Cell Reports, Volume *42*

Supplemental information

Development of white matter fiber covariance networks supports executive function in youth

Joëlle Bagautdinova, Josiane Bourque, Valerie J. Sydnor, Matthew Cieslak, Aaron F. Alexander-Bloch, Maxwell A. Bertolero, Philip A. Cook, Raquel E. Gur, Ruben C. Gur, Fengling Hu, Bart Larsen, Tyler M. Moore, Hamsanandini Radhakrishnan, David R. Roalf, Russel T. Shinohara, Tinashe M. Tapera, Chenying Zhao, Aristeidis Sotiras, Christos Davatzikos, and Theodore D. Satterthwaite

SUPPLEMENTAL FIGURES

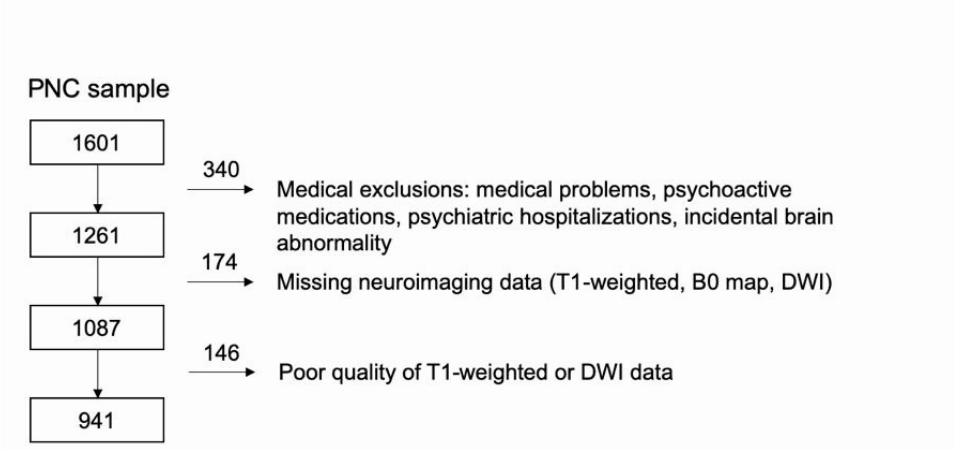

Figure S1. **Sample construction; related to STAR Methods.** Among the original 1,601 participants from the PNC, 340 participants were excluded due to clinical factors, including medical disorders that could affect brain function, current use of psychoactive medications, prior inpatient psychiatric hospitalizations, or an incidentally encountered structural brain abnormality. Among the 1,261 participants eligible for inclusion, 174 participants were excluded for missing either a B0 field map, and/or diffusion images. The remaining 1,087 participants underwent a rigorous manual and automated quality assurance protocol for DWI datasets, which excluded 146 participants for poor data quality. This set of exclusion criteria resulted in a final sample of 941 participants.

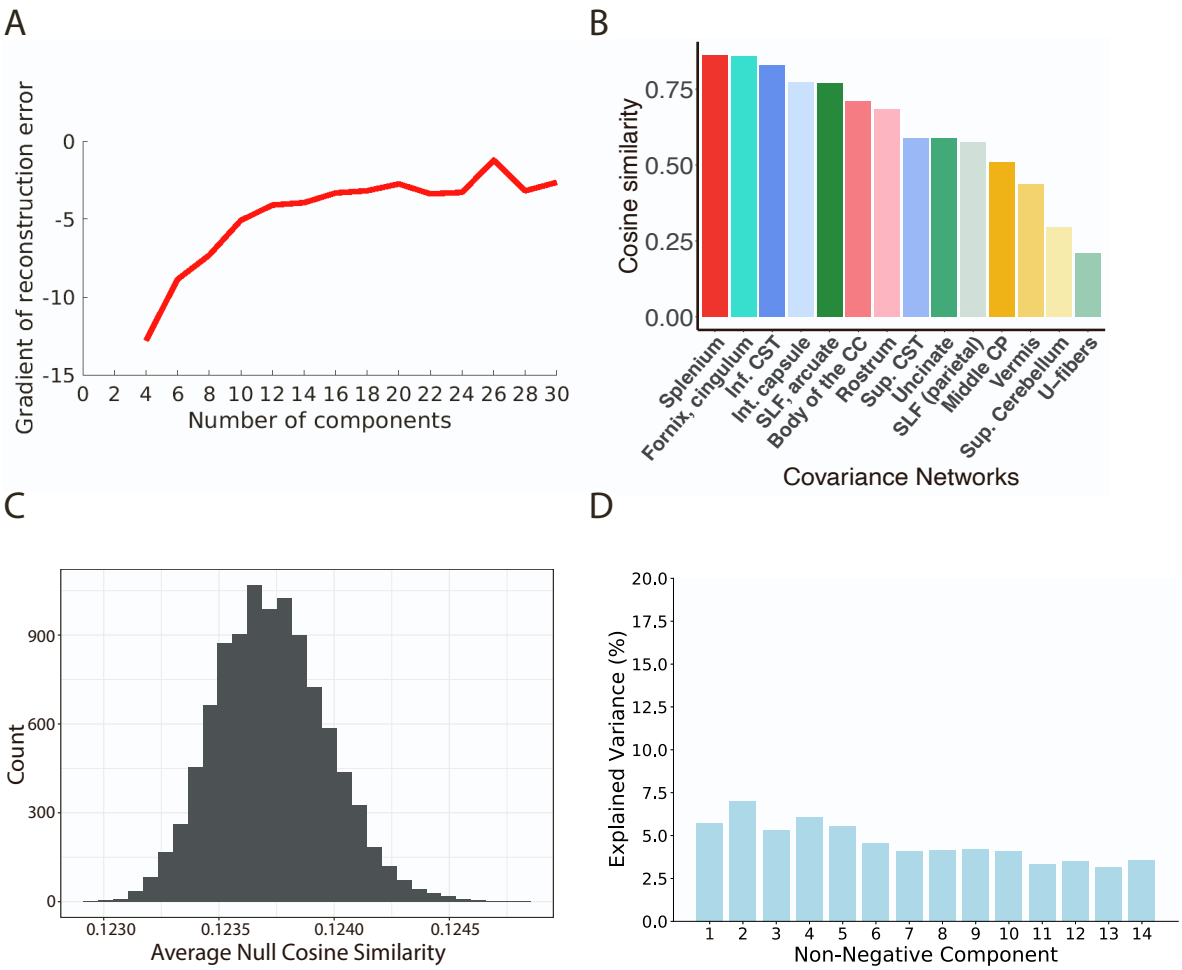

Figure S2. Evaluation of opNMF performance; related to STAR Methods. A) Gradient of reconstruction error for opNMF solutions. Reconstruction error is plotted for opNMF solutions ranging from two to thirty components. The gradient is the difference in reconstruction error of the X matrix (input data) as the opNMF solution increases by 2 components. The y-axis of the plot ranges from -100 to 0. However, to better visualize the differences in reconstruction errors between the different solutions, the y-axis was cropped to -30. As expected, reconstruction error plateaus as the number of components increases. The reconstruction error between the 10- to 30-components are fairly similar. We chose the 14-components solution as it is the most parsimonious solution before a small drop in reconstruction error. Accordingly, the 14-network solution was used for all subsequent analyses. **B)** Cosine similarity scores from the split-half analysis indicating the stability of FDC covariance networks. Cosine similarity ranges from 0 to 1, with higher values indicating greater stability of a given component across different data splits. **C**) Histogram of null cosine similarity values computed on permuted W matrices from data splits 1 and 2. The observed cosine similarity value (0.61) was 1995.3 standard deviations from the null distribution of cosine similarity values (< 0.125), confirming that the selected 14 components are far more stable than would be expected by chance. **D)** Variance explained by each of the 14 fiber covariance networks**.** Non-negative matrix factorization produces a parts-based representation of the data, where the variance of the data is distributed fairly evenly across each component.

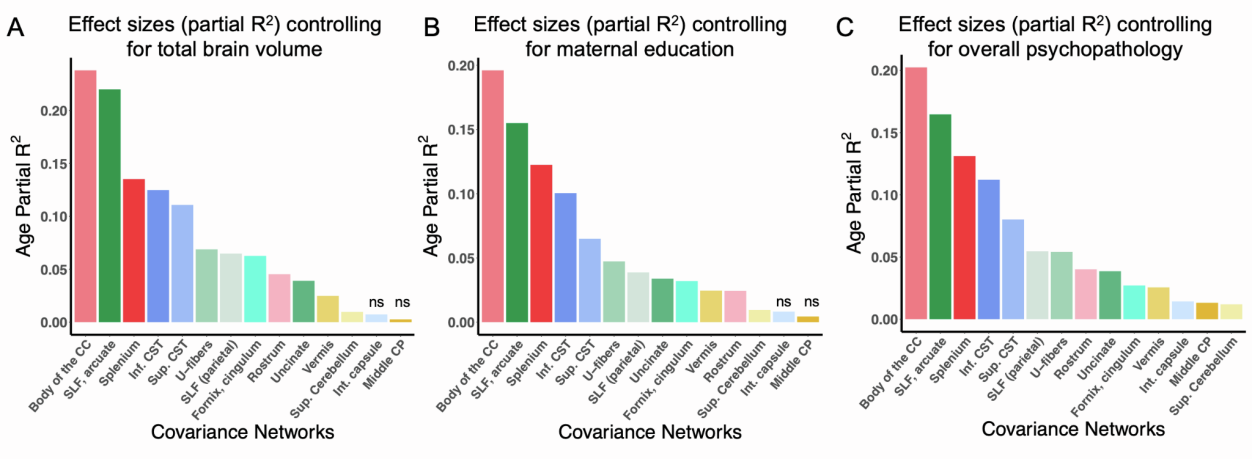

Figure S3. Effect sizes (partial R^2) of age in each fiber covariance network controlling for total brain volume, maternal education, and overall psychopathology; related to Figure 3. **A**) Bar graph depicting the effect size of the developmental effect for each network (partial R^2) while controlling for total brain volume. **B**) Bar graph depicting the effect size of the developmental effect for each network (partial R^2) while controlling for maternal education. **C**) Bar graph depicting the effect size of the developmental effect for each network (partial R^2) while controlling for overall psychopathology. Non-significant associations are marked by "ns". Abbreviations: CC, corpus callosum; SLF, superior longitudinal fasciculus; CST, cortico-spinal tract; Sup, superior; Int, internal; CP, cerebellar peduncle.

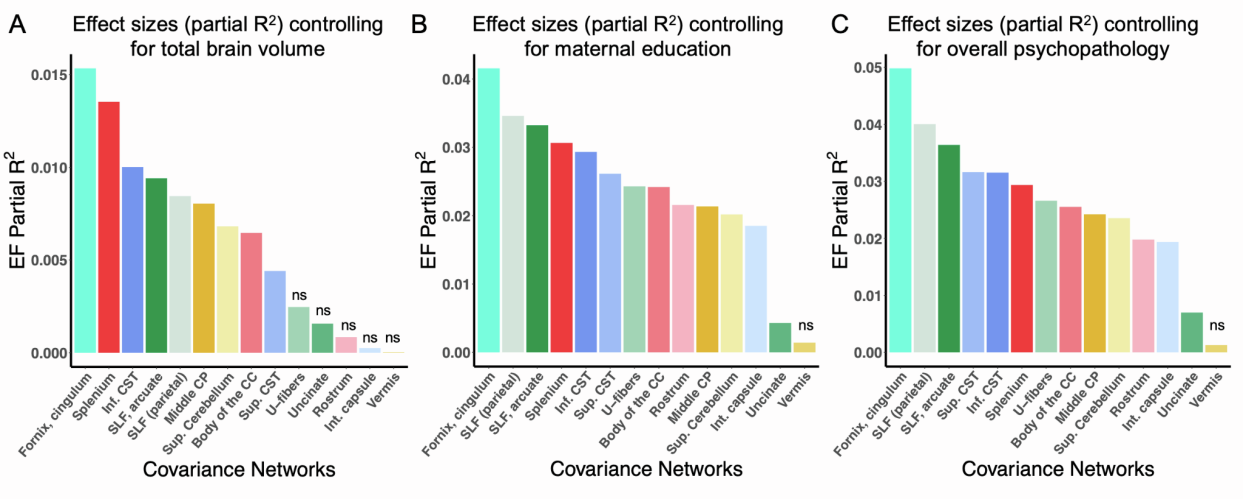

Figure S4. Effect sizes (partial R^2) of executive function in each fiber covariance network controlling for total brain volume, maternal education, and overall psychopathology; related to Figure 5. A) Bar graph depicting the effect size of executive function for each network (partial R^2) while controlling for total brain volume. **B**) Bar graph depicting the effect size of executive function for each network (partial R^2) while controlling for maternal education. **C**) Bar graph depicting the effect size of executive function for each network (partial R^2) while controlling for overall psychopathology. Non-significant associations are marked by "ns". Abbreviations: EF, executive function; CC, corpus callosum; SLF, superior longitudinal fasciculus; CST, cortico-spinal tract; Sup, superior; Int, internal; CP, cerebellar peduncle.

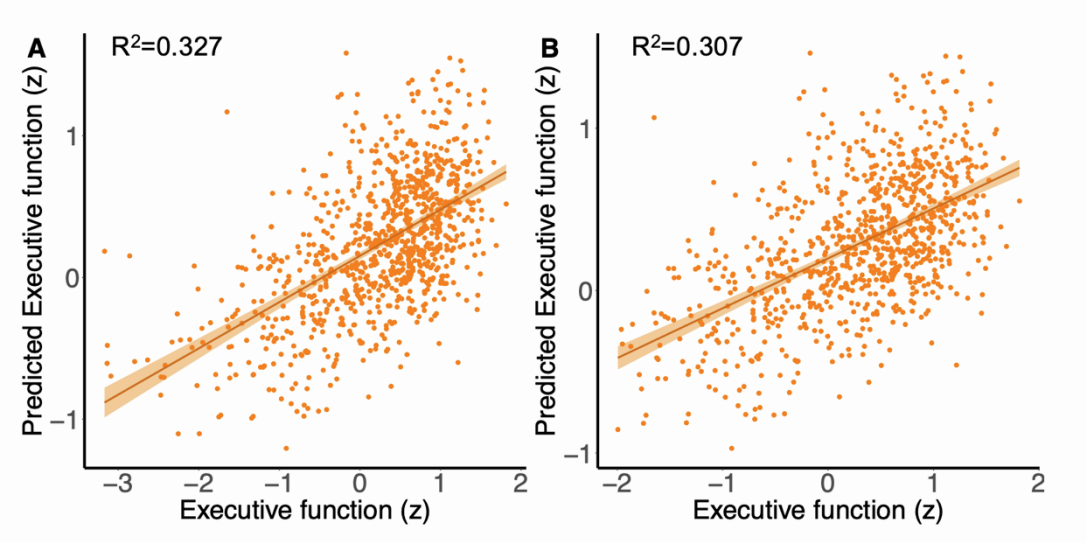

Figure S5. Assessment of potential outliers in the multivariate fiber covariance networks prediction of executive function; related to Figure 5. **A)** Executive function prediction with actual executive function scores on the x axis and predicted executive function scores on the y axis, including all data. In the original model including the full sample, we found a significant difference between a reduced covariate-only model (i.e., sex, motion, and image quality) and a full model that included both the fiber covariance networks and covariates (F=6.56, df=14, p<0.001). The proportion of variance in executive function explained by the 14 covariance networks was R^2=0.327. **B)** Prediction of executive function excluding participants with z-scores below -2. F-test results (F=6.54, df=14, p<0.001) and proportion of explained variance (R^2=0.307) were quite similar, indicating that the outlying data points did not play an outsized role in model predictions.