Genetic and environmental influence on the human functional connectome

Andrew E. Reineberg*, Alexander S. Hatoum, John K. Hewitt
Institute for Behavioral Genetics
Department of Psychology and Neuroscience
University of Colorado Boulder

Marie T. Banich
Institute of Cognitive Science
Department of Psychology and Neuroscience
University of Colorado Boulder

Naomi P. Friedman
Institute for Behavioral Genetics
Department of Psychology and Neuroscience
University of Colorado Boulder

Detailed mapping of genetic and environmental influences on the functional connectome is a crucial step toward developing intermediate phenotypes between genes and clinical diagnoses or cognitive abilities. Historical attempts to estimate the genetic etiology of the connectome have focused on large-scale brain networks - obscuring possible heterogeneity among or novel communities of small network subcomponents. In the current study, we analyze resting-state data from two, adult twin samples - 198 twins from the Colorado Longitudinal Twin Sample and 422 twins from the Human Connectome Project - to examine genetic and environmental influence on all pairwise functional connections between 264 brain regions (~35,000 functional connections). We find high non-shared environmental influence across the entire connectome, moderate heritability in roughly half of all connections, and weak-to-moderate shared environmental influences. The pattern of genetic influence across the connectome is related to *a priori* notions of functional brain networks but also highly heterogeneous as confirmed by a hierarchical clustering analysis of the genetic profile of all 264 regions. Additionally, we confirm genetic influences on connections are independent of genetic influences shared with a global summary measure of brain connectivity - an important validation analysis for future, high-dimensionality genetic neuroimaging studies. Together, our analyses reveal a novel genetic taxonomy of brain regions and have implications for studies employing multivariate signals for prediction purposes. Variation across the population in those neurobiological signals is influenced by genes and the environment in different spatial locations and to different degrees suggesting genetic risk factors may be limited to a subset of the connectome.

Keywords: functional connectivity, resting-state, heritability, genetic neuroimaging

The functional connectome refers to intrinsically correlated activity between brain regions when individuals are not engaged in a particular task (i.e., measured during the "resting state"; [Fox and Raichle, 2007]). Patterns within the functional connectome are associated with clinical diagnoses (for reviews see [Greicius, 2008, Zhang and Raichle, 2010]) and individual differences in cognitive abilities (for a broad review of 125 studies see [Vaidya and Gordon, 2013]). Recent work has showcased reliable and generalizable predictive models of individual differences in behavior that utilize many measurements of the connectome as features [Rosenberg et al., 2016, Finn et al., 2015] suggesting patterns of connectivity may be candidate intermediate phenotypes (i.e., endophenotypes; [Hall and Smoller, 2010, Kendler and Neale, 2010]) between genes and traits if genetic influences exist. However, no work has quantified genetic and environmental influences on the connectome at the level of analysis of small regions of interest as used in multivariate predictive models.

The unit of analysis for studies investigating the functional connectome spans several orders of magnitude - from functional connections between a small number of large networks with correlated activity [Yeo et al., 2011] and related function [Smith et al., 2009] to functional connections between nearly a million individual voxels. Due to the computational power needed to perform classic twin models at the level of voxels or small regions, current efforts have focused on quantifying genetic and environmental influence on either global summary measures of functional connectivity, resting-

* Corresponding Author: andrew.reineberg@colorado.edu

state networks (i.e., large and spatially-separated groups of regions that are all moderately correlated at rest and thus appropriate to model as a single unit), or large regions of interest (ROIs). At the coarsest level of detail, several studies have revealed moderate heritability (i.e., "h^2" or the proportion of phenotypic variance explained by genetic variance; h^2 = 0.43 - 0.64) of the degree to which an individual's connectome is globally efficient (i.e., maximizes information transfer while reducing long path lengths and unnecessary connections; [van den Heuvel et al., 2013, Fornito et al., 2011, Sinclair et al., 2015]). However, while global efficiency may be an informative phenotype, it may not be a thorough summary of the entire connectome, possibly summarizing only connections amongst the brains' densely connected and metabolically costly hub regions [Heuvel et al., 2012].

At the level of networks, functional connectivity of a network implicated in internal mentation functions, the default network [Andrews-Hanna, 2011], is moderately heritable as a whole (h^2 = 0.42), while connectivity of subcomponents of the default network are weakly-to-moderately heritable (h^2 range: 0.10 to 0.42; [Glahn et al., 2010]), replicating observations from the anatomical literature that heritability estimates may be higher for large versus small pieces of cortex [Eyler et al., 2012]. Other work has reported moderate heritability of a precuneus-dorsal posterior cingulate network, visual network, default network, frontoparietal network, and dorsal attention network (range h^2: 0.23 - 0.65), non-significant heritability for the salience and sensory-somatomotor networks, and evidence of environmental effects on functional connectivity between networks [Yang et al., 2016]. Finally, a recent study investigated the genetic etiology of functional connections among seven networks and pairwise connections between 51 brain areas, finding moderate-to-strong heritability of seven networks (range h^2: ~0.60 - ~0.75) using a linear mixed effects model approach to account for unreliability across multiple resting-state scans [Ge et al., 2017]. At the level of the 51-region parcellation, the authors found heritability estimates for components of some network, such as the default network, were consistent, but also found evidence of heterogeneity for regions of other networks such as the limbic and cognitive control networks.

In summary, existing studies have provided heritability estimates for functional connectivity at global, network, or large ROI levels of analysis. Generally speaking, parcellations that include many regions (e.g., 200 - 500) claim to divide the brain into units with distinct specific functions (as opposed to vague overarching functional labels assigned to large networks, such as vision). Although coarser levels of analysis are undoubtedly informative, they preclude an examination of key questions that can only be addressed by taking a finer-grained approach. In particular, examining heritability at the network level assumes that areas within the networks are homogeneous in terms of their genetic connections to areas in other networks. Moreover, individual differences in within-network connectivity cannot be examined, and these individual differences may have important implications for behavior (i.e., as contributors to "fingerprints" of cognitive processes or psychopathology [Rosenberg et al., 2016]). Important questions that can be examined at a finer level of analysis are 1) Do within- and between-network connections show similar levels of genetic and environmental influences? 2) Are networks homogeneous in terms of their genetically influenced connectivity to other networks? That is, is the best way to conceptualize genetic influences on connectivity really in terms of *a priori* networks? Finally, 3) Does a summary measure of the connectome like global efficiency capture most of the genetic variance in local connectivity (c.f., as a global measure of cognition captures most of the genetic variance in specific cognitive abilities [Petrill, 1997, Panizzon et al., 2014])? Answering these questions speaks to recent efforts to "carve nature at its joints" and thus has important implications for how we conceptualize resting state connectivity as a biomarker or candidate endophenotype for behaviors of interest.

To answer these questions, we analyzed resting state data from two comparably-aged adult twin samples: the Colorado Longitudinal Twin Study (LTS; N = 251, including 54 complete monozygotic [MZ] and 45 complete same-sex dizygotic [DZ] pairs), and the Human Connectome Project (HCP; N = 422, including 136 complete MZ and 75 complete same-sex and opposite-sex DZ pairs). The inclusion of both samples allowed us to examine replicability of general patterns rather than significance of specific effects, which is important given that the size of each sample is not large for a twin study, although it is large for a neuroimaging study. We decomposed the functional connectome of each individual into pairwise correlations between 264 individual regions (referred to as connections) from a widely used and independently derived brain parcellation ([Power et al., 2011]; i.e., 34,716 functional connections). This parcellation was developed to reflect functional distinctions between small parts of cortex [Wig et al., 2011], is accompanied by metadata assigning each region to one of 14 function-specific resting-state communities (e.g., visual network, default network, etc.), and is within a window of optimum dimensionality that maximizes reproducibility [Thirion et al., 2014].

We addressed the first two questions by applying a classic univariate twin model to each connection (see **Materials and Methods - Genetic Models**) to estimate the proportion of variance in connection strength explained by additive genetic influence (A or heritability; the sum of a large number of genetic variants that additively influence a trait), shared environmental influence (C; influences that increase similarity of siblings), and non-shared environmental influence (E; influences that decrease similarity of siblings). The result-

ing high-resolution genetic and environmental maps allowed us to investigate differences between within-network and between-network connections (*question 1*) and also investigate the distribution and patterns of genetic influence within *a priori* resting state networks compared to novel clusters of regions (*question 2*). We addressed *question 3* by examining genetic and environmental decompositions of the correlation between each local connection and a single, global summary measure of the connectome (global efficiency) using bivariate genetic models. The bivariate models allow us to depict where there are genetic influences on connections that are independent of the genetic influences on global efficiency as well as genetic influences on connections that are shared with global efficiency. If residual genetic influence is present across the connectome, this analysis supports high-resolution analysis approaches as independent and complementary to analyses that utilize summary measures of the connectome. Together, these analyses elucidate differences in the genetic and environmental etiology of connections of different type and function, and demonstrating that these genetic influences are independent of genetic influences that have been previously described.

Materials and Methods

The current study is a parallel analysis of resting-state data from a sample of adults recruited from the Colorado Longitudinal Twin Study (referred to as LTS throughout the manuscript) and 422 adults from a publicly available data set from the Human Connectome Project (referred to as HCP throughout the manuscript).

Participants

Participants from the LTS sample were 251 individuals (M_{age} = 28.7 years, SD_{age} = 0.57 years; 97 males) after 15 participants were removed due to excessive movement during the scanning session based on the criteria of greater than 2 mm translation (motion in X, Y, or Z plane) or 2 degrees rotation (roll, pitch, or yaw motion) (n = 14), and failure of the presentation computer to display a fixation cross during the resting scan (n = 1). Of the 251 individuals, there were 54 pairs of MZ twins, 45 pairs of DZ twins, 24 MZ twin singletons, and 28 DZ twin singletons. Singletons are members of twin pairs whose co-twins either did not participate or were excluded from analysis. We did not utilize singletons in the genetic analyses but did utilize them when investigating the relationship between individual differences in local connection strength and global efficiency. All participants were recruited from the Colorado Longitudinal Twin Study which recruited from the Colorado Twin Registry based on birth records. Comparisons with normative data on several measures suggests that the sample is cognitively, academically, and demographically representative of the state of Colorado. Based on self report, the LTS sample is 92.6% White, 5.0% "more than one race", <1% American Indian/Alaskan Native, <1% Pacific Islander, or did not report their race (1.2%). Hispanic individuals composed 9.1% of the sample. Additional information about the LTS sample can be found in [Rhea et al., 2006, Rhea et al., 2013]. Participants were paid $150 for participation in the study or $25 per half hour for those who did not finish the entire three-hour session. The study session involved the administration of behavioral tasks that measured EF ability as well as acquisition of anatomical and functional brain data via magnetic resonance imaging.

HCP participants were 422 individuals (M_{age} = 29.2 years, SD_{age} = 3.46 years, Range = 22 - 35 years; 171 males) selected from the most recent HCP data release because they were part of complete pairs of twins who completed the anatomical and functional imaging components of the study. This subset of HCP participants were 136 MZ pairs and 75 DZ pairs with race reported as 82.7% White, 11.3% Black/African American, 4.5% Asian/Nat. Hawaiian/Other Pacific Is., and <1% Unknown/Not reported, More than one, or Am. Indian/Alaskan Nat. each.

Procedure

Testing took place in a single three-hour session. Following review and obtainment of informed consent, participants were familiarized with the imaging procedures. If both twins of a pair participated on the same day, the twins completed the protocol sequentially (twin order randomized) with the same ordering of behavioral testing and imaging acquisition. The resting-state scan always occurred first in the imaging protocol, before tasks. All study procedures were fully approved by the Institutional Review Board of the University of Colorado Boulder. All participants read and agreed to the informed consent document prior to their initial enrollment in the study and at each follow-up assessment. Testing for the HCP sample participants has been explained thoroughly in prior work [Van Essen et al., 2013].

Brain Imaging

Participants from the LTS sample were scanned in a Siemens Tim Trio 3T scanner. Neuroanatomical data were acquired with T1-weighted MP-RAGE sequence (acquisition parameters: repetition time (TR) = 2400 ms, echo time (TE) = 2.07, matrix size = 320 x 320 x 224, voxel size = 0.80 mm x 0.80 mm x 0.80 mm, flip angle (FA) = 8.00 deg., slice thickness = 0.80 mm). Resting state data was acquired with a 6.25 minute T2*-weighted echo-planar functional scan (acquisition parameters: number of volumes = 816, TR = 460 ms, TE = 27.2 ms, matrix size = 82 x 82 x 56, voxel size = 3.02 mm x 3.02 mm x 3.00 mm, FA = 44.0 deg., slice thickness = 3.00 mm, field of view (FOV) = 248 mm). During the resting-state scan, participants were instructed to relax and stare at a fixation cross while blinking as they normally would. Resting-state acquisition in the HCP sam-

ple is described in detail elsewhere [Smith et al., 2013], but briefly, each participant completed an anatomical and four, 15-minute resting state scans (eyes fixated) in the context of a large imaging and behavioral testing battery. In the current study, the first two 15-minute resting state scans were utilized.

Preprocessing and Connectome Extraction

All processing of LTS brain data was performed in a standard install of FSL build 509 [Jenkinson et al., 2012]. To account for signal stabilization, the first 10 volumes of each individual functional scan were removed, yielding 806 volumes per subject for additional analysis. The functional scans were corrected for head motion using MCFLIRT, FSL's motion correction tool. Brain extraction (BET) was used to remove signal associated with non-brain material (e.g., skull, sinuses, etc.). FSL's FLIRT utility was used to perform a boundary-based registration of each participant's functional scan to his or her anatomical volume and a 6 degree of freedom affine registration to MNI152 standard space. LTS scans were subjected to AROMA, an automated independent components analysis-based, single-subject denoising procedure [Pruim et al., 2014]. Signal was extracted from masks of the lateral ventricles, white matter, and whole brain volume and regressed out along with a set of 6 motion regressors and associated first and second derivatives. Finally, the scans were band-pass filtered (.001 - .08 Hz band).

Preprocessing for HCP data is described in [Glasser et al., 2013]. Briefly, HCP scans were subjected to a minimal preprocessing and FIX, a semi-automated single-subject denoising procedure [Salimi-Khorshidi et al., 2014]. Additionally, we regressed out the mean greyordinate time series from each scan as a proxy for the global signal (as suggested by [Burgess et al., 2016]). HCP scans were band-pass filtered (0.001 - 0.080 Hz band).

For each participant, we extracted the BOLD time series from each of 264, 1 cm spherical ROIs, drawn from [Power et al., 2011], which serve as the nodes for the present analysis. We used these nodes as they are drawn from a meta-analysis of functional activations and have a community structure that agrees with task-based functional networks (i.e., are organized into networks such as default mode network and frontoparietal task control network). One-centimeter spherical ROIs were chosen, as they provide the largest possible size for a given ROI but preclude overlap with neighboring ROIs. Within each participant, all pairwise Pearson's *r* correlations were calculated, yielding a 264 x 264 correlation matrix. All Pearson's *r*-values were subjected to the Fisher's *z* transformation to normalize the variance in correlation values. All analyses used the per-participant 264 x 264 z-correlation matrices as input. Bivariate analyses utilized a global summary measure of each participant's connectivity matrix which was calculated as the reciprocal of the average shortest path length between all 264 regions as calculated on a proportionally thresholded (15%) connectivity matrix using the Python package networkx [Hagberg et al., 2008]. Calculation and manipulation of connectivity matrices as well as plotting was also done in Python using the Pandas [McKinney, 2010], Seaborn (http://seaborn.pydata.org/), and Matplotlib packages [Hunter, 2007].

Genetic Models

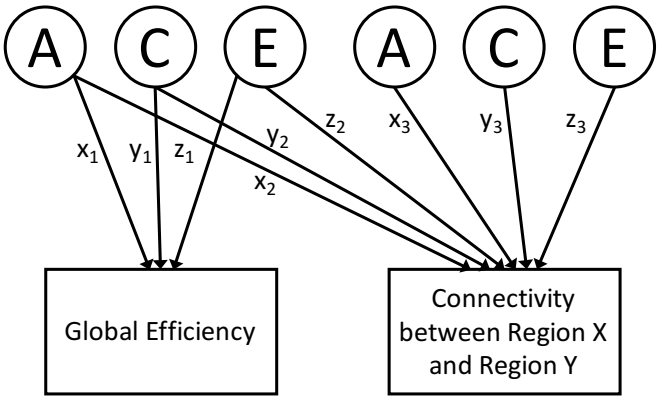

Figure 1. **Bivariate Cholesky model**. Additive genetic (A), shared environmental (C), and non-shared environmental (E) latent variables (left side) predicting global efficiency (via paths x_1, y_1, and z_1) and functional connectivity (via paths x_2, y_2, and z_2). Functional connectivity has residual A, C, and E influences (right side). The variance explained by each influence is obtained by squaring the paths (e.g., x_3, y_3, and z_3). The phenotypic correlation of global efficiency and local functional connectivity that is predicted by share genetic influences (Gr) is obtained by multiplying $x_1 * x_2$. The univariate model of global efficiency is equivalent to left side of the figure (i.e., removing the local connectivity measure).

All genetic analyses conducted were run as structural equation models in R through the OpenMx [Boker et al., 2011] and UMX packages [Bates, 2017]. As all measures were continuous, these models utilized maximum likelihood estimation [Bentler and Weeks, 1980]. Univariate genetic models were run on each connection. A univariate model decomposes total phenotypic variation in a connection into additive genetic (A), shared environmental (C), and non-shared environmental (E) components. MZ twins share all of their genes, whereas DZ twins share on average 50% of their genes by descent, and both types are reared together. Genetic influences (A) are indicated when the MZ twin correlation is higher than the DZ correlation; shared environmental influences (C) are indicated when the DZ correlation is greater than half the MZ correlation; and non-shared environmental influences (E), which can include

measurement error, are indicated when the MZ correlation is less than unity.

For the association between connections and global efficiency, we utilized bivariate Cholesky decompositions (pictured in **Figure 1**). The Cholesky decomposition is a common form of bivariate twin analysis, and can be used to calculate the genetic correlation and correlation predicted from A, C, and E overlap (for more see [Neale and Cardon, 1992]). In this Cholesky decomposition, the first set of A, C, and E latent variables predicting global efficiency are also allowed to predict the local connection (via paths x_2, y_2, and z_2), and the local connection also has residual A, C, and E variances (obtained by squaring paths x_3, y_3, and z_3). Two results from this analysis are of interest. First, the matrix of residual A variances (i.e., squared x_3 paths) enables us to ascertain whether the finely detailed genetic map of the connectome is simply a redescription of a global summary measure of the connectome. That is, it depicts where there are genetic influences on connections that are independent of the genetic influences on global efficiency. If residual genetic influence is present across the connectome, this analysis supports high-resolution analysis approaches as independent and complementary to analyses that utilize summary measures of the connectome. Second, the phenotypic correlation between global efficiency and a connection can be decomposed into the correlation explained by genetic influences common to both traits ($Gr = x_1 * x_2$), that explained by common shared environmental influences ($Cr = y_1 * y_2$), and that explained by common non-shared environmental influences ($Er = z_1 * z_2$), with these three components summing to the total phenotypic correlation. The Gr matrix thus reveals where the phenotypic correlations between global efficiency and local connections are explained by shared genetic influences.

Clustering Analysis

We clustered patterns of heritability estimates - rows/columns of **Figure 2a**. Ward clustering was implemented in Python using the Fastcluster [Mullner, 2013] and Scikit-learn packages [Pedregosa et al., 2011]. We applied clustering to the 264 x 264 matrix of A estimates to find 2-100 clusters of regions. To estimate the stability of each clustering solution, we calculated the silhouette score for each sample and averaged all scores for each clustering solution (**Figure 4a**). The silhouette score compares the distance between a region and other members of its cluster to the distance between that region and the nearest neighboring cluster in similarity space. In the current analysis, similarity was defined as the Euclidean distance between two regions' vectors of heritability estimates. The silhouette analysis The silhouette analysis revealed several "stable" solutions in which the average silhouette score reached a local maximum, as seen in the peaks of **Figure 4a** at solutions of k = 3, 11, 26, 46, and 65. We describe the clustering results at the coarsest levels in the main body because they are a valid demonstration of novel genetic communities without the added complexity of describing 26, 46, and 65 clusters. We provide a description of a high dimensionality solution in **Supporting Information**.

Data Availability

Upon publication, A, C, and E estimates for all pairwise connections for the HCP sample will be available for download at https://github.com/AReineberg/genetic_connectome

Results

Group Average Connectomes

Visual comparison of mean phenotypic connectivity matrices for each sample to one another and to matrices reported in prior work using independent samples (e.g., Figure 3 of [Cole et al., 2014]; Figure 2 of [Reineberg and Banich, 2016]) reveals striking similarity, especially in the prominence of resting-state networks along the diagonals (**Figure S1** [LTS] and **Figure S2** [HCP]).

Univariate Twin Models

Connection-wise estimates of additive genetic influence for the LTS and HCP samples are shown in **Figure 2a, b** (lower triangles). In the LTS sample, additive genetic influence was moderate and bimodally distributed across the connectome such that 18,077 of 34,716 unique connections were estimated as having approximately zero heritability while a separate, positively-skewed distribution described the heritability of 16,639 connections (M = 0.140, SD = 0.097, Skew = 0.748, Range = 0.010 - 0.574). Similarly, for the HCP sample, 13,546 of 34,716 unique connections were estimated as having approximately zero heritability while a separate, positively-skewed distribution described the heritability of 21,170 connections (M = 0.131, SD = 0.083, Skew = 0.578, Range = 0.010 - 0.504).

Shared environmental influences generally explained less variance than genetic influences, as shown in **Figure 2a, b** (upper triangles). In the LTS sample, shared environmental influence was weak to moderate and bimodally distributed across the connectome such that 22,520 of 34,716 unique connections were estimated as having approximately zero shared environmental influence while a separate, positively-skewed distribution described the shared environmental influence of 12,196 connections (M = 0.106, SD = 0.076, Skew = 0.834, Range = 0.01 - 0.470). Similarly, for the HCP sample, 20,598 of 34,716 unique connections were estimated as having approximately zero shared environmental influence while a separate, positively-skewed distribution described the shared environmental influence of 14,118 connections (M = 0.098, SD = 0.065, Skew = 0.690, Range = 0.01 -

0.368). Although C estimates were low, several pieces of the connectome have moderate shared environmental influence such as within-default network connections, within-sensory somatomotor connections, and default-to-other connections.

In both samples, non-shared environmental influences, which include measurement error, were high across the entire connectome (M_{LTS} = 0.893, SD_{LTS} = 0.102; M_{HCP} = 0.877, SD_{HCP} = 0.087) and negatively skewed ($Skew_{LTS}$ = -0.832, $Skew_{HCP}$ = -0.468). Connection-wise estimates of non-shared environmental influences are shown in the lower and upper triangles of **Figure S3** for the HCP and LTS samples, respectively. It should be noted that even though E estimates include measurement error, reliability of connections was tested for the HCP sample and found to be high (M = 0.849, SD = 0.061; see **Supporting Information - Reliability**), suggesting that the high E estimates across the connectome are unlikely to solely reflect random measurement error.

Within- and Between-Network Connections

To examine whether high resolution mapping of genetic influence reveals differences in within- versus between-network connections, (*question 1*), we investigated heritability estimates for connections of those types. First, we considered average heritability across all connections considered to be within the same *a priori* network versus all between-network connections. In both samples, within-network connectivity was more heritable than between network connectivity (**Table S1a**; whole connectome results). This effect was present even when controlling for the estimated test-retest reliability of each connection in the HCP sample (see **Supporting Information - Reliability**).

We also quantified differences in heritability for within and between network connections at the level of each *a priori* network, as shown in **Figure 3**. In both samples, within-network connections tended to be more heritable on average than between-network connections. In both samples, the default, sensory-somatomotor hand, dorsal attention, visual network had significantly higher heritability for within- than between-network connections (**Table S1b**), but the subcortical and uncategorized networks had significantly higher heritability for between- than within-network connections. Although within-network connections tended to be more heritable on average than between-network connections, the distributions of between-network connections tended to be more positively skewed, perhaps suggesting there are a minimal number of highly heritable between-network connections.

Finally, we quantified differences in heritability for within- and between-network connections at the level of regions (each of the 264 regions of the parcellation). Of regions that had significantly higher heritability for within- than between-network connections in the LTS (n = 61) and HCP (n = 83) samples, 32 regions showed the effect in both samples (See **Table S2**). These regions were predominantly from the sensory-somatomotor (6), default (15), and visual (10) networks, as well as a single salience network region. None of the regions that had significantly higher heritability for between-network connections in either sample showed the effect in both samples.

Clustering Reveals Unique Genetic Communities

The distribution of A estimates within *a priori* networks (**Figure 3**) partially answers the question of whether genetic influences are homogeneous within any given resting-state community (*question 2*). While these distributions of heritability estimates suggest heterogeneity in the magnitudes of genetic influence, patterns of genetic influence for the different regions of each network are also of interest. Thus, we conducted a data-driven hierarchical (Ward) clustering analysis to group together regions with similar patterns of heritable connectivity. Ward clustering groups rows of the additive genetic influence matrix (**Figure 2**) that show similar patterns of heritable connectivity with all other regions, while attempting to minimize within-cluster variance. This analysis could reveal that the 264 regions cluster together in a manner similar to *a priori* networks or in a novel way (e.g., a cluster of regions with highly heritable connectivity to some default and frontoparietal network regions, but minimally heritable connectivity to other regions). We analyzed average silhouette scores for clustering solutions (i.e., *k*-values) from 2 to 100, shown in **Figure 4a**, and discovered stable solutions at *k*-values of 3, 11, 26, 46, and 65 in the HCP sample and *k*-values of 2, 3, and 4 in the LTS sample.

Because stable solutions existed at a *k*-value of 3 in both samples, we explored that clustering solution first. This level provides the highest level overview of patterns of genetic influence across the connectome. The three clusters from the $k = 3$ stable solution will be referred to as super-clusters throughout the remainder of the manuscript. The 3-cluster solution for the HCP sample is shown in **Figure 4b**. **Figure 4c** provides a summary of how the three super-clusters differ on genetic influence. **Figure 5** provides an overview of both the spatial location of the regions in each super-cluster (a-c) and also the composition of those super-clusters in terms of the regions assignments to *a priori* networks (right-most column of **Figure 4a-c**).

Overall, the results revealed that regions do not genetically cluster based on an *a priori* notion of resting-state community structure. Super-cluster 1 was composed of 42 regions with connectivity very minimally influenced by genes (less than 5% variance explained by genetic sources, on average). While regions in super-cluster 1 belonged to all *a priori* networks, subcortical and unassigned regions made up a large percentage of the super-cluster. Super-cluster 2 regions had especially heritable connectivity to default, frontoparietal, salience, dorsal attention, and visual network re-

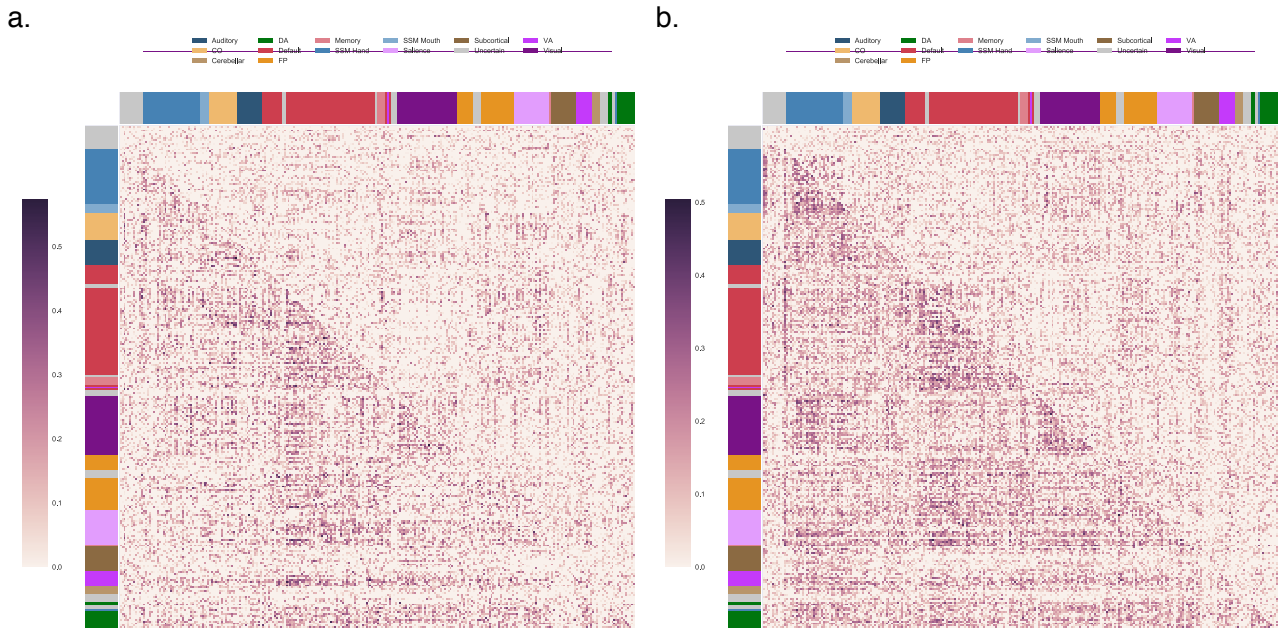

Figure 2. **Connection-wise estimates of additive genetic (A) and shared environmental (C) influences**. Matrices contain estimates from univariate twin models with the spatial location of each cell (estimate) corresponding to the functional connection between two regions. Assignment to *a priori* networks is represented by colored bars along x and y axes. Because triangles are redundant, different estimates are displayed in upper and lower triangles. a. LTS sample A and C estimates are in the lower and upper triangles, respectively. b. HCP sample A and C estimates are in the lower and upper triangles, respectively. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

gions. Super-cluster 2 was composed of 100 regions that can best be summarized as the majority of the default network as well many fronto-parietal regions, among others. Super-cluster 3 regions had especially heritable connectivity to pure sensory and sensory-somatomotor regions as well as moderately heritable connectivity to regions from cognitive networks (default, cingulo-opercular, frontoparietal, and attention networks). Super-cluster 3 was composed of 121 regions from a variety of pure sensory (e.g., visual) and sensory-somatomotor networks. The $k = 3$ solution of the LTS sample maps closely on to the $k = 3$ solution of the HCP sample with only a few notable difference (discussed further in **Supporting Information - Clustering**).

Higher-order clustering solutions from the HCP sample give insight into how these large genetic communities break down into more specific patterns of genetic influence. **Figure SX** shows the composition of sub-clusters from the 11-cluster solution. Super-cluster 1 did not break down in this higher dimensionality solution. Super-cluster 2 broke down into three sub-clusters (2-4). Notably, each of these sub-clusters (2, 3, and 4) contained a roughly equal percentage of the default network, supporting the conclusion that the default network contains several sets of regions that have unique patterns of heritable connectivity across the connectome. Super-cluster 3 broke into several smaller sub-clusters in the 11-cluster solution. Some of these sub-clusters (7, 8, 10, and 11) contained regions from a single modality (e.g., visual and sensory-somatomotor regions), while others were a conglomeration of regions from several *a priori* networks mixed between higher-level cognition and lower-level cognition network types.

Bivariate Analyses

To ascertain whether global efficiency captures most of the genetic variance in local connectivity (*question 3*), we conducted a bivariate genetic analysis of local connections and global efficiency. These bivariate analyses quantify the degree to which local functional connectivity is genetically separable from a summary measure of the connectome.

We found global efficiency was more heritable in the HCP sample than in the LTS sample. In the HCP sample, 36.6% of the variance in global efficiency was attributed to additive genetic influences. In the LTS sample, 6.3% of the variance in global efficiency was attributed to additive genetic influences. Given this difference, estimates of global efficiency may be sensitive to the short scan length or the smaller sample size of the LTS data.

Regarding *question 3*, residual additive genetic influence after accounting for genetic influence shared with global efficiency (i.e., residual A variances or specific A) for each func-

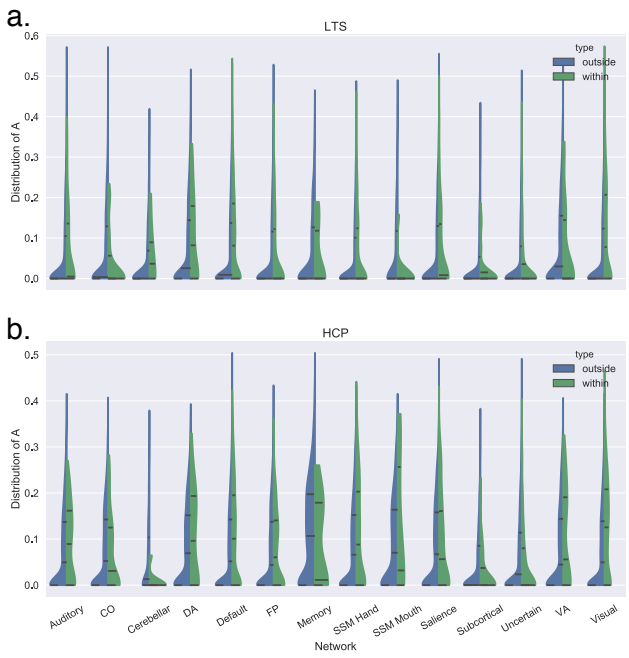

Figure 3. **Connection-wise estimates of additive genetic influence summarized by *a priori* networks**. Violin plots of distribution of additive genetic (A) estimates grouped by *a priori* networks reveal higher heritability for within- than between-network connections in both LTS (a) and HCP (b) samples. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

tional connection (i.e., squared x_3 paths; see **Figure 1**) was distributed bimodally in a manner similar to the univariate A estimates. In the HCP sample, approximately 64.5% of connections (22,400) had 0 residual A variance with a separate positively skewed distribution describing the remaining 12,316 connections (M = 0.117, SD = 0.073, Skew = 0.726, Range = 0.010 - 0.524). In the LTS sample, approximately 75.8% of connections (26,318) had 0 residual A variance, with a separate positively skewed distribution describing the remaining 8,398 connections (M = 0.127, SD = 0.087, Skew = 0.926, Range = 0.010 - 0.520). The matrix of specific A estimates is displayed in **Figure 6** for the HCP and LTS samples (lower and upper triangle respectively). These results suggest that there are genetic influences specific to local functional connections. That is, the genetic influences on finer-grained measures of functional connections are not simply reflecting the genetic influences on global efficiency. Future work should synthesize the spatial distribution of these global and specific effects.

The bivariate models also serve as an example of how a multivariate signal (here, 34,716 connections) associated with individual differences in a trait (here, global efficiency)

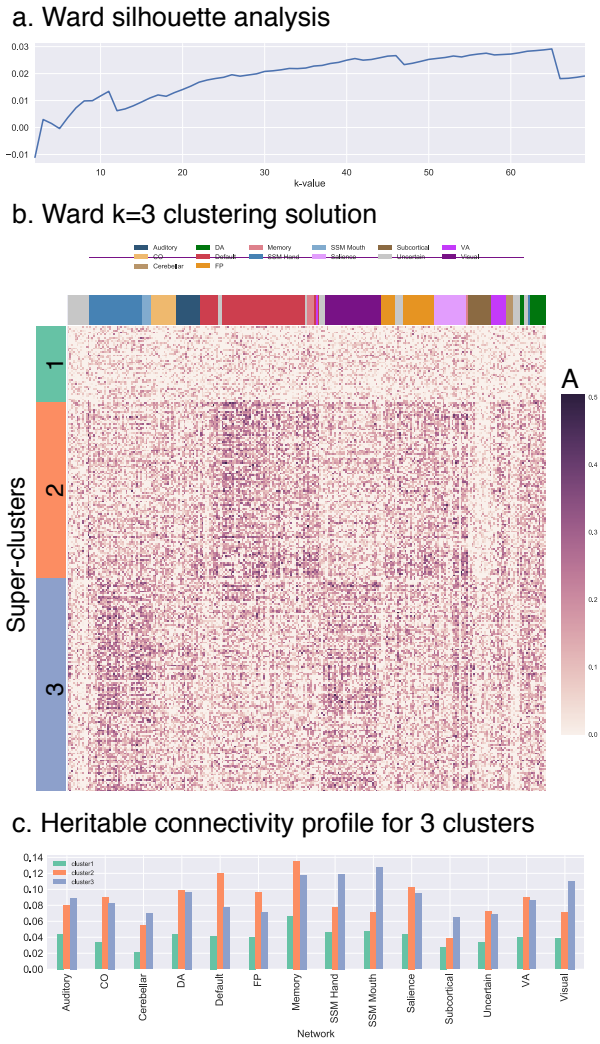

Figure 4. **Ward 3-cluster solution**. Row-wise clustering of HCP additive genetic (A) estimates reveals several stable clustering solutions of regions with similar patterns of connectivity heritability. Super-clusters (k = 3) are described in detail. a. Silhouette analysis revels stable clustering solutions at k-values of 3, 11, 26, 46, and 65. b. Clustered version of HCP A estimates for k = 3 solution. c. Average connectivity heritability for super-clusters 1, 2, and 3 organized by *a priori* network of connection. Super-cluster 1 is characterized by weakly heritable connections. Super-clusters 2 and 3 have high connectivity heritability globally with particularly high heritability of higher level cognitive (Super-cluster 2) and sensory-somatomotor (Super-cluster 3) connections. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

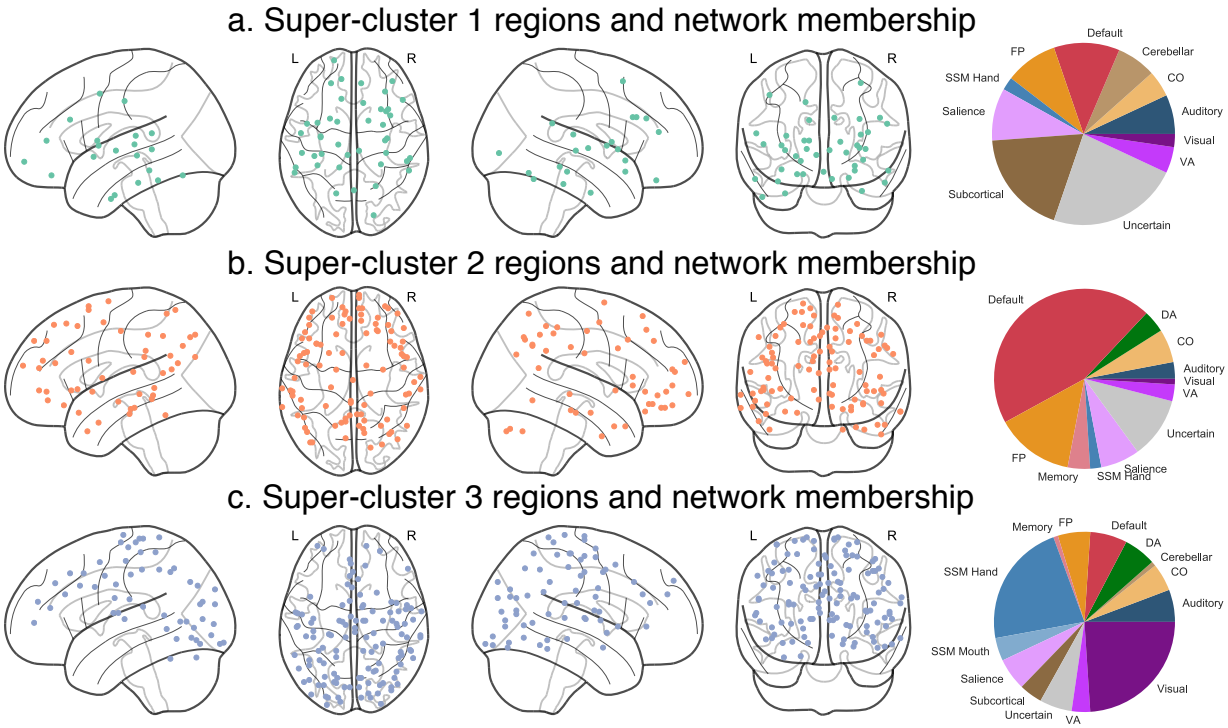

Figure 5. **Ward 3-cluster summary**. Spatial location of regions from super-clusters 1-3 of $k = 3$ solution. a. Super-cluster 1 regions were widely distributed across the brain. b. Super-cluster 2 regions were located across lateral prefrontal, lateral parietal, mid and anterior temporal, midline frontal, and cingulate areas. c. Super-cluster 3 regions were located primarily in sensory and somatomotor areas. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

can be broken into genetic and environmental components. We describe this decomposition as an exploratory follow-up to *question 3* in **Supporting Information - Bivariate Analysis**, which can serve as a model for future work investigating the etiology of the relationship between functional connectivity and other phenotypes. Briefly, we found individual differences in a subset of the connectome were phenotypically associated with variation in global efficiency, so we focused our exploration on those connections. Genetic influences were highly heterogeneous in their influence on the correlation between local and global measures of connectivity, non-shared environmental influences were homogeneous in their influence on the correlation between global efficiency and local functional connectivity, and shared environment did not play as large a role as genetic and non-shared environmental influence but globally predicted negative correlations between connection strength and global efficiency, even for connections showing positive phenotypic correlations between connection strength and global efficiency.

Discussion

Across all analyses, we found converging evidence of etiological heterogeneity in the functional connectome. High-resolution mapping reveals a distribution of genetic and environmental influence that may be missed by approaches that summarize functional connectivity at the level of larger ROIs, networks, and global summary measures of the connectome. More specifically, we found differences in genetic influences for connections of different type (i.e., higher heritability of connections between regions of the same functional network versus between regions of different functional networks). This pattern was present across the whole connectome and especially for the default, sensory/somatomotor, dorsal attention, and visual networks. This result provides preliminary evidence that the organization of the brain into networks based on function may be driven by genetic influences on connections between regions involved in the same processes. Interestingly, prior work has established specific patterns of gene expression within functional networks [Richiardi and Altmann, 2015], a possible mechanism linking these observations of genetic influence to specific functions.

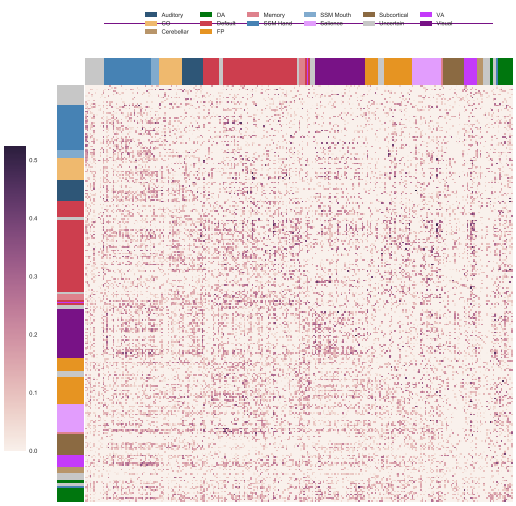

Figure 6. **Connection-wise estimates of residual additive genetic (A) variance.** Estimates of residual genetic variance for each connection after accounting for genetic influence shared with global efficiency. Estimates for HCP and LTS samples are in lower and upper triangles, respectively. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

Although a number of our analyses were summarized by *a priori* functional networks, the broad range of genetic estimates across the connectome led us to question whether alternative groupings could better describe patterns of heritability in the connectome. A clustering procedure revealed a highly novel finding: Regions grouped together based on patterns of heritable connectivity at a level that was superordinate to that of classic resting-state communities. In both samples, we found stable super-clusters of regions. Most notably, a set of "cognitive" and "sensory/somatomotor" regions had characteristic patterns of highly heritable connections to regions of higher-level (e.g., default network) and lower-level (e.g., visual) functions, for the cognitive and sensory/somatomotor clusters respectively. Although the description of the these super-clusters as higher and lower-level is likely an oversimplification, it is a worthwhile descriptive tool until future work dissects the role of these sets of brain regions. In both samples, we found evidence of a super-cluster of regions that was almost entirely environmentally influenced. Analyses of connection-wise reliability (very high in all connections in the HCP sample) suggest that these non-shared environmental influence estimates do not simply reflect random measurement error. Thus, future work should seek a more thorough understanding of the environmental factors influencing these primarily sensory, subcortical, and cerebellar connections. Stable clustering solutions were also found at levels of granularity similar to classic resting-state communities, but, interestingly, these genetic clusters were quite dissimilar to the *a priori* networks. Notably, regions from the default network broke into several sub-clusters which were differentiated on heritability of connectivity to other default network regions and to regions of other networks such as the frontoparietal network. Future work should dissect these finer-grained parcels in more detail as well as explore the other stable clustering solutions. The latter may reveal small communities with highly characteristic patterns of heritable connectivity which may not align with known clusters of regions based on community detection performed on phenotypic functional connectivity.

A final novel finding was that the local connections showed genetic influences independent from genetic influences on a global summary measure of the brain. Residual genetic influence justifies analysis at the level of small regions and is an important commentary on an ongoing debate about the proper level of analysis of connectivity, suggesting all levels may be complementary. A practical application of this evidence of residual genetic influence would be to the study of multivariate functional connectivity signals as a predictor of individual differences in some cognitive ability or clinical variation (i.e., "fingerprinting" or "connecto-typing"). Our results suggest influence on a whole connectome signal will be diverse and not accurately represented in global summary measures of the connectome. Future work could, for example, apply the same bivariate analysis we describe above and in **Supporting Information - Bivariate Analyses** to the relationship between local connections and clinical symptomatology and/or cognitive abilities. Such an application could identify novel brain-based candidate endophenotypes or focus intervention studies to novel locations, similar to studies from the neuroanatomical/clinical endophenotyping literature in which bivariate genetic models have been used to identify the genetically-influenced neurobiological underpinning of disorders such as major depressive disorder [Glahn et al., 2012] and the genetically-influenced neurobiological underpinnings of schizophrenia that are shared with other psychiatric disorders [Lee et al., 2016].

Our approach is not without caveats. Although we highlighted overlapping results in both samples, we also observed some sample-specific results. Notably, shared environmental influence was lower across the entire connectome in the LTS sample than in the HCP sample. Although it is not unusual to find a lack of shared environmental influence (e.g., in the anatomical MRI [Eyler et al., 2011] and cognitive literatures), sample differences could be due to reliability differences in the measurement of resting-state functional connectivity (e.g., scan time of 30 minutes [HCP] versus 6 minutes [LTS] is known to produce more reliable results [Gordon et al., 2017]) and/or due to demographic differences in the two samples (e.g, the Colorado sample is less racially diverse and sampled from higher socioeconomic sta-

tus communities than the HCP sample). Socioeconomic status differences could certainly explain sample differences in the current study given prior work showing elevated shared environmental influence on variation in IQ for individuals near or below the poverty line [Turkheimer et al., 2003]. Effects of SES in a subset of the HCP sample have been partially explored previously and shown to influence brains connectivity [Smith et al., 2015]. Regarding modeling of genetic influences, a small literature suggests classic twin modeling procedures may bias estimates (upward in the case of A and downward in the case of E) when compared to models that do not impose boundary constraints on parameters [Carey, 2005]. Future work should compare these approaches and report notable differences, if any, in the genetic profile of affected connections.

Overall, we demonstrate the utility of fine-grained A, C, and E estimates by showing the genetic organization of the brain is diverse and not as one would expect based solely off the classic functional organization of the phenotypic connectome. Our analysis sits in a continuum of dimensionality reductions that spans multiple levels of brain organization (i.e., from global summary measures to voxels), so, obviously one must ask if genetic neuroimaging studies should continue to assess the etiology of finer grained parcellations in the future. Our demonstration of residual genetic variance for local connections in the bivariate analyses certainly demonstrates the added value of a fine-grained approach in addition to a single summary measure of the connectome. But, our results also suggest a trade-off between reliability and interpretability/application: Large networks maximize heritability estimates but are of imprecise function and cannot be used to dissect the etiology of highly dimensional signals that are most useful for predictive modeling. Parcellations in the range of 200-500 might be recommended for region-based approaches in the future because there are numerous well-vetted atlases [Power et al., 2011, Craddock et al., 2012, Gordon et al., 2016] designed to differentiate homogeneous functional brain units while maximizing reliability (which could become an issue in voxel-based approaches). There is still room for determining the best functional parcellation scheme among these possible alternatives, with genetic etiology as one possible mechanism for evaluating the quality/usefulness of the parcellations. In conclusion, our approach has important implications for investigations of neuroimaging-based biomarkers by 1. quantifying which pieces of the connectome are heritable and thus can be investigated as a potential endophenotype or marker of genetic risk; 2. serving as a model for future studies seeking a greater understanding of a broad literature of traits; and 3. establishing the foundation of a taxonomy of functional connections based on genetic influence.

Acknowledgements

This work was supported by NIH R01MH063207 awarded to PI Friedman.

References

[Andrews-Hanna, 2011] Andrews-Hanna, J. R. (2011). The brain's default network and its adaptive role in internal mentation. *The Neuroscientist*, pages 1–20.

[Bates, 2017] Bates, T. C. (2017). umx: A help package for structural equation modeling in openmx.

[Bentler and Weeks, 1980] Bentler, P. M. and Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, 45(3):289–308.

[Boker et al., 2011] Boker, S., Neale, M. C., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T. C., Paras, M., and Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, 76(2):306–317.

[Burgess et al., 2016] Burgess, G. C., Kandala, S., Nolan, D., Laumann, T. O., Power, J., Adeyemo, B., Harms, M. P., Petersen, S. E., and Barch, D. M. (2016). Evaluation of Denoising Strategies To Address Motion-Correlated Artifact in Resting State fMRI Data from the Human Connectome Project. *Brain Connectivity*, page brain.2016.0435.

[Carey, 2005] Carey, G. (2005). Cholesky problems. *Behavior Genetics*, 35(5):653–665.

[Cole et al., 2014] Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., and Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, 83(1):238–51.

[Craddock et al., 2012] Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., and Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8):1914–1928.

[Eyler et al., 2012] Eyler, L. T., Chen, C.-h., Panizzon, M. S., Fennema-notestine, C., Neale, M. C., Jak, A., Jernigan, T. L., Fischl, B., Franz, C. E., Lyons, M. J., Grant, M., Prom-wormley, E., Seidman, L. J., Tsuang, M. T., Fiecas, M. J. A., Dale, A. M., and Kremen, W. S. (2012). A comparison of heritability maps of cortical surface area and thickness and the influence of adjustment for whole brain measures: A magnetic resonance imaging twin study. *Twin research and human genetics : the official journal of the International Society for Twin Studies*, 15(3):304–314.

[Eyler et al., 2011] Eyler, L. T., Prom-Wormley, E., Panizzon, M. S., Kaup, A. R., Fennema-Notestine, C., Neale, M. C., Jernigan, T. L., Fischl, B., Franz, C. E., Lyons, M. J., Grant, M., Stevens, A., Pacheco, J., Perry, M. E., Schmitt, J. E., Seidman, L. J., Thermenos, H. W., Tsuang, M. T., Chen, C. H., Thompson, W. K., Jak, A., Dale, A. M., and Kremen, W. S. (2011). Genetic and environmental contributions to regional cortical surface area in humans: A magnetic resonance imaging twin study. *Cerebral Cortex*, 21(10):2313–2321.

[Finn et al., 2015] Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671.

[Fornito et al., 2011] Fornito, A., Zalesky, A., Bassett, D. S., Meunier, D., Ellison-Wright, I., Yucel, M., Wood, S. J., Shaw, K., O'Connor, J., Nertney, D., Mowry, B. J., Pantelis, C., and Bullmore, E. T. (2011). Genetic influences on cost-efficient organization of human cortical functional networks. *J Neurosci*, 31(9):3261–3270.

[Fox and Raichle, 2007] Fox, M. D. and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews. Neuroscience*, 8(9):700–11.

[Ge et al., 2017] Ge, T., Holmes, A. J., Buckner, R. L., Smoller, J. W., and Sabuncu, M. R. (2017). Heritability analysis with repeat measurements and its application to resting-state functional connectivity. *Proceedings of the National Academy of Sciences*, 114(21):201700765.

[Glahn et al., 2012] Glahn, D. C., Curran, J. E., Winkler, A. M., Carless, M. A., Kent, J. W., Charlesworth, J. C., Johnson, M. P., Göring, H. H. H., Cole, S. A., Dyer, T. D., Moses, E. K., Olvera, R. L., Kochunov, P., Duggirala, R., Fox, P. T., Almasy, L., and Blangero, J. (2012). High dimensional endophenotype ranking in the search for major depression risk genes. *Biological Psychiatry*, 71(1):6–14.

[Glahn et al., 2010] Glahn, D. C., Winkler, A. M., Kochunov, P., Almasy, L., Duggirala, R., Carless, M. a., Curran, J. C., Olvera, R. L., Laird, a. R., Smith, S. M., Beckmann, C. F., Fox, P. T., and Blangero, J. (2010). Genetic control over the resting brain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(3):1223–8.

[Glasser et al., 2013] Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M. A., Polimeni, J. R., Van Essen, D. C., and Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–124.

[Gordon et al., 2016] Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., and Petersen, S. E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, 26(1):288–303.

[Gordon et al., 2017] Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M., and Dosenbach, N. U. F. (2017). Precision Functional Mapping of Individual Human NeuroResource Precision Functional Mapping of Individual Human Brains. *Neuron*, 95(4):791–807.e7.

[Greicius, 2008] Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology*, 21(4):424–30.

[Hagberg et al., 2008] Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, (SciPy):11–15.

[Hall and Smoller, 2010] Hall, M.-H. and Smoller, J. W. (2010). A new role for endophenotypes in the GWAS era: Functional characterization of risk variants. *Harvard Review of Psychiatry*, 18(1):67–74.

[Heuvel et al., 2012] Heuvel, M. P. V. D., Kahn, R. S., Goñi, J., and Sporns, O. (2012). High-cost, high-capacity backbone for global brain communication. *Proceedings of the National Academy of Sciences*, 109(28):11372–11377.

[Hunter, 2007] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

[Jenkinson et al., 2012] Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *NeuroImage*, 62(2):782–90.

[Kendler and Neale, 2010] Kendler, K. S. and Neale, M. C. (2010). Endophenotype: a conceptual analysis. *Molecular Psychiatry*, 15(8):789–797.

[Lee et al., 2016] Lee, P. H., Baker, J. T., Holmes, A. J., Jahanshad, N., Ge, T., Jung, J.-y., Cruz, Y., Manoach, D. S., Hibar, D. P., Faskowitz, J., Mcmahon, K. L., de Zubicaray, G. I., Martin, N. H., Wright, M. J., Buckner, R., Roffman,

J., Thompson, P. M., and Smoller, Jordan, W. (2016). Partitioning heritability analysis reveals a shared genetic basis of brain anatomy and schizophrenia. *Molecular Psychiatry*, 21(12):1680–1689.

[McKinney, 2010] McKinney, W. (2010). Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, pages 51–56.

[Mullner, 2013] Mullner, D. (2013). fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *Journal of Statistical Software*, 53(9).

[Neale and Cardon, 1992] Neale, M. C. and Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families*.

[Panizzon et al., 2014] Panizzon, M. S., Vuoksimaa, E., Spoon, K. M., Jacobson, K. C., Lyons, M. J., Franz, C. E., Xian, H., Vasilopoulos, T., and Kremen, W. S. (2014). Genetic and environmental influences on general cognitive ability: Is g a valid latent construct? *Intelligence*, 43(1):65–76.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., and Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Petrill, 1997] Petrill, S. A. (1997). Molarity versus modularity of cognitive functioning? A behavioral genetics perspective. *Perspectives on Psychological Science*, 6(4):96–99.

[Power et al., 2011] Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., and Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, 72(4):665–78.

[Pruim et al., 2014] Pruim, R. H., Mennes, M., van Rooij, D., Llera Arenas, A., Buitelaar, J. K., and Beckmann, C. F. (2014). ICA-AROMA: A robust ICA-based strategy for removing motion artifact from fMRI data. *Submitted for Publication*.

[Reineberg and Banich, 2016] Reineberg, A. E. and Banich, M. T. (2016). Functional connectivity at rest is sensitive to individual differences in executive function: A network analysis. *Human Brain Mapping*, 00.

[Rhea et al., 2006] Rhea, S.-A., Gross, A. a., Haberstick, B. C., and Corley, R. P. (2006). Colorado twin registry. *Twin research and human genetics : the official journal of the International Society for Twin Studies*, 9(6):941–949.

[Rhea et al., 2013] Rhea, S.-A., Gross, A. A., Haberstick, B. C., and Corley, R. P. (2013). Colorado Twin Registry - An Update. *Twin research and human genetics : the official journal of the International Society for Twin Studies*, 16(1):1–14.

[Richiardi and Altmann, 2015] Richiardi, J. and Altmann, A. (2015). Correlated gene expression supports synchronous activity in brain networks. *Science*, 348(6240):11–14.

[Rosenberg et al., 2016] Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., and Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature neuroscience*, 19(1).

[Salimi-Khorshidi et al., 2014] Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90(0):449–468.

[Shen et al., 2017] Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., and Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, 12(3):506–518.

[Sinclair et al., 2015] Sinclair, B., Hansell, N. K., Blokland, G. A., Martin, N. G., Thompson, P. M., Breakspear, M., de Zubicaray, G. I., Wright, M. J., and McMahon, K. L. (2015). Heritability of the network architecture of intrinsic brain functional connectivity. *NeuroImage*, 121:243–252.

[Smith et al., 2013] Smith, S. M., Andersson, J., Auerbach, E. J., Beckmann, C. F., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. a., Griffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J. D., Salimi-Khorshidi, G., Snyder, A. Z., Vu, A., Woolrich, M. W., Xu, J., Yacoub, E., Ugurbil, K., Van Essen, D. C., and Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage*.

[Smith et al., 2009] Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., and Beckmann, C. F. (2009). Correspondence of the brain' s functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):13040–13045.

[Smith et al., 2015] Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E. J., Glasser, M. F., Ugurbil, K., Barch, D. M., Essen, D. C. V., and Miller, K. L.

(2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, (September):1–7.

[Thirion et al., 2014] Thirion, B., Varoquaux, G., Dohmatob, E., and Poline, J. B. (2014). Which fMRI clustering gives good brain parcellations? *Frontiers in Neuroscience*, 8(8 JUL):1–13.

[Turkheimer et al., 2003] Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., and Gottesman, I. I. (2003). Socioeconomic status modified heritability of IQ in young children. *Psychological Science*, 14(6):623–628.

[Vaidya and Gordon, 2013] Vaidya, C. J. and Gordon, E. M. (2013). Phenotypic variability in resting-state functional connectivity: current status. *Brain connectivity*, 3(2):99–120.

[van den Heuvel et al., 2013] van den Heuvel, M. P., van Soelen, I. L. C., Stam, C. J., Kahn, R. S., Boomsma, D. I., and Hulshoff Pol, H. E. (2013). Genetic control of functional brain network efficiency in children. *European Neuropsychopharmacology*, 23(1):19–23.

[Van Essen et al., 2013] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., and Consortium, W.-M. H. (2013). The WU-Minn Human Connectome Project: An Overview. *NeuroImage*, 15(18):62–79.

[Wig et al., 2011] Wig, G. S., Schlaggar, B. L., and Petersen, S. E. (2011). Concepts and principles in the analysis of brain networks. *Annals of the New York Academy of Sciences*, 1224:126–46.

[Yang et al., 2016] Yang, Z., Zuo, X.-N., McMahon, K. L., Craddock, R. C., Kelly, C., de Zubicaray, G. I., Hickie, I., Bandettini, P. A., Castellanos, F. X., Milham, M. P., and Wright, M. J. (2016). Genetic and Environmental Contributions to Functional Connectivity Architecture of the Human Brain. *Cerebral cortex (New York, N.Y. : 1991)*, 26(5):2341–52.

[Yeo et al., 2011] Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., and Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(3):1125–65.

[Zhang and Raichle, 2010] Zhang, D. and Raichle, M. E. (2010). Disease and the brain's dark energy. *Nature reviews. Neurology*, 6(1):15–28.

Supporting Information

Mean Phenotypic Connectivity

Figures S1 and **S2** show functional connectivity averaged across all individuals in the LTS and HCP samples, respectively.

Univatiate Twin Models - Non-shared Environmental Influence

Connection-wise estimates of non-shared environmental influence for the LTS and HCP samples are shown in **Figure S3**.

Within versus Between-network Connectiity

We quantified differences in heritability for within and between network connections summazired at various levels of analysis - averaged across the whole connectome (**Table S1a**), averaged for within- and between-network connections for 14 *a propri* networks of interest (**Table S1b**), and for each of 264 regions that are part of the parcellation utilized in the current study (**Table S2**). These averages included all estimates, including those estimated as zero.

Clustering

Figure S4 shows the configuration of the $k = 11$ clustering solution for the HCP sample. These 11 clusters are subcomponents of the $k = 3$ supercluster solution as described in the main text.

We subjected the LTS sample heritability matrix to the same clustering procedure as the HCP sample's. We found stable solutions at k-values of 2, 3 and 4 (**Figure S5a**), with some evidence of stable, higher-order solutions evidenced by local maxima at k-values of 28 and 42 (although it should be noted these solutions were substantially less stable than the lower-order solutions). For the most direct comparison to the HCP $k = 3$ super-cluster solution, we investigated the $k = 3$ solution in the LTS sample in great detail. The $k = 3$ clustering solution is shown in **Figure S5b** and heritable connectivity profile is described by **Figure S5c**. The super-clusters of the LTS sample were similar in nature to those of the HCP sample, with one super-cluster containing regions with no particualrly clear pattern of heritable connectivity (1), another super-cluster containing regions with particularly heritable connectivity to default and other cognitive regions (2), and a third super-cluster with particularly heritable connectivity to sensory areas (predominantly visual regions). Super-cluster 1 is larger in the LTS sample and super-cluster 3 is smaller in the LTS sample, consistent with the lower heritability estimates in the LTS sample. **Figure S6** shows the spatial locations of and *a priori* network assignments for regions that are part of the three LTS super-clusters.

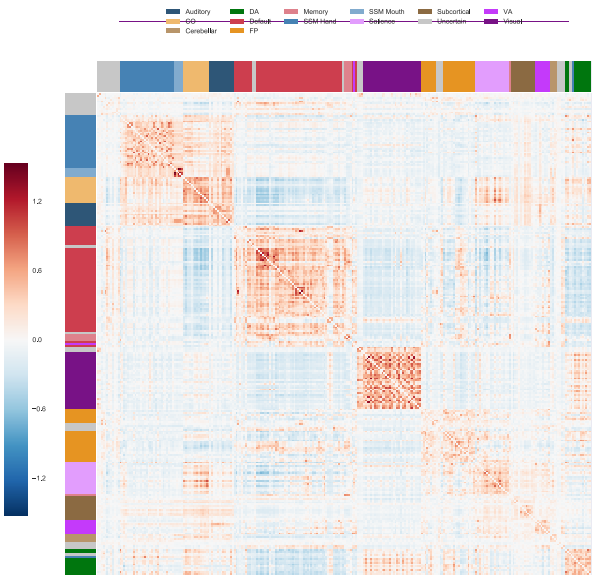

Figure S1. **Mean phenotypic connectivity across LTS participants**. Group average connectome for LTS sample. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

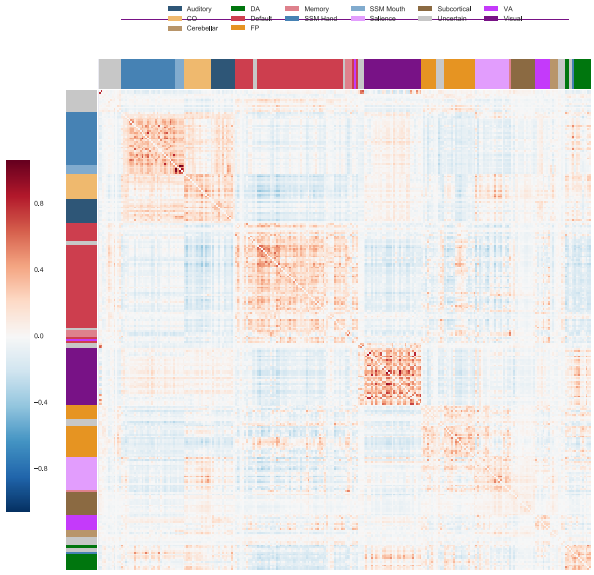

Figure S2. **Mean phenotypic connectivity across HCP participants**. Group average connectome for HCP sample. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

Bivariate Analyses

We found individual differences in global connectivity are associated with variation in local functional connectivity by correlating each edge within the connectome with the global efficiency measure across individuals (a procedure similar to one described in prior work to develop predictive models ([Shen et al., 2017]; see **Figure S7a** for phenotypic correlations > 0.15 or < -0.15). The phenotypic correlation pattern largely conforms to the *a priori* network structure with large portions of sensory-somatomotor, default, visual, frontoparietal, and dorsal attention network connectivity significantly associated with global efficiency. Additionally, the pattern is compatible with our understanding of global efficiency in that many connections that are positively associated with global efficiency are between-network connections, suggesting that one way to increase global efficiency is to decrease the functional distance between communities that typically communicate mostly via connector hubs. A notable exception to this rule is dorsal attention/visual-to-sensory/somatomotor/cingulo-opercular connections, which are between-network connections negatively correlated with

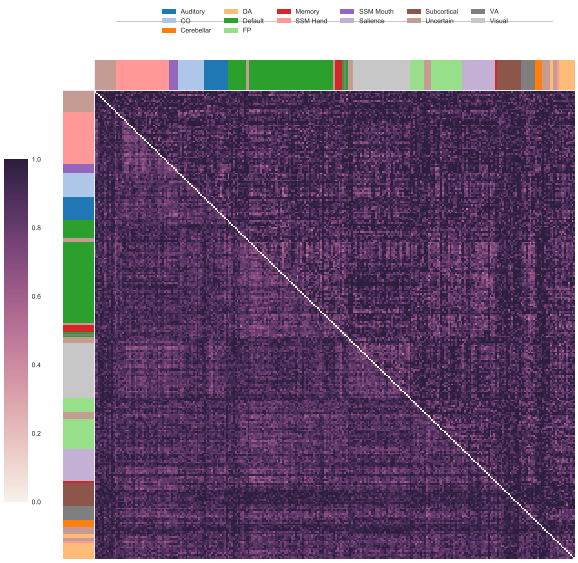

Figure S3. **Connection-wise estimates of non-shared environmental influence in LTS sample**. Estimates of non-shared environmental influence (E) are displayed for the HCP and LTS samples in the lower and upper triangles, respectively. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

Table 1
Within- versus between network connectivity by network

	LTS				HCP			
	m_{within}	$m_{between}$	*t*	*p*	m_{within}	$m_{between}$	*t*	*p*
a. Heritability (A)$_{\text{Whole Connectome}}$ -	0.090	0.065	12.883	<0.001	0.106	0.078	15.413	<0.001
b. Heritability (A) by network								
Frontoparietal	0.072	0.065	1.130	ns	0.082	0.075	1.443	ns
Ventral Attention	0.082	0.085	-0.167	ns	0.116	0.079	2.005	0.053
Subcortical	0.023	0.040	-2.954	0.004	0.033	0.050	-2.646	0.010
Cingulo-opercular	0.042	0.072	-3.798	0.000	0.075	0.078	-0.323	ns
Auditory	0.082	0.060	1.798	ns	0.100	0.077	2.361	0.021
Default	0.110	0.076	11.585	0.000	0.117	0.080	13.323	0.000
Memory	0.073	0.071	0.061	ns	0.110	0.117	-0.223	ns
SSM Hand	0.068	0.058	2.251	0.025	0.119	0.089	5.544	0.000
Dorsal Attention	0.114	0.080	2.354	0.022	0.119	0.088	2.360	0.022
Visual	0.115	0.068	8.169	0.000	0.132	0.078	10.709	0.000
Cerebellar	0.082	0.043	1.173	ns	0.011	0.057	-4.164	0.007
Salience	0.080	0.071	0.946	ns	0.097	0.089	0.950	ns
SSM Mouth	0.016	0.066	-3.149	0.011	0.164	0.092	1.483	ns
Uncertain	0.036	0.050	-3.491	0.001	0.055	0.066	-2.528	0.012

global efficiency.

Multiplying the first two paths from the bivariate genetic models (i.e., $x_1 * x_2$, $y_1 * y_2$, $z_1 * z_2$), provides estimates of the phenotypic correlation between connection strength and global efficiency predicted by genetics, shared environments, and non-shared environments, respectively. These three estimates sum to the total phenotypic correlation, and dividing by the phenotypic correlation provides information about the relative contributions of genes and environments to the association between local and global connectivity. Possible results of this analysis include: 1. globally homogeneous genetic or environment influence on the phenotypic correlation, 2. homogeneous genetic or environmental influence within any given *a priori* network, or 3. a heterogeneous patten of genetic and environmental influence. Briefly, our results indicate that genetic influence on the edges most phe-

Figure S4. **Ward 11-cluster configurations**. Higher dimensionality clustering solution (k = 11) of additive genetic influences (A) reveals super-cluster 2 breaks into sub-clusters (2, 3, 4) with mixed domain organization based on *a priori* functional characterization. Super-cluster 3 breaks into 2 sub-clusters (5, 6) of mixed domains and 5 sub-clusters (7-11) with more pure membership (e.g., visual regions only in sub-cluster 7) according to *a priori* functional characterization. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

Table 2
Within- versus between-network connectivity by region. *Region is 0-indexed.*

		LTS					HCP				
region	network	Mean A	Mean A_{within}	Mean $A_{between}$	t	p	Mean A	Mean A_{within}	Mean $A_{between}$	t	p
17	SSM Hand	0.024	0.044	0.021	2.200	0.029	0.066	0.125	0.059	3.902	0.000
18	SSM Hand	0.041	0.071	0.037	2.325	0.021	0.121	0.176	0.114	2.895	0.004
23	SSM Hand	0.070	0.105	0.065	2.339	0.020	0.123	0.171	0.117	2.571	0.011
26	SSM Hand	0.060	0.026	0.064	-2.316	0.021	0.109	0.147	0.104	2.104	0.036
29	SSM Hand	0.075	0.118	0.069	2.662	0.008	0.112	0.165	0.106	3.291	0.001
31	SSM Hand	0.042	0.102	0.034	4.955	0.000	0.080	0.129	0.074	3.274	0.001
74	Default	0.080	0.120	0.068	3.308	0.001	0.092	0.135	0.079	4.047	0.000
90	Default	0.107	0.146	0.097	2.815	0.005	0.107	0.156	0.094	3.950	0.000
98	Default	0.106	0.132	0.098	2.156	0.032	0.116	0.162	0.102	4.142	0.000
99	Default	0.074	0.102	0.066	2.544	0.012	0.116	0.166	0.102	4.262	0.000
103	Default	0.099	0.146	0.086	3.844	0.000	0.103	0.157	0.087	4.709	0.000
104	Default	0.119	0.173	0.104	3.687	0.000	0.065	0.091	0.057	2.965	0.003
106	Default	0.086	0.114	0.078	2.190	0.029	0.110	0.144	0.100	2.916	0.004
107	Default	0.084	0.127	0.072	3.475	0.001	0.080	0.111	0.071	3.278	0.001
108	Default	0.069	0.097	0.061	2.647	0.009	0.077	0.139	0.060	6.621	0.000
114	Default	0.085	0.123	0.075	3.008	0.003	0.111	0.156	0.099	4.130	0.000
115	Default	0.105	0.139	0.095	2.711	0.007	0.057	0.096	0.047	4.561	0.000
120	Default	0.092	0.131	0.080	3.454	0.001	0.091	0.112	0.086	2.044	0.042
127	Default	0.095	0.147	0.081	4.101	0.000	0.109	0.172	0.091	5.123	0.000
131	Default	0.081	0.105	0.074	2.085	0.038	0.111	0.175	0.093	5.797	0.000
137	Default	0.094	0.140	0.081	3.510	0.001	0.084	0.152	0.064	6.765	0.000
145	Visual	0.068	0.124	0.060	3.646	0.000	0.114	0.174	0.106	3.466	0.001
146	Visual	0.075	0.132	0.067	3.422	0.001	0.103	0.147	0.098	2.710	0.007
148	Visual	0.078	0.166	0.066	5.405	0.000	0.092	0.199	0.078	7.151	0.000
149	Visual	0.094	0.172	0.084	4.019	0.000	0.067	0.129	0.059	4.277	0.000
152	Visual	0.095	0.136	0.090	2.351	0.019	0.103	0.166	0.095	3.855	0.000
165	Visual	0.060	0.094	0.055	2.315	0.021	0.080	0.157	0.070	5.073	0.000
166	Visual	0.074	0.112	0.069	2.302	0.022	0.106	0.157	0.099	2.992	0.003
168	Visual	0.088	0.227	0.069	8.455	0.000	0.075	0.127	0.068	4.000	0.000
170	Visual	0.083	0.153	0.074	3.991	0.000	0.124	0.175	0.118	2.807	0.005
173	Visual	0.066	0.134	0.057	4.329	0.000	0.066	0.136	0.056	5.429	0.000
212	Salience	0.076	0.011	0.081	-3.021	0.003	0.039	0.080	0.036	2.988	0.003
255	SSM Hand	0.080	0.043	0.084	-2.179	0.030	0.118	0.160	0.113	2.455	0.015

notypically correlated with global efficiency are heterogeneous while environmental influences are much more cohesive across the entire connectome.

Specifically, genetic, common environmental, and unique environmental influence collectively determines the phenotypic relationship between local and global connectivity. **Figure S7b** shows the correlation that genetic influences predict between local connectivity and global efficiency scaled by the phenotypic correlation; thus units are proportions of the phenotypic correlation predicted by genes/environment with directionality indicating prediction in the same (+) or opposite (-) direction of the phenotypic correlation. Notably, correlations predicted by genetics were both positive and negative for connections between different spatial locations, even within connections of the same *a priori* network. That is, the estimates of the predicted correlations varied both within regions and between regions, demonstrating a complex architecture of genetic effects across the connectome. For example, shared additive genetic influence predicted a relationship between default-to-sensory-somatomotor connectivity and global efficiency in the same direction as the phenotypic correlation for some default network connections, but predicted a relationship in the opposite direction of the phenotypic correlation for other default network connections.

Figure S7c, upper triangle shows proportion of correlations between local connectivity and global efficiency predicted by shared environmental influence. There are clearly structured differences between the spatial distribution of positive and negative correlations predicted by shared environmental influences; however, the absolute direction of the predicted correlation was almost universally negative. That is, C influences shared by local connectivity and global efficiency predicted negative correlations between local and

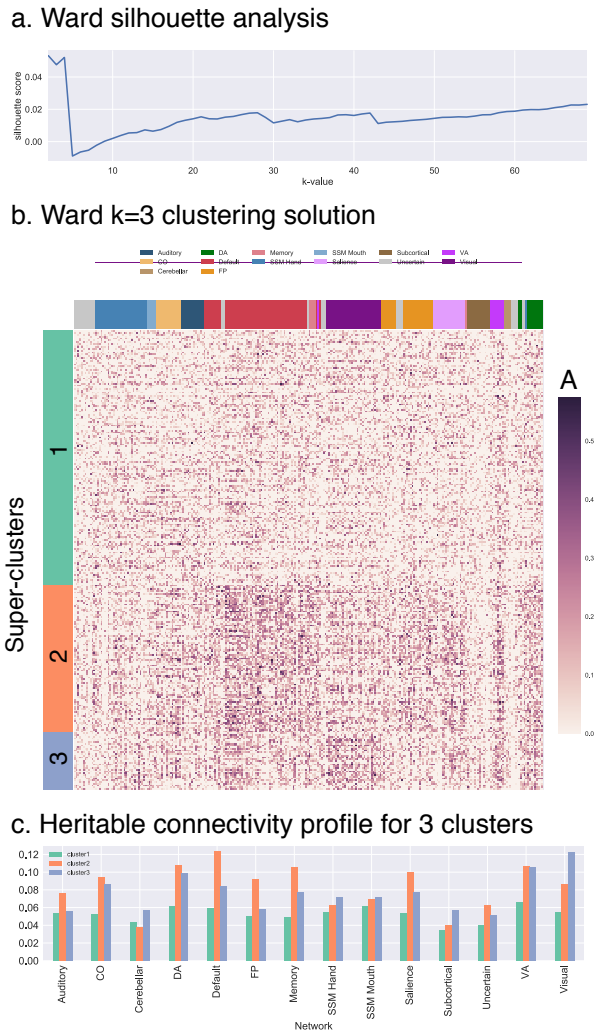

Figure S5. **Ward 3-cluster solution (LTS sample)**. Row-wise clustering of LTS additive genetic (A) estimates reveals several stable clustering solutions of regions with similar patterns of connectivity heritability. Super-clusters (k = 3) are described in detail for comparison to resuls in the HCP sample. a. Silhouette analysis revels stable clustering solutions at k-values of 2, 3, and 4. b. Clustered version of LTS A estimate matrix for k = 3 solution. c. Average connectivity heritability for super-clusters 1, 2, and 3 organized by *a priori* network of connection. Super-cluster 1 is characterized by weakly heritable connections. Super-clusters 2 and 3 have high connectivity heritability globally with particularly high heritability of higher level cognitive (Super-cluster 2) and visual (Super-cluster 3) connections. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.
global measures of the connectome, even for places where positive correlations are present in the phenotypic correlation matrix, such as most connections between default network areas and other networks.

Figure S7c, lower triangle shows proportions of correlations between local connectivity and global efficiency predicted by non-shared environmental influences. Non-shared environmental influences predominantly predicted relationships between local connectivity and global efficiency in the same direction as the phenotypic correlations.

Reliability

To examine the effect of test-retest reliability on estimation of genetic influences, we calculated edge-wise estimates of reliability utilizing multiple resting-state scans from the HCP dataset. Although a total of four resting-state scans are available for each HCP participant, we utilized only the two from the first day of scanning due to storage constraints of our computing infrastructure. Because each HCP participant's scans were collected with different phase encoding (LR or RL with the intention of combining the two scans to minimize susceptibility artifact when compared to a single long scan using AP phase encoding, for example), we created two different estimates of reliability for each edge, 1. split scan reliability - correlation (corrected using the Spearman-Brown prophecy formula) between homologous edges from the first half of scan 1 concatenated to the second half of scan 2 and the first half of scan 2 concatenated to the second half of scan 1 - to minimize any confounding relationship between measurement error, lateralized susceptibility artifact due to LR or RL phase encoding, and start-of-scan/end-of-scan effects, and 2. scan 1/scan 2 reliability - correlation (corrected using the Spearman-Brown prophecy formula) between homologous edges from scan 1 and scan 2 - in which measurement error could be confounded with susceptibility artifact or start-of-scan/end-of-scan effects.

Split scan reliability was high and normally distributed (M = 0.849, SD = 0.061). Scan 1/scan 2 reliability was lower than split scan reliability (M = 0.449, SD = 0.215). The two reliability metrics were highly correlated (Pearson's r = 0.849, $p < 0.001$). We utilized these reliability metrics to test whether or not differences in within versus between-network connectivity persisted above and beyond any effects of reliability. We found a moderate and highly significant positive correlation between A estimates and split scan reliability (Pearson's r = 0.298, $p < 0.001$). A similar moderate and highly significant positive correlation is found between A estimates and scan 1/scan 2 reliability (Pearson's r = 0.319, p < 0.001). We then regressed A estimates on within versus between network connectivity controlling for either reliability measure (in separate models) and found that differences in heritability estimates for within versus between-network connections persisted in both models.

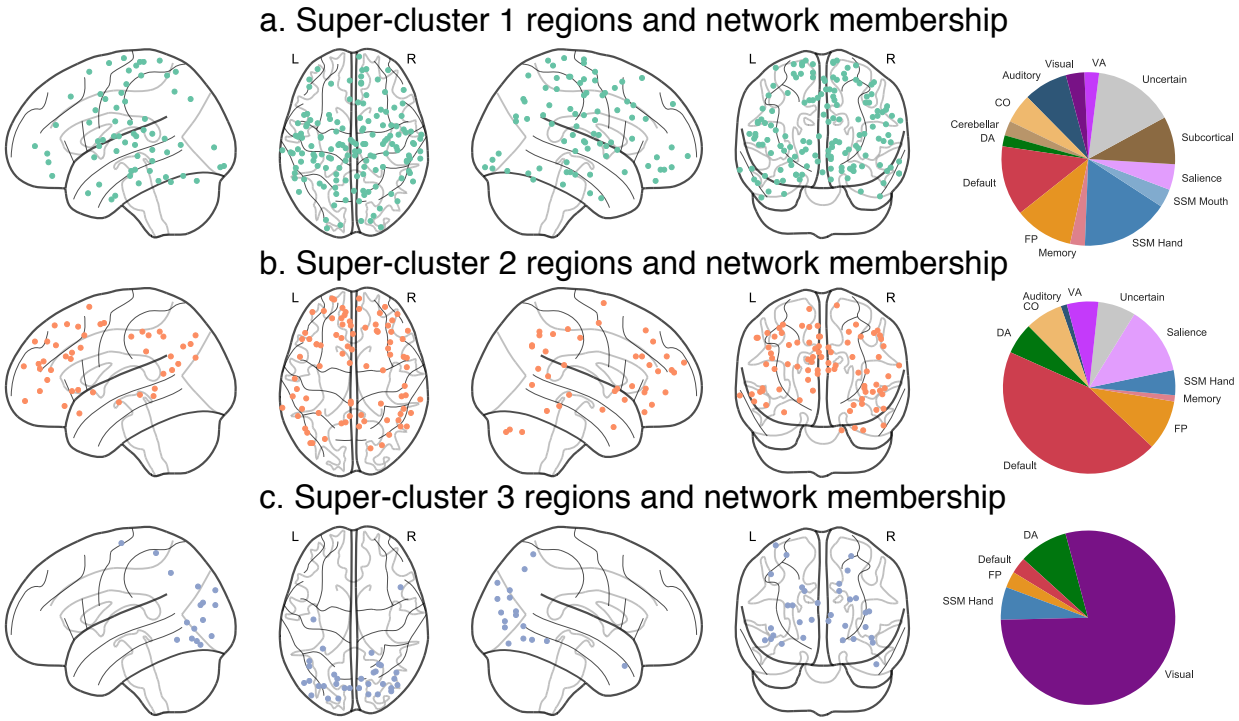

Figure S6. **Ward 3-cluster summary (LTS sample)**. Spatial location of regions from super-clusters 1-3 of k = 3 solution in the LTS sample. a. Super-cluster 1 regions are widely distributed across the brain. b. Super-cluster 2 regions are located across lateral prefrontal, lateral parietal, mid and anterior temporal, midline frontal, and cingulate areas. c. Super-cluster 3 regions are located primarily in visual areas. CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.

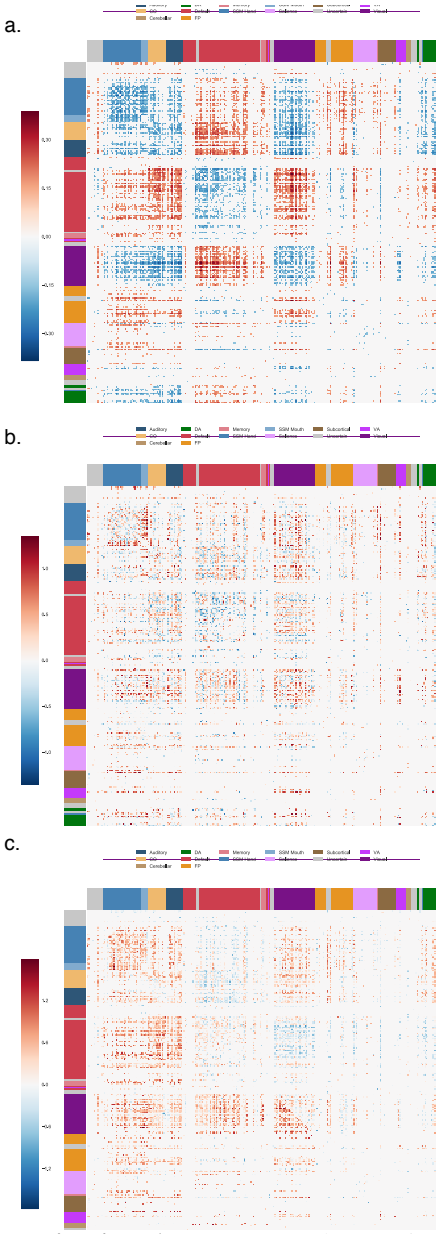

Figure S7. **Bivariate Analyses**. a. Phenotypic correlation between local connections and global efficiency. Only phenotypic correlations > 0.15 or < -0.15 are displayed. b. Correlation between local connections and global efficiency predicted by shared additive genetic influence (Gr). Gr is divided by signed phenotypic correlation, thus units are proportions of the phenotypic correlation with directionality indicating prediction in the same (+) or opposite (-) direction of the phenotypic correlation. c. Correlation between local connections and global efficiency predicted by common environmental influence (Cr; upper triangle) and unique environmental influence (Er; lower triangle). CO = cingulo-opercular, DA = dorsal attention, FP = frontoparietal, SSM = sensory/somatomotor, VA = ventral attention.