Supporting Information

Meda et al. 10.1073/pnas.1313093111

SI Materials and Methods

Study Sample. The study protocol was approved by the Institutional Review Board at each local site. After a complete description of the study was given to volunteers, their written informed consent was obtained. In addition to the resting state functional MRI (RS-fMRI), blood samples from each volunteer were collected and stored for genetic analysis.

Data Acquisition and Pre-Processing. ***Imaging.*** All subjects at each site underwent a single 5-min run of RS-fMRI on a 3-T scanner (1). Participants were instructed to keep their eyes open, focus on a crosshair displayed on a monitor, and remain still during the entire scan. In addition, head motion was restricted with a custom-built head-coil cushion. Alertness during the scan was confirmed immediately afterward. If necessary, the scan was repeated. These instructions helped reduce head motion, prevented subjects from falling asleep, and served as an experimental control over visual input. Scanning protocols were mostly consistent; a few differences between sites noted in Table S4 were accounted for during analysis.

The initial six images, during which T2 effects stabilized, were discarded. Remaining images were reconstructed offline and realigned with INRIAlign (www-sop.inria.fr/epidaure/software/INRIAlign/index.php), implemented in Statistical Parametric Mapping (SPM8; www.fil.ion.ucl.ac.uk/spm/software/spm8/). Realigned images then were corrected for differences in slice timing, were spatially normalized to MNI space, and were smoothed by an 8-mm isotropic Gaussian kernel. At each stage the output was validated visually, and scans were discarded if they did not meet quality control (QC) standards (e.g., regarding artifacts, noise, excessive motion, missing data, and so forth). After QC, 1,305 subjects were selected for further imaging analysis.

Genetics. Blood samples were drawn after fMRI scanning. A subset ($n = 620$) of those who underwent fMRI scanning was genotyped using an Illumina's Human Omni1-Quad Bead Chip for 1,140,419 common SNPs at Genomas Inc., Hartford Hospital. Genomic markers (SNPs) were interrogated although a two-step detection process. Carefully designed 50-mer probes were selectively hybridized to the loci of interest, stopping one base before the interrogated marker. Marker specificity was conferred by enzymatic single-base extension to incorporate a labeled nucleotide. Subsequent dual-color fluorescent staining allowed the labeled nucleotide to be detected by the Illumina's iScan imaging system.

Genotyped data were preprocessed in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) following a published workflow reported by Anderson et al. (2), combining both per-individual and per-marker QC (Fig. 1). After preprocessing, SNPs in high linkage disequilibrium were removed (window of 50 SNPs; $r^2 > 0.5$) to increase independence between markers and then were subjected to a principal component analysis (PCA) using custom Matlab scripts to identify stratifying factors using an algorithm similar to EIGENSTRAT. Data were adjusted using the top three eigenvectors to exclude any underlying stratification, and SNPs then were statistically prioritized for parallel independent component analysis (para-ICA) using logistic regression across each proband group separately. All SNPs that were both unique and common to both disorders and associated significantly at a nominal $P < 0.05$ uncorrected threshold were retained for further analysis, achieving three important goals: (*i*) effectively restricting core analyses to potentially disease-related genetic data in the current sample; (*ii*) reducing potential noise from other interacting genes with minimal relationship to the disease model, thereby providing hypothesis-free data-driven enhancement; and (*iii*) improving accuracy and the linking coefficient of the para-ICA algorithm. After genetic QC, 549 subjects and 10,136 SNPs remained for the final para-ICA analysis. Fig. 1 illustrates the general processing workflow.

Primary Data Analysis. ***ICA.*** fMRI data were analyzed using a group–ICA algorithm (GIFT v1.3f; http://mialab.mrn.org/software/gift) to identify spatially independent networks by pooling data from all participants ($n = 1,305$) into a single ICA analysis. After data reduction through two PCA stages, that typically retains 99% of data variance (3), 20 mutually independent components were determined using minimum description length (MDL) criteria adjusted to account for correlated samples (4). Time courses and spatial maps (SM) were then back-reconstructed for each participant, maintaining interindividual variability. The SM loading coefficients from a group ICA represent the degree of connectivity of each voxel's time course with the aggregate/overall time course of the network, thus inherently representing a functional connectivity map. Each network's consistency metrics were derived from multiple ICA runs using a clustering approach called ICASSO (5).

Default mode network selection. Because we were specifically interested in default mode networks (DMNs), all aggregate SMs derived from the group ICA were spatially correlated to a default mode intrinsic network map generated using an ICA decomposition of modeled activation images achieved using the recently published Brainmap behavioral metadatabase coordinates (6). We identified DMNs by assessing all SMs that correlated highly (and that also were visually verified using a scree method) with the previously published metanetwork DMN map from the above study.

Conventional analysis of DMNs. Each DMN was subjected to an SPM8 full-factorial model comprising two factors (diagnosis and site), each with five levels. F-contrasts were generated to detect voxels showing significant diagnosis-by-site interaction, which, if found, were removed from subsequent analyses. Remaining voxels were subjected to a permutation-testing model used in FMRIB Software Library's (FSL) Randomize (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise) to test for a voxelwise main effect of diagnosis across all groups, using the threshold-free cluster enhancement (TFCE) option that minimizes potential bias toward threshold dependence and localization of cluster inference (7). Voxels surviving a $P < 0.05$ familywise error adjusted for multiple comparisons threshold from an omnibus F-test contrast were averaged and imported to SPSS18 (www-01.ibm.com/software/analytics/spss/) for further post hoc tests. Statistical tests were conducted on the regional measurements extracted in the previous step to investigate our biomarker and/or endophenotype hypothesis using an ANOVA model and subsequent post hoc *t* tests. All these analyses were adjusted for age, sex, and site.

Heritability. Heritability (h^2) estimates for the regional connectivity measurements for each DMN were computed using Sequential Oligogenic Linkage Analysis Routines (SOLAR) (8), using maximum likelihood variance-decomposition methods to modeling covariance among family members as a function of kinship.

Correlation analyses with clinical measures. Average DMN connectivity measures were correlated against symptom measures obtained on scanning day, including the Young Mania Rating Scale (YMRS) (9), Montgomery-Asberg Depression Rating Scale (MADRS) (10), Social Functioning Scale (SFS) (11), Schizo-

Bipolar Scale (SBS) (12), and Positive and Negative Symptom Scale (PANSS) (13).

Parallel-ICA (imaging–genetics). Because no relatives were genotyped, genetic relationships and biological processes were derived only for controls and probands. To identify genetic underpinnings of DMNs identified using ICA, we used separate para-ICAs (14–19) on each DMN using Fusion ICA Toolbox (FIT: http://mialab.mrn.org/software/fit/) in MATLAB 7.7, an algorithm validated in imaging–genetic studies encompassing a range of sample sizes (20–23). Para-ICA is a second-level analysis that essentially allows investigation of cross information between two different data modalities. The goal of para-ICA was threefold: (*a*) to run an ICA on the subject-specific back-reconstructed DMN to extract fMRI subnetworks that are maximally spatially independent and measure variation/modulation in connectivity across individuals; (*b*) simultaneously to run another ICA on the SNP-subject matrix to extract distinct, linear combinations of SNP data (covarying genetic networks across subjects); and (*c*) iteratively to update the unmixing matrices derived from *a* and *b* to maximize correlations between the derived networks from steps *a* and *b* until a stopping criterion is reached. This process results in a number of components for each feature set that are variably expressed across subjects and quantified by a subject-level loading coefficient for each data type. The variation of expression (estimated loading parameter) of a single component for one data type is correlated across subjects with the expression of a single component from the second data type, resulting in pairs of correlated components from each data modality. The number of genetic components in the current study was estimated to be 14, and the number of components estimated for the fMRI networks varied between two and three depending on the input DMN. Component estimation used standard MDL criteria (4). The residual loading parameters for each feature (imaging, genetic) were tested post hoc for between-group differences across probands using one-way ANOVA after adjusting for age, sex, and site effects in SPSS-18, per prior publications (21–23). Fig. S3 illustrates the para-ICA procedure.

Gene annotation, visualization, and enrichment. During para-ICA, voxels within each DMN subnetwork and SNPs within each genetic subnetwork were assigned Z-scores according to their contribution toward a particular network's overall signal. All gene components were arbitrarily thresholded at a $|Z|>2$ cutoff, and SNPs surpassing this threshold were deemed to be contributing significantly to that particular genetic network. All SNPs were mapped to their respective genes using the genome variation server database (http://gvs.gs.washington.edu/GVS137/).

Before enrichment analysis all SNPs along with their respective *P* values for each para-ICA network were fed into the VEGAS software (http://gump.qimr.edu.au/VEGAS/) to derive a gene-based trait association value that allowed us to account for any bias in gene size. Only genes with a $P < 0.05$ value were input for further enrichment analysis. Enrichment and visualization of significant SNPs mapped onto their respective genes were carried out using Metacore, (GeneGo Inc: www.genego.com/metacore.php) to interpret results in the context of a curated biological knowledge base. Functional enrichment was derived in multiple ontologies, including pathways, network processes, diseases, gene ontology processes, and metabolic networks. Quantitative enrichment scores were calculated using a hyper-geometric approach to estimate the likelihood that significant genes were overrepresented in particular biological pathways, networks, or processes. Significance values were adjusted using false discovery rate (FDR) correction for multiple comparisons (24). This enrichment analysis was done using genes both (*i*) mapped within each network separately and (*ii*) pooled across all DMN subnetworks. The latter strategy provided a global perspective of possible biological processes mediating disease risk via DMN connectivity.

1. Tamminga CA, et al. (2013) Clinical phenotypes of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *Am J Psychiatry* 170 (11):1263–1274.
2. Anderson CA, et al. (2010) Data quality control in genetic case-control association studies. *Nat Protoc* 5(9):1564–1573.
3. Calhoun VD, Liu J, Adali T (2009) A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* 45(1, Suppl):S163–S172.
4. Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001) A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp* 14(3):140–151.
5. Himberg J, Hyvärinen A, Esposito F (2004) Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage* 22(3): 1214–1222.
6. Laird AR, et al. (2011) Behavioral interpretations of intrinsic connectivity networks. *J Cogn Neurosci* 23(12):4022–4037.
7. Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* A44(1):83–98.
8. Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62(5):1198–1211.
9. Young RC, Biggs JT, Ziegler VE, Meyer DA (1978) A rating scale for mania: Reliability, validity and sensitivity. *Br J Psychiatry* 133:429–435.
10. Montgomery SA, Asberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* 134:382–389.
11. Birchwood M, Smith J, Cochrane R, Wetton S, Copestake S (1990) The Social Functioning Scale. The development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients. *Br J Psychiatry* 157:853–859.
12. Keshavan MS, et al. (2011) A dimensional approach to the psychosis spectrum between bipolar disorder and schizophrenia: The Schizo-Bipolar Scale. *Schizophr Res* 133(1-3):250–254.
13. Kay SR, Fiszbein A, Opler LA (1987) The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 13(2):261–276.
14. Allen EA, et al. (2011) A baseline for the multivariate comparison of resting-state networks. *Front Syst Neurosci* 5:2.
15. Calhoun VD, Eichele T, Adalı T, Allen EA (2012) Decomposing the brain: Components and modes, networks and nodes. *Trends Cogn Sci* 16(5):255–256.
16. Guo Y, Pagnoni G (2008) A unified framework for group independent component analysis for multi-subject fMRI data. *Neuroimage* 42(3):1078–1093.
17. Joel SE, Caffo BS, van Zijl PC, Pekar JJ (2011) On the relationship between seed-based and ICA-based measures of functional connectivity. *Magn Reson Med* 66(3):644–657.
18. Khadka S, et al. (2013) Is aberrant functional connectivity a psychosis endophenotype? A resting state functional magnetic resonance imaging study. *Biol Psychiatry* 74(6):458–466.
19. Meda SA, et al. (2010) A pilot multivariate parallel ICA study to investigate differential linkage between neural networks and genetic profiles in schizophrenia. *Neuroimage* 53(3):1007–1015.
20. Jamadar S, et al. (2011) Genetic influences of cortical gray matter in language-related regions in healthy controls and schizophrenia. *Schizophr Res* 129(2-3):141–148.
21. Liu J, et al. (2009) Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Hum Brain Mapp* 30(1):241–255.
22. Rzepecki-Smith CI, et al. (2010) Disruptions in functional network connectivity during alcohol intoxicated driving. *Alcohol Clin Exp Res* 34(3):479–487.
23. Meda SA, et al. (2012) A large scale multivariate parallel ICA method reveals novel imaging-genetic relationships for Alzheimer's disease in the ADNI cohort. *Neuroimage* 60(3):1608–1621.
24. Reiner-Benaim A (2007) FDR control by the BH procedure for two-sided correlated tests with implications to gene expression data analysis. *Biom J* 49(1):107–126.

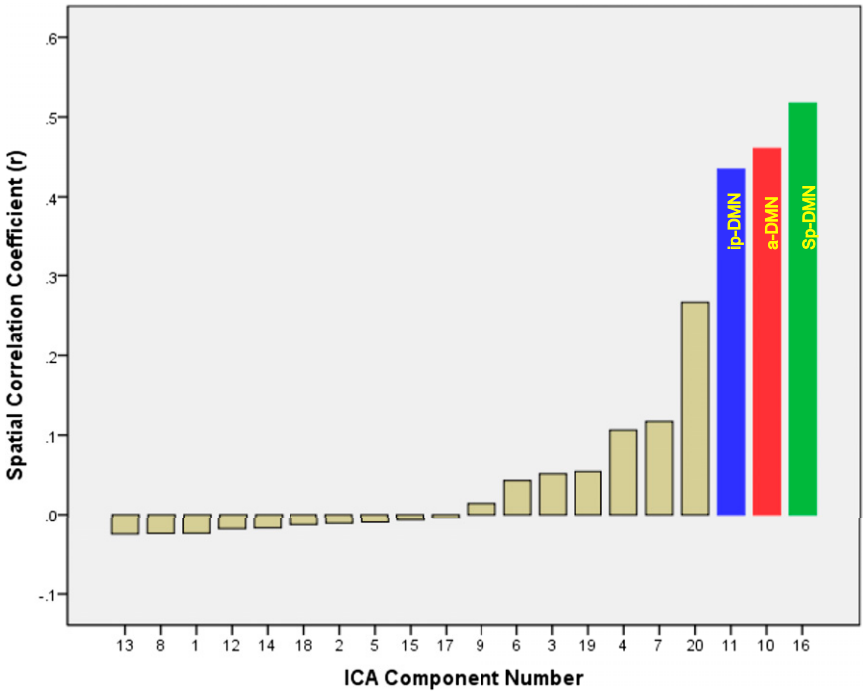

Fig. S1. Distribution of spatial correlation coefficients (Pearson *r* values) between the DMN reported in Laird et al. (6) and all intrinsic ICA-derived networks from the initial global ICA. Intrinsic DMN s identified and reported in the present study are color coded per designations in the main text.

Enriched Ontology	Para-ICA Sub-Network
Development	• aDMNS1, aDMNS2, ipDMNS1, spDMNS1, spDMNS2
Cell Adhesion	• aDMNS1, ipDMNS1, spDMNS1
Smooth Muscle Contraction	• aDMNS2, spDMNS1, spDMNS2
Signal Transduction	• aDMNS1, ipDMNS1
Neurophysiologic process	• aDMNS1, spDMNS1
G-Protein Signalling	• aDMNS1, ipDMNS1
Transport	• spDMNS1
Transcription	• ipDMNS1
Apoptosis	• ipDMNS1
Cytoskeleton	• aDMNS1
Cell Cycle	• ipDMNS1
Immune Response	• ipDMNS1
cAMP/Ca2+	• spDMNS1

Fig. S2. A bird's-eye view of different processes significantly enriched (FDR <0.05) within each genetic subnetwork identified using para-ICA. Note the overall presence of certain processes such as development (mostly genes regulating neurodevelopment, guidance, and so forth) common to all subnetworks.

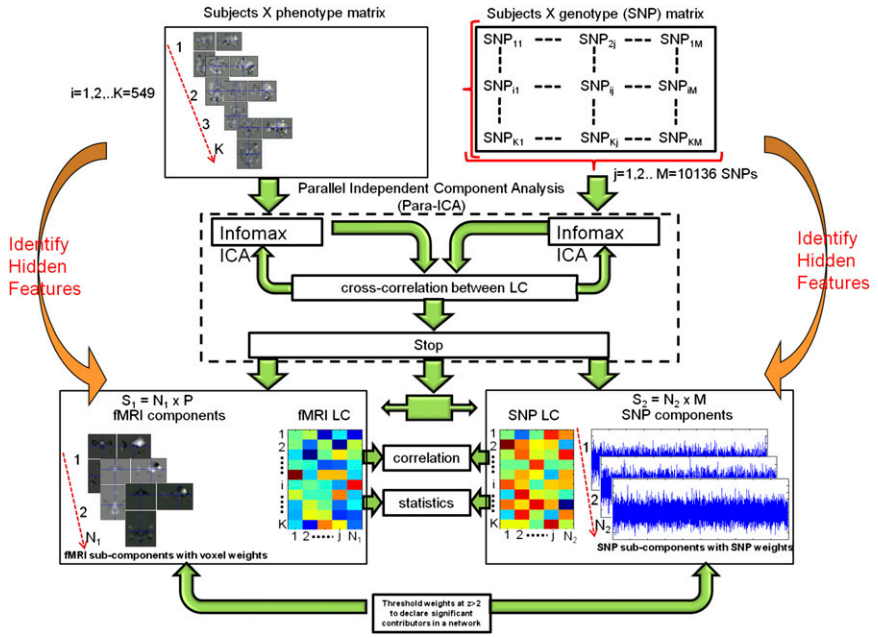

Fig. S3. A pictorial depiction of the para-ICA (fusion) procedure used to derive genotype–phenotype relationships in the current study.

Other Supporting Information Files

Table S1 (DOC)
Table S2 (DOC)
Table S3 (DOC)
Table S4 (DOC)