Supplementary Method S1

Diffusion weighted image (DWI) analysis.

b-value selection

The signal in DW MR images reflects movement of water molecules. In tissues, there are two main sources of water movement, diffusion caused by Brownian motion of water molecules, and convection caused by fluid flow in blood vessels, *i.e.* perfusion. According to the IVIM model introduced by Bihan and colleagues (1), diffusion and convection can be separated due to relatively fast signal loss by perfusion compared to diffusion.

$$\frac{S(b)}{S_0} = fBV \cdot e^{-b \cdot D^*} + (1 - fBV) \cdot e^{-b \cdot ADC} \quad \text{(I)}$$

The assumption, $D^* >> ADC$, is fulfilled for *b*-values where the signal loss is linear in a logarithmic plot, yielding a simplified IVIM model:

$$\ln\left(\frac{S(b)}{S_0}\right) = -b \cdot ADC + \ln(1 - fBV) \quad \text{(II)}$$

Hence, the signal from the low *b*-values is a result of both diffusion and perfusion, leading to a non-linear signal loss in a logarithmic plot, while the signal from higher *b*-values is mainly a result of diffusion, yielding a linear signal loss in a logarithmic plot (Fig. 1A).

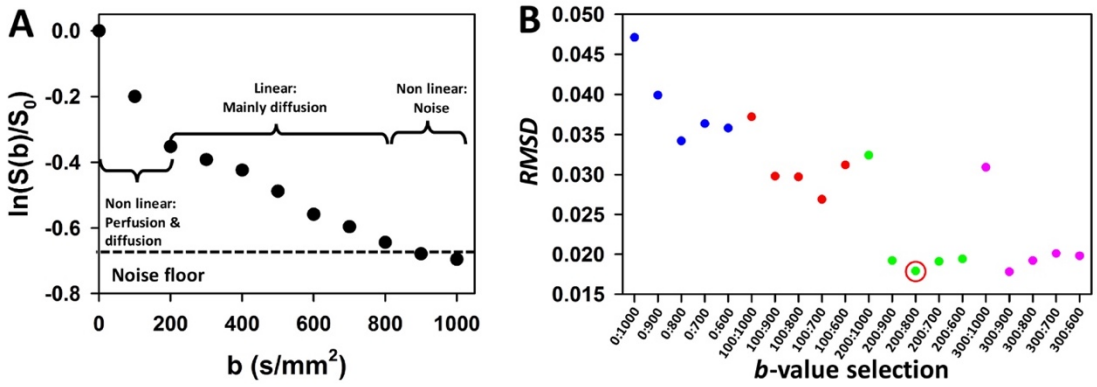

Figure 1. Considerations for selection of b-values. A, Typical plot of relative signal, $\ln(S(b)/S_0)$, *versus* *b*-value for a tumor pixel. Areas of non-linearity due to perfusion and noise, and area of linearity due to diffusion alone are indicated. The horizontal dashed line represents the noise floor. B, Overall mean model fit error, RMSD, *versus* *b*-value combination used in the linear model (equation II). The mean RMSD of each tumor was averaged for 50 cohort 1 patients in our study. The chosen *b*-value selection is outlined by a red circle.

When using the simplified IVIM model to quantify ADC and fBV, it is vital to avoid *b*-values affected by perfusion. Inclusion of too low *b*-values would overestimate *ADC* and underestimate *fBV*. For high *b*-values, non-linearity may exist because the signal reaches the measuring limit of the scanner, *i.e.* the noise floor (Fig. 1A), leading to an underestimation of *ADC* and overestimation of *fBV*. *b*-values above 200 have been suggested as the optimal choice (2). However, this may be tissue type dependent, and must be investigated for each type.

To find the *b*-values that led to a linear signal loss in logarithmic plots of our data, and thereby enabled use of the simplified IVIM model, we visually inspected plots of a large number of pixels. This indicated that a *b*-value selection from 200 to 800 s/mm^2 was optimal. Furthermore, we performed linear fits of equation (II) to plots with a range of *b*-value combinations and evaluated the root mean squared deviation, RMSD, as a measure of fit error:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad \text{(III)}$$

where $\hat{y}$ is the predicted value from the fit, y is the measured value, and n is the total number of values. The lowest overall mean RMSD was found for a *b*-value selection of 200-800 s/mm^2, in line with our impression from the visual inspection (Fig. 1B). From Figure 1B it is evident that including *b*-values of 0 and 100 s/mm^2 increased RMSD considerably. Similarly, including *b*-values of 1000 and 900 s/mm^2 also increase RMSD. *b*-values below 200 and above 800 s/mm^2 were therefore excluded, yielding a final *b*-value selection of 200-800 s/mm^2.

Comparison of data from the IVIM model and simplified IVIM model

The considerations above were performed by using the simplified linear IVIM model in equation (II). To ensure that this model provided reliable *ADC* and *fBV* measures, the data were compared with those obtained with the bi-exponential IVIM model. For the simplified IVIM model, we used the *b*-value selection of 200-800 s/mm^2, while *b*-values of 900 and 1000 s/mm^2 were excluded in the IVIM model to avoid noisy measurements. We found a strong correlation between the results obtained with the two models (Fig. 2), indicating that they provided similar information. However, the IVIM model contains an additional parameter, D*, which is calculated from the first, non-linear, part of the

curve. In our study, where only the signal at *b*-values of 0 and 100 s/mm^2 was evidently influenced by perfusion, it would be difficult to obtain reliable estimates of this parameter due to this sparse number of low *b*-values. The use of the IVIM model could then lead to uncertainties also in the calculation of *ADC* and *fBV*. Application of the simplified IVIM model in our study was therefore justified.

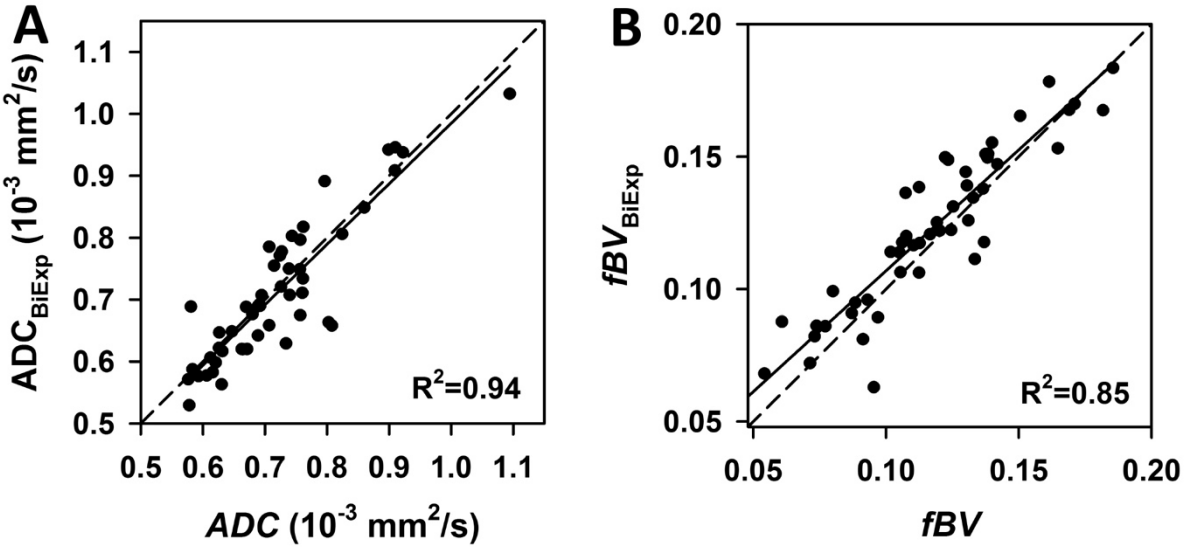

Figure 2. Comparison of data achieved with the bi-exponential IVIM model, ADC_{BiExp} (A) and fBV_{Biexp} (B), and the simplified, linear, IVIM model, *ADC* (A) and *fBV* (B) for 50 cohort 1 patients in our study. Line of unity is shown as a dashed line and the linear regression line is shown as a solid line. Pearson correlation coefficient (R^2) is indicated.

Impact of model fit accuracy on DWI parameters

Since the accuracy of fit can influence parameter estimation, we first performed a tumor based evaluation by investigating if the model fit error, RMSD, showed any relationship to the median DWI parameters of cohort 1 patients.

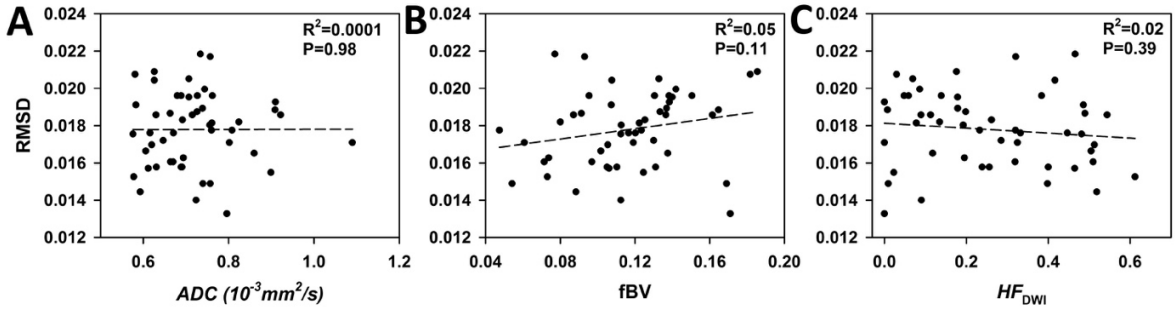

Figure 3. Scatter plot of *ADC* (A), *fBV* (B) and HF_{DWI} (C) *versus* RMSD for 50 cohort 1 patients in our study. Each point represents mean RMSD and median DWI parameter of each tumor. Linear regression line and correlation coefficient (R^2) and *P*-value from Pearson correlation analysis are indicated.

We further evaluated the fit accuracy at the pixel level by the standard deviation of the regression line, to achieve a better understanding of the parameter error. Figure 4A presents a pixel-wise plot of *fBV versus ADC* for a more hypoxic and less hypoxic tumor with standard deviation of the fit visualized for randomly selected points. Such plots showed that the standard deviation of individual pixels was low compared to the parameter value and not related to the discrimination of hypoxic and non-hypoxic pixels. Furthermore, the pixel-wise standard deviation of all patients pooled together was relatively stable for all *ADC* and *fBV* values, in particular for values within the 5 and 95 percentile (Fig. 4B, C). All together, these analyses showed that the DWI parameters estimated in our study were not dependent on the model fit accuracy, neither at the tumor nor pixel level.

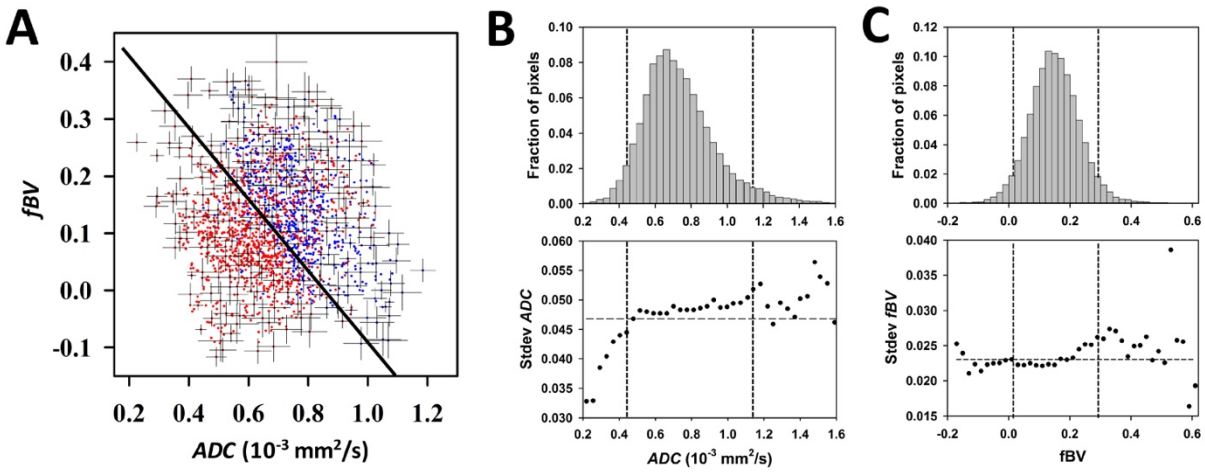

Figure 4. Pixel-wise plot of *fBV versus ADC* from a more hypoxic (red dots) and less hypoxic (blue dots) tumor (A) with the standard deviation of the regression line shown by black lines. The tumors correspond to those presented in Figures 3 and 4 in the paper. For visualization purposes, standard deviation is indicated only for some randomly selected points. The separation line between hypoxic and non-hypoxic pixels are shown. Pixel-wise histogram (upper) and average standard deviation of the corresponding percentiles (lower) for *ADC* (B) and *fBV* (C) of 50 cohort 1 patients in our study pooled together. The 5 and 95 percentiles are indicated by vertical dashed lines. The horizontal dashed line in the lower panels represents the average standard deviation of all pixels.

Validation of *ADC* and *fBV* independence

The use of suboptimal *b*-values may lead to cross-correlations between *ADC* and *fBV* that are not biologically founded. To ensure that our data were not confounded by such cross-correlations, we investigated whether the *ADC* and *fBV* values were independent of each other. First, we tested all correlations between *ADC* percentiles 1-99 and *fBV* percentiles 1-99 in cohort 1 patients. The analysis was performed for all patients

combined, and for patients with less and more hypoxia separately. Overall, we found very weak correlations that reached significance only at extreme percentiles (Fig. 5). At the 50% percentile (median) used in our study, the correlations were R=-0.087, R=-0.44, R=-0.18 for all patients and for patients with less and more hypoxia, respectively. These results show that *ADC* and *fBV* were independent across patients.

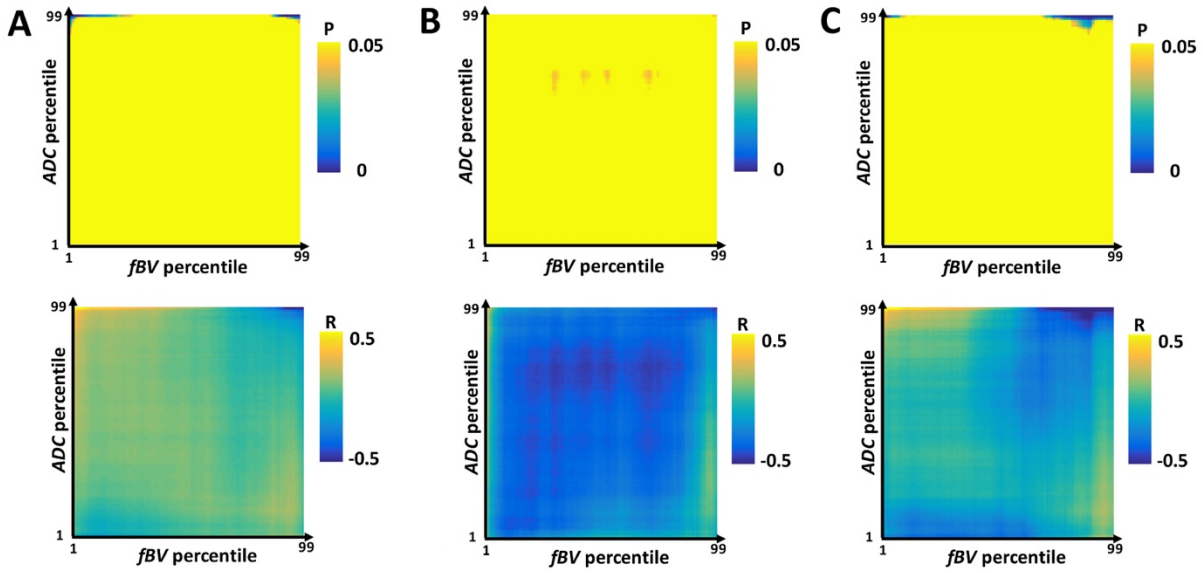

Figure 5. Correlation analysis of *ADC versus fBV* based on tumor percentile 1-99. The percentiles were calculated from the parameter histogram of all pixels in each tumor. Pearson correlation matrices with *P*-values (upper panels) and *R*-values (lower panels) for all 50 cohort 1 patients (A), 21 patients with less hypoxia (HS_{Pimo} ≤2) (B) and 20 patients with more hypoxia (HS_{Pimo}>2) (C). B, C, Cohort 1 patients with paired DWI and pimonidazole data (n=41) were included.

Although there were almost no correlations between *ADC* and *fBV* percentiles across patients, significant correlations occurred within individual tumors in some cases, as exemplified by patient 1 in Figure 3 in the paper. We therefore further investigated if the correlation coefficient, *R*, obtained in the pixel-wise analyses of *ADC versus fBV* within tumors was related to the median *ADC*, *fBV* or HF_{DWI} values across patients. Both positive and negative correlations between *ADC* and *fBV* were found within tumors, but there was no significant relationship between *R* and any of the DWI parameters (Fig. 6). This and the above analysis (Fig. 5) strongly indicate that the correlation between *ADC* and *fBV* observed for some tumors was not due to cross-correlation caused by our model, but rather a result of biological variation within the tumors.

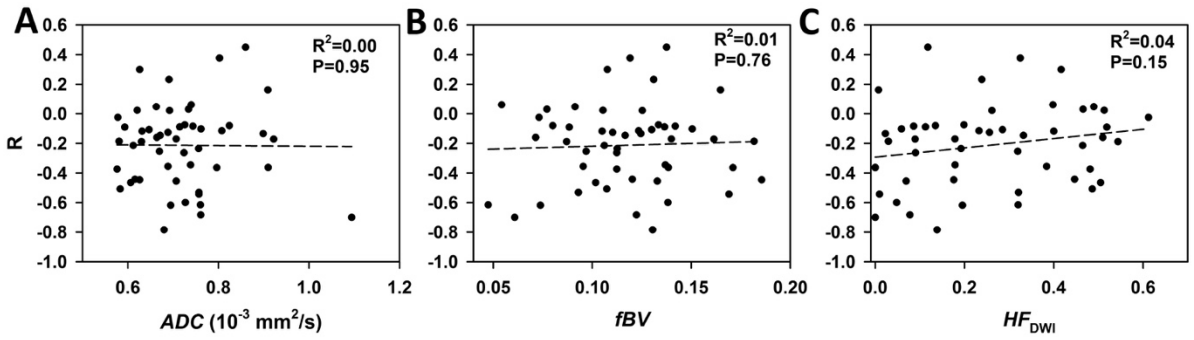

Figure 6. Scatter plot of *ADC* (A), *fBV* (B) and HF_{DWI} (C) *versus* Pearson correlation coefficient (*R*). Each point represents mean *R* and median DWI parameter of each tumor. *R* was obtained from pixel-wise analysis of *ADC versus fBV* of individual tumors for 50 cohort 1 patients in our study. Linear regression line and correlation coefficient (R^2) and *P*-value from Pearson correlation analysis are indicated.

References

1. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988;168:497–505.

2. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia. 2009;11:102–25.