Tractography algorithms: how do they affect connectomes?

Dmitry Petrov, Yulia Dodonova

January 28
IITP

Overview

— Why tractography is important?

— Main algorithms to fiber tracking

1. Tensor
2. Orientation distribution function (ODF)
3. Probability

— Open questions

1. Is there “best” tractography algorithm in terms of predictive power?
2. Can we combine data from them to boost predictive power?
3. How does field strength affect connectome weights?
4. How do spatial/angular resolution affect weights?

— Conclusion

Overview

— Why tractography is important?

— Main algorithms to fiber tracking

1. Tensor-based
2. Orientation distribution function (ODF)
3. Probability

— Open questions

1. Is there “best” tractography algorithm in terms of predictive power? (Spoiler: maybe no)
2. Can we combine data from them to boost predictive power? (Spoiler: maybe yes)
3. How does field strength affect connectome weights? (Spoiler: it looks like 7T is slightly better than 3T)
4. How do spatial/angular resolution affect weights? (Spoiler: looks like thinner tracts are the victims)

— Conclusion

Why is tractography important?

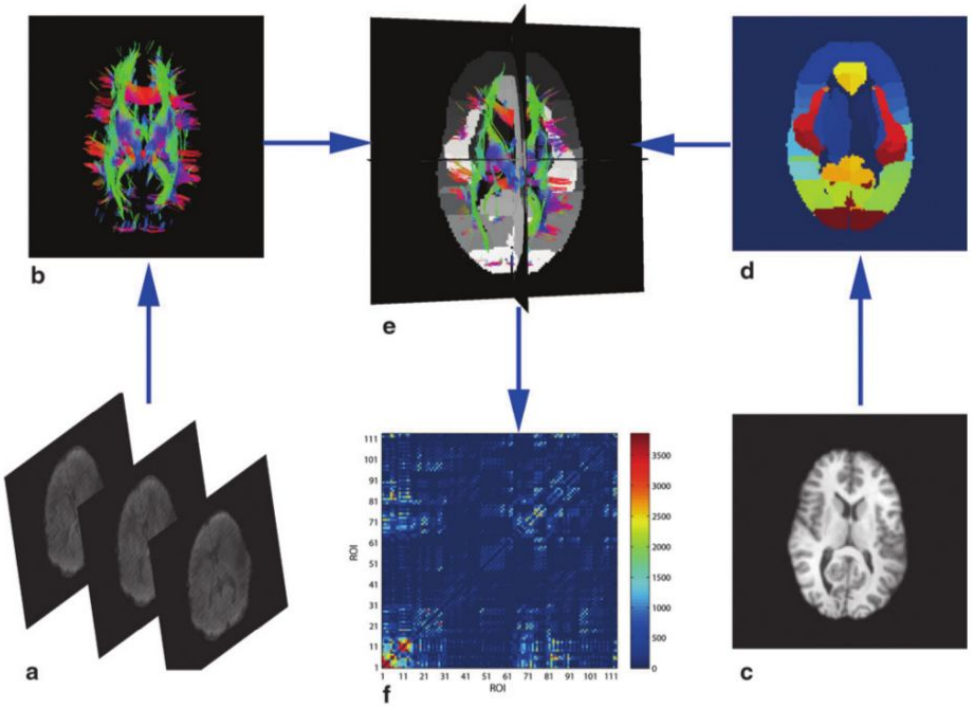


FIG. 1. Flowchart of steps to compute brain connectivity. Diffusion weighted images **(a)** are used as the basis to compute maps of whole-brain tractography **(b)**; in parallel, the standard T1-weighted anatomical magnetic resonance image from the same subject **(c)** is parcellated using the Harvard/Oxford Cortical and Subcortical probabilistic atlases, to define the regions of interest (ROIs) **(d)** by counting the number of detected fibers connecting each pair of ROIs **(e)**, and expressing them as a proportion of all fibers recovered in the entire brain, we can create the anatomical connectivity matrix **(f)**, for each subject in the study, and for each type of scan they had.

— tractography algorithms provide connection weights

— there are two main steps in tractography: to fit a diffusion model into each voxel and track fiber across the voxels

— there are two main classes of fiber tracking: probabilistic and deterministic

Image source: “Boosting brain connectome classification accuracy in Alzheimer’s disease using higher-order singular value decomposition», (Zhan et al.)

Algorithms overview

— Fiber assignment by continuous tracking (tensor-FACT).

— Second-order Runge-Kutta (tensor-RK2).

— Interpolated streamline (tensor-SL).

— Tensorline (tensor-TL).

— ODF-FACT

— ODF-RK2

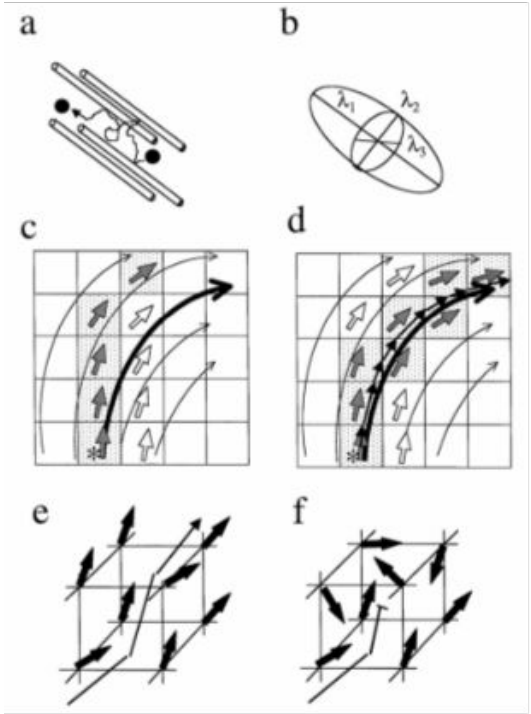
— PICo

— Hough

— Probtrackx

Tensor-FACT

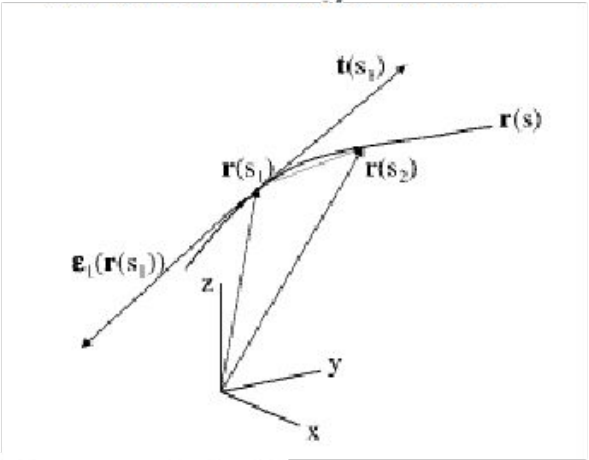
Fiber assignment by continuous tracking



Mori et al., 1999

Tensor-RK2

2d-order Runge-Kutta



$$\mathbf{t}(s) = \boldsymbol{\varepsilon}_1(\mathbf{r}(s))$$

$$\frac{d\mathbf{r}(s)}{ds} = \boldsymbol{\varepsilon}_1(\mathbf{r}(s)) \qquad \mathbf{r}(0) = \mathbf{r}_0$$

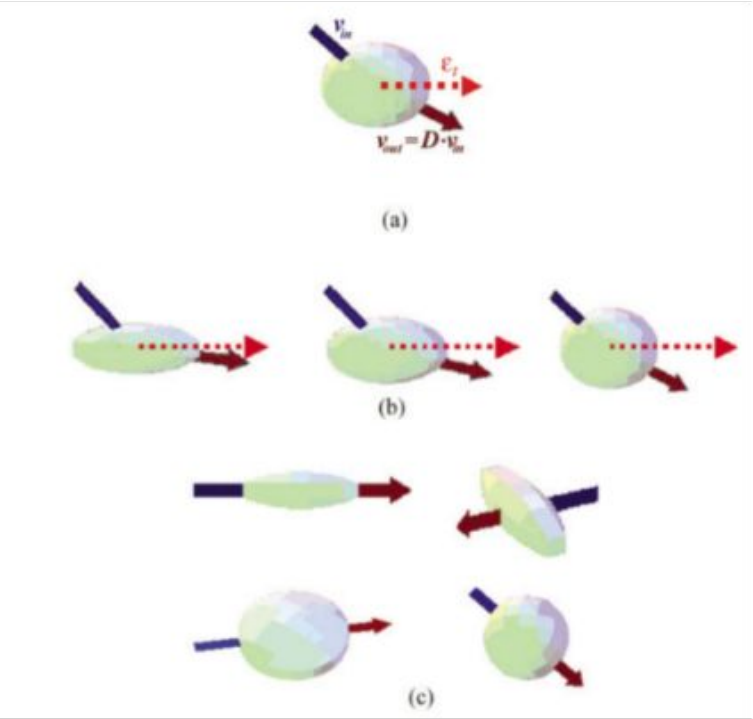
$$\mathbf{r}(s_1) = \mathbf{r}(s_0) + \mathbf{r}'(s_0)\,(s_1 - s_0) +$$

$$\mathbf{r}(s_1) \sim \mathbf{r}(s_0) + \alpha\boldsymbol{\varepsilon}_1(\mathbf{r}(s_0))$$

Basser et al., 2000

Tensor-TL

Tensorline (instead of streamline)



$$\mathbf{v}_{out} = \mathbf{D} \cdot \mathbf{v}_{in}$$

$$\mathbf{v}_{in} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{e}_3$$

$$\mathbf{v}_{out} = \lambda_1 \left(\alpha_1 \mathbf{e}_1 + \frac{\lambda_2}{\lambda_1} \alpha_2 \mathbf{e}_2 + \frac{\lambda_3}{\lambda_1} \alpha_3 \mathbf{e}_3 \right)$$

$$\mathbf{v}_{out} = f\mathbf{e}_1 + (1 - f)((1 - g)\mathbf{v}_{in} + g\mathbf{D} \cdot \mathbf{v}_{in})$$

Lasar et al., 2003

Tensor-SL

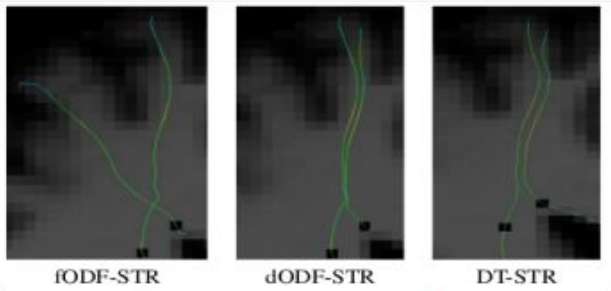
Interpolated streamline

The tensor D and its eigenvector corresponding to the largest eigenvalue were calculated at each step, from interpolated DT-MRI data, to define the direction of the next step.

Conturo et al., 1999

ODF-based methods

Orientation Distribution Function



In our analytical QBI solution, the signal at position p is first estimated as

$$S(\mathbf{u})_p = \sum_{j=1}^{R} c_j Y_j(\mathbf{u}), \qquad (1)$$

where $S(\mathbf{u})$ is the measured diffusion weighted signal in each of the N gradient directions $\mathbf{u} := (\theta, \phi)$ on the sphere (θ, ϕ obey physics convention, $\theta \in [0, \pi], \phi \in [0, 2\pi]$), c_j are the SH coefficients describing the signal, Y_j is the j^{th} element of the SH basis and $R = (1/2)(\ell + 1)(\ell + 2)$ is the number of terms in the basis of order ℓ when choosing only even orders.

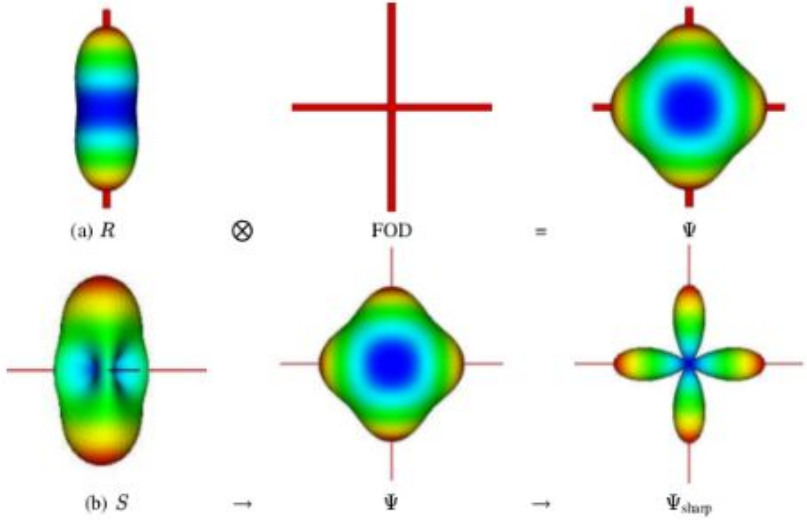


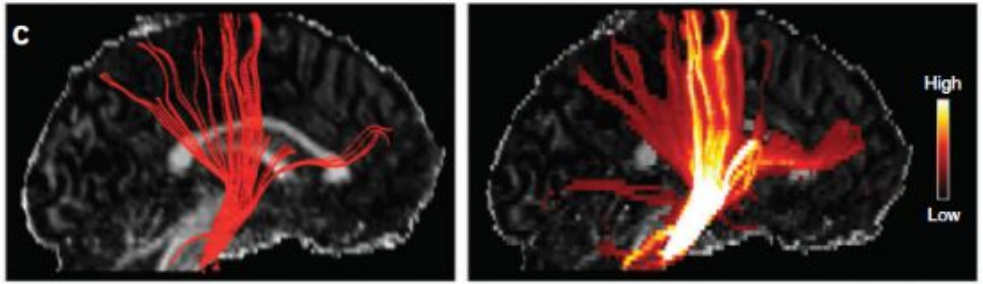
Fig. 1. Sketch of the convolution/deconvolution process. In (a), the convolution between the diffusion ODF kernel R and the fibre orientation function (FOD) produces a smooth diffusion ODF Ψ. In (b), we show a sketch of the deconvolution sharpening. The Funk-Radon Transform of the simulated HARDI signal on the sphere S produces a smooth diffusion ODF Ψ. This diffusion ODF is transformed into a sharper fibre ODF Ψ_{sharp} by deconvolution with the diffusion ODF kernel R of (a).

Descoteaux et al., 2008

Table 3
Summary of three major dimensions along which most tractography algorithms can be classified.

Dimension	Approach	Description
Probabilistic vs deterministic	Deterministic	Propagates single trajectories in accordance with the principal direction of water diffusion (e.g., Basser et al., 2000). Does not estimate the spatial uncertainty of the trajectory.
	Probabilistic	Samples a direction distribution function at each step to determine the propagation direction. Allows estimation of a probability density of the most likely location of the tract, and thus its spatial uncertainty (e.g., Behrens et al., 2003).
Local vs global	Locally greedy	Trajectories propagate incrementally using a near-sighted, voxel-by-voxel approach (Basser et al., 2000; Behrens et al., 2003). Can be affected by noisy voxels.
	Globally optimal	Estimates the globally optimal path between two regions, typically by representing voxel-wise water diffusion as a connected graph and finding the shortest path between seed and target voxels (Iturria-Medina et al., 2007; Iturria-Medina et al., 2008; Zalesky, 2008; Zalesky and Fornito, 2009). More robust to noise.
Single vs multi-direction	Single direction	The direction of water diffusion in each voxel is represented using the primary eignvector of the diffusion tensor (Basser et al., 2000; Behrens et al., 2003). Does not distinguish crossing fibers.
	Multi-direction	The direction of water diffusion in each voxel is represented using an orientation distribution function (Behrens et al., 2007; Tournier et al., 2004). Allows resolution of crossing fibers, but requires good quality, high angular resolution data.

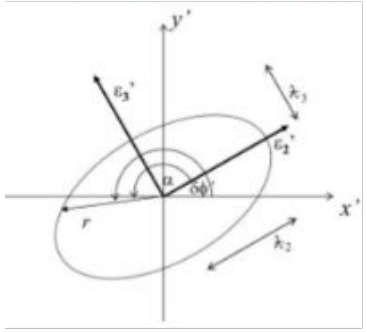
Diffusion model	Single-direction		Multi-direction	
Tractography method	Deterministic	Probabilistic	Deterministic	Probabilistic
Local tractography	**SDD** - **DTI streamline** [1] - Tensor deflection [2,3]	**SDP** - **PICo** [4] - PROBTRACK [5]	**MDD** - ODF streamline [6,7,8] - **CSD streamline** [9]	**MDP** - **iFOD** [10] - PROBTRACKX [11] - PAS-PICO [12]
Global tractography	**SDG** - **DTI Graph-tractography** [13] - ConTrack [14]		**MDG** - **CSD Multigraph-tractography** [15] - BlueMatter [16]	

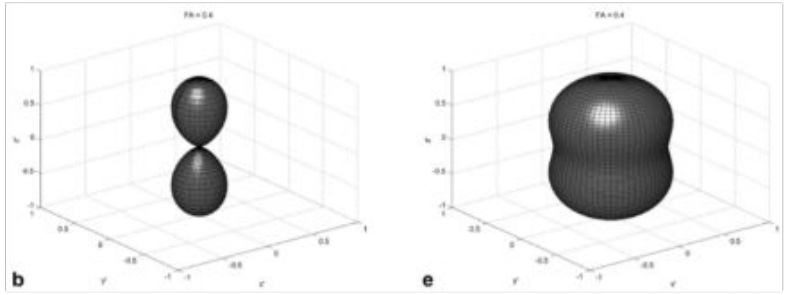


Bastiani et al., 2012

PICo

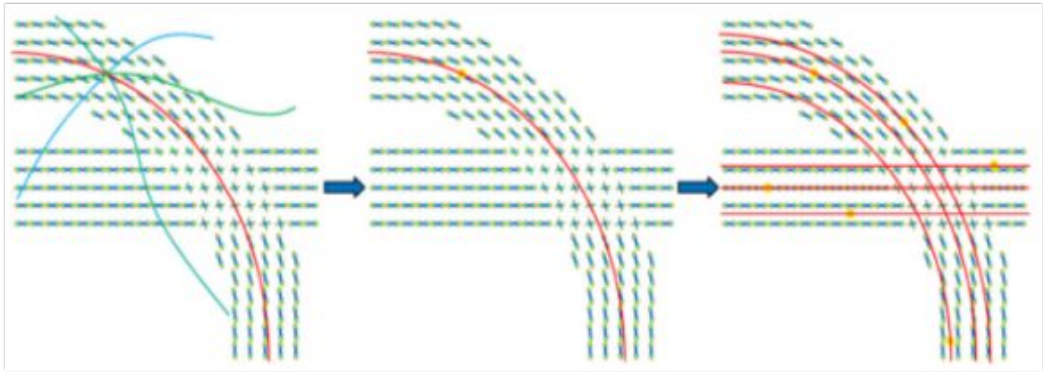
Probabilistic Index of Connectivity





Geoffrey et al., 2003

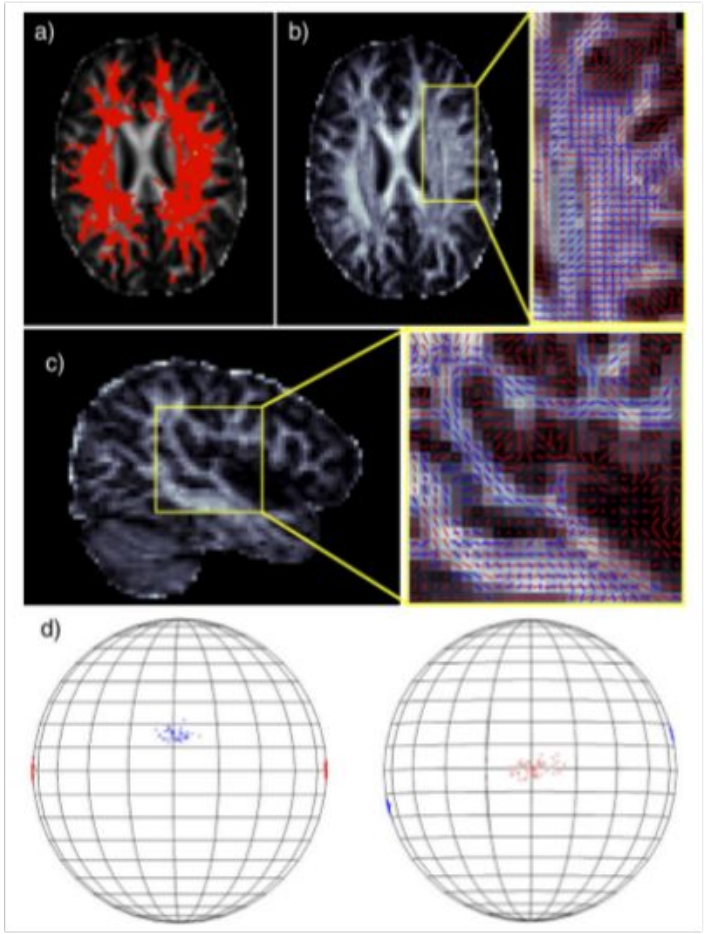
Hough



Aganj et al., 2011

ProbtrackX

X for crossing



Behrens et al., 2007

Is there “best” algorithm in terms of predictive power?

— It looks like there is no universally optimal method according to “Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer’s disease» (Zhan et al., 2015);

— Authors also didn’t found universally helpful method of dimensionality reduction;

— However, they did found, that some diagnostic groups were easier to discover than others (surprize-surprize).

Connectome AD comparison: pipeline and demographics

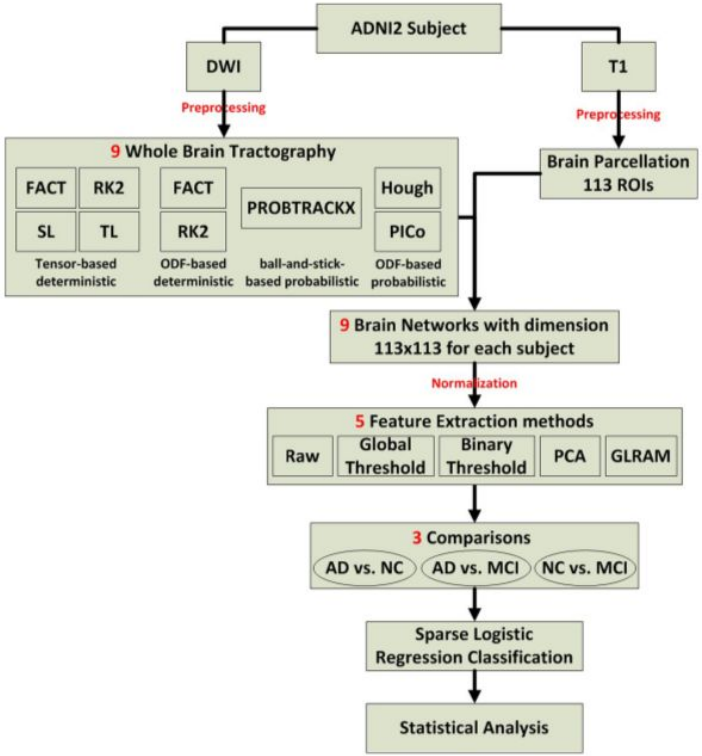


TABLE 1 | Summary of ADNI data used in this study.

	Normal control (NC)	MCI (MCI)	AD	Total
Number	51	112	39	202
Age (y)	69.69 ± 15.43	71.68 ± 9.89	75.56 ± 9.11	71.92 ± 11.54
Sex	29F	41F	14F	84F

There is no age difference among these groups based on a one-way ANOVA ($p = 0.0536$) but the proportion of women in HC group (56.86%) was higher than that of the AD (35.90%) or MCI groups (36.61%).

Parcellation atlas: Harvard Oxford Cortical and Subcortical probabilistic atlas (113×113)

Image auisition info: Each subject underwent whole-brain MRI scanning on Tesla GE Medical Systems scanners. T1-weighted SPGR(spoiled gradient echo) sequences (256 x 256 matrix; voxel size = 1.2 × 1.0 × 1.0 mm 3 ; TI = 400 ms; TR = 6.98 ms; TE = 2.85 ms; flip angle = 11 ◦), were collected as well as DWI (128 × 128 matrix; voxel size: 2.7 × 2.7 × 2.7 mm 3 ; scan time = 9 min; more imaging details may be found at http://adni.loni.usc.edu/wp-content/uploads/2010/05/ADNI2_GE_3T_22.0_T2.pdf). 46 separate images were acquired for each DWI scan: five T2-weighted images with no diffusion sensitization (b 0 images) and 41 DWIs (b = 1000 s/mm 2). The DWI protocol for ADNI was chosen after a detailed evaluation of different protocols that could be performed in a reasonable amount of time; we reported these comparisons previously (Jahanshad et al., 2010; Zhan et al., 2013a). All T1-weighted MR and DWI images were checked visually for quality assurance toexclude scans with excessive motion and/or artifacts; all scans were included.

FIGURE 1 | Flow chart describing the steps taken in this study to create, analyze, and compare structural networks.

Source: "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease" (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: features

1. Raw features. All the values in the upper triangle of the matrices — 6328 (113 × 112/2) features overall;
2. Global threshold. Subset of previous features based on Top 5-40% largest elements of mean matrix across all subjects (6328 features);
3. Individual binary threshold. Top largest values mask of the top 5-40% of individual matrix (6328 features);
4. PCA. First k principal components on raw features matrix, where k varies from 10 to 150. The reduced input matrix is then used to perform classification;
5. GLRAM. Generalized low-rank approximation, which factorizes all the subject matrices into three components. That is, for the matrix of each subject M_i , it factorizes as $M_i = L \times X_i \times R$, where $L \in R^{d \times k}$ and $R \in R^{k \times d}$ are shared orthonormal transformations for all matrices, and $X_i \in R^{k \times k}$ is a reduced matrix. X_i is the new feature representation for classification. One important parameter in GLRAM is the reduced row/column dimensionality. Again, a range of parameter values was investigated to seek the “best” option for the classification. The feature space is m^2.

Source: “Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer’s disease” (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: classification pipeline

— z-score normalization performed for each feature;

— To avoid class imbalance, authors constructed 20 balanced training/testing sample splits, as follows:
(a) Randomly draw 85% of the data from the smaller class for training, and the remaining 15% for testing.
(b) In the larger class, match the same number of training samples by a random subsampling, and the rest are put in the test set;

— Classifier: sparse logistic regression;

— Main metric: ROC AUC.

Source: “Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer’s disease” (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: *post hoc* statistical analysis

— The 95% confidence interval (CI) for the AUC was computed and one-way analysis of variance (ANOVA) was performed on the AUCs;

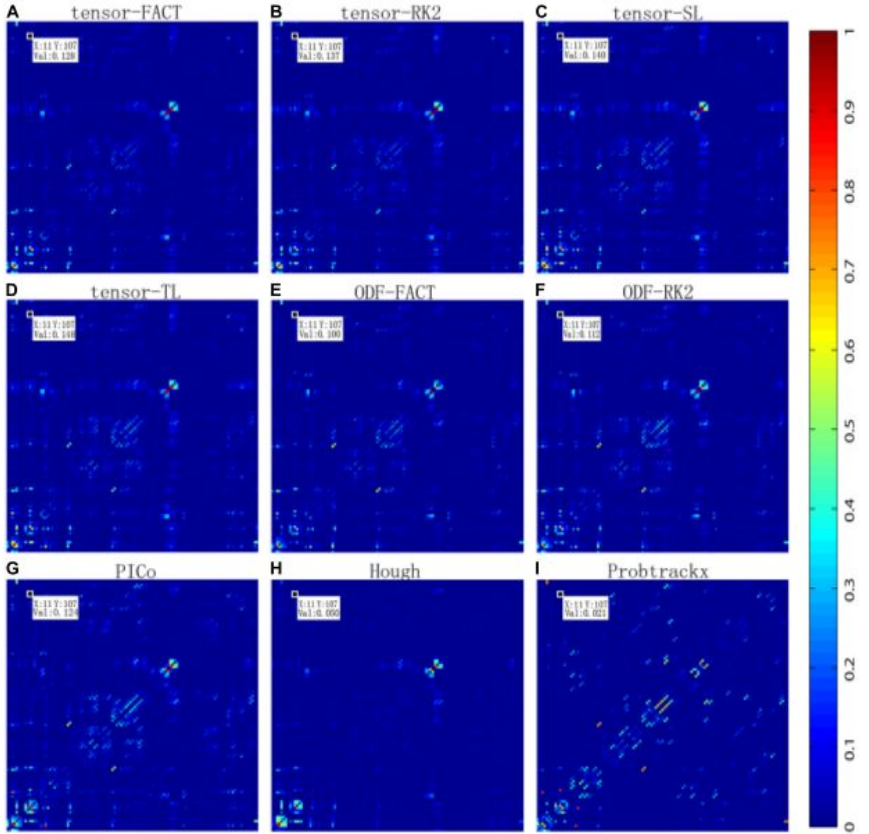
— Null hypothesis (H_0) was: there is no significant difference in the AUCs from different tractography algorithms;

— The experiment-wise alpha threshold was set to $p < 0.05$;

— All the p-values reported have been adjusted by SPSS with the appropriate correction for the effective number of multiple comparisons used (Bonferroni correction). For instance, for a three-group experiment, a pairwise comparison (i.e., a t-test) that yields a p-value of 0.016 would be considered significant at the 0.05 level, because $0.016 < (0.05/3)$.

Source: “Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer’s disease” (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: matrices visualisation



Comparison of averaged normalized brain networks from nine different tractography algorithms, including (A) tensor-based FACT; (B) tensor-based RK2; (C) tensor-based SL; (D) tensor-based TL; (E) ODF-based FACT; (F) ODF-based RK2; (G) ODF-based PICo; (H) ODF-based Hough, and (I) Ball-and-stick model based Probtrackx, from all 202 subjects.

In each network, each cell represents the connectivity between each pair of ROIs; the ROI index runs from 1 to 113 from left to right and from bottom to top. ROI names are detailed in the Supplement. Visually, brain networks from different tractography algorithms may have similar patterns but in reality, the recovered brain network varies, as shown by the value in the randomly selected cell (11,107).

Selected pixel weights (x=11, y=107): 0.128, 0.137, 0.140, 0.148, 0.1, 0.112, 0.124, 0.05, 0.021.

Source: "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease" (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: raw matrices

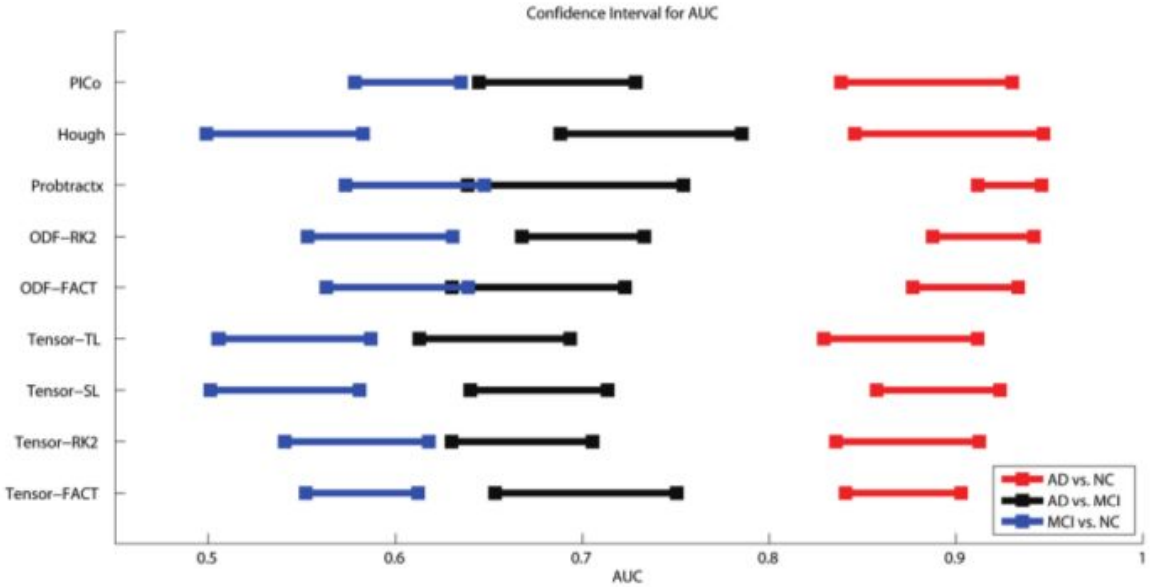


FIGURE 3 | Ninety five percent confidence intervals (CI) for the AUC (classification accuracy) for three diagnostic tasks represented by the three color bars (blue, black, and red) for nine tractography algorithms. The red color means the AD vs. NC classification task, black colors denote AD vs. MCI, and blue colors indicate MCI vs. NC. The y-axis indicates the tractography algorithms and x-axis shows the AUC value. In theory, the higher the AUC value, the better the classification performance. However, if the CI of AUCs has some overlap, we cannot conclude that one algorithm is better than the others, even if the mean AUCs are numerically different. As is evident from the three color bars' horizontal positions, some classifications are more difficult: AD v. NC is the easiest, and MCI v. NC is the most difficult, perhaps in line with expectation.

Source: "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease" (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: raw matrices and thresholded

TABLE 3 | One-way ANOVA test on the classification performance of nine tractography algorithms for three diagnostic tasks when using the raw matrices as features.

Diagnostic task		Degrees of freedom	*F*	Sig.
AD vs. NC	Between groups	8	1.111	0.358
	Within groups	171		
AD vs. MCI	Between groups	8	1.348	0.223
	Within groups	171		
MCI vs. NC	Between groups	8	1.945	0.056
	Within groups	171		

The "F" column presents computed F score and the "Sig." column gives the p-value. Results with Sig. value < 0.05 are treated as nominally significant, so no differences were detectable. "Between Groups" represents sum of the squared deviations from the mean between groups, which captures variability between each group. "Within Groups" represents sum of the squared deviations from the mean within groups, which captures variability within each group. We have nine tractography algorithms, so the number of degrees of freedom for the Between Groups comparison is 9−1 = 8. And since we have 20 splits for each algorithm, the number of degrees of freedom for the Within Groups comparison is 20x9−9 = 171. Since $\alpha = 0.05$ and the number of degrees of freedom = (8,171), we accept H_0 if $F_{8,171} \leq 1.9929$. All our three F-values (1.111, 1.348, and 1.945) are all less than 1.9929, so we accept our H_0. In other words, there are no significant group differences in these nine tractography algorithm-derived networks using the raw matrices as features.

Supplementary Table 2. One-Way ANOVA on AUCs computed from 9 tractography algorithm-derived thresholded matrices. The threshold value for each tractography was different and we chose the one with the largest average AUC from 8 possible threshold values (0.05~0.40). Again the degree of freedom for "Between Groups" is 9-1=8 and the degree of freedom for "Within Groups" is 9x20-9=171, so our critical F value at α=0.05 level is 1.9929. Our computed F values in this table are all less than 1.9929, which means there is no evidence to reject the *H0*, in other words, there are no statistical differences among the AUCs from these 9 tractography algorithm-derived thresholded matrices in each diagnostic task, no matter Global Threshold or Individual Binary Threshold.

Diagnostic Task		Degree of freedom	Global Threshold		Individual Binary Threshold	
			F	Sig.	F	Sig.
AD vs NC	Between Groups	8	1.296	.248	.531	.832
	Within Groups	171				
AD vs MCI	Between Groups	8	.857	.554	1.583	.133
	Within Groups	171				
MCI vs NC	Between Groups	8	1.590	.131	1.906	.062
	Within Groups	171				

Source: "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease" (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: individual threshold

TABLE 4B | *Post hoc* comparison results for the Individual Binary Threshold method.

Diagnostic tasks	Tractography algorithm	(I) Threshold	(J) Threshold	Mean difference (I-J)	Sig.	95% confidence interval	
						Lower bound	Upper bound
AD vs. MCI	Tensor-RK2	0.05	0.30	−0.09325	0.023	−0.01800	−0.0065
			0.40	−0.08961	0.035	−0.1763	−0.0029
	Hough	0.05	0.25	−0.10897	0.034	−0.2142	−0.0038
MCI vs. NC	Tensor-TL	0.05	0.35	−0.10285	0.014	0.0109	0.1948
	PICo	0.05	0.15	0.11037	0.028	0.0056	0.2151
			0.35	0.12817	0.004	0.0234	0.2329
			0.40	0.12654	0.005	0.0218	0.2313

The "Sig." column shows the SPSS adjusted p-value and only values below 0.05 are treated as nominally significant (Please refer to the footnote for detailed explanation). Only comparisons that passed Bonferroni correction are shown here. 95% confidence interval is on the mean difference (I-J). Using the Individual Binary Threshold as the feature extraction method, the AUCs from some tractography algorithms may be statistically affected by the threshold values chosen for specific diagnostic tasks.

Source: "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease" (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: PCA

TABLE 5B | *Post hoc* comparisons results.

Diagnostic tasks	Tractography algorithm	(I) PC number	(J) PC number	Mean difference (I-J)	Sig.	95% confidence interval	
						Lower bound	Upper bound
AD vs. NC	Tensor-RK2	15	75	0.16667	0.003	0.0293	0.3040
			150	0.16759	0.003	0.0302	0.3050
		20	75	0.15370	0.011	0.0163	0.2911
			150	0.15463	0.010	0.0173	0.2920
AD vs. MCI	Tensor-FACT	10	150	0.14378	0.022	0.0092	0.2784
	Tensor-RK2	10	150	0.17437	0.016	0.0150	0.3338
	Tensor-SL	10	40	0.12890	0.017	0.0105	0.2473
			100	0.12099	0.038	0.0026	0.2394
			150	0.19536	0.000	0.0770	0.3137
		15	150	0.17532	0.000	0.0569	0.2937
		20	150	0.12342	0.030	0.0050	0.2418
	Tensor-TL	10	150	0.17099	0.001	0.0383	0.3037
		15	150	0.14747	0.013	0.0148	0.2802
	ODF-FACT	10	150	0.11424	0.019	0.0084	0.2200
	Probtrackx	10	100	0.14219	0.000	0.0519	0.2325
		15	100	0.12236	0.000	0.0320	0.2127
		20	100	0.10876	0.004	0.0184	0.1991
		25	100	0.12500	0.000	00347	0.2153
		30	100	0.13544	0.000	0.0451	0.2258
		35	100	0.13397	0.000	0.0436	0.2243
		40	100	0.10506	0.006	0.0147	0.1954
		45	100	0.09726	0.019	0.0069	0.1876
		50	100	0.09515	0.026	0.0048	0.1855
	Hough	10	150	0.12416	0.008	0.0154	0.2329
		15	150	0.11994	0.014	0.0112	0.2287
MCI vs. NC	ODF-RK2	40	150	0.11920	0.012	0.0125	0.2259
	Probtrackx	10	100	0.12047	0.002	0.0247	0.2162
		15	100	0.09728	0.041	0.0016	0.1930

Only tests that passed Bonferroni correction are shown here. Using PCA as a feature extraction method, the AUCs for some tractography algorithms are statistically affected by the number of PCs for specific diagnostic tasks. Moreover, a smaller number of PCs tends to give better performance (higher AUC) than higher numbers of PCs for these tractography algorithms when using PCA.

TABLE 6A | Statistical analysis results for classification performances from nine tractography algorithms using PCA. (A) One-way ANOVA.

Task		Degrees of freedom	*F*	Sig.
AD vs. NC	Between groups	8	3.144	0.002
	Within groups	171		
AD vs. MCI	Between groups	8	2.191	0.030
	Within groups	171		
MCI vs. NC	Between groups	8	2.728	0.007
	Within groups	171		

We have nine tractography algorithms, so the number of degrees of freedom for the Between Groups comparison is 9−1 = 8. And since we have 20 splits for each algorithm, the number of degrees of freedom for the Within Groups comparison is 20x9−9 = 171. Since $\alpha = 0.05$ and the number of degrees of freedom = (8,171), we accept H_0 if $F_{8,171} \leq 1.9929$. All our three F-values (3.144, 2.191, and 2.728) are larger than 1.9929, so we reject our H_0; in other words, there are significant differences in these nine tractography algorithms in classification using PCA to extract features. However, for the task AD vs. MCI, no group comparison passes Bonferroni correction in the post hoc tests.

Source: "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease" (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: significant algorithms difference

TABLE 6B | *Post hoc* group comparisons.

Task	(I) Tractography algorithm	(J) Tractography algorithm	Mean difference (I-J)	Sig.	95% confidence interval	
					Lower bound	Upper bound
AD vs. NC	Tensor-SL	ODF-FACT	−0.09444	0.006	−0.1741	−0.0148
		ODF-RK2	−0.09028	0.011	−0.1700	−0.0106
CMI vs. NC	Probtrackx	Tensor-FACT	0.10109	0.011	0.0119	0.1903
		Tensor-RK2	0.10091	0.011	0.0117	0.1901
		Tensor-TL	0.09339	0.030	0.0042	0.1826
		ODF-RK2	0.09348	0.030	0.0043	0.1827

The "Sig." column show the SPSS adjusted p-value; only values 0.05 are treated as significant. Only comparisons that passed Bonferroni correction are listed here. For the task AD vs. NC, the classification performance of tensor-SL is significantly poorer than that of ODF-FACT or ODF-RK2. Interestingly, for the task MCI vs. NC, Probtrackx has statistically better performance than the four deterministic tractography algorithms (tensor-FACT, RK2, TL, and ODF-RK2).

Source: "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease" (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Connectome AD comparison: main feature-algorithm differences

TABLE 8B | *Post hoc* comparisons.

Diagnostic tasks	Tractography algorithm	(I) Feature extraction method	(J) feature extraction method	Mean difference (I-J)	Sig.	95% confidence interval	
						Lower bound	Upper bound
AD vs. NC	Tensor-SL	Individual binary threshold	PCA	0.09259	0.005	0.0186	0.1666
	Probtrackx	Raw feature	GLRAM	0.05833	0.042	0.0011	0.1155
		Global threshold	GLRAM	0.06296	0.021	0.0058	0.1202
AD vs. MCI	Tensor-TL	Individual binary threshold	Raw feature	0.09900	0.020	0.0096	0.1884
	ODF-RK2	PCA	GLRAM	0.10348	0.007	0.0183	0.1887
	Hough	Individual binary threshold	Raw feature	0.10612	0.007	0.0193	0.1930
			Global threshold	0.08966	0.038	0.0028	0.1765
			GLRAM	0.10232	0.010	0.0155	0.1892
		PCA	Raw feature	0.09768	0.017	0.0108	0.1845
			GLRAM	0.09388	0.025	0.0070	0.1807
MCI vs. NC	PICo	Individual binary threshold	Raw feature	0.09697	0.036	0.0035	0.1904

The "Sig." column show the SPSS adjusted p-value; only values 0.05 are treated as significant. Only comparisons that passed Bonferroni correction are listed here. Although some methods show better performance for some tractography algorithms in some specific tasks, the trend is not consistent, and there is no universally optimal method.

Source: "Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease" (Liang Zhan, Jiayu Zhou, Yalin Wang, Yan Zhin et al., 2015)

Can we combine tractography data to boost predictive power?

— Yes, according to “Discriminative fusion of multiple brain networks for early mild cognitive impairment detection» (Wang et al.);

— Convex sum of 9 tractography algorithms boosted AUC to 0.89 from 0.66 by best algorithm (Probtrackx);

— Averaged connectome from 9 algorithms performed worse than each tractography individually (0.50 AUC);

— Concatenated data performed about the same as most algorithms (0.58 AUC).

Connectome fusion: pipeline

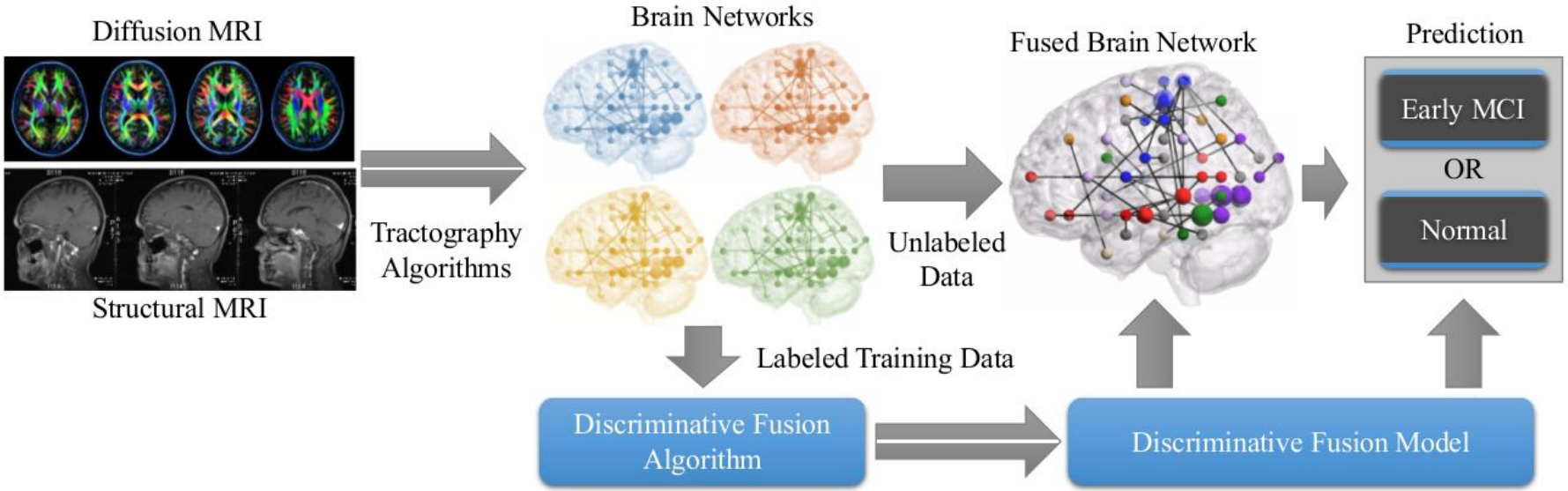


Fig. 2. Overview of our network fusion framework. Multiple types of brain networks are computed by applying different tractography methods to the participants' diffusion MRI data [22]. Different brain networks are combined using a sparse learning method and the optimal convex combination is used for classification. The combination coefficients and the classifiers are simultaneously learned from the training data and cross-validated.

Source: "Discriminative fusion of multiple brain networks for early mild cognitive impairment detection" (Qi Wang, Liang Zhan, Paul M. Thompson, Hiroko H. Dodge, Jiayu Zhou)

Connectome fusion: experiment setting

— Dataset: 124 subjects including 51 normal elderly controls (NCs), 73 individuals diagnosed with early mild cognitive impairment (eMCI);

— Atlas: Harvard Oxford Cortical and Subcortical Probabilistic Atlas (113×113);

— Algorithms: tensor (FACT, RK-2, TL, SL), ODF (FACT, RK2) and probabilistic (PICo, Hough, Probtrackx);

— Data fusion method: convex sum of each of nine networks;

— Classifier: Logistic Regression with L_1-regularization (sparse LR) with loss depending on sum weights;

$$\min_{\mathbf{w},c,\boldsymbol{\tau}} \sum_{i=1}^{N} \ell(\mathbf{w}, c, \boldsymbol{\tau}; \mathbf{x}_i, y_i) + \lambda \|\mathbf{w}\|_1, \qquad \ell(\mathbf{w}, c, \boldsymbol{\tau}; \mathbf{x}_i, y_i) = \log\left(1 + exp\left(-y_i(x_i(\boldsymbol{\tau})^T \mathbf{w} + c)\right)\right)$$

$$\text{s.t.} \quad \sum_{m=1}^{M} \tau_m = 1; \tau_m \geq 0, \forall \tau_m$$

— Cross validation: 10 iterations of 10-fold;

— Baselines: B-CON (concatenated networks) and B-AVG (mean of all networks).

Source: “Discriminative fusion of multiple brain networks for early mild cognitive impairment detection” (Qi Wang, Liang Zhan, Paul M. Thompson, Hiroko H. Dodge, Jiayu Zhou)

Connectome fusion: results

	AUC	Sensitivity	Specificity
DFUSE	**0.89 ± 0.09**	**0.84 ± 0.16**	**0.77 ± 0.07**
B-CON	0.58 ± 0.11	0.54 ± 0.11	0.54 ± 0.10
B-AVG	0.50 ± 0.15	0.56 ± 0.18	0.48 ± 0.07
T-FACT	0.56 ± 0.13	0.66 ± 0.19	0.40 ± 0.09
T-RK2	0.54 ± 0.10	0.58 ± 0.18	0.49 ± 0.06
T-SL	0.59 ± 0.14	0.42 ± 0.20	0.79 ± 0.06
T-TL	0.51 ± 0.13	0.48 ± 0.19	0.51 ± 0.07
O-FACT	0.58 ± 0.12	0.58 ± 0.24	0.43 ± 0.09
O-RK2	0.56 ± 0.13	0.60 ± 0.27	0.47 ± 0.06
PICo	0.58 ± 0.11	0.54 ± 0.16	0.54 ± 0.10
Hough	0.58 ± 0.07	0.42 ± 0.24	0.61 ± 0.08
Probt	0.66 ± 0.10	0.48 ± 0.25	0.69 ± 0.11

Main findings:

— DFUSE algorithm significantly outperformed all other competing methods (p-value < 0.001);

— Probtrackx has the heaviest weight of 0.87 (all elements of τ range from 0 to 1), averaged over 10 iterations;

— The weights of T-TL, O-FACT, O-RK2 are consistently zeros;

— the predictive performance of the feature concatenation (B-CON) does not even perform as well as the best individual brain network. This may be because there are too many features presented to the classifier (over 56k), relative to the number of subjects (samples) available to train it (~110).

Source: “Discriminative fusion of multiple brain networks for early mild cognitive impairment detection” (Qi Wang, Liang Zhan, Paul M. Thompson, Hiroko H. Dodge, Jiayu Zhou)

How does field strength affect connectome weights?

— Connectivity matrices from whole-brain tractography tended to pick up a greater density for some subcortical connections at 7T than at 3T, according to “Magnetic Resonance Field Strength Effects on Diffusion Measures and Brain Connectivity Networks” (Zhang et al., 2012)

— We can expect that 7T data should better represent reality of neural pathways, but further validation is needed. Field itself is not enough. Voxel size, numbers of diffusion gradients, and diffusion weighting schemes on the local diffusion model are also important.

— Paper is rather old, so there may be more up-to-date analysis.

3T vs 7T connectomes: protocols and demographics

MRI machine name				*Siemens TIM Trio 3T*	*Siemens Magnetom 7T*
PAT mode				GRAPPA	GRAPPA
Acceleration factor PE				2	2
Isotropic voxel size (mm)				2.0	2.0
TR/TE (ms)				7800/82	5700/57
FOV (mm)				192×192	256×256
Diffusion weighting, b (sec/mm^2)				1000	1000
Number of diffusion weighted images (DWI)				128	128
Number of non-diffusion weighted reference images (b_0 images)				15	15
Total scan time (seconds)				1138	832
Dataset 1 23 subjects	Age	23.75±2.62	Field strength	3T	7T
	Sex	11 female	Reconstruction method	AC	SOS
Dataset 2 9 subjects	Age	73.95±12.79	Field strength	3T	3T
	Sex	7 female	Reconstruction method	AC	SOS
Dataset 3 5 subjects	Age	78.35±9.39	Field strength	7T	7T
	Sex	5 female	Reconstruction method	SOS	SENSE1

All scan protocols used single-spin echo DTI sequences, to allow for shorter TE times. Other consistently applied sequence parameters included an acquisition of 64 slices, 2-mm isotropic voxels, a b-value of 1000 sec/mm^2, 128 diffusion directions and 15 $b=0$ scans. TE and TR times were set to be the fastest possible allowed by the system. The superior gradient performance of the 7T scanner allowed for significantly shorter TE and TR times than could be achieved at 3T.

DW-MRI, diffusion-weighted magnetic resonance imaging; SOS, sum-of-squares; AC, adaptive recombine; DTI, diffusion tensor imaging; TR, time to repetition; TE, time to echo.

Source: “Magnetic Resonance Field Strength Effects on Diffusion Measures and Brain Connectivity Networks” (Zhang et al., 2012)

3T vs 7T connectomes: methods

— Atlas: Harvard Oxford Cortical and Subcortical probabilistic atlas (113×113);

— Tractography algorithm: ODF-RK2 (all voxels with FA > 0.2);

— Spline filter was applied to each generated fiber. All duplicate and very short fibers (<10 mm) were removed;

— SNR of the non-diffusion-sensitized images (b_0)comparison between 3T and 7T;

— ROI-based DTI-derived measures comparisons (fractional anisotropy, mean diffusivity, axial diffusivity, radial diffusivity) — **I won't elaborate on that today**;

— Network metrics depending on sparsity: Characteristic Path Length, Global Efficiency, Mean Clustering Coefficient, Modularity, degree of small-worldness.

Source: "Magnetic Resonance Field Strength Effects on Diffusion Measures and Brain Connectivity Networks" (Zhang et al., 2012)

3T vs 7T connectomes: SNR differences

TABLE 3. SNR DIFFERENCES IN HEAD-TO-HEAD PROTOCOL COMPARISONS

Dataset 1	3T-AC 4.0829 ± 1.1647	7T-SOS 4.8307 ± 1.5126	Paired T test p $SNR_{3T\text{-}AC} < SNR_{7T\text{-}SOS}$ **0.0158**
Dataset 2	3T-AC 4.0923 ± 1.5521	3T-SOS 4.0954 ± 1.6011	Paired T test p $SNR_{3T\text{-}AC} < SNR_{3T\text{-}SOS}$ 0.5707
Dataset 3	7T-SOS 4.0783 ± 1.1242	7T-SENSE1 4.0172 ± 0.9190	Paired T test p $SNR_{7T\text{-}SOS} < SNR_{7T\text{-}SENSE1}$ 0.6412

We listed the mean SNR for two protocols in each dataset, and the p value was computed from the Student's paired T test. When this p value <0.05 (e.g., 0.0158 for dataset 1), it means the SNR of 7T-SOS is significantly higher than the SNR of 3T-AC in dataset 1; there were no detectable differences between the two protocols in dataset 2 ($p=0.57$) and dataset 3 ($p=0.64$). These results suggest that the field strength is likely to play an important role in boosting the SNR. Mathematically, the different reconstruction methods should give rise to some differences in SNR, but our failure to detect an SNR difference in datasets 2 and 3 suggests that reconstruction methods alone do not explain the observed boost in SNR at the higher field strength. The bold values highlight comparisons that passed the significance threshold.

SNR, signal-to-noise ratio.

Source: "Magnetic Resonance Field Strength Effects on Diffusion Measures and Brain Connectivity Networks" (Zhang et al., 2012)

3T vs 7T connectomes: fiber number and length

TABLE 5. COMPARISON OF WHOLE-BRAIN FIBER TRACTOGRAPHY SUMMARY PARAMETERS BETWEEN SCANNING PROTOCOLS, ACROSS THREE DATASETS

	Dataset 1		
	Fiber number	*Max fiber length*	*Mean fiber length*
3T-AC	34582 ± 2089	170.16 ± 18.11 mm	35.29 ± 2.38 mm
7T-SOS	35618 ± 2216	167.90 ± 15.63 mm	36.86 ± 2.56 mm
Paired T test p value (3T-AC < 7T-SOS)	**0.0333**	0.9349	**0.0411**
	Dataset 2		
	Fiber number	*Max fiber length*	*Mean fiber length*
3T-AC	30736 ± 4080	173.02 ± 20.94 mm	34.57 ± 2.49 mm
3T-SOS	30262 ± 4008	176.08 ± 21.98 mm	34.64 ± 2.58 mm
Paired T test p value (3T-AC < 3T-SOS)	0.5964	0.3831	0.4765
	Dataset 3		
	Fiber number	*Max fiber length*	*Mean fiber length*
7T-SOS	31513 ± 2867	161.19 ± 16.39 mm	31.94 ± 1.96 mm
7T-SENSE1	31907 ± 3416	170.04 ± 20.92 mm	31.93 ± 2.05 mm
Paired T test p value (7T-SOS < 7T-SENSE1)	0.1035	0.0841	0.5465

Results are averaged across the subjects in each dataset. The minimum fiber length was set to 10 mm for each. To correct for multiple comparisons, the Bonferroni corrected significance threshold was set to $p < 0.05/3$. All results are null if properly corrected for multiple comparisons. The bold values highlight comparisons that passed the significance threshold.

Source: "Magnetic Resonance Field Strength Effects on Diffusion Measures and Brain Connectivity Networks" (Zhang et al., 2012)

3T vs 7T connectomes: connectivity

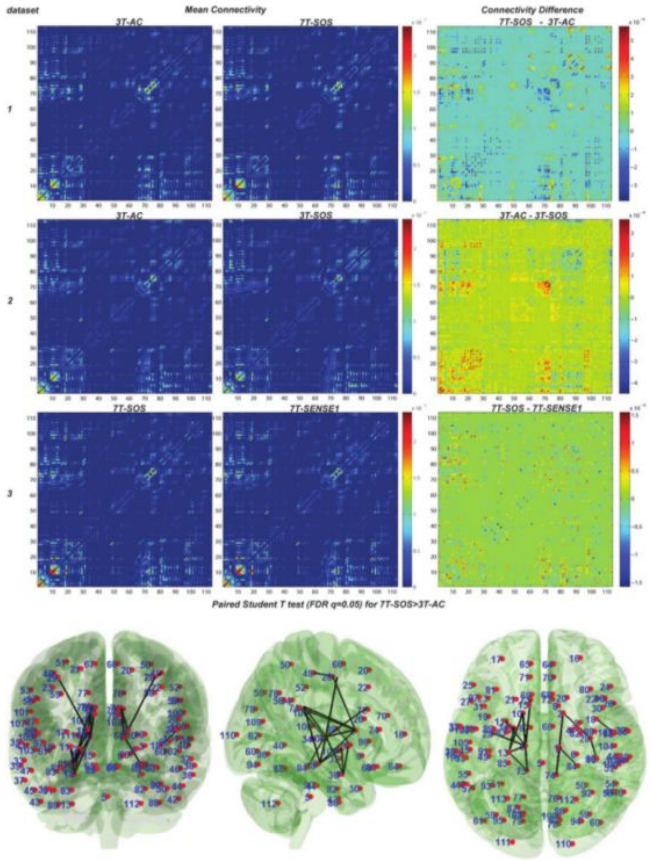
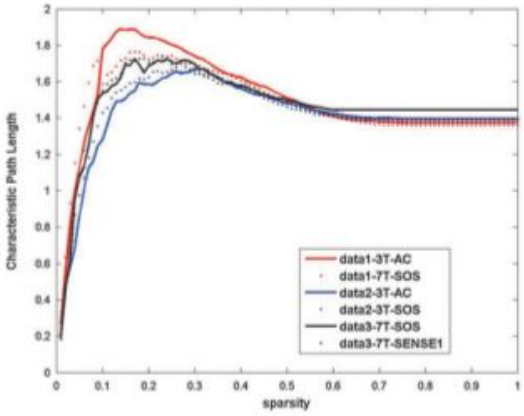
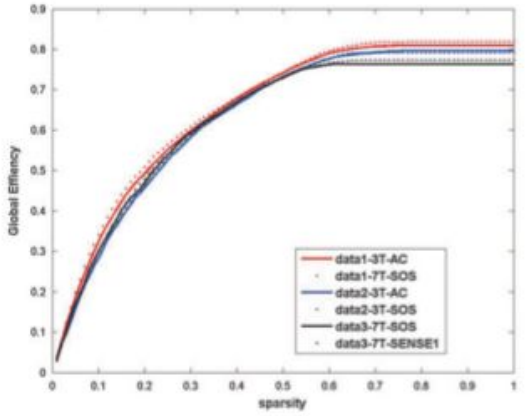


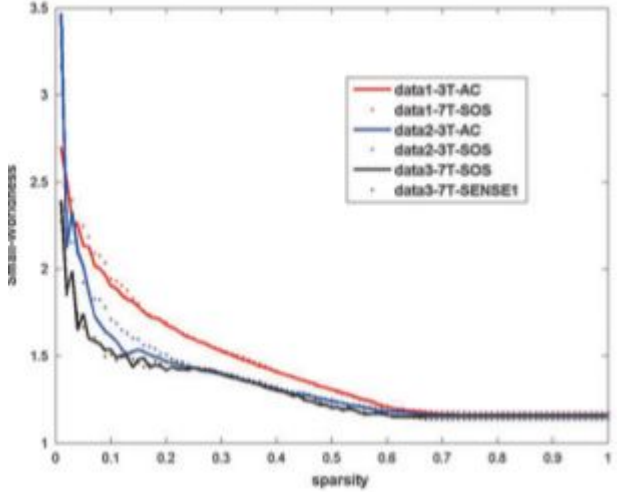
FIG. 3. Differences in measured brain connectivity patterns. The first three rows show the mean connectivity pattern (*first two columns*) and connectivity difference between protocols for the three datasets (in Table 1). Within each row, the exact same *subjects* are scanned—only the scanner or the reconstruction methods differ. All fiber counts are normalized to the whole brain fiber count, so this difference only refers to the proportional representation of connections, which leads to the assignment of weights in the overall network. In general, connectivity patterns are very similar across protocols. In the maps of mean connectivity across all subjects, red colors indicate a *stronger* connection (more fibers detected) and blue colors denote a weaker connection (fewer fibers); in the connectivity difference maps, a red color indicates a positive difference and a blue color represents a negative difference. The last row shows connections that passed false discovery rate (FDR) ($q=0.05$) in paired Student's *t*-tests when comparing 7T-SOS > 3T-AC in dataset 1 (i.e., connection density was higher at 7T-SOS). No connections passed FDR ($q=0.05$) in datasets 2 or 3—the tests of effects of reconstruction methods—or when comparing 3T-AC > 7T-SOS (i.e., where connection density was higher at 3T-AC) in dataset 1. Each red dot in the plot represents one ROI, numbered according to the index in Table 2. The line between two red dots represents the fiber connection between them (in reality, these are curved 3D lines, but a straight line is used for visual clarity). Overall, the higher field strength (7T) enhanced the apparent strength of some subcortical connections, that is, proportionally more fibers were detected in the whole brain tractography.

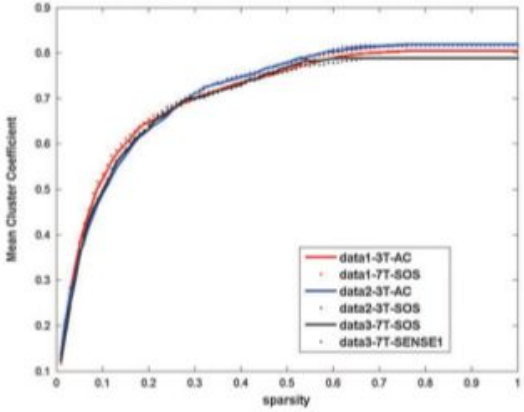
Source: "Magnetic Resonance Field Strength Effects on Diffusion Measures and Brain Connectivity Networks" (Zhang et al., 2012)

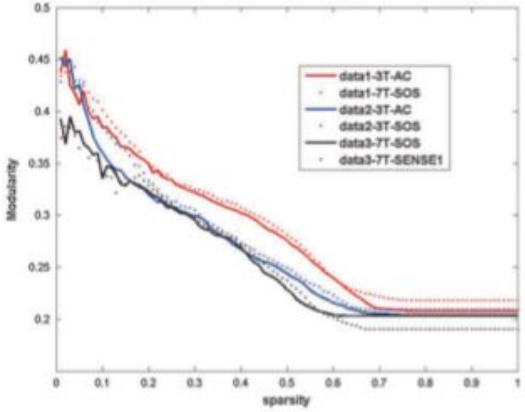
3T vs 7T connectomes: network metrics











Source: “Magnetic Resonance Field Strength Effects on Diffusion Measures and Brain Connectivity Networks” (Zhang et al., 2012)

How does spatial and angular resolution affect connectome weights?

— Spatial and angular resolution affected the computed connectivity for narrower tracts (internal capsule and cerebellum), but also for the corticospinal tract, according to “How do spatial and angular resolution affect brain connectivity maps from diffusion MTI” (Zhang et al.);

— Data resolution affected the apparent role of some key structures in cortical anatomic networks;

— Care is needed when comparing network data across studies, and interpreting apparent disagreements among findings.

Spatial/angular resolution: data used

— Dataset 1 (Mayo). 8 healthy subjects (age: 32.0 years ± 3.9SD; 4 males).

Table 1. Imaging protocols for the Mayo dataset

	Protocol 1 (P1)	Protocol 2 (P2)
Isotropic voxel size (mm)	3.0	2.5
Prescribed matrix	128 x 128	128 x 128
Number of slices	40	48
Number of DWI	48	37
Number of b_0 images	4	4
TR (ms)	7750	9825
b-value (s/mm^2)	1000	1000

— Dataset 2. Healthy male subject (32 years old) was scanned at multiple spatial resolutions (2x2x2, 2.5x2.5x2.5, 3x3x3 and 4x4x4 mm 3) with axial DTI using the following acquisition parameters: TR=8000 ms, TE=83 ms, 128x128 matrix, 64 slices, b-value=1000 s/mm 2 , one baseline (b_0) scan and 12 (low!) gradient directions.

Spatial/angular resolution: methods

— Angular resolution (dataset 1). DWIs subsampled from 48 to 15. Sub-sampling was based on maximizing the total angular distribution energy of the remaining set of k gradients, to optimize the uniformity of the spherical sampling;

— Spatial resolution (dataset 1, P2 protocol). Authors gradually reduced data spatial resolution by downsampling its isotropic voxels of side 2.5 mm to 10 mm (0.1 mm step) with linear interpolation;

— Spatial resolution (dataset 2). Variations across 4 voxel sizes;

— White matter connectivity: 50 ROIs ICBM young adult DTI-81 atlas (there is some criticism of this atlas). Connectivity was then computed from a fast-marching based method

— Cortical connectivity: 70 ROIs standard FreeSurfer Atlas. Tensor-FACT fiber tracking.

Spatial/angular resolution: angular effects

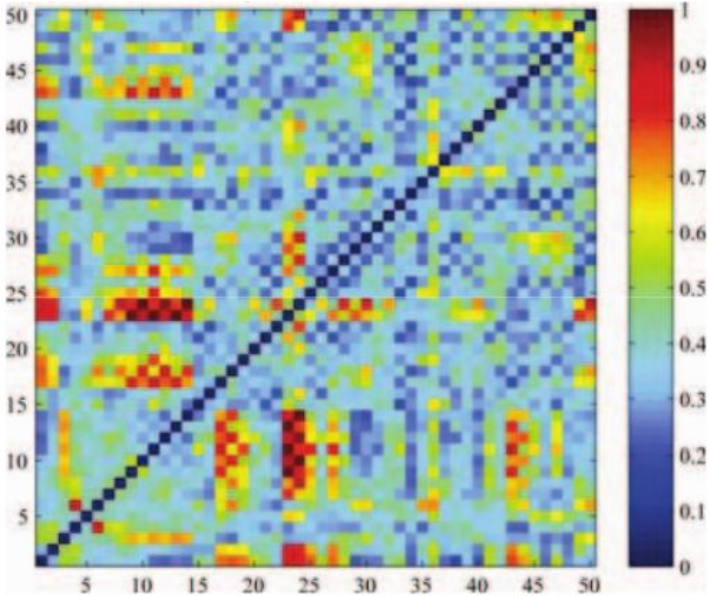


Fig. 2. Angular resolution affects white matter connectivity measures. The names of the ROIs are listed in **Table 2**. In the *red cells,* varying the angular resolution of the scan affected the proportion of fibers apparently connecting the two regions of interest (on the *x* and *y* axes). Data show the standard deviation of the computed proportion of fibers.

— Figure shows the standard deviation of the connectivity matrix elements among the connectivity maps calculated from subset 15 to subset 48; this standard deviation was computed in each of the 8 subjects, and then averaged across all 8, to infer general patterns.

— Some of the thinnest (narrowest) fiber tracts – the cerebellar ICP and SCP, and the internal capsules – were strongly affected by altering the angular resolution. Even some of the major pathways, including the apparent connections of the cortico-spinal tract with the ACR, ALIC and SFO were also quite severely influenced

Spatial/angular resolution: isolating spatial resolution effect

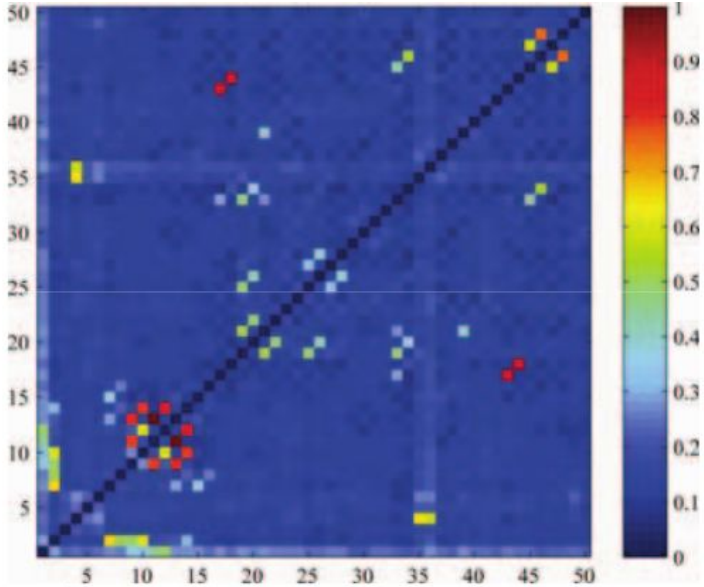


Fig. 3. White matter connectivity measures depend on the spatial resolution of the scans. The names of the ROIs are listed in **Table 2**. Here the thinnest tracts – the internal capsules and cerebellar peduncles – are among those whose connectivity is least stable as the spatial resolution of the DTI scan is changed. The least stable tracts are shown *in red*.

— Figure shows the standard deviation of connectivity matrix elements across connectivity maps calculated at all voxel sizes in the range 2.5-10 mm, averaged across all eight subjects.

— The computed WM connectivity in all tracts and all regions is affected by partial volume effects. Greatest differences were found in the connections of the medial lemniscus, cerebellar peduncles, internal capsules, which are among the thinnest tracts.

Spatial/angular resolution: spatial effects from dataset 2

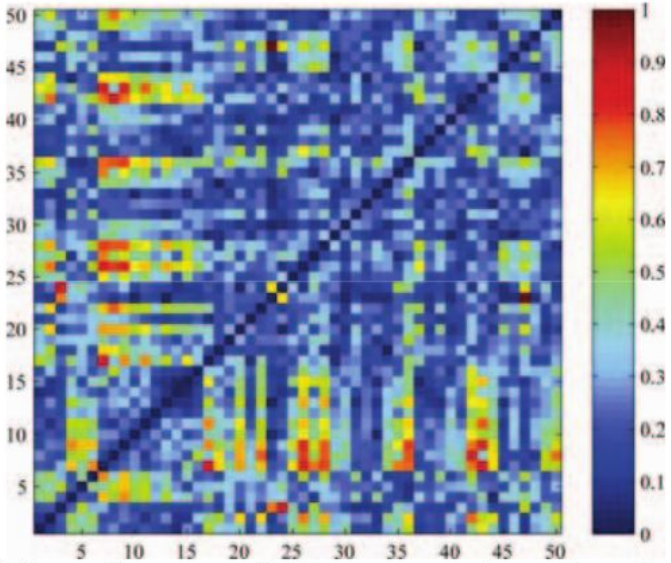


Fig. 4. White matter connectivity measures depend on the SNR and spatial resolution of the scans. *The names of the ROIs are listed in* ***Table 2***. *Red matrix entries* show connections that vary the most as spatial resolution was changed, in one subject scanned at 4 spatial resolutions. Many connections differ with spatial resolution; unlike **Fig. 3**, which downsampled the scan data without SNR varying.

— Figure shows the standard deviation of connectivity among connectivity maps calculated at 4 different isotropic spatial resolutions (2, 2.5, 3 and 4);

— These maps show more differences than those in previous figure, as signal averaging was used to boost the SNR for the scans with smaller voxels.

Spatial/angular resolution: spatial effects from dataset 2

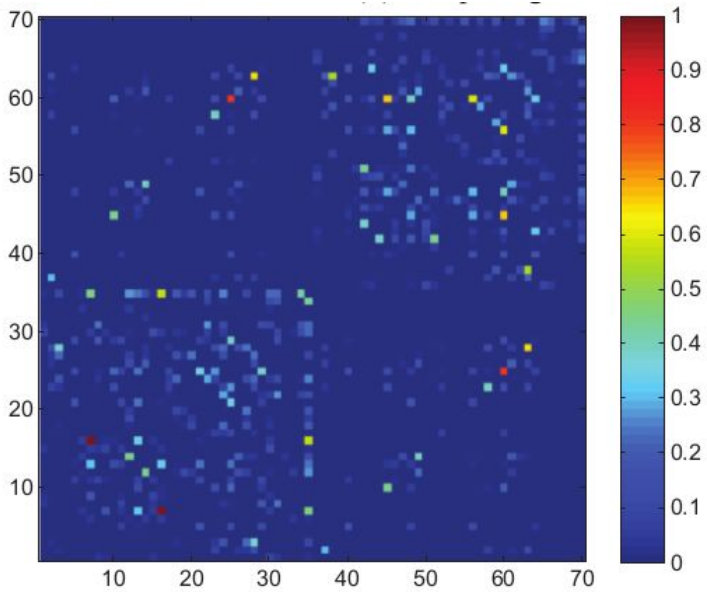


Fig. 5. Cortical connectivity variation within a single subject scanned at 4 spatial resolutions. The names of the ROIs are listed in **Table 3** (1-35, *left hemisphere*; 36-70, *right hemisphere*, e.g., ROIs 2 and 37 are the caudal anterior cingulate in the left and right hemispheres, respectively).

— Figure shows the standard deviation of elements in cortical connectivity matrices for 70 ROIs in the 12-direction dataset, at 4 different spatial resolutions.

— The computed pattern of cortical connectivity heavily depends on the spatial resolution, with less apparent connectivity in scans with large voxels. The cortical connection between parahippocampal and fusiform gyri, and between corresponding structures in the left and right hemispheres were most affected by spatial resolution

Overall conclusion

— Tractography algorithm choice is important;

— There is no best tractography algorithm in terms of predictive power;

— We can combine tractography data to improve predictions;

— Field strength matters, but we can work with current 3T tracks;

— Spatial and angular resolution affects connectome weights.

Thank you!

to.dmitry.petrov@gmail.com

ya.dodonova@mail.ru