Histology-informed liver diffusion MRI: biophysical model design and demonstration in cancer immunotherapy

Francesco Grussu[1]*, Kinga Bernatowicz[1], Marco Palombo[2,3], Irene Casanova-Salas[1], Ignasi Barba[1,4], Sara Simonetti[1], Garazi Serna[1], Athanasios Grigoriou[1,5], Anna Voronova[1,5], Valezka Garay[6], Juan Francisco Corral[7,8], Marta Vidorreta[9], Pablo García-Polo García[10], Xavier Merino[7,8], Richard Mast[7,8], Núria Roson[7,8], Manuel Escobar[7,8], Maria Vieito[11], Rodrigo Toledo[1], Paolo Nuciforo[1], Joaquin Mateo[11], Elena Garralda[11], Raquel Perez-Lopez[1]*

*Corresponding authors. Email: fgrussu@vhio.net (F.G.), rperez@vhio.net (R.P.L.)

[1] Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Barcelona Hospital Campus, Spain
[2] Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, United Kingdom
[3] School of Computer Science and Informatics, Cardiff University, Cardiff, UK
[4] University of Vic - Central University of Catalonia (UVic-UCC), Vic, Spain
[5] Department of Biomedicine, Faculty of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain
[6] PET/MR Unit, CETIR-ASCIRES, Barcelona, Spain
[7] Department of Radiology, Hospital Universitari Vall d'Hebron, Barcelona, Spain
[8] Institut de Diagnòstic per la Imatge (IDI), Barcelona, Spain
[9] Siemens Healthineers, Madrid, Spain
[10] GE HealthCare, Madrid, Spain
[11] Medical Oncology Service, Vall d'Hebron Barcelona Hospital Campus, Vall d'Hebron Institute of Oncology (VHIO), Spain

Abstract

Innovative diffusion Magnetic Resonance Imaging (dMRI) models enable *in vivo* mapping of biologically meaningful properties such as cell size, potential biomarkers in cancer. However, while cancers frequently spread to the liver, models tailored for liver applications and easy to deploy in the clinic are still sought. We tackle this unmet need by delivering a practical and clinically viable liver dMRI modelling framework. Through direct comparison of candidate dMRI approaches in mouse models and cancer patients' data, we select a model of intra-cellular diffusion fitted to highly diffusion-weighted images, as it provides the strongest radiological-histological correlates. We demonstrate the potential application of the proposed model in cancer immunotherapy, stratifying the risk of progression based on baseline tumour cell size/density from dMRI. This result, heretofore unreported and not achievable with standard dMRI indices (e.g., apparent diffusion coefficient), suggests that our approach may become a useful tool for precision imaging in oncology.

Introduction

Routine clinical Magnetic Resonance Imaging (MRI) focusses on visualising macroscopic pathological anatomical features, as presence of tumours. Nonetheless, MRI also offers the possibility of measuring statistics of key biological properties within each pixel of a three-dimensional (3D) scan – known as *voxel*. This approach, usually referred to as quantitative MRI (qMRI) (*1*), involves the acquisition of multiple images, each featuring a different contrast, which are then analysed jointly with a mathematical model. qMRI provides promising metrics, which could become quantitative biomarkers complementing the qualitative assessment by the expert radiologist (*2*). Within the qMRI landscape, diffusion MRI (dMRI) refers to a subset of methods that sensitise the signal to water diffusion with magnetic field gradients (*1*, *3*, *4*). Since diffusion in biological tissues is influenced by the cellular microenvironment where diffusion takes place, dMRI ultimately enables the indirect estimation of properties at the micrometric scale (*5*), such as the size of cells restricting water (*6*, *7*), thus bridging the gap between macroscopic and microscopic imaging. Promising dMRI approaches have been demonstrated in virtually any region of the body, including but not limited to: brain (*5*), spinal cord (*8*), prostate (*6*), or breast (*9*).

Innovative dMRI techniques are also urgently needed in abdominal imaging, as in liver MRI (*10*, *11*). The liver is a frequent site for cancer metastasisation (*12*), and liver tumours are common targets for treatment response assessment in oncology. However, current response criteria such as RECIST (*13*) have limitations, in that they rely on MRI or computed tomography (CT) merely to measure tumour size, without accounting for changes under therapy at the cellular level. Novel dMRI metrics could enable the characterisation of cancer microenvironment over large areas, shedding light on the composition of tumours that cannot be biopsied. The new readouts could also provide information on tumour heterogeneity, relevant in the development of treatment resistance (*14*, *15*), and could better stratify patients eligible for treatments as immunotherapy (*16*), given the challenge of predicting which patients can practically benefit from this innovative class of drugs (*17*). This would be a major advancement in oncology, as it may allow for personalised treatment planning, reductions in sample sizes in clinical trials, and ultimately improve patient outcomes (*18*).

The most recent biophysical dMRI techniques describe the non-vascular liver tissue signal as the sum of contributions from intra-cellular and extra-cellular water (*19–21*). While these models provide promising readouts of tumour microenvironments (*22*), their practical use in real-world settings is made unfeasible by i) the high number of dMRI images (and hence long scan time) required for model fitting, and by ii) the availability of specialised dMRI acquisitions (*23*), beyond standard examinations that MRI manufactures make available in the scanner console. In this study we aim to tackle this unmet need by delivering a practical liver dMRI signal model that is truly feasible in hospital settings, i.e., on 1.5T or 3T systems, with scan time that does not exceed 15 minutes, and using vendor-provided dMRI sequences. With this objective in mind, we embraced the latest "histology-informed" dMRI development framework, which is based on informing signal model design with co-localised histological information. The framework has shown promise in delivering dMRI approaches with unprecedented fidelity to cytoarchitecture (*24*, *25*), ensuring the biological specificity of the newly designed techniques (*26*).

In this article, we aimed to identify a practical mathematicl models that maximises the agreement of dMRI estimates of metrics such as cell size to their underlying biological counterparts as seen in histology. We analysed a rich data set of dMRI scans and co-

localised hematoxylin and eosin (HE)-stained images from excised mouse livers and patients' liver biopsies. We used this data to select the model maximising radiological-histological correlations, corroborating results with computer simulations. Afterwards, we demonstrated the clinical utility of the designed approach in one exemplificatory response assessment task, in patients suffering from advanced solid tumours and treated with immunotherapy. This involved the stratification of the probability of progression under treatment from baseline dMRI scans, as this is still a urgent need, given the lack of robust predictors of treatment response for this class of drugs (*17*). In summary, our study delivers a liver dMRI approach that offers metrics with high fidelity to histopathology, and which is feasible in the clinic. The proposed method, based on a signal model consisting of a single compartment of restricted, intra-cellular diffusion fitted to highly diffusion-weighted (DW) images, identified which patients progress faster given baseline dMRI scans acquired before starting immunotherapy. These results, while exploratory and requiring confirmation by future studies, suggest that our dMRI framework could provide complementary information to current standard-of-care imaging, playing a key role in oncology research and practice.

Results

Overview: data set

Fig. 1 illustrates the data used in this study. We will refer to data obtained in fixed mouse livers as *preclinical*. We will instead refer to data obtained in cancer patients as *clinical*.

Preclinical data consists of pulsed gradient spin echo (PGSE) DW MRI scans of seven fixed mouse livers, performed *ex vivo* on a 9.4T Bruker system. It also includes whole-organ HE-stained sections, obtained at known radiographic position. We studied the livers of mice sacrificed as part of xenograft model development in prostate cancer studies. Six had been implanted with biopsies of prostate cancer patients, while one had not had any implantation. While the livers from the implanted mice did not grow any tumours, they feature a variety of pathologies that generated three unique histopathological phenotypes (Fig. S1). The liver from the mouse with no implantation instead features normal liver structures, and we will refer to it as *Control*. Of the six implanted cases, two also show normal liver tissue, with representation of all normal hepatic structures. We will refer to these two cases as Pat_{NA1} and Pat_{NA2} (patient biopsy implantation, but normal appearing). Another case exhibits generalised necrosis and diffuse acute and chronic inflammation surrounding necrotic areas, with presence of occluded thrombotic vessels. This specimen will be identified as Pat_{nec} (patient biopsy implantation, with necrosis). Finally, three specimens feature an immature, lymphoproliferative process, with various degrees of infiltration of small, lymphoid, atypical cells with abundant mitosis, which infiltrate portal vessels and sinusoidal capillaries, but without producing tumours. These cases will be referred to as Pat_{inf1} to Pat_{inf3} (patient biopsy implantation, with lymphoid cell infiltration).

We obtained clinical data on cancer patients suffering from advanced solid tumours, participating in an ongoing imaging study at the Vall d'Hebron Institute of Oncology (VHIO) of Barcelona (Spain). The study involves the acquisition of MRI data, alongside clinical and biological information (e.g., biopsies), in patients potentially eligible for a phase I immunotherapy trial. We included data from 33 patients with liver malignancies (mean/std of age: 62.91/12.34 year; 16 male, 17 female). dMRI was based on diffusion-weighted (DW) echo planar imaging (EPI) scans performed with clinical systems, namely: Twice-Refocussed Spin Echo (TRSE) DW-EPI on a 1.5T Siemens Prisma system; PGSE DW-EPI on a 3T GE SIGNA Pioneer system. We also obtained digitised HE-stained images, from

biopsies taken from one of the liver tumours. Biopsies were available for 18 patients, and were collected at baseline, immediately after dMRI (before starting immunotherapy). Of the 33 patients, clinical outcome was available for 30 patients, who effectively entered the immunotherapy trial after screening. For 15 of them, a biopsy was also available. The outcome was in the form of standard progression-free survival (PFS).

We used coupled dMRI-histology to design an optimal dMRI signal model (N = 25), while coupled dMRI-clinical outcome for a clinical demonstration (immunotherapy response assessment, N = 30), as illustrated in Fig. 1.

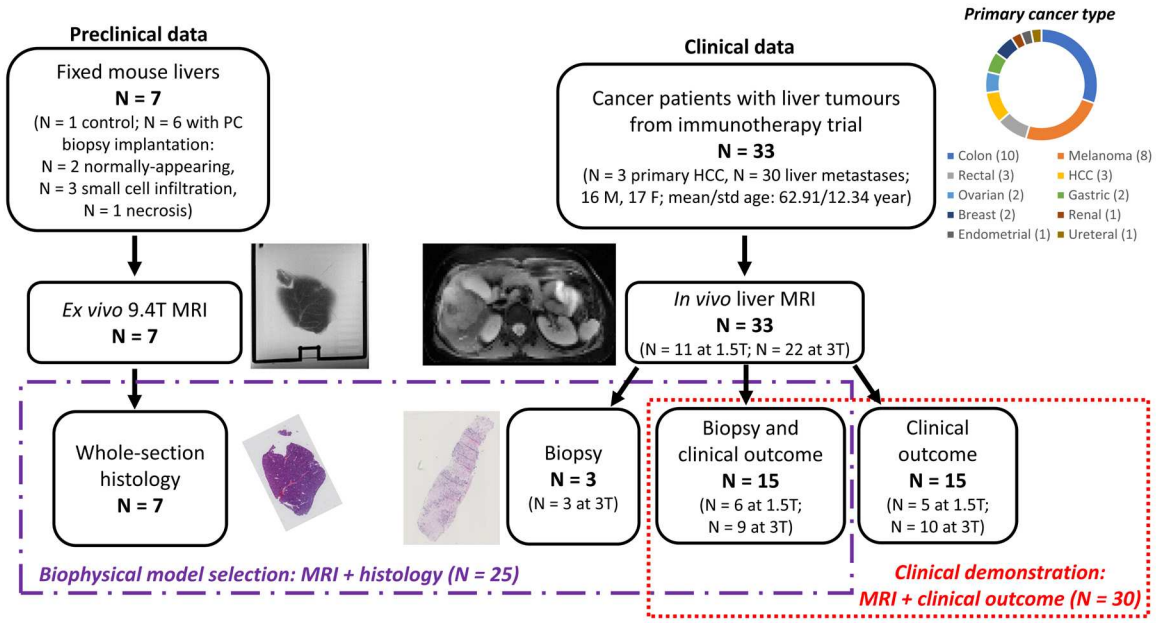

Fig. 1. Illustration of the liver MRI and histology data used in the study. Our data set consisted of preclinical and clinical data. The preclinical data encompasses dMRI scans of seven fixed livers from mice (six implanted with tissue from biopsies of patients suffering from prostate cancer; one without any implantation). We scanned the livers *ex vivo* on a 9.4T system, and obtained HE histological sections at known position. The clinical data includes *in vivo* liver dMRI scans performed on 33 patients suffering from advanced liver tumours. Scans were performed on clinical 1.5T and 3T MRI systems. For 18 patients, HE-stained material from a biopsy taken from one of the imaged liver tumours was also available. 30 out of 33 patients effectively participated in a phase I immunotherapy trial, and clinical outcome was available as PFS. We used a total sample size of N = 25 for dMRI model design and of N = 30 for response assessment. In the figure, PC standard fort prostate cancer and HCC for hepatocellular carcinoma.

Overview: dMRI signal models

We continue by presenting the different liver dMRI biophysical models examined in this article. All models account for restricted diffusion inside spherical cells and hindered diffusion in the extra-cellular space – a common modelling framework in body dMRI (*6*, *19*, *22*). Models can be grouped into two families (Fig. 2.A; see Material and Methods for implementation details).

The first family of models is more general, in that it does not make any assumption on which is higher between intra-cellular and extra-cellular ADC (referred to as ADC_I and ADC_E from now on). It includes the following models:

i. *Diff-in-exTD*: it accounts for restricted intra-cellular diffusion within spheres, modelling cells (*6*), and hindered diffusion in the extra-cellular space, with diffusion time dependence (TD) (*27*) in both intra-cellular and extra-cellular spaces (*28*). Note that the diffusion time is describes for how long water molecules can sense cellular barriers before the MR image is acquired.
ii. *Diff-in-ex*: as previous model *Diff-in-exTD*, but neglecting TD in the extra-cellular space. Popular techniques such as IMPULSED (*19*) or VERDICT (*20*) are practical implementations of this model, with intra-/extra-cellular diffusivities fixed to *ad hoc* values.

Conversely, the second family of models explicitly assumes that $ADC_E > ADC_I$, and assumption made in related dMRI techniques (e.g., Restriction Spectrum Imaging in the prostate (*29*); power-law axon radius mapping in the brain (*7*)). This family includes:

i. *Diff-in-exTDFast*: equivalent to *Diff-in-exTD* but ensuring that $ADC_E > ADC_I$.
ii. *Diff-in-exFast*: equivalent to *Diff-in-ex* but again ensuring that $ADC_E > ADC_I$.
iii. *Diff-in*: a model where it is hypothesised that the signal is dominated by intra-cellular water contributions (or, in other words, that the extra-cellular signal is negligible due to fast diffusion).

Overview: dMRI metrics

We fitted all models to the dMRI data, obtaining voxel-wise estimates of key tissue parameters, namely: *volume-weighted mean cell size* (vCS_{MRI}, expressed in µm) and *intra-cellular signal fraction* (F_{MRI}, dimensionless). We combined those two into a surrogate metric of *cell density per unit volume* ($CD_{MRI} = F_{MRI}/vCS_{MRI}^3$, expressed in cell mm^{-3}), as shown in other studies (*20*). For reference, we benchmarked our dMRI metrics against routine ADC (in µm^2 ms^{-1}) and apparent diffusion excess kurtosis K (dimensionless) from diffusion kurtosis imaging (DKI) (*30*). These are popular dMRI indices sensitive to cancer cellularity, which can be computed from short dMRI acquisitions (*20*, *31*). The details of the fitting procedures can be found in the Materials and Methods.

We processed HE-stained histological data with automatic cell detection (*32*) to derive histological counterparts of vCS_{MRI} and F_{MRI} at known radiographic location (details in the Materials and Methods section). The histological metrics were: histological volume-weighted mean cell size (vCS_{histo}, in µm), intra-cellular area fraction (F_{histo}, dimensionless), and cell density per unit area (CD_{histo}, in cell mm^{-2}). The direct comparison between vCS_{MRI} and vCS_{histo} and between F_{MRI} and F_{histo} enables histology-informed dMRI model selection (Fig. 2B).

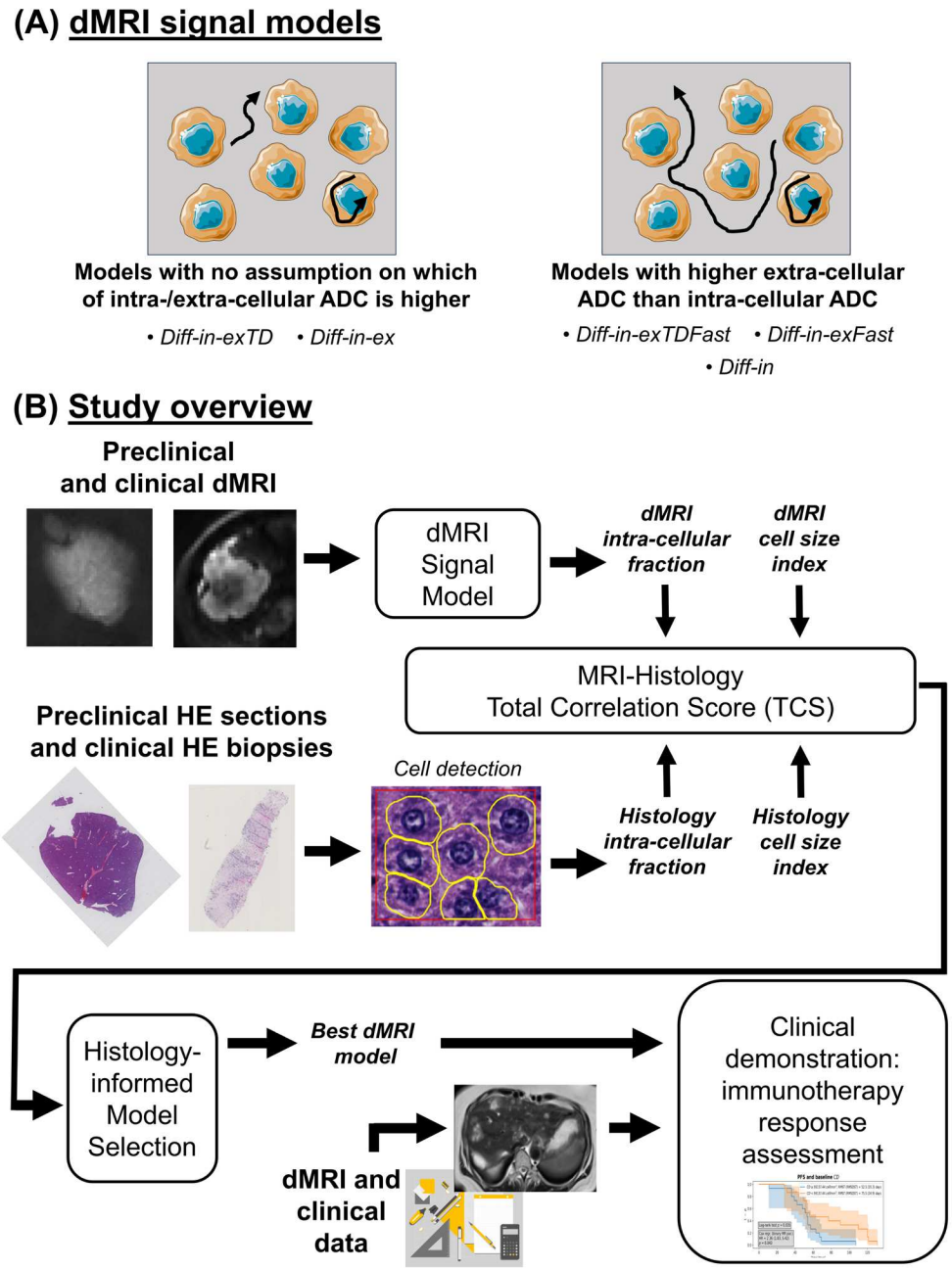

Fig. 2. Description of the dMRI signal models and study overview. (**A**), top: cartoon illustrating the two families of dMRI models considered in this study, consisting of 1) models with no assumption of which of intra/extra-cellular ADC is higher, and 2) models where the extra-cellular ADC is hypothesised to be higher than the intra-cellular ADC. (**B**), bottom: study overview. We analysed dMRI data from fixed mouse livers (preclinical data) and from cancer patients imaged *in vivo* (clinical data) to derive estimates of intra-cellular fraction and of cell size. In parallel, we processed histological material from the same tissues (whole-liver sections for the preclinical mouse data; biopsies from one of the imaged tumours for the clinical data), and derived the histological counterparts of such dMRI metrics. We compared dMRI and histological cell size and intra-cellular fraction to select the dMRI model featuring the best fidelity to histology. The utility of the model was then demonstrated in immunotherapy response assessment *in vivo*. In Fig. 2, pictures from Servier Medical Art have been used. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0).

A one-pool model of intra-cellular diffusion provides the most histologically meaningful metrics

Fig. 3A summarises the different dMRI models, while Fig. 3B reports values of the MRI-Histology *Total Correlation Score* (TCS) for all models. TCS measures the overall correlation between histological and radiological readouts of cell size and intra-cellular fraction. It was computed by summing Pearson's correlation coefficients between vCS_{MRI} and F_{MRI} with their respective histological counterparts vCS_{histo} and F_{histo} (see Materials and Methods). Higher values of TCS point towards stronger histological-radiological correlation. Negative correlations reduce TCS, so they are penalised.

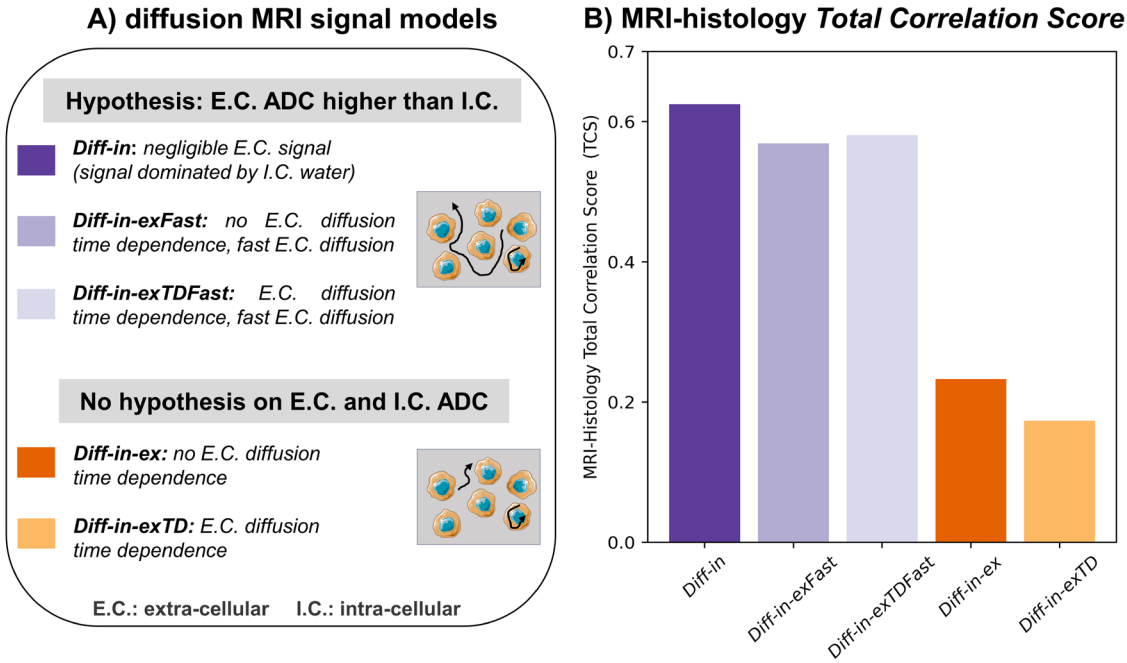

Fig. 3. Biophysical dMRI signal model selection based on the MRI-histology Total Correlation Score (TCS). (**A**): panel summarising the salient differences between the biophysical dMRI models compared in this study. Models can be divided in two families, i.e.: i) models where it is hypothesised that the extra-cellular ADC is higher than the intra-cellular ADC, and ii) models with no hypothesis on which, between intra-/extra-cellular ADC is higher. Violet shades are used for the first family (models *Diff-in*, *Diff-in-exFast* and *Diff-in-exTDFast*), while orange shades for the second family (models *Diff-in-ex* and *Diff-in-exTD*). (**B**): values of TCS for all models, as obtained by fitting them on high b-value images (b > 1800 s/mm^2 in the fixed mouse livers; b > 900 s/mm^2 *in vivo*). We performed model selection using a sample size of N = 25 (see Fig. 1). In Fig. 3, pictures from Servier Medical Art have been used. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0).

The bar plot in Fig. 3B highlights that dMRI models in which that extra-cellular ADC is higher than the intra-cellular one ($ADC_E > ADC_I$), shown in violet shades, provide consistently higher TCS values than models that do not make such an assumption (orange shades). All models assuming fast extra-cellular diffusion provide similar TCS. However, we observe the highest TCS for model *Diff-in*. *Diff-in* is, in practice, a compact representation where the extra-cellular DW signal is neglected compared to the intra-

cellular one, due to the much faster diffusion occurring in the former. In other words, *Diff-in* assumes that the DW signal can be described as dominated by intra-cellular water, characterised by diffusion restriction within cells of spherical shape. Note that Fig. 3 refers to TCS values obtained by fitting dMRI models only to hih b-value images (see Materials and Methods), as this provided the highest TCS figures. Fig. S2 reports TCS for model fitting performed to the whole set of diffusion images. In this case, TCS is lower, and again, models where $ADC_E > ADC_I$ provide higher TCS than models that do not make this assumption. In Fig. S2, we observe the highest TCS for model *Diff-in-exFast*.

Fig. S3 reports rankings according to additional model selection criteria. These were: the *Histology Fidelity Criterion* (HFC), measuring the sum of absolute errors in histological intra-cellular fraction F and cell size vCS estimation via MRI, and the *Bayesian Information Criterion* (BIC). BIC (*33*), commonly used in dMRI model development (*34*, *35*), quantifies the overall model fitting quality (penalizing model complexity), without accounting for histological information. Lower HFC, as well as lower BIC, imply better model performance. Fig. S3 reports the number of times, in percentage term, that a model provides the overall lowest HFC and BIC across our sample of N = 25 MRI-histology cases. Results essentially confirm rankings seen on TCS: models hypothesising $ADC_E > ADC_I$ are selected more frequently than models that do not do, according to HFC. The model *Diff-in* is the most selected model according to both BIC and HFC (fig. S3.B; fitting to high b-value images). Fig. S4 splits HFC and BIC rankings depending on the MRI scanner (9.4T for *ex vivo* mouse liver imaging; 1.5T or 3T for *in vivo* clinical imaging). In all cases, models with $ADC_E > ADC_I$ (*Diff-in*, *Diff-in-exFast*, *Diff-in-exFastTD*) are selected more frequently than models *Diff-in-ex* and *Diff-in-ex-TD*. When fitting is performed only on high b-value images, *Diff-in* is again the most selected model according to both BIC and HFC in all cases. Data S1 reports all TCS, HFC and BIC rankings for the complete set of MRI-histology observations.

Computer simulations confirm model selection from actual MRI measurements

We performed Monte Carlo computer simulations to corroborate the model selection performed on *ex vivo* and *in vivo* dMRI data. The simulations consisted in generating synthetic dMRI signals according to the protocols implemented in the three MRI scanners used in this study. We synthesised signals for a substrate made of packed spherical cells (Fig. S5), a common geometric tissue model used for body dMRI imaging (*6*, *19*, *22*, *23*), and performed model selection on such syntethic signals as illustrated above (additional details in the Materials and Methods). Supplementary Tables S1, S2, and S3 report TCS, HFC and BIC rankings obtained on simulated MRI signals. Results confirm that model *Diff-in* enables the best estimation of cell size and intra-cellular fraction among the candidate models in the majorities of cases, thus confirming results from *ex vivo* and *in vivo* MRI.

Our proposed approach: fitting of a one-compartment model of intra-cellular diffusion to high b-value images

In view of all model rankings reported above for *ex vivo*, *in vivo* and synthetic dMRI signals, we propose to *fit a one-compartment model of restricted intra-cellular diffusion within spherical cells to high b-values images* ($\gtrapprox$ 1800 s/mm^2 *ex vivo*, $\gtrapprox$ 900 s/mm^2 *in vivo*). This model, referred to as *Diff-in* in this article, is hence our recommended implementation.

Cell size and density estimates from the proposed dMRI model correlate with histology

In the following, we report on the Pearson's correlation coefficient r between metrics from the proposed *Diff-in* model with their direct histological counterparts. We consider correlations to be weak, moderate, and strong when $|r| < 0.4$, $|r| \geq 0.4$ but $|r| < 0.6$, and $|r| \geq 0.6$ respectively.

Table 1 reports r between metrics from the proposed model *Diff-in* with their histological counterparts. For reference, we also report correlation coefficients between more standard DKI metrics ADC and K. The correlation between *Diff-in* metrics F_{MRI} (intra-cellular fraction), vCS_{MRI} (cell size index) and CD_{MRI} (cell density) with their counterparts F_{histo}, vCS_{histo} and CD_{histo} are respectively weak ($r = 0.19$ between F_{MRI} and F_{histo}), moderate ($r = 0.44$ between vCS_{MRI} and vCS_{histo}) and strong ($r = 0.70$ between CD_{MRI} and CD_{histo}). The weak correlation between F_{MRI} and F_{histo} can be explained, at least in part, with the fact that F_{MRI} is a signal fraction, rather than an actual volume/area fraction (unlike F_{histo}), i.e., it is influenced by T2/T1 differences between intra-cellular and residual extra-cellular signals (*36*). Moreover, F_{MRI} estimation can be biased by unaccounted exchange between intra-cellular and extra-cellular water (*21*, *37*), a biophysical characteristic that is not accounted for in our signal models. Conversely, the much higher correlations between vCS_{MRI} and vCS_{histo} and between CD_{MRI} and CD_{histo} suggest that vCS_{MRI} and CD_{histo} may be biologically-specific tools to monitor the cytoarchitecture of the liver *in vivo*.

Table 1 also reports correlation coefficients between standard dMRI ADC and kurtosis K with histological F_{histo}, vCS_{histo} and CD_{histo}. Both ADC and K exhibit significant, moderate correlations with histological properties. For example, negative/positive correlation of ADC/K with cell density CD_{histo} ($r = -0.47$ and 0.43 respectively) are seen – a result entirely consistent with previous studies (*38*, *39*). Significant correlations are also seen with F_{histo} ($r = 0.40$, $p = 0.048$ between K and F_{histo}), again in line with known literature: ADC and K are sensitive to the underlying tissue microstructure. Nevertheless, they are also unspecific, being surrogate metrics difficult to interpret, which conflate different histopathological characteristics into a single number.

Fig. S6 and Fig. S7 show Pearson's correlation coefficients for all possible paird of metrics, in the form of correlation matrices. The figures report correlations for all models and all fitting strategies (fitting to the whole set of dMRI images, or only to those acquired at high b-value). Correlations among dMRI metrics are seen, as for example a strong negative correlation between CD_{MRI} and vCS_{MRI} ($r = -0.84$ for model *Diff-in-exFast* fitted at high b-value). This finding, which indicates that tighter cell packings per unit volume are achieved with smaller cells, is biophysically plausible, being mirrored by histological CD_{histo} and vCS_{histo} ($r = -0.88$ between CD_{histo} and vCS_{histo}). Other weak-to-moderate correlations are seen, e.g., between K and vCS_{MRI} ($r = -0.47$) and CD_{MRI} ($r = 0.38$), which agree with the correlations observed between K and histological vCS_{histo} ($r = -0.31$) and CD_{histo} ($r = 0.43$). In general, metrics from dMRI models where $ADC_E > ADC_I$ (models *Diff-in-exTDFast*, *Diff-in-exFast*, *Diff-in*) show stronger correlations with their histological counterparts, as compared to models without such an assumption (*Diff-in-exTD* and *Diff-in-ex*). We observe the strongest dMRI-histology correlations for model *Diff-in* fitted to high b-value images. Fitting models to high b-values images provides stronger dMRI-histology correlations in all cases for models where $ADC_E > ADC_I$ (*Diff-in-exTDFast*, *Diff-in-exFast*, *Diff-in*), while it does not necessarily do for the other models (*Diff-in-exTD* and *Diff-in-ex*).

Table 1. Correlation between dMRI metrics and histological metrics. The table reports Pearson's correlation coefficients r and corresponding p-values p of dMRI metrics F_{MRI} (intra-cellular fraction), vCS_{MRI} (volume-weighted cell size index) and CD_{MRI} (cell density per unit volume) with their histological pairs (F_{histo}, vCS_{histo} and CD_{histo} respectively) for the selected dMRI model (*Diff-in*, fitted to high b-value images). The table also reports correlation coefficients between routine ADC and K from DKI and each of F_{histo}, vCS_{histo} and CD_{histo}. The sample size was N = 25, so that that $p < 0.05$ if $|r| > 0.3961$. When $p < 0.05$, grey shadowing is used.

dMRI technique	Histology F_{histo}	Histology vCS_{histo}	Histology CD_{histo}
Model *Diff-in*	With F_{MRI}: $r = 0.19$; $p = 0.37$	With vCS_{MRI}: $r = 0.44$; $p = 0.029$	With CD_{MRI}: $r = 0.70$; $p = 0.0001$
Routine DKI	With ADC: $r = -0.28$; $p = 0.18$	With ADC: $r = 0.49$; $p = 0.014$	With ADC: $r = -0.47$; $p = 0.017$
	With K: $r = 0.40$; $p = 0.048$	With K: $r = -0.31$; $p = 0.13$	With K: $r = 0.43$; $p = 0.033$

Metrics from the proposed dMRI model reveal intra-/inter-tumour characteristics

Fig. 4 shows maps from the proposed dMRI model *Diff-in*, as well as corresponding histological metrics in 3 mouse livers. These are representative of the 3 histopathological phenotypes seen in our mouse data, namely: sample *Control*, for normal liver structures; sample Pat_{infl}, for generalised small cell infiltration; sample Pat_{nec}, for necrosis.

Fig. 4 shows excellent co-localisation between MRI slices and histology sections. The histological details reveal the higher cellularity of sample Pat_{infl} compared to *Control*, due to packing of small cells in between larger hepatocytes. They also show that sample Pat_{nec} features an alternation of areas with extremely low cell density (i.e., necrosis), with areas with higher abundance of cells. These qualitative trends are confirmed in the histological maps F_{histo}, vCS_{histo}, CD_{histo}: Pat_{infl} exhibits the highest F_{histo} and CD_{histo} as well as lowest vCS_{histo} of all samples; Pat_{nec} shows a patchy structure, with locally-low F_{histo} and CD_{histo}, indicating necrosis. *Control* sample maps are in physiologically plausible ranges, as for example intra-cellular fractions around 0.75 and cell sizes of the order of 25 μm (*40*, *41*). Maps F_{MRI}, vCS_{MRI} and CD_{MRI} replicate the contrasts seeing in their histological counterparts F_{histo}, vCS_{histo} and CD_{histo}. Fig. S8 shows standard dMRI metrics, namely ADC and kurtosis excess K, in the same mouse livers. Visual trends highlight that the higher cell density of sample Pat_{infl} translated to remarkably reduced ADC and increased K compared to the *Control*, is in line with known correlations with cellularity (*38*, *39*). Lastly, Fig. S9 shows F_{MRI}, vCS_{MRI} and CD_{MRI} maps from another model (*Diff-in-exFast*, providing the highest TCS for fitting performed to the whole set of images). Fig. S10 instead shows maps of the other metrics provided by models *Diff-in-exFast* and *Diff-in* (intrinsic cytosol diffusivity $D_{0,I}$ and asymptotic ADC_E, $D_{E,\infty}$). The figures highlight that overall, spatial trends seen in maps from the selected model *Diff-in* agree with those seen in *Diff-in-exFast*. Metrics $D_{0,I}$ and $D_{E,\infty}$ show limited between-sample contrast, and are difficult to validate histologically.

Table S4 and Data S2 report qualitative per-sample mean and standard deviation of all MRI and histology metrics, highlighting that F_{MRI} slightly underestimates F_{histo}, while vCS_{MRI} slightly overestimates the numerical value of vCS_{histo}. We speculate that the discrepancies may be due, at least in part, to unaccounted factors such as variability in intrinsic cell shape/cytosol diffusivity (*42*) or water exchange (*21*), and by the intrinsic challenge of relating histology metrics derived from 2D data to 3D MRI (*43*).

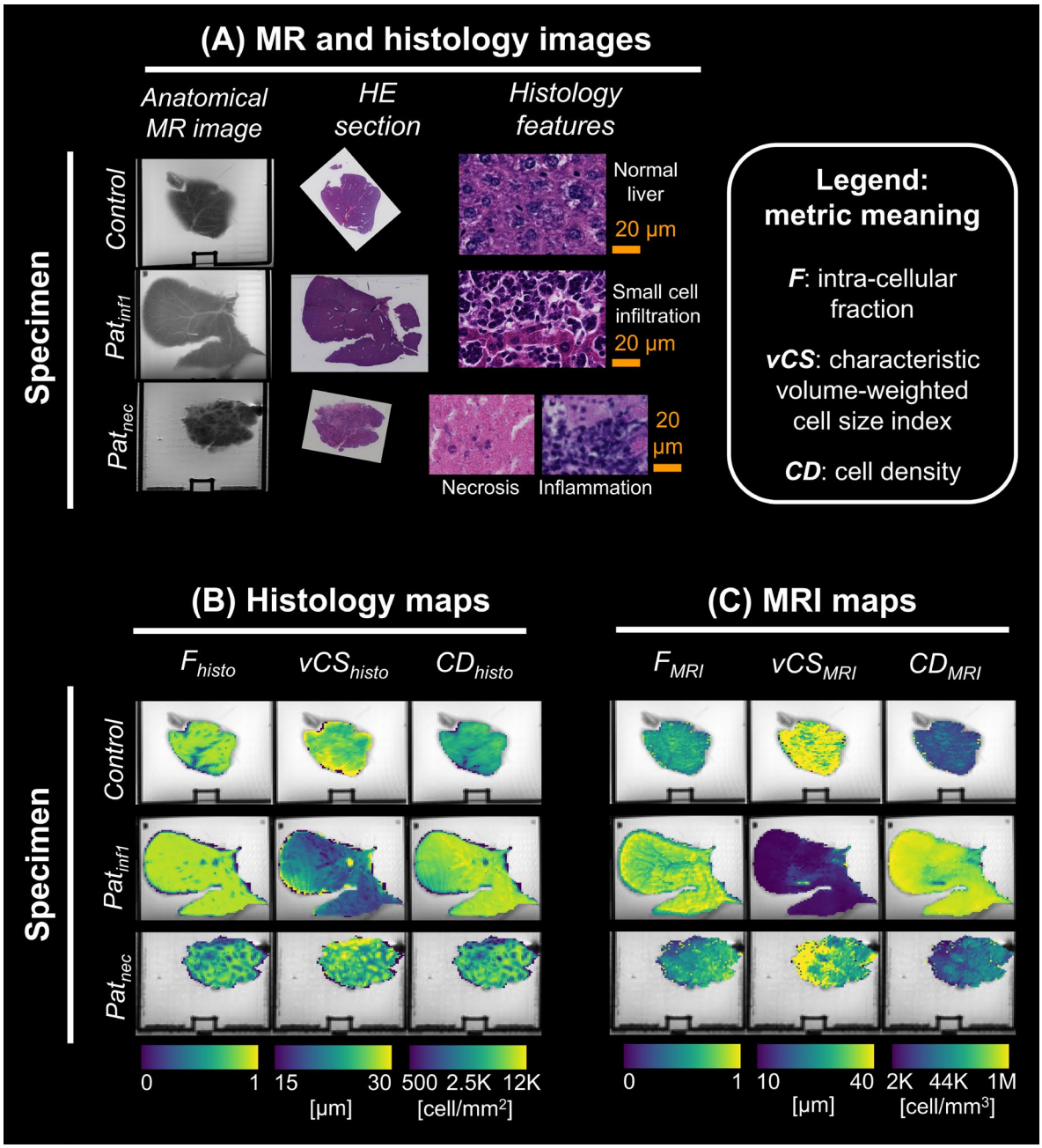

Fig. 4. Maps from the selected dMRI model *Diff-in* with their histological counterparts in the fixed mouse livers scanned at 9.4T *ex vivo*. The figure reports MRI and histology data for 3 specimens, representative of the 3 microstructural phenotypes observed in our *ex vivo* data set, namely: normal liver structures (*Control* case); a proliferative process, characterized by infiltration of small cells (Pat_{infl} case); necrosis and inflammation (Pat_{nec} case). For all specimens, the following is shown. **(A)**, *top left*: a high-resolution T2-w anatomical scan is shown next to the corresponding HE section, with histological details. **(B)**, *bottom left*: histological maps warped to the dMRI space (intra-cellular patch area

fraction F_{histo}; volume-weighted mean cell size index vCS_{histo}; cell density per unit patch area CD_{histo}). **(D)**, *bottom right*: dMRI maps F_{MRI}, vCS_{MRI} and CD_{MRI} from the selected dMRI signal model (model *Diff-in*, fitted to high *b*-value images, i.e., $b > 1800$ s/mm^2;).

Fig. 5 shows F_{MRI}, vCS_{MRI} and CD_{MRI} maps in patients, alongside biopsies. Assessment of histopathology highlights the variety of characteristics that can coexist within advanced solid tumours, e.g.: areas of fibrosis; areas with tightly packed cancer cells embedded within stromal scaffolding; extensive areas of necrosis. Notably, dMRI maps F_{MRI}, vCS_{MRI}, CD_{MRI} show contrasts that are plausible with the histopathological features seen on the biopsies, as for example: a core of low intra-cellular fraction F_{MRI} and low cell density CD_{MRI} in a breast cancer case, corresponding to necrosis; local areas of high F_{MRI} and low vCS_{MRI} in the HCCs, indicative of tightly packed cancer cells. Fig. S11 shows routine dMRI ADC and K in the same tumours. ADC and K trends are compatible with the histology seen on the biopsies, e.g., a core of high ADC and low K in the necrotic core of the BC tumour. Supplementary Fig. S12 shows F_{MRI}, vCS_{MRI} and CD_{MRI} from model *Diff-in-exFast*. Image contrasts match visually those seen in the same metrics from model *Diff-in* (the proposed approach), giving confidence of the overall robustness of the biophysical modelling framework. Fig. S13 shows intra-cellular cytosol diffusivity $D_{0,I}$ asymptotic ADC_E ($D_{E,\infty}$) in the same tumours. Their speckled appearance suggests that these metrics are difficult to measure accurately *in vivo* (*19*, *44*).

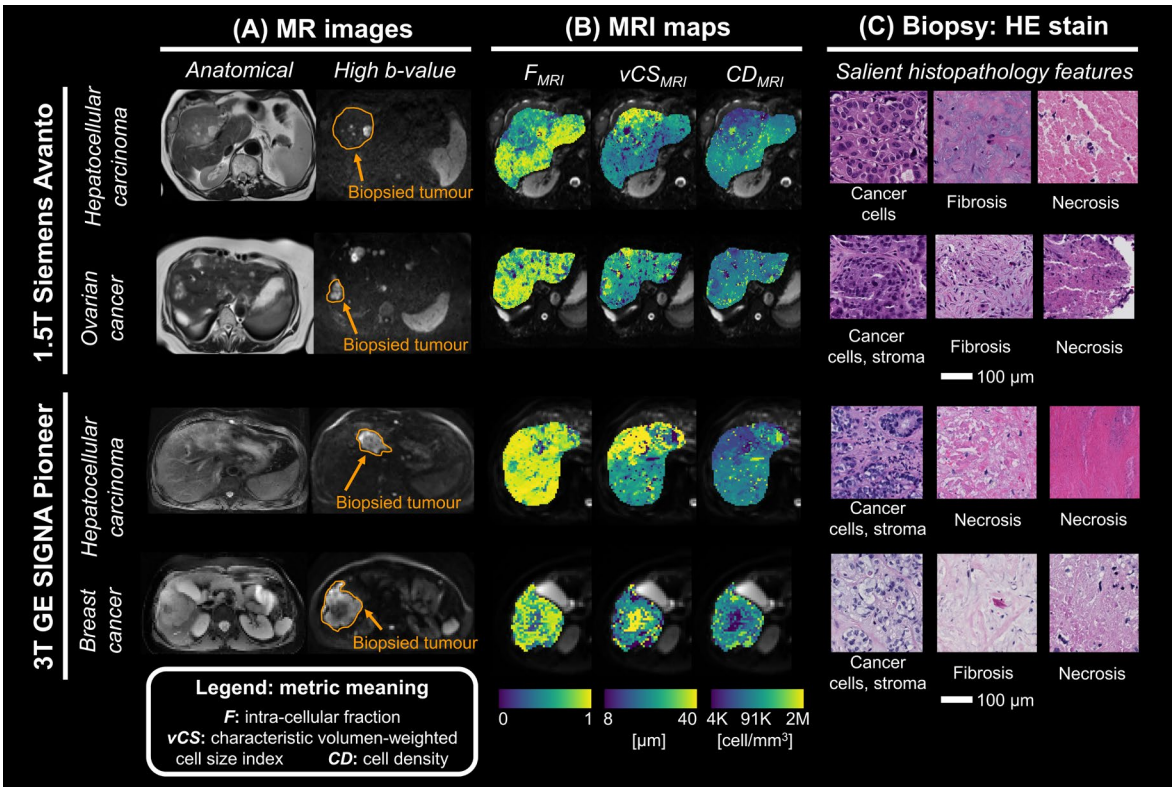

Fig. 5. Examples of maps from the proposed dMRI model *Diff-in* in liver tumours of patients scanned at 1.5T and 3T *in vivo*, with co-localised biopsies. MRI maps are shown in a biopsied liver tumour in two patients for each MRI scanner, arranged along rows. **(A)**: examples of slices from the high-resolution anatomical T2-w image and from a high b-value image, with biopsied tumour outlined. **(C)**: maps from the selected model (*Diff-in*, fitted to high b-value images $b > 900$ s/mm^2). From left to right: intra-cellular signal fraction F_{MRI}; volume-weighted mean cell size index vCS_{MRI}; cell density per unit volume CD_{MRI}. **(C)**: histological details from the HE-stained biopsy. For the 1.5T Siemens scanner (first and

second rows from top) we report: patient 6 (primary hepatocellular carcinoma) and patient 3 (liver metastases from ovarian cancer). For the 3T GE scanner (third and fourth rows from top) we report: patient 24 (primary hepatocellular carcinoma (HCC)) and patient 30 (liver metastases from breast cancer).

Metrics from the proposed dMRI model stratify the risk of cancer progression in immunotherapy

Finally, we demonstrate the potential utility of the proposed liver dMRI model in an exemplificatory response assessment task in immunotherapy. This consisted of stratifying the risk of cancer progression, given baseline values of dMRI metrics within liver tumours obtained immediately prior to treatment.

Fig. 6 reports response assessment based on metrics from the proposed model *Diff-in*, (intra-cellular fraction F, volume-weighted cell size index vCS, and cell density per unit volume CD). Panels on the left report results from Kaplan-Meier analysis, log-rank testing and Cox regression performed after binarising dMRI metrics as higher/lower than the median of the cohort. Panels on the right report instead results from Cox regressions assessing the continuous dependence of PFS on F, vCS and CD. We do not observe any statistically significant associations between baseline F_{MRI} and probability of progression. Conversely, we detect a dependence of the probability of progression on baseline vCS_{MRI} and CD_{MRI}. The Kaplan-Meier curves of the two groups differ significantly for both vCS_{MRI} (log-rank test: p = 0.047, Fig. 6C) and CD_{MRI} (log-rank test: p = 0.035, Fig. 6E). These differences correspond to statistically significant Hazard Ratios (HRs) from Cox regression (HR = 0.47, p = 0.050 for binarised vCS_{MRI}; HR = 2.36, p = 0.043 for binarised CD_{MRI}). These findings suggest that smaller values of cell size vCS_{MRI} or, alternatively, higher values of cell density CD_{MRI} within liver tumours at baseline, may be indicative of faster progression risk under immunotherapy. In practical terms, the risk of progression is about twice as high in patients whose baseline cell size vCS_{MRI} is smaller than the median vCS_{MRI}, compared to patients where it is larger. Similarly, the risk of progression is about twice as high in patients whose baseline cell density CD_{MRI} is higher than the median CD_{MRI}, compared to those where it is lower. Importantly, we found similar associations between the risk of progression and baseline vCS_{MRI} and CD_{MRI} when these two metrics are not binarised, but rather used as continuous predictors in Cox regression (right panels in Figure 6). In this latter case, the HR was statistically significant for vCS_{MRI} (HR = 0.65, p = 0.034, Fig. 6D), and it approached statistical significance for CD_{MRI} (HR = 1.40, p = 0.055, Fig. 6F). Noteworthily, these association are not confounded by age, sex or baseline tumour volume (Supplementary Table S5; HR = 0.59, p = 0.02 for vCS_{MRI}; HR = 1.65, p = 0.01 for CD_{MRI}).

Fig. 7 and supplementary Fig. S14, Fig. S15 and Fig. S16 report results for the response assessment based on all other dMRI metrics considered in this study (e.g., routine ADC and K in Fig. 7; cytosol diffusivity $D_{0,I}$ for model *Diff-in* in Fig. S14; vascular fraction f_V in Fig. S15; all metrics from model *Diff-in-exFast* in Fig. S16). While the estimated HRs for metrics vCS_{MRI}, F_{MRI} and CD_{MRI} from dMRI model *Diff-in-exFast* go in the same direction as those from model *Diff-in*, their association with the probability of progression is weaker, and only reaches borderline statistical significance for CD_{MRI} (HR = 1.53; p = 0.05; Table S5). This finding is likely due to the intrinsic higher variability (and hence, lower statistical power) of metrics from model *Diff-in-exFast* compared to model *Diff-in*, since the latter model has one parameter less to estimate, making it easier to fit. Importantly, baseline

values of routine ADC and K do not show any association with the probability of progression (Fig. 7 and Table S5).

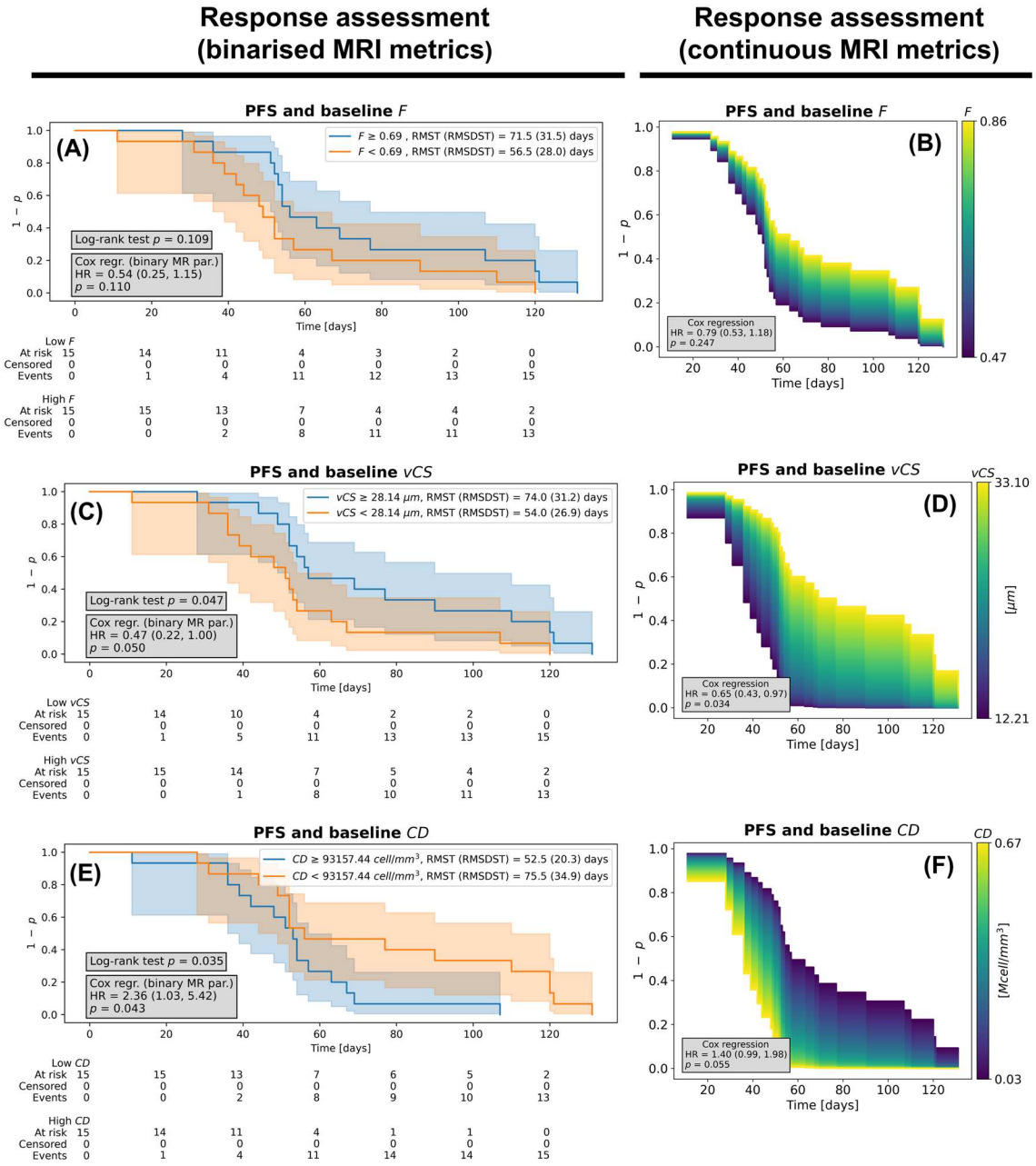

Fig. 6. Immunotherapy response assessment based on metrics from the proposed model *Diff-in* within liver tumours at baseline. This figure reports on the dependence of patients' progression-free survival (PFS) on the average value of F, vCS and CD within liver tumours at baseline (i.e., before starting immunotherapy), as obtained by fitting model *Diff-in* at high b-value. *Left*: Kaplan-Meier (KM) survival curves of two groups obtained by splitting patients based on baseline F (panel A), vCS (panel C) and CD (panel D) (lower/higher than the sample median). The grey panel reports the p-values of a log-rank sum test comparing the KM curves, and of a Cox regression based on the binarised MRI metric (with the corresponding hazard ratio (HR) estimate and 95% confidence interval). The legend reports the Restricted Mean Survival Time (RMST) and Restricted Standard Deviation of Survival Time (RSDST) for each KM curve. *Right*: results from univariate

Cox regression where the baseline F (panel B), vCS (panel D) and CD (panel F) is a continuous predictor of the survival. The panel shows how changes in baseline F, vCS and CD modulate the survival curve, given the HR estimated for each metric. In the grey box, the p-value and HR (with 95% CI) corresponding to the baseline MRI metric are reported. In all panels, the y-axis shows $1 - p$, with p being the probability of progression, while the x-axis shows the time to progression (in days). We performed the response assessment using a sample size of N = 30 (Fig. 1).

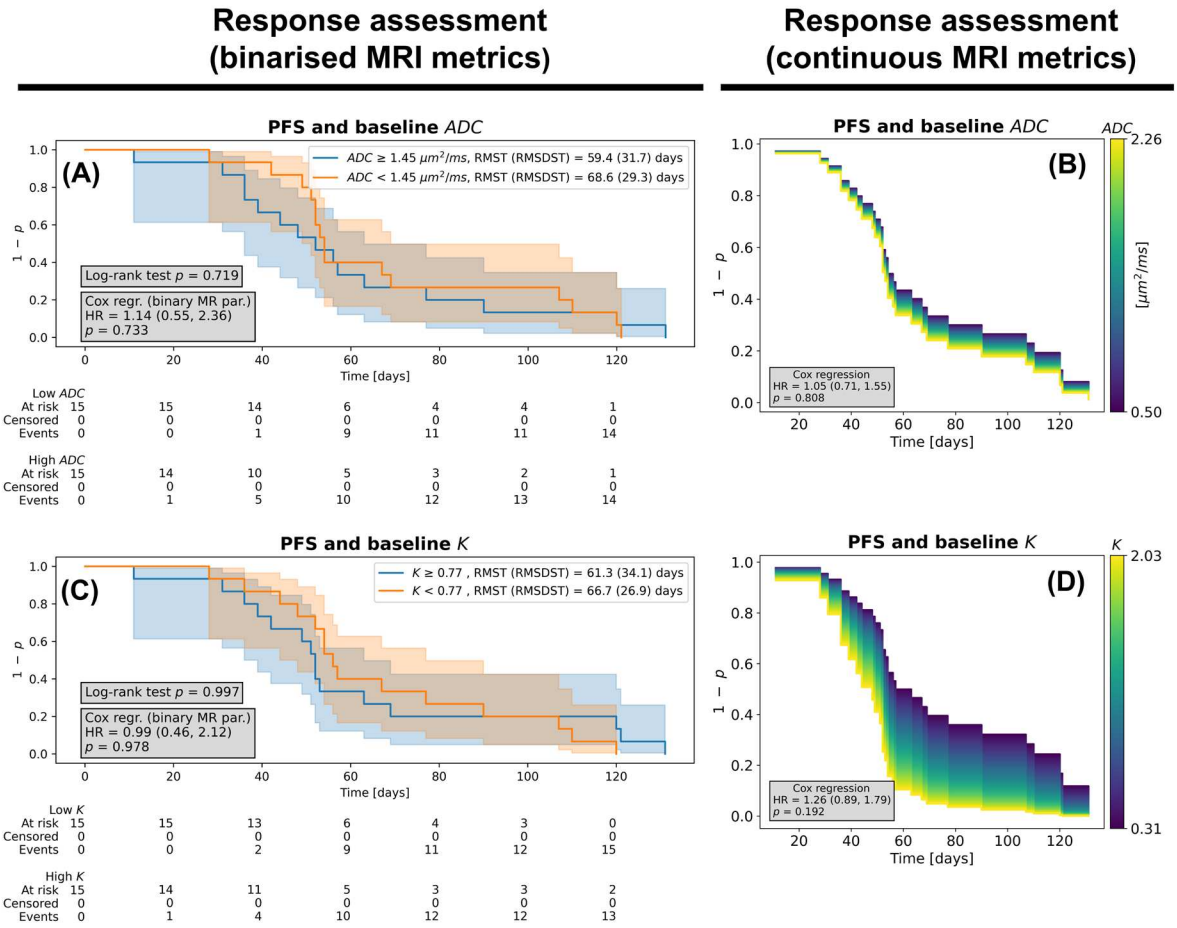

Fig. 7. Immunotherapy response assessment based on mean *ADC* and kurtosis *K* within liver tumours at baseline. This figure reports on the dependence of patients' progression-free survival (PFS) on the average value of *ADC* and *K* within liver tumours at baseline (i.e., before starting immunotherapy). *Left*: Kaplan-Meier (KM) survival curves of two groups obtained by splitting patients based on baseline *ADC* (panel A) and *K* (panel C) (lower/higher than the sample median). The grey panel reports the p-values of a log-rank sum test comparing the KM curves, and of a Cox regression based on the binarised MRI metric (with the corresponding hazard ratio (HR) estimate and 95% confidence interval). The legend reports the Restricted Mean Survival Time (RMST) and Restricted Standard Deviation of Survival Time (RSDST) for each KM curve. *Right*: results from univariate Cox regression where the baseline *ADC* (panel B) and *K* (panel D) is a continuous predictor of the survival. The panel shows how changes in baseline *ADC* and *K* modulate the survival curve, given the HR estimated for each metric. In the grey box, the p-value and HR (with 95% CI) corresponding to the baseline MRI metric are reported. In all panels, the y-axis

shows 1 – p, with p being the probability of progression, while the x-axis shows the time to progression (in days). We performed the response assessment using a sample size of N = 30 (Fig. 1).

Discussion

The latest liver dMRI signal models account for two key signal compartments, in that they attempt to disentangle intra-cellular and extra-cellular water contributions (*6*, *19*, *22*, *23*). This powerful approach enables the estimation of innovative tissue property maps, which may become useful biomarkers in oncological applications. However, the clinical deployment of advanced models of this type is hampered by their high number of unknown tissue parameters. Estimating accurately several tissue parameters at once requires large sets of MR images or, potentially, specialised dMRI acquisitions. This makes the use of such models impractical in real-world radiology settings (*19*, *45*), where scan time is limited and where most hospitals only have access to vendor-provided acquisitions. With this unsolved challenge in mind, this paper aims to deliver a practical implementation of a two-compartment dMRI signal model, tailored for liver imaging, which is truly feasible in the clinic. Through histology-informed model selection, we design a compact dMRI framework consisting of fitting a one-pool model of restricted intra-cellular diffusion to highly DW images. The framework provides pixel-by-pixel estimates of cell size and density that are correlated with the underlying histology. These metrics outperform well-establish ADC and diffusion kurtosis in patient stratification in immunotherapy, suggesting that they could become useful biomarkers in oncological applications.

To find the optimal dMRI signal implementation, we analysed co-localised dMRI and histology data (N = 25). These consisted of: i) MR images and HE sections of fixed mouse livers processed *ex vivo*, and ii) *in vivo* dMRI scans of cancer patients and HE-stained biopsies from one of the imaged liver tumours. We compared 5 signal models, each fitted according to two distinct strategies, and ranked them for their ability to accurately estimate intra-cellular fraction and cell size metrics, compared to reference histology. Rankings, confirmed by Monte Carlo computer simulations, unequivocally suggest the highest radiological-histological agreement is obtained by fitting a single-compartment model of diffusion (restricted diffusion within spherical cells – a model here referred to as *Diff-in*) to images acquired with b-values higher than approximately 900 s/mm^2 *in vivo* and 1800 s/mm^2 *ex vivo*. *Diff-in*, a model whose deployment is feasible in the clinic with scan times of under 15 minutes, and with default dMRI examinations made available by MRI vendors, provides cell size and density estimates that correlate moderately and strongly with histology, and which may therefore become biologically-specific biomarkers.

Interestingly, our central result points towards the fact that simpler models of diffusion (e.g., single- vs two-compartment), can provide the highest fidelity to the underlying histology, if deployed in appropriate measurement regimes. This is the case, in our settings, in presence of strong diffusion-weighting, which minimises extra-cellular signal contributions. Notably, our finding partially disagrees with some of the latest trends in body dMRI development. Recent literature has been characterised by the introduction of ever-complex biophysical models, featuring increasingly high numbers of tissue parameters to estimate (*21*, *36*, *37*, *46*). While more complex models better capture the true degrees of freedom on *in vivo* dMRI signals, they have the disadvantage of requiring longer and/or more sophisticated acquisition protocols to enable their fitting (*26*), a fact that negatively impacts on their clinical applicability. Conversely, our approach suggests that focussing on measurement regimes where the signal is dominated by intra-cellular diffusion may enable the

deployment of simpler models, which still suffice to capture salient features of tissue microstructure. This conclusion is in line, for example, to recent modelling approaches seen in brain imaging, which provide axon size indices from dMRI measurements where the extra-axonal signal has been suppressed (*7*, *47–49*). Notably, our model selection results can also be framed within the context of recent literature, since recently extra-cellular liver ADC values as high as 2.5-2.8 $\mu m^2\ ms^{-1}$ have been reported (*19*). Such a high ADC_E implies that the extra-cellular signal would decay to roughly 5% or less of its non-DW value even for b-values as low as approximately 1200 s/mm^2 ($\exp(-b\ ADC_E) \approx 0.05$ for $b = 1200\ s/mm^2 = 1.2\ ms^{-1}\ \mu m^2$ and $ADC_E = 2.5\ \mu m^2\ ms^{-1}$), making the use of a model of intra-cellular diffusion feasible as a proxy for the total DW signal (*7*), simplifying considerably parameter estimation (*44*, *50*).

Importantly, we conducted dMRI model selection using a variety of criteria. These were: the TCS, measuring the overall correlation between dMRI metrics and their histological counterparts; HFC, quantifying the accuracy in estimating histological properties via dMRI; and BIC, a common index of dMRI model quality of fit, which does not take into account any agreement with histology (*35*). In general, all criteria point towards the same direction, with model *Diff-in* fitted to high b-value images being the top-ranking model. We would also like to emphasise that the same liver dMRI model implementation (i.e., fitting model *Diff-in* on high b-value images) would have been delivered had we only looked at the *ex vivo* mouse data, or at *in silico* signals (Fig. S4, panels A and D). This fact gives confidence on the robustness and generalisability of our dMRI model selection, justifying our choice of using 15 *in vivo* dMRI scan for both histology-informed model design and for the clinical demonstration in immunotherapy.

Another important observation is that two distinct histology-informed model selection criteria, i.e., TCS and HFC, unequivocally suggest that models where the extra-cellular ADC is constrained to be higher than the intra-cellular ADC, outperform more general models. The better performance of these models, compatible with values of extra-cellular ADC of approimxately 2.5 $\mu m^2\ ms^{-1}$ reported in (*19*), is not apparent when looking at BIC rankings (e.g., Fig. S3.A, Fig. S4.A, Fig. S4.B, Fig. S4.C). This minor discrepancy can be understood considering that a good signal fitting quality may not necessary imply accurate parameter estimation, especially in presence of noisy measurements (*26*). dMRI signal model fitting is a highly degenerate inverse problem, since strikingly different combinations of tissue parameters (and hence, different estimated cytoarchitectures) can provide similar fitting quality (*44*). This finding stresses the importance of informing dMRI modelling with information from histology, striving for biological specificity.

After selecting a practical dMRI model implementation, we investigated its utility in an exemplificatory immunotherapy response assessment task. Immunotherapy treatments, such as immune checkpoint inhibitors, have shown promise in several cancers (*16*). However, only a small fraction of patients truly benefits from these treatments, and their identification prior to drug administration remains an unsolved challenge (*17*). With this in mind, we investigated on a sample of N = 30 patients whether it is possible to stratify the probability of progression, given baseline values of dMRI metrics within liver tumours. Several, independent statistical analyses consistently point towards the fact that smaller baseline dMRI cell sizes and higher dMRI cell densities, are associated to faster cancer progression under immunotherapy, as measured by PFS. To our knowledge, this is the first time that markers of cell size and density, derived non-invasively in patients *in vivo* with dMRI, have been tested in patient stratification in the context of immunotherapy. Notably,

the deleterious impact of higher cell density on clinical outcome is in line with other studies focussing on different treatments, where higher tumour cellularity has been associated with higher cancer aggressiveness or worse prognosis. In CRC liver metastasis resection, for example, high cell density in resected metastases has been associated to shorter disease-free survival (*51*). Conversely, in breast cancer, lower tumour cellularity has been associated to pathologic complete response in chemo-free dual HER2 blockade treatment (for HER2-positive BC) (*52*), as well as longer survivals in neoadjuvant chemotherapy (*53*). The significant association between immunotherapy PFS and baseline dMRI cell size/density reported here is promising, and motivates future imaging studies. However, we remark that such a finding, while encouraging, comes from an exploratory proof-of-concept demonstration of our biophysical liver dMRI approach in immunotherapy, which requires validation in larger and independent cohorts.

In this study, we benchmarked the proposed dMRI approach against well-established DKI apparent diffusion and kurtosis coefficients (ADC and K) (*30*). ADC and K have shown utility in clinical settings, being easy to compute with compact dMRI acquisitions, and sensitive, for example, to cancer cellularity (negative/positive correlation for ADC/K (*46*), a result confirmed by the correlations detected in this study, namely: r = –0.47 between ADC and CD_{histo}, r = 0.43 between K and CD_{histo}). However, in our cohort, neither of ADC and K showed statistically significant associations with the clinical outcome (PFS). This may due, at least partly, to the fact that ADC and K are semi-quantitative, protocol-dependent metrics, whose value can change as function of, for example, the diffusion time (*46*, *54*). Here, we did not perform any inter-scanner harmonisation delibterately, to stress the quantitative nature of our cell size/density mapping approach, which inherently accounts for inter-scanner protocol differences. However, we acknowledge that better performances for semi-quantitative metrics such as ADC and K could be obtained by incorporating inter-scanner harmonisation (*55*, *56*) in the MRI analysis pipeline.

We would like to acknowledge the following potential limitations of our study. Firstly, our sample size (N = 25 paired dMRI-histology measurements for model development; N = 30 paired dMRI-outcome measurements in patients) is relatively small. This paper provides a first demonstration of the potential utility of a single-compartment model of restricted diffusion fitted to high b-value data in the liver. The demonstration is unique of its kind, since it reports, for example, heretofore undescribed dMRI-based stratification in immunotherapy, providing useful reference standards for future biomarker development studies. Moreover, to our knowledge there are no freely available multi b-value liver dMRI data sets with coupled histological and/or clinical information that could serve as independent test beds for our findings. Nonetheless, while works proposing related dMRI techniques relied on similar (*23*, *36*), if not even smaller (*19*, *20*), sample sizes, we acknowledge that our findings are exploratory. Further confirmation is required in diseases beyond liver cancer, treatments other than immunotherapy, as well as from a larger number of dMRI scanners, protocols, and patients.

Secondly, we acknowledge that results from dMRI-histology comparisons should be always taken with care. For example, here we related dMRI metrics obtained *in vivo* to histological indices from biopsies. While we were able to identify the liver tumours from which the biopsies were taken, we could not identify exactly the area, within a tumour, that was biopsied. This may imply that we may have underestimated the true correlation between dMRI and histological metrics. Also, and most importantly, histology has its own limitations, since it provides cell property estimates that may not be, *per se*, fully accurate.

For example, routine HE histology is an inherently 2D technique, unlike 3D MRI. Moreover, it is affected by artifacts (e.g., due to dehydration, paraffin embedding, imperfect staining, cutting, etc (*57*)), and the automatic processing of large histological fields-of-view requires trading off between sensitivity and specificity. We took steps to mitigate these issues, e.g., by accounting for biases due to tissue shrinkage. Nonetheless, we acknowledge that our histology-derived estimates of cell properties may be biased versions of the true figures. However, we find encouraging that the same dMRI models provided the best fidelity to histology according to independent rankings (e.g., HFC and TCS criteria), giving confidence on the histological validity of the proposed dMRI approach.

We would also like to acknowledge that the proposed dMRI approach neglects other potentially relevant microstructural properties, such as mapping of water exchange between the intra-/extra-cellular space (*21*, *37*), full intra-voxel cell size/cytosolic diffusivity distributions (*42*, *58*), or intra-compartmental T2 or T1 (*36*). One the one hand, this can lead to biases in our estimates of cell size, intra-cellular fraction or cell density. It could also explain, for example, the weak correlation seen between dMRI-derived and histology-derived intra-cellular fraction: dMRI intra-cellular fraction estimates have been shown to be biased by unaccounted exchange (*21*, *37*). On the other hand, properties such as exchange rates may be relevant *per se*, as they could be useful markers of cellular stress. However, we would like to remark that increased model complexity may lead to higher overall variability of the estimated tissue parameters, especially when short, clinically feasible dMRI protocols are considered, as done in this study. In future work we aim to incorporate additional tissue properties to our biophysical model implementations, ensuring that these could be deployed reliably in clinical settings.

In conclusion, *this study delivers a practical liver dMRI signal model consisting of a single-compartment of restricted diffusion within spherical cells, which should be fitted to b-values higher than, approximately, 900 s/mm^2* in vivo. In our data set, this model offers estimates of cell size and cell density that are, respectively, moderately and highly correlated to the underlying histology. Moreover, these metrics may provide complementary information to routine volumetric tumour burden measurements in applications such as immunotherapy, since they enable the stratification of the risk of progression from baseline dMRI scans. Striving to bringing precision imaging one step closer to the clinic, we release our approach as an easy-to-use, open-source Python implementation, which will be freely accessible online.

Materials and Methods

Experimental design

In this study we collected and analysed co-localised liver dMRI and HE-stained histological data to design a practical signal model with high fidelity to histology, and feasible in the clinic. Moreover, we analysed dMRI scans acquired *in vivo* on cancer patients suffering from advanced solid tumours, and participating in a phase I immunotherapy trial. We analysed these scans to assess whether metrics from the proposed dMRI approach enable the stratification of the risk of progression under immunotherapy. In the following sections, we will first introduce the dMRI models object of this paper. Afterwards, we will provide details of the data acquisition and statistical analysis. We will refer to dMRI and histology images acquired in fixed mouse livers *ex vivo* as *preclinical data*. We will instead refer to data acquired in patients *in vivo*, encompassing dMRI, biopsies, and clinical information, as *clinical data*.

dMRI models

Common biophysical dMRI signal models used in body applications accounts for three, non-exchanging water pools, namely: vascular water, characterised by incoherent flow (pseudo-diffusion) in fluid-filled conduits; restricted, intra-cellular water; hindered, extra-cellular, extra-vascular water. This general three-pool representation has found application in several tissues and organs, e.g., colorectal cancer xenografts (*6*), human prostate (*20*), liver (*19*) or breast (*59*). According to this representation, the dMRI signal for a standard PGSE experiment measured at a b-value b, diffusion gradient duration δ, separation Δ, and echo time TE is written as (*6*, *19*, *20*, *60*)

$$s = s_0\left(f_V\, e^{-\frac{TE}{T2_V}} a_V + (1-f_V)\left(f_I\, e^{-\frac{TE}{T2_I}} a_I + (1-f_I)\, e^{-\frac{TE}{T2_E}} a_E\right)\right). \quad (1)$$

Above, s_0 is the apparent proton density, f_V is the vascular signal voxel fraction, f_I is the intra-cellular signal tissue fraction, $T2_V/T2_I/T2_E$ and $a_V/a_I/a_E$ are the T2 constants and diffusion-weighting factors of the vascular, intra-cellular and extra-cellular signals.

The vascular diffusion-weighting factor a_V captures the intra-voxel incoherent motion (IVIM) effect (*61*), i.e., signal attenuating arising from blood perfusing within intricated networks of capillaries. *In vivo*, the IVIM signal decay occurs at a rate that is considerably higher than that of the intrinsic fluid diffusivity: the vascular ADC has been reported to range in intervals as high as [15; 60] $\mu m^2\ ms^{-1}$ (*62*). For this reason, for sufficiently high b-values (no less than $b > 100$ s/mm^2 *in vivo*), the vascular signal vanishes. For example, a_V reduces to approximately 5% of its non-diffusion-weighted value for $b = 150$ s/mm^2 and for a vascular ADC of 20 $\mu m^2\ ms^{-1}$, since $\exp(-150\ s\ mm^2 \times 20\ \mu m^2\ ms^{-1}) = \exp(-0.15 \times 20) = 0.0498$. In such a measurement regime, Eq. (3) reduces to a two-pool representation (*19*), which will be the starting point of our modelling framework, i.e.,

$$s = s_0\ (1-f_V)\left(f_I\, e^{-\frac{TE}{T2_I}} a_I + (1-f_I)\, e^{-\frac{TE}{T2_E}} a_E\right). \quad (2)$$

A common model for the intra-cellular diffusion-weighting factor a_I in Eq. 4 is that of restricted diffusion within spheres of diameter L (*19*, *20*). In this case, a_I depends on the b-value b, gradient duration δ and gradient separation Δ as

$$a_I = e^{-b\, ADC_I(\delta,\Delta,D_{0,I},L)}, \quad (3)$$

where

$$ADC_I = \frac{2}{D_{0,I}\delta^2(\Delta-\delta/3)} \sum_{m=1}^{\infty} \frac{\alpha_m^{-4}}{\alpha_m^2 R^2 - 2}\left(2\delta - \frac{2+e^{-\alpha_m^2 D_{0,I}(\Delta-\delta)} - 2e^{-\alpha_m^2 D_{0,I}\delta} - 2e^{-\alpha_m^2 D_{0,I}\Delta} + e^{-\alpha_m^2 D_{0,I}(\Delta+\delta)}}{\alpha_m^2 D_{0,I}}\right) \quad (4)$$

is the Gaussian phase distribution approximation of the intra-cellular ADC (*63*). Above, α_m is the m-th root of $\alpha_m R\, J'_{3/2}(\alpha_m R) - 0.5\, J_{3/2}(\alpha_m R) = 0$, $J_{3/2}(x)$ is the Bessel function of the first kind and order 3/2, and $J'_{3/2}(x)$ its first-order derivative. ADC_I depends on the intrinsic intra-cellular diffusivity $D_{0,I}$ (i.e., the diffusivity of the cytosol), and on the cell size $L = 2R$

(with R being the cell radius, whereas L is the diameter). Note that in practice, large cells contribute substantially more to the total dMRI signal than small cells, since they contain more water and feature stronger diffusion time dependence. For this reason, dMRI estimates of cell size L in fact represent volume-weighted mean cell size statistics (*7*, *42*). Therefore, throughout this manuscript, dMRI cell size L has been referred to as a *volume-weighted cell size* (vCS) index.

Conversely, the extra-cellular, extra-vascular signal may be described in terms of hindered diffusion in a tortuous space (*19*, *27*, *60*), i.e.,

$$a_E(b,\Delta) \;=\; e^{-b\,ADC_E\left(\Delta, D_{E,\infty}, \beta\right)}. \quad (5)$$

A general form of ADC_E depends on the gradient separation Δ via

$$ADC_E\left(\Delta, D_{E,\infty}, \beta\right) \;=\; D_{E,\infty} \;+\; \frac{\beta}{\Delta}, \quad (6)$$

where in Eq. 6 $D_{E,\infty}$ is the asymptotic, long-time value of ADC_E for $\Delta \to \infty$, while β controls the degree of diffusion time dependence (*60*).

The 5 implementations of the two-compartment model

It is not yet clear to what extent the complete set of extra-cellular and intra-cellular diffusion properties can be resolved jointly with the minimal dMRI encodings that are feasible in the clinic (*19*, *27*, *44*). Here we investigated 5 different practical implementations of Eq. 2, with different levels of detail and complexity. The 5 implementations can be divided into two families, namely: models in which there are no assumptions on which of ADC_I and ADC_E is higher; models hypothesising that $ADC_E > ADC_I$, an assumption compatible, for example, with the values of ADC_E as high as approximately 2.5–2.8 $\mu m^2\ ms^{-1}$ measured *in vivo* in the liver in recent studies (*19*).

The first family of models is more general, and includes two implementations:

i. *Diff-in-exTD*: the most general implementation, as it accounts explicitly for diffusion TD in the extra-cellular space. It relies on the full expression of ADC_E in Eq. 6, and the search range for $D_{E,\infty}$ is large enough to include values for which ADC_E can either be larger or smaller than ADC_I.
ii. *Diff-in-ex*: a slightly simpler implementation where extra-cellular TD is considered negligible, i.e., such that $\beta = 0$ in Eq. 6. Again, the search range for $D_{E,\infty}$ is large enough to include values for which ADC_E can either be larger or smaller than ADC_I.

The second family of models, where $ADC_E > ADC_I$, includes 3 implementations:

i. *Diff-in-exTDFast*: equivalent to *Diff-in-exTD*, but such that the lower bound for $D_{E,\infty}$ ensures that that $ADC_E > ADC_I$ even for the largest cell size L.
ii. *Diff-in-exFast*: equivalent to *Diff-in-ex*, but again such that the lower bound for $D_{E,\infty}$ ensures that $ADC_E > ADC_I$ for any cell size L.
iii. *Diff-in*: a model where the extra-cellular signal is assumed to be negligible compared to the intra-cellular one, due to ADC_E being much larger than ADC_I. In this implementation we assume that $f_I a_I + (1 - f_I) a_E \approx f_I a_I$ for physiological meaningful intra-cellular signal fractions f_I, implying that Eq. 2 reduces to

$$s = s_0 (1 - f_V) f_I e^{-\frac{TE}{T2_I}} a_I(b, \delta, \Delta, D_{0,I}, L). \quad (7)$$

For the practical implementation of the 5 dMRI models described above, we assumed that $T2_I \approx T2_E \doteq T2_T$, given the challenge of resolving accurately multiple T2 constants with minimal TE samplings (*20*, *36*), jointly with diffusion properties.

Fitting strategy

We fitted the 5 models custom-written Python routines, based on objective function minimisation initialised by a grid search. The objective function f_{obj} was defined as $f_{obj} = -ln(\lambda)$, where λ is the offset-Gaussian likelihood function, as in (*34*). Fitting was performed twice. The first fitting was performed on the set of DW measurements where the vascular signal can be considered suppressed, supporting the use of the two-pool representation of Eq. 2. For the sake of simplicity, we will refer to this fitting as "*fitting to the whole image set*", although in practice we mean the whole set of images *for which vascular contributions can be neglected*. Afterwards, we repeated the fitting on a high b-value subset of the former data set, in which contributions from extra-cellular signals are also minimised. For practical model fitting, we fixed f_V and $T2_I$ to pre-computed values, obtained by fitting a separate diffusion-relaxation model disentangling vascular vs non-vascular (i.e., tissue) signals (*64*).

Fitting provided voxel-wise estimates of vCS ($vCS = L$), as well as the intra-cellular voxel signal fraction F, defined as

$$F = (1 - f_V) f_I. \quad (8)$$

We combined vCS and F to derive a cell density per unit volume, borrowed from previous literature (*20*), and defined as

$$CD = \frac{F}{vCS^3}. \quad (9)$$

Preclinical data

Animals

We obtained data from 7 fixed livers of NOD.Cg-Prkdc^{scid} IL2rg^{tm1WjI}/SzJ mice. All experimental protocols were approved and monitored by the Vall d'Hebron Institute of Research Animal Experimentation Ethics Committee (CEEA; registration number 68/20) in accordance with relevant local and EU regulations. We studied six livers from mice implanted with cells derived from biopsies of prostate cancer patients, as part of an ongoing study aiming to develop patient-derived xenografts models, plus an additional liver from a mouse without any implantation. We implanted one tumor biopsy core with growth factor-enriched Matrigel (Corning) subcutaneously in the flank of each mice. We derived tissue from the following biopsies: iliac bone metastasis biopsy of a patient with metastatic castration-resistant prostate cancer, presenting with bone metastasis and Gleason score 3+4 adenocarcinoma; prostate biopsy of a patient with metastatic hormone-sensitive prostate cancer, presenting with bone metastasis and Gleason score 5+4 adenocarcinoma; liver biopsy of a patient with metastatic castration-resistant prostate cancer, presenting with bone and visceral metastasis and Gleason score 4+4 acinar adenocarcinoma; liver biopsy of a patient with metastatic hormone-sensitive prostate cancer, presenting with bone and liver metastasis and Gleason score 4+4 adenocarcinoma. After implantation, we measured

tumour size using calipers, and monitored mouse weight weekly. We sacrificed animals by cervical dislocation under general anesthesia when tumour volume exceeded 2000 mm^3. We collected the livers and fixed them overnight in formalin. Finally, we transferred them to phosphate-buffered saline (PBS) solution for downstream MRI acquisition.

MRI

The mouse livers were washed in PBS solution for 24 hours and scanned immersed in PBS on a 9.4T Bruker Avance system with 200 mT/m gradient inserts at room temperature. Samples were tightened with sewing thread to a histology cassette, which was then placed into a Falcon® tube filled with PBS. A 1-channel birdcage coil was used for both excitation and signal detection. The protocol included a high-resolution, anatomical T2-weighted RARE fast spin echo sequence with parameters: TR = 2500 ms; TE = 12 ms; resolution: 144 µm × 144 µm; matrix size: 230 × 230; 4 slices, 2.216 mm-thick; NEX = 4; RARE factor: 8. We acquired dMRI using the vendor's PGSE (Fig. S17A) sequence (TR = 2700 ms; resolution: 386 µm × 386 µm; matrix size: 86 × 86; 4 slices, 2.216 mm-thick, 2 of which containing the liver, NEX = 1). The protocol featured: δ = 10 ms, Δ = {15, 30} ms, 10 linearly spaced b-values for each Δ (minimum/maximum b of 0 and roughly 2800 s/mm^2). DW images corresponding to Δ = 15 ms were acquired at each of TE = {31, 45, 65} ms, while DW images corresponding to Δ = 30 ms at each of TE = {45, 65} ms. The b-value as a function of the gradient strength G and of δ and Δ can be calculated with the well-known expression $b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3)$.

We post-processed dMRI scans as follows. Firstly, we merged all DW images and denoised them with the freely available Python implementation of Marchenko-Pastur Principal Component Analysis (MP-PCA) (*65*) (kernel size: 7×7×3). Afterwards, we mitigated Gibbs ringing with the local sub-voxel shift method (*66*) based on MrTrix3, and corrected temporal signal drifts by assessing changes of signal level in a small region of PBS, while accounting for TE changes (T2 of PBS solution: 500 ms). Finally, we fitted the 5 dMRI models (*Diff-in-exTD*, *Diff-in-exTDFast*, *Diff-in-ex*, *Diff-in-exFast* and *Diff-in*) voxel-by voxel. Fitting bounds for tissue parameters were: [0; 1] for f_I; [0.8; 2.6] µm^2 ms^{-1} for $D_{0,I}$; [8; 40] µm for the *vCS* parameter L; [0.8; 2.6] µm^2 ms^{-1} for $D_{E,\infty}$ in models *Diff-in-ex* and *Diff-in-exTD,* and [1.75; 2.6] µm^2 ms^{-1} in models *Diff-in-exFast* and *Diff-in-exTDFast*; [0; 10] µm^2 for β in models *Diff-in-ex-TD* and *Diff-in-exTDFast*.

For practical fitting, we fixed f_V and $T2_T$ in each voxel to values obtained by fitting a two-pool model disentangling vascular vs tissue signals (*64*) (fitting bounds: [0; 1] for f_V; [5; 80] ms for $T2_T$). Fitting was performed twice: once, on all images with b > 1000 s/mm^2 (fitting on whole image set); afterwards, for b > 1000 s/mm^2 (high b-value fitting). In our *ex vivo* preclinical data, the vascular signal captures partial volume effects with the PBS solution, e.g., at the PBS-liver interface or in large vessels, filled with the liquid, due to lack of blood flow and pulsation. Note that in our *ex vivo* data, the vascular ADC represents the ADC of PBS solution (i.e., free water) at room temperature (fitting bounds: [2.2; 2.8] µm^2 ms^{-1}), hence much smaller than the IVIM-dominated vascular ADC of *in vivo* imaging. For this reason, we adopted a b-value threshold of 1000 s/mm^2 to achieve a satisfactory suppression of the PBS signal, while we used a minimum b-value of 1800 s/mm^2 for fitting on high b-value images, minimising the contribution of extra-cellular water.

For reference, we also used in-house Python routines to compute standard ADC and apparent diffusion excess kurtosis K, by fitting

$$s = s_0\, e^{-b\,ADC + \frac{1}{6}K(b\,ADC)^2} \quad (10)$$

to the set of DW images acquired at TE = 45 ms and Δ = 30 ms.

Histology

After MRI, samples underwent standard histological procedures (i.e., dehydration, paraffine embedding) within the same histology cassette in which they were imaged. We then cut two 4 µm-thick histological sections at known radiographic position (one per MRI slice) with a microtome, and stained them with HE. We performed cuts following the same direction of the virtual cuts performed by MRI slicing. We placed stained material on standard microscopy slides, and acquired digital images with a Hamamatsu C9600-12 slide scanner at a pixel resolution of 0.227 µm (40× magnification). An experienced pathologist (S.S.) inspected the images and assessed them qualitatively. We processed digital histological images with the automatic cell detection tool of QuPath (*32*), obtaining per-cell characteristic area A and diameter $l = \sqrt{\frac{4}{\pi}A}$. Afterwards, we split histological images into 386 µm × 386 µm patches, matching the in-plane MRI resolution, and computed characteristic volume-weighted cell size vCS_{histo} and intra-cellular area fraction F_{histo} patch-by-patch, similarly to (*42*). vCS_{histo}, defined as

$$vCS_{histo} = \left(\frac{<l^7>}{<l^3>}\right)^{\frac{1}{4}}, \quad (11)$$

is a better counterpart of dMRI cell size metrics than the arithmetic mean cell size $aCS_{histo} =< l >$ (*7*, *42*), since the dMRI signal is intrinsically volume-weighted (i.e., larger cells contribute more to the signal than smaller cells). Conversely, F_{histo} is the fraction of patch area occupied by the intra-cellular space, and is therefore the counterpart of quantity F_{MRI} in Eq. 9. We accounted for biases coming from: i) estimating the size of 3D objects from 2D views (bias 1), ii) tissue shrinkage (bias 2), by rescaling vCS_{histo} and CD_{histo}. The final estimate of vCS_{histo} was 1.4616 times larger than the value obtained from direct image processing (1.148×1.2732 = 1.4616, where 1.2732 was derived from the theory of spherical caps and accounts for bias 1, while 1.148 accounts for bias 2, and corresponds to a plausible shrinkage of 12.9% following dehydration, clearing and paraffin embedding (*57*)). The final estimate of CD_{histo} was instead 1.318 times smaller than the value derived from direct image processing, since 1 mm^2 of shrunk tissue corresponds to 1.148×1.148 mm^2 = 1.318 mm^2 of unprocessed tissue for a shrinkage factor of 12.9% (*57*). Lastly, we co-registered the L_{histo} and F_{histo} maps to their corresponding MRI slice, using a warping transformation estimated via symmetric diffeomorphic registration of the specimen manual outlines (*42*) in DiPy (*67*).

Clinical data

Cohort description

We obtained data from patients suffering from advanced solid tumours, recruited for an ongoing imaging study approved by the Vall d'Hebron University Hospital Ethics committee (PR(AG)29/2020). Patients, potentially eligible for a phase I immunotherapy trial at VHIO (Barcelona, Spain), provided informed written consent to participate in the imaging study. We included data from 33 patients with liver malignancies (mean/std of age: 62.91/12.34 year; 16 male, 17 female), of which 3 suffered from primary HCC, while 30 had liver metastases from different primary cancers (10 colon, 8 melanoma, 3 rectal, 2 ovarian, 2 gastric, 2 breast, 1 renal, 1 endometrial, 1 ureteral). We scanned 11 patients on a

1.5T Siemens Avanto system, while 22 on a 3T GE SIGNA Pioneer system. We obtained baseline dMRI scans (i.e., acquired immediately before starting immunotherapy), and digitised HE-stained biopsies from one of the imaged liver tumours. In this study, we used biopsies from 18 patients (6 scanned at 1.5T, 12 at 3T), which we collected after baseline dMRI. 30 out of 33 patients effectively entered the immunotherapy trial after screening. For this subset of 30 patients, we obtained clinical outcome information in the form of standard PFS. The PFS represents the temporal lag between therapy starting date and progression or death (whichever occurs first), with progression determined via RECIST (*13*) on standard-of-care CT imaging, or in case of established clinical worsening.

MRI

We imaged patients at the level of the abdomen on two MRI machines. We scanned 11 patients on a 1.5T Siemens Avanto scanner using the vendor 18-channel body coil for detection. The protocol included an anatomical, T2-weighted RARE fast spin echo scan, with salient parameters: resolution: 1.4 × 1.4 × 5 mm^3; 32 slices; TR = 4500 ms; TE = 82 ms; echo train length: 29; NEX = 8; GRAPPA = 2. It also included a fat-suppressed DW TRSE (Fig. S17B) EPI scan (total scan time: 16 minutes). It featured: resolution: 1.9 × 1.9 × 6 mm^3; 32 slices; TR = 7900 ms; bandwidth 1430 Hz/pixel; averaging of 3 orthogonal diffusion directions × 2 signal averages (effective NEX = 6); GRAPPA factor of 2; 6/8 partial Fourier imaging. The dMRI protocol consisted of b = {0, 50, 100, 400, 900, 1200, 1600} s/mm^2 images, each acquired for TE = {93, 105, 120} ms, for a total of 21 images. The phase encoding direction was anterior-posterior, and one additional image (b = 0 s/mm^2; TE = 93 ms) was acquired with reversed phase encoding polarity. The gradient timings (Fig. S17B) were: δ_1 = 8.9 ms, δ_2 = 17.6 ms, δ_3 = 20.4 ms, δ_4 = 6.0 ms, $\Delta_{1,2}$ = 17.4 ms and $\Delta_{1,4}$ = 63.9 ms when TE = 93 ms; δ_1 = 13.2 ms, δ_2 = 19.3 ms, δ_3 = 24.8 ms, δ_4 = 7.7 ms, $\Delta_{1,2}$ = 21.7 ms and $\Delta_{1,4}$ = 74.2 ms when TE = 105 ms; δ_1 = 18.9 ms, δ_2 = 21.0 ms, δ_3 = 30.5 ms, δ_4 = 9.5 ms, $\Delta_{1,2}$ = 27.5 ms and $\Delta_{1,4}$ = 87.5 ms when TE = 120 ms. The b-value as a function of the gradient timings, the gradient strength G and the proton gyromagnetic ration γ can be expressed as

$$b = \gamma^2 G^2 \left(\delta_1^2 (\Delta_{1,2} - \delta_1) + \frac{2}{3}(\delta_1 + \delta_2)^3 + (\delta_1 + \delta_2 - \delta_3)^2 (\Delta_{1,4} - \Delta_{1,2} - \delta_2 - \delta_3) \right). \quad (12)$$

Moreover, we scanned 22 patients on a 3T GE SIGNA Pioneer scanner, using the vendor 48-channel torso coil for signal reception, with 32 channels enabled for detection. The protocol included an anatomical, respiratory-gated T2-weighted RARE fast spin echo scan, with salient parameters: resolution: 1.4 × 1.4 × 6 mm^3; 32 slices; TR = 4615 ms; TE = 52.86 ms; echo train length: 16. It also included a respiratory-gated, fat-suppressed PGSE (Fig. S10A) EPI scan (total scan time: 16 minutes). It featured: resolution: 2.4 × 2.4 × 6 mm^3; 32 slices; TR = 6000 ms; bandwidth 1953 Hz/pixel; averaging of 3 orthogonal diffusion directions × 2 signal averages (effective NEX = 6); ASSET factor of 2. The dMRI protocol consisted of b = {0, 50, 100, 400, 900, 1200, 1500} s/mm^2 images, each acquired for TE = {75, 90, 105} ms, for a total of 21 images. The gradient timings (Fig. S17A) were: gradient duration δ of δ = {0.0, 3.9, 5.2, 9.2, 15.0, 18.2, 21.0} ms for the b-values acquired at TE = 75 ms, and δ = {0.0, 3.9, 5.2, 9.2, 13.0, 15.8, 18.5} ms for those acquired at both TE = 90 ms and 105 ms; gradient separation Δ of Δ = {0.0, 27.8, 29.0, 33.0, 28.7, 31.8, 34.7} ms for the b-values acquired at TE = 75 ms and Δ = {0.0, 27.8, 29.0, 33.0, 37.0, 39.6, 42.3} ms for those acquired at both TE = 90 ms and TE = 105 ms.

dMRI post-processing consisted of slice-wise Python MP-PCA denoising (kernel size: 5 × 5) (*65*); Gibbs unringing (*66*) based on MRTrix3; motion correction via affine co-registration (*68*); EPI distortion correction based on FSL (*69*) (on the 1.5T data only). An experienced radiologists (R.P.L.) segmented tumours on the T2-w anatomical scan, enabling per-patient tumour volume calculation. Afterwards, we warped the tumour mask to dMRI space using a non-linear transformation estimated with ANTs (*70*). Then, we fitted the 5 dMRI models object of this article on a voxel-by-voxel basis, fixing again f_V and $T2_T$ to previously computed values (*64*) (fitting bounds: [0; 1] for f_V; [20; 140] ms for $T2_T$). Fitting bounds for tissue parameters were: [0; 1] for f_I; [0.8; 3.0] μm^2 ms^{-1} for $D_{0,I}$; [8; 40] μm for the *vCS* parameter L; [0.8; 3.0] μm^2 ms^{-1} for $D_{E,\infty}$ in models *Diff-in-ex* and *Diff-in-exTD,* and [1.75; 3.0] μm^2 ms^{-1} in models *Diff-in-exFast* and *Diff-in-exTDFast*; [0; 10] μm^2 for β in models *Diff-in-ex-TD* and *Diff-in-exTDFast*.

We fitted the 5 dMRI models twice: once, on images acquired at a b-value b > 100 s/mm^2, to suppress vascular signals (fitting to the whole image set); afterwards, to b > 900 s/mm^2 images, to also minimise extra-cellular signal contributions (high b-value fitting). For scans performed on the 1.5T Siemens system, based on a DW TRSE sequence (Fig. S10B), we used $\Delta_{1,2} + \delta_2$ in place of Δ in Eq. 6 for the extra-cellular ADC. Moreover, we replaced Eq. 4 with a numerical implementation of the intra-cellular model of restricted diffusion within spheres. This implementation was based on a Radial Basis Function regressor trained on synthetic signals generated with Monte Carlo random walks within meshed spheres. We performed the simulations with the freely available MCDC simulator (*71*), varying the intra-cellular diffusivity $D_{0,I}$ in [0.8; 3.0] μm^2 ms^{-1} and the cell diameter L in [8; 40] μm. We used 12 linearly spaced values for each of $D_{0,I}$ and L, for a total of 12×12 = 144 simulations. Finally, we computed routine ADC and diffusion excess kurtosis K by fitting Eq. 10 on b > 100 s/mm^2 images acquired at the shortest TE with in-house Python code.

Histology

In patients, we performed ultrasound-guided biopsies of one liver tumour at the Vall d'Hebron University Hospital of Barcelona (Spain). The biological material underwent standard histological processing and staining with HE. We acquired digital images of the stained biopsies using a Hamamatsu C9600-12 slide scanner (resolution: 0.454 μm; 20× magnification). An experienced pathologist (S.S.) assessed the images and drew manually a region-of-interest (ROI) outlining the tumour tissue in each HE image. In parallel, an experienced radiologist (R.P.L.) inspected the ultrasound scan jointly with MR images, and outlined on MRI the tumour from which the biopsy was taken. We processed HE images with QuPath and computed per-biopsy volume-weighted cell size vCS_{histo}, intra-cellular area fraction F_{histo} and cell density per unit area CD_{histo}, as previously described for the mouse data. Again, we corrected vCS_{histo} and CD_{histo} for biases. The final estimate of vCS_{histo} was 1.503 times larger than the value obtained from direct image processing (1.1806×1.2732 = 1.503, where 1.2732 accounts again for biases from 2D sectioning, while 1.1806 accounts for a plausible tissue shrinkage of 15.3% following fixation, dehydration, clearing and paraffin embedding (*57*)). The final estimate of CD_{histo} was instead 1.3938 times smaller than the value derived from direct image processing, since 1 mm^2 of shrunk tissue corresponds to 1.1806×1.1806 mm^2 = 1.3938 mm^2 of unprocessed tissue for a shrinkage factor of 15.3% (*57*).

Statistical analyses

We performed three analyses to i) select the dMRI model providing metrics with the highest fidelity to histology; ii) assess the correlation between MRI and histology metrics; iii)

demonstrate the potential of the new dMRI model in the context of cancer immunotherapy. For the analyses listed above, we computed mean and standard deviation of all MRI and histology metrics within the mouse liver samples. Moreover, we computed the mean and standard deviation of all MRI metrics within a mask containing all liver tumours, as well as on the tumour from which the biopsy was taken.

dMRI model selection

We performed histology-informed model selection using the *MRI-histology Total Correlation Score* (TCS). We carried out the selection independently on dMRI metrics obtained i) fitting the signal models to the whole image set, or ii) fitting the signal models only to high b-value images.

TCS selects the model providing the highest Pearson's correlation coefficients between vCS_{MRI} and vCS_{histo}, and between F_{MRI} and F_{histo}. TCS focusses on the overall sensitivity of a dMRI model towards the underlying histology (i.e., correlation), rather than on the accuracy of the numerical estimation of the histological properties themselves. This is justified by noting that in certain applications it may be relevant to detect, for example, changes in MRI apparent cell size from normative reference values, even if the numerical value of such an apparent cell size contains systematic biases with respect to the true histological cell size. Practically, TCS selects the model that maximises

$$\mathrm{TCS} = r(vCS_{MRI}, vCS_{histo}) + r(F_{MRI}, F_{histo}), \quad (13)$$

where $r(vCS_{MRI}, vCS_{histo})$ and $r(F_{MRI}, F_{histo})$ are the Pearson's correlation coefficients of vCS_{MRI} and F_{MRI} with their histological counterparts vCS_{histo} and F_{histo}. The correlation between CD_{MRI} and CD_{histo} was not included in Eq. 13 since CD_{MRI} is fully determined analytically from vCS_{MRI} and F_{MRI}. For the computation of the TCS index, we pooled together mouse and human data (N = 25).

For reference, we also performed model selection using a *Histology Fidelity Criterion* (HFC) and the well-known *Bayesian Information Criterion* (BIC) (*33*), a common model selection technique in dMRI model development (*34*). HFC rewards the models providing the best accuracy in the numerical estimation of histological cell size and intra-cellular fraction estimation, i.e., the model providing vCS_{MRI} as close as possible to vCS_{histo}, and F_{MRI} as close as possible to F_{histo}. Practically, HFC selects the candidate model that minimises

$$\mathrm{HFC} = \frac{|vCS_{MRI} - vCS_{histo}|}{vCS_{histo}} + \frac{|F_{MRI} - F_{histo}|}{F_{histo}}. \quad (14)$$

Information on CD_{MRI} and CD_{histo} was not included in Eq. 14 since CD_{MRI} is not a degree of freedom of the dMRI models (it is fully determined analytically once vCS_{MRI} and F_{MRI} are fitted), and also because CD_{MRI} and CD_{histo} have different units, so they are not directly comparable.

BIC selects the model providing the best goodness of fit, penalising complexity. BIC can be computed from the value of the minimised fitting objective function $f_{obj} = -ln(\lambda)$, where λ is maximum likelihood, as

$$BIC = Pln(N) - ln(\lambda), (15)$$

where P is the number of estimated model parameters, and N the number of dMRI measurements used for the fitting. We performed model selection based on BIC on a voxel-by-voxel basis, and selected a winning model for each mouse liver and patient by majority voting across voxels.

Simulated dMRI model selection

We corroborated results from *ex vivo* and *in vivo* dMRI with computer simulations, in which dMRI model selection was performed on synthetic dMRI signals. We synthesised signals via Monte Carlo diffusion random walks for each of the three protocols considered in this study, i.e., for the PGSE protocols used in the 9.4T scanner *ex vivo* and in the 3T *in vivo*, and for the TRSE protocol used in the 1.5T scanner *in vivo*.

We performed Monte Carlo simulations with the freely available, open-source MCDC simulator (*71*), seeding random walkers in a substrate made of packed spherical cells of identical diameter (Fig. S5), a common geometric model in body dMRI (*6*, *19*, *22*, *23*). We controlled the intra-sphere fraction F by adding gaps of increasing size in-between abutting spheres, which we packed in an ideal cubic lattice. We probed 4 different values of F (approximately equal to 0.197, 0.323, 0.406, 0.523; notice that the maximum theoretical value of F for cubic lattice packing is equal to 0.5236). For each value of F, we varied the cell diameter (8, 16, 22 and 30 μm), intra-sphere diffusivity (10 linearly-spaced values in the ranges [0.8; 2.6] $\mu m^2\,ms^{-1}$ and [0.8; 3.0] $\mu m^2\,ms^{-1}$ for the *ex vivo* and *in vivo* protocols respectively) and extra-sphere intrinsic diffusivity (again, 10 linearly-spaced values in the ranges [0.8; 2.6] $\mu m^2\,ms^{-1}$ and [0.8; 3.0] $\mu m^2\,ms^{-1}$ for the *ex vivo* and *in vivo* protocols respectively), generating a total of 1600 synthetic voxels.

We corrupted synthetic signals with Rician noise (signal-to-noise ratio at $b = 0$ of 30), and then processed them in the same way as done for actual MRI measurements, Briefly, we first fitted the five candidate models for both fitting strategy (whole image set; high b-value image subset), and then performed dMRI model selection according to the TCS, HFC and BIC criteria. For TCS ranking, we calculated correlation coefficients between reference and estimated vCS and F pooling results from all 1600 synthetic voxels. Conversely, for HFC and BIC ranking, we selected in each synthetic voxel the model minimising each of HFC and BIC, and then counted the proportion of voxels in which each model was selected.

dMRI-histology correlation analysis

We computed mean and standard deviation of all MRI and histology metrics within the mouse liver samples. Moreover, we computed the mean and standard deviation of all MRI metrics within a mask containing all liver tumours, as well as on the tumour from which the biopsy was taken. We pooled together MRI and histology metrics from mice and patients to calculate the Pearson's linear correlation coefficient r among each possible pair of metrics, visualising these are correlation matrices. ADC was normalised to the ADC of the PBS solutions (mouse livers) and to the free water diffusivity at 37 °C (3.0 $\mu m^2\,ms^{-1}$), to account for differences in temperature between the preclinical and clinical data set. The sample size was N = 25 (N = 7 data points from mice plus N = 18 from patients). We consider correlations to be weak, moderate, and strong when $|r| < 0.4$, $|r| \geq 0.4$ but $|r| < 0.6$, and $|r| \geq 0.6$ respectively.

Response assessment in immunotherapy

We studied mean values of *in vivo* dMRI metrics within liver tumours at baseline to assess whether these could stratify the probability of progression under immunotherapy, as measured by our clinical outcome PFS (measured in days). The sample size was N = 30. Firstly, we performed the stratification after binarising all MRI metrics as lower/higher than the median of the cohort. This allowed us to obtain two groups of patients, i.e., patients whose dMRI metric at baseline was either low or high. We evaluated group-wise survival curves with the Kaplan-Meier estimator, and compared the two curves with a log-rank test. We confirmed results from such a statistical model with a proportional hazard Cox regression, where the binarised MRI metric was the only regressor. Afterwards, we performed a second patient stratification, but without binarising MRI metrics, i.e., using them as continuous regressors. We fitted a proportional hazard Cox model using each dMRI metric in turn as the only regressor (in the form of a z-score), and then fitted the statistical model again but including including age, sex and total baseline tumour volume as confounding factors. We considered p-values of 0.05 or lower as significant. We performed all survival analyses in Python, using the freely available *lifelines* library.

Acknowledgments

We thank the whole medical oncology, radiology, pathology, molecular biology, clinical trial, and IT teams at the Vall d'Hebron University Hospital and at the Vall d'Hebron Institute of Oncology in Barcelona (Spain), without whom this study would not have been possible. We are also thankful to the Vall d'Hebron Radiology department and to the ASCIRES CETIR clinical team for their assistance, and to past and present members of the Radiomics group for useful discussion and advice. Finally, we would like to express our sincere gratitude to all patients and their families for dedicating their time to research.

Funding: 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.). 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 Generitat de Catalunya Agency for Management of University and Research Grants of Catalonia (AGAUR) (2023PROD00178). This research has been funded by the CaixaResearch Advanced Oncology Research Program supported by "La Caixa" Foundation (to R.P.L.). 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) and "LCF/BQ/PI20/11760033" (funding I.C.S). I.C.S. also receives the support of the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 847648. 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).

AstraZeneca: PREdICT (R.P.L, E.G., P.N., R.T.)
CRIS Foundation: TALENT19-05 (R.P.L.)
Instituto de Salud Carlos III: PI18/01395 and PI21/01019 (R.P.L.)
Prostate Cancer Foundation: 18YOUN19 (R.P.L.)
Fero Foundation (R.P.L.)
"la Caixa" Foundation: CaixaResearch Advanced Oncology Research Program (R.P.L.)
"la Caixa" Foundation: LCF/BQ/PR22/11920010 (F.G., A.V., A.G.)
"la Caixa" Foundation: LCF/BQ/PI20/11760033 (I.C.S.)
European Union's Horizon 2020 Marie Sklodowska-Curie: 847648 (I.C.S.)
Generalitat de Catalunya: BP3 801370, 2019 BP 00182 (K.B.), 2023PROD00178 (R.P.L.)
Agencia Estatal de Investigación: CEX2020-001024-S/AEI/10.13039/501100011033, PRE2022-102586 (A.G.)
UK Research and Innovation: MR/T020296/2 (M.P.)

Author contributions:
Conceptualization: F.G., R.P.L., K.B., M.P., E.G., R.T., P.N., J.M.
Methodology: F.G., R.P.L., K.B., M.P., A.G.
Investigation: F.G., R.P.L., K.B., I.C.S., I.B., S.S., G.S., A.G., V.G., J.F.C., X.M., R.M., N.R., M.E., M.Vie., R.T., P.N., J.M., E.G.
Resources: R.P.L., F.G., P.N., J.M., E.G., N.R., M.E., V.G., M.Vid., P.G.PG, I.B.
Formal analysis: F.G.
Visualization: F.G.
Software: F.G., K.B., A.G.
Data curation: F.G., R.P.L., K.B., A.V., G.S., I.C.S., A.G.
Project administration: F.G., R.P.L., K.B., E.G., P.N., R.T., J.M., I.C.S.
Funding acquisition: R.P.L., E.G., R.T., P.N., J.M., F.G., K.B., I.C.S.
Supervision: F.G., R.P.L., E.G., R.T., P.N., J.M.
Writing—original draft: F.G., R.P.L., K.B., M.P.
Writing—review & editing: all authors

Competing interests: This study received funding from AstraZeneca. M.Vid. works for Siemens Healthineers. P.G.P.G works for GE HealthCare. K.B. worked as a researcher at the Vall d'Hebron Institute of Oncology (Barcelona), and is now an employee of AstraZeneca. AstraZeneca, Siemens and GE did not influence the acquisition and analysis of the data, the interpretation of the results, or the decision to submit the manuscript in its current form for publication.

References

1. N. Weiskopf, L. J. Edwards, G. Helms, S. Mohammadi, E. Kirilina, Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. *Nature Reviews Physics*. **3**, 570–588 (2021).

2. A. M. Jarrett, A. S. Kazerouni, C. Wu, J. Virostko, A. G. Sorace, J. C. DiCarlo, D. A. Hormuth 2nd, D. A. Ekrut, D. Patt, B. Goodgame, S. Avery, T. E. Yankeelov, Quantitative

magnetic resonance imaging and tumor forecasting of breast cancer patients in the community setting. *Nat. Protoc.* **16**, 5309–5338 (2021).

3. V. G. Kiselev, Fundamentals of diffusion MRI physics. *NMR Biomed.* **30** (2017), doi:10.1002/nbm.3602.

4. D. S. Novikov, E. Fieremans, S. N. Jespersen, V. G. Kiselev, Quantifying brain microstructure with diffusion MRI: Theory and parameter estimation. *NMR Biomed.* **32**, e3998 (2019).

5. D. Le Bihan, Looking into the functional architecture of the brain with diffusion MRI. *Nat. Rev. Neurosci.* **4**, 469–480 (2003).

6. E. Panagiotaki, S. Walker-Samuel, B. Siow, S. P. Johnson, V. Rajkumar, R. B. Pedley, M. F. Lythgoe, D. C. Alexander, Noninvasive quantification of solid tumor microstructure using VERDICT MRI. *Cancer Res.* **74**, 1902–1912 (2014).

7. J. Veraart, D. Nunes, U. Rudrapatna, E. Fieremans, D. K. Jones, D. S. Novikov, N. Shemesh, Noninvasive quantification of axon radii using diffusion MRI. *Elife*. **9**, e49855 (2020).

8. J. Cohen-Adad, E. Alonso-Ortiz, M. Abramovic, C. Arneitz, N. Atcheson, L. Barlow, R. L. Barry, M. Barth, M. Battiston, C. Büchel, M. Budde, V. Callot, A. J. E. Combes, B. De Leener, M. Descoteaux, P. L. de Sousa, M. Dostál, J. Doyon, A. Dvorak, F. Eippert, K. R. Epperson, K. S. Epperson, P. Freund, J. Finsterbusch, A. Foias, M. Fratini, I. Fukunaga, C. A. M. G. Wheeler-Kingshott, G. Germani, G. Gilbert, F. Giove, C. Gros, F. Grussu, A. Hagiwara, P.-G. Henry, T. Horák, M. Hori, J. Joers, K. Kamiya, H. Karbasforoushan, M. Keřkovský, A. Khatibi, J.-W. Kim, N. Kinany, H. Kitzler, S. Kolind, Y. Kong, P. Kudlička, P. Kuntke, N. D. Kurniawan, S. Kusmia, R. Labounek, M. M. Laganà, C. Laule, C. S. Law, C. Lenglet, T. Leutritz, Y. Liu, S. Llufriu, S. Mackey, E. Martinez-Heras, L. Mattera, I. Nestrasil, K. P. O'Grady, N. Papinutto, D. Papp, D. Pareto, T. B. Parrish, A. Pichiecchio, F. Prados, À. Rovira, M. J. Ruitenberg, R. S. Samson, G. Savini, M. Seif, A. C. Seifert, A. K. Smith, S. A. Smith, Z. A. Smith, E. Solana, Y. Suzuki, G. Tackley, A. Tinnermann, J. Valošek, D. Van De Ville, M. C. Yiannakas, K. A. Weber 2nd, N. Weiskopf, R. G. Wise, P. O. Wyss, J. Xu, Generic acquisition protocol for quantitative MRI of the spinal cord. *Nat. Protoc.* **16**, 4611–4632 (2021).

9. G. C. Baxter, M. J. Graves, F. J. Gilbert, A. J. Patterson, A Meta-analysis of the Diagnostic Performance of Diffusion MRI for Breast Lesion Characterization. *Radiology*. **291**, 632–641 (2019).

10. H. Donato, M. França, I. Candelária, F. Caseiro-Alves, Liver MRI: From basic protocol to advanced techniques. *Eur. J. Radiol.* **93**, 30–39 (2017).

11. E. B. Tapper, A. S.-F. Lok, Use of Liver Imaging and Biopsy in Clinical Practice. *N. Engl. J. Med.* **377**, 756–768 (2017).

12. D. I. Tsilimigras, P. Brodt, P.-A. Clavien, R. J. Muschel, M. I. D'Angelica, I. Endo, R. W. Parks, M. Doyle, E. de Santibañes, T. M. Pawlik, Liver metastases. *Nat Rev Dis Primers*. **7**, 27 (2021).

13. E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D.

Lacombe, J. Verweij, New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer*. **45**, 228–247 (2009).

14. J. Y. Buikhuisen, A. Torang, J. P. Medema, Exploring and modelling colon cancer inter-tumour heterogeneity: opportunities and challenges. *Oncogenesis*. **9**, 66 (2020).

15. I. Vitale, E. Shema, S. Loi, L. Galluzzi, Intratumoral heterogeneity in cancer progression and response to immunotherapy. *Nat. Med.* **27**, 212–224 (2021).

16. A. D. Waldman, J. M. Fritz, M. J. Lenardo, A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat. Rev. Immunol.* **20**, 651–668 (2020).

17. C. Pilard, M. Ancion, P. Delvenne, G. Jerusalem, P. Hubert, M. Herfs, Cancer immunotherapy: it's time to better predict patients' response. *Br. J. Cancer*. **125**, 927–938 (2021).

18. J. P. B. O'Connor, E. O. Aboagye, J. E. Adams, H. J. W. L. Aerts, S. F. Barrington, A. J. Beer, R. Boellaard, S. E. Bohndiek, M. Brady, G. Brown, D. L. Buckley, T. L. Chenevert, L. P. Clarke, S. Collette, G. J. Cook, N. M. deSouza, J. C. Dickson, C. Dive, J. L. Evelhoch, C. Faivre-Finn, F. A. Gallagher, F. J. Gilbert, R. J. Gillies, V. Goh, J. R. Griffiths, A. M. Groves, S. Halligan, A. L. Harris, D. J. Hawkes, O. S. Hoekstra, E. P. Huang, B. F. Hutton, E. F. Jackson, G. C. Jayson, A. Jones, D.-M. Koh, D. Lacombe, P. Lambin, N. Lassau, M. O. Leach, T.-Y. Lee, E. L. Leen, J. S. Lewis, Y. Liu, M. F. Lythgoe, P. Manoharan, R. J. Maxwell, K. A. Miles, B. Morgan, S. Morris, T. Ng, A. R. Padhani, G. J. M. Parker, M. Partridge, A. P. Pathak, A. C. Peet, S. Punwani, A. R. Reynolds, S. P. Robinson, L. K. Shankar, R. A. Sharma, D. Soloviev, S. Stroobants, D. C. Sullivan, S. A. Taylor, P. S. Tofts, G. M. Tozer, M. van Herk, S. Walker-Samuel, J. Wason, K. J. Williams, P. Workman, T. E. Yankeelov, K. M. Brindle, L. M. McShane, A. Jackson, J. C. Waterton, Imaging biomarker roadmap for cancer studies. *Nat. Rev. Clin. Oncol.* **14**, 169–186 (2017).

19. X. Jiang, J. Xu, J. C. Gore, Mapping hepatocyte size in vivo using temporal diffusion spectroscopy MRI. *Magn. Reson. Med.* **84**, 2671–2683 (2020).

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

21. R. Gardier, J. L. Villarreal Haro, E. J. Canales-Rodríguez, I. O. Jelescu, G. Girard, J. Rafael-Patiño, J.-P. Thiran, Cellular Exchange Imaging (CEXI): Evaluation of a diffusion model including water exchange in cells using numerical phantoms of permeable spheres. *Magn. Reson. Med.* (2023), doi:10.1002/mrm.29720.

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

23. X. Jiang, S. Dudzinski, K. E. Beckermann, K. Young, E. McKinley, O. J McIntyre, J. C. Rathmell, J. Xu, J. C. Gore, MRI of tumor T cell infiltration in response to checkpoint inhibitor therapy. *J Immunother Cancer*. **8** (2020), doi:10.1136/jitc-2019-000328.

24. Z. Ye, R. L. Price, X. Liu, J. Lin, Q. Yang, P. Sun, A. T. Wu, L. Wang, R. H. Han, C. Song, R. Yang, S. E. Gary, D. D. Mao, M. Wallendorf, J. L. Campian, J.-S. Li, S. Dahiya, A. H. Kim, S.-K. Song, Diffusion Histology Imaging Combining Diffusion Basis Spectrum Imaging (DBSI) and Machine Learning Improves Detection and Classification of Glioblastoma Pathology. *Clin. Cancer Res.* **26**, 5388–5399 (2020).

25. A. F. Howard, J. Mollink, M. Kleinnijenhuis, M. Pallebage-Gamarallage, M. Bastiani, M. Cottaar, K. L. Miller, S. Jbabdi, Joint modelling of diffusion MRI and microscopy. *Neuroimage*. **201**, 116014 (2019).

26. D. S. Novikov, V. G. Kiselev, S. N. Jespersen, On modeling. *Magn. Reson. Med.* **79**, 3172–3193 (2018).

27. J. Xu, J. Xie, N. B. Semmineh, S. P. Devan, X. Jiang, J. C. Gore, Diffusion time dependency of extracellular diffusion. *Magn. Reson. Med.* **89**, 2432–2440 (2023).

28. C. H. Neuman, Spin echo of spins diffusing in a bounded medium. *J. Chem. Phys.* **60**, 4508–4511 (1974).

29. C. C. Conlin, C. H. Feng, A. E. Rodriguez-Soto, R. A. Karunamuni, J. M. Kuperman, D. Holland, R. Rakow-Penner, M. E. Hahn, T. M. Seibert, A. M. Dale, Improved Characterization of Diffusion in Normal and Cancerous Prostate Tissue Through Optimization of Multicompartmental Signal Models. *J. Magn. Reson. Imaging*. **53**, 628–639 (2021).

30. J. H. Jensen, J. A. Helpern, A. Ramani, H. Lu, Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *in Medicine: An …* (2005) (available at https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.20508).

31. S. J. Hectors, S. Semaan, C. Song, S. Lewis, G. K. Haines, A. Tewari, A. R. Rastinehad, B. Taouli, Advanced Diffusion-weighted Imaging Modeling for Prostate Cancer Characterization: Correlation with Quantitative Histopathologic Tumor Tissue Composition-A Hypothesis-generating Study. *Radiology*. **286**, 918–928 (2018).

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

33. G. Schwarz, Estimating the Dimension of a Model. *aos*. **6**, 461–464 (1978).

34. E. Panagiotaki, T. Schneider, B. Siow, M. G. Hall, M. F. Lythgoe, D. C. Alexander, Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. *Neuroimage*. **59**, 2241–2254 (2012).

35. U. Ferizi, T. Schneider, E. Panagiotaki, G. Nedjati-Gilani, H. Zhang, C. A. M. Wheeler-Kingshott, D. C. Alexander, A ranking of diffusion MRI compartment models with in vivo human brain data. *Magn. Reson. Med.* **72**, 1785–1792 (2014).

36. M. Palombo, V. Valindria, S. Singh, E. Chiou, F. Giganti, H. Pye, H. C. Whitaker, D. Atkinson, S. Punwani, D. C. Alexander, E. Panagiotaki, Joint estimation of relaxation and diffusion tissue parameters for prostate cancer with relaxation-VERDICT MRI. *Sci. Rep.* **13**, 2999 (2023).

37. X. Jiang, S. P. Devan, J. Xie, J. C. Gore, J. Xu, Improving MR cell size imaging by inclusion of transcytolemmal water exchange. *NMR Biomed.* **35**, e4799 (2022).

38. A. B. Rosenkrantz, E. E. Sigmund, A. Winnick, B. E. Niver, B. Spieler, G. R. Morgan, C. H. Hajdu, Assessment of hepatocellular carcinoma using apparent diffusion coefficient and diffusion kurtosis indices: preliminary experience in fresh liver explants. *Magn. Reson. Imaging*. **30**, 1534–1540 (2012).

39. W.-J. Tang, W. Yao, Z. Jin, Q.-C. Kong, W.-K. Hu, Y.-S. Liang, L.-X. Chen, S.-Y. Chen, Q.-Q. Zhang, X.-H. Wei, X.-D. Xu, Y. Guo, X.-Q. Jiang, Evaluation of the Effects of Anti-PD-1 Therapy on Triple-Negative Breast Cancer in Mice by Diffusion Kurtosis Imaging and Dynamic Contrast-Enhanced Imaging. *J. Magn. Reson. Imaging*. **56**, 1912–1923 (2022).

40. N. C. Martin, C. T. McCullough, P. G. Bush, L. Sharp, A. C. Hall, D. J. Harrison, Functional analysis of mouse hepatocytes differing in DNA content: volume, receptor expression, and effect of IFNgamma. *J. Cell. Physiol.* **191**, 138–144 (2002).

41. Z. Zhou, M.-J. Xu, B. Gao, Hepatocytes: a key cell type for innate immunity. *Cell. Mol. Immunol.* **13**, 301–315 (2016).

42. F. Grussu, K. Bernatowicz, I. Casanova-Salas, N. Castro, P. Nuciforo, J. Mateo, I. Barba, R. Perez-Lopez, Diffusion MRI signal cumulants and hepatocyte microstructure at fixed diffusion time: Insights from simulations, 9.4T imaging, and histology. *Magn. Reson. Med.* (2022), doi:10.1002/mrm.29174.

43. M. Morawski, E. Kirilina, N. Scherf, C. Jäger, K. Reimann, R. Trampel, F. Gavriilidis, S. Geyer, B. Biedermann, T. Arendt, N. Weiskopf, Developing 3D microscopy with CLARITY on human brain tissue: Towards a tool for informing and validating MRI-based histology. *Neuroimage*. **182**, 417–428 (2018).

44. I. O. Jelescu, J. Veraart, E. Fieremans, D. S. Novikov, Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue. *NMR Biomed.* **29**, 33–47 (2016).

45. M. Nilsson, G. Eklund, F. Szczepankiewicz, M. Skorpil, K. Bryskhe, C.-F. Westin, C. Lindh, L. Blomqvist, F. Jäderling, Mapping prostatic microscopic anisotropy using linear and spherical b-tensor encoding: A preliminary study. *Magn. Reson. Med.* **86**, 2025–2033 (2021).

46. E. Fokkinga, J. A. Hernandez-Tamames, A. Ianus, M. Nilsson, C. M. W. Tax, R. Perez-Lopez, F. Grussu, Advanced Diffusion-Weighted MRI for Cancer Microstructure Assessment in Body Imaging, and Its Relationship With Histology. *J. Magn. Reson. Imaging* (2023), doi:10.1002/jmri.29144.

47. J. H. Jensen, G. Russell Glenn, J. A. Helpern, Fiber ball imaging. *Neuroimage*. **124**, 824–833 (2016).

48. E. T. McKinnon, J. A. Helpern, J. H. Jensen, Modeling white matter microstructure with fiber ball imaging. *Neuroimage*. **176**, 11–21 (2018).

49. W. Warner, M. Palombo, R. Cruz, R. Callaghan, N. Shemesh, D. K. Jones, F. Dell'Acqua, A. Ianus, I. Drobnjak, Temporal Diffusion Ratio (TDR) for imaging restricted diffusion: Optimisation and pre-clinical demonstration. *Neuroimage*. **269**, 119930 (2023).

50. O. Reynaud, Time-Dependent Diffusion MRI in Cancer: Tissue Modeling and Applications. *Frontiers in Physics*. **5** (2017), doi:10.3389/fphy.2017.00058.

51. A. Serrablo, P. Paliogiannis, F. Pulighe, S. S.-M. Moro, V. Borrego-Estella, F. Attene, F. Scognamillo, C. Hörndler, Impact of novel histopathological factors on the outcomes of liver surgery for colorectal cancer metastases. *Eur. J. Surg. Oncol.* **42**, 1268–1277 (2016).

52. P. Nuciforo, T. Pascual, J. Cortés, A. Llombart-Cussac, R. Fasani, L. Paré, M. Oliveira, P. Galvan, N. Martínez, B. Bermejo, M. Vidal, S. Pernas, R. López, M. Muñoz, I. Garau, L. Manso, J. Alarcón, E. Martínez, V. Rodrik-Outmezguine, J. C. Brase, P. Villagrasa, A. Prat, E. Holgado, A predictive model of pathologic response based on tumor cellularity and tumor-infiltrating lymphocytes (CelTIL) in HER2-positive breast cancer treated with chemo-free dual HER2 blockade. *Ann. Oncol.* **29**, 170–177 (2018).

53. D. Gentile, A. Sagona, C. De Carlo, B. Fernandes, E. Barbieri, S. Di Maria Grimaldi, F. Jacobs, G. Vatteroni, L. Scardina, E. Biondi, V. Vinci, R. M. Trimboli, D. Bernardi, C. Tinterri, Pathologic response and residual tumor cellularity after neo-adjuvant chemotherapy predict prognosis in breast cancer patients. *Breast*. **69**, 323–329 (2023).

54. H.-H. Lee, A. Papaioannou, D. S. Novikov, E. Fieremans, In vivo observation and biophysical interpretation of time-dependent diffusion in human cortical gray matter. *Neuroimage*. **222**, 117054 (2020).

55. J.-P. Fortin, D. Parker, B. Tunç, T. Watanabe, M. A. Elliott, K. Ruparel, D. R. Roalf, T. D. Satterthwaite, R. C. Gur, R. E. Gur, R. T. Schultz, R. Verma, R. T. Shinohara, Harmonization of multi-site diffusion tensor imaging data. *Neuroimage*. **161**, 149–170 (2017).

56. S. Cetin Karayumak, S. Bouix, L. Ning, A. James, T. Crow, M. Shenton, M. Kubicki, Y. Rathi, Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters. *Neuroimage*. **184**, 180–200 (2019).

57. H. Boonstra, J. W. Oosterhuis, A. M. Oosterhuis, G. J. Fleuren, Cervical tissue shrinkage by formaldehyde fixation, paraffin wax embedding, section cutting and mounting. *Virchows Arch. A Pathol. Anat. Histopathol.* **402**, 195–201 (1983).

58. J. Xu, X. Jiang, S. P. Devan, L. R. Arlinghaus, E. T. McKinley, J. Xie, Z. Zu, Q. Wang, A. B. Chakravarthy, Y. Wang, J. C. Gore, MRI-cytometry: Mapping nonparametric cell size distributions using diffusion MRI. *Magn. Reson. Med.* **85**, 748–761 (2021).

59. J. Xu, X. Jiang, H. Li, L. R. Arlinghaus, E. T. McKinley, S. P. Devan, B. M. Hardy, J. Xie, H. Kang, A. B. Chakravarthy, J. C. Gore, Magnetic resonance imaging of mean cell size in human breast tumors. *Magn. Reson. Med.* **83**, 2002–2014 (2020).

60. O. Reynaud, K. V. Winters, D. M. Hoang, Y. Z. Wadghiri, D. S. Novikov, S. G. Kim, Pulsed and oscillating gradient MRI for assessment of cell size and extracellular space (POMACE) in mouse gliomas. *NMR Biomed.* **29**, 1350–1363 (2016).

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

62. Y. T. Li, J.-P. Cercueil, J. Yuan, W. Chen, R. Loffroy, Y. X. J. Wáng, Liver intravoxel incoherent motion (IVIM) magnetic resonance imaging: a comprehensive review of published data on normal values and applications for fibrosis and tumor evaluation. *Quant. Imaging Med. Surg.* **7**, 59–78 (2017).

63. B. Balinov, B. Jonsson, P. Linse, O. Soderman, The NMR Self-Diffusion Method Applied to Restricted Diffusion. Simulation of Echo Attenuation from Molecules in Spheres and between Planes. *J. Magn. Reson. A*. **104**, 17–25 (1993).

64. N. P. Jerome, J. A. d'Arcy, T. Feiweier, D.-M. Koh, M. O. Leach, D. J. Collins, M. R. Orton, Extended T2-IVIM model for correction of TE dependence of pseudo-diffusion volume fraction in clinical diffusion-weighted magnetic resonance imaging. *Phys. Med. Biol.* **61**, N667–N680 (2016).

65. J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-Aron, J. Sijbers, E. Fieremans, Denoising of diffusion MRI using random matrix theory. *Neuroimage*. **142**, 394–406 (2016).

66. E. Kellner, B. Dhital, V. G. Kiselev, M. Reisert, Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magn. Reson. Med.* **76**, 1574–1581 (2016).

67. E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. van der Walt, M. Descoteaux, I. Nimmo-Smith, Dipy Contributors, Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* **8**, 8 (2014).

68. S. Ourselin, A. Roche, G. Subsol, X. Pennec, N. Ayache, Reconstructing a 3D structure from serial histological sections. *Image Vis. Comput.* **19**, 25–31 (2001).

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

70. B. B. Avants, C. L. Epstein, M. Grossman, J. C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).

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

Supplementary Materials

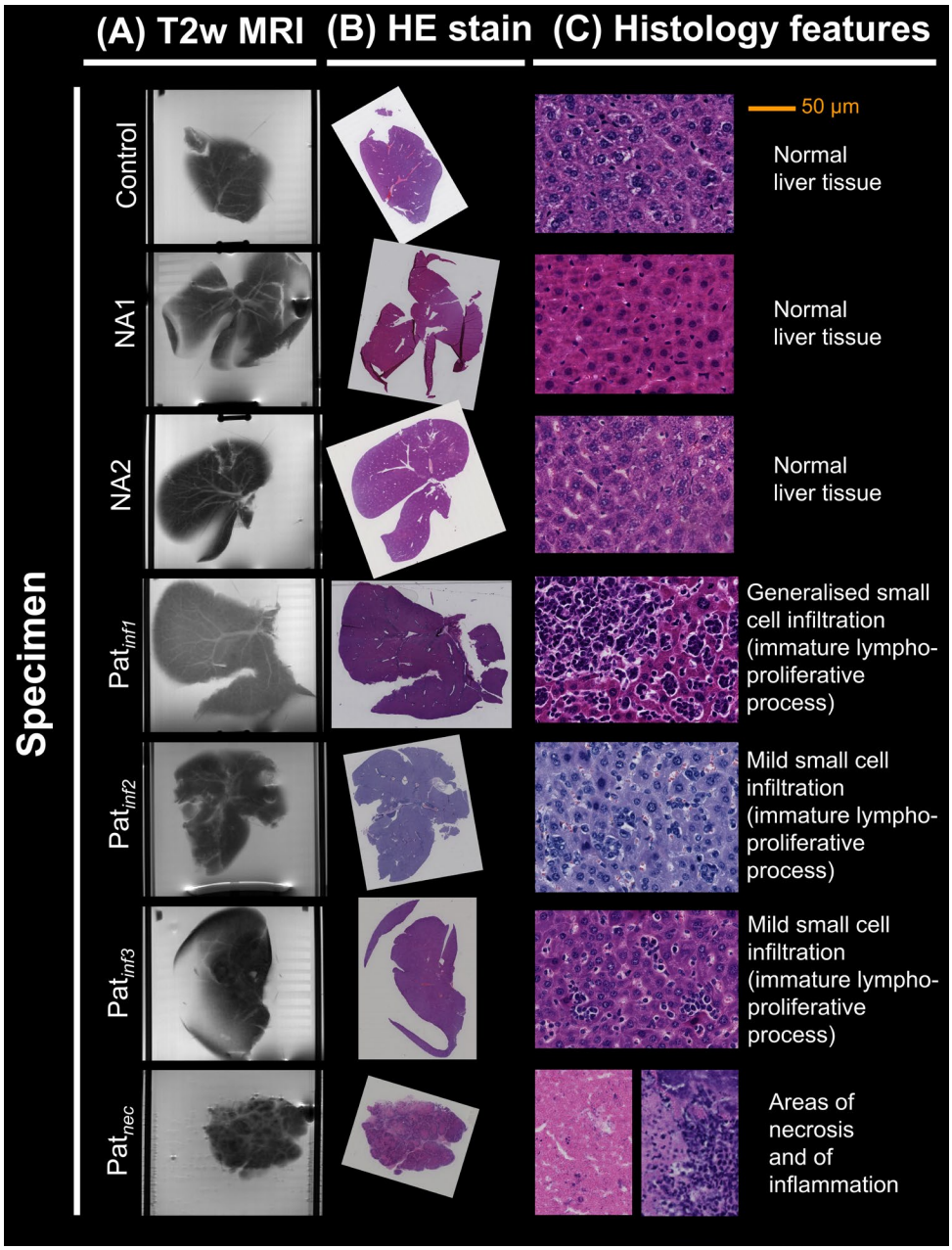

Fig. S1: radiological-histological co-localisation of the *ex vivo* mouse liver data.

Illustration of the radiological-histological co-localisation on the 7 fixed mouse livers obtained from mice implanted with a biopsy from a prostate cancer patient. **(A)**, left: illustrative slice of the high-resolution anatomical T2-weighted fast spin echo. **(B)**, centre: hematoxylin and eosin (HE)-stained section, taken from the MRI slice shown to the left. **(C)**, right: detail of the microstructure characterising each specimen, as assessed by an experienced pathologist (SS). Different specimens are arranged along different rows. From top to bottom: *Control*, normal liver structures (no biopsy implantation); Pat_{NA1} and Pat_{NA2}, normal appearing normal liver structures after prostate cancer biopsy implantation; Pat_{infl}, Pat_{inf2} and Pat_{inf3}: pathology following implantation, consisting of an immature, lympho-proliferative process (infiltration of small cells in sinusoidal spaces); Pat_{nec}, pathology following implantation, consisting of necrosis and inflammation.

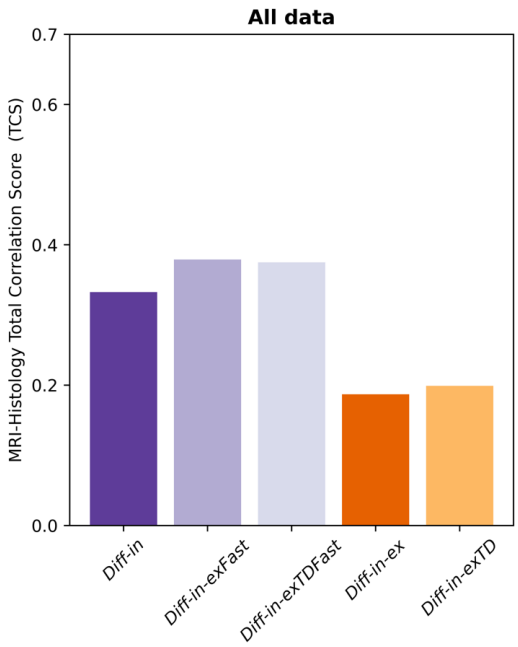

Fig. S2: MRI-histology Total Correlation Score (TCS) for biophysical dMRI model selection, as obtained when fitting dMRI signal models on the whole image set.

Values of TCS for all models, as obtained by fitting models on the whole image set (set of images with negligible vascular signal contributions, i.e., $b > 1000$ s/mm^2 in the fixed mouse livers and $b > 100$ s/mm^2 *in vivo*). We evaluated TCS for histology-informed model selection using a sample size of N = 25.

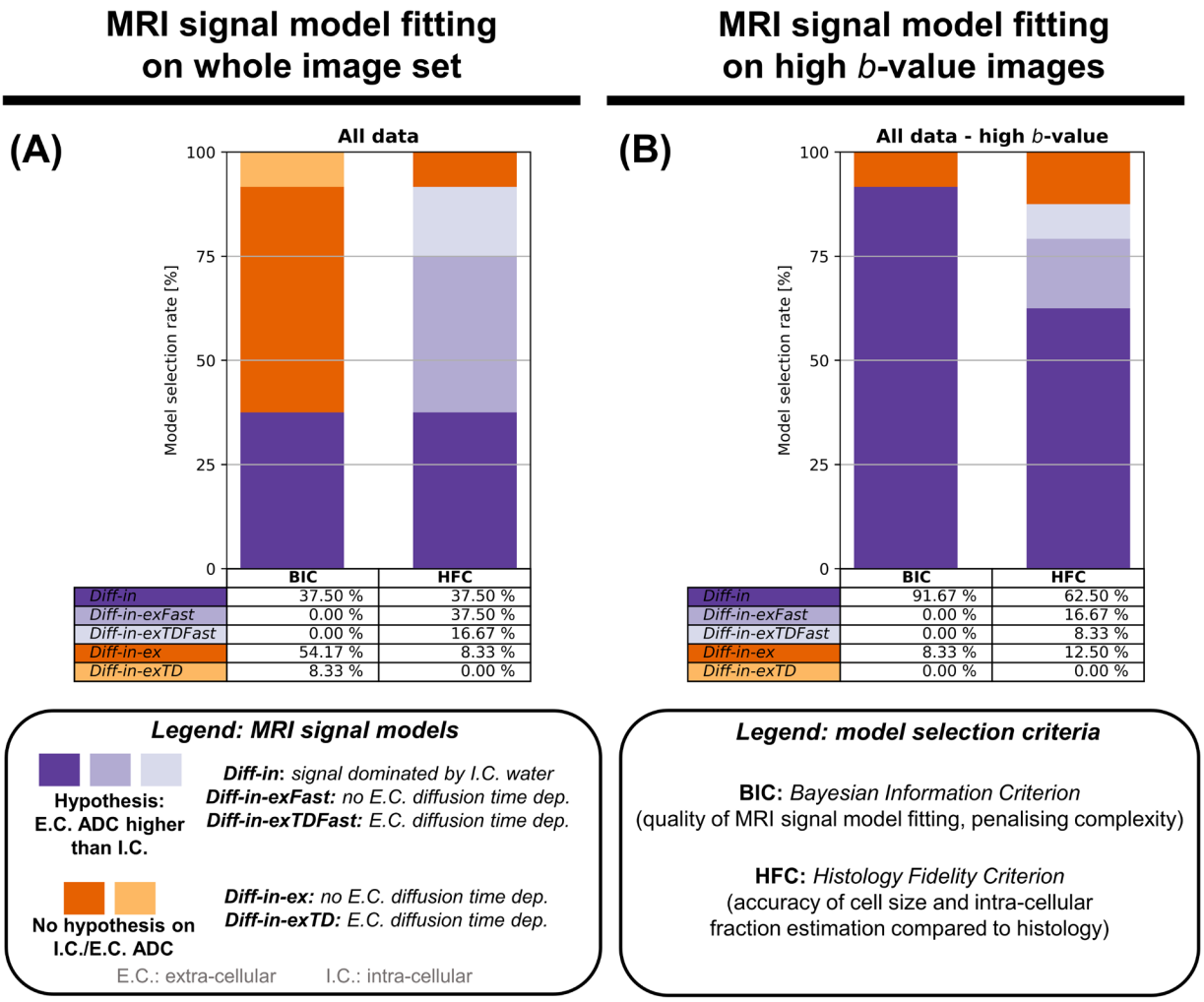

Fig. S3: Biophysical dMRI signal model selection based on BIC and HFC criteria.

Frequency of model selection based on the *Bayesian Information Criterion* (BIC, quantifying how well a model fits the dMRI signals, penalising model complexity) and on the *Histology-fidelity Criterion* (HFC, quantifying how accurately a dMRI models estimates the intra-cellular fraction and the volume-weighted cell size as seen on histology). (**A**) reports results when models are fitted to the entire set of measurements with negligible vascular signal constributions ($b > 1000$ s/mm^2 for suppression of PBS fluid within vessels in the fixed mouse livers; $b > 100$ s/mm^2 for IVIM signal suppression *in vivo* on clinical systems), while (**B**) reporting results obtained when fitting models only on high b-value images ($b > 1800$ s/mm^2 in the fixed mouse livers; $b > 900$ s/mm^2 *in vivo*).

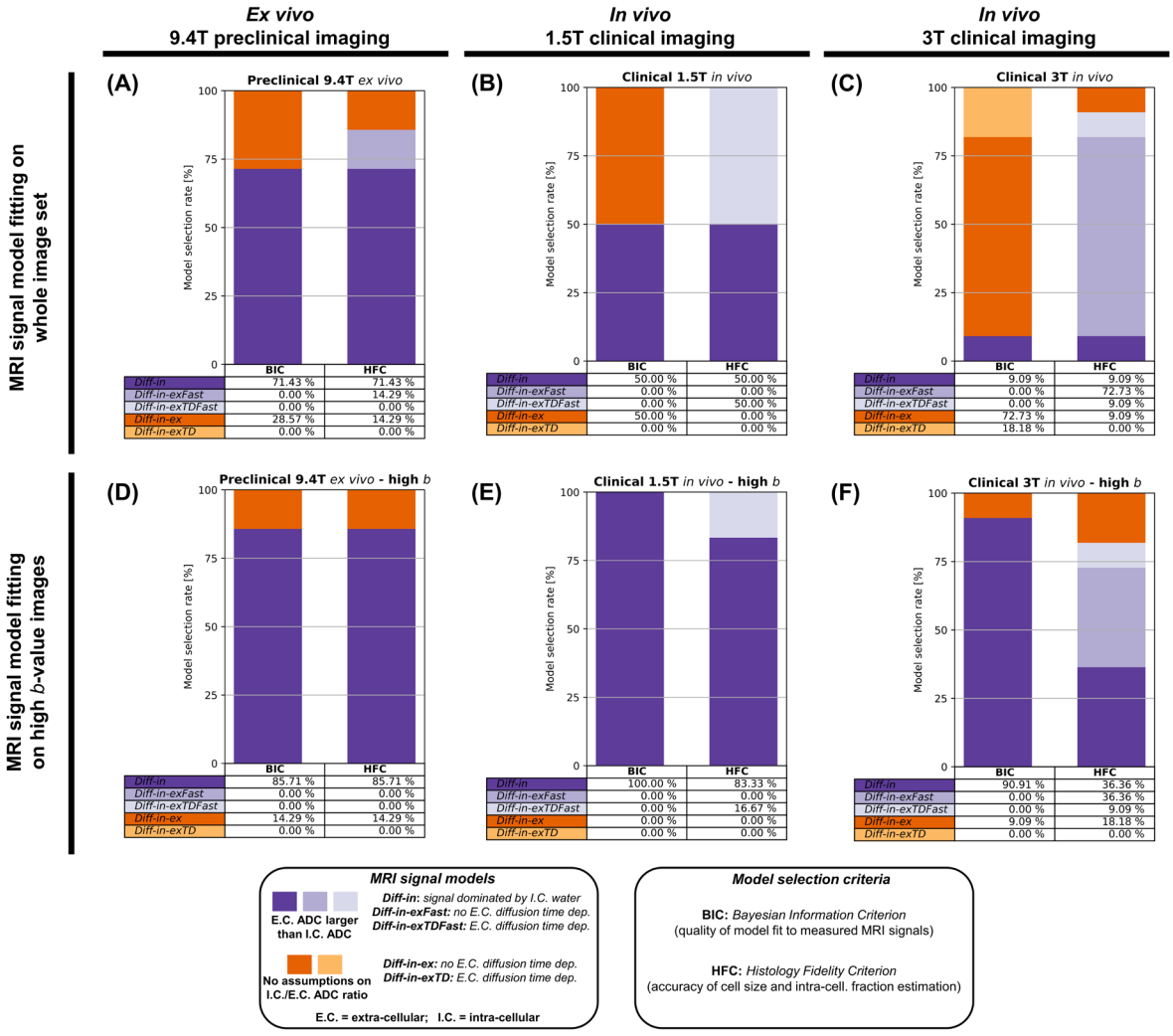

Fig. S4: biophysical model selection across different MRI scanners and data subsets.

Frequency of selection of each of 5 biophysical dMRI models on 3 MRI-histology data subsets. First column: selection on 7 fixed mouse livers scanned *ex vivo* on a preclinical 9.4T Bruker system (**A** and **D**, left); Second column: selection on 6 liver tumours imaged *in vivo* on a clinical 1.5T Siemens system (**B** and **E**, middle); Third column: selection on 12 liver tumours imaged *in vivo* on a clinical 3T GE system (**C** and **F**, right). Plots on top (**A** to **C**) refer to dMRI model fitting performed on images where the vascular signal was suppressed ("whole image set fitting", $b >$ 1000 s/mm^2 for suppression of PBS fluid within vessels on the 9.4T; $b >$ 100 s/mm^2 for IVIM signal suppression on clinical systems). Plots to the bottom (**D** to **F**) refer to dMRI model fitting performed on images where both vascular and extra-cellular, extra-vascular signals were suppressed ("high b-value fitting", $b >$ 1800 s/mm^2 on the 9.4T; $b >$ 900 s/mm^2 on clinical systems). Violet: models where extra-cellular ADC is larger than intra-cellular ADC; orange: models with no constraints on which is larger between intra-/extra-cellular ADC. The *Bayesian Information Criterion* (BIC) selects a model depending on the goodness of MRI signal fitting. The *Histology Fidelity Criterion* (HFC) selects a model depending on the overall agreement between MRI volume-weighted Cell Size (*vCS*) and intra-cellular fraction (*F*) with their histology counterparts.

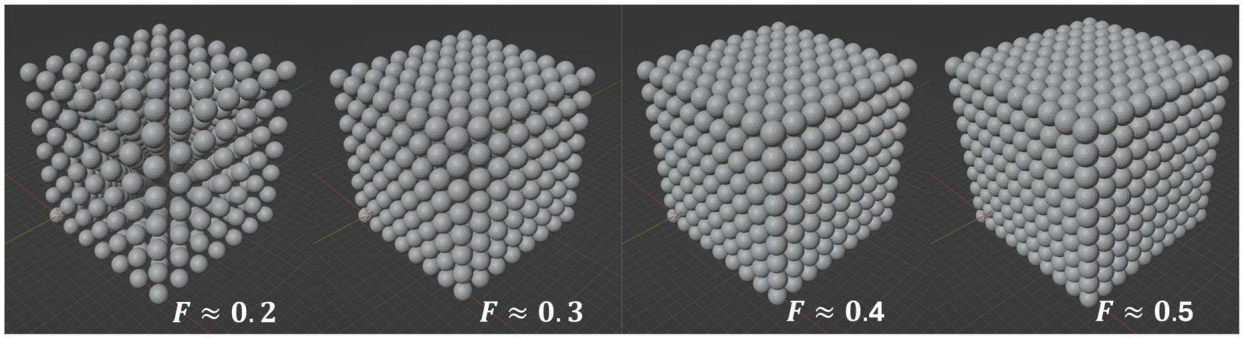

Fig. S5: environments used to generate synthetic dMRI signals in computer simulations.
The synthetic environment consisted of meshed spheres of fixed diameter, representing cells, as this is a common biophysical model used in several dMRI techniques (e.g., VERDICT, IMPULSED). We used the synthetic environment to generate dMRI signals via Monte Carlo simulations for each of the 3 dMRI protocols considered in this study (the PGSE protocol used on the *ex vivo* mouse livers at 9.4T; the PGSE protocol used in patients *in vivo* at 3T; the DW TRSE protocol used in patients *in vivo* at 1.5T). Afterwards, we performed performed dMRI model selection on the synthetic signals, following the same procedures implemented for actual *ex vivo* and *in vivo* dMRI data. We controlled the intra-sphere fraction F by adding gaps of increasing size in-between abutting spheres packed in an ideal cubic lattice. We probed 4 different values of F (approximately equal to 0.197, 0.323, 0.406, 0.523; notice that the maximum theoretical value of F for cubic lattice packing is equal to 0.5236). For each value of F, we varied the cell diameter (8, 16, 22 and 30 μm), intra-sphere diffusivity (10 linearly-spaced values in the ranges [0.8; 2.6] $\mu m^2\ ms^{-1}$ and [0.8; 3.0] $\mu m^2\ ms^{-1}$ for the *ex vivo* and *in vivo* protocols respectively) and extra-sphere intrinsic diffusivity (again, 10 linearly-spaced values in the ranges [0.8; 2.6] $\mu m^2\ ms^{-1}$ and [0.8; 3.0] $\mu m^2\ ms^{-1}$ for the *ex vivo* and *in vivo* protocols respectively), generating a total of 1600 synthetic voxels. Before dMRI signal model fitting, we corrupted synthetic signal with Rician noise at a signal-to-noise ratio (SNR) of 30 on the b = 0 signal $s(b=0)$ (SNR = $s(b=0)/\sigma$, where σ is the noise standard deviation).

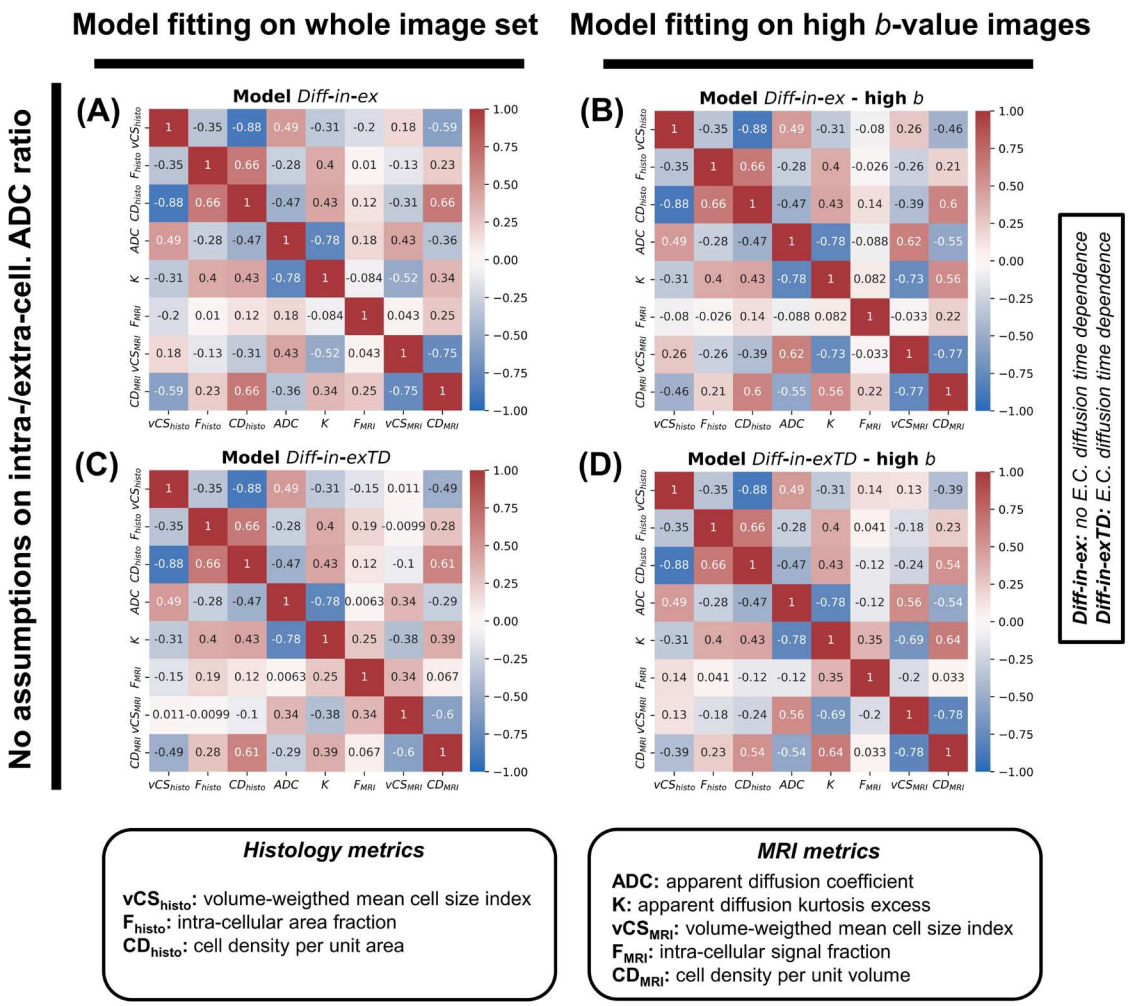

Fig. S6: MRI-histology correlations for models with no assumptions on which is larger between intra-cellular and extra-cellular ADC. Matrices illustrating Pearson's correlation coefficients among all possible pairs of MRI and histology metrics. Histological metrics are: intra-cellular area fraction F_{histo}; volume-weighted mean cell size index vCS_{histo}; cell density per unit area CD_{histo}. MRI metrics are: apparent diffusion coefficient ADC; apparent diffusion excess kurtosis K; intra-cellular area fraction F_{MRI}; volume-weighted mean cell size index vCS_{MRI}; cell density per unit area CD_{MRI}. Metrics F_{MRI}, vCS_{MRI} and CD_{MRI} were obtained by fitting models with no assumptions on which is larger between intra-cellular and extra-cellular ADC (*Diff-in-ex* and *Diff-in-exTD*). The 4 panels refer to models *Diff-in-ex* and *Diff-in-exTD* fitted according to 2 different strategies. Panel (**A**): model *Diff-in-ex* fitted on the whole set of measurements with vascular signal suppression (b > 100 s/mm^2 *in vivo*, b > 1000 s/mm^2 *ex vivo*); panel (**B**): model *Diff-in-ex* fitted on high b-value measurements (b > 900 s/mm^2 *in vivo*, b > 1800 s/mm^2 *ex vivo*); panel (**C**): model *Diff-in-exTD* fitted on the whole set of measurements with vascular signal suppression; panel (**D**): model *Diff-in-exTD* fitted on high b-value measurements. We calculated correlation coefficients using a sample size of N = 25 (Fig. 1).

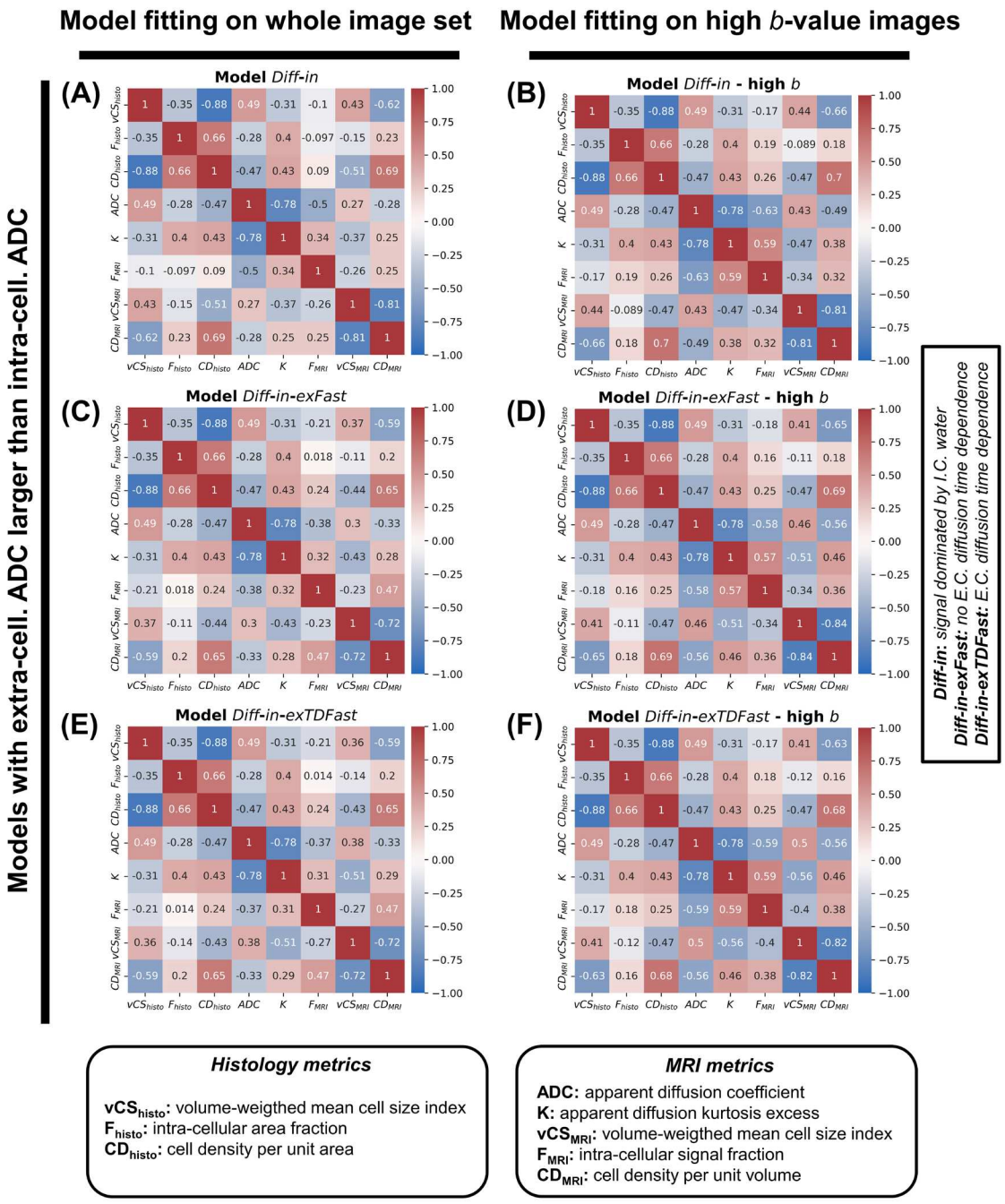

Fig. S7: MRI-histology correlations for models where the extra-cellular ADC is constrained to be larger than the intra-cellular ADC. Matrices illustrating Pearson's correlation coefficients among all possible pairs of MRI and histology metrics. Histological metrics are: intra-cellular area fraction F_{histo}; volume-weighted mean cell size index vCS_{histo}; cell density per unit area CD_{histo}. MRI metrics are: apparent diffusion coefficient ADC; apparent diffusion excess kurtosis K; intra-cellular area fraction F_{MRI}; volume-weighted mean cell size index vCS_{MRI}; cell density per unit area CD_{MRI}. Metrics F_{MRI}, vCS_{MRI} and CD_{MRI} were obtained by fitting models that assume that the extra-cellular ADC is always larger than the intra-cellular ADC (*Diff-in*, *Diff-in-exFast* and *Diff-in-exTDFast*). The 6 panels refer to models *Diff-in*, *Diff-in-exFast* and *Diff-in-exTDFast* fitted

according to 2 different strategies. Panel (**A**): model *Diff-in* fitted on the whole set of measurements with vascular signal suppression ($b > 100$ s/mm^2 *in vivo*, $b > 1000$ s/mm^2 *ex vivo*); panel (**B**): model *Diff-in* fitted on high b-value measurements ($b > 900$ s/mm^2 *in vivo*, $b > 1800$ s/mm^2 *ex vivo*); panel (**C**): model *Diff-in-exFast* fitted on the whole set of measurements with vascular signal suppression; panel (**D**): model *Diff-in-exFast* fitted on high b-value measurements; panel (**E**): model *Diff-in-exTDFast* fitted on the whole set of measurements with vascular signal suppression; panel (**F**): model *Diff-in-exTDFast* fitted on high b-value measurements. We calculated correlation coefficients using a sample size of N = 25 (Fig. 1).

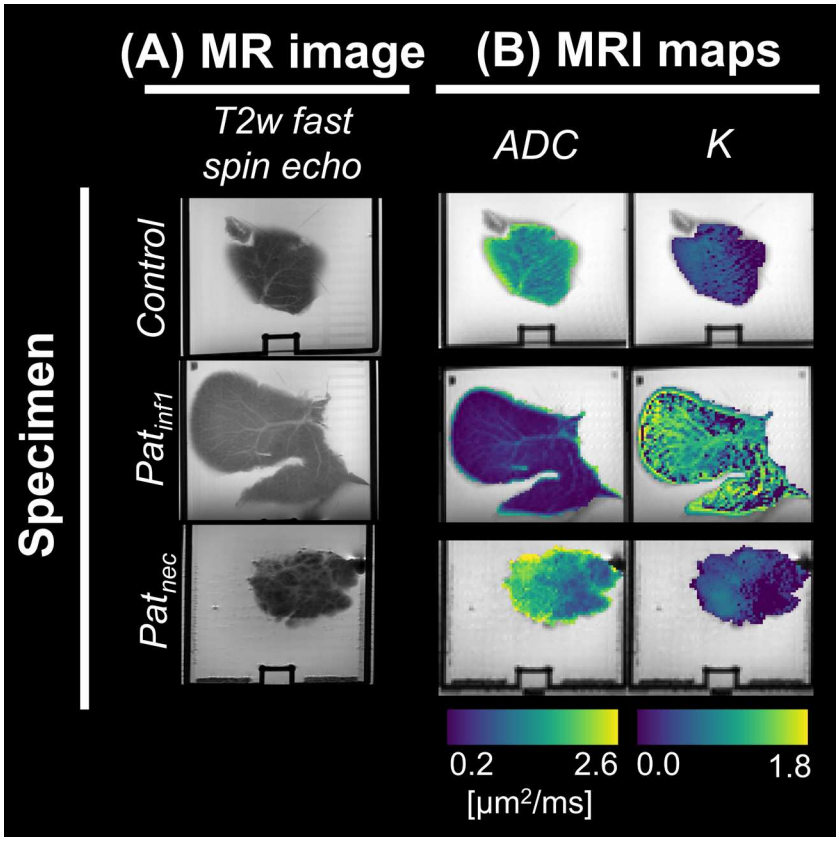

Fig. S8: standard diffusion MRI metrics in fixed mouse livers *ex vivo*.

(A): high-resolution fast spin echo scan acquired in fixed mouse livers scanned *ex vivo* on the 9.4T Bruker system. **(B)**: standard diffusion metrics, namely *ADC* (apparent diffusion coefficient) and *K* (apparent diffusion kurtosis excess). These metrics were obtained by fitting the standard diffusion kurtosis signal representation $s = s_0 \exp(- b\ ADC + K\ (b\ ADC)^2/6)$ to the set of measurements at fixed TE = 45 ms and Δ = 30 ms. From top to bottom, the figure reports maps from 3 specimens, representative of the 3 different microstructural phenotypes seen in our mouse liver data. These are: normal liver structures (illustrated by the *Control* case, e.g., mouse with no biopsy implantation); pathology following biopsy implantation, consisting of an immature, lympho-proliferative process (infiltration of small cells in sinusoidal spaces, illustrated by case Pat_{inf}); pathology following biopsy implantation, consisting of necrosis and inflammation (illustrated by case Pat_{nec}).

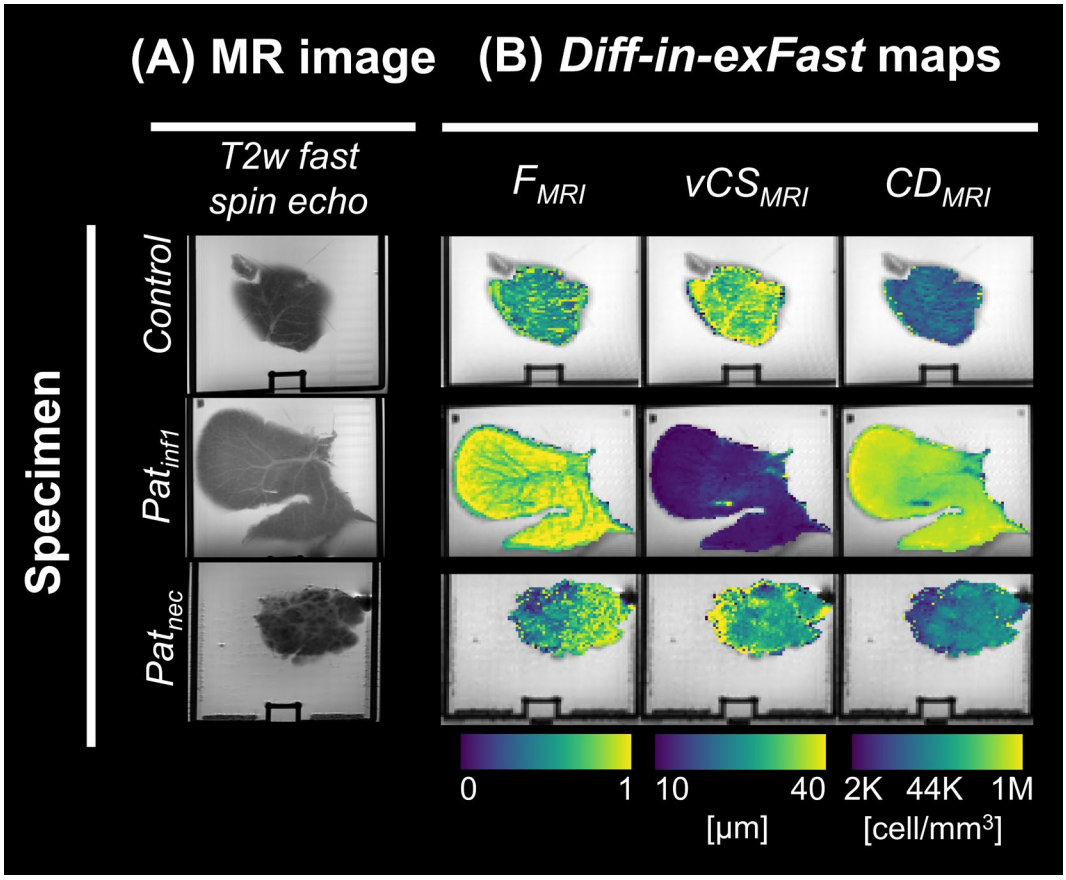

Fig. S9: key parametric maps of the *Diff-in-exFast* model on *ex vivo* mouse livers.

(A): high-resolution fast spin echo scan acquired in fixed mouse livers scanned *ex vivo* on the 9.4T Bruker system. **(B)**: metrics from the *Diff-in-exFast* model fitted to the whole DW image set (b-values with negligible vascular signal contributions, i.e., $b > 1000$ s/mm^2 on fixed *ex vivo* tissue, to suppress signal from PBS-filled vessels). From left to right: intra-cellular signal fraction F_{MRI}; volume-weighted cell size index vCS_{MRI} index; cell density per unit volume CD_{MRI}. Maps from 3 specimens are reported along different rows. The specimens are representative of the 3 different microstructural phenotypes seen in our mouse liver data. From top to bottom, these are: normal liver structures (illustrated by the *Control* case, e.g., mouse with no biopsy implantation); pathology following biopsy implantation, consisting of an immature, lympho-proliferative process (infiltration of small cells in sinusoidal spaces, illustrated by case Pat_{inf}); pathology following biopsy implantation, consisting of extended necrosis and inflammation (illustrated by case Pat_{nec}).

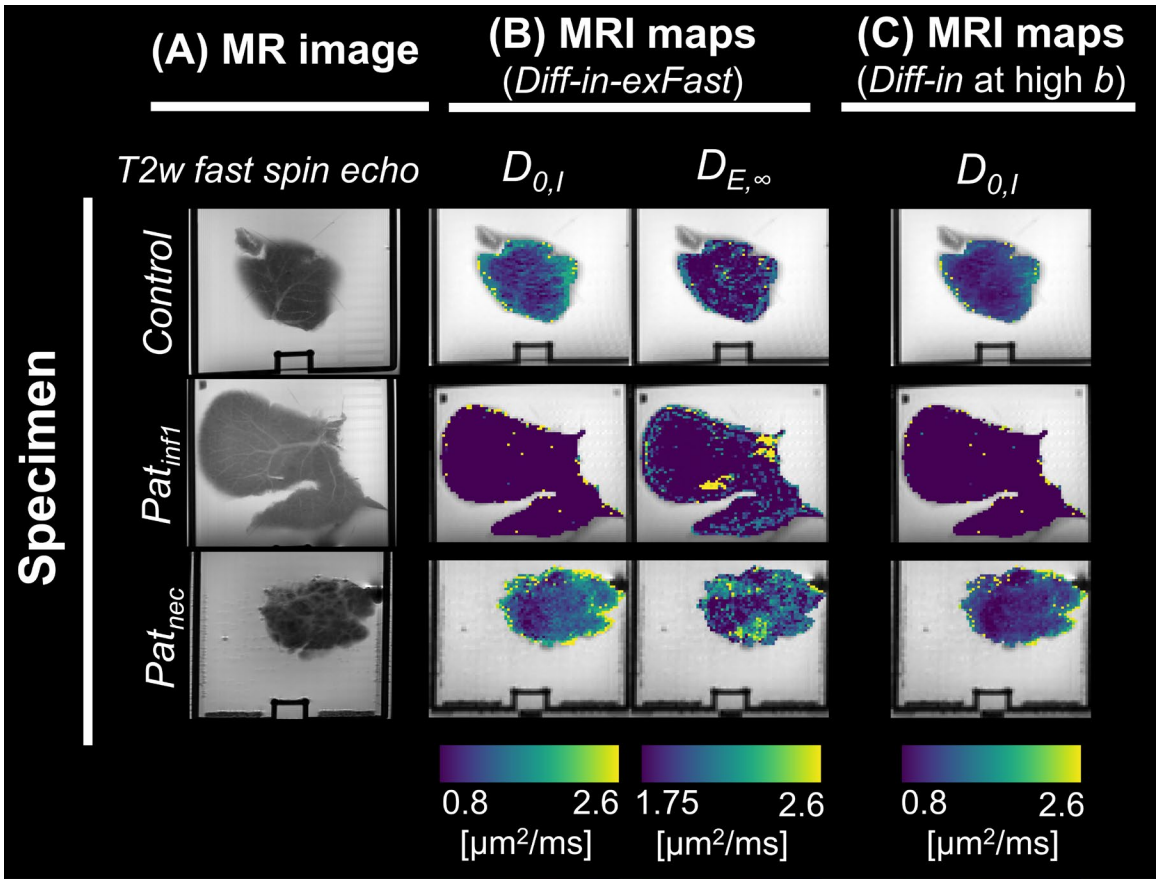

Fig. S10: diffusivity metrics from biophysical MRI models in fixed *ex vivo* mouse livers.

(A): high-resolution fast spin echo scan acquired in fixed mouse livers scanned *ex vivo* on the 9.4T Bruker system. **(B)**: diffusivity metrics from biophysical model *Diff-in-exFast*, namely: intrinsic intra-cellular cytosolic diffusivity $D_{0,I}$; asymptotic extra-cellular diffusion coefficient $D_{E,\infty}$. **(C)**: intrinsic intra-cellular cytosolic diffusivity $D_{0,I}$ from model *Diff-in* fitted to high b-value images (b > 1800 s/mm^2). Maps from 3 specimens are reported along different rows. The specimens are representative of the 3 different microstructural phenotypes seen in our mouse liver data. From top to bottom, these are: normal liver structures (illustrated by the *Control* case, e.g., mouse with no biopsy implantation); pathology following biopsy implantation, consisting of an immature, lympho-proliferative process (infiltration of small cells in sinusoidal spaces, illustrated by case Pat_{inf}); pathology following biopsy implantation, consisting of extended necrosis and inflammation (illustrated by case Pat_{nec}).

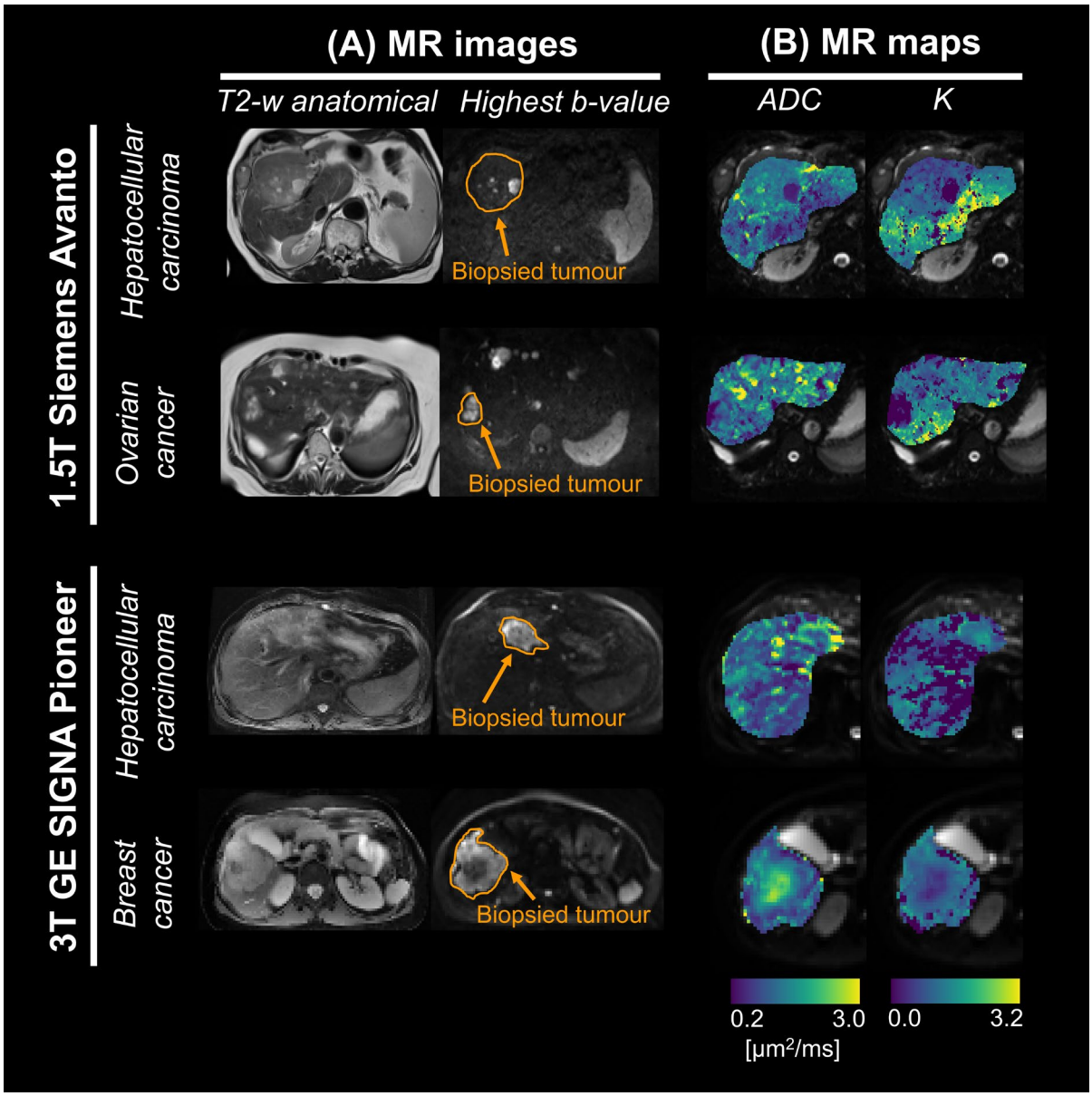

Fig. S11: standard diffusion MRI metrics in patients *in vivo*.

(A): high-resolution fast spin echo scan as well as a high b-value diffusion image, with biopsied tumour outlined. **(B)**: standard diffusion metrics in the biopsied tumour. Metrics are: apparent diffusion coefficient (*ADC*) and apparent diffusion kurtosis excess (*K*). These were obtained by fitting the standard diffusion kurtosis signal representation $s = s_0 \exp(-b\,ADC + K\,(b\,ADC)^2/6)$ to the set of measurements at fixed, minimum TE and $b > 100$ s/mm^2. Maps are shown in four representative patients (two patients for each MRI scanner), along different rows. For the 1.5T Siemens scanner (first and second rows from top): patient 6 (primary hepatocellular carcinoma) and patient 3 (liver metastases from ovarian cancer). For the 3T GE scanner (third and fourth rows from top): patient 24 (primary hepatocellular carcinoma) and patient 30 (liver metastases from breast cancer).

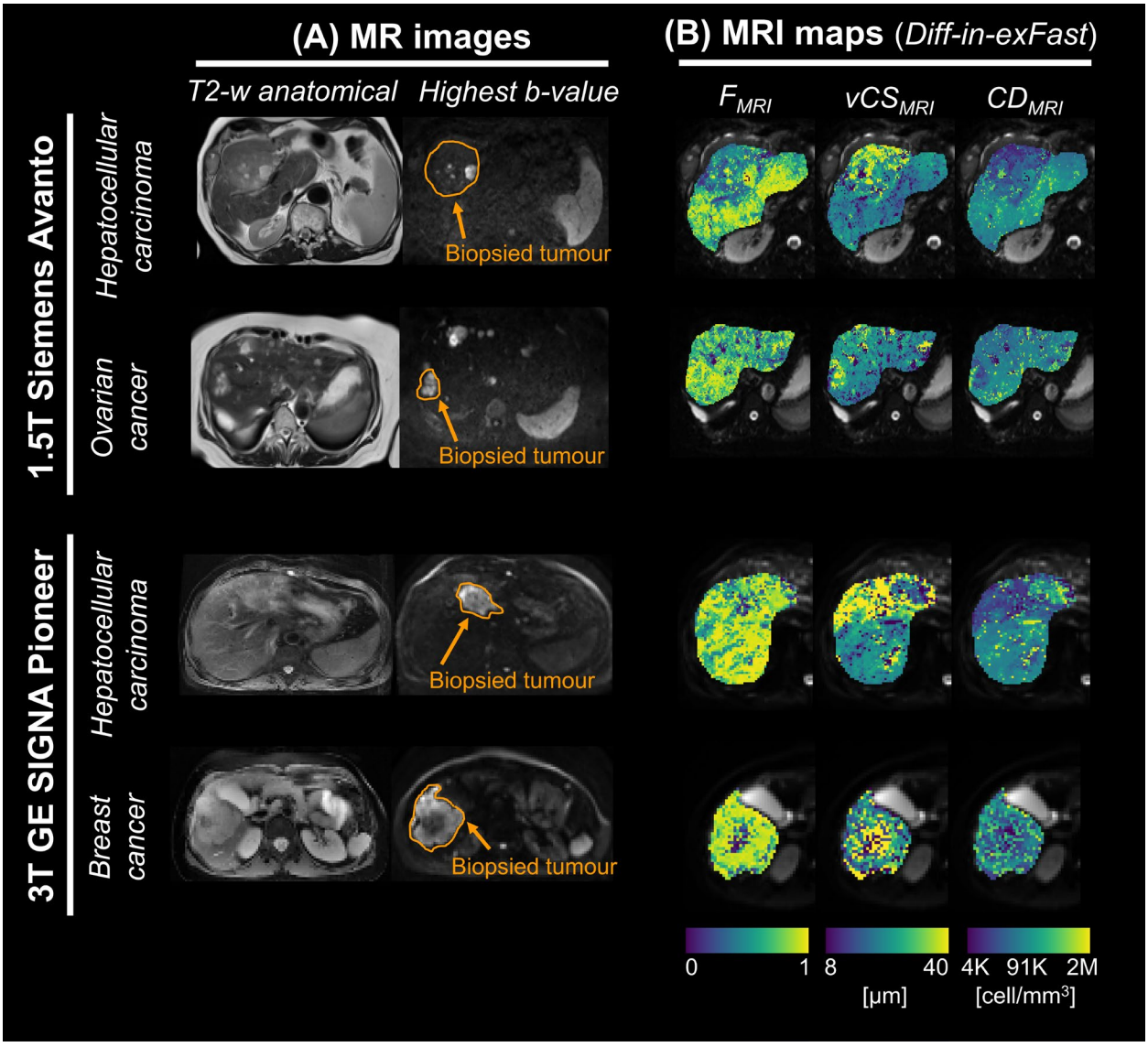

Fig. S12: key parametric maps of the *Diff-in-exFast* model in patients *in vivo*.

(A): high-resolution fast spin echo scan as well as a high b-value diffusion image, with biopsied tumour outlined. **(B)**: salient metrics of the *Diff-in-exFast* model fitted to the whole set of images with negligible vascular signal contributions (b > 100 s/mm^2). Metrics are shown in the biopsied tumour. From left to right: intra-cellular signal fraction F_{MRI}; volume-weighted cell size index vCS_{MRI} index; cell density per unit volume CD_{MRI}. Metrics are shown in four representative patients (two patients for each MRI scanner), along different rows. For the 1.5T Siemens scanner (first and second rows from top): patient 6 (primary hepatocellular carcinoma) and patient 3 (liver metastases from ovarian cancer). For the 3T GE scanner (first and second rows from bottom): patient 24 (primary hepatocellular carcinoma) and patient 30 (liver metastases from breast cancer).

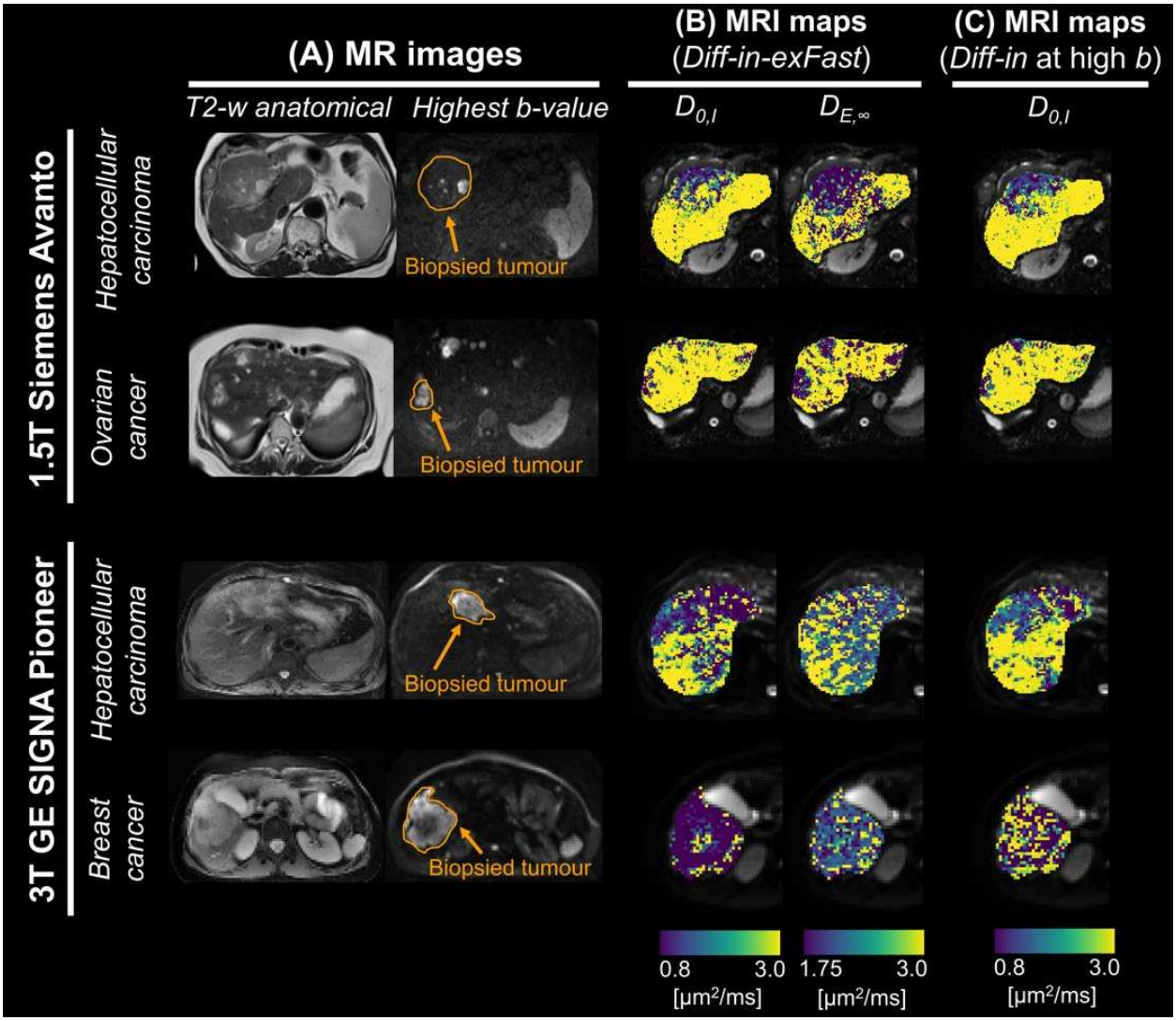

Fig. S13: diffusivity metrics from biophysical MRI models in patients *in vivo*.

(A): high-resolution fast spin echo scan as well as a high b-value diffusion image, with biopsied tumour outlined. **(B)**: diffusivity maps from biophysical model *Diff-in-exFast* in the biopsied tumour. Metrics are: intra-cellular cytosolic diffusivity $D_{0,I}$ and asymptotic extra-cellular diffusion coefficient $D_{E,\infty}$. **(C)**: intra-cellular cytosolic diffusivity $D_{0,I}$ for biophysical model *Diff-in* fitted being fitted only to high b-value images ($b > 900$ s/mm^2). Metrics are shown in four representative patients (two patients for each MRI scanner), along different rows. For the 1.5T Siemens scanner (first and second rows from top): patient 6 (primary hepatocellular carcinoma) and patient 3 (liver metastases from ovarian cancer). For the 3T GE scanner (first and second rows from bottom): patient 24 (primary hepatocellular carcinoma) and patient 30 (liver metastases from breast cancer).

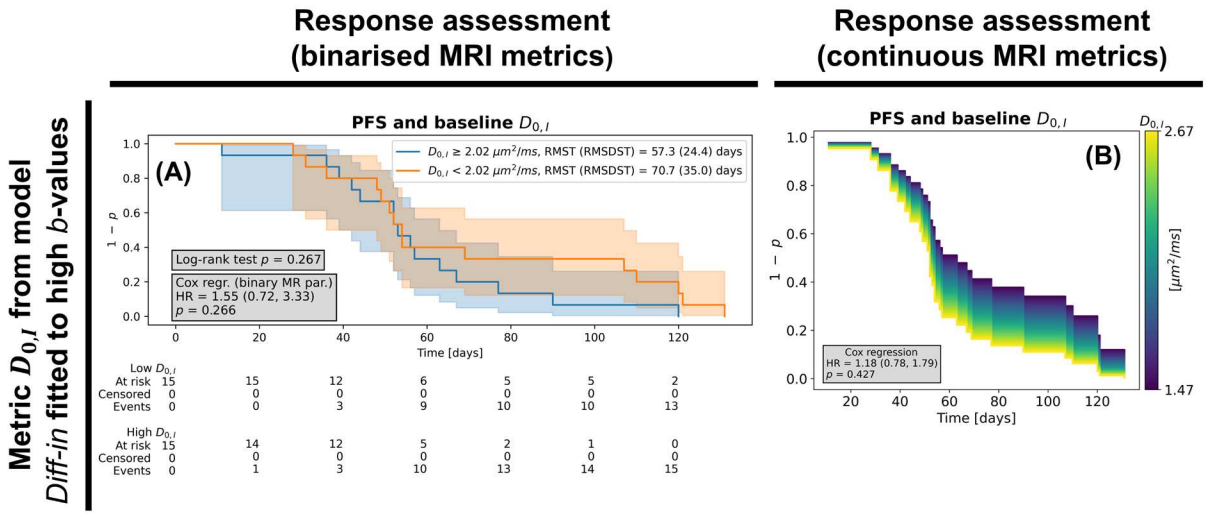

Fig. S14: immunotherapy response assessment based on *Diff-in* cytosol diffusivity estimates.

This figure reports on the dependence of patients' progression-free survival (PFS) on the average value of the intrinsic intra-cellular cytosol diffusivity $D_{0,I}$ within liver tumours at baseline (i.e., before starting immunotherapy), as obtained by fitting model *Diff-in* at high b-value. Left (panel **A**): Kaplan-Meier (KM) survival curves of two groups obtained by splitting patients based on baseline $D_{0,I}$ (lower or higher than the sample median). The grey panel reports the p-values of a log-rank sum test comparing the KM curves, and of a Cox regression based on the binarised MRI metric (with the corresponding hazard ratio (HR) estimate and 95% confidence interval). The legend reports the Restricted Mean Survival Time (RMST) and Restricted Standard Deviation of Survival Time (RSDST) for each KM curve. Right (panel **B**): results from univariate Cox regression where the baseline $D_{0,I}$ is a continuous predictor of the survival. The panel shows how changes in baseline $D_{0,I}$ modulate the survival curve, given the HR estimated for each metric. In the grey box, the p-value and HR (with 95% CI) corresponding to the baseline MRI metric are reported. In all panels, the y-axis shows $1 - p$, with p being the probability of progression, while the x-axis shows the time to progression (in days).

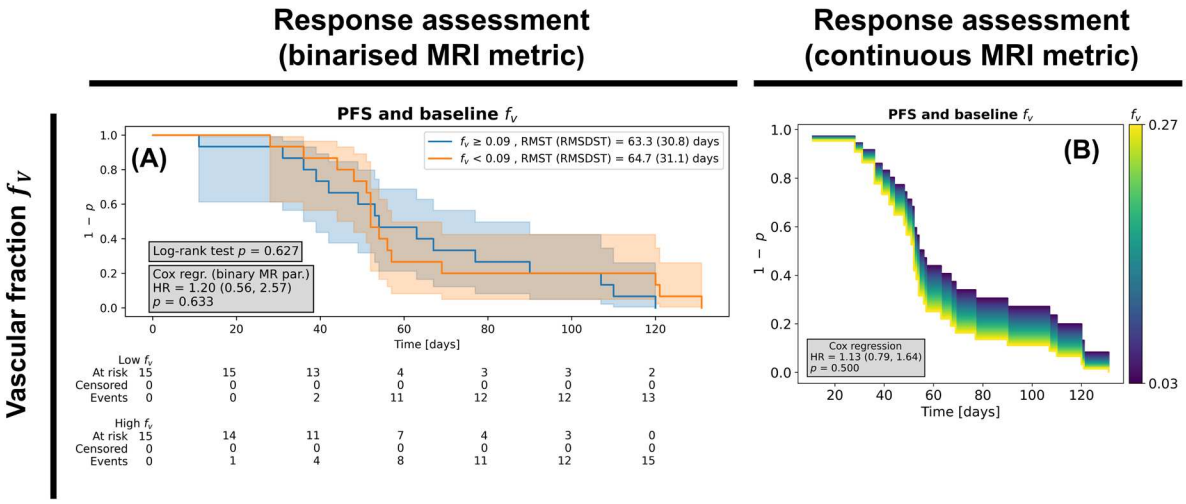

Fig. S15: immunotherapy response assessment based on vascular fraction estimates.

This figure reports on the dependence of patients' progression-free survival (PFS) on the average value of the vascular signal fraction f_v within liver tumours at baseline (i.e., before starting immunotherapy). The vascular signal fraction f_v was computed in the initial fitting step, which disentangles the vascular from the non-vascular (tissue) signal, before the latter is split into intra-/extra-cellular contributions in the biophysical model fitting step. The same representation layout as in Fig. S6 was used. Left (**A**): Kaplan-Meier (KM) analysis, log-rank sum test and Cox regression based on the binarised f_v (higher/lower than the sample median). Right (**B**): Cox regression modelling the probability of survival as a continuous function of baseline f_v. In all panels, the y-axis shows $1 - p$, with p being the probability of progression, while the x-axis shows the time to progression (in days).

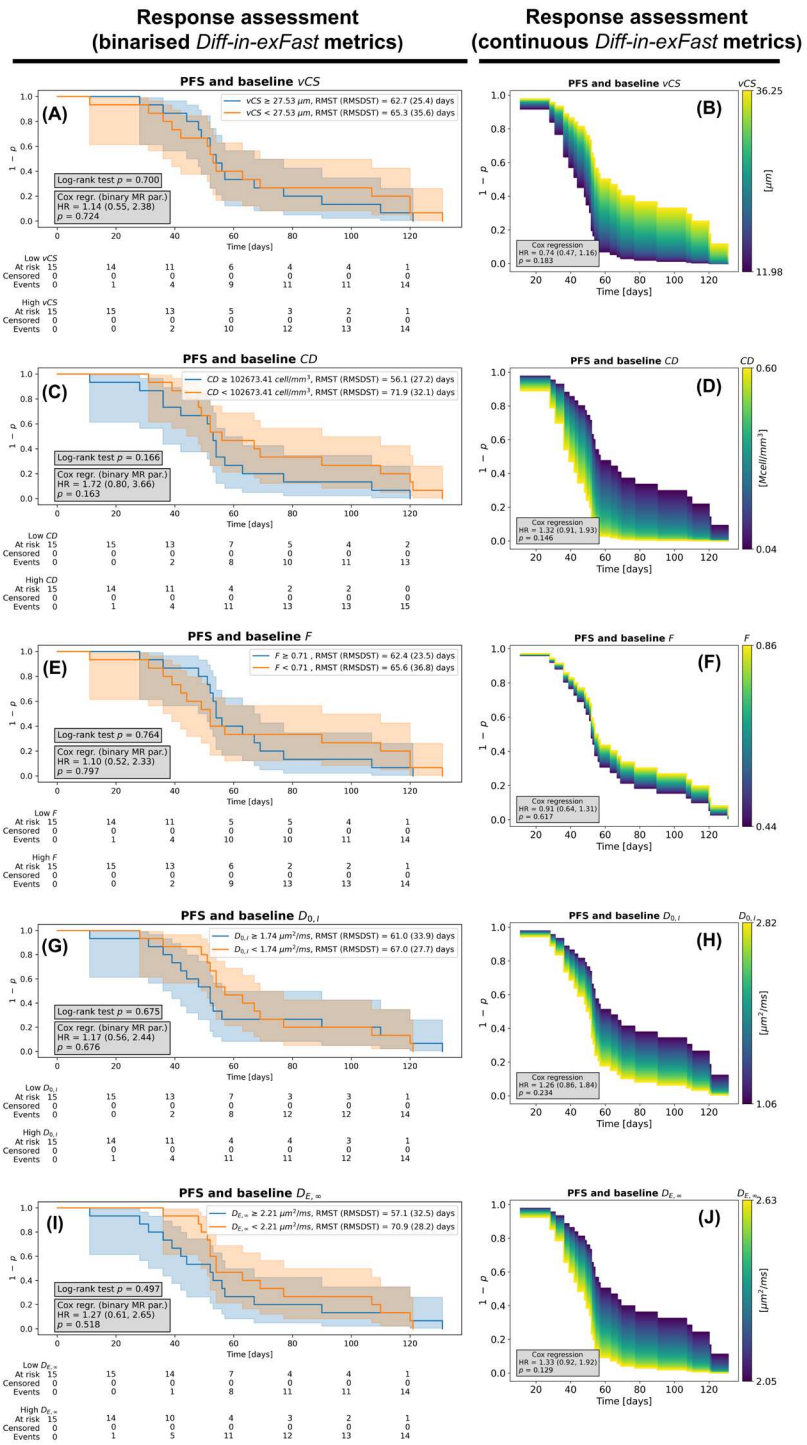

Fig. S16: immunotherapy response assessment based on *Diff-in-exFast* MRI metrics.

This figure reports on the dependence of patients' progression-free survival (PFS) on the average value of all *Diff-in-exFast* metrics within liver tumours at baseline (i.e., before starting immunotherapy). In each row, from top to bottom: PFS based on baseline volume-weighted Cell Size vCS (panels **A** and **B**), Cell Density CD (panels **C** and **D**), intra-cellular fraction F (panels **E**

and **F**), intrinsic intra-cellular cytosol diffusivity $D_{0,I}$ (panels **G** and **H**), asymptotic extra-cellular diffusion coefficient $D_{E,\infty}$ (panels **I** and **J**). The same representation layout as in Fig. S6 was used. Left (**A**, **C**, **E**, **G**, **I**): Kaplan-Meier (KM) analysis, log-rank sum test and Cox regression based on the binarised MRI metrics (higher/lower than the sample median). Right (**B**, **D**, **F**, **H**, **J**): Cox regression modelling the probability of survival as a continuous function of baseline MRI metrics. In all panels, the y-axis shows $1 - p$, with p being the probability of progression, while the x-axis shows the time to progression (in days).

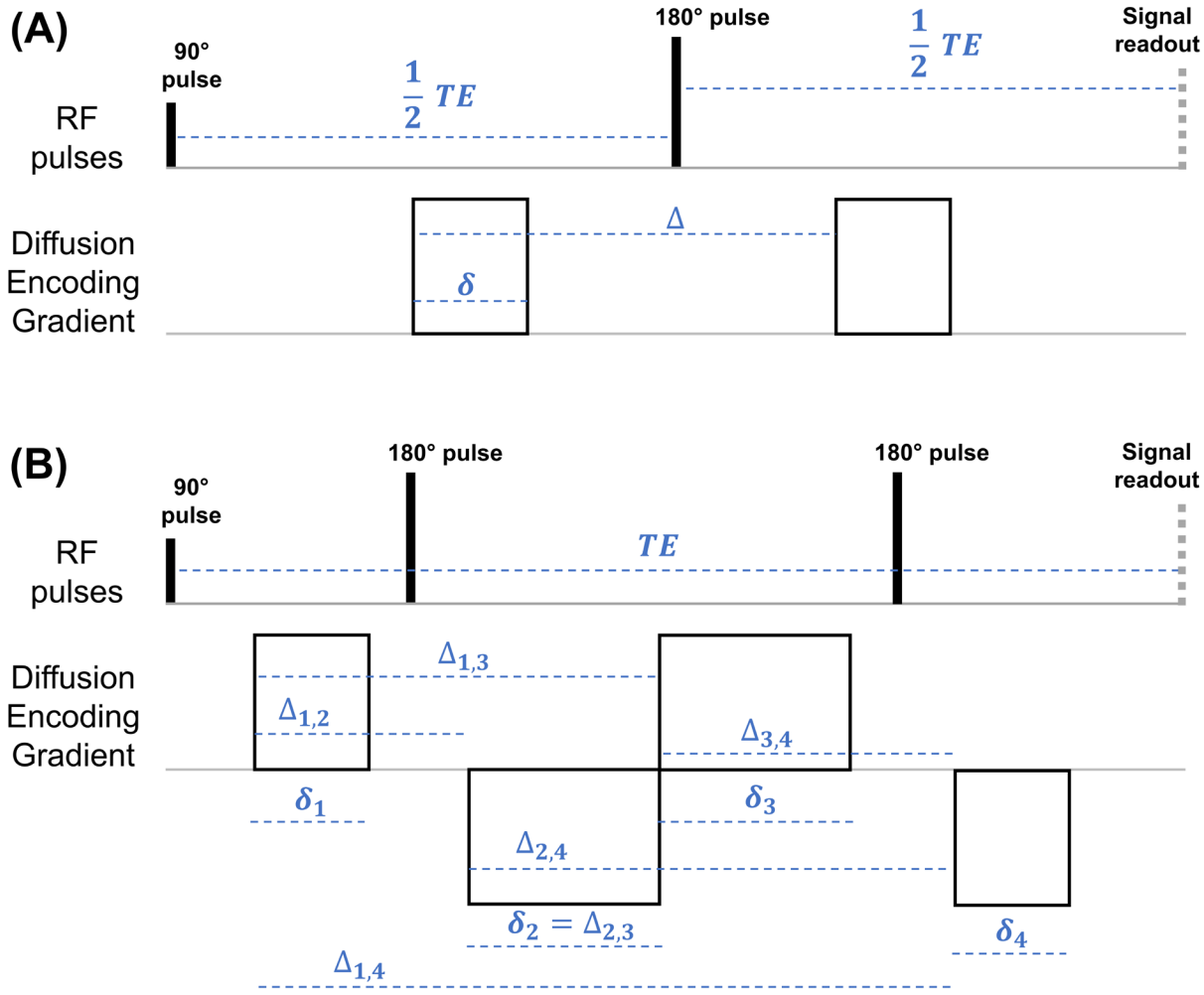

Fig. S17: schematic of the dMRI sequences used in this study.
(**A**): pulsed gradient spin echo (PGSE sequence, also known as Stejskal-Tanner sequence, pulsed-field gradient (PFG), or single linear diffusion encoding) used to acquire data on the 9.4T Bruker system on fixed mouse livers *ex vivo* and on the 3T GE system on patients *in vivo*. δ and Δ respectively indicate the diffusion gradient duration and separation, while TE is the echo time. (**B**): twice-refocussed diffusion-weighted spin echo sequence used to acquire data on the 1.5T Siemens system on patients *in vivo*. δ_n and $\Delta_{n,m}$ respectively indicate the duration of the *n*-th gradient lobe and the separation time between the *n*-th and *m*-th gradient lobes, for *n,m* = 1, …, 4. TE is again the echo time. In both panels, "Signal readout" corresponds to sampling the center of the k-space (i.e., zero spatial frequency).

Table S1: results of the model selection based on the Total Correlation Score (TCS) as obtained on simulated dMRI signals.

We performed model selection on synthetic signals simulated for all the dMRI protocols considered in this study (*ex vivo* PGSE, used on fixed mouse livers; *in vivo* PGSE and DW TRSE, used in patients *in vivo*; see Methods for a full description of the protocols). We fitted the models on protocol subsets obtained with the same b-value thresholds used when analysing actual MRI signals ("Regular fit": fitting on all b-values with negligible vascular contributions; "High *b* only fit": fitting on b-values minimising extra-cellular signal contributions). The Table reports the value of $TCS = r(vCS_{est},vCS_{gt}) + r(F_{est},F_{gt})$, where *vCS* is the cell size, *F* the intra-cellular fraction, $r(x,y)$ the Pearson's correlation between variables *x* and *y* computed pooling together all synthetic voxels, and where subscripts *est* and *gt* respectively indicate estimated and ground truth values. Higher values of *TCS* point towards better model performance. For each protocol and fitting strategy, the model with the highest *TCS* is flagged by gray shadowing and bold font.

	Protocol: *ex vivo*		**Protocol: *in vivo* PGSE**		**Protocol: *in vivo* TRSE**	
	Regular fit	High *b* only fit	Regular fit	High *b* only fit	Regular fit	High *b* only fit
Model						
Diff-in-exTD	0.217	–0.111	0.312	0.202	0.604	0.483
Diff-in-ex	0.406	0.089	0.472	0.335	**0.700**	0.550
Diff-in-exTDFast	0.948	0.827	0.536	0.336	0.618	0.544
Diff-in-exFast	0.952	0.850	0.563	0.349	0.626	0.547
Diff-in	**1.222**	**0.977**	**0.773**	**0.462**	0.543	**0.630**

Table S2: results of the model selection based on the Histology Fidelity Criterion (HFC) as obtained on simulated dMRI signals.

We performed model selection on synthetic signals simulated for all the dMRI protocols considered in this study (*ex vivo* PGSE, used on fixed mouse livers; *in vivo* PGSE and DW TRSE, used in patients *in vivo*; see Methods for a full description of the protocols). We fitted the models on protocol subsets obtained with the same b-value thresholds used when analysing actual MRI signals ("Regular fit": fitting on all b-values with negligible vascular contributions; "High *b* only fit": fitting on b-values minimising extra-cellular signal contributions). For each model, the table reports the percentage of synthetic voxels where $HFC = |vCS_{est} - vCS_{gt}|/vCS_{gt} + |F_{est} - F_{gt}|/F_{gt}$ was the lowest across all models. Above, vCS is the cell size, F the intra-cellular fraction, and subscripts *est* and *gt* respectively indicate estimated and ground truth values. Higher percentages indicate smaller estimation errors, and therefore point towards better model performance. For each protocol and fitting strategy, the model with the highest proportion of synthetic voxels with minimum HFC is flagged by gray shadowing and bold font.

	Protocol: *ex vivo*		**Protocol: *in vivo* PGSE**		**Protocol: *in vivo* TRSE**	
	Regular fit	High *b* only fit	Regular fit	High *b* only fit	Regular fit	High *b* only fit
Model						
Diff-in-exTD	12.62%	13.56%	21.69%	22.10%	26.69%	20.69%
Diff-in-ex	21.19%	16.62%	**25.94%**	20.35%	**28.94%**	24.31%
Diff-in-exTDFast	15.50%	14.94%	10.75%	12.27%	19.69%	14.69%
Diff-in-exFast	21.81%	10.50%	22.25%	11.33%	19.12%	12.12%
Diff-in	**28.88%**	**44.38%**	19.38%	**33.94%**	5.56%	**28.19%**

Table S3: results of the model selection based on the Bayesian Information Criterion (BIC) as obtained on simulated dMRI signals.

We performed model selection on synthetic signals simulated for all the dMRI protocols considered in this study (*ex vivo* PGSE, used on fixed mouse livers; *in vivo* PGSE and DW TRSE, used in patients *in vivo*; see Methods for a full description of the protocols). We fitted the models on protocol subsets obtained with the same b-value thresholds used when analysing actual MRI signals ("Regular fit": fitting on all b-values with negligible vascular contributions; "High *b* only fit": fitting on b-values minimising extra-cellular signal contributions). For each model, the table reports the percentage of synthetic voxels where the Bayesian Information Criterion (BIC, a standard metric of model fitting quality that penalises model complexity) was the lowest across all models. Higher percentages indicate smaller BIC values across synthetic voxels, and therefore point towards better model fitting quality. For each protocol and fitting strategy, the model with the highest proportion of synthetic voxels with minimum HFC is flagged by gray shadowing and bold font.

	Protocol: *ex vivo*		**Protocol: *in vivo* PGSE**		**Protocol: *in vivo* TRSE**	
	Regular fit	High *b* only fit	Regular fit	High *b* only fit	Regular fit	High *b* only fit
Model						
Diff-in-exTD	0.12%	0.00%	0.00%	0.19%	0.00%	0.00%
Diff-in-ex	31.25%	15.88%	17.44%	4.69%	**40.69%**	13.69%
Diff-in-exTDFast	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Diff-in-exFast	3.44%	0.06%	6.69%	0.00%	21.38%	1.06%
Diff-in	**65.19%**	**84.06%**	**75.88%**	**95.12%**	38.00%	**85.25%**

Table S4: descriptive statistics of histology and MRI metrics in the fixed mouse livers.

The table reports mean and standard deviation (within brackets) of histology and dMRI metrics in the 7 fixed mouse livers that were scanned on a 9.4T Bruker system. Histological maps were computed within patches matching the in-plane MRI resolution and then warped non-linearly to dMRI space. The histological maps are: per-patch intra-cellular area fraction F_{histo}, per-patch arithmetic mean cell size aCS_{histo}; per-patch volume-weighted mean cell size vCS_{histo}, cell density per unit patch area CD_{histo}. dMRI metrics are: apparent diffusion coefficient ADC, apparent diffusion kurtosis excess K, intra-cellular signal fraction F_{MRI}, volume-weighted cell size index vCS_{MRI}, apparent cell density per unit volume CD_{MRI}. Metrics F_{MRI}, vCS_{MRI} and CD_{MRI} are reported for both models *Diff-in-exFast* and model *Diff-in*, with *Diff-in* fitted only to high b-value images (b > 1800 s/mm^2). In model *Diff-in-exFast*, the extra-cellular ADC does not feature diffusion time dependence and is constrained to be larger than the intra-cellular ADC. In model *Diff-in*, the extra-cellular signal is modelled as negligible compared to the intra-cellular one (i.e., total signal dominated by intra-cellular water). Specimens are: *Control* (normal liver structures); *NA1* and *NA2* (normal appearing cases, i.e., normal liver structures despite sub-cutaneous biopsy implantation); $Pat_{inf1\text{-}3}$ (cases developing liver pathology following sub-cutaneous biopsy implantation, consisting of small cell infiltration in sinusoidal spaces, in between larger hepatocytes); Pat_{nec} (case developing liver pathology following sub-cutaneous biopsy implantation, consisting of necrosis and inflammation). aCS_{histo}, always considerably lower than vCS_{histo}, was included to highlight the impact of the largest cells in the computation of statistics based on weighting by cell volume (vCS_{histo}).

	Histology				Standard diffusion metrics		*Diff-in-exFast* model			*Diff-in* model at high b		
Specimen	F_{histo}	aCS_{histo} [μm]	vCS_{histo} [μm]	$CD_{histo}/10^2$ [cell/mm^2]	ADC [μm^2/ms]	K	F_{MRI}	vCS_{MRI} [μm]	$CD_{MRI}/10^5$ [cell/mm^3]	F_{MRI}	vCS_{MRI} [μm]	$CD_{MRI}/10^5$ [cell/mm^3]
Control	0.75 (0.18)	21.63 (1.85)	27.1 (1.6)	3.2 (1.1)	1.50 (0.32)	0.33 (0.20)	0.60 (0.18)	33.2 (5.7)	0.38 (1.69)	0.56 (0.11)	36.5 (5.1)	0.28 (1.44)
NA1	0.59 (0.25)	23.97 (3.05)	27.8 (3.2)	2.1 (1.1)	1.61 (0.37)	0.10 (0.15)	0.66 (0.25)	30.5 (8.4)	0.69 (2.18)	0.54 (0.16)	32.2 (7.5)	0.45 (1.68)
NA2	0.76 (0.14)	23.46 (1.41)	29.4 (1.2)	2.7 (0.7)	1.43 (0.44)	0.51 (0.29)	0.80 (0.16)	22.2 (4.8)	1.01 (1.35)	0.71 (0.12)	22.8 (5.0)	0.83 (1.15)
Pat_{inf1}	0.80 (0.20)	15.73 (2.18)	20.8 (2.7)	6.7 (2.2)	0.58 (0.41)	0.98 (0.49)	0.83 (0.13)	13.4 (3.1)	4.20 (2.22)	0.77 (0.14)	12.4 (3.0)	4.86 (2.27)
Pat_{inf2}	0.79 (0.20)	21.41 (2.29)	26.8 (2.7)	3.5 (1.2)	1.67 (0.39)	0.17 (0.17)	0.37 (0.22)	31.2 (7.5)	0.44 (2.06)	0.41 (0.14)	37.1 (6.0)	0.30 (1.57)
Pat_{inf3}	0.70 (0.27)	20.95 (1.75)	27.6 (1.7)	3.1 (1.3)	1.57 (0.62)	0.43 (0.31)	0.63 (0.23)	23.6 (8.8)	1.27 (2.62)	0.59 (0.20)	24.2 (8.0)	0.97 (2.01)
Pat_{nec}	0.52 (0.25)	19.10 (3.33)	25.9 (2.8)	3.4 (2.3)	1.49 (0.44)	0.31 (0.22)	0.61 (0.23)	28.2 (6.6)	0.58 (1.93)	0.54 (0.16)	31.6 (6.7)	0.42 (1.66)

Table S5. Hazard ratios obtained from Cox regression models controlling for sex, age, and baseline tumour volume.

The table reports the Hazard Ratios (HR) for different MRI metrics, with relative 95% confidence interval and p-value, estimated through Cox proportional hazard regressions. The models assessed the dependence of the probability of progression on the baseline mean value of MRI metrics within liver tumours, accounting for sex, age and tumour volume. Results are shown for standard diffusion metrics (apparent diffusion and excess kurtosis coefficients, ADC and K), for the vascular signal fraction f_v, for metrics from model *Diff-in-exFast* (intra-cellular fraction F, volume-weighted cell size vCS, cell density per unit volume CD, intrinsic intra-cellular cytosolic diffusivity $D_{0,I}$, extra-cellular asymptotic diffusion coefficient $D_{E,\infty}$) and for metrics from model *Diff-in* (intra-cellular fraction F, volume-weighted cell size vCS, cell density per unit volume CD, intrinsic intra-cellular cytosolic diffusivity $D_{0,I}$) fitted at high b-value (b > 900 s/mm^2). Grey shadowing highlights HRs whose p-value is ≤ 0.05.

	HR of MRI metric	**HR of male sex**	**HR of age**	**HR of tumour volume**
Standard diffusion MRI metrics				
ADC	1.00 (0.67; 1.49); p = 1.00	0.71 (0.32; 1.55); p = 0.39	0.97 (0.94; 0.99); p = 0.02	0.32 (0.06; 1.70); p = 0.18
K	1.36 (0.94; 1.97); p = 0.11	0.64 (0.29; 1.41); p = 0.27	0.97 (0.94; 1.00); p = 0.02	0.35 (0.08; 1.54); p = 0.17
Other metrics (from vascular vs non-vascular fitting initialisation step)				
f_v	1.16 (0.79; 1.71); p = 0.44	0.77 (0.35; 1.68); p = 0.51	0.96 (0.93; 0.99); p = 0.01	0.34 (0.17; 2.56); p = 0.17
Metrics from MRI model *Diff-in-exFast*				
F	0.87 (0.61; 1.26); p = 0.46	0.78 (0.35; 1.72); p = 0.54	0.97 (0.94; 1.00); p = 0.02	0.31 (0.06; 1.58); p = 0.16
vCS	0.70 (0.44; 1.12); p = 0.14	0.78 (0.36; 1.68); p = 0.53	0.97 (0.94; 0.99); p = 0.02	0.35 (0.08; 1.46); p = 0.15
CD	1.53 (1.00; 2.34); p = 0.05	0.63 (0.29; 1.38); p = 0.25	0.96 (0.93; 0.99); p = 0.01	0.32 (0.06; 1.81); p = 0.20
$D_{0,I}$	1.45 (0.98; 2.14); p = 0.06	0.68 (0.31; 1.47); p = 0.33	0.96 (0.93; 0.99); p = 0.01	0.36 (0.09; 1.35); p = 0.13
$D_{E,\infty}$	1.41 (0.97; 2.05); p = 0.07	0.69 (0.32; 1.48); p = 0.34	0.96 (0.94; 0.99); p = 0.01	0.40 (0.13; 1.26); p = 0.12
Metrics from MRI model *Diff-in* fitted to high *b*-value images				
F	0.79 (0.54; 1.17); p = 0.24	0.80 (0.36; 1.74); p = 0.57	0.97 (0.94; 0.99); p = 0.02	0.33 (0.07; 1.49); p = 0.15
vCS	0.59 (0.37; 0.93); p = 0.02	0.66 (0.30; 1.43); p = 0.29	0.96 (0.94; 0.99); p = 0.01	0.40 (0.12; 1.37); p = 0.14
CD	1.65 (1.12; 2.44); p = 0.01	0.61 (0.28; 1.35); p = 0.22	0.96 (0.93; 0.99); p = 0.01	0.36 (0.09; 1.52); p = 0.17
$D_{0,I}$	1.28 (0.82; 1.98); p = 0.28	0.67 (0.31; 1.45); p = 0.31	0.96 (0.94; 0.99); p = 0.01	0.37 (0.09; 1.48); p = 0.16