Supplemental Figures:

Figure S1

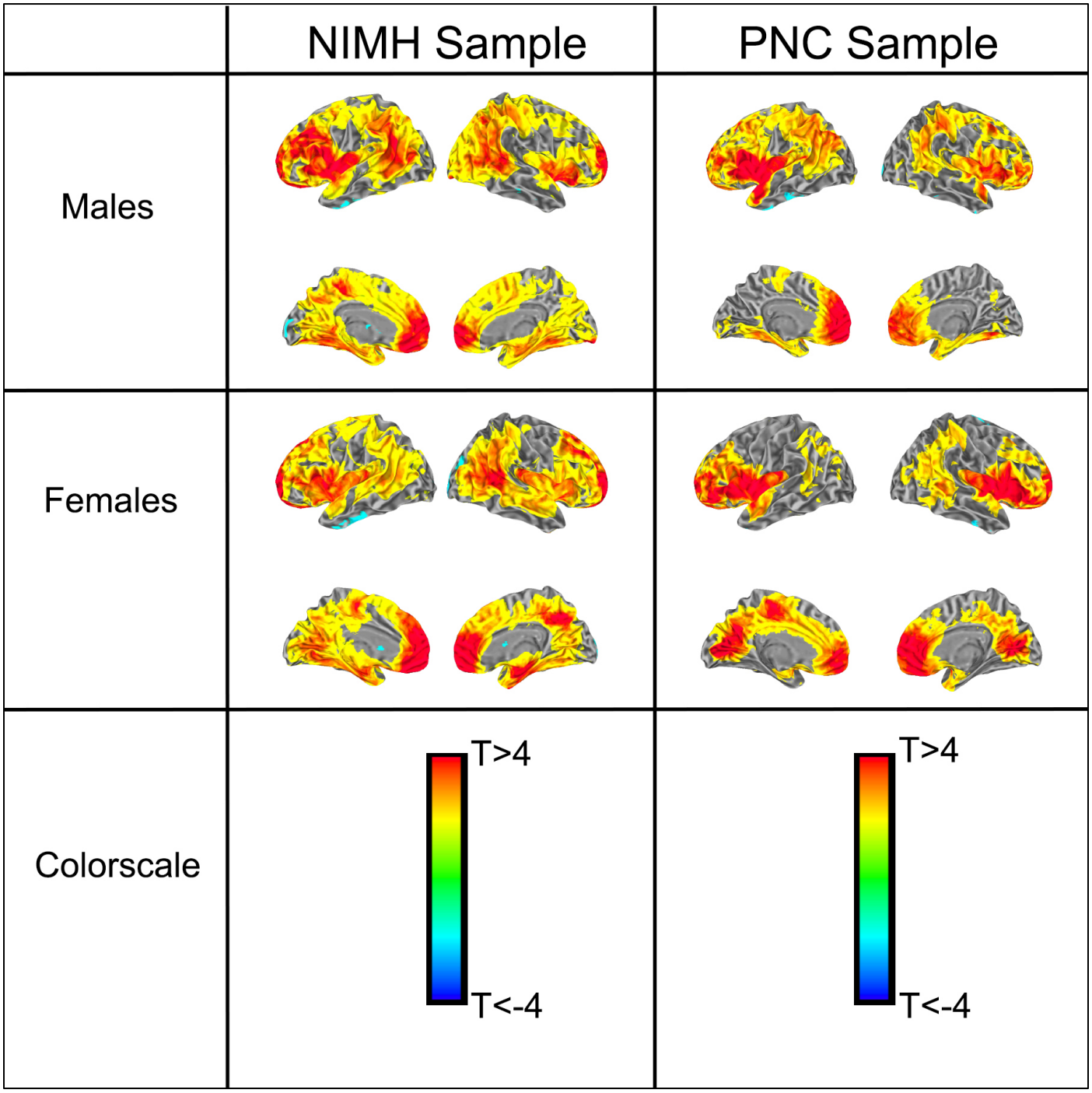

Figure S2

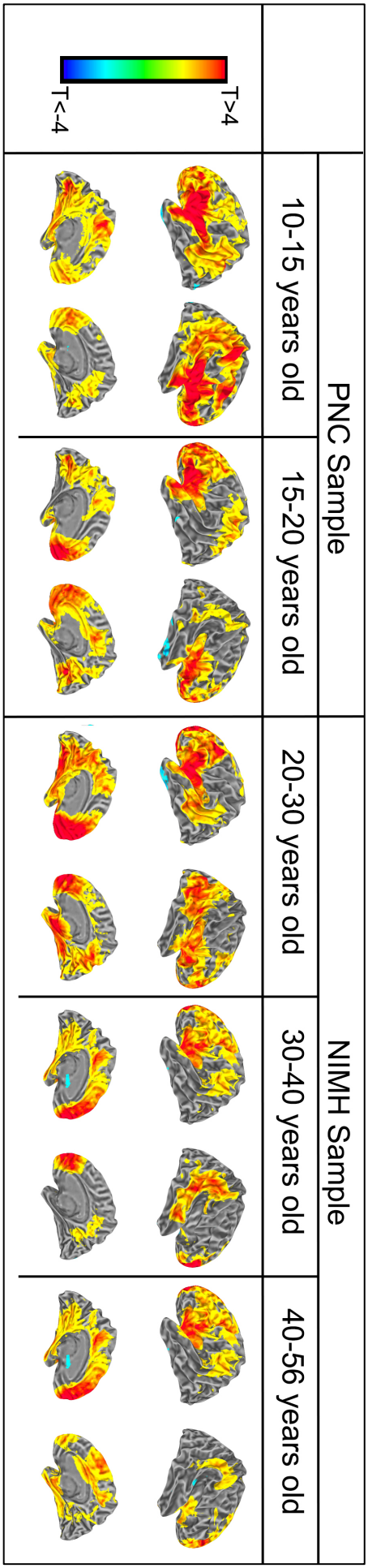
T>4
T<-4
PNC Sample
NIMH Sample
10-15 years old
15-20 years old
20-30 years old
30-40 years old
40-56 years old

Figure S1, Related to Figure 1: The LGI-*g* relationship in males and females separately. Figure shows the LGI-*g* relationship in males (Top) and females (Bottom) for the NIMH (Left) and PNC (Right) samples separately. No regions showed significant differences in this relationship across sexes. Statistical maps were thresholded at FDR-corrected $p<0.05$. The equally-weighted composite was used as the estimate of *g.* Note the similar spatial pattern to the LGI-*g* relationship seen in the entire group, as shown in Figure 1.

Figure S2, Related to Figure 3: Age-specific relationship of the LGI-*g* relationship. Figure shows the LGI-*g* relationship in each of the age subgroups. Images were thresholded at $p<0.05$. The equally-weighted composite was used as an estimate of *g* in this analysis.

Supplemental Tables

Table S1, Related to Figures 1 and 2: List of neuropsychological tests in the NIMH and PNC samples used in the composite z-score method to calculate general cognitive ability.

NIMH	PNC
WAIS Digit Symbol Scaled Score	**Conditional Exclusion** Number of Categories
Digit Span Forward Raw Score	**Emotional Differentiation** Percent Correct
Digit Span Backwards Raw Score	**Emotional Identification** Correct Responses
WAIS Letter Number Sequencing Raw Score	**Face Memory** Correct Responses
WMS Logical Memory 1 Raw Score	**Letter N-Back** Correct Responses (2-Back)
WMS Logical Memory 2 Raw Score	**Line Orientation** Correct Responses
WMS Verbal Paired Associates Raw Score	**Matrix Reasoning** Correct Responses
CVLT Trials 1-5 Raw Score	**Verbal Reasoning** Correct Responses
WMS Visual Reproduction 1 Raw Score	**Visual Object Learning** Correct Responses
WMS Visual Reproduction 2 Raw Score	**Word Memory** Correct Responses
Wisconsin Card Sorting Percent Perseverative Errors	**WRAT Reading** Correct Responses
Wisconsin Card Sorting Categories / Trials	
Wisconsin Card Sorting Correct Responses / Trials	
Number N-Back (1-Back) Percent Correct	
Number N-Back (2-Back) Percent Correct	
Number N-Back (3-Back) Percent Correct	
WAIS Arithmetic Scaled Score	
WAIS Similarities Scaled Score	
WAIS Picture Completion Scaled Score	

WRAT Reading Standard Score	
Benton Line Orientation Raw Score	
Letter Fluency Number of Words	
Category Fluency Number of Words	
Trails A Seconds to Complete	
Trails B Seconds to Complete	

Table S2, Related to Figures 1 and 2: Component weights for the first three principal components from an analysis of the PNC neuropsychological data. The first un-rotated component was used to estimate general cognitive ability in Figures 1 and 2. Associations between the second and third components with LGI were tested, but did not show any significant results.

Neuropsychological Test Measure	**Component #1 Weights**	**Component #2 Weights**	**Component #3 Weights**
WRAT Reading Correct Responses	.766	.254	
Verbal Reasoning Correct Responses	.711	.326	.123
Word Memory Reaction Time to Correct Responses	-.689	.344	.158
Finger Tapping Dominant + Non-dominant Hands	.650		-.352
Emotional Identification Reaction Time of Correct Responses	-.645	.354	
Continuous Performance Reaction Time to Correct Responses	-.634	.110	.426
Verbal Reasoning Reaction Time of Correct Responses	-.593	.263	-.144
Emotional Differentiation Percent Correct	.584	.382	
Matrix Reasoning Percent Correct	.583	.487	
Line Orientation Percent Correct	.582	.362	
Age Differentiation Percent Correct	.571	.328	
Letter N-Back Correct Responses	.559	.204	.158
Face Memory Correct Responses	.527	.222	.242
Letter N-Back Reaction Time of Correct Responses	-.521		.321
Continuous Performance Correct Responses	.517	.134	

Motor Praxis Reaction Time of Correct Responses	-.506	.221	.257
Emotional Identification Correct Responses	.375	.139	
Conditional Exclusion Number of Categories	.352	.191	.269
Word Memory Correct Responses	.343	.195	.273
Age Differentiation Reaction Time of Correct Responses	-.295	.657	-.141
Visual Object Learning Reaction Time of Correct Responses	-.327	.633	.189
Emotional Differentiation Reaction Time	-.404	.586	-.227
Face Memory Reaction Time of Correct Responses	-.456	.582	.102
Matrix Reasoning Reaction Time of Correct Responses	.131	.556	-.300
Line Orientation Reaction Time of Correct Responses	-.205	.528	-.209
Visual Object Learning Correct Responses	.311	.226	.447
Conditional Exclusion Reaction Time of Correct Responses	-.354	.339	-.369

Table S3, Related to Figure 1: Dice Similarity Coefficients (DSC) and Correlation Coefficients (R^2) comparing the regional pattern of LGI-*g* associations found using the equally-weighted composite score with the pattern found using the three other methods of *g* calculation in both NIMH and PNC samples.

		Weighted-Component Score	WRAT Reading Subtest	WAIS Full-Scale IQ
NIMH	DSC	0.97	0.82	0.93
	R^2	0.99	0.88	0.92
PNC	DSC	0.91	0.86	Not Available
	R^2	0.99	0.79	

Supplemental Experimental Procedures

Participants:

NIMH Study: Four hundred forty healthy individuals underwent 3T MRI scanning, and completed a comprehensive neuropsychological battery, structured clinical interview for the DSM-IV (SCID-IV), neurological exam, clinical MRI, urine drug screen and laboratory evaluation to ensure they were free from any medical or neuropsychiatric disease. Participants were 18-56 years of age (mean 31.3 +/- 9.4 years) and included 250 females (F:M ratio=1.3:1).

Philadelphia Neurodevelopment Cohort (PNC): Data were accessed from the freely available resource on dbGaP (http://www.ncbi.nlm.nih.gov/gap; accession number phs000607.v1.p1) [S1]. Nine hundred ninety-seven participants from the PNC sample underwent neuroimaging and neuropsychological testing. After quality control procedures to ensure that participants were free from neuropsychiatric disease, that 3T structural imaging scans were of high quality by visual inspection, and that complete neuropsychological data were available, 662 participants remained for further analysis. The remaining participants were 8-21 years of age (mean 14.7 +/- 3.3 years) and included 378 females (F:M ratio=1.3:1).

Neuropsychological Data:

Several methods were used to estimate *g* from the neuropsychological batteries in both samples. These included an equally-weighted composite z-score method, a component-weighted factor analysis, using the Wide-Range Achievement

Test (WRAT) reading subtest, and a four item short-form of the WAIS-R (only available in the NIMH sample).

NIMH Study: Participants were administered a comprehensive neuropsychological battery. Estimates of *g* were calculated in four ways. First, guided by factor analytic results, 25 variables were z-score normalized and averaged to form an equally-weighted composite index of *g*, as previously reported [S2]. Table S1.1 lists the variables included in the composite calculation of *g* in the NIMH sample. Second, a component-weighted index of *g* was calculated from a principal component analysis of the neuropsychological data with SPSS (version 21, IBM corp., Armonk, NY). The first un-rotated principal component was used to represent *g* [S3] and weighted individual scores for *g* were derived using standard least squares regression methodology [S4]. Third, the NIMH neuropsychological battery included the WRAT reading subtest, which has previously been shown to estimate *g* [S5, 6]. Finally, this data set also included a validated estimate of full-scale IQ derived from a four item short-form of the WAIS-R including Similarities, Arithmetic, Picture Completion and Digit Symbol tests [S7-9].

Philadelphia Neurodevelopment Cohort (PNC): All participants in the PNC sample completed the widely-used Penn Computerized Neurocognitive Battery [S10]. Estimates of *g* for the PNC sample were generated in three ways. First, as in the NIMH sample, an equally-weighted composite was constructed by averaging z-transformed scores for individual accuracy measures. Also, as in the NIMH sample, the WRAT reading subtest was also available in the PNC neuropsychological battery and used as an estimate of *g*. Finally, the first component from an SPSS principal

components analysis of the full battery yielded a sample-specific, component-loading-weighted estimate of *g*, derived in the same manner as in the NIMH sample. In performing this analysis, we found that, together, the first three components accounted for roughly 46% of the overall variance. No subsequent components accounted for more than 5% of the variance. Examination of the three components revealed the first to have predominantly high loadings from accuracy measures. The un-rotated first principal component was subsequently used as the metric of *g* above for examination of the LGI-*g* relationship using the component-weighted factor method. The second component most strongly reflected reactions time (RT) measures, weighted towards simple speediness of information processing during task performance. The third component included a mix of positive and negative loadings, which may reflect a speed/accuracy tradeoff dimension for performance on some tasks. This is consistent with prior analyses of these same data by other groups [S11]. In contrast to the significant associations found between LGI and the first component used to estimate *g*, the second and third components revealed no significant associations with LGI at an FDR-corrected p-value threshold of 0.05.

MRI Acquisition and Preprocessing:

For the NIMH study, T1-weighted 3D magnetization prepared rapid acquisition gradient echo (MPRAGE) structural images (TR=7.28 s, TE=2.74 ms, 120-146-136 slices, 0.859x0.859x1.2 mm) were collected on a GE 3T MRI scanner (Milwaukee, WI). For the PNC study, T1-weighted 3D MPRAGE structural images

were collected on a Siemens Tim Trio 3T MRI scanner (TR=1.81 s, TE=3.51 ms, 160 slices, 0.9375x0.9375x1 mm).

MRI data from both studies were processed in the same way. Raw images were visually inspected for artifacts and excluded if not of high quality. Data were corrected for intensity non-uniformity using n3 normalization [S12], and then processed with Freesurfer, version 5.3 (http://surfer.nmr.mgh.harvard.edu) [S13], to create 3-dimensional surface representations of the pial and white matter surfaces. Resulting surfaces were further processed with Freesurfer using the local gyrification index (LGI) function to create a representation of the cortical hull with measurements of LGI for each node of the surface [S14]. All surfaces for each individual were aligned to a standard mesh using AFNI/SUMA tools (http://afni.nimh.nih.gov) [S15] to allow for comparisons across individuals. Additionally, the surface areas of both the pial surface and the cortical hull were calculated from resulting surfaces and 3D global gyrification index (GGI) was measured for each participant by taking the ratio of total pial surface area to the total cortical hull surface area.

Statistical Analyses:

Statistics for the NIMH and PNC datasets were computed separately, following the same steps. To test our primary hypothesis, linear regressions were performed to examine associations between LGI and *g*, controlling for age and sex, then for age, sex and TBV, in each sample. All resulting maps were corrected for multiple comparisons using a false discovery rate correction method (FDR) [S16].

Maps were thresholded at an FDR-corrected p-value < 0.01. These whole-brain, surface-based calculations were performed using AFNI/SUMA tools (http://afni.nimh.nih.gov).

To quantify the amount of variance explained by the regression analyses using the equally-weighted component *g* estimate, for each subject we calculated an average of the LGI values from all nodes that displayed significant associations with *g* (FDR-corrected p-value<0.01) in the group-level analysis. Using SPSS (version 21, IBM corp., Armonk, NY), a two-step multiple regression analysis was performed using *g* as the dependent variable. In the first step, effects due to age and sex were calculated, and the second step tested for effects due to *g*. The change in R^2 for the second step relative to the first was taken as the percentage of variance explained. Similar analyses were also performed to determine the variance of *g* explained by TBV.

For GGI, global measures of GI for each participant were calculated from Freesurfer surfaces and imported into SPSS version 21 (IBM Corp., Armonk, NY) for correlation with other variables, including age, sex, total brain volume and *g*.

Comparisons of Statistical Maps:

To compare correlational maps (1) between the two groups, and (2) within the groups and between methods of estimating *g*, we used both Pearson correlation coefficients and Dice-similarity coefficients (DSC). For Pearson correlation coefficients, we correlated the T-scores of all 198,812 nodes per hemisphere of one map to another (e.g., all nodes of the NIMH sample to all nodes of the PNC sample).

DSC values were computed to determine the similarity of the regions found to exhibit significant association between LGI and *g*. The DSC computes the similarity, or agreement, of categorical data; is a specific application of the statistical coefficient kappa; and has been previously applied to neuroimaging data [S17-19]. To compute DSC values, statistical maps were transformed to binary masks indicating for each node whether it exhibited significance above the statistical thereshold (FDR corrected p-value<0.01). The number of significant nodes in each mask was counted (N_{mask1} and N_{mask2} in the equation below), as was the number of nodes showing significance in both masks (N_{agree} in the equation). The DSC is then computed by dividing twice the number of nodes showing significant association in both analyses (N_{agree}) by the number of nodes showing significance in each mask (N_{mask1} and N_{mask2}):

$$DSC = \frac{2 * N_{agree}}{N_{mask1} + N_{mask2}}$$

Values for DSC greater than 0.6 indicate “substantial” agreement between maps and values greater than 0.8 indicate “almost perfect” agreement [S20].

Supplemental References:

S1. Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M.E., Hopson, R., Jackson, C., Keefe, J., Riley, M., et al. (2014). Neuroimaging of the Philadelphia neurodevelopmental cohort. NeuroImage *86*, 544-553.

S2. Dickinson, D., Goldberg, T.E., Gold, J.M., Elvevag, B., and Weinberger, D.R. (2011). Cognitive factor structure and invariance in people with schizophrenia, their unaffected siblings, and controls. Schizophrenia bulletin *37*, 1157-1167.

S3. Jensen, A.R. (1998). The g factor : the science of mental ability, (Westport, Conn.: Praeger).

S4. Distefano, C., Zhu, M., and Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. Practical Assessment, Research and Evaluation *14*, 1-11.

S5. Ball, J.D., Hart, R.P., Stutts, M.L., Turf, E., and Barth, J.T. (2007). Comparative utility of Barona Formulae, Wtar demographic algorithms, and WRAT-3 reading for estimating premorbid ability in a diverse research sample. The Clinical neuropsychologist *21*, 422-433.

S6. Griffin, S.L., Mindt, M.R., Rankin, E.J., Ritchie, A.J., and Scott, J.G. (2002). Estimating premorbid intelligence: comparison of traditional and contemporary methods across the intelligence continuum. Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists *17*, 497-507.

S7. Kaufman, A.S., and Lichtenberger, E.O. (2002). Assessing adolescent and adult intelligence, 2nd Edition, (Boston, MA: Allyn and Bacon).

S8. Nagle, R.J., and Bell, N.L. (1995). Clinical utility of Kaufman's "amazingly" short forms of the WAIS-R with educable mentally retarded adolescents. Journal of clinical psychology *51*, 396-400.

S9. Ward, L.C., and Ryan, J.J. (1997). Validity of quick short forms of the Wechsler Adult Intelligence Scale-Revised with brain-damaged patients. Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists *12*, 63-69.

S10. Gur, R.C., Richard, J., Hughett, P., Calkins, M.E., Macy, L., Bilker, W.B., Brensinger, C., and 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*, 254-262.

S11. Moore, T.M., Reise, S.P., Gur, R.E., Hakonarson, H., and Gur, R.C. (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. Neuropsychology *29*, 235-246.

S12. Sled, J.G., Zijdenbos, A.P., and Evans, A.C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE transactions on medical imaging *17*, 87-97.

S13. Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage *9*, 179-194.

S14. Schaer, M., Cuadra, M.B., Tamarit, L., Lazeyras, F., Eliez, S., and Thiran, J.P. (2008). A surface-based approach to quantify local cortical gyrification. IEEE transactions on medical imaging *27*, 161-170.

S15. Saad, Z.S., and Reynolds, R.C. (2012). Suma. NeuroImage *62*, 768-773.

S16. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

S17. Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., and Petersen, S.E. (2014). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. Cerebral cortex.

S18. Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., and Palmer, A.C. (1994). Morphometric analysis of white matter lesions in MR images: method and validation. IEEE transactions on medical imaging *13*, 716-724.

S19. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells, W.M., 3rd, Jolesz, F.A., and Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index. Academic radiology *11*, 178-189.

S20. Landis, J.R., and Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics *33*, 159-174.