bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Local Connectome Fingerprinting Reveals the Uniqueness of Individual White Matter Architecture

Fang-Cheng Yeh[1*], Jean M. Vettel[2,3], Aarti Singh[4], Barnabas Poczos[4], Scott Grafton[3], Kirk I. Erickson[5], Wen-Yih I. Tseng[6], and Timothy D. Verstynen[1*]

[1]Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pennsylvania, USA.

[2]U.S. Army Research Laboratory, Aberdeen Proving Ground, Aberdeen, Maryland, USA.

[3]Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, California, USA.

[4]Department of Machine Learning, Carnegie Mellon University, Pennsylvania, USA.

[5]Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania USA.

[6]Institute of Medical Device and Imaging and Molecular Imaging Center, National Taiwan University, Taipei, Taiwan.

*Correspondence to:

Timothy D. Verstynen Ph.D., Fang-Cheng Yeh Ph.D., M.D.

Department of Psychology and Center for the Neural Basis of Computation,

Carnegie Mellon University

Pittsburgh, PA, USA

Email: timothyv@andrew.cmu.edu, frankyeh@cmu.edu

*Keywords: Connectome, Connectomics, Connectome fingerprint, Diffusion MRI

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

It is generally assumed that the uniqueness of individual identity is reflected in the connective architecture of the human brain. Here we introduce local connectome fingerprinting, a noninvasive method that uses diffusion MRI to characterize white matter bundles as "fingerprints". Using four independently acquired datasets (total n=213), we show that the local connectome fingerprint is highly specific to an individual, achieving 100% accuracy across 17,398 identification tests with an estimated classification error at 10^{-6}. This uniqueness profile is higher than fingerprints derived from local fractional anisotropy or region-to-region connectivity patterns. We further illustrate that local connectome fingerprinting allows for quantifying similarity between genetically-associated individuals, e.g., monozygotic twins (12% connectomic similarity), and neuroplasticity with time, e.g., fingerprint uniqueness decreases 0.02% per day. This approach opens a new door for probing the influence of pathological, genetic, social, or environmental factors on the unique configuration of the human connectome.

*Keywords: Connectome, Connectomics, Connectome fingerprint, Diffusion MRI

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The specific personality characteristics that define an individual are encoded by the unique pattern of connections between the billions of neurons in the brain[1]. This complex wiring system, termed the connectome [2, 3], reflects the necessary connective architecture for the neural dynamics that give rise to nearly all cognitive functions [4, 5]. Substantial attention has been paid to the mapping of region-to-region connectivity [6], aiming to advance the understanding of what makes individuals unique. Recently, it has been shown that functional connections measured by functional MRI (fMRI) could be used as the "functional connectome fingerprint", achieving subject identification with accuracy up to 99% [7, 8]. Since the hemodynamic signals acquired by fMRI are determined by the unique organization of the underlying physical connections between neurons, it is likely that structural connectivity may exhibit higher uniqueness to individuals. However, a robust method to measure such uniqueness in the white matter architecture remains to be developed.

Diffusion MRI (dMRI) is a noninvasive method that characterizes white matter structure by measuring its microscopic diffusion patterns of water molecules [9, 10]. dMRI has had some success in mapping the trajectories of white matter fascicles in the human brain and define the graph structure of region-to-region connectivity[11, 12]; however the efficiency of tractography approaches has recently come into question[13, 14]. Instead of mapping region-to-region connectivity, the concept of the *local* connectome has recently been proposed as an alternative measure of macroscopic white matter pathways that overcomes the limitations of diffusion MRI fiber tracking [13, 14]. The local connectome is defined as the degree of connectivity between adjacent voxels within a white matter fascicle defined by the density of the diffusing water (see Methods). A collection of these density measurements provides a high dimensional feature vector that can describe the unique configuration of the brain connectome.

Here we use this local connectome feature vector as a local connectome fingerprint to represent the unique white matter properties within an individual. Using four independently collected dMRI datasets (n=11, 25, 60, 118, see Methods) with repeat scans at different time intervals (ranging from the same day to a year), we examined whether this local connectome fingerprint is a unique identifier of an individual person. This uniqueness was compared with fingerprints derived from fractional-anisotropy (FA)[15] and conventional region-to-region connectivity

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

methods. Follow-up analysis reveals on how local connectome fingerprints can quantify the similarity between genetically-related individuals as well measure longitudinal changes in within an individual across time.

Results

Uniqueness of local connectome fingerprint

We first illustrate the uniqueness of local fascicular structure within an individual. Figure 1A shows the spin distribution functions (SDFs) [16] estimated from dMRI scans at the mid-sagittal section of the corpus callosum. SDF represents the density of water diffusing at any orientation, and its magnitudes at the axonal directions can be used as the density-based measurements to quantify the local connectome (see Methods). An example of the local connectome quantified at the corpus callosum is illustrated for three subjects in Fig. 1B. Here the anterior and posterior portion of corpus callosum exhibit substantial diversity between these three subjects. A repeat scan several months later reveals a qualitative within-subject consistency. This high individuality appears to be specific to diffusion density estimates, as conventional FA measures calculated from diffusivity do not yield this qualitative between-subject diversity (Fig. 1C). To sample the local density measurements across all major white matter pathways, dMRI data was reconstructed into a standard space, and the fiber directions of a common atlas was used to sample a SDF value for each fiber direction (see Methods and Fig. 2A). This approach yields, for each dMRI scan, a local connectome fingerprint consisting of a high-dimensional feature vector with a total of 513,316 density estimates (Fig. 2B). Fig. 2C shows the fingerprints of the same three subjects in Fig. 1B and the fingerprints from their repeat scans. Consistent with the qualitative measurements in Fig. 1B, each local connectome fingerprint in Fig. 2C shows, at a coarse level, a highly similar pattern for within-subject scans and also high variability across subjects, suggesting that the local connectome fingerprint may exhibit individuality of the human connectome.

To quantify how well the local connectome fingerprint captures unique individual identity we used four independently collected dMRI datasets (n=11, 24, 60, 118) with repeat scans for a subset of the subjects (n=11×3, 24×2, 14×2, 44×2, respectively). Euclidian distance (i.e., root-mean-squared error) was used as a distant estimate between any two fingerprints. For each

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

dataset, we computed within-subject distances (n=33, 24, 14, 44, respectively) and between-subject distances (n=495, 1104, 2687, and 12,997, respectively). Figure 3 shows the within- and between-subject distances of the four datasets. All four datasets show a clear separation between the within-subject and between-subject distance distributions, with no single within-subject distance pairing being as large as any of the between-subject distances. The d-prime sensitivity index also shows high separability between the two distance distributions, with d-prime values of 14.84, 12.80, 7.21, and 8.12, for dataset I, II, III, and IV respectively.

This consistency suggests that the local connectome fingerprint could be used as a unique identifier of an individual subject. To assess this, we used a linear discriminant analysis (LDA) classifier [17] to classify whether two fingerprints came from the same individual using only the distance between fingerprints as the classification feature. For each dataset, the classification error was estimated using leave-one-out cross-validation. Out of a total of 17,398 cross-validation rounds from four datasets (17,283 different-subject and 115 same-subject pairings), there was not a single misclassification. To estimate the true classification error, we modeled the distributions of within-subject and between-subject distances by the generalized extreme value distribution [18], a continuous probabilistic function often used to assess the probability of extreme values (smallest or largest) appearing in independent identically distributed random samples (last row of Fig. 3B). The classification error can be quantified by the probability of a within-subject distance greater than a between-subject distance. Our analysis shows that the classification error was 4.25×10^{-6} for dataset I, 9.97×10^{-7} for dataset II, 5.3×10^{-3} for dataset III, and 5.5×10^{-3} for dataset IV. The larger error in dataset III and IV could be due to their longer scan interval (6 months and one year). Thus, the probability of mistaking two samples of the same subject's local connectome fingerprint as coming from two different individuals is low enough so as to consider the local connectome fingerprint a highly reliable measure of individual subject uniqueness.

Uniqueness exhibited by the corpus callosum

Since the gyral and sucal patterns are also unique to an individual, it is possible that the uniqueness we observed reflects an artifact of the normalization process. Therefore we retested the uniqueness measure within a restricted white matter mask that only covers the median sagittal sections of corpus callosum defined by the JHU white matter atlas. The same analysis procedures were applied, and the result showed that the d-prime values were 5.97, 5.85, 3.78,

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and 4.08, for dataset I, II, III, and IV, respectively. The leave-one-out cross-validation analysis showed that classification error was 0%, 0.089%, 1.26%, and 0.63%, for dataset I, II, III, and IV, respectively. The classification error modeled by the generalized extreme value distribution was 9.13×10^{-4}, 5.6×10^{-3}, 6.9×10^{-3}, and 7.2×10^{-3}, for dataset I, II, III, and IV, respectively. The median sagittal sections of the corpus callosum itself already can achieve more than 99% accuracy in subject identification, suggesting that the high individuality of the local connectome fingerprint is not due to the process of normalizing to the unique gyral and sulcal patterns of each subject.

Comparison with FA-based and global connectivity-based fingerprints

Using the same analysis pipeline as was used for the local connectome fingerprint, we replaced the SDF-based measures with FA values of the corresponding voxels. Our analysis showed that the d-prime values were 4.84, 4.70, 4.56, and 3.60, for dataset I, II, III, and IV, respectively. All values are substantially smaller than those of the local connectome fingerprint. The leave-one-out cross-validation analysis showed that classification error was 0%, 0.18%, 0.22%, and 0.87%. While FA-based fingerprints also have high uniqueness with classification error less than 1%, it is clear from these results that the greatest reliability at characterizing connectomic uniqueness comes from local connectome measures.

We further compared the local connectome fingerprint with region-to-region connectivity estimates from diffusion MRI fiber tracking. The same analysis pipeline as was used for the local connectome fingerprint was used to calculate leave-one-out cross-validation error for the traditional connectivity matrix. The d-prime values for the region-to-region connectivity matrices in dataset I, II, III, and IV were at 3.44, 2.06, 2.41, and 2.25, respectively. The classification error for datasets I, II, III, and IV were 3.6%, 13.65%, 11.81%, and 9.48%, respectively (estimated by leave-one-out cross validation). While the classification accuracy for the traditional connectivity matrices is still quite high and similar to what has previously been observed in resting state functional connectivity estimates [7], it is clear from these results that the greatest reliability at characterizing connectomic uniqueness comes from local connectome measures.

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Similarity among genetically-related individuals and repeated scans

The high uniqueness of the local connectome fingerprint opens the possibility for comparing not only differences but also the similarities between individuals. To this end, we used a publicly available dMRI dataset of 486 subjects from Human Connectome Project (2014, Q3 release), including 49 pairs of monozygotic (MZ) twins, 43 pairs of dizygotic twins (DZ) twins, and 96 pairs of non-twin siblings. While the local connectome fingerprints of MZ twins show generally similar patterns at the coarse level (Fig. 4), there are also substantial individual differences between the twins that can still be observed along the fingerprints. Consistent with these qualitative comparisons, we found that MZ twins have smaller distances between fingerprints, followed by DZ twins, siblings, and unrelated subjects (Fig. 5A). It is noteworthy that all distance distributions have large overlapping regions (Fig. 5B), indicating that the distance between twins or siblings may often fall within the distribution of distances from genetically-unrelated subjects. We further compared the similarity between twins and siblings. The similarity index of two local connectome fingerprints was calculated as a percentage of the mean distance between unrelated subjects (see Methods). On average, MZ twins have a similarity index of 12.51±1.09%, whereas similarity for DZ twins and siblings is 5.14±1.34% and 4.47±0.59%, respectively (Fig. 5C). The difference in similarity index was significant across MZ twins, DZ twins, non-twin siblings, and genetically-unrelated subjects ($F[3,51224] = 93.64$, $p < 0.001$). Post-hoc comparisons using Tukey-Kramer tests showed significant differences between any two groups (all $p < 0.001$), except between DZ twins and non-twin siblings ($p = 0.9348$). We further compared the similarity between genetically related subjects with that of repeat scans. On average, the similarity index of repeated scans was 72.43±0.74% in dataset I, 69.82±1.05% in dataset II, 67.02±2.75% in dataset III, and 65.85±1.39% in dataset IV (Fig. 5C), suggesting a high similarity. Although MZ twins have a higher similarity index compared with DZ twins and siblings, the similarity is not as high as what is observed in repeat scans of the same individuals.

Influence of time interval on the similarity between repeated scans

Finally, we examined how time impacts the uniqueness of local connectome fingerprints. The timeframe between repeat scans varied across the four datasets. The subjects in dataset I (n=11) were scanned three times, with all sessions occurring within 16 days. The subjects in dataset II (n=24) were scanned twice with 1-3 months apart. In dataset III (n=60), 14 of the subjects were

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

scanned twice about six months apart, and 44 of the subjects in dataset IV (n=118) were scanned a second time one year later. Consistent with our hypothesis, the similarity in repeat scans of the same individual was strongest within 16 days (72.43% similarity in dataset I), whereas the similarity was lower at 1~3 months (69.82% dataset II), six months (67.02% dataset III) and one year (65.85% for dataset IV)(Fig. 5C). The similarity index was significantly different across all four datasets ($F[3,111] = 5.26$, $p = 0.002$). Using linear regression, we next modeled the influence of time on connectomic uniqueness (Fig. 5D). As expected, the impact of the interval on similarity index was negative (-0.01989, $p < 0.001$) meaning that within-subject similarity dropped at a rate of 0.02% per day or 7.26% per year. The longer intervals between repeat scans led to reduced similarity index of the same subject, suggesting that local connectome fingerprint may capture structural changes due to life experience.

Discussion

Local white matter architecture is so unique and highly consistent within an individual that it can be used as a reliable personal identifier or fingerprint. This uniqueness is more salient than what is observed when looking at the region-to-region connectivity reported by either dMRI or fMRI, as typically done in human connectomic studies [2, 7, 19]. In comparison, the region-to-region connectivity achieved a classification accuracy around 90~97%. This accuracy is very close to its functional counterpart[7], that was recently reported to have an accuracy of 92-94% in whole brain identification and 98-99% in frontoparietal network. Although both region-to-region connectivity approaches have accuracy greater than 90%, the performance remains substantially lower than the perfect classification in 17,398 leave-one-out rounds and an estimated error of 10^{-6} achieved by local connectome fingerprint.

It is important to point out that the local connectome fingerprint reflects different aspects of underlying white matter architecture than more popular diffusivity-based metrics such as fractional anisotropy, axial diffusivity, and radial diffusivity. Diffusivity quantifies *how fast* water diffuses in tissue [20] and is sensitive to the structural integrity of the underlying fiber bundles [15], such as axonal loss and demyelination [21-24]. By contrast, SDF quantifies *how much* water diffuses along the fiber pathways[16, 25] and is sensitive to density characteristics of white matter such as the compactness of the fiber bundles. As illustrated in our qualitative analysis

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Fig. 1C), while the density characteristics vary substantially among normal populations, the diffusivity measurements do not show obvious differences between subjects. This highlights how diffusivity-based metrics may not be as sensitive to individual uniqueness as density-based measures, suggesting that our local connectome fingerprint is picking up on very specific structural characteristics of underlying white matter systems.

Local connectome fingerprints also allow us to explore various factors that may have an influence on the uniqueness of the connectome. For example, several studies have already shown high heritability in cortical connections [26, 27] and white matter integrity [28-31]; however, high heritability may not necessarily imply that a large portion of the differences or similarity observed in phenotypes are due to genetic factors [32]. Here local connectome fingerprint can provide this complementary information by quantifying the similarity between genetic-related subjects. This is illustrated in our twin analysis: monozygotic twins share only 12.51% similarity in local white matter architecture while repeat scans for the same subject have a much high similarity hovering around 60-70%. A considerable portion of the individuality in local connectome could be driven by environmental factors such as life experience and learning, and monozygotic twins still exhibit high individuality in their connectome.

The fact that uniqueness of the local connectome fingerprint decreases at a rate of 0.02% per day, or 7.26% per year, raises many questions about which factors (genomic, social, environmental, or pathological) sculpt local white matter systems. Of course, white matter integrity also varies with normative development [33-35], a portion of which may be determined genetically. This warrants more longitudinal and genetic analysis to identify specific contributions of genetic and environmental factors on the uniqueness of connectomic structure, with an aim to understand how those factors interact with abnormal brain circuits in neurological and psychiatric disorders.

Methods

Five independently collected dMRI datasets

The first dataset included a total of 11 subjects (9 males and 2 females, age 20~42). Each subject had three diffusion MRI scans within 16 days on a Siemens Trio 3T system at the University of California, Santa Barbara. The diffusion MRI was acquired using a twice-refocused spin-echo

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

EPI sequence. A 257-direction full-sphere grid sampling scheme was used. The maximum b-value was 5000 s/mm^2. TR = 9916 ms, TE = 157 ms, voxel size = 2.4×2.4×2.4 mm, FoV = 231×231 mm.

The second set of data included a total of 24 subjects (8 males and 16 females, age 22 ~ 74). All participants were scanned on a Siemens Tim Trio 3T system at National Taiwan University, and all subjects had their second scan at 1~3 months. The diffusion MRI was also acquired using a twice-refocused spin-echo EPI sequence. The diffusion scheme is a 101-direction half-sphere grid sampling scheme with b-max = 4000 s/mm^2 (b-table available at http://dsi-studio.labsolver.org). TR = 9600 ms, TE = 130 ms, voxel size = 2.5×2.5×2.5 mm.

The third set of data included a total of 60 subjects (30 males and 30 females, age 18 ~ 46). All participants were scanned on a Siemens Verio 3T system at Carnegie Mellon University, and 14 of the 60 subjects had their second scan at 6 months. The diffusion MRI was also acquired using a twice-refocused spin-echo EPI sequence. A 257-direction full-sphere grid sampling scheme was used. The maximum b-value was 5000 s/mm^2. TR = 9916 ms, TE = 157 ms, voxel size = 2.4×2.4×2.4 mm, FoV = 231×231 mm.

The fourth set of diffusion data included a total of 118 subjects (91 males and 27 females, age 22 ~ 55) scanned on a Siemens Verio 3T system at the University of Pittsburgh. 44 of them had another scan after one year. The diffusion images were acquired on a Siemens Verio scanner using a 2D EPI diffusion sequence. TE=96 ms, and TR=11100 ms. A total of 50 diffusion sampling directions were acquired. The b-value was 2000 s/mm^2. The in-plane resolution was 2.4 mm. The slice thickness was 2.4 mm.

The fifth dataset was from the Human Connectome Projects (Q3, 2014) acquired by Washington University in Saint Louis and University of Minnesota. The diffusion MRI data were acquired on a Siemens 3T Skyra scanner using a 2D spin-echo single-shot multiband EPI sequence with a multi-band factor of 3 and monopolar gradient pulse. A total of 486 subjects (195 males and 291 females, age 22 ~ 36) received diffusion scans. The spatial resolution was 1.25 mm isotropic. TR=5500 ms, TE=89.50 ms. The b-values were 1000, 2000, and 3000 s/mm^2. The total number of diffusion sampling directions was 90, 90, and 90 for each of the shells in addition to 6 b0

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

images. The total scanning time was approximately 55 minutes. The scan data included 49 pairs of monozygotic twin, 43 pairs of dizygotic twins, and 96 pairs of siblings.

The procedures for this archival study were approved by the local institutional review board at Carnegie Mellon University.

Local connectome fingerprinting

As shown in Fig. 2A, the diffusion MRI data of each subject were reconstructed in a common stereotaxic space using q-space diffeomorphic reconstruction (QSDR)[36], a white matter based nonlinear registration approach that directly reconstructed diffusion information in a standard space:

$$\psi(\hat{\mathbf{u}}) = \left|J_{\varphi}(\mathbf{r})\right| Z_0 \sum_i \mathrm{W}_i(\varphi(\mathbf{r}))\mathrm{sinc}\left(\sigma\sqrt{6Db_i} < \hat{\mathbf{g}}_i, \frac{J_{\varphi}(\mathbf{r})\hat{\mathbf{u}}}{\left\|J_{\varphi}(\mathbf{r})\hat{\mathbf{u}}\right\|} >\right) \qquad (1)$$

$\psi(\hat{\mathbf{u}})$ is a spin distribution function (SDF)[16] in the standard space, defined as the density of diffusing spins that have displacement oriented at direction $\hat{\mathbf{u}}$. φ is a function that maps a coordinate **r** from the standard space to the subject's space, whereas J_{φ} is the Jacobian matrix of φ, and $\left|J_{\varphi}\right|$ is the Jacobian determinant. W_i is the diffusion signals acquired by a b-value of b_i with diffusion sensitization gradient oriented at $\hat{\mathbf{g}}_i$. σ is the diffusion sampling ratio controlling the displacement range of the diffusing spins sampled by the SDFs. Lower values allow for quantifying more from restricted diffusion. D is the diffusivity of free water diffusion, and Z_0 is the constant estimated by the diffusion signals of free water diffusion in the brain ventricle[36]. A σ of 1.25 was used to calculate the SDFs, and 1 mm resolution was assigned to the output resolution of the QSDR reconstruction.

A common axonal directions atlas, derived from the HCP dataset (this HCP-488 atlas is freely available at http://dsi-studio.labsolver.org), was used to provide the sampling direction $\hat{\mathbf{u}}$ to sample the magnitude of SDFs along axonal directions in the cerebral white matter. Gray matter was excluded using the ICBM-152 white matter mask (MacConnel Brain Imaging Centre, McGill University, Canada). The cerebellum was also excluded due to different slice coverage in

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cerebellum across subjects. Since each voxel in the cerebral white matter may have more than one axonal direction, multiple measurements can be extracted from the SDF of the same voxel. The density measurements were sampled by the left-posterior-superior voxel order and compiled into a sequence of scalar values (Fig. 2B). Since the density measurement has arbitrary units, the local connectome fingerprint was scaled to make the variance equal to 1. The computation was conducted using DSI Studio (http://dsi-studio.labslover.org), an open-source diffusion MRI analysis tool for connectome analysis. The source code and the local connectome fingerprint data are publicly available on the same website.

Estimation of classification error

For each dMRI dataset, the root-mean-squared error between any two connectome fingerprints was calculated to obtain a matrix of paired-wise distance. The calculated distance was used as the feature to classify whether two connectome fingerprints are from the same or different person. The default linear discriminant analysis (LDA) classifier provided in MATLAB (MathWorks, Natick, MA) was used, and for each dataset, the classification error was estimated using leave-one-out cross-validation. We also used a modeling method to calculate the classification error if the leave-one-out cross- validation did not yield any classification error. The histograms of the within-subject and between-subject distances were fitted by the generalized extreme value distribution using the maximum likelihood estimator (gevfit) provided in MATLAB. To consider the non-negativity of the distribution, the estimated k parameter of the generalized extreme value distribution was set to be greater than 0. The classification error was estimated by the probability of a within-subject distance greater than a between-subject distance estimated using the generalized extreme value distribution.

Comparison with traditional connectivity matrix

To compare local connectome fingerprint with region-to-region connectivity matrix, deterministic fiber tracking[37] was applied using a 100,000 uniform white matter seeding points, a maximum turning angle of 60 degrees, and a default anisotropy threshold determined using Otsu's threshold[38]. The cortical regions were defined through a nonlinear registration between the subject anisotropy map and the HCP-488 anisotropy map in DSI Studio and parcellated using the Automated Anatomical Labeling (AAL) atlas. The matrix entries were quantified by the

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

number of tracks ending in each of the region pairs. The root-mean-squared error can also be calculated from any two connectivity matrices. The classification error was also estimated and compared with local connectome fingerprint.

Similarity index

The similarity index between two local connectome fingerprints was calculated by $100\%\times(1-d_1/d_0)$, where d_1 was the distance between two fingerprints, and d_0 was the expected value of the distances between unrelated subjects scanned by the same imaging protocol. The similarity between MZ twins, DZ twins, non-twin siblings, and repeated scans was calculated and compared. To further study the similarity between repeat scans, the similarity indices were regressed against their scanning time intervals to study the effect of time interval on the local connectome fingerprints.

Acknowledgments

The research was sponsored by the Army Research Laboratory and accomplished under Cooperative Agreement Number W911NF-10-2-0022. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. This research was supported by and NSF BIGDATA (1247658). Part of the data used this study were from the Human Connectome Project, WU-Minn Consortium (1U54MH091657). This research was supported in part by the Ruentex Group and the Ministry of Economic Affairs, Taiwan (101-EC-17-A-19-S1-175). This research was supported in part by National Institutes of Health (R01 DK095172).

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Seung, S. Connectome: How The Brain's Wiring Makes Us Who We Are. (Houghton Mifflin Harcourt Publishing Company, New York, 2012).
2. Hagmann, P. et al. Mapping the structural core of human cerebral cortex. *PLoS Biol* **6**, e159 (2008).
3. Sporns, O., Tononi, G. & Kotter, R. The human connectome: A structural description of the human brain. *PLoS Comput Biol* **1**, e42 (2005).
4. Bota, M., Sporns, O. & Swanson, L.W. Architecture of the cerebral cortical association connectome underlying cognition. *Proc Natl Acad Sci U S A* **112**, E2093-2101 (2015).
5. Jarrell, T.A. et al. The connectome of a decision-making neural network. *Science* **337**, 437-444 (2012).
6. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* **10**, 186-198 (2009).
7. Finn, E.S. et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci* (2015).
8. Miranda-Dominguez, O. et al. Connectotyping: model based fingerprinting of the functional connectome. *PLoS One* **9**, e111048 (2014).
9. Le Bihan, D. & Johansen-Berg, H. Diffusion MRI at 25: exploring brain tissue structure and function. *Neuroimage* **61**, 324-341 (2012).
10. Jbabdi, S., Sotiropoulos, S.N., Haber, S.N., Van Essen, D.C. & Behrens, T.E. Measuring macroscopic brain connections in vivo. *Nat Neurosci* **18**, 1546-1555 (2015).
11. Hagmann, P. et al. White matter maturation reshapes structural connectivity in the late developing human brain. *Proc Natl Acad Sci U S A* **107**, 19067-19072 (2010).
12. Sporns, O. Contributions and challenges for network models in cognitive neuroscience. *Nat Neurosci* **17**, 652-660 (2014).
13. Reveley, C. et al. Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion MR tractography. *Proc Natl Acad Sci U S A* **112**, E2820-2828 (2015).
14. Thomas, C. et al. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proc Natl Acad Sci U S A* **111**, 16574-16579 (2014).
15. Pierpaoli, C. & Basser, P.J. Toward a quantitative assessment of diffusion anisotropy. *Magn Reson Med* **36**, 893-906 (1996).
16. Yeh, F.C., Wedeen, V.J. & Tseng, W.Y. Generalized q-sampling imaging. *IEEE Trans Med Imaging* **29**, 1626-1635 (2010).
17. Krzanowski, W.J. Principles of Multivariate Analysis: A User's Perspective. (New York: Oxford University Press, 1988).
18. Jenkinson, A.F. The Frequency Distribution of the Annual Maximum (or Minimum) of Meteorological Elements. *Quarterly Journal of the Royal Meteorological Society,* **81**, 158-171 (1955).
19. Hagmann, P. et al. Mapping human whole-brain structural networks with diffusion MRI. *PLoS One* **2**, e597 (2007).
20. Le Bihan, D. et al. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology* **161**, 401-407 (1986).
21. Budde, M.D. et al. Toward accurate diagnosis of white matter pathology using diffusion tensor imaging. *Magn Reson Med* **57**, 688-695 (2007).
22. Sun, S.W. et al. Noninvasive detection of cuprizone induced axonal damage and demyelination in the mouse corpus callosum. *Magn Reson Med* **55**, 302-308 (2006).
23. Song, S.K. et al. Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. *Neuroimage* **17**, 1429-1436 (2002).
24. Song, S.K. et al. Demyelination increases radial diffusivity in corpus callosum of mouse brain. *Neuroimage* **26**, 132-140 (2005).
25. Callaghan, P.T. Principles of Nuclear Magnetic Resonance Microscopy. (Oxford University Press, 1994).
26. Shen, K.K. et al. Investigating brain connectivity heritability in a twin study using diffusion imaging data. *Neuroimage* **100**, 628-641 (2014).
27. Chiang, M.C. et al. Gene network effects on brain microstructure and intellectual performance identified in 472 twins. *J Neurosci* **32**, 8732-8745 (2012).
28. Kochunov, P. et al. Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data. *Neuroimage* **111**, 300-311 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

29. Kochunov, P. et al. Genetics of microstructure of cerebral white matter using diffusion tensor imaging. *Neuroimage* **53**, 1109-1116 (2010).
30. Chiang, M.C. et al. Genetics of brain fiber architecture and intellectual performance. *J Neurosci* **29**, 2212-2224 (2009).
31. Sinclair, B. et al. Heritability of the network architecture of intrinsic brain functional connectivity. *Neuroimage* **121**, 243-252 (2015).
32. Gray, J.R. & Thompson, P.M. Neurobiology of intelligence: science and ethics. *Nat Rev Neurosci* **5**, 471-482 (2004).
33. Craik, F.I. & Bialystok, E. Cognition through the lifespan: mechanisms of change. *Trends Cogn Sci* **10**, 131-138 (2006).
34. Tardif, C.L. et al. Advanced MRI techniques to improve our understanding of experience-induced neuroplasticity. *Neuroimage* (2015).
35. Simmonds, D.J., Hallquist, M.N., Asato, M. & Luna, B. Developmental stages and sex differences of white matter and behavioral development through adolescence: a longitudinal diffusion tensor imaging (DTI) study. *Neuroimage* **92**, 356-368 (2014).
36. Yeh, F.C. & Tseng, W.Y. NTU-90: a high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction. *Neuroimage* **58**, 91-99 (2011).
37. Yeh, F.C., Verstynen, T.D., Wang, Y., Fernandez-Miranda, J.C. & Tseng, W.Y. Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS ONE* **8**, e80713 (2013).
38. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber* **9**, 62-66 (1979).

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

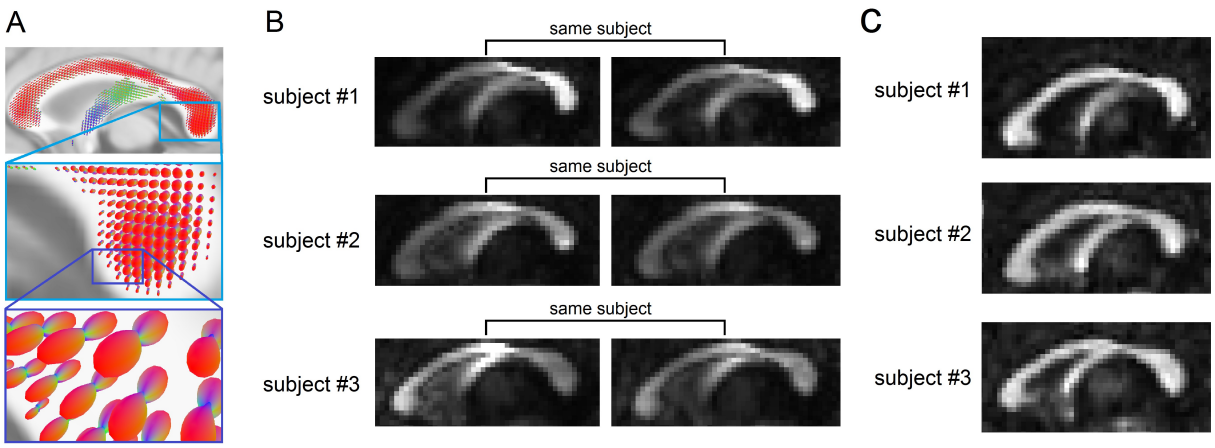

Fig. 1

The uniqueness of local connectome structure revealed by the density of diffusing water. (A) The spin distribution function (SDF) calculated from diffusion MRI quantifies the density of diffusing water along axonal fiber bundles. The magnitudes of SDF at axonal directions provide density-based measurements to characterize axonal fiber bundles. (B) The density measurements obtained from SDF show individuality between-subjects #1, #2, and #3 (intensity scaled between [0 0.8]). The density of diffusing water varies substantially across different portions of the corpus callosum. The repeat measurements after 238 (subject #1), 191 (subject #2), and 198 (subject #3) days still present a consistent pattern that captures individual variability. (C) In contrast to the SDF shown in (B), the fractional anisotropy derived from diffusivity shows no obvious individuality between the same subjects #1, #2, and #3 (intensity also scaled between [0 0.8]). This is due to the fact that diffusivity, which quantifies how *fast* water diffuses, does not vary a lot in normal axonal bundles.

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

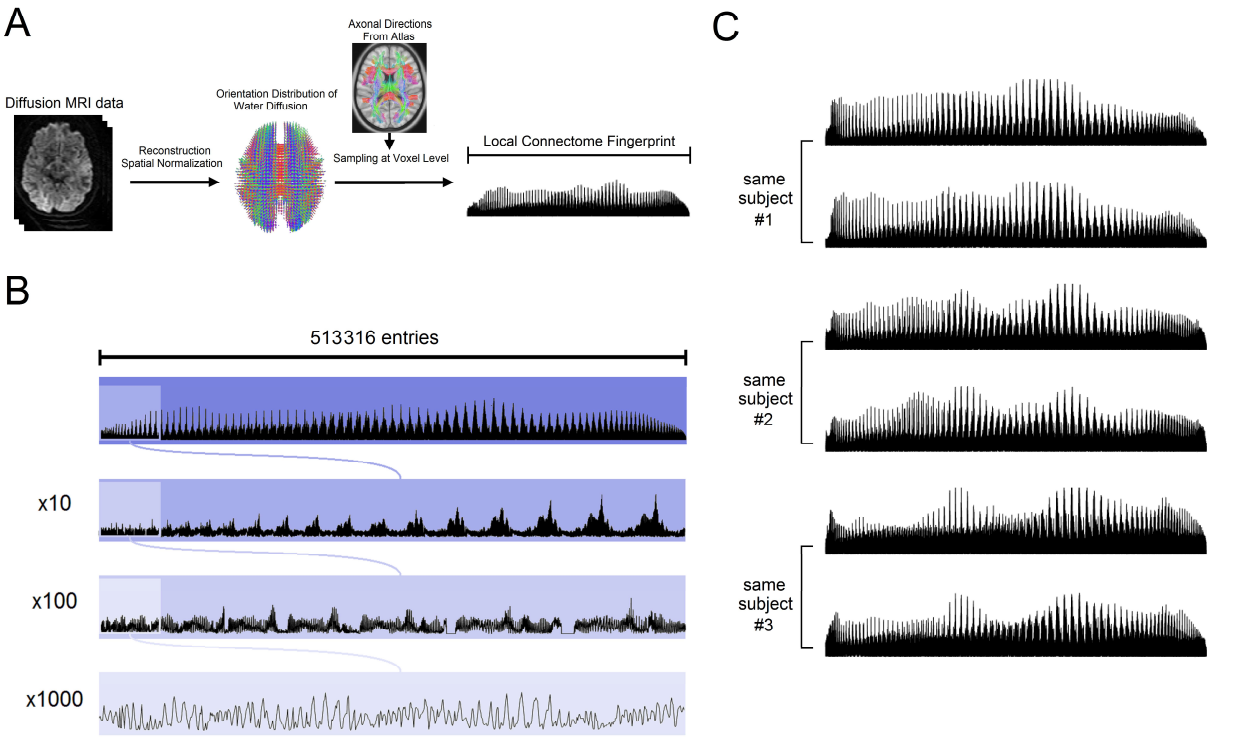

Fig. 2

Local connectome fingerprinting. (A) Local connectome fingerprinting is conducted by first reconstructing diffusion MRI data into a standard space to calculate the spin distribution functions (SDFs). A common fiber direction atlas is then used to sample the density of diffusing water along the principle directions in the cerebral white matter. The sampled measurements are compiled in a left-posterior-superior order to form a sequence of characteristic values as the local connectome fingerprint. (B) One local connectome fingerprint is shown in different zoom-in resolutions. A local connectome fingerprint has a total of 513,316 entries of scalar values presenting a unique pattern that can characterize the individuality of the human brain connections. (C) The local connectome fingerprint of subject #1, #2, and #3 and their repeat measurements (lower row) after 238, 191, and 198 days, respectively. At a coarse level, the local connectome fingerprint differs substantially between three subjects, whereas those from the repeat scans show a remarkably identical pattern, indicating the uniqueness and reproducibility of the local connectome fingerprint.

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

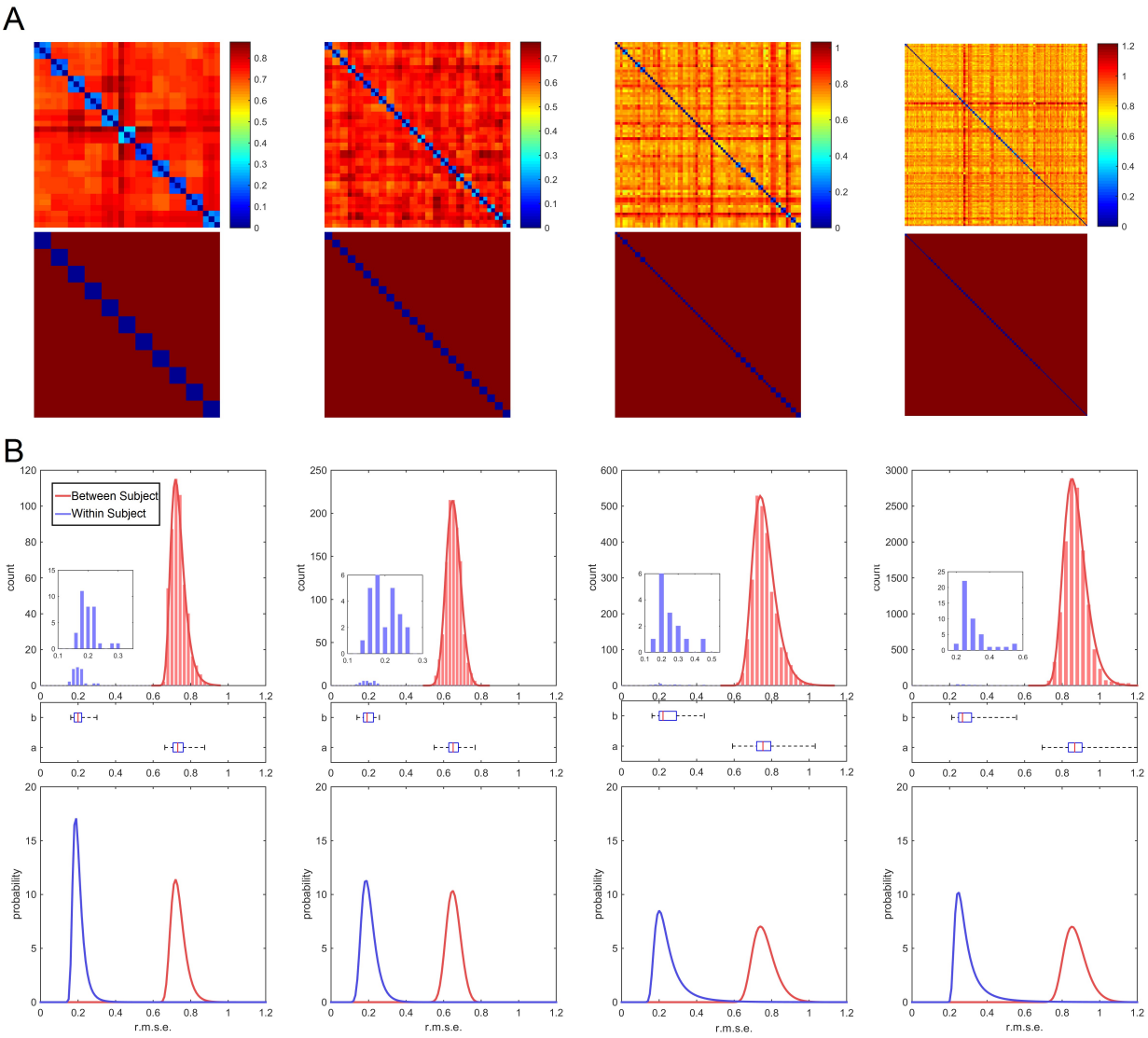

Fig. 3

Within-subject versus between-subject distance in the local connectome fingerprints calculated from four datasets. (A) The first row shows the matrix of pair-wise distance between any two local connectome fingerprints is calculated for dataset I, II, III, and IV (column 1, 2, 3, and 4). The second row shows the location of the within-subject (blue) and between-subject distance (red). The distance matrix shows substantially high between-subject distance and low within-subject distance. (B) The histograms of within-subject (blue) and between-subject (red) distance in the connectome fingerprints calculated from the dataset I, II, III, and IV (column 1, 2, 3, and 4). The first row shows the histograms, and the second row shows the box plot of their quartiles. In these four datasets, within-subject (blue) and between-subject (red) distances have perfect separation. In the last row, the histograms are fitted with generalized extreme value distribution

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(also shown by solid curves in the second row) to estimate the classification error of the connectome fingerprint. The estimated classification error was 4.25×10^{-6}, 9.97×10^{-7}, 5.3×10^{-3}, and 5.5×10^{-3} for dataset I, II, III, and IV, respectively. The larger error in dataset III and IV could be due to their longer scanning interval (6 months and one year).

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

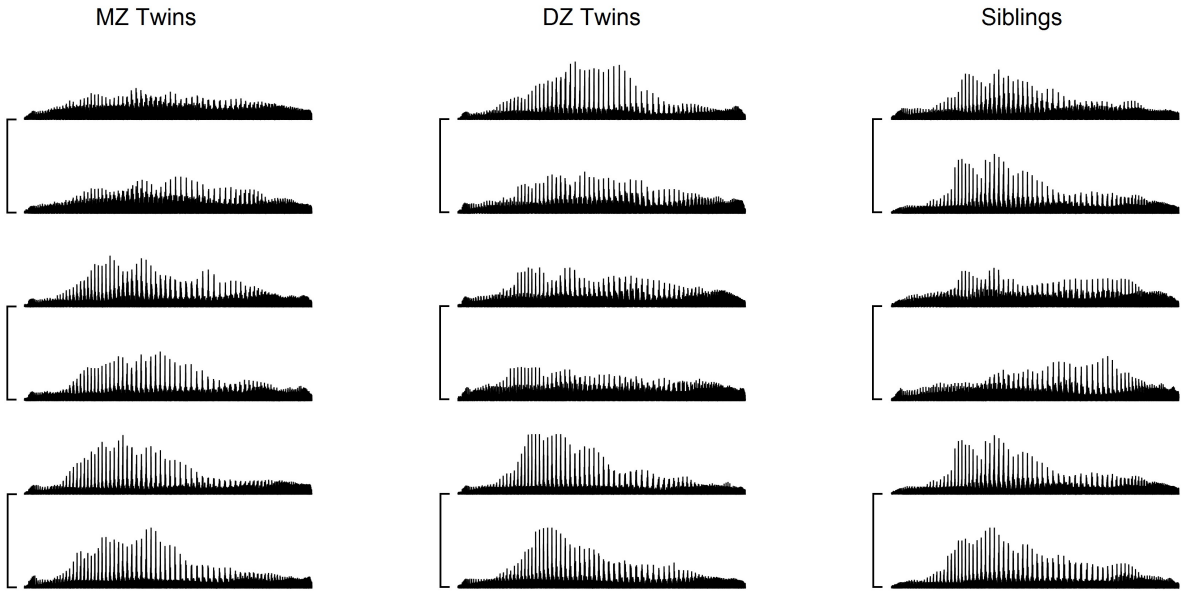

Fig. 4

The local connectome fingerprints of monozygotic (MZ) twins, dizygotic (DZ) twins, and non-twin siblings. 3 pairs of connectome fingerprints are shown for each population group, and each pair is annotated by a connecting line. The connectome fingerprints between MZ twins show the grossly similar patterns though between-subject difference can still be observed. DZ twins and siblings also have a similar pattern, but the between-subject difference becomes more prominent.

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

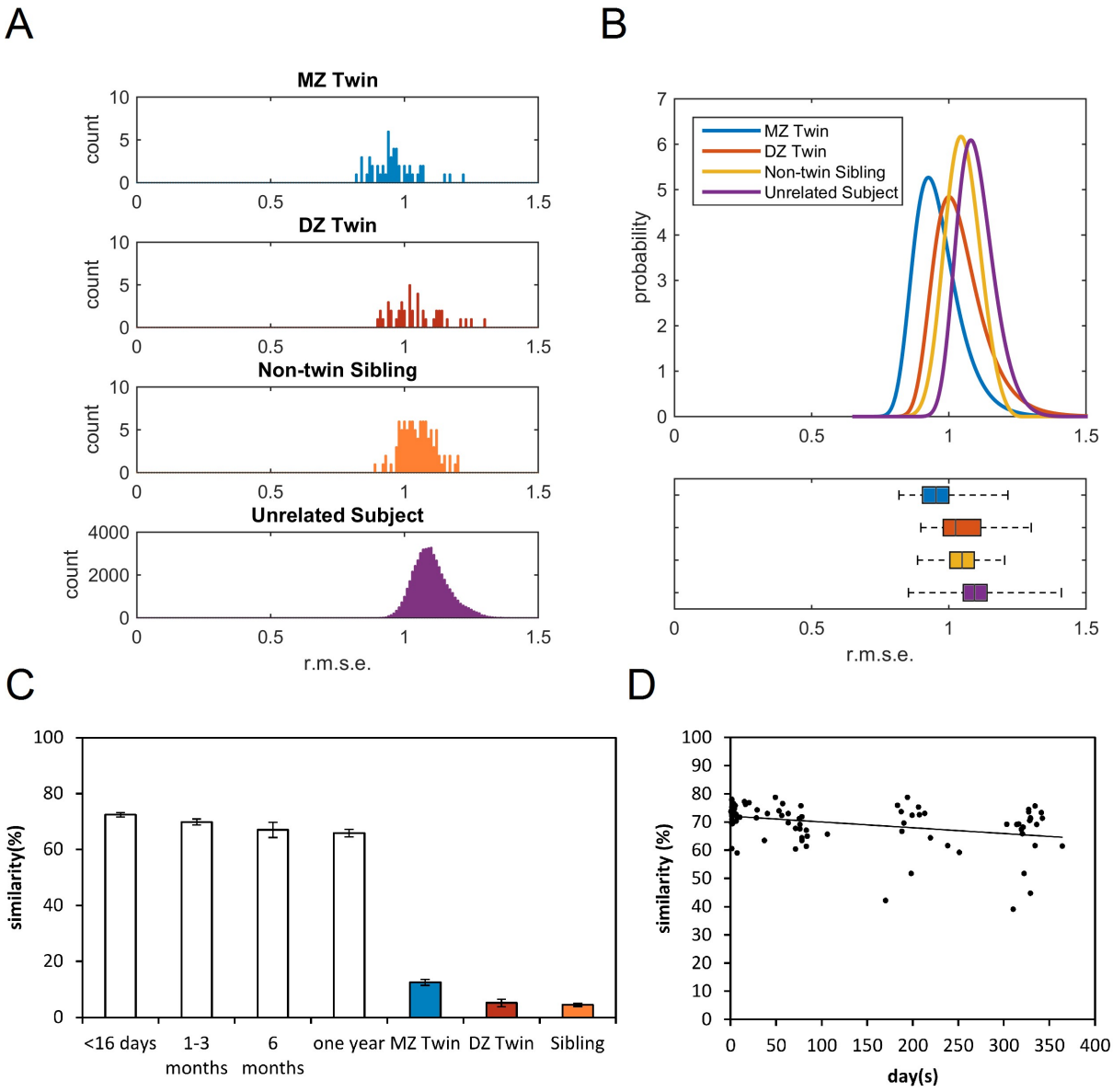

Fig. 5

The similarity between genetically related subjects versus similarity between repeat scans. (A) The histograms show the distribution of the distance between monozygotic (MZ) twins, dizygotic (DZ) twins, non-twin siblings, and genetically unrelated subjects calculated from their local connectome fingerprints. In average, MZ twins have the lowest distance between each twin pair, followed by DZ twins and siblings. However, the distance still ranges around 1, and there are unrelated subjects whose distance is less than several MZ twins. (B) The upper figure shows the histograms of the distance fitted with generalized extreme value distribution, whereas lower figure shows the box plot of the distribution to facilitate comparison. The four distributions are

bioRxiv preprint doi: https://doi.org/10.1101/043778; this version posted March 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

mostly overlapped, indicating that twins and siblings still have high individuality similar to genetically-unrelated subjects. (C) The similarity between genetically related subjects (blue, red, orange bars) is substantially lower than the similarity between repeated scans (white bars), suggesting that genetically-related subjects still have high uniqueness in their local connectome fingerprints. (D) The scatter plot shows the similarity of repeated scans with respect to the scan interval. The longer intervals between repeat scans result in lower similarity index, suggesting that local connectome fingerprint may capture structural changes due to life experience.