section, and high-resolution histological patches. Left: parametric maps from forward model 2. First row: IC fraction f_{in} (a); mean CS index mCS (b); variance of CS varCS (c). Second row: skewness of the CS distribution skewCS (d); intrinsic IC diffusivity $D_{0|in}$ (e); intrinsic EC diffusivity $D_{0|ex}$ (f). For each metric, we show results on both samples: Normal spleen (top), splenomegaly (bottom).

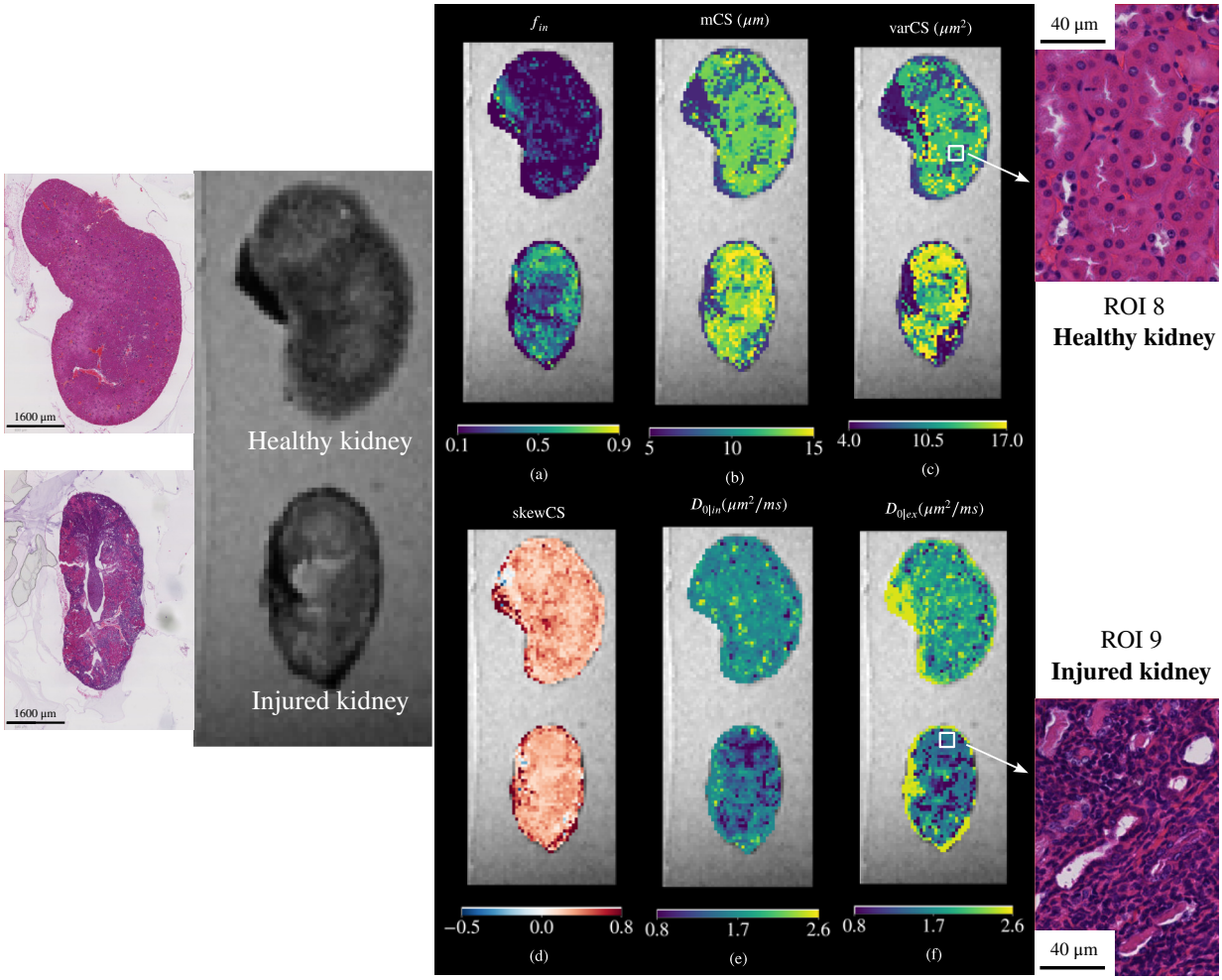


Figure S9: MRI and histological results as obtained on two mouse kidneys scanned *ex vivo* on a 9.4T system. Left: $b = 0$ image, co-localised HE-stained section, and high-resolution histological patches. Right: parametric maps from forward model 2. First row: IC fraction f_{in} (a); mean CS index mCS (b); variance of CS varCS (c). Second row: skewness of the CS distribution skewCS (d); intrinsic IC diffusivity $D_{0|in}$ (e); intrinsic EC diffusivity $D_{0|ex}$ (f). We show again results on both samples: Normal kidney (top), folic-acid induced kidney injury (bottom).

Patient information / MRI scanner	**Primary cancer / Biopsied area**	f_{in}	mCS (μm)	varCS (μm^2)	skewCS	$D_{0\|in}$ ($\mu m^2/ms$)	$D_{0\|ex}$ ($\mu m^2/ms$)
Case A, [61-65] y.o. F/ 1.5T system	Endometrial / Vagina	0.318 [0.103]	10.4 [2.3]	11.3 [3.6]	0.347 [0.300]	1.79 [0.39]	1.91 [0.50]
Case B, [81-85] y.o. F/ 1.5T system	Melanoma / Liver tumour	0.330 [0.169]	8.4 [2.3]	8.0 [3.2]	-0.064 [0.451]	2.18 [0.66]	2.57 [0.50]
Case C, [51-55] y.o. F/ 1.5T system	Ovarian / Liver tumour	0.245 [0.182]	7.4 [1.8]	6.4 [2.8]	0.052 [0.404]	1.95 [0.55]	2.52 [0.57]
Case D, [65-70] y.o. M/ 1.5T system	Melanoma / Liver tumour	0.392 [0.142]	8.9 [2.9]	9.9 [2.7]	-0.033 [0.446]	2.24 [0.69]	2.64 [0.58]
Case E, [75-80] y.o. F/ 3T system	Colorectal / Liver tumour	0.264 [0.129]	10.7 [2.9]	9.5 [3.5]	0.221 [0.168]	1.79 [0.33]	2.02 [0.66]
Case F, [65-70] y.o. M/ 3T system	Colorectal / Liver tumour	0.175 [0.084]	7.7 [2.5]	6.5 [3.5]	0.162 [0.264]	1.81 [0.34]	2.49 [0.59]
Case G, [61-65] y.o. F/ 3T system	Ovarian / Liver tumour	0.214 [0.094]	9.9 [2.9]	8.5 [3.8]	0.180 [0.163]	1.81 [0.23]	1.98 [0.59]
Case H, [51-55] y.o. M/ 3T system	Colorectal / Liver tumour	0.2 [0.112]	9.2 [2.8]	7.5 [4.1]	0.106 [0.220]	1.92 [0.21]	2.10 [0.68]
Case I, [81-85] y.o. F/ 3T system	Endometrial / Liver tumour	0.279 [0.162]	10.2 [3.1]	8.9 [4.1]	0.183 [0.268]	1.86 [0.42]	2.30 [0.72]

Table S1: Mean and standard deviation of metrics from forward model 2 computed from *in vivo* dMRI scans, within manually segmented tumours. The table reports results from 9 cancer patients suffering from advanced solid tumours and participating in ongoing imaging studies. The table also includes summary clinical information, namely: patients' age, sex, primary cancer, biopsy location, and MRI scanner used for *in vivo* imaging. Patients' age is given in 5 year intervals.