Local Connectome Phenotypes Predict Social, Health, and Cognitive Factors

Michael A. Powell
Department of Mathematical Sciences, United States Military Academy, West Point, NY

Javier O. Garcia
U.S. Army Research Laboratory, Aberdeen Proving Ground, MD
Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

Fang-Cheng Yeh
Department of Neurological Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA
Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA

Jean M. Vettel
U.S. Army Research Laboratory, Aberdeen Proving Ground, MD
Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA
Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

Timothy Verstynen
Department of Psychology & Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA

Corresponding Author:
Name: Timothy Verstynen
Address: Department of Psychology, 342C Baker Hall, Carnegie Mellon University, Pittsburgh, PA 15213
Telephone: 412-533-2961
Fax:

Short Title: Local Connectome Phenotypes

Keywords: local connectome, white matter, individual differences, behavior prediction, structural connectivity

Acknowledgements: 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.

Abstract

The unique architecture of the human connectome is defined initially by genetics and subsequently sculpted over time with experience. Thus similarities in predisposition and experience that lead to similarities in social, biological, and cognitive attributes should also be reflected in the local architecture of white matter fascicles. Here we employ a method known as local connectome fingerprinting that uses diffusion MRI to measure the fiber-wise characteristics of macroscopic white matter pathways throughout the brain. This fingerprinting approach was applied to a large sample (N=841) of subjects from the Human Connectome Project, revealing a reliable degree of between-subject correlation in the local connectome fingerprints, with a relatively complex, low-dimensional substructure. Using a cross-validated, high-dimensional regression analysis approach, we derived local connectome phenotype (LCP) maps that could reliably predict 14 out of 36 subject attributes measured, including a large set of health and cognitive measures. These LCP maps were highly specific to the attribute being predicted but also sensitive to correlations between attributes. Collectively, these results indicate the sensitivity of the local connectome to predict both individualized and shared structural variability between subjects related to genetic and experiential factors.

Author Summary

The local connectome is the pattern of fiber systems (i.e., number of fibers, orientation, and size) within a voxel, and reflects the characteristics of white matter fascicles distributed throughout the brain. Here we show how variability in the local connectome is correlated in a principled way across individuals. This inter-subject correlation is reliable enough that unique phenotype maps can be learned to predict between-subject variability in a range of social, health, and cognitive attributes. This work shows, for the first time, how the local connectome has both the sensitivity and specificity to be used as a phenotypic marker for subject-specific attributes.

Introduction

The unique pattern of connections among the billions of neurons in the brain is termed the connectome (Sporns, Tononi, & Kotter, 2005) this pattern encapsulates a fundamental constraint on neural computation and cognition (Gu et al., 2015; Thivierge & Marcus, 2007). This connective architecture is initially structured by genetics and then sculpted by experience over time (Kochunov, Fu, et al., 2016; Kochunov, Thompson, et al., 2016; F.-C. Yeh, Vettel, et al., 2016). Recent advancements in neuroimaging techniques, particularly diffusion MRI (dMRI), have opened the door to mapping the macroscopic-level properties of the structural connectome in vivo (see Le Bihan & Johansen-Berg, 2012). As a result, a growing body of research has focused on quantifying how variability in structural connectivity associates with individual differences in functional properties of brain networks (Muldoon et al., 2016; Passingham, Stephan, & Kötter, 2002), as well as associating with differences in social (Gianaros, Marsland, Sheu, Erickson, & Verstynen, 2013; Molesworth, Sheu, Cohen, Gianaros, & Verstynen, 2015), biological (Arfanakis et al., 2013; Miralbell et al., 2012; Verstynen et al., 2013), and cognitive (Muraskin et al., 2016; Verstynen, 2014; Ystad et al., 2011) attributes.

DMRI works by measuring the microscopic diffusion pattern of water trapped in white matter, allowing for a full characterization of its characteristics, such as axonal fiber direction and integrity (for review see Jbabdi, Sotiropoulos, Haber, Van Essen, & Behrens, 2015; Le Bihan & Johansen-Berg, 2012). Previous studies have used dMRI to map the global properties of the macroscopic connectome by determining end-to-end connectivity between brain regions (P Hagmann et al., 2010; Patric Hagmann et al., 2008, 2010; Sporns, 2014). The resulting connectivity estimates can then be summarized, often using graph theoretic techniques that are then associated with variability across individuals (Bullmore & Sporns, 2009; Rubinov & Sporns, 2010). While dMRI acquisition and reconstruction approaches have improved substantially in recent years (Fan et al., 2016; D C Van Essen et al., 2012), the reliability and validity of many popular fiber tractography algorithms have come into question (Daducci, Dal Palú, Descoteaux,

& Thiran, 2016; Reveley et al., 2015; Thomas et al., 2014). As a result, the reliability of subsequent inter-region connectivity estimates may be negatively impacted.

Instead of mapping end-to-end connectivity between regions, we recently introduced the concept of the local connectome as an alternative measure of structural connectivity that doesn't rely on fiber tracking (F.-C. Yeh, Badre, & Verstynen, 2016). The local connectome is defined as the pattern of fiber systems (i.e., number of fibers, orientation, and size) within a voxel, as well as immediate connectivity between adjacent voxels, and can be quantified by measuring the fiber-wise density of microscopic water diffusion within a voxel. This voxel-wise measure shares many similarities with the concept a "fixel" proposed by others (Raffelt et al., 2015). The complete collection of these multi-fiber diffusion density measurements within all white matter voxels, termed the local connectome fingerprint, provides a high-dimensional feature vector that can describe the unique configuration of the structural connectome (F.-C. Yeh, Vettel, et al., 2016). In this way, the local connectome fingerprint provides a diffusion-informed measure along the fascicles that supports inter-regional communication, rather than determining the start and end positions of a particular fiber bundle.

We recently showed that the local connectome fingerprint is highly specific to an individual, affording near-perfect accuracy on within-versus-between subject classification tests among hundreds of participants (F.-C. Yeh, Badre, et al., 2016). Importantly, this demonstrated that a large portion of an individual's local connectome is driven by experience. Whole-fingerprint distance tests revealed only a 12.51% similarity between monozygotic twins, relative to almost no similarity between genetically unrelated individuals. In addition, within-subject uniqueness showed substantial plasticity, changing at a rate of approximately 12.79% every 100 days (F.-C. Yeh, Vettel, et al., 2016). Thus the unique architecture of the local connectome appears to be initially defined by genetics and then subsequently sculpted over time with experience.

The plasticity of the local white matter architecture suggests that it is important to consider how whole-fingerprint uniqueness may mask more subtle similarities arising from common experiences. If experience, including common social or environmental factors, is a major force impacting the structural connectome, then common experiences between individuals may also lead to increased similarity in their local connectomes. In addition, since the white matter is a fundamental constraint on cognition, similarities in local connectomes are expected to associate with similarities in cognitive function. Thus, we hypothesized that shared variability in certain social, biological, or cognitive attributes can be predicted from the local connectome fingerprints.

To test this, we reconstructed multi-shell dMRI data from the Human Connectome Project (HCP) to produce individual local connectome fingerprints from 841 subjects. A set of 36 subject-level attributes was used for predictive modeling, including many social, biological, and cognitive factors. A model between each fiber in the local connectome fingerprint and a target attribute was learned using a cross-validated, sparse version of principal component regression. The predictive utility of each attribute map, termed a local connectome phenotype (LCP), was evaluated by predicting a given attribute in an independent data set. Our results show that specific characteristics of the local connectome are sensitive to shared variability across individuals, as well as being highly reliable within an individual (F.-C. Yeh, Vettel, et al., 2016), confirming its utility for understanding how network organization reflects genetic and experiential factors.

Materials and Methods

Participants

We used publicly available dMRI data from the S900 (2015) release of the Human Connectome Project (HCP; (David C. Van Essen et al., 2013)), acquired by Washington University in St. Louis and the University of Minnesota. Out of the 900 participants released, 841 genetically

unrelated participants (370 male, ages 22-37, mean age 28.76) had viable dMRI datasets. Our analysis was restricted to this subsample. All data collection procedures were approved by the institutional review boards at Washington University in St. Louis and the University of Minnesota. The post hoc data analysis was approved as exempt by the institutional review board at CMU, in accordance with 45 CFR 46.101(b)(4) (IRB Protocol Number: HS14-139).

Diffusion MRI Acquisition

The dMRI 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. 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 for each of the three shells in addition to 6 b0 images. The total scanning time was approximately 55 minutes.

Local Connectome Fingerprint Reconstruction

An outline of the pipeline for generating local connectome fingerprints is shown in the top panel of Figure 1. The dMRI data for each subject was reconstructed in a common stereotaxic space using q-space diffeomorphic reconstruction (QSDR) (F. C. Yeh & Tseng, 2011), a nonlinear registration approach that directly reconstructs water diffusion density patterns into a common stereotaxic space at 1 mm resolution.

A publicly available atlas of axonal direction in each voxel was derived from the HCP dataset. This atlas is publicly available (http://dsi-studio.labsolver.org). A spin distribution function (SDF) sampling framework was used to provide a consistent set of directions $\hat{u}$ to sample the magnitude of SDFs along axonal directions in the cerebral white matter. Since each voxel may have more than one axonal direction, multiple measurements were extracted from the SDF for voxels that contained crossing fibers, while a single measurement was extracted for voxels with fibers in a single direction. The appropriate number of density measurements from

each voxel was sampled by the left-posterior-superior voxel order and compiled into a sequence of scalar values. Gray matter was excluded using the ICBM-152 white matter mask (MacConnell Brain Imaging Centre, McGill University, Canada). The cerebellum was also excluded due to different slice coverage in cerebellum across participants. Since the density measurement has arbitrary units, the local connectome fingerprint was scaled to make the variance equal to 1 (F.-C. Yeh, Vettel, et al., 2016).

The local connectome fingerprint construction was conducted using DSI Studio (http://dsi-studio.labsolver.org), an open-source diffusion MRI analysis tool for connectome analysis. The source code, documentation, and local connectome fingerprint data are publicly available on the same website.

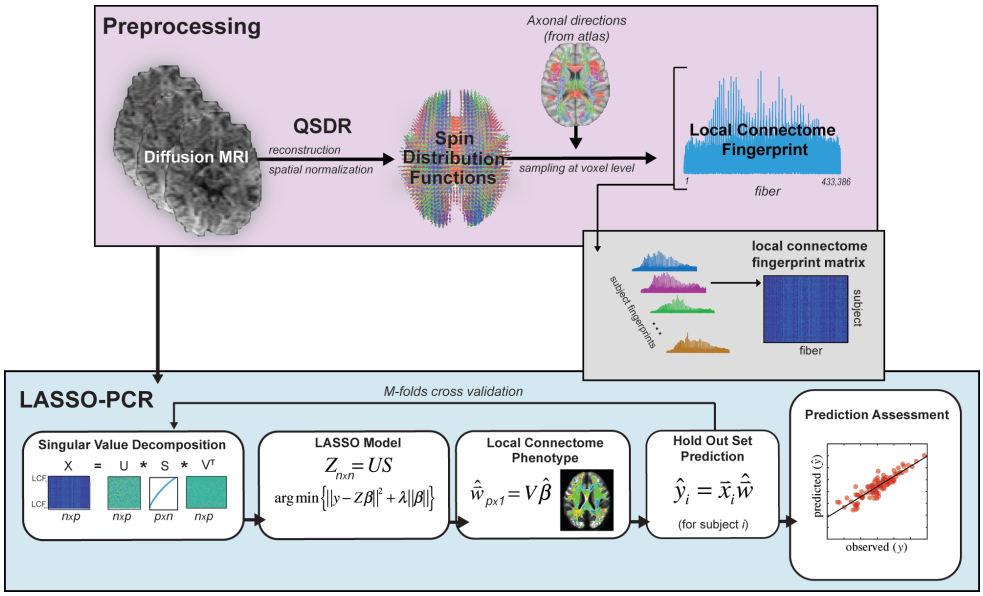

Figure 1. *Data analysis pipeline.* dMRI from the HCP dataset were preprocessed consistent with previous research investigating the local connectome fingerprint (top panel) and included registration via QSDR and estimation of SDF using an axonal directional atlas derived from the HCP dataset. Once fingerprints were estimated for each individual, the pipeline consisted of four major steps: 1) a PCA-based dimensionality reduction, 2) a LASSO model based on the lower-dimensional components of the local connectome

fingerprint, 3) local connectome phenotype estimation from projection of the contributing components of the LASSO model, and 4) prediction on the held-out dataset.

Response Variables

A total of 36 response variables across social, health, and cognitive factors were selected from the public and restricted data sets released as part of the HCP. Each variable is summarized below, but additional details can be found in the HCP Data Dictionary (https://wiki.humanconnectome.org/display/PublicData/HCP+Data+Dictionary+Public-+500+Subject+Release).

Demographic and social factors included age (years), gender, race (restricted to white and black subsets of total population), ethnicity (Hispanic vs. non-Hispanic), handedness, income (from the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) scale), education (SSAGA), and relationship status (SSAGA).

Health factors included height (inches), weight (pounds), body mass index, two hematocrit samples, blood pressure (diastolic and systolic), hemoglobin A1c, and sleep quality (Pittsburgh Sleep Quality Index).

Cognitive measures included eleven tests that sampled a broad spectrum of domains: (1) the NIH Picture Sequence Memory Test assessed episodic memory performance, (2) NIH Dimensional Change Card Sort tested executive function and cognitive flexibility, (3) NIH Flanker Inhibitory Control and Attention Test evaluated executive function and inhibition control, (4) Penn Progressive Matrices examined fluid intelligence and was measured using three performance metrics (number of correct responses, total skipped items, and median reaction time for correct responses), (5) NIH Oral Reading Recognition Test assessed language and reading performance, (6) NIH Picture Vocabulary Test examined language skills indexed by vocabulary comprehension, (7) NIH Pattern Comparison Processing Speed Test evaluated processing speed, (8) Delay Discounting tested self-regulation and impulsivity control using two

different financial incentives (Area Under the Curve (AUC) for discounting of $200, AUC for discounting of $40,000), (9) Variable Short Penn Line Orientation assessed spatial orientation performance and was measured using three metrics (total number correct, median reaction time divided by expected number of clicks for correct, total positions off for all trials), (10) Penn Word Memory Test evaluated verbal episodic memory using two performance metrics (total number of correct responses, median reaction time for correct responses), and (11) the NIH List Sorting Task tested working memory performance.

LASSO Principal Components Regression (LASSO-PCR).

The primary goal of our analysis pipeline was to identify specific patterns of variability in the local connectome that reliably predict individual differences in a specific attribute. These unique patterns would reflect a local connectome phenotype for that attribute. The LASSO-PCR pipeline used to generate local connectome phenotype (LCP) maps is illustrated in the lower panel of Figure 1. This process relied on a 5-fold cross-validation scheme in which a unique 20% of the participants were assigned to each of five subsamples. For each cross-validation fold, we trained models using 80% of the participants in order to make predictions on the held-out 20% of participants. The analysis pipeline consisted of four major steps.

Step 1: Dimensionality Reduction. The matrix of local connectome fingerprints (841 participants x 433,386 features) contains many more features than participants ($p >> N$), thereby posing a problem for fitting virtually any type of model. To efficiently develop and evaluate predictive models in a cross-validation framework, on each fold we first performed an economical singular value decomposition (SVD) on the matrix of training subjects' local connectome fingerprints (Wall, Andreas, and Rocha, n.d.) :

$$X = USV^T \qquad \text{(Eq. 1)}$$

where X is an $n \times p$ matrix containing local connectome fingerprints for *n* participants in the cross validation fold (~673 subjects × 433,386 elements per fingerprint), V^T is an $n \times p$ matrix with row

vectors representing the orthogonal principal axes of X, and the matrix product US is an $n\times n$matrix with rows corresponding to the principal components required to reproduce the original matrix X when multiplied by the principal axes matrix V^T.

Step 2: LASSO Model. To reduce the chance of overfitting and improve the generalizability of the model for a novel test set, we employed LASSO regression, a technique that penalizes the multivariate linear model for excessive complexity (i.e., number and magnitude of nonzero coefficients) (Tibshirani, 2011). The penalty in this approach arises from the L1 sparsity constraint in the fitting process, and this combined method, known was LASSO-PCR, has been used successfully in similar high-dimensional prediction models from neuroimaging data sets (Wager et al., 2013; Wager, Atlas, Leotti, & Rilling, 2011). In short, the LASSO-PCR approach identifies a sparse set of components that reliably associate individual response variables (see Figure 1) and takes the following form:

$$\hat{\beta} = arg\ min_{\beta}\{||y - Z\beta||^2 + \lambda||\beta||\} \qquad \text{(Eq. 2)}$$

where $Z = US$ as defined above. Using a cross-validation approach, we estimated the optimal λ parameter and associated $\hat{\beta}$ coefficients using the "glmnet" package in R (Friedman & Hastie, 2009). For each response-specific regression model, the model inputs included the principal components estimated from Eq. 1, i.e., US (see Figure 2), and intracranial volume (ICV). For continuous variables, e.g., reaction times, a linear regression LASSO was used. For binarized categorical variables, e.g., gender, a logistic regression variant of LASSO was used. Finally, the LASSO-produced $\hat{\beta}$ vector was truncated ($\hat{\beta}^*$) to exclude ICV and thereby restrict interpretation to the relationship between the response variables and the principal components.

The inclusion of ICV while building a model serves to restrict any predictive power of a model to only the local connectome fingerprint and not to head size, which is a common adjustment used when attempting to understand structural differences between individuals or groups to reduce the possibility of type-I errors (O'Brien et al., 2011). Our LASSO-PCR

procedure considers ICV in every model, and in some cases, ICV is deemed a significant contributor to variance in the response variable. In other cases, ICV is assigned a regression coefficient of zero. Regardless of the coefficient assigned to ICV, we ultimately want to make predictions without any knowledge of ICV by excluding the ICV coefficient and associated participant measurements from the model prediction step. While the quality of the resulting predictions (Step 4 below) may be negatively impacted by removing ICV as a potentially significant predictor in a model, controlling for ICV in this manner ensures that any observed correlation is not related to intracranial volume.

Step 3: Local Connectome Phenotype Map. For each response variable, we expect $\hat{\beta}^*$ to contain non-zero weights on a subset of the orthogonal principal components (US, or equivalently, XV), and these weights were used to construct a local connectome phenotype map, defined as the weighted influence of each fiber in the local connectome on the modeled response variable. To convert the regression coefficients into the dimensions of the local connectome, the sparse vector of regression coefficients $\hat{\beta}^*$ was multiplied by the principal axes matrix V to produce a weighted linear combination of the principal axes deemed relevant to a particular subject attribute.

$$\hat{\vec{w}} = V\hat{\beta}^* \qquad \text{(Eq. 3)}$$

This linear combination of principal axes, $\hat{\vec{w}}$, represents a $p \times 1$ vector reflecting the white matter substructure of the local connectome fingerprint vector relevant to a particular observed response. We refer to the vector $\hat{\vec{w}}$ as the local connectome phenotype for the associated response variable.

Step 4: Prediction. Finally, we use the reconstructed local connectome phenotype map to predict a variety of social, biological, and cognitive responses for participants in the test set. Ultimately, we sought a model that predicted a response variable $\hat{y}_i$ for subject *i* in the test set such that $\hat{y}_i = \vec{x_i}\hat{\vec{w}}$ where $\hat{\vec{w}}$ is the response-related local connectome phenotype and $\vec{x_i}$ is the

individual participant's local connectome fingerprint. A prediction was generated for all participants in the hold out set on each validation fold. Once predictions for all participants were generated for a given response variable, the performance of the model was evaluated as either percent correct (categorical variables) or as a correlation between predicted and observed values (continuous variables).

Results

Covariance Structure and Dimensionality of Local Connectome Fingerprints

Inter-voxel white matter architecture, reflected in the local connectome fingerprint, has been shown to be unique to an individual and sculpted by both genetic predisposition and experience (F.-C. Yeh, Vettel, et al., 2016); however, it is not yet clear whether the local connectome also exhibits reliable patterns of shared variability across individuals. To illustrate this, Figure 2A shows three exemplar fingerprints from separate subjects in the sample. These exemplars reveal the sensitivity of the method to capture both common and unique patterns of variability. For example, the highest peaks in the three fingerprints are similar in terms of their size and location. This pattern appears to exist across subjects and is generally expressed in the mean fingerprint (Fig. 2C). However, there are also clear differences between participants. For example, consider the sharpness and location of the rightmost peaks in the three exemplar fingerprints in Figure 2A. This uniqueness supports our previous work highlighting single subject classification from the fingerprint across varying temporal intervals (F.-C. Yeh, Vettel, et al., 2016).

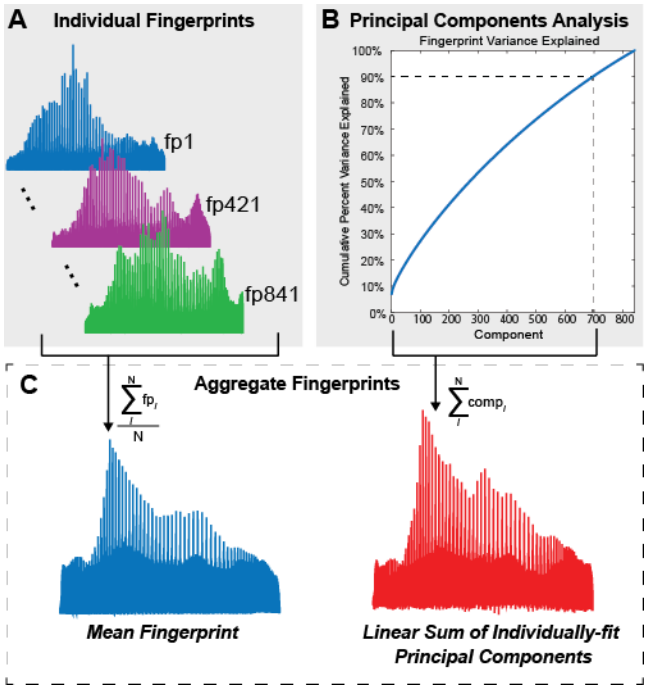

Figure 2. *Lower dimensional structure of the local connectome fingerprints.* (A) Three individual local connectome fingerprints, from three separate subjects, show coarse commonalities and unique patterns of variability when connection density is reshaped in a left-posterior-superior vector. (B) Cumulative summation of variance explained from each component, sorted by the amount of variance explained by each component. Dotted lines indicate the number of components (697) needed to explain 90% of the variability in the fingerprint dataset. (C) Mean fingerprint across participants (blue, left) and linear summation of principal components that explain 90% of the variance (red, right).

In order to explicitly test for covariance across participants, we looked at the distribution of pairwise correlations between fingerprints. The histogram in Figure 3 shows the total distribution of pairwise inter-subject correlations, revealing a tight spread of correlations such that the middle 95% of the distribution lies between 0.32 and 0.50. This confirms that inter-subject correlations are substantially lower, averaging a correlation of 0.42 across all pairs of 841 HCP participants, than intra-subject correlations, found to be well above 0.90 (F.-C. Yeh,

Vettel, et al., 2016). Thus the local connectome fingerprint exhibits a moderate but reliable covariance structure across participants, indicating its utility to examine shared structural variability across subjects that capture similarity in social, health, and cognitive factors.

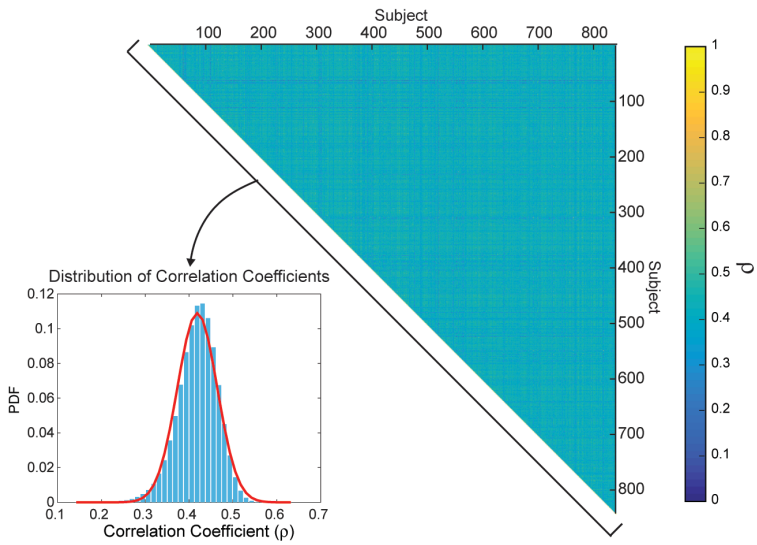

Figure 3. *Correlations between fingerprints.* The matrix of between-subject correlations in local connectome fingerprints, sorted by participant index is shown on the right. The distribution (inset) is the histogram of the upper triangle of the correlation matrix and the best fit kernel density estimate (red line).

A major challenge to examining the predictive value of the local connectome for group similarity is the dimensionality of the fingerprint itself (841 participants x 433,386 elements). It contains many more features than subjects ($p >> N$), posing a strong risk of overfitting. We employed a dimensionality reduction routine that isolates independent principal components from the entire local connectome fingerprint matrix to decompose the variance within the set of fingerprints. This analysis found that the dimensionality of the local connectome fingerprint matrix was still relatively high and complex, requiring 697 of 841 components to explain 90% of the variance (Figure 2B). While it appears that many components are required to meaningfully

explain fingerprint variance, the pattern of the mean fingerprint could be successfully recovered by a linear combination of the principal components (Figure 2C), confirming that this lower dimensional projection is adequate to represent the much larger dimensional fingerprint.

Predicting Inter-Subject Variability

After identifying a covariance structure in the group fingerprint matrix, we fit regression models to test how well the fingerprints could predict participant attributes, including social, biological, and cognitive factors. Although we used the principal components as predictor variables, the underlying dimensionality of the local connectome fingerprint matrix (697 components for 90% variance) is still quite high relative to the sample size (841 participants). Therefore, we applied an L1 sparsity constraint (i.e., LASSO) in the fitting process of a principal components regression (LASSO-PCR), as this approach identifies a sparse set of components that reliably predict individual response variables (see Figure 1).

Table 1 shows the logistic LASSO-PCR results for the four binary categorical participant attributes: gender, race, ethnicity, and relationship status. Although statistically significant fits were observed for both race and relationship status at both the training and testing stages, none of the models outperformed chance on the hold out classification tests. A possible explanation may be the saturation effects in the base rate probabilities. For example, the ethnicity response variable has a base rate of 90.5% non-Hispanic participants, meaning that a model that chooses non-Hispanic for every sample will have a classification accuracy of ~90%.

Model Response (Significant Responses Italicized)	Sample Size	Prediction Accuracy (Training)	Prediction Accuracy (CV-Test)	Confidence Interval [Lower, Upper]		Non-Zero Model Coefficients	Significant Correlation with Intracranial Volume	p-value	Confidence Interval Contains Observed Base Rate
Gender[1]	841	0.5589	0.5589	0.5262	0.5957	106	Yes	0.2835	Yes
Race[2]	760	0.4105*	0.3855*	0.3540	0.4197	115	Yes	<10E-4	No
Ethnicity[3]	841	0.9049	0.9049	0.8859	0.9239	50	No	0.4950	Yes
Relationship Status[4]	840	0.3476*	0.4429*	0.4095	0.4750	13	No	<10E-4	Yes

* The modeled response was statistically significant after correcting for multiple comparisons.

[1] The female-male split in the 841 subjects was 56%-44%, respectively.

[2] Two races were sufficiently represented in the population to build models. The white and black subpopulations made up 80% and 20%, respectively, of the 760 subjects

[3] The Not Hispanic/Latino and Hispanic split in the 841 subjects was 90.5%-9.5%, respectively.

[4] Relationship status included 44.3% of the population in a "Married or Live-in Relationship" and 55.7% not in such a relationship.

Table 1. Logistic LASSO-PCR results for the four categorical attributes.

In contrast to the binary participant attributes, we observed many reliable prediction models with the continuous variables. Table 2 (third column) shows the training results for the corresponding linear models. As expected, nearly all models were statistically significant at the training stage, even after adjusting for multiple comparisons. Only two variables, the Pittsburgh Sleep Quality Index and systolic blood pressure, were not significant at the training stage, largely because the LASSO model did not contain any non-zero coefficients.

Model Response (Siginificant Test Set Results Italicized)	**Sample Size**	**Observed vs. Predicted Correlation (Training)**	**Observed vs. Predicted Correlation (CV-Test)**	**Confidence Interval [Lower, Upper]**		**Non-Zero Model Coefficients**	**Significant Correlation with Intracranial Volume**	**p-value**
Age (in years)	841	0.1430*	0.0311	-0.0378	0.1007	6	Yes	0.1805
Handedness	841	0.5581*	-0.0594	-0.1208	0.0017	21	No	0.0312
Total Household Income	836	0.1604*	-0.0029	-0.0753	0.0632	4	Yes	0.4617
Years of Education Completed	840	0.4377*	0.0729*	0.0127	0.1343	55	No	0.0096
Height	840	0.7593*	0.5362*	0.4953	0.5773	117	Yes	<10E-4
Weight	840	0.7731*	0.1744*	0.1074	0.2405	38	Yes	<10E-4
Body Mass Index	840	0.4976*	0.2736*	0.2067	0.3421	111	No	<10E-4
Hematocrit Sample 1	772	0.4676*	0.1282*	0.0579	0.1889	10	Yes	<10E-4
Hematocrit Sample 2	740	0.3928*	0.1280*	0.0560	0.1931	49	Yes	0.0001
Mean Hematocrit Sample	740	0.4348*	0.1324*	0.0654	0.1939	21	Yes	<10E-4
Diastolic Blood Pressure	830	0.2058*	0.0615	-0.0154	0.1378	81	No	0.0561
Systolic Blood Pressure	830	0.3596*	0.1396*	0.0745	0.2076	9	Yes	<10E-4
Systolic-Diastolic Blood Pressure Ratio	830	0.0000	-0.0240	-0.0926	0.0474	12	Yes	0.2547
Hemoglobin A1C	566	0.2130*	0.0098	-0.0794	0.1071	0	No	0.4009
Pittsburgh Sleep Quality Index	841	0.0000	-0.0314	-0.0966	0.0415	12	No	0.2003
NIH Picture Sequence Memory Test	840	0.5964*	0.0977*	0.0290	0.1618	3	No	0.0023
NIH Dimensional Change Card Sort Test	839	0.2381*	-0.0299	-0.0945	0.0379	0	No	0.1853
NIH Flanker Inhibitory Control and Attention Test	841	0.1285*	-0.0001	-0.0706	0.0651	4	Yes	0.4807
Penn Progressive Matrices: Number of Correct Responses	838	0.2027*	0.0849*	0.0187	0.1502	30	Yes	0.0048
Penn Progressive Matrices: Total Skipped Items	838	0.2090*	0.0733*	0.0120	0.1383	197	Yes	0.0105
Penn Progressive Matrices: Median Reaction Time for Correct Responses	838	0.1078*	0.0086	-0.0619	0.0754	1	Yes	0.4178
NIH Oral Reading Recognition Test	841	0.1665*	0.0008	-0.0702	0.0660	53	Yes	0.4838
NIH Picture Vocabulary Test	841	0.5206*	0.0481	-0.0187	0.1142	21	Yes	0.0828
NIH Toolbox Pattern Comparison Processing Speed Test	841	0.1814*	-0.0569	-0.1260	0.0061	0	No	0.0438
Delay Discounting: Area Under the Curve for Discounting of $200	838	0.3010*	0.0275	-0.0311	0.0891	84	Yes	0.1991
Delay Discounting: Area Under the Curve for Discounting of $40,000	838	0.2056*	0.0802*	0.0132	0.1527	2	No	0.0107
Variable Short Penn Line Orientation: Total Number Correct	838	0.4490*	0.0951*	0.0279	0.1589	6	Yes	0.0029
Variable Short Penn Line Orientation: Median Reaction Time Divided by Expected Number of Clicks for Correct	838	0.4449*	-0.0572	-0.1302	0.0141	4	Yes	0.0603
Variable Short Penn Line Orientation: Total Positions Off for All Trials	838	0.4695*	0.0014	-0.0621	0.0735	8	Yes	0.4811
Penn Word Memory Test: Total Number of Correct Responses	838	0.2382*	0.0474	-0.0228	0.1189	45	No	0.0881
Penn Word Memory Test: Median Reaction Time for Correct Responses	838	0.2354*	-0.0391	-0.0965	0.0191	0	No	0.1119
NIH List Sorting Working Memory Test	841	0.4140*	0.0793*	0.0097	0.1540	1	Yes	0.0137

* The modeled response was statistically significant after correcting for multiple comparisons.

Table 2. Linear LASSO-PCR results for the thirty-two continuous attributes.

To complement the model training results, we examined the predictive performance of the model on an independent, hold-out test sample. This was done by projecting the regression weights in component space back into local connectome space in order to provide a weight map for each fiber in the local connectome to the target response variable. These maps reflect the local connectome phenotype for that attribute and were multiplied against a full local connectome fingerprint for each participant in the hold-out set to generate a prediction for that participant (see bottom panel, Figure 1).

We assessed the generalizability of 32 continuous response models in a cross-validation paradigm and, as shown in Table 2 (column 4), 14 (44%) of these attributes were significant predictors after correcting for multiple comparisons. These factors included years of education, measures of body type (height, weight, BMI), physiology (hematocrit samples, blood pressure measures), and several cognitive measures including episodic memory (NIH Picture Sequence Memory Test), fluid intelligence (Penn Progressive Matrices: Number of Correct Responses & Total Skipped Items), self-regulation (Delay Discounting: Area Under the Curve for Discounting of $40,000), spatial orientation (Variable Short Penn Line Orientation: Total Number Correct), and working memory (NIH List Sorting Working Memory Test).

Specificity of Phenotypes to Response Variables

Some of the significant models for the hold-out test set evaluation reflect correlated response variables (e.g., two hematocrit samples) or response variables generated from other response variables (e.g., BMI is calculated from height and weight). It should be noted, however, that a large portion of the significant test predictions came from largely independent attributes. The ability to predict some of the correlated participant attributes raises the possibility that a local connectome phenotype map learned for one attribute may not be unique to the attribute being modeled, but instead might reflect one or a few generalized maps that explain variance expressed across multiple attributes. Therefore, in our final analysis, we examined the

specificity of a local connectome phenotype map by considering whether or not the predictive maps were unique for each participant attribute being predicted. In other words, we tested whether a single map could capture a generalized predictive relationship for multiple, response variables, indicating that the models themselves may lack specificity. If so, any given model may perform suitably well at predicting any participant attribute (e.g., BMI), even if derived from training on a different participant factor (e.g., years of education completed).

To explicitly test this, we looked at the correlation between the 14 significant phenotype maps from the hold-out test shown in Table 2. This correlation is shown in Figure 4. There were four clusters of moderately correlated models. Three of these clusters of associations are not surprising. For example, BMI is calculated from height and weight, the mean hematocrit sample is calculated from the two hematocrit samples, and people with more correct responses skipped fewer items in the Penn. Progressive Matrices test. Only the correlation between the phenotypes for the Variable Short Penn Line Orientation task and the NIH List Sorting Working Memory Test was unexpected. We elaborate on this association in the discussion section.

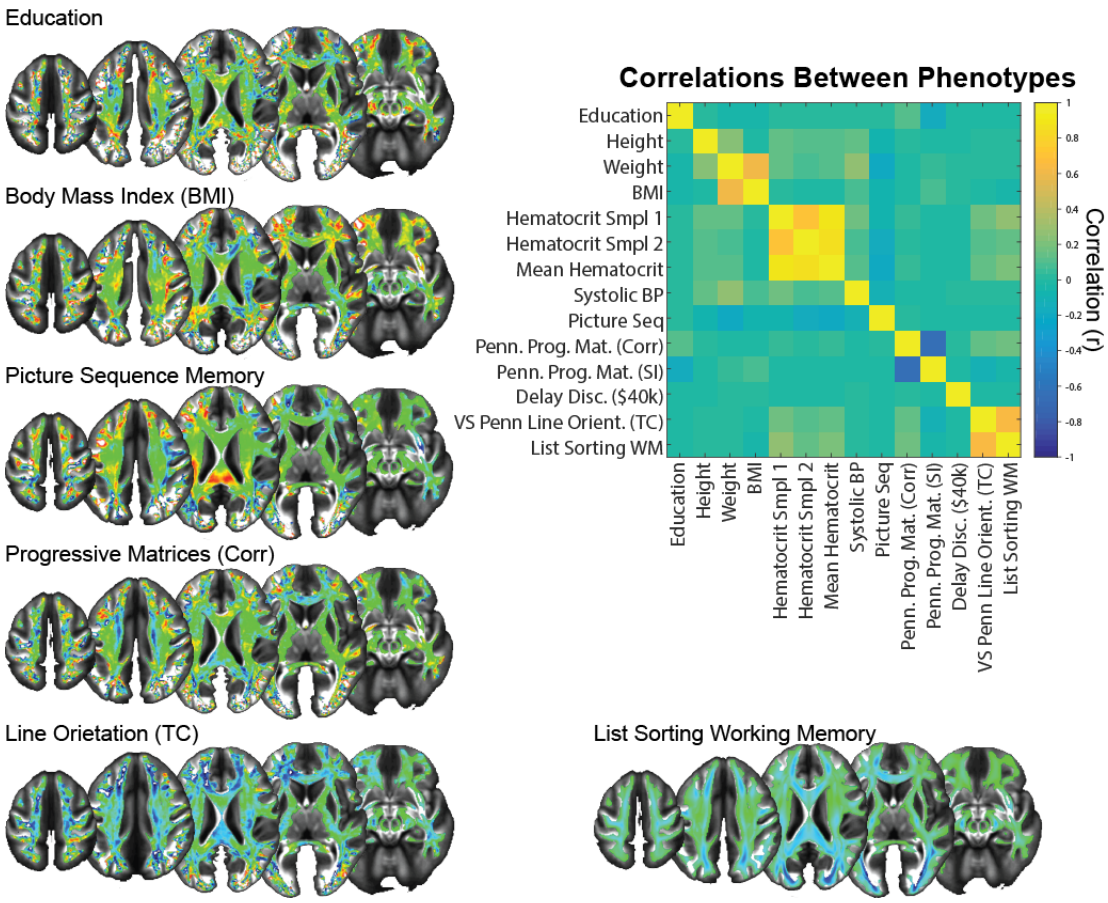

Figure 4. *Local connectome phenotypes.* Matrix inset is a correlation matrix displaying the similarity between phenotypes of the local connectome. Example phenotype maps are shown around the correlation matrix, representing the areas of the local connectome that are most predictive of the labeled response variable.

While it is perhaps not surprising that phenotypes for correlated response variables are themselves correlated, the majority (91%) of phenotype maps were uncorrelated. We visualized the uniqueness of these phenotype maps by projecting the local connectome phenotypes into voxel space. A subset of these maps is shown in Figure 4. Visual inspection of these six example phenotype maps reveals large heterogeneity between models. For instance, strong positive loadings are observed in portions of the splenium of the corpus callosum and frontal association fiber systems for the Picture Sequence Memory Task, while these same regions

load negatively for the Variable Short Penn Line Orientation test and NIH List Sorting Working Memory Test. Bilateral corona radiata pathways appear to negatively load for the Penn Progressive Matrices and Variable Short Penn Line Orientation test, but not for any of the other attributes. These qualitative comparisons, along with the direct correlation tests, confirm that the phenotype maps for predicting inter-subject variability are highly specific to the variable being modeled.

Discussion

Our analysis revealed, for the first time, that the local connectome fingerprint exhibits a moderate, but reliable, correlation between participants that can be leveraged to predict at the level of the individual along dimensions of social, biological, and cognitive attributes. Although the between-subject correlation is much smaller than the within-participant correlation reported previously (F.-C. Yeh, Vettel, et al., 2016), it was robust enough to capture inter-subject similarities. Much to our surprise, the lower dimensional structure of this inter-subject covariance was still relatively complex, with hundreds of principal components required to explain most of the variance in the sample. Using a cross-validation regression approach that is optimized for ultra-high dimensional data sets, we show how patterns of variability in the local connectome not only correlated with nearly all participant-level social, health, and cognitive attributes (i.e., strong and significant training accuracy) but could also independently predict variability in almost half of the features tested (i.e., hold-out test accuracy). Finally, we were able to show how the local connectome phenotype maps for individual attributes were highly specific to the variable being modeled. This suggests that there is not some unique, generalizable feature of local white matter that predicts inter-subject variability, but instead there are highly specific patterns that predict variance in specific inter-subject attributes. Taken together, the current results confirm our hypothesis that shared variability across participants is reflected in the local connectome itself. This opens the door for leveraging the local connectome fingerprint, along

with functional measures of connectomic architecture (Shen et al., 2017), as a reliable marker for individual differences in behavior.

The current findings clearly show how it is possible to recover a portion of variability in social, biological, or cognitive attributes from the white matter signal itself. The novelty of this predictive approach opens the door to a new way of assessing brain-behavior links that moves beyond association testing. By building a reliable phenotype map that predicts a portion of behavioral variability, we highlight that neuroimaging tools have both the sensitivity and specificity to describe individual differences in features of interest. For example, in our study, structural similarity in the local connectome fingerprint reliably predicted five of the tested cognitive performance measurements, including a list sorting task that captures individual variability in working memory performance (R C Gur et al., 2001; Ruben C Gur et al., 2010). The associated local connectome phenotype for working memory identified portions of what appear to be frontoparietal pathways (Figure 4). Our results nicely complement a recent study of working memory that focused on direct and indirect connectivity in the frontoparietal networks (Ekman, Fiebach, Melzer, Tittgemeyer, & Derrfuss, 2016). In their work, the authors found that the network centrality of focal structural connections in the frontal, temporal, and parietal cortices could predict individual differences in working memory capacity using linear regression. When considered in the context of the current study, our findings augment previous *correlative* findings between frontoparietal regions and working memory capacity (Bender, Prindle, Brandmaier, & Raz, 2016; Klingberg, 2006; Nagy, Westerberg, & Klingberg, 2004; Takeuchi et al., 2010) by showing that the integrity of the pathway of these white matter fascicles reliably *predicts* working memory performance.

The existence of reliable and predictive inter-subject covariance patterns in the white matter fascicles of the human brain begs the question of mechanism: are these similarities genetically determined, experientially sculpted, or developed through gene-by-environment interactions? Emergent findings in genetics are suggesting that at least a portion of macroscopic

white matter structure is guided by genetics (Kochunov, Fu, et al., 2016; Kochunov, Thompson, et al., 2016; F.-C. Yeh, Vettel, et al., 2016).. For example, recent work by Kochunov and colleagues (2016a) examined a heritability relationship between whole-brain fractional anisotropy (FA) and information processing speed in two interesting participant populations, the HCP twins cohort and an Old Order Amish cohort. The cohorts both had well-characterized genetic properties, but they differed in the amount of experiential variability since the Amish have more high environmental homogeneity compared to the urban/suburban HCP cohort. Kochunov and colleagues (2016a) argued that the replication of the genetic contribution to processing speed and FA of cerebral white matter despite the experiential variability in the cohorts suggested a strong phenotypic association for the trait. Our analysis would be able to pick up such genetically mediated brain-behavior phenotypes.

While genetics may contribute to white matter architecture, overwhelming evidence suggests that experience sculpts these pathways over time. For example, variability in the white matter signal has been shown to covary with several social (Gianaros et al., 2013; Molesworth et al., 2015), biological (Arfanakis et al., 2013; Miralbell et al., 2012; Verstynen et al., 2013), and cognitive (Muraskin et al., 2016; Verstynen, 2014; Ystad et al., 2011) attributes. In many cases, it is difficult to extract or identify specific pathways or systems that link white matter pathways to these shared experiential factors. However, several intervention studies have targeted more specific experience-white matter associations. For example, prolonged training on a variety of tasks has been shown to induce changes in the diffusion MRI signal (Blumenfeld-Katzir, Pasternak, Dagan, & Assaf, 2011; Sampaio-Baptista et al., 2013; Scholz, Klein, Behrens, & Johansen-Berg, 2009; Steele, Scholz, Douaud, Johansen-Berg, & Penhune, 2012). In some cases, the particular change in the diffusion signal is consistent with alterations in the underlying myelin (Sampaio-Baptista et al., 2013), for which there is emerging support from validation studies in non-human animal models (Budde, Janes, Gold, Turtzo, & Frank, 2011; Budde, Xie, Cross, & Song, 2009; Klawiter et al., 2011). One consistency in these reports of training-induced

plasticity in white matter pathways is that the effects are task-specific (i.e., training in a specific task appears to impact specific white matter fascicles). This specificity of experiential factors on white matter pathways is necessary in order to be able to build reliable prediction models from the diffusion MRI signal.

Our previous work showed that the local connectome fingerprint reflects both genetic and experiential factors that contribute to between-subject variability in white matter architecture (F.-C. Yeh, Vettel, et al., 2016). We found that monozygotic twins expressed a modest degree of similarity in their local connectome fingerprints, with ~12% of the local connectome pattern being similar between monozygotic twins. This similarity was much higher than what was detected in siblings or dizygotic twins; however, genetic similarities overall seemed to contribute very little to similarities in the local connectome. In contrast, most of the structure in the local connectome fingerprint appeared to be driven by experience. By comparing changes in the fingerprint over time, average intra-subject similarity changed linearly with time. While it can be argued that part of this change simply reflects aspects of the normal aging process (Simmonds, Hallquist, Asato, & Luna, 2014; Westlye et al., 2010), we should point out that the intra-subject changes seen in our previous study happen at a much faster rate than typical age-related changes in white matter pathways (i.e., days and weeks vs. years, respectively). Thus we expect that much of this plasticity is likely due to experiential factors.

One of the strengths of the local connectome fingerprint approach used here is that it does not rely on fiber tracking algorithms. Recent evidence indicates a false positive bias when mapping white matter pathways (Daducci et al., 2016; Reveley et al., 2015; Thomas et al., 2014). This is due in large part to the difficulty that tracking algorithms have when distinguishing between a crossing and turning fiber pathway. Our approach does not rely on a deterministic or probabilistic tracking algorithm; instead, we analyze the entire set of reconstructed fibers throughout the brain as a unitary data object. This eliminates the false positive identification of white matter fascicles by not attempting fascicular classification at all. However, without tracking

along pathways we cannot say whether specific pathways positively or negatively predict a specific response variable. In the future, exploration of the local connectome phenotype maps with careful pathway labeling, e.g., expert-vetted fiber labeling, can identify general regions that positively or negatively contribute to the prediction.

Another limitation of the approach used here arises from the fact that, by necessity, the local connectome fingerprints must be computed from a common, atlas-defined space. The nonlinear transformations required in order to transform brains of various shapes and sizes into a stereotaxic space through the QSDR procedure invariably introduce a degree of noise in the SDFs. The number and orientation of fibers in each voxel determine the local connectome fingerprint, and these measurements could possibly be distorted during QSDR. Such a transformation is unavoidable because the dimensionality of each fingerprint must be identical, and each element of a fingerprint must represent the same brain micro-region as the corresponding element in any other fingerprint. Only with this common, atlas-aligned representation of the local connectome fingerprint can we apply LASSO-PCR to explore common substructures. The potential price for this convenience is an introduction of noise in the local connectome fingerprint itself, likely increasing the possibility of a false-negative error, e.g., failing to recognize a true phenotypic relationship.

Despite these limitations, the current work clearly shows that the local connectome fingerprint reliably reflects shared variance between individuals in the macroscopic white matter pathways of the brain. For the first time, we not only show how global white matter structure associates with different participant features, but we also show how the entire local connectome itself can predict a portion of the variability in independent samples. While the overall variance explained by the local connectome fingerprint may at first seem small, it is consistent or even stronger than effect sizes of genetic risk scores used in behavioral medicine (Plomin, DeFries, Knopik, & Neiderhiser, 2016). Thus our local connectome phenotyping approach may also be predictive of not only normal, but also pathological variability (see also Yeh et al., 2013). Future

work in clinical populations should focus on applying this approach to generate diagnostic local connectome phenotypes for neurological and psychiatric disorders, thereby leveraging the full potential of this approach.

References

Arfanakis, K., Fleischman, D. a, Grisot, G., Barth, C. M., Varentsova, A., Morris, M. C., … Bennett, D. a. (2013). Systemic inflammation in non-demented elderly human subjects: brain microstructure and cognition. *PloS One*, *8*(8), e73107. http://doi.org/10.1371/journal.pone.0073107

Bender, A. R., Prindle, J. J., Brandmaier, A. M., & Raz, N. (2016). White matter and memory in healthy adults: Coupled changes over two years. *NeuroImage*, *131*, 193–204. http://doi.org/10.1016/j.neuroimage.2015.10.085

Blumenfeld-Katzir, T., Pasternak, O., Dagan, M., & Assaf, Y. (2011). Diffusion MRI of structural brain plasticity induced by a learning and memory task. *PLoS ONE*, *6*(6), e20678. http://doi.org/10.1371/journal.pone.0020678

Budde, M. D., Janes, L., Gold, E., Turtzo, L. C., & Frank, J. a. (2011). The contribution of gliosis to diffusion tensor anisotropy and tractography following traumatic brain injury: validation in the rat using Fourier analysis of stained tissue sections. *Brain*, *134*, 2248–60. http://doi.org/10.1093/brain/awr161

Budde, M. D., Xie, M., Cross, A. H., & Song, S. (2009). Axial Diffusivity Is the Primary Correlate of Axonal Injury in the Experimental Autoimmune Encephalomyelitis Spinal Cord : A Quantitative Pixelwise Analysis. *The Journal of Neuroscience*, *29*(9), 2805–2813. http://doi.org/10.1523/JNEUROSCI.4605-08.2009

Bullmore, E. T., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews. Neuroscience*, *10*(3), 186–98. http://doi.org/10.1038/nrn2575

Daducci, A., Dal Palú, A., Descoteaux, M., & Thiran, J.-P. (2016). Microstructure Informed Tractography: Pitfalls and Open Challenges. *Frontiers in Neuroscience*, *10*, 247. http://doi.org/10.3389/fnins.2016.00247

Ekman, M., Fiebach, C. J., Melzer, C., Tittgemeyer, M., & Derrfuss, J. (2016). Different Roles of Direct and Indirect Frontoparietal Pathways for Individual Working Memory Capacity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *36*(10), 2894–903. http://doi.org/10.1523/JNEUROSCI.1376-14.2016

Fan, Q., Witzel, T., Nummenmaa, A., Van Dijk, K. R. A., Van Horn, J. D., Drews, M. K., … Rosen, B. R. (2016). MGH-USC Human Connectome Project datasets with ultra-high b-value diffusion MRI. *NeuroImage*, *124*(Pt B), 1108–14. http://doi.org/10.1016/j.neuroimage.2015.08.075

Friedman, J., & Hastie, T. (2009). Regularization Paths for Generalized Linear Models via Coordinate Descent, 1–24.

Gianaros, P. J., Marsland, A. L., Sheu, L. K., Erickson, K. I., & Verstynen, T. D. (2013). Inflammatory pathways link socioeconomic inequalities to white matter architecture. *Cerebral Cortex*, *23*(9), 2058–2071. http://doi.org/10.1093/cercor/bhs191

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., … Bassett, D. S. (2015). Controllability of structural brain networks. *Nature Communications*, *6*, 8414. http://doi.org/10.1038/ncomms9414

Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., … Gur, R. E. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *25*(5), 766–76. http://doi.org/10.1016/S0893-133X(01)00278-0

Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., … Gur, R. E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *Journal of Neuroscience Methods*, *187*(2), 254–62. http://doi.org/10.1016/j.jneumeth.2009.11.017

Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Grant, P. E., Wedeen, V., … Sporns, O. (2010). MR connectomics: Principles and challenges. *Journal of Neuroscience Methods*,

194(1), 34–45. http://doi.org/10.1016/j.jneumeth.2010.01.014

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, *6*(7), e159. http://doi.org/10.1371/journal.pbio.0060159

Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V. J., … Grant, P. E. (2010). White matter maturation reshapes structural connectivity in the late developing human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(44), 19067–72. http://doi.org/10.1073/pnas.1009073107

Jbabdi, S., Sotiropoulos, S. N., Haber, S. N., Van Essen, D. C., & Behrens, T. E. (2015). Measuring macroscopic brain connections in vivo. *Nature Neuroscience*, *18*(11), 1546–55. http://doi.org/10.1038/nn.4134

Klawiter, E. C., Schmidt, R. E., Trinkaus, K., Liang, H.-F., Budde, M. D., Naismith, R. T., … Benzinger, T. L. (2011). Radial diffusivity predicts demyelination in ex-vivo Multiple Sclerosis spinal cords. *Neuroimage*, *55*(4), 1454–1460. http://doi.org/10.1016/j.neuroimage.2011.01.007.Radial

Klingberg, T. (2006). Development of a superior frontal-intraparietal network for visuo-spatial working memory. *Neuropsychologia*, *44*(11), 2171–7. http://doi.org/10.1016/j.neuropsychologia.2005.11.019

Kochunov, P., Fu, M., Nugent, K., Wright, S. N., Du, X., Muellerklein, F., … Hong, L. E. (2016). Heritability of complex white matter diffusion traits assessed in a population isolate. *Human Brain Mapping*, *37*(2), 525–35. http://doi.org/10.1002/hbm.23047

Kochunov, P., Thompson, P. M., Winkler, A., Morrissey, M., Fu, M., Coyle, T. R., … Hong, L. E. (2016). The common genetic influence over processing speed and white matter microstructure: Evidence from the Old Order Amish and Human Connectome Projects. *NeuroImage*, *125*, 189–97. http://doi.org/10.1016/j.neuroimage.2015.10.050

Le Bihan, D., & Johansen-Berg, H. (2012). Diffusion MRI at 25: exploring brain tissue structure

and function. *NeuroImage*, *61*(2), 324–41. http://doi.org/10.1016/j.neuroimage.2011.11.006

Miralbell, J., Soriano, J. J., Spulber, G., López-Cancio, E., Arenillas, J. F., Bargalló, N., … Mataró, M. (2012). Structural brain changes and cognition in relation to markers of vascular dysfunction. *Neurobiology of Aging*, *33*(5), 1003.e9-17. http://doi.org/10.1016/j.neurobiolaging.2011.09.020

Molesworth, T., Sheu, L. K., Cohen, S., Gianaros, P. J., & Verstynen, T. D. (2015). Social network diversity and white matter microstructural integrity in humans. *Social Cognitive and Affective Neuroscience*, *10*(9), 1169–76. http://doi.org/10.1093/scan/nsv001

Muldoon, S. F., Pasqualetti, F., Gu, S., Cieslak, M., Grafton, S. T., Vettel, J. M., & Bassett, D. S. (2016). Stimulation-Based Control of Dynamic Brain Networks. *PLoS Computational Biology*, *12*(9), e1005076. http://doi.org/10.1371/journal.pcbi.1005076

Muraskin, J., Dodhia, S., Lieberman, G., Garcia, J. O., Verstynen, T., Vettel, J. M., … Sajda, P. (2016). Brain dynamics of post-task resting state are influenced by expertise: Insights from baseball players. *Human Brain Mapping*, *37*(12), 4454–4471. http://doi.org/10.1002/hbm.23321

Nagy, Z., Westerberg, H., & Klingberg, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *Journal of Cognitive Neuroscience*, *16*(7), 1227–33. http://doi.org/10.1162/0898929041920441

O'Brien, L. M., Ziegler, D. A., Deutsch, C. K., Frazier, J. A., Herbert, M. R., & Locascio, J. J. (2011). Statistical adjustments for brain size in volumetric neuroimaging studies: some practical implications in methods. *Psychiatry Research*, *193*(2), 113–22. http://doi.org/10.1016/j.pscychresns.2011.01.007

Passingham, R. E., Stephan, K. E., & Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nature Reviews. Neuroscience*, *3*(8), 606–16. http://doi.org/10.1038/nrn893

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 Replicated

Findings From Behavioral Genetics. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *11*(1), 3–23. http://doi.org/10.1177/1745691615617439

Raffelt, D. a., Smith, R. E., Ridgway, G. R., Tournier, J.-D., Vaughan, D. N., Rose, S., … Connelly, A. (2015). Connectivity-Based Fixel Enhancement: Whole-Brain Statistical Analysis of Diffusion MRI Measures in the Presence of Crossing Fibres. *NeuroImage*, *117*, 40–55. http://doi.org/10.1016/j.neuroimage.2015.05.039

Reveley, C., Seth, A. K., Pierpaoli, C., Silva, A. C., Yu, D., Saunders, R. C., … Ye, F. Q. (2015). Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion MR tractography. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(21), E2820-8. http://doi.org/10.1073/pnas.1418198112

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, *52*(3), 1059–69. http://doi.org/10.1016/j.neuroimage.2009.10.003

Sampaio-Baptista, C., Khrapitchev, A. a, Foxley, S., Schlagheck, T., Scholz, J., Jbabdi, S., … Johansen-Berg, H. (2013). Motor skill learning induces changes in white matter microstructure and myelination. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(50), 19499–503. http://doi.org/10.1523/JNEUROSCI.3048-13.2013

Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience*, *12*(11), 1370–1. http://doi.org/10.1038/nn.2412

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518. http://doi.org/10.1038/nprot.2016.178

Simmonds, D. J., Hallquist, M. N., Asato, M., & Luna, B. (2014). Developmental stages and sex differences of white matter and behavioral development through adolescence: a longitudinal diffusion tensor imaging (DTI) study. *NeuroImage*, *92*, 356–68. http://doi.org/10.1016/j.neuroimage.2013.12.044

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Publishing Group*, *17*(5), 652–660. http://doi.org/10.1038/nn.3690

Sporns, O., Tononi, G., & Kotter, R. (2005). The Human Connectome : A Structural Description of the Human Brain. *PLoS Computational Biology*, *1*(4). http://doi.org/10.1371/journal.pcbi.0010042

Steele, C. J., Scholz, J., Douaud, G., Johansen-Berg, H., & Penhune, V. B. (2012). Structural correlates of skilled performance on a motor sequence task. *Frontiers in Human Neuroscience*, *6*(October), 289. http://doi.org/10.3389/fnhum.2012.00289

Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., … Kawashima, R. (2010). Training of working memory impacts structural connectivity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(9), 3297–303. http://doi.org/10.1523/JNEUROSCI.4611-09.2010

Thivierge, J.-P., & Marcus, G. F. (2007). The topographic brain: from neural connectivity to cognition. *Trends in Neurosciences*, *30*(6), 251–9. http://doi.org/10.1016/j.tins.2007.04.004

Thomas, C., Ye, F. Q., Irfanoglu, M. O., Modi, P., Saleem, K. S., Leopold, D. A., & Pierpaoli, C. (2014). Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proceedings of the National Academy of Sciences*, *111*(46), 16574–9. http://doi.org/10.1073/pnas.1405672111

Tibshirani, R. (2011). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology*, *73*(3), 267–288. http://doi.org/10.1111/j.1467-9868.2011.00771.x

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K.

(2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. http://doi.org/10.1016/j.neuroimage.2013.05.041

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D. M., Behrens, T. E. J., Bucholz, R., … Yacoub, E. (2012). The Human Connectome Project: a data acquisition perspective. *NeuroImage*, *62*(4), 2222–31. http://doi.org/10.1016/j.neuroimage.2012.02.018

Verstynen, T. D. (2014). The organization and dynamics of corticostriatal pathways link the medial orbitofrontal cortex to future behavioral responses. *Journal of Neurophysiology*, *112*(10), 2457–69. http://doi.org/10.1152/jn.00221.2014

Verstynen, T. D., Weinstein, A. M., Erickson, K. I., Sheu, L. K., Marsland, A. L., & Gianaros, P. J. (2013). Competing physiological pathways link individual differences in weight and abdominal adiposity to white matter microstructure. *NeuroImage*, *79*, 129–137. http://doi.org/10.1016/j.neuroimage.2013.04.075

Wager, T. D., Atlas, L. Y., Leotti, L. A., & Rilling, J. K. (2011). Predicting individual differences in placebo analgesia: contributions of brain activity during anticipation and pain experience. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(2), 439–452. http://doi.org/10.1523/JNEUROSCI.3420-10.2011

Wager, T. D., Atlas, L. Y., Lindquist, M. a, Roy, M., Woo, C.-W., & Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *The New England Journal of Medicine*, *368*(15), 1388–97. http://doi.org/10.1056/NEJMoa1204471

Westlye, L. T., Walhovd, K. B., Dale, A. M., Bjørnerud, A., Due-Tønnessen, P., Engvig, A., … Fjell, A. M. (2010). Life-span changes of the human brain white matter: diffusion tensor imaging (DTI) and volumetry. *Cerebral Cortex*, *20*(9), 2055–68. http://doi.org/10.1093/cercor/bhp280

Yeh, F.-C., Badre, D., & Verstynen, T. (2016). Connectometry: A statistical approach harnessing the analytical potential of the local connectome. *NeuroImage*, *125*, 162–71. http://doi.org/10.1016/j.neuroimage.2015.10.053

Yeh, F.-C., Tang, P.-F., & Tseng, W.-Y. I. (2013). Diffusion MRI connectometry automatically reveals affected fiber pathways in individuals with chronic stroke. *NeuroImage. Clinical*, *2*, 912–21. http://doi.org/10.1016/j.nicl.2013.06.014

Yeh, F.-C., Vettel, J. M., Singh, A., Poczos, B., Grafton, S. T., Erickson, K. I., … Verstynen, T. D. (2016). Quantifying Differences and Similarities in Whole-Brain White Matter Architecture Using Local Connectome Fingerprints. *PLoS Computational Biology*, *12*(11), e1005203. http://doi.org/10.1371/journal.pcbi.1005203

Yeh, F. C., & Tseng, W.-Y. Y. I. (2011). NTU-90: A high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction. *NeuroImage*, *58*(1), 91–99. http://doi.org/10.1016/j.neuroimage.2011.06.021

Ystad, M., Hodneland, E., Adolfsdottir, S., Haász, J., Lundervold, A. J., Eichele, T., & Lundervold, A. (2011). Cortico-striatal connectivity and cognition in normal aging: a combined DTI and resting state fMRI study. *NeuroImage*, *55*(1), 24–31. http://doi.org/10.1016/j.neuroimage.2010.11.016