Diverse types of genetic variation converge on functional gene networks involved in schizophrenia

Sarah R. Gilman, Jonathan Chang, Bin Xu, Tejdeep S. Bawa,

Joseph A. Gogos, Maria Karayiorgou, Dennis Vitkup

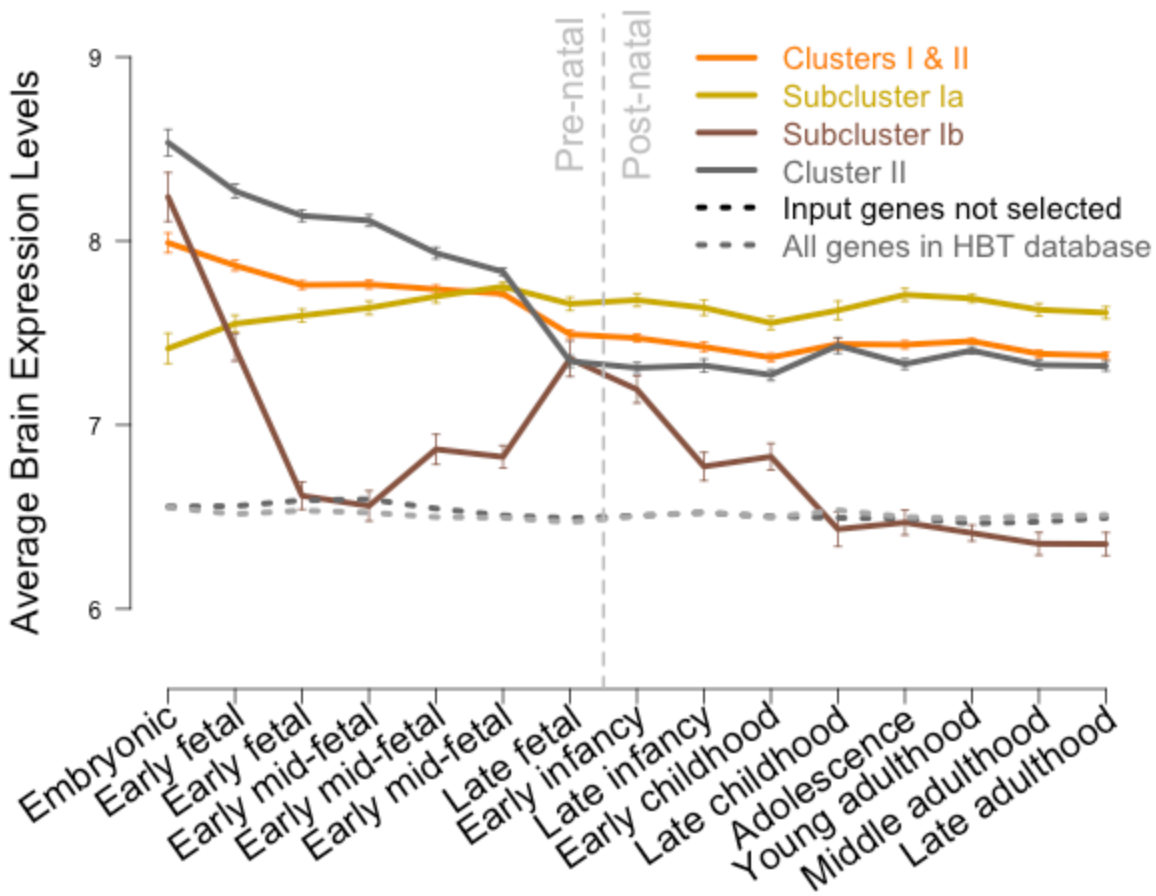


Figure S1. Temporal brain expression profiles for genes forming the identified clusters (average expression). Gene expression profiles in the brain across developmental stages. Gene expression data were obtained from the Human Brain Transcriptome database (hbatlas.org). Average gene expression across all samples at each developmental stage is shown. Error bars in the figure represent standard error of the mean across samples, and where applicable all genes, at a given developmental stage.

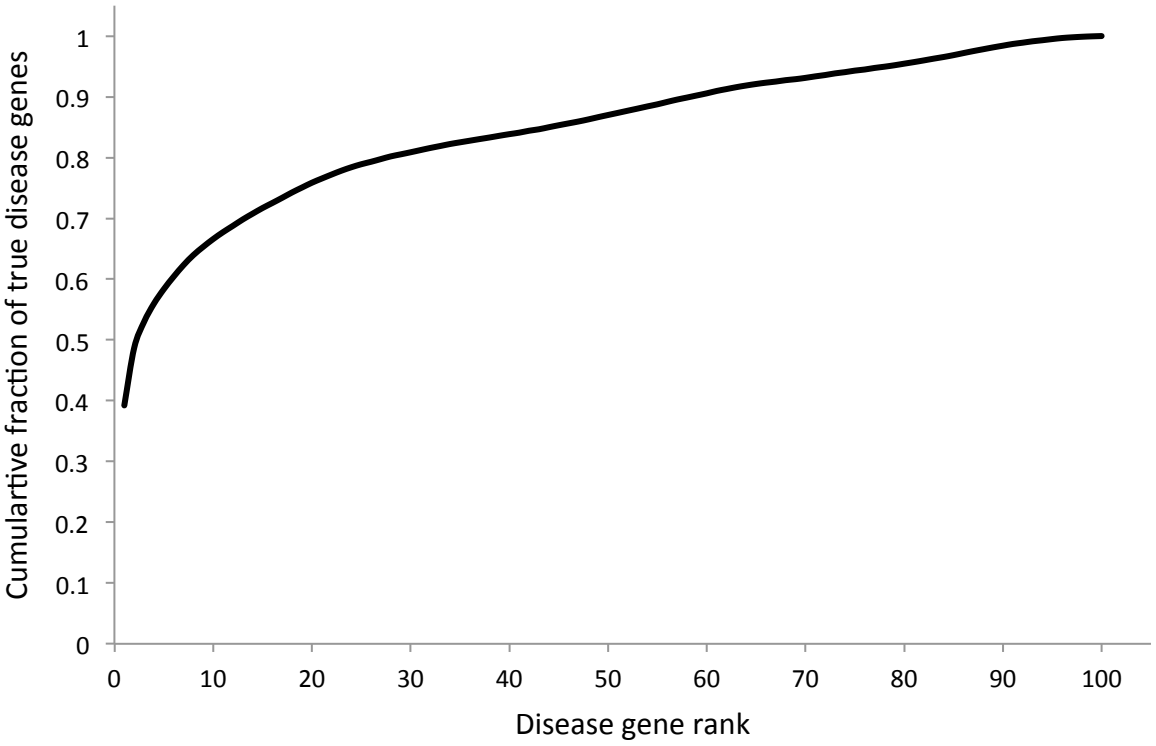
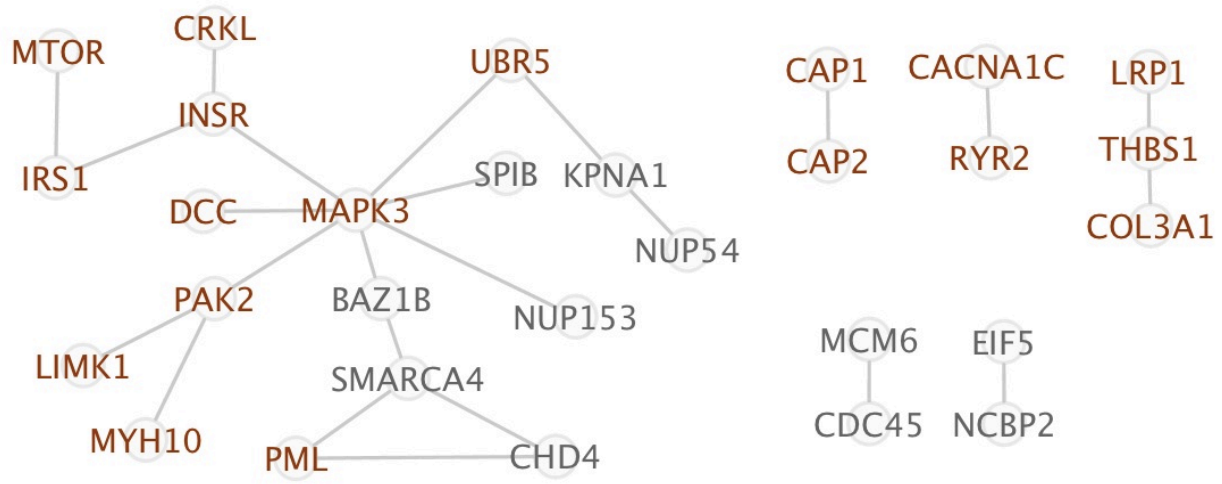


Figure S2. Gene ranking for a diverse collection of phenotypes from OMIM. We demonstrated the ability of the underlying phenotype network and our scoring method to prioritize candidate genes responsible for a diverse set of genetic phenotypes. For each gene in the test set – composed of 74 genetic phenotypes with 338 associated genes (see Supplementary Table S12) - we randomly selected as decoys 99 human genes with comparable network connectivity. We then evaluated how high the correct gene could be ranked, among the decoys, based on the strength of network connections to other OMIM genes responsible for the same phenotype. For each disease gene in the test set we repeated the ranking test 100 times and calculated the average prioritization rank. The cumulative results for the test set are presented in the figure. The correct gene was ranked as the top gene (out of 100 genes) in 39% of the cases, within the top 3 in 53% of the cases, and within the top 10 in 66% of the cases.



Supplementary Figure S3. Protein-protein between cluster genes. The figure shows direct protein-protein interactions among the genes forming cluster I (brown) and II (grey) in Figure 1, main text. Interactions were obtained using three sources: BioGRID, DIP, and HPRD.

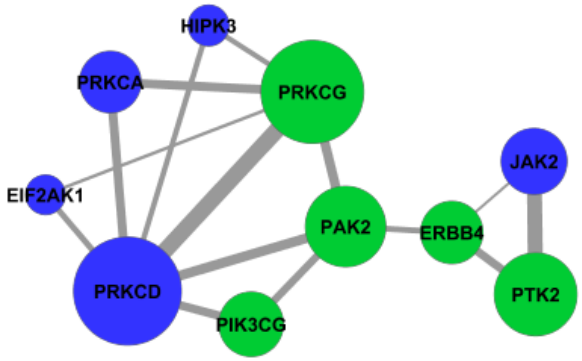


Figure S4. Gene cluster based on the independent set of schizophrenia-associated CNVs. The cluster shown in the figure was identified by our algorithm using an independent set of schizophrenia-associated CNVs, i.e. CNVs not used in the original paper. This set included 35 rare inherited CNVs observed in schizophrenia patients and 13 *de novo* CNVs associated with childhood onset schizophrenia (COS) . The 10-gene cluster (*P*-value=0.05, multiple hypothesis corrected) contained genes both from inherited CNVs (green) and COS CNVs (blue). Edge width in the figure represents the strength of the likelihood score between two genes, and node size represents the gene's contribution to the overall cluster score. The GO functional terms associated with the cluster are given Supplementary Table S13.

Table S1: Clusters identified by the NETBAG+ algorithm.		
Input Events	**Number of genes in cluster**	**Cluster *P*-value**
SNVs + CNVs + GWAS	47	< 0.001
CNVs + GWAS	29	0.013
SNVs + CNVs	37	0.014
SNVs	16	0.056
CNVs	6	0.057
Masked SNVs + CNVs + GWAS	42	0.071
SNVs + GWAS	23	0.079
GWAS	10	0.463
Control SNVs	10	0.568
Control SNVs + CNVs	18	0.572
SNVs + CNVs + GWAS (100kb)	51	< 0.001
SNVs + CNVs + GWAS (450kb)	48	< 0.001

Table S1. Clusters identified by the NETBAG+ algorithm and their significances. The left column shows various combinations of input events, the middle column shows the number of genes in the best cluster, and the right column shows the best cluster *P*-value. The rows in the table are sorted by the *P*-values. "Masked" indicates an input set where the genes forming cluster I were removed. Control SNVs were taken from three sources: *de novo* mutations in unaffected individuals and *de novo* synonymous mutations in probands from the recent paper by Xu *et al.*, and *de novo* mutations in unaffected siblings from two recent autism exome sequencing studies by O'Roak *et al.* and Sanders *et al.* In total, this combined set contained 161 genes. Control *de novo* CNVs were taken from unaffected siblings in a recent autism study, and contained 67 genes in 14 events.

Tables S2. Gene Ontology (GO) terms associated with cluster genes using FuncAssociate. We searched for over-represented GO terms describing the schizophrenia clusters in Figure 1. In the table, N is the number of cluster genes annotated with a given term, X is the number of all human genes with that GO term. All *Padj*-values have been corrected for multiple hypothesis testing based on the number of considered GO terms.

Table S3. Gene Ontology (GO) terms associated with cluster genes using DAVID. We searched for over-represented GO terms describing the schizophrenia clusters in Figure 1, Figure S4 and genes associated to schizophrenia in a recent study of gene expression. This latter set included 596 genes from the recent study by Brennand *et al.* that were differentially expressed in neuron-differentiated human induced pluripotent stem cells compared to matched controls from healthy individuals. In the table, N is the number of cluster genes annotated with a given term, X is the number of all human genes with that GO term. All *Padj*-values in the table were determined using the Benjamini-Hochberg procedure.

Table S4. Likely impact of CNV genes on growth of dendrites or dendritic spines. We performed a literature review to assess the impact of cluster genes associated with *de novo* CNVs on the growth of dendrites and dendritic spines – increase (inc) or decrease (dec). CNV polarity, i.e. deletion (del) or duplication (dupl), allowed us to determine a corresponding change in the gene dosage. CNV-associated genes were taken from either the schizophrenia clusters or the autism cluster identified in our previous work[6]. For the two genes with both duplication and deletion events (CRKL and PIAS3), we used the frequency of CNVs reported in Malhotra *et al.*[9] to determine the predominant polarity associated with each disease. We included Pubmed identifiers for key references.

Table S5. Disease phenotypes used in the prioritization test. Shown are disease phenotypes and associated genes from OMIM. To compile the test set for prioritization we excluded phenotypes used in training our network, phenotypes with less than 3 associated genes, and phenotypes with somatic mutations (such as cancer). In total, the training set contained 74 phenotypes with 338 associated genes.

Table S6. Literature review of genes with *de novo* SNVs. We performed a manual literature review of information for the 159 genes with *de novo* SNVs from the recent study by Xu *et al*. In the table we indicate whether each gene is a member of the cluster I or II (main text, Figure 1), and whether the gene has a known neural or brain-related function. Functional descriptions in the table were primarily taken from GeneBank and NCBI. We also included Pubmed identifiers for key references.