**Abbreviations**

Multi-parametric Magnetic resonance imaging (MP-MRI)

Diffusion weighted imaging (DWI)

Apparent Diffusion Coefficient (ADC)

Digital reference object (DRO)

Monoexponential apparent diffusion coefficient (MEADC)

Kurtosis (K)

Diffusion kurtosis (DK)

Bi-exponential diffusion (BID)

Pseudo-diffusion (BID\*)

Perfusion fraction (F)

Receiver operating characteristic area under the curve (ROC AUC)

Gleason 3, 4 (G3, G4)

**Supplemental Table 1.** Percent difference data comparing site submission datasets (Manuscript Figure 3, Top).

**Percent Difference**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Cancer** | Mean | STD | | | Min | | Max |
| BID | 38.25 | | 24.56 | 3.76 | | 105.41 | | |
| BID\* | 107.17 | | 55.96 | 10.70 | | 196.49 | | |
| F | 79.59 | | 44.50 | 3.36 | | 190.21 | | |
| DK | 5.54 | | 4.50 | 0.02 | | 17.97 | | |
| K | 3.93 | | 2.30 | 0.00 | | 7.54 | | |
| MEADC | 18.66 | | 13.27 | 0.00 | | 52.80 | | |
|  |  | |  |  | |  | | |
| **Benign** | Mean | STD | | | Min | | Max |
| BID | 42.67 | | 26.64 | 4.69 | | 113.57 | | |
| BID\* | 106.50 | | 54.16 | 17.29 | | 196.50 | | |
| F | 93.04 | | 56.17 | 10.36 | | 195.17 | | |
| DK | 3.86 | | 3.31 | 0.02 | | 12.93 | | |
| K | 3.43 | | 2.01 | 0.01 | | 6.53 | | |
| MEADC | 19.04 | | 14.05 | 0.00 | | 52.17 | | |

**Supplemental Table 2.** Pearson correlation coefficient data comparing site submission datasets (Manuscript Figure 3, Bottom). Site submission count: BID, BID\* and MEADC, n=11, F, n=12, K and DK, n=9.

**Correlation Coefficient**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Cancer** | Mean | STD | | | Min | | Max |
| BID | 0.48 | | 0.22 | -0.06 | | 0.98 | | |
| BID\* | 0.20 | | 0.29 | -0.41 | | 0.94 | | |
| F | 0.14 | | 0.27 | -0.57 | | 0.77 | | |
| DK | 0.91 | | 0.05 | 0.80 | | 1.00 | | |
| K | 0.72 | | 0.28 | 0.06 | | 1.00 | | |
| MEADC | 0.93 | | 0.06 | 0.75 | | 1.00 | | |
|  |  | |  |  | |  | | |
| **Benign** | Mean | STD | | | Min | | Max |
| BID | 0.37 | | 0.26 | -0.24 | | 0.97 | | |
| BID\* | 0.25 | | 0.25 | -0.27 | | 0.88 | | |
| F | 0.10 | | 0.27 | -0.59 | | 0.75 | | |
| DK | 0.84 | | 0.07 | 0.70 | | 0.98 | | |
| K | 0.70 | | 0.22 | 0.14 | | 0.99 | | |
| MEADC | 0.84 | | 0.14 | 0.53 | | 1.00 | | |

**Supplemental Table 3.** Receiver operating characteristic area under the curve by image contrast (Manuscript Figure 4).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cancer vs Benign Atrophy** | Median | Min | Max |  | **Grade 3 vs Grade 4+** | Median | Min | Max |
| MEADC | 0.78 | 0.76 | 0.80 |  | MEADC | 0.67 | 0.66 | 0.68 |
| DK | 0.78 | 0.76 | 0.81 |  | DK | 0.67 | 0.65 | 0.70 |
| K | 0.75 | 0.72 | 0.76 |  | K | 0.64 | 0.63 | 0.65 |
| BID | 0.71 | 0.53 | 0.80 |  | BID | 0.60 | 0.52 | 0.68 |
| BID\* | 0.56 | 0.51 | 0.81 |  | BID\* | 0.54 | 0.50 | 0.69 |
| F | 0.61 | 0.52 | 0.80 |  | F | 0.59 | 0.51 | 0.69 |

**Supplemental Table 4.** Receiver operating characteristic area under the curve varied by prostate peripheral (PZ) and transition zone (TZ) and by image contrast (Supplemental Figure 3).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PZ** | Median | Min | Max |  | **TZ** | Median | Min | Max |
| MEADC | 0.81 | 0.74 | 0.82 |  | MEADC | 0.84 | 0.73 | 0.85 |
| DK | 0.77 | 0.72 | 0.80 |  | DK | 0.74 | 0.68 | 0.81 |
| K | 0.77 | 0.76 | 0.79 |  | K | 0.86 | 0.84 | 0.87 |
| BID | 0.73 | 0.51 | 0.82 |  | BID | 0.72 | 0.51 | 0.84 |
| BID\* | 0.58 | 0.52 | 0.76 |  | BID\* | 0.60 | 0.51 | 0.77 |
| F | 0.63 | 0.54 | 0.76 |  | F | 0.62 | 0.50 | 0.80 |

**Supplemental Table 5.** Receiver operating characteristic area under the curve considering only the index lesion on each slide compared to the largest atrophic region (Supplemental Figure 4).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PZ** | Median | Min | Max |  |
| MEADC | 0.82 | 0.75 | 0.83 |  |
| DK | 0.75 | 0.57 | 0.82 |  |
| K | 0.58 | 0.51 | 0.77 |  |
| BID | 0.62 | 0.52 | 0.76 |  |
| BID\* | 0.79 | 0.78 | 0.80 |  |
| F | 0.77 | 0.72 | 0.81 |  |

**Supplemental Table 6.** Receiver operating characteristic area under the curve for all institutions grouped by fit, repeated varying the maximum lesion size.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cancer vs Benign Atrophy** | | | | |  |  | **High Grade vs Low Grade** | | | |
|  | Cluster  Limit | Median | Min | Max |  |  | Cluster  Limit | Median | Min | Max |
| MEADC | 100 | 0.74 | 0.67 | 0.75 |  | MEADC | 100 | 0.61 | 0.60 | 0.62 |
| 200 | 0.80 | 0.74 | 0.81 |  | 200 | 0.67 | 0.66 | 0.68 |
| 300 | 0.81 | 0.74 | 0.82 |  | 300 | 0.67 | 0.67 | 0.68 |
| 400 | 0.84 | 0.75 | 0.84 |  | 400 | 0.72 | 0.71 | 0.73 |
| 500 | 0.87 | 0.77 | 0.87 |  | 500 | 0.72 | 0.71 | 0.74 |
| DK | 100 | 0.69 | 0.65 | 0.73 |  | DK | 100 | 0.61 | 0.60 | 0.63 |
| 200 | 0.76 | 0.71 | 0.79 |  | 200 | 0.67 | 0.65 | 0.70 |
| 300 | 0.76 | 0.71 | 0.80 |  | 300 | 0.68 | 0.65 | 0.70 |
| 400 | 0.78 | 0.72 | 0.82 |  | 400 | 0.71 | 0.69 | 0.72 |
| 500 | 0.80 | 0.74 | 0.85 |  | 500 | 0.73 | 0.70 | 0.75 |
| K | 100 | 0.71 | 0.63 | 0.72 |  | K | 100 | 0.58 | 0.57 | 0.59 |
| 200 | 0.77 | 0.76 | 0.78 |  | 200 | 0.64 | 0.63 | 0.65 |
| 300 | 0.78 | 0.77 | 0.81 |  | 300 | 0.65 | 0.64 | 0.67 |
| 400 | 0.81 | 0.80 | 0.83 |  | 400 | 0.71 | 0.70 | 0.72 |
| 500 | 0.85 | 0.83 | 0.85 |  | 500 | 0.71 | 0.69 | 0.72 |
| BID | 100 | 0.68 | 0.53 | 0.75 |  | BID | 100 | 0.58 | 0.51 | 0.62 |
| 200 | 0.72 | 0.53 | 0.81 |  | 200 | 0.60 | 0.52 | 0.68 |
| 300 | 0.73 | 0.53 | 0.82 |  | 300 | 0.61 | 0.51 | 0.70 |
| 400 | 0.76 | 0.52 | 0.84 |  | 400 | 0.65 | 0.53 | 0.73 |
| 500 | 0.77 | 0.52 | 0.87 |  | 500 | 0.66 | 0.51 | 0.73 |
| BID\* | 100 | 0.51 | 0.50 | 0.67 |  | BID\* | 100 | 0.53 | 0.51 | 0.62 |
| 200 | 0.56 | 0.52 | 0.73 |  | 200 | 0.54 | 0.50 | 0.69 |
| 300 | 0.58 | 0.53 | 0.75 |  | 300 | 0.56 | 0.50 | 0.71 |
| 400 | 0.59 | 0.54 | 0.77 |  | 400 | 0.57 | 0.50 | 0.71 |
| 500 | 0.60 | 0.53 | 0.80 |  | 500 | 0.58 | 0.50 | 0.74 |

**Supplemental Table 7.** Receiver operating characteristic area under the curve for cancer vs benign annotations varying the method used to calculate values within each region (median, mean, and 10th percentile).

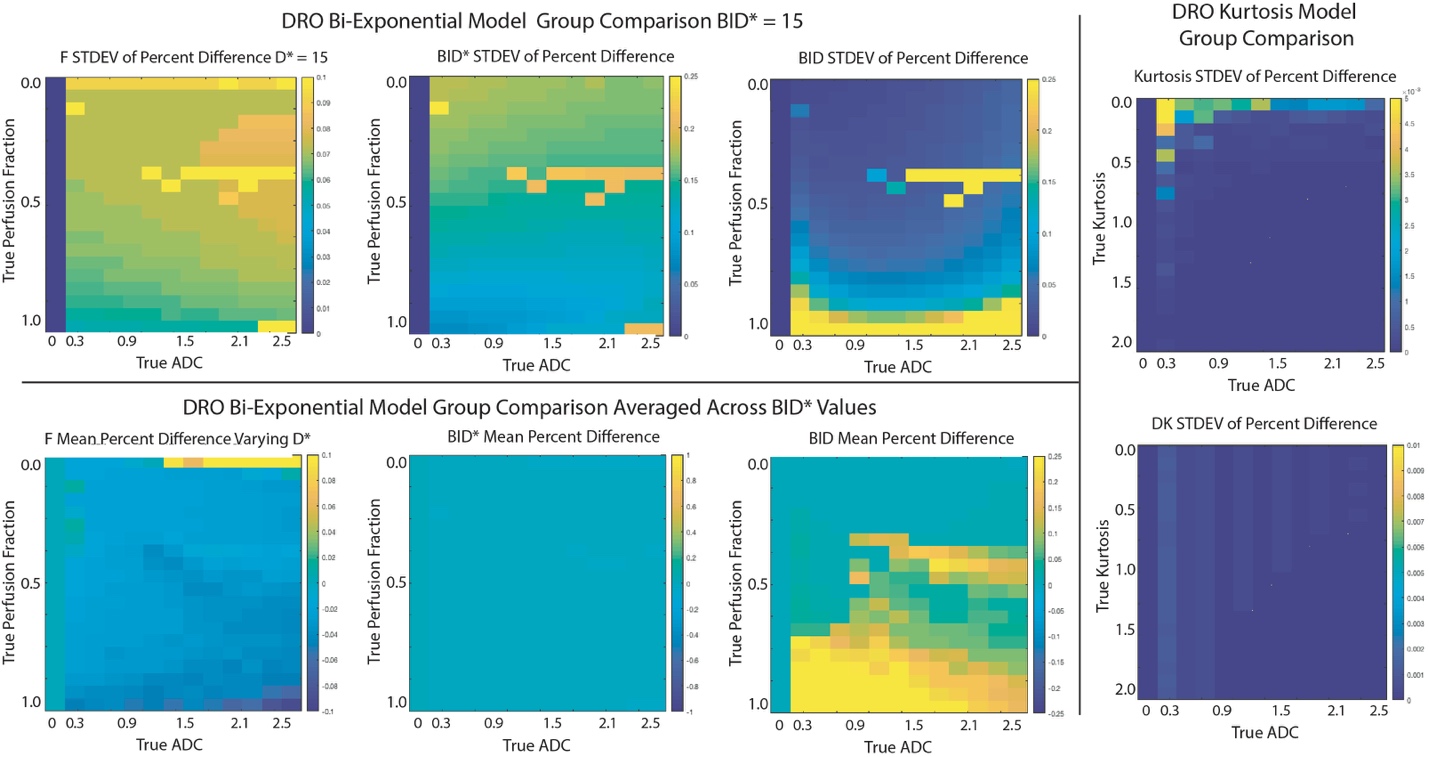
|  |  |  |  |
| --- | --- | --- | --- |
| **BAvC Median** | Median | Min | Max |
| MEADC | 0.80 | 0.74 | 0.81 |
| DK | 0.76 | 0.71 | 0.79 |
| K | 0.77 | 0.76 | 0.78 |
| BID | 0.72 | 0.53 | 0.81 |
| BIDS | 0.56 | 0.52 | 0.73 |
| F | 0.62 | 0.51 | 0.76 |
|  |  |  |  |
| **BAvC Mean** | Median | Min | Max |
| MEADC | 0.80 | 0.71 | 0.81 |
| DK | 0.74 | 0.69 | 0.79 |
| K | 0.77 | 0.65 | 0.78 |
| BID | 0.71 | 0.52 | 0.80 |
| BIDS | 0.53 | 0.50 | 0.73 |
| F | 0.62 | 0.50 | 0.75 |
|  |  |  |  |
| **BAvC 10th** | Median | Min | Max |
| MEADC | 0.80 | 0.77 | 0.81 |
| DK | 0.75 | 0.73 | 0.78 |
| K | 0.80 | 0.79 | 0.81 |
| BID | 0.69 | 0.54 | 0.81 |
| BIDS | 0.55 | 0.51 | 0.71 |
| F | 0.59 | 0.52 | 0.75 |
|  |  |  |  |
| **LGvHG Median** | Median | Min | Max |
| MEADC | 0.67 | 0.66 | 0.68 |
| DK | 0.67 | 0.65 | 0.70 |
| K | 0.63 | 0.63 | 0.65 |
| BID | 0.60 | 0.52 | 0.68 |
| BIDS | 0.54 | 0.50 | 0.69 |
| F | 0.59 | 0.51 | 0.69 |
|  |  |  |  |
| **LGvHG Mean** | Median | Min | Max |
| MEADC | 0.67 | 0.66 | 0.67 |
| DK | 0.66 | 0.65 | 0.69 |
| K | 0.63 | 0.62 | 0.65 |
| BID | 0.60 | 0.55 | 0.67 |
| BIDS | 0.55 | 0.50 | 0.68 |
| F | 0.59 | 0.50 | 0.70 |
|  |  |  |  |
| **LGvHG 10th** | Median | Min | Max |
| MEADC | 0.68 | 0.67 | 0.68 |
| DK | 0.67 | 0.65 | 0.70 |
| K | 0.65 | 0.65 | 0.67 |
| BID | 0.58 | 0.51 | 0.69 |
| BIDS | 0.53 | 0.50 | 0.63 |
| F | 0.57 | 0.52 | 0.71 |

**Supplemental Table 8.** Receiver operating characteristic area under the curve for cancer vs benign annotations varying the pathologist annotating the slides.

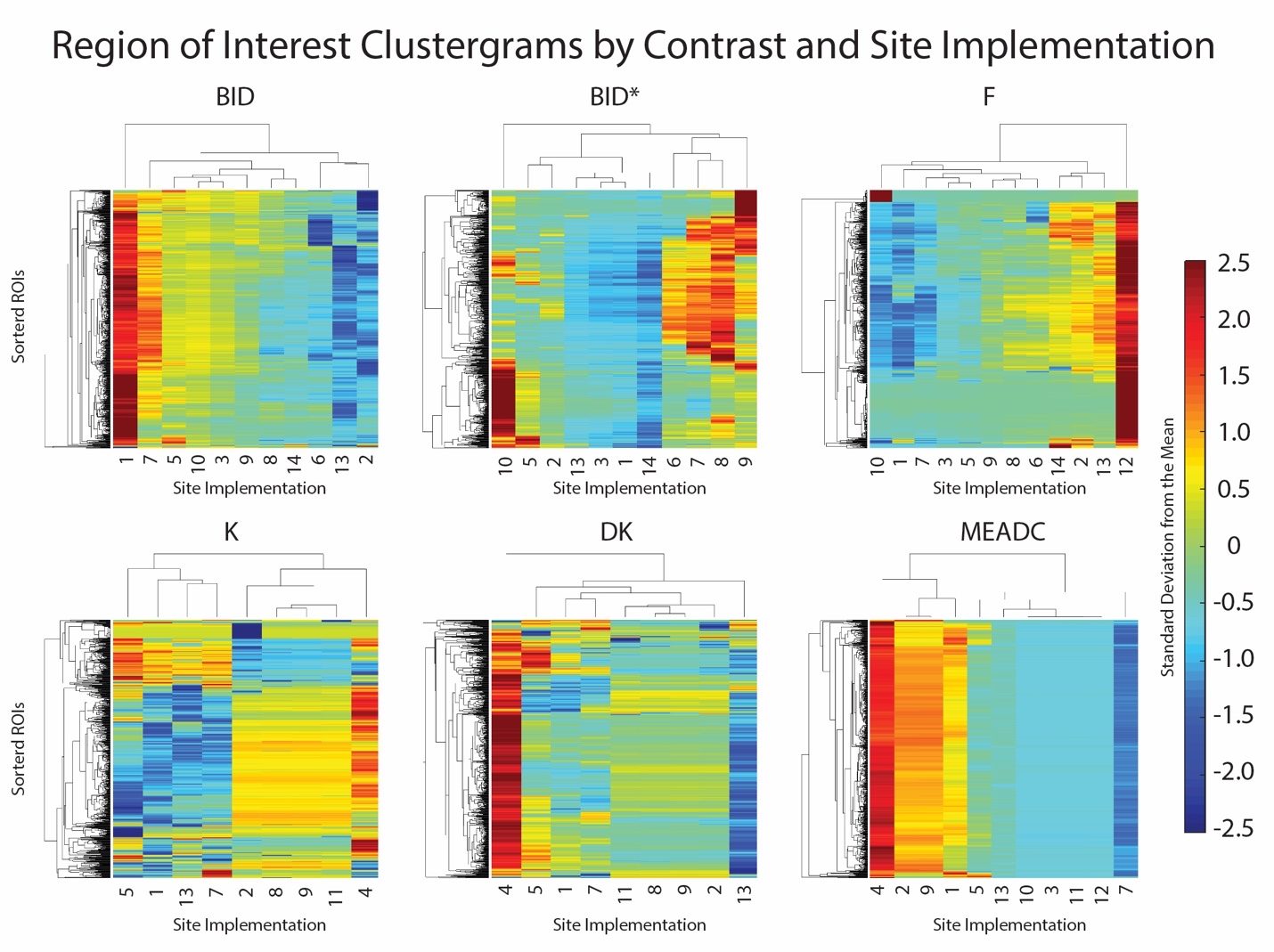
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cancer vs Benign Atrophy** | | | | |
|  | Observer | Median | Min | Max |
| MEADC | 1 | 0.83 | 0.79 | 0.88 |
| 2 | 0.78 | 0.74 | 0.83 |
| 3 | 0.79 | 0.75 | 0.84 |
| 4 | 0.81 | 0.78 | 0.85 |
| 5 | 0.87 | 0.86 | 0.88 |
| DK | 1 | 0.82 | 0.77 | 0.86 |
| 2 | 0.78 | 0.72 | 0.81 |
| 3 | 0.81 | 0.74 | 0.83 |
| 4 | 0.82 | 0.76 | 0.84 |
| 5 | 0.87 | 0.83 | 0.89 |
| K | 1 | 0.82 | 0.78 | 0.83 |
| 2 | 0.77 | 0.75 | 0.83 |
| 3 | 0.79 | 0.76 | 0.87 |
| 4 | 0.80 | 0.79 | 0.86 |
| 5 | 0.88 | 0.87 | 0.91 |
| BID | 1 | 0.77 | 0.57 | 0.83 |
| 2 | 0.76 | 0.62 | 0.90 |
| 3 | 0.77 | 0.63 | 0.87 |
| 4 | 0.79 | 0.57 | 0.88 |
| 5 | 0.83 | 0.68 | 0.88 |
| BID\* | 1 | 0.64 | 0.52 | 0.83 |
| 2 | 0.67 | 0.50 | 0.79 |
| 3 | 0.69 | 0.52 | 0.82 |
| 4 | 0.67 | 0.54 | 0.83 |
| 5 | 0.69 | 0.52 | 0.87 |
| F | 1 | 0.63 | 0.51 | 0.77 |
| 2 | 0.59 | 0.51 | 0.68 |
| 3 | 0.61 | 0.51 | 0.83 |
| 4 | 0.61 | 0.50 | 0.74 |
| 5 | 0.65 | 0.51 | 0.82 |

**Supplemental Figures**

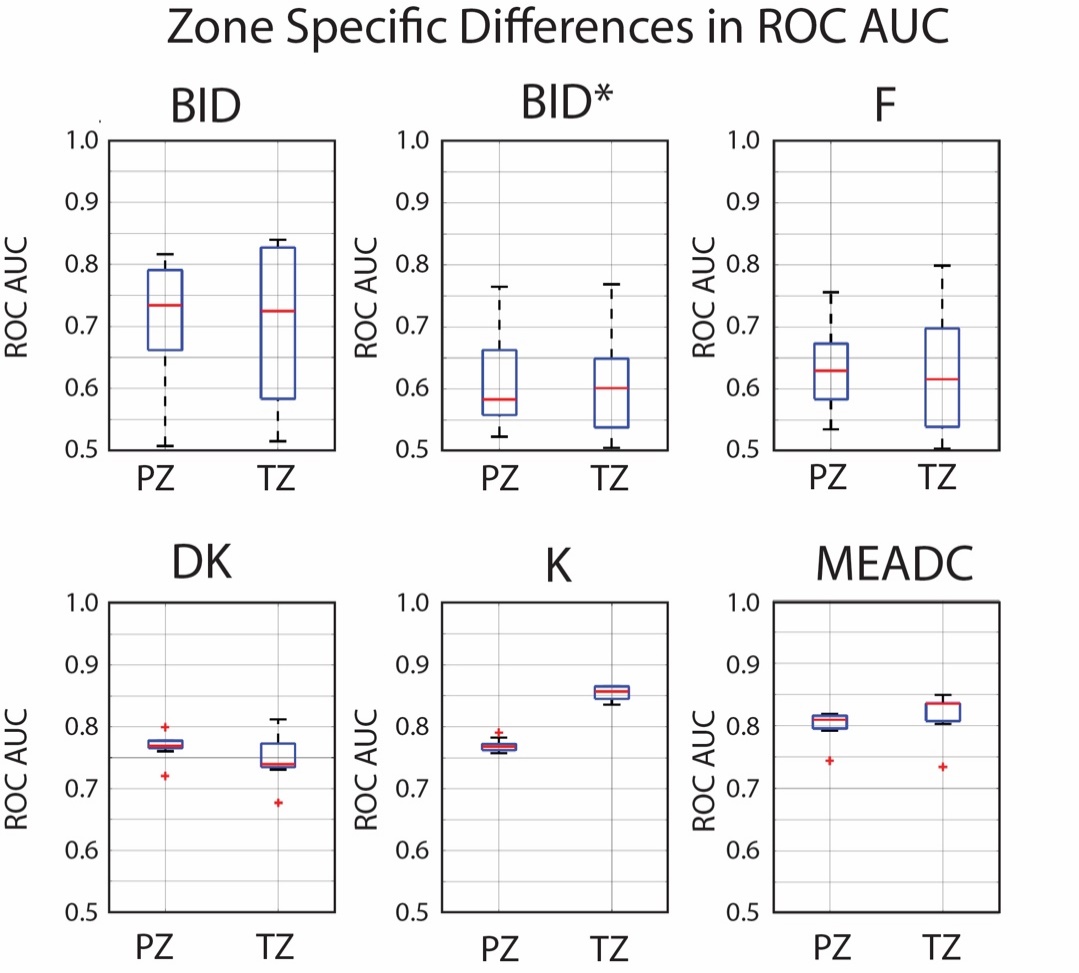
**Supplemental Figure 1.** Results from the DRO analysis. F, BID\*, and BID are shown on the left while K and DK are shown on the far right. Heat maps indicate the standard deviation of the percent difference across sites varying true perfusion, true ADC, and BID\* for the bi-exponential model, and true kurtosis and true ADC for the kurtosis model.



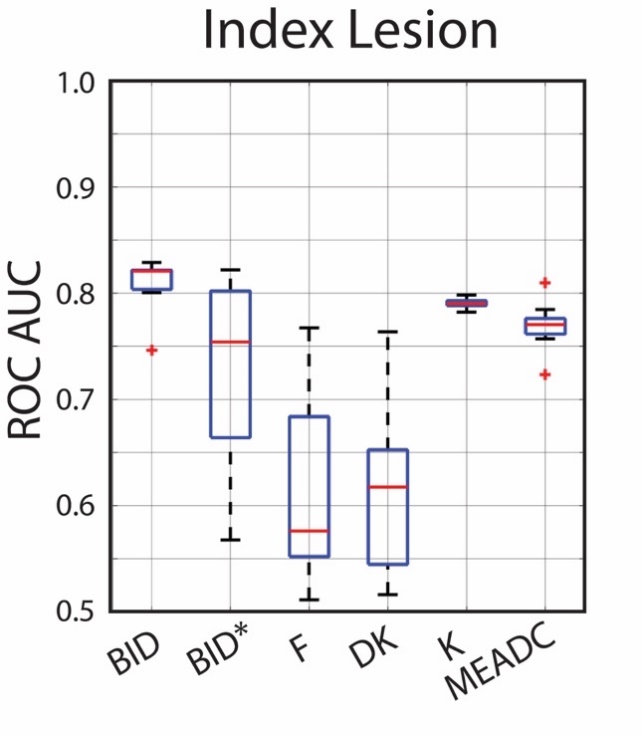
**Supplemental Figure 2.** Clustergram analysis for each DWI contrast. Heat maps with dendrograms showing hierarchical clustering between sites and by region of interest (ROI). Heat maps indicate standard deviation from the mean for each value. Note that more consistency and grouping is seen in the MEADC, K and DK site fit implementations than the BID\*, F and to some extent BID.



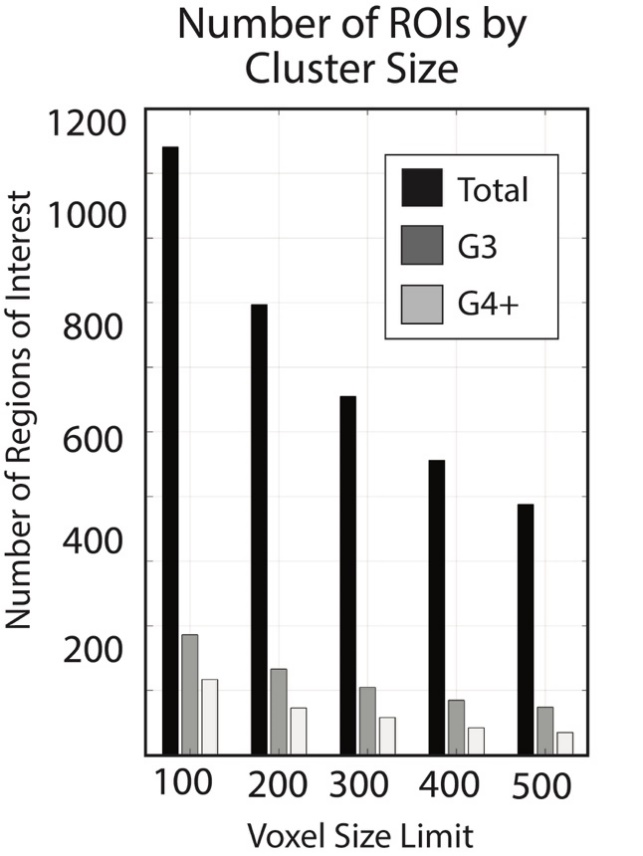
**Supplemental Figure 3.** Analysis of the difference in peripheral zone (PZ) and transition zone (TZ) of the prostate on the receiver operating characteristic area under the curve (ROC AUC) between DWI contrasts. Lesions which crossed zones were counted in the zone with more voxels.



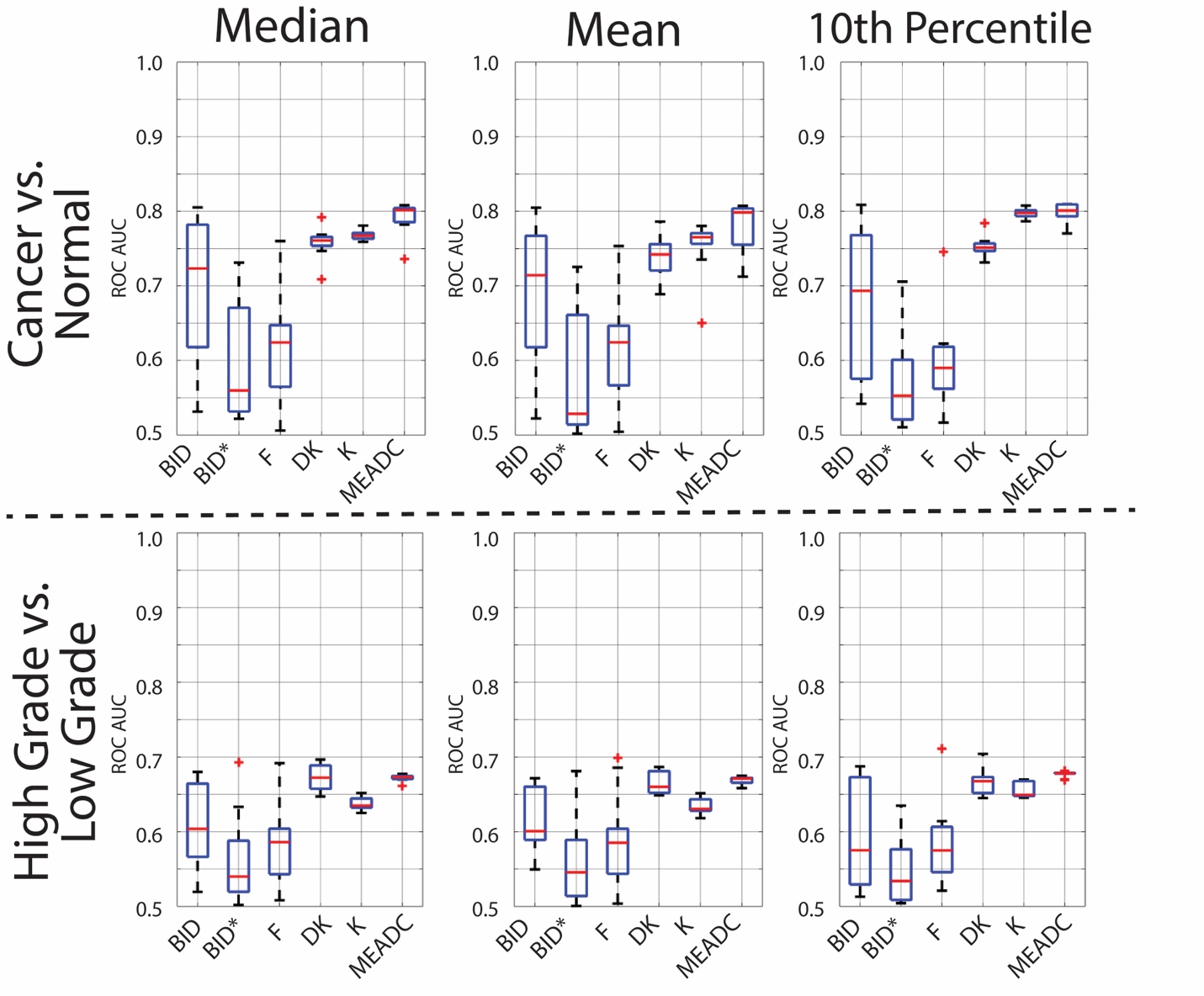
**Supplemental Figure 4.** Boxplot showing receiver operating characteristic area under the curve (ROC AUC) evaluating the largest tumor (index lesion) on each slide compared to the largest region of benign atrophy.



**Supplemental Figure 5.** Number of pathologist annotations that met defined continuous voxel size thresholds. As the size grew, the number of regions of interest included in the analysis dropped.



**Supplemental Figure 6.** Comparison of the area under the receiver operating characteristic while varying the metric used to extract the MRI contrast values from the pathologist annotated ROIs. Top: G3+ vs benign atrophy. Bottom: G3 vs G4+.



**Supplemental Methods**

**DRO Design**

Two separate digital reference objects (DROs) were created for the IVIM and Kurtosis models35,36. Methods for the DRO analysis can be found in the supplement. Each contained simulated trace-DWI DICOM (magnitude images) series generated for bi-exponential intra-voxel incoherent motion (IVIM), and diffusion kurtosis (K) models without noise (infinite SNR) for a range of tissue-relevant diffusion and model-specific parameter (BID\*, F, and K) values to evaluate performance of quantitative DWI analysis algorithms against ground truth36. Each DRO image contained magnitude DWI intensities for a single b-values including b = [0, 10, 25, 50, 80, 100, 200, 500, 800, 1000, 1500, 2000] s/mm2. The simulated image object consisted of 240 zones assembled in 20 rows by 12 columns, where each rectangular zone contained 20x30 pixels. The extra leftmost 30-pixel-wide column of each image contains signal-free sample. Model-specific parameter (F and K) and diffusion coefficient (BID, DK) were monotonically varied across rows and columns, respectively. For IVIM, BID\* parameters were also varied along the 10 "slices". Input diffusion increased as [0.3,0.5,0.7,…,2.5]x10-3mm2/s going from image left-to-right (12 columns) (i.e. patient’s right-to-left). The model-specific (dimensionless) parameter values (20 rows) increased going from image top-to-bottom (i.e. patient’s posterior-to-anterior), as follows: F: [0.02/0.0,0.05,0.1,...,0.95] and K: [0.1,0.2,...,2.0]. For bi-exponential model DWI DRO (10 slices), BID\* varied as [5,10,...30,40,50,70,100]x10-3mm2/s. For KEM DRO, model convergence constraint [REF Jensen 2010] was implemented by explicitly excluding (zeroing-out) signal voxels for DKbmaxK>3. DWI intensity values for each DICOM series were scaled by maximum b=0 DWI to fill 15-bit (unsigned integer) range. The DICOM header was in "classic" (single-frame) format with public tags following standard DWI Macro (<http://dicomlookup.com/lookup.asp?sw=Ttable&q=C.8.13-23>). b-value information was captured in public DICOM tag [0018,9087]. Private tags were intentionally not used. Axial DWI geometry was assumed for all DROs. The true input DRO parametric maps were provided in MHD format.

***DRO Group Analysis*** After each of the site submissions was reoriented into the same space and scaled to be consistent between submissions, the DRO values submitted from each site were analyzed to determine how well they matched the ground truth. To visualize where site implementations varied the largest, the standard deviation of percent difference was mapped (Supplemental Figure 1).

Site Specific Processing:

***Site Implementation 1 Local Methods*** Raw DICOM data were downloaded and further processing was performed in MATLAB (v. 2018a). No smoothing or filtering was done during pre-processing. ME, IVIM, and KEM models were applied to all datasets using the following select b-value ranges, respectively: b-values = [0, 10, 25, 50, 80, 100, 200, 500, 1000], b-values = [0, 10, 50, 100, 200], and b-values = [0, 100, 200, 500, 1000, 2000]. ME parameters were computed by a linear fit of the logarithmic signal data for the b-values below 1000. IVIM parameters were computed using results from the ME fit as initial conditions for the bi-exponential fit using the fit function. KEM parameters were computed by fitting signal to the equation using the lsqcurvefit function.

***Site Implementation 1 Preprocessing at the Central Site***  Submitted NIFTI results were first reoriented to match the direction cosines of the NIFTI converted raw DICOM files that resulted from *mri\_convert (*[*www.surfer.nmr.mgh.harvard.edu*](http://www.surfer.nmr.mgh.harvard.edu)*, v5.3)* being applied to the raw DICOMs. The orientation was manually checked for each patient, and the resulting maps were scaled to match group-level units accordingly.

***Site Implementation 2 Local Methods*** Raw DICOM data were downloaded and pre-processed by sorting files corresponding to the individual series into separate folders, and converting each series into the NRRD multivolume format specific to 3D Slicer and suitable for the analysis by the 3D Slicer DWModeling module of the SlicerProstate extension1,2. Fitting was performed on the non-linearized data as described in detail earlier1. Default initial conditions hard-coded in the DWModeling module were used without modifications.

***Site Implementation 2 Preprocessing at the Central Site***  Submitted NRRD format images were first converted to NIFTI files using 3dSlicer (slicer.org). Values were then scaled to match group-level units accordingly.

***Site Implementation 3 Local Methods*** Raw DICOM data were downloaded. DWI were post-processed using OsiriX MD version 7.5 (Pixmeo SARL, Bernex, Switzerland) and IB Diffusion version 2.0.1263 (Imaging Biometrics LLC, Elm Grove, WI) plugin. All derived maps were masked using a 1% noise threshold. Monoexponential ADC maps were generated in DICOM format with b-values = [200, 500, 1000, and 2000]. IVIM maps of BID, BID\*, and F were generated in DICOM format using a 3 parameter biexponential model3 with b-values = [0,10, 25, 50, 80, 100, 200, 500, 1000, and 2000].

***Site Implementation 3 Preprocessing at the Central Site***  Submitted DICOM results were first converted to NIFTI format using *mri\_convert* and then reoriented and resliced to match the raw datasets using a nearest neighbor interpolation and float data type as implemented in *mri\_convert* and scaled to match group level units accordingly.

***Site Implementation 4 Local Methods*** Raw DICOM data were downloaded and further processing was performed in MATLAB (Mathworks Inc. Natick MA, version 2016b).  Monoexponential ADC maps were calculated using a linear fit through the logarithmic signal data at the different b-values. The DKI parameter maps were calculated as previously described4. Results were submitted in DICOM format.

***Site Implementation 4 Preprocessing at the Central Site***  Submitted DICOM results were first converted to NIFTI format using *mri\_convert*, and scaled to match group-level units accordingly. No reorientation was necessary.

***Site Implementation 5 Local Methods*** Raw DICOM data were downloaded and converted to NIfTI format using DICOM converter “dcm2niix” (<https://github.com/rordenlab/dcm2niix>). All DWI data postprocessing and quantitative imaging metric map generation, detailed below, were performed using in-house–developed software entitled MRI-QAMPER (MRI Quantitative Analysis of Multi-Parametric Evaluation Routines) as previously published5 and Technical Benchmark (Level 3)6 approved by NCI/QIN. The raw images were smoothed with 3D gaussian filter prior to running fitting routines (kernel size = 5 voxels). Multi- b-value DWI data sets were analyzed using four models. For the mono-exponential (ADC) model and non-monoexponential diffusion kurtosis imaging (DKI) model, the DWI data fitting was determined from the logarithmic signal using a linear matrix solution. Bi-exponential modeling of the data was performed using intravoxel incoherent motion (IVIM) which were fit to the logarithmic data using non-linear least squares routines. The choice of upper and lower reference bounds for IVIM nonlinear least squares solver were selected based on mean metric values in tumor and normal prostate tissue as previously published7. Noise in each voxel was corrected from standard deviation of a small patch of background voxels in the b=0 (s/mm2) DWI image8-10. Values were constrained 0<DK<5 (x10-3mm2/s), 0<K<3, 0<BID<3.5 (x10-3mm2/s), BID<BIDS<500 (x10-3mm2/s), and 0<F<1. Randomized seeds were set for each parameter within bounding limits and the equation was fit in each voxel with four iterations to determine the solution with the best fit.

***Site Implementation 5 Preprocessing at the Central Site***  Submitted NIFTI results were first reoriented and resliced to the mri\_convert oriented version of the raw DICOMs. Images were then scaled to match group-level units accordingly.

***Site Implementation 6 Local Methods*** Raw DICOM data were downloaded and processed using in-house written MATLAB code. The ADC maps were calculated using standard monoexponential expressions11. Biexponential intravoxel incoherent motion (IVIM) fit using published equations12 produced diffusion (BID), pseudodiffusion (BID\*), and perfusion fraction (f) maps. Results were submitted in NIFTI format.

***Site Implementation 6 Preprocessing at the Central Site***  Submitted NIFTI results were first reoriented and resliced to the mri\_convert oriented version of the raw DICOMs. Images were then scaled to match group-level units accordingly.

***Site Implementation 7 Local Methods*** Raw DICOM data were downloaded and processed using in-house Matlab tools and published quantitative DWI models13-15. Before model fit, all DWI images were masked off for high-b SNR<1.5 (using GUI) and b>0 exceeding b=0 artifacts. Same mask was used independent of fit model.  For mono-exponential (perfusion-suppressed ADC15) and isotropic kurtosis models, the log-DWI signal dependence on b-value was fit by linear-least-squares (LLS) minimization using Matlab *lscov*-function (for b>150s/mm2). Both intercept and slope were allowed to vary for MEADC LLS fit. For kurtosis model, the maximum allowed b-value (bmax) used by the fit was constrained (using 2nd iteration) bmax<3/(K\*DK) to ensure model convergence14. For bi-exponential model13, the DWI intensities were fit for all b-value using *fminsearch* (Nelder-Mead) non-linear error minimization (NLM) function with maximum iterations set at 1000, function tolerance at 1e-8, and fit-parameter tolerance at 1e-5. The “physical” BID and BID\* values, where then re-assigned from the unconstrained diffusion fit parameters D1 and D2 by using the “lower” D1<3.5x10-3mm2/s, fraction(D1)>0.5 thresholds for the BID voxels. BIPF map was derived as (1-Fraction(BID)). The remaining “unphysical” BID\*<BID voxels were then masked-off (set to zero). The resulting fit maps for the MEADC, DK, and BID, were constrained between zero and max (=5x10-3mm2/s); =500 x10-3mm2/s for BID, =3 for K, and=1 for fractions. The output parametric maps were scaled by 1000 and converted to MHD format.

***Site Implementation 7 Preprocessing at the Central Site*** Submitted MHD results were first converted to NIFTI format using 3dSlicer (slicer.org), then reoriented and resliced to the mri\_convert oriented version of the raw DICOMs. Images were then scaled to match group-level units accordingly.

***Site Implementation 8 Local Methods*** Raw DICOM data were downloaded and processed using in-house Matlab (Mathworks Inc, Natick MA) scripts and published quantitative DWI models16. A custom Matlab script was used to read in the DRO and patient DICOM files and sort the DICOM files by slice position and *b*-value. We applied our image processing and model calibration pipeline to three common DWI models: IVIM, PfIVIM, and Kurtosis. First, a binary mask was created for each patient to indicate voxels in the lowest *b-*value image with signal greater than zero. This mask was applied for all DWI model fits to accelerate the parameter estimation process. Second, each model was fit to the signal intensity time course using the Matlab function *‘lsqcurvefit*’. We used the following additional settings: *‘TolFun’ =* 1e-12, *‘Tolx*’ = 1e-12, *‘MaxIter’* = 500, ‘MaxFunEvals’ = 400.  Third, we used the Matlab function ‘niftiwrite’ so save the model parameters in individual files.  For the biexponential or IVIM model, we fit for four parameters: *S0* (inherent signal intensity), *f* (perfusion fraction), *BID\**(pseudodiffusion coefficient), *BID* (diffusion coefficient). *S0* was bounded from 0 to 5 ´ max(signal intensity), *f* was bounded from 0 to 0.99, *D\**  was bounded from 0 to 400 ´ 103  mm2/s, and *D* was bounded from 0 to 3 ´ 103 mm2/s. For the pfIVIM model, we assumed *D\** was 400 ´ 103 mm2/s, and the remaining parameters were bounded as described above. For the kurtosis model, we fit for *S0*, *D*, and *K* (apparent diffusional kurtosis). *S0* and *D* were bounded as above, while *K* was bounded as from 0 to 3.

***Site Implementation 8 Preprocessing at the Central Site***  Submitted NIFTI results lacked header information. A custom Matlab script (Mathworks Inc. Natick MA) was developed to match headers with the images from the *mri\_convert* version of the converted DICOMs. The resulting images were then checked manually to ensure alignment. Images were then scaled to match group-level units accordingly.

***Site Implementation 9 Local Methods*** Raw DICOM data were downloaded and processed using in-house written Matlab scripts. The data were fitted using three models including standard mono-exponential (M0\*e-MEADC\*b), bi-exponential3 (M0\*[(1-F)\*e-BID\*b + F\*e-BID\*\*b])and kurtosis17 (M0\*e(-DK\*b + (DK\*b)^2\*K/6)). Non-linear fitting was performed for each of the three models using a 'trust-region-reflective' algorithm18.  The initial values for the non-linear fitting were derived using a simple regression, of the log of signal intensity to all b-values for the MEADC, of the log of signal intensity to the b-values ≥200 s/mm2 for BID, K and DK, and of the log of signal intensity to the b-values ≤50 s/mm2 for BID\* and F. The signal intensity values for b=0 s/mm2 was used as the initial value for M0 in the non-linear fit of mono-exponential model, while an initial nonlinear fit of the bi-exponential model was used to obtain the initial M0­ value for the non-linear fit of all other parameters. The output values were constrained to the following bounds: 0≤MEADC≤0.05, 0≤BID≤0.0032, 0.0032≤BID\*≤0.2, 0≤F≤1, 0≤DK≤0.0032, 0≤K≤3. Finally, the data were scaled to prevent excessive quantization error and written out in DICOM format.

***Site Implementation 9 Preprocessing at the Central Site***  Submitted DICOM results were first converted to NIFTI format using *mri\_convert*. Images were then scaled to match group-level units accordingly; no reorientation was needed.

***Site Implementation 11 Local Methods*** Raw DICOM data were downloaded and processed by in-house software *imFIAT*. The DICOM diffusion weighted images as well as headers were directly read by imFIAT. The diffusion weighted images were fitted to a mono-exponential function to obtained ADC maps, and to the intravoxel kurtosis diffusion model17 to yield K and DK maps.  The parameter maps were saved in the DICOM format and submitted the project lead team for further analysis.

***Site Implementation 11 Preprocessing at the Central Site***  Submitted DICOM results were first reorganized, then converted to NIFTI format using 3dSlicer, then reoriented and resliced to the mri\_convert oriented version of the raw DICOMs. Images were then scaled to match group-level units accordingly.

***Site Implementation 12 Local Methods*** The original data in the native DICOM format were downloaded and imported into ImageJ (NIH) software under which our in-house developed in C++ language with double-precision accuracy for mono and bi-exponential fitting modules that run as a plug-in to generate desired maps and save them.in DICOM or NIFTI format.  The plug-in, automatically sorts out the DWI into a multi b-value serial stack of data based on the b-values. This stacking helps to verify if the input data could be properly organized by b-values before the fitting routine is applied. Following that step, sample fittings of the mono- and bi-exponential models were tested at different points in the image to verify that the fitting looks appropriate by using plots of the signal intensity versus b-values. If necessary, the cut-off b-values between fast and slow components can be adjusted to improve the fitting. Also, if desired, a particular b value, such as for example b0, can be marked to get excluded during the fitting calculation. In the case of analyzing several datasets as part of a study, any such exceptions and exclusions will be repeated across all datasets, to maintain consistency. After this verification, the plug-in filter can be run on the entire volume, to generate the required mono- and bi-exponential maps, the mathematical models of which, are explained in detail in reference19.

***Site Implementation 12 Preprocessing at the Central Site***  Submitted DICOM results were first converted to NIFTI format using *mri\_convert*, then reoriented and resliced to the *mri\_convert* oriented version of the raw DICOMs. Images were then scaled to match group-level units accordingly.

***Site Implementation 13 Local Methods*** Raw DICOM data were downloaded into the Osirix image processing platform (Pixmeo.com, Berne, CH) and processed using the Osirix plugin UMM Diffusion (Version 2.3). The Osirix plugin was created by Dr Frank Zöllner at the Medical University Center, Heidelberg University, Mannheim, Baden-Württemberg, DE (http://ikrsrv1.medma.uni-heidelberg.de/redmine/projects/ummdiffusion)20. Briefly, multiple b-value diffusion MRI images were loaded as 4D objects in Osirix where the frames are multiple b-value image sets. The software determines the b-value from the DICOM header (tag 0018,9087).  ADC maps were calculated by 1) a standard mono-exponential fit11, and 2) a bi-exponential intravoxel incoherent motion (IVIM) fit for more than 2 multiple b-value frames with an option for a cut-off perfusion b-value, that produced diffusion (D = slow diffusion), pseudodiffusion (D\* = fast diffusion) and perfusion fraction (f) maps3. Kurtosis maps were computed by the method of Jensen et al.17, and all parametric image maps were saved in DICOM format. MRIconvert (Version 2.0) was used to convert DICOM to NifTi format for image submission.

***Site Implementation 13 Preprocessing at the Central Site***  Images were scaled to match group-level units accordingly, no reorientation was needed.

***Site Implementation 10 and 14 Local Methods*** Raw DICOM data were downloaded and image processing for determination of diffusion parameters, used the publicly available Osirix with the installed plugin module ADCmap (V1.9, Stanford University, <https://github.com/mribri999/ADCmap>). Briefly, DICOM images were loaded as 4D objects in Osirix where the frames are multiple b-value image sets. The software permits manual entry of b-values for each frame of data, as ADCmap has limited ability to detect b-values in the DICOM header. Standard ADC maps were calculated from the mono-exponential decay curve of signal intensity as a function of diffusion b values by performing a nonlinear least-squares fit. Fast and slow diffusion maps and perfusion fraction were derived for intravoxel incoherent motion (IVIM) data using two methods. A two-stage b-value fit beginning with higher b-values (b>200 s/mm2) to produce diffusion (D = slow diffusion), pseudodiffusion (D\* = fast diffusion), perfusion fraction (f) and residual error maps. The Site Implementation 14 method used a bi-exponential fit with non-linear optimization of all parameters simultaneously, and has a non-negativity constraint. All parametric image maps were saved in DICOM format, and converted to NifTi using MRIconvert (v2.1, University of Oregon, <https://lcni.uoregon.edu/downloads/mriconvert>).

***Site Implementation 10 & 14 Preprocessing at the Central Site*** Submitted NIFTI results were first reoriented and resliced to the *mri\_convert* oriented version of the raw DICOMs. Images were then scaled to match units accordingly.

**Supplemental References (also cited in the main manuscript)**

1. Langkilde F, Kobus T, Fedorov A, et al. Evaluation of fitting models for prostate tissue characterization using extended-range b-factor diffusion-weighted imaging. *Magn Reson Med.* 2018;79(4):2346-2358.

2. <https://github.com/SlicerProstate/SlicerProstate>. 2020.

3. 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(2):497-505.

4. Hectors SJ, Semaan S, Song C, et al. Advanced Diffusion-weighted Imaging Modeling for Prostate Cancer Characterization: Correlation with Quantitative Histopathologic Tumor Tissue Composition-A Hypothesis-generating Study. *Radiology.* 2018;286(3):918-928.

5. Paudyal R, Konar AS, Obuchowski NA, et al. Repeatability of Quantitative Diffusion-Weighted Imaging Metrics in Phantoms, Head-and-Neck and Thyroid Cancers: Preliminary Findings. *Tomography.* 2019;5(1):15-25.

6. Farahani K, Tata D, Nordstrom RJ. QIN Benchmarks for Clinical Translation of Quantitative Imaging Tools. *Tomography.* 2019;5(1):1-6.

7. Pang Y, Turkbey B, Bernardo M, et al. Intravoxel incoherent motion MR imaging for prostate cancer: an evaluation of perfusion fraction and diffusion coefficient derived from different b-value combinations. *Magn Reson Med.* 2013;69(2):553-562.

8. Constantinides CD, Atalar E, McVeigh ER. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magn Reson Med.* 1997;38(5):852-857.

9. Kristoffersen A. Optimal estimation of the diffusion coefficient from non-averaged and averaged noisy magnitude data. *J Magn Reson.* 2007;187(2):293-305.

10. Lu Y, Jansen JF, Mazaheri Y, Stambuk HE, Koutcher JA, Shukla-Dave A. Extension of the intravoxel incoherent motion model to non-gaussian diffusion in head and neck cancer. *J Magn Reson Imaging.* 2012;36(5):1088-1096.

11. Stejskal EO, Tanner JE. Spin diffusion measurements: spin echoes in the presence of a time‐dependent field gradient. *The journal of chemical physics.* 1965;42(1):288-292.

12. Du J, Li K, Zhang W, et al. Intravoxel Incoherent Motion MR Imaging: Comparison of Diffusion and Perfusion Characteristics for Differential Diagnosis of Soft Tissue Tumors. *Medicine (Baltimore).* 2015;94(25):e1028.

13. Dyvorne HA, Galea N, Nevers T, et al. Diffusion-weighted imaging of the liver with multiple b values: effect of diffusion gradient polarity and breathing acquisition on image quality and intravoxel incoherent motion parameters--a pilot study. *Radiology.* 2013;266(3):920-929.

14. Jensen JH, Helpern JA. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed.* 2010;23(7):698-710.

15. Lewin M, Fartoux L, Vignaud A, Arrive L, Menu Y, Rosmorduc O. The diffusion-weighted imaging perfusion fraction f is a potential marker of sorafenib treatment in advanced hepatocellular carcinoma: a pilot study. *European radiology.* 2011;21(2):281-290.

16. Koh DM, Collins DJ, Orton MR. Intravoxel incoherent motion in body diffusion-weighted MRI: reality and challenges. *AJR Am J Roentgenol.* 2011;196(6):1351-1361.

17. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med.* 2005;53(6):1432-1440.

18. Branch MA, Coleman TF, Li Y. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing.* 1999;21(1):1-23.

19. Newitt DC, Malyarenko D, Chenevert TL, et al. Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network. *J Med Imaging (Bellingham).* 2018;5(1):011003.

20. Zöllner F, Kaiser S, Weisser G, Schad L. UMMDiffusion: an OsiriX plug-in for ADC and IVIM analysis in clinical routine. *Salt Lake CIty, USA.* 2013:3115.