Supplemental Information

Summary of Supplemental Information

Figure S1, related to Figure 1. Further characterization and validation in independent data of co-expression network and modules.

Figure S2, related to Figure 3. Direct and indirect protein-protein interaction networks for RDNV-affected genes in M2 and M3.

Figure S3, related to Figure 3. Additional characterization RDNV-affected genes in M2 and M3, including robustness of RDNV enrichment, P(HI) analysis, and common variant enrichment.

Figure S4, related to Figure 5. Additional enrichment analyses in fetal and primate laminar expression data, including PCW 21 enrichment and adult primate cell-type marker enrichment analyses.

Table S1, related to Figure 1. Additional information relevant to the identification and characterization of transcriptional co-expression networks for cortical development.

Table S2, related to Figure 2, 3, and 4. Gene lists, variant data, and statistics relevant to gene set enrichment analyses.

Table S3, related to Figure 4. TF binding site analysis results. A) For ASD associated modules. B) Individual TF-motif information driving the module-level enrichments.

Table S4, related to Figure 7. Example prioritization of RDNV affected genes from M2 and M3.

Supplemental Data

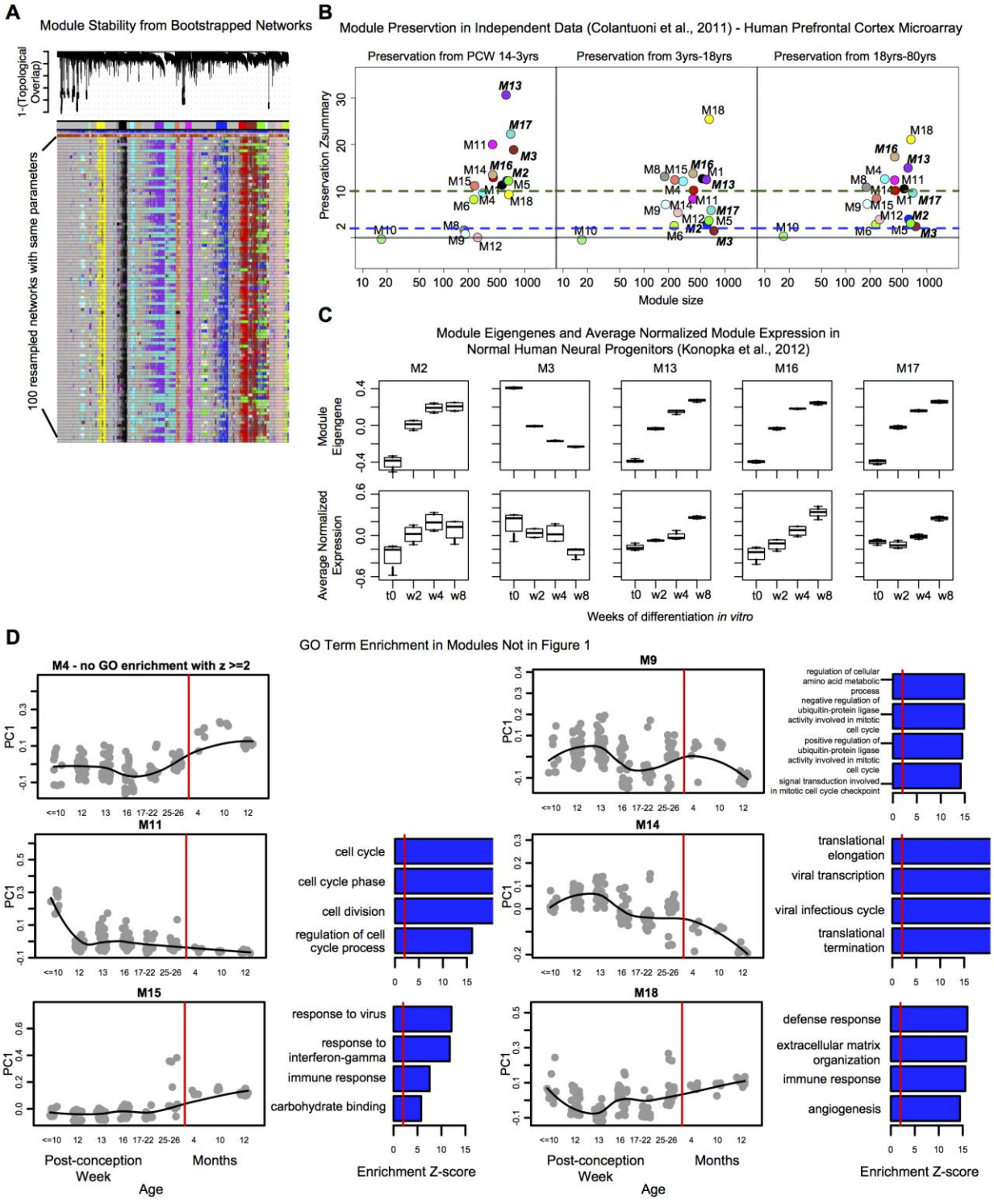


Figure S1, related to Figure 1. Further characterization and validation in independent data of co-expression network and modules.

A) Module robustness analysis was carried out by reconstructing networks with the same parameters and randomly resampling from the initial set of samples (as described in Langfelder and Horvath, 2012). Modules were found to be reproducible with perturbations to the initial individual subject, regional, and temporal structure, and the fraction of times each gene was assigned to the same module is reported in Table S1A.

B) Module preservation analysis was used to calculate the $Z_{summary}$ statistic for each module (Langfelder et al., 2011). This permutation test assesses whether module density and intramodular connectivity are preserved. An advantage of the $Z_{summary}$ statistic is that it allows us to rigorously argue that a module is not preserved ($Z_{summary} < 2$), if it is moderately preserved ($Z_{summary} > 2$), or if it is highly preserved ($Z > 10$). We applied the module preservation analysis to assess whether our modules are preserved in three epochs from an independent dataset of prefrontal cortex microarray spanning development (Colantuoni et al., 2011). At the first time window (Panel B, left), many modules were preserved, while at later time points some modules were not preserved. Of note, M13 and M16 are highly preserved at all time points, while M2 and M3 are highly preserved at the earliest time point, and moderately or weakly preserved thereafter. This non-preservation of M2 and M3 in later time points is also seen in the adult BrainSpan data (TableS1B). Bonferonni-corrected p-values for these Z-scores are reported in Table S1B. Modules that were enriched for ASD risk genes are shown in bolded italics.

C) Module preservation analysis in Table S1B indicates that modules are preserved during normal human neural progenitor (NHNP) development (Konopka et al., 2012), despite differences in *in vivo* and *in vitro* neural development. Plots of expression trajectories based on the eigengene and the average normalized expression in modules enriched for ASD risk genes show that NHNPs show a similar temporal trend as early *in vivo* development (PCW 8-20). Compare modules from Figure S1D to Figure 1C. This suggests that modules, at least in part, reflect the differentiation of neuronal progenitors to neurons.

D) Module eigengenes and gene ontology for remaining 6 modules that pass 2/3 replication criteria (Extended Experimental Methods). For each plot, the eigengene trajectory and gene ontology are plotted, with top 4 GO terms passing $Z > 2$ (for FDR values see Table S1B). See Table S1B for eigengenes and additional details for GO term enrichment.

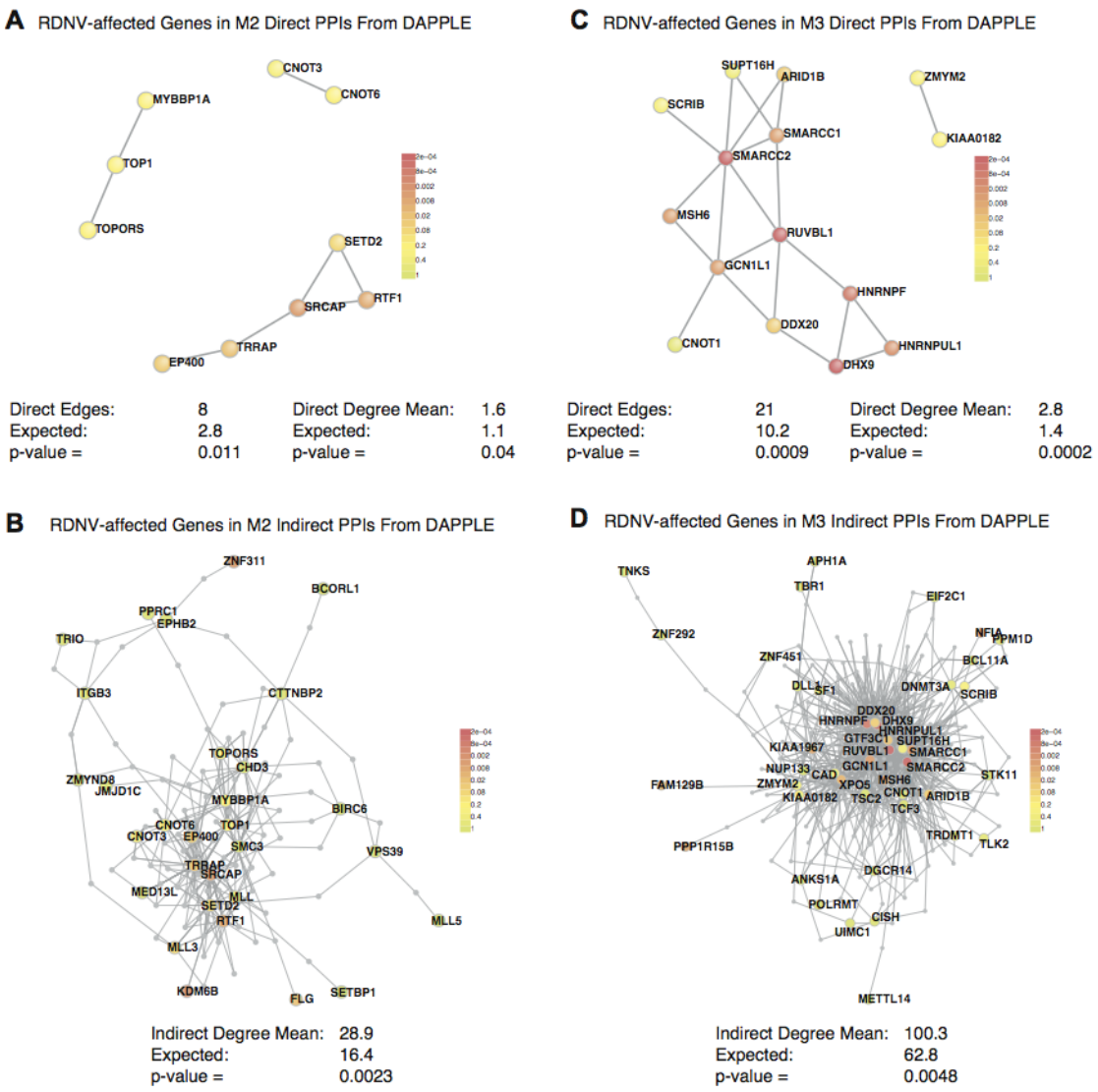


Figure S2, related to Figure 3. Direct and indirect protein-protein interaction networks for RDNV-affected genes in M2 and M3 as obtained from InWeb PPIs via DAPPLE (Rossin et al., 2011).

A) M2 direct interaction network obtained by inputting genes in M2 with a variant identified in the combined set of protein disrupting and missense RDNVs.

B) M2 indirect interaction network, which allows for one node to be skipped when calculating enrichment and network relationships.

C-D) Same as A-B, for M3.

Expected values and p-values are calculated via the DAPPLE's within-degree within-node permutation methodology that allows ranking of PPI hubs by p-value. PPI hubs in the RDNV sub-network from M2

with p < 0.01 include: *KDM6B, SRCAP, ZNF311, and RTF1*. PPI hubs for the RDNV sub-network from M3 with p < 0.01 include: *DHX9, RUVBL1, SMARCC2, HNRNPF, HNRNPUL1, MSH6, SMARCC1, GCN1L1, NFIA, KIAA1967, PPP1R15B, FAM129B, XPO5*.

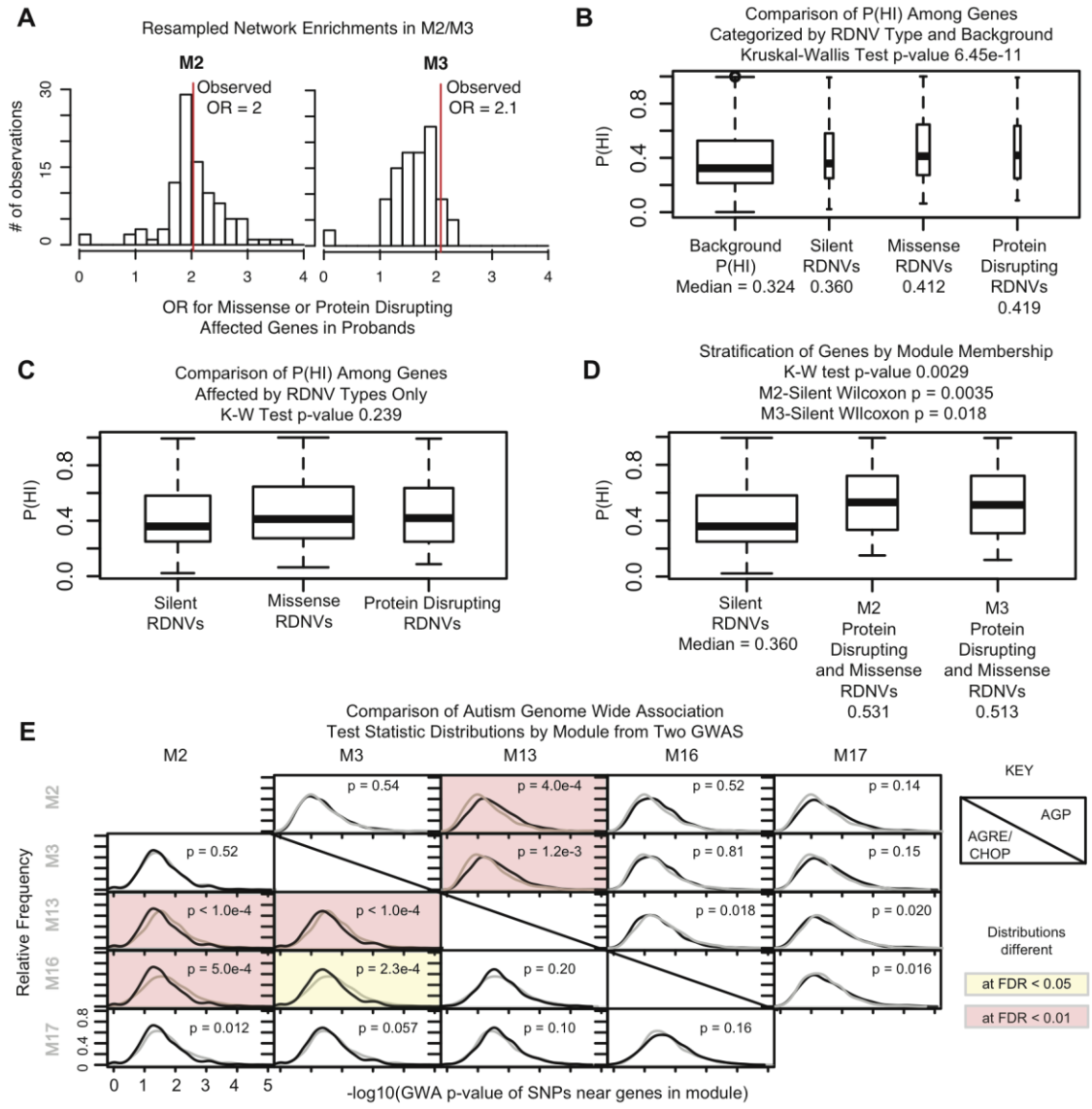


Figure S3, related to Figure 3. Further characterization of ASD risk gene enrichment in modules. P(HI) score comparisons.

A) Enrichment for Protein Disrupting and Missense RDNV affected genes from probands in M2 and M3 from the 100 resampled networks shows that the enrichment is robust to perturbations in network structure.

B-C) In B), comparison of P(HI) scores among background and three mutation-affected gene categories reveals a significant difference, but in C) a comparison of scores excluding background reveals no significant difference, suggesting that all RDNV-affected gene sets, including those affected by silent variants, are predicted to be more deleterious than background. Thus, stratification of protein disrupting and missense RDNV affected genes distinguish them from the silent set requires additional information.
D) Stratifying the RDNV affected gene sets by co-expression relationships, based on membership in M2 or M3, yields a significantly elevated P(HI) score both sets compared to the silent set.
E) Differential enrichment for common variants suggests that M13 is enriched for common variant genome-wide association (GWA) signal compared to M2 and M3 (see Results and Extended Experimental Procedures for details). M16 is also preferentially affected by common variation, but only in one GWA. AGP refers to the GWA performed in by Anney et al. (Anney et al., 2012), while AGRE/CHOP refers to a GWA by Wang et al. (Wang et al., 2009). Of note, the latter set was re-analyzed with samples overlapping the AGP cohort removed to establish independent sets.

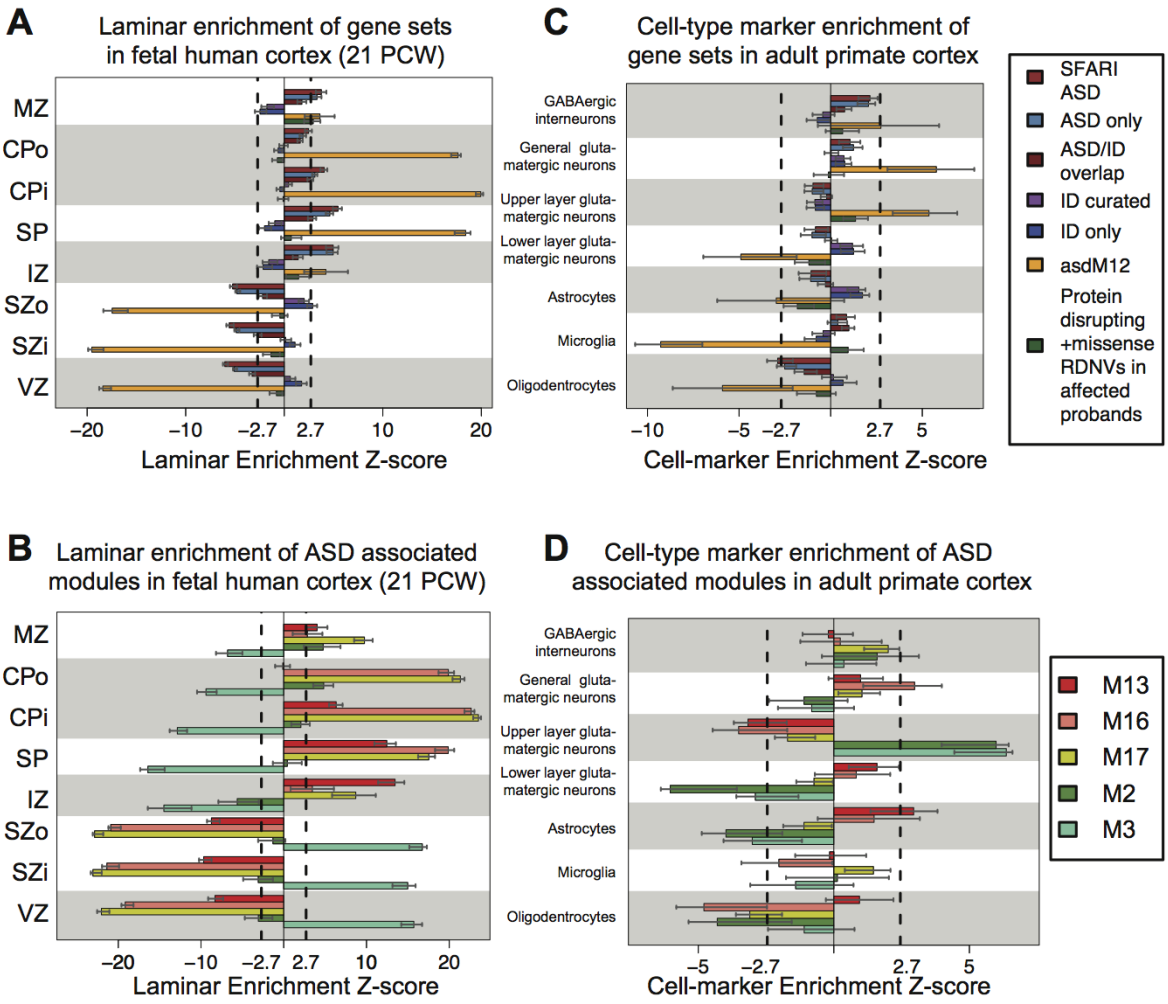


Figure S4, related to Figure 5. Laminar enrichment for PCW 21 and cell-type marker enrichment in adult primate cortex.

A-B) Laminar enrichment at PCW 21 shows an identical pattern of enrichment to PCW 15/16 for candidate genes and B) modules.

C-D) Cell-type marker enrichment in the adult primate cortex for candidate genes and D) modules. This enrichment was calculated using the distribution of Pearson correlation values to the first principal components of the set of cell markers for each cell-type as delineated in Table S1.

Z > 2.7 is equivalent to an FDR < 0.01, and 95% CIs were derived from bootstrapping the underlying expression data 10,000 times, as in Figure 5. For detailed t-values for each gene as well as cell markers used for Figures S4C and D, see Table S1.

See associated Excel file for all tables.

<TableS1.xlsx>

Table S1, related to Figures 1. Network analysis details and analysis results for genes and modules compiled across analyses.

A) Network statistics, gene set membership, laminar specificity, and cell-type enrichment information for all 22,084 protein-coding and noncoding transcripts. For each gene, this includes (in order from left to right) gene ID and associated information from GENCODEV10, module assignment (by color and label) and robustness (P(Assigned to same module in resampled network)), average expression level at multiple temporal epochs, square root of the adjusted R^2 when using neocortical region as the factor in a linear model with expression values as the outcome, correlation of the transcript to RIN, connectivity to the module eigengene (kME, correlation to the ME, a measure of centrality in the module) across all modules (all related to Figure 1), membership in candidate and RDNV gene sets (related to Figures 2 and 3), t-values for layer-specificity from limma differential expression analysis (Figure 5, Figure S3A,B), and r values from correlation to marker genes from cell-type marker analysis (Figure S3C,D). In the Module Label column, genes assigned to M7, the grey module, are marked with a "-" as the grey module represents the set of genes not strongly co-expressed with other genes.

B) Module-level statistics, including GO term enrichment, enrichment of PPIs, module preservation, and module eigengenes across all samples. This includes (in order from left to right), a summary of the independent levels of validation supporting each module, the proportion of variance explained by the ME, the top 5 GO terms from GO Elite, PPI enrichment statistics (p-value, observed interactions, and percentiles of randomized distributions), module preservation in 4 independent datasets (including the $Z_{summary}$ and the associated Bonferonni-corrected p-values), pairwise correlation between MEs, and the ME value across all samples for each module. For details, see Extended Experimental Methods.

C) List of BrainSpan neocortical samples used to construct developmental networks.

<TableS2.xlsx>

Table S2, related to Figure 2, 3, and 4. Gene sets and enrichment analysis for curated lists and RDNV lists.

A) Gene lists used for enrichment analysis in this study.

B) Network-wide enrichment for candidate gene sets, RDNV-associated gene sets, and FMRP interactors. Contains number of genes in each module overlapping with each set. The background set was all 15,585 cortex-expressed protein coding genes, except in the case of asdM12/asdM16 where we restricted the background to genes Illumina probes used in Voineagu et al., 2012, and in the case of FMRP interactors where we used the background set of all protein coding one-one human-mouse orthologs. Enrichments with OR > 1, FDR < 0.05 are bolded and italicized while enrichments with OR > 1, p < 0.05 are bolded. FDR values are reported as corrected across enrichments performed for the candidate ASD and ID gene sets, the RDNV discovery set, and the FMRP target set.

C) Compilation of RDNV data from four studies – Iossofiv et al. 2012, Sanders et al. 2012, O'Roak et al. 2012, and Neale at al. 2012.

D) Assessment of the contribution of RDNVs from four studies to the enrichments in the network analysis.

<TableS3.xlsx>

Table S3, related to Figure 4. TF binding site analysis results for enrichments connecting two or more modules. A) Summary of enrichment by module. A "Y" is marked for enrichment if the motif for the TF from TRANSFAC is enriched (p < 0.05) above the three background sets. If the TF is associated with neuronal function or neuronal development, a reference is provided. If ChIP data exists for the TF, we report the fraction of overlapping sites. B) Individual TF-motif enrichments driving module-level enrichment above background.

<TableS4.xlsx>

Table S4, related to Discussion. Example prioritization of M2 and M3 RDNV affected genes using information from Table S1, as described in the discussion.

Extended Experimental Procedures

Developmental expression data

BrainSpan developmental RNA-seq data (publicly available via www.brainspan.org) summarized to GENCODE10 (Harrow et al., 2006) gene-level reads per kilobase million mapped reads (RPKM) values were used (Table S7 for sample details, BrainSpan website for data collection methods). We used the RNA-seq level data instead of microarray data as it better reflects the dynamic range of transcripts across development, particularly low-expressed transcription factors. Furthermore, networks based on RNA-seq have the advantage of including relevant intermediary non-coding transcripts in the co-expression relationships and will allow for future studies to investigate non-coding regions for putative function by mapping relevant mutations to these networks.

Count-level RNA-seq data from 52,376 transcripts across 528 samples was normalized for GC content (Hansen et al., 2012) followed by batch effect (Johnson et al., 2006) and outlier removal. Only the neocortical regions were used in our analysis – dorsolateral prefrontal cortex (DFC), ventrolateral prefrontal cortex (VFC), medial prefrontal cortex (MFC), orbitofrontal cortex (OFC), primary motor cortex (M1C), primary somatosensory cortex (S1C), primary association cortex (A1C), inferior parietal cortex (IPC), superior temporal cortex (STC), inferior temporal cortex (ITC), and primary visual cortex (V1C).

Genes were defined as expressed if they were present at an RPKM of 1 in 80% of the samples from at least one neocortical region at one major temporal epoch (based on the BrainSpan periods), resulting in 22,084 coding and non-coding transcripts. Of these, 15,591 (representing 15,585 unique gene symbols) were protein coding as annotated by GENCODE10 (which corresponds to ENSEMBL65), a similar number to those observed in a microarray analysis of a subset of these brain samples (Kang et al., 2011).

The samples were split into development (PCW 8 to 3 years of age) and later maturation (after 3 years of age to 40 years of age). RNA integrity number (RIN, a surrogate marker for RNA quality) was highly correlated ($r = -0.33$, $p = 1.3 \times 10^{-6}$) to age during early development, so we filtered for RIN >=9, this left 146 samples for the developmental time points and reduced the RIN effect ($r = -0.10$, $p = 0.24$). This resulted in the 146 high-quality samples ranging from PCW 8 to 1 year of age that were used to construct the developmental network. Time points prior to PCW 10 and between 1 year of age and 8 years of age

were not used as the anatomy of earlier regions is less well defined for the former, and samples did not pass our RIN threshold for the latter. Finally, expression values were log-transformed (log_2[RPKM+1]). The processed data used for network analysis are available with the supplemental code.

Weighted Gene Co-expression Network Analysis

All analyses were carried out in R (version 2.15.1) on a 64-bit Linux system equipped with a 2xIntel Xeon X5690 with Westmere 3.47Ghz processors and 96GB RAM. All network plots were constructed using the igraph package in R (Csárdi and Nepusz, 2006).

Briefly, correlations were estimated in a robust manner using the biweight midcorrelation (Langfelder and Horvath, 2012). Next a signed weighted correlation network was used to identify co-expression modules comprised of positively correlated genes with high topological overlap (Zhang and Horvath, 2005). Modules were defined as branches of a hierarchical cluster tree using the hybrid dynamic tree cut method (Langfelder et al., 2008). For each module, the expression patterns were summarized by the module eigengene (ME), defined as the right singular vector of the standardized expression patterns. Pairs of modules with high module eigengene correlations ($r > 0.85$) were merged. This maintains a level of decorrelation among MEs (Table S1B). MEs for modules are plotted in Figures 1 and S1, with trajectories visualized using the scatter.smooth function in R with a second order polynomial fit to the points (otherwise default parameters were used) after grouping by age as shown on the axes.

In more detail, the biweight midcorrelation, which is more robust to outliers compared to Pearson correlation, was implemented as defined in the default settings of the bicor function in the WGCNA package. A weighted signed network was computed based on a fit to scale-free topology, and a thresholding power of 26 was chosen (as it was the smallest threshold that resulted in a scale-free R^2 fit of 0.8), and the pair-wise topological overlap (TO) between genes was calculated (Zhang and Horvath, 2005). These transformations effectively monotonically transform pair-wise correlation values from [-1,1] to TO co-expression values from [0,1], where values close to 1 represent highly shared neighborhoods of co-expression. The TO captures relationships among neighborhoods of genes, and is therefore more robust than pairwise correlation alone for clustering genes by similarity. In fact, the TO approach has

been shown to be as effective as mutual information in defining modules for large-scale gene networks (Allen et al., 2012).

This TO dendrogram was used to define modules using the cutreeHybrid function in WGCNA (Langfelder et al., 2008), with a minimum module size set to 200 genes and the deepSplit parameter set to 2. The connectivity of every gene in every module (assessed by correlation to the ME, kME) is available in Table S1A. We tested additional parameters and found that the modules we focus on were robustly identified under variations of these parameters.

Further network characterization – permutation, resampling, and preservation analyses

Further characterization of the co-expression network and modules was carried out by asking 1) which modules represented co-expression above chance, 2) whether modules were robustly defined in the current set of samples, 3) whether modules reproduced in independent data. We first compared the summed correlation of genes in each module with 10,000 randomly drawn gene sets representing modules of the same size, and found that every module exhibited co-expression above chance (all modules, $p < 1x10^{-4}$).

Next, we asked whether module structure was highly sensitive to removing samples involved in the initial calculation of the network. We reconstructed networks 100 times with the same parameters but by randomly resampling from the initial sample set. Modules were found to be reproducible with perturbations to the initial individual subject, regional, and temporal structure, and the fraction of times each gene was assigned to the same module is reported in Table S1A. To validate co-expression in independent data, we asked whether our modules represented co-expressed sets of genes that could be found in other datasets. Module preservation analysis was used to calculate the $Z_{summary}$ statistic for each module. This measure combines module separability, module density, and intramodular connectivity metrics to give a composite statistic where $Z > 2$ suggests moderate preservation and $Z > 10$ suggests high preservation (Langfelder et al., 2011). The preservation analysis was performed in three epochs from an independent dataset of prefrontal cortex microarray spanning development (Colantuoni et al., 2011). M13, M16, and M17 are well-preserved at all time points, while M2 and M3 are highly preserved only at the earliest time point (Figure S1B). The preferential preservation of M2 and M3 during early

development was also observed in the BrainSpan adult data (Table S1B). This analysis demonstrated that our data was highly reproducible. Modules that were enriched for ASD risk genes are set off in bolded italics, and M2, M3, M13, M16, and M17 are all –log10(p-values) < -40 in the independent data from early development, indicating that these modules are highly reproducible in human fetal cortex.

We also assessed preservation in normal human neural progenitor (NHNP) development (Konopka et al., 2012). Plots of MEs and the average normalized expression in modules enriched for ASD risk genes show that NHNPs show a similar temporal trend as early *in vivo* development (PCW 8-20). Comparing Figure 1C to Figure S1C suggests that at least part of the transcriptional trajectory captured by these modules reflects the differentiation of neurons, and conversely, that the differentiation of neurons can model these transcriptional trajectories *in vitro*.

Gene Ontology Analysis

Genes in network modules were characterized using GO Elite (version 1.2.5 updated on 7/7/12) to control the network-wide false discovery rate using the cortex-expressed genes as background (Zambon et al., 2012). GO Elite uses a Z-score approximation of the hypergeometric distribution to assess term enrichment, and removes redundant GO terms to give a concise output. We used 10,000 permutations and required at least 10 genes to be enriched in a given pathway at a Z-score of at least 2. Gene Ontology enrichment results fulfilling these criteria are reported for all modules in Figure 1C, Figure S1D, and Table S1B.

Protein-Protein Interaction Analyses

Protein-protein interactions were compiled from two resources, InWeb (Rossin et al., 2011) and BioGRID (Stark, 2006). Data were downloaded for InWeb via the DAPPLE web resource at http://web.mit.edu/~erossin/Public/ on 3/13/2013. BIOGRID 3.2.98 data were downloaded on 3/21/2013 and restricted to physical interactions observed in *Homo sapiens*. We used only non-redundant interactions, and defined all interactions as undirected edges in a binary network. A union of the two networks was taken, and a degree-matched permutation analysis was applied in order to control for biological and methodological biases in PPI data. For every module, the subset of compiled PPIs

between genes in that module was extracted and all edges were counted. The entire PPI dataset was split into percentiles based on the degree of connectivity of every gene to other genes, and equally sized null modules matching the degree percentiles in the observed module were generated, and their interactions were counted over 10,000 iterations. A p-value was calculated based on the rank of the observed module count among the null module counts (Table S1B).

We also assessed whether the enriched subsets of RDNVs are interconnected by PPIs above chance using DAPPLE (Rossin et al., 2011), which uses a within-degree within-node permutation method that allows us to rank PPI hubs by p-value. RDNV-affected genes in both M2 and M3 show increased PPI connectivity (Figure S2). Thus, independent data supports the coherent nature of these co-expressed RDNV affected genes, as they are highly connected at the protein-protein interaction level.

Criteria for Shared Function by Multiple Systems Biology Resources

For downstream characterization of modules, we kept modules that fulfilled two of the following three criteria: 1) significant preservation in independent developmental expression data after Bonferonni correction; 2) enrichment for protein-protein interaction after Bonferonni correction; 3) enrichment for GO terms at an FDR < 0.01. From the initial set of 17 modules, 12 passed these criteria of reproducibility and independent functional validation. These 12 modules were used for downstream analyses.

Gene set over-representation

All enrichments of gene sets were performed using a two-sided Fisher exact test with 95% confidence calculated according to the R function fisher.test. We preferred p-values from this two-sided approach to the one-sided test (which is equivalent to the hypergeometric p-value) as we do not *a priori* assume there will be enrichment (Rivals et al., 2007). To reduce the likelihood of false positives, we focus on FDR adjusted p-values (Benjamini and Hochberg, 1995). We computed the false discovery rate for all gene set enrichments relevant to our primary analysis (candidate genes, RDNV discovery set, and FMRP interactors) based on 204 tests (Table S3), and focused on enrichments with OR > 1 passing FDR < 0.05. For the RDNV replication set, given the smaller sample size and differing methodology of the replication study, we required an OR > 1, p < 0.05 or validation. The stricter FDR threshold may result in false

negatives when the claim is made that a gene set is not enriched in a given module, e.g. when we claim ID genes are not enriched in M2, M3, M13, M16, or M17. Therefore, to reduce the risk of false negatives for such claims we require $p > 0.05$ for non-enrichment. We make note of the enrichment trends that do not reach significance where applicable, as these processes or pathways may be significant once additional data is available. Thus, in favoring stronger enrichment when claiming enrichment and in favoring weaker enrichment when claiming lack of enrichment, we ensure our claims are more accurate.

It is critical that the background set in an over-representation analysis reflect the claim being made. We use a cortex-expressed protein coding gene set for enrichment analyses unless otherwise specified in Table S2B. Protein coding is defined by the biotype annotation "protein_coding" in GENCODE. Ensembl Gene IDs in GENCODEV10 were overlapped according to the HUGO symbol, and all conversions among identifiers were performed using the R package biomaRt. This set of 20,007 genes was intersected with the 22,084 cortex-expressed transcripts, resulting in a "cortex-expressed background" set of 15,585 unique gene symbols. Gene membership in each set is delineated in Table S1A. In the case of asdM12 and asdM16, we used the set of 8,108 protein coding genes that had probes called as expressed in Voineagu et al., 2012. In the case of FMRP interactors, we use one-to-one human-mouse protein coding orthologs as the background set.

ASD and ID implicated gene sets

The ASD SFARI list was compiled using the online SFARI gene database, AutDB (https://gene.sfari.org/autdb/, accessed 8/20/2012). The database contains ASD candidates based on varying levels of evidence from the published literature (Basu et al., 2009). We restricted our list to genes with strong genetic evidence by filtering by the category S (syndromic) and evidence levels 1-4 (high confidence - minimal evidence). The minimal evidence category encompasses any gene in an ASD-associated multigenic CNV, genes near GWAS variants, convincing but not replicated association study results, and genes with multiply identified mutations that were not identified in a genome-wide statistical context. Importantly, this prioritization excludes genes with equivocal evidence (one of multiple genes found under a linkage peak, for example) and genes that functionally interact with higher-confidence

genes (PPI, co-expression, or other network-based categorizations). This resulted in 155 total candidate ASD risk genes that have observed genetic evidence.

We obtained asdM12 (444 genes) and adsM16 (386 genes) from an independent gene expression study that identified reproducible gene expression changes in ASD post-mortem cortex and applied WGCNA to identify modules of dysregulated genes ASD (Voineagu et al., 2011). An important rationale behind our using asdM12 is that it is possible that the SFARI ASD gene set is biased by the likely over-representation of studies investigating candidate synaptic genes in ASD, but asdM12 is agnostic to such biases.

We also considered a set of 72 CNV-affected genes that were highly interconnected by a published functional network analysis called NETBAG, which we refer to as NETBAG CNV genes (Gilman et al., 2011). Although this gene set exhibited elevated ORs (>1.5) in M2, M3, M16, and M17, we did not observe significant enrichment for the NETBAG genes in any module due to the small size of the gene set. Since the NETBAG network constructed without considering brain tissue specificity or molecular relationships from brain development among genes (shared phenotype, shared PPI, shared annotated terms from many resources), we asked if it was coalescing genes from different neurobiological pathways into one network. Post-hoc analysis testing enrichment of NETBAG genes in our ASD modules (M2, M3, M13, M16, or M17 as one set) confirmed enrichment for NETBAG genes in this larger non-specific set ($p = 5.3x10^{-3}$, OR = 2.1 [1.2-3.5]), suggesting that incorporating brain gene expression added specificity above previous methods in our analysis.

The ID set was compiled based on four reviews (Inlow and Restifo, 2004; Lubs et al., 2012; Ropers, 2008; van Bokhoven, 2011) of genes that have been associated with ID. After removing redundant gene symbols, this resulted in 401 genes. Notably, GO enrichment of the ID set using DAVID implicates the terms disease mutation, mental retardation, and epilepsy as the top three enriched terms, but many other terms associated with syndromic disorders and brain disorders are also enriched, suggesting this gene set agrees well with systematic annotation.

We also analyzed enrichment for overlapping ASD and ID genes (ASD/ID overlap, 38 genes) and ASD genes with ID-implicated genes removed (ASD only, 117 genes) and ID genes with ASD implicated genes removed (ID only, 363 genes). These were also used in Figure 5 and Figure S4 to ensure the

overlapping genes did not confound the layer enrichment analyses. The intersection of candidate gene sets with genes expressed in cortex is available in Table S1A, while the full lists are in Table S2A.

We obtained RDNVs from four publications (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012). Sanders et al., O'Roak et al., and Neale et al. were used as a discovery set as they shared similar criteria for calling variants, while Iossifov et al. was used as a replication set as more stringent filters were used to avoid false positives which may have increased the false negative rate. We note that the two studies predominantly sequenced trios (Neale et al., 2012; O'Roak et al., 2012), thus due to this and the reduced hit rate of protein-disrupting and missense RDNVs in siblings, we have fewer total genes available from unaffected siblings which decreases our power to assess gene set enrichment in siblings (Table S2B). We compiled all rare *de novo* mutation affected genes reported in these studies and categorized them as protein disrupting (there is a protein-coding change that induces a nonsense, splice-site, or frameshift mutation), missense (an amino acid change), or silent (no amino acid change). Contributions of each variant type from each study are delineated in Tables S2C-D. In total, these studies identify 125 protein disrupting, 559 missense, and 236 silent RDNV hit genes in 965 affected individuals and 36 protein disrupting, 307 missense, and 126 silent RDNV hit genes in 565 unaffected siblings, though the variant counts are higher as some genes are affected by recurrent RDNVs. We mapped the position of mutation for these variants to ENSEMBL gene models using biomaRt to ensure that all symbols complied with GENCODEV10.

Robustness of RDNV enrichments

One concern when taking variants across multiple studies is the difference in exome capture, DNA sequencing, and bioinformatic analyses that could lead to results being driven by only one study (Gratten et al., 2013; Leek et al., 2010). Table S2D shows that all studies contribute to the observed results. The ratio of protein disrupting and missense variants to silent variants is similar across studies (2.4-4.7, combined across studies 3.0), while the ratio of protein disrupting, missense, and silent variants in probands compared to siblings is similar in the two studies with matching probands-siblings pairs. Finally, M2 and M3 were enriched for protein disrupting and missense variation in M2 and M3 in the vast

majority of resampled networks (Figure S2A), demonstrating that enrichment for RDNVs in M2 and M3 is extremely robust.

Comparative Enrichment of Common Variants from Genome-Wide Association Studies

In order to test whether common variants differentially affect ASD implicated modules M2, M3, M13, M16, and M17, we compared the distribution of p-values from two genome-wide association (GWA) studies, one from the Autism Genome Project (AGP) and another from the Autism Genome Resource Exchange/Children's Hospital Philadelphia (AGRE/CHOP). Both have been published previously, though no single finding was replicated between the two at a genome-wide significant level (Anney et al., 2012; Wang et al., 2009). Of note, the published AGRE cohort overlaps with the AGP cohort. We used a modified set of AGRE subjects that is independent of the AGP cohort for our analysis, and obtained GWA p-values by re-running the association using permutation-based tests in PLINK (Purcell et al., 2007). It had previously been shown that common variants from the AGRE/CHOP GWA were enriched in asdM12 (Voineagu et al., 2011) using a modified Kolmogorov-Smirnov (K-S) test (Wang et al., 2010). For each gene model and the 30kB upstream from that gene, the SNP with a minimum p-value from the GWA is taken to tag the gene, and an enrichment statistic is calculated. Genes were tagged by SNPs using the ENSEMBL gene model and dbSNP137 SNPs and their locations in on hg19. The distribution of SNP p-values near genes is then calculated for a given pathway (or in this case, a module).

We utilized a similar approach, but applied a permutation-based procedure to assess pair-wise differential enrichment for low p-value SNPs from GWA in order to control or differences in gene size, since longer genes are more likely to contain a lower p-value SNP by our definition of tagging. First, we compared the distribution of the SNPs in a pair of modules, and calculated the Kolmogorov-Smirnov test statistic. Next, we drew 10,000 pairs of distributions sharing the same number of tagging SNPs as the initial pair (controlling for gene size and haplotype structure), and re-computed the K-S statistic each time. Finally, we calculated a p-value based on the rank of the observed K-S statistic in this distribution. We also conducted this test with p-values from a GWAS in psoriasis (Nair et al., 2009), and found that none of the pairwise comparisons passed FDR < 0.05.

Transcription Factor Binding Site (TFBS) Enrichment

TFBS enrichment analysis was performed by scanning the canonical promoter region (1000bp upstream of the transcription start site) for the genes in each co-expression module. For each TF, we assessed the top 200 connected genes (ranked by kME) in each module using the following steps: 1) putative motifs bound by the TF were obtained from TRANSFAC (Matys et al., 2003). 2) upstream sequences of these 200 genes were scanned with the Clover algorithm (Frith et al., 2004) to calculate motif enrichment; and iii) enrichment above background was calculated using the MEME algorithm (Bailey and Elkan, 1994). We compared enrichment for each TF motif in three different background datasets to ensure robustness: 1000 bp sequences upstream of all human genes, human CpG islands, and the sequence of human chromosome 20. We calculated p-values by drawing 1000 sequences of the same length and testing them for enrichment using MEME, and computed the p-value based on the observed motif enrichment rank versus the randomized sets. We report TFs with $p < 0.05$ across all 3 backgrounds (Table S6).

We also asked whether existing ChIP data for TFs from either ENCODE (Consortium, 2011) or other compiled genome-wide ChIP data (Lachmann et al., 2010) (ENCODE ChIP data website: http://genome.ucsc.edu/ENCODE/dataMatrix/encodeChipMatrixHuman.html, accessed 2/6/2013; ChEA data website: http://amp.pharm.mssm.edu/lib/chea.jsp, accessed 5/8/2013) supported our predicted binding sites. For TFs where a dataset was readily available, we report the fraction of sites that overlap in the existing data. Most of these TF binding sites come from non-neuronal cell lines, and many come from proliferating cells (most ENCODE lines are cancer cell lines). We therefore cautiously interpret this experimental support for our binding sites, but find it encouraging that a moderate to large fraction of predicted binding sites have been observed in experiments, suggesting that it is at least possible for the TF to bind near the predicted target gene.

Layer-specific and Cell-type Marker Enrichment

To quantify layer-specific gene expression during development, we utilized micro-dissected human fetal neocortical laminar gene expression datasets from BrainSpan, two for each of the earlier and later fetal periods. The 15 PCW and 16 PCW data together comprises 351 samples in total, including 6

regions and 8 layers, while the two 21 PCW brains' data comprises 337 samples. Entrez gene IDs corresponding to array probes were mapped to GENCODEV10 gene symbols using biomaRt. Since multiple probes can cover each gene, we picked the probe with maximum mean expression level to represent each gene. For adult layers, we used primate neocortical laminar gene expression data from macaque data comprising 10 cortical regions and 5 layers within each region (Bernard et al., 2012). We used adult cortical dissections for our cell-type analyses, as we found that laminar differential expression exhibits gradients in the fetal data. In contrast, differential expression of genes in adult primate cortical data at $t > 2$ reflected well ISH patterns of laminar specificity seen in human (compare specific genes where overlaps occur with Zeng et al., 2012).

For laminar enrichment, the limmar package in R was used to calculate the t-statistics of differential expression for all genes in each layer against all the other layers. Then, for each gene set, the difference in the distribution of t-values in each layer for that set versus background was computed using a Z statistic. This quantifies the skew of differential expression t-values of each gene set in each layer. If there is no layer specificity, the distribution of t-values from a gene set is expected to follow the same distribution as the background set (with $Z = 0$), while a significant skew toward differential expression in a layer results in a positive Z score. We calculated an FDR cut-off across all enrichments in all layers ($Z = 2.7$, FDR = 0.01) and computed bootstrapped confidence intervals for each enrichment. To quantify cell-marker relationships, we used the same method, with the t-value replaced by the correlation of each gene to the first principal component of a set of known cell marker genes in the adult layer data (Table S1A lists cell-type marker genes and r values). We reported both strong, FDR-corrected enrichment, as well as nominal enrichment to emphasize trends. Statistical comparison of enrichment trends across layers between ASD and ID gene sets set was performed by 1) computing the difference in the Z score between the two sets for each layer, 2) summing this difference across all layers, and 3) comparing this to the distribution of summed differences in layers of 10,000 permuted pairs of sets matched for gene set size (see Extended Experimental Methods for details).

Network analysis resources and R code parameters for network analysis

Table S10 shows an example of prioritizing RDNV affected genes from a module using information from our analyses (see Discussion for details). We also provide an interactive network interface for further exploration ***<link pending for finalized version>***.

Future work using the same methods here can increase the temporal tiling (which could result in more specific modules) and expand the pool of mutations implicated in ASD (this would improve the signal-to-noise in the enrichments). We have provided the code used in this analysis that will allow future work to easily update and incorporate this level of analysis. We provide a template for other to build upon by providing the R code and processed expression data at:
http://labs.genetics.ucla.edu/horvath/htdocs/CoexpressionNetwork/developingcortex
to allow reconstruction of networks with additional data and mapping of new genes as they are discovered.

Supplemental References

Allen, J.D., Xie, Y., Chen, M., Girard, L., and Xiao, G. (2012). Comparing Statistical Methods for Constructing Large Scale Gene Networks. PLoS ONE *7*, e29348.

Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bölte, S., Bolton, P.F., Bourgeron, T., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Human Molecular Genetics *21*, 4781–4792.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28–36.

Basu, S.N., Kollu, R., and Banerjee-Basu, S. (2009). AutDB: a gene reference resource for autism research. Nucleic Acids Res *37*, D832–D836.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 289–300.

Bernard, A., Lubbers, L.S., Tanis, K.Q., Luo, R., Podtelezhnikov, A.A., Finney, E.M., McWhorter, M.M.E., Serikawa, K., Lemon, T.A., Morgan, R.J., et al. (2012). Transcriptional Architecture of the Primate Neocortex. Neuron *73*, 1083–1099.

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., Colantuoni, E.A., Elkahloun, A.G., Herman, M.M., Weinberger, D.R., et al. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature *478*, 519–523.

Consortium, E.P. (2011). A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biol *9*, e1001046.

Csárdi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, Complex Systems *1695*.

Frith, M.C., Fu, Y., Yu, L., Chen, J.-F., Hansen, U., and Weng, Z. (2004). Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res *32*, 1372–1381.

Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D. (2011). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron *70*, 898–907.

Gratten, J., Visscher, P.M., Mowry, B.J., and Wray, N.R. (2013). Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. Nat Genet *45*, 234–238.

Hansen, K.D., Irizarry, R.A., and WU, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics *13*, 204–216.

Harrow, J., Denoeud, F., Frankish, A., Reymond, A., Chen, C.-K., Chrast, J., Lagarde, J., Gilbert, J.G., Storey, R., Swarbreck, D., et al. (2006). GENCODE: producing a reference annotation for ENCODE. - Abstract - UK PubMed Central. Genome Biol *7*, S4.

Inlow, J.K., and Restifo, L.L. (2004). Molecular and comparative genetics of mental retardation. Genetics *166*, 835–881.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.-H., Narzisi, G., Leotta, A., et al. (2012). De Novo Gene Disruptions in Children on the Autistic Spectrum. Neuron *74*, 285–299.

Johnson, W.E., Li, C., and Rabinovic, A. (2006). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483–489.

Konopka, G.G., Wexler, E.E., Rosen, E.E., Mukamel, Z.Z., Osborn, G.E.G., Chen, L.L., Lu, D.D., Gao, F.F., Gao, K.K., Lowe, J.K.J., et al. (2012). Modeling the functional genomics of autism using human neurons. Mol Psychiatry *17*, 202–214.

Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438–2444.

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics *24*, 719–720.

Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of Statistical Software *46*, –.

Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is My Network Module Preserved and Reproducible? PLoS Comput Biol *7*, e1001057.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., and Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet *11*, 733–739.

Lubs, H.A., Stevenson, R.E., and Schwartz, C.E. (2012). Fragile X and X-Linked Intellectual Disability: Four Decades of Discovery. The American Journal of Human Genetics *90*, 579–590.

Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res *31*, 374–378.

Nair, R.P., Duffin, K.C., Helms, C., Ding, J., Stuart, P.E., Goldgar, D., Gudjonsson, J.E., Li, Y., Tejasvi, T., Feng, B.-J., et al. (2009). Genome-wide scan reveals association of psoriasis with IL-23 and NF-|[kappa]|B pathways. Nat Genet *41*, 199–204.

Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.-F., Stevens, C., Wang, L.-S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 1–5.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 1–7.

Purcell, S., Neale, B., Todd-Brown, K., and Thomas, L. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of ….

Rivals, I., Personnaz, L., Taing, L., and Potier, M.-C. (2007). Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics *23*, 401–407.

Ropers, H.H. (2008). Genetics of intellectual disability. Current Opinion in Genetics & Development *18*, 241–250.

Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Constortium, I.I.B.D.G., Cotsapas, C., and Daly, M.J. (2011). Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology. PLoS Genet. *7*, e1001273.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., Dilullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

Stark, C. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res *34*, D535–D539.

van Bokhoven, H. (2011). Genetic and epigenetic networks in intellectual disabilities. Annu. Rev. Genet. *45*, 81–104.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.

Wang, K., Wang, K., Dickson, S.P., Dickson, S.P., Stolle, C.A., Stolle, C.A., Krantz, I.D., Krantz, I.D., Goldstein, D.B., Goldstein, D.B., et al. (2010). Interpretation of Association Signals and Identification of Causal Variants from Genome-wide Association Studies. The American Journal of Human Genetics *86*, 730–742.

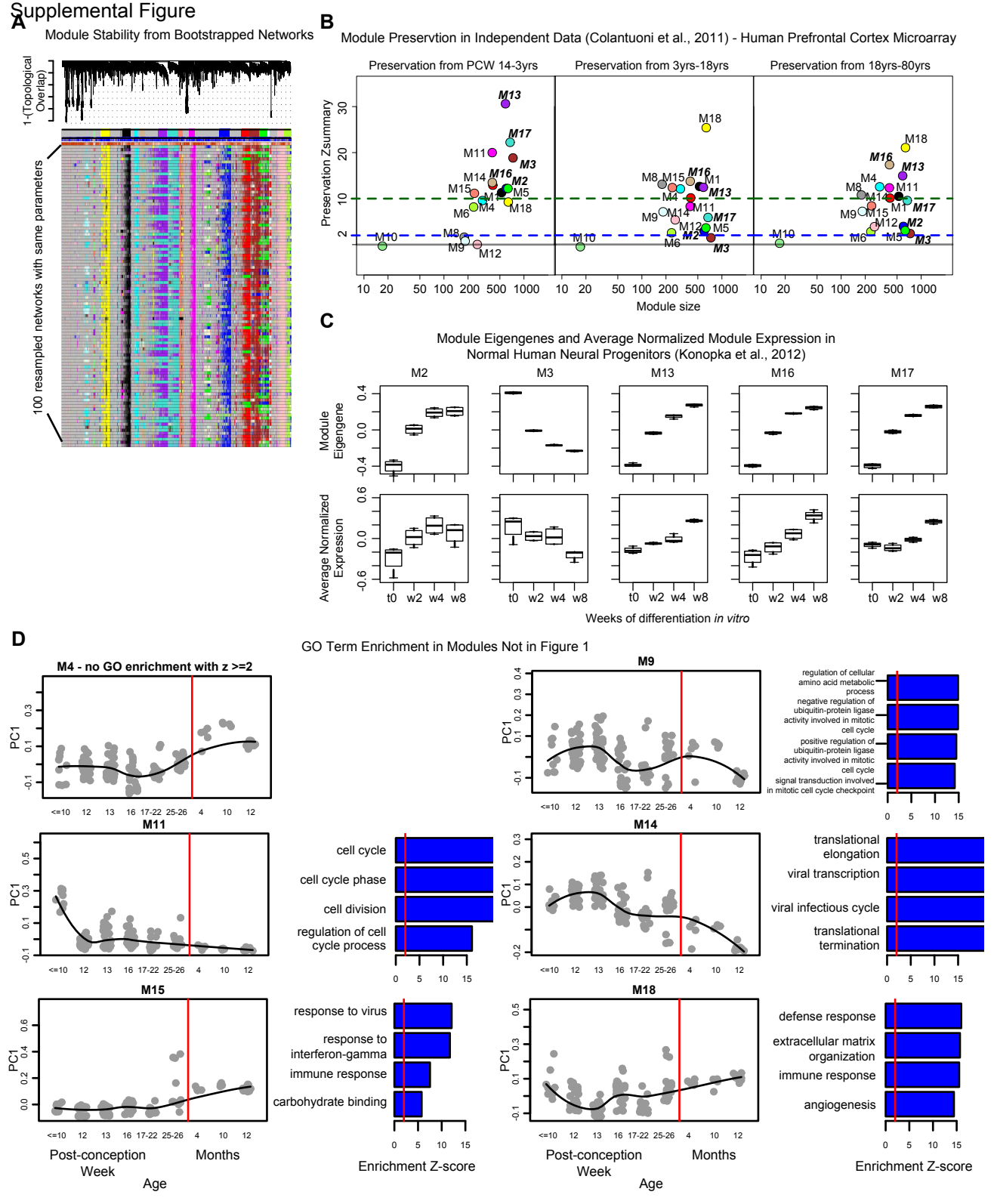
Wang, K., Zhang, H., Ma, D., Bućan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imielinski, M., Bradfield, J.P., Sleiman, P.M.A., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature *459*, 528–533.

Zambon, A.C., Gaj, S., Ho, I., Hanspers, K., Vranizan, K., Evelo, C.T., Conklin, B.R., Pico, A.R., and Salomonis, N. (2012). GO-Elite: A Flexible Solution for Pathway and Ontology Over-Representation. Bioinformatics.

Zeng, H., Shen, E.H., Hohmann, J.G., Oh, S.W., Bernard, A., Royall, J.J., Glattfelder, K.J., Sunkin, S.M., Morris, J.A., Guillozet-Bongaarts, A.L., et al. (2012). Large-Scale Cellular-Resolution Gene Profiling in Human Neocortex Reveals Species-Specific Molecular Signatures. Cell *149*, 483–496.

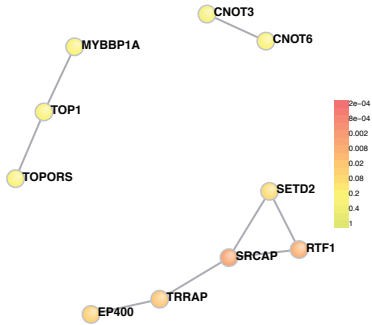
Zhang, B., and Horvath, S. (2005). A General Framework for Weighted Gene Co-Expression Network Analysis. Statistical Applications in Genetics and Molecular Biology *4*.

Supplemental Figure



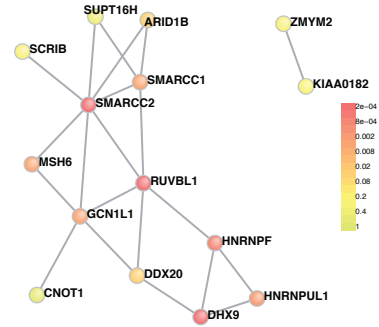
Supplemental Figure

A RDNV-affected Genes in M2 Direct PPIs From DAPPLE



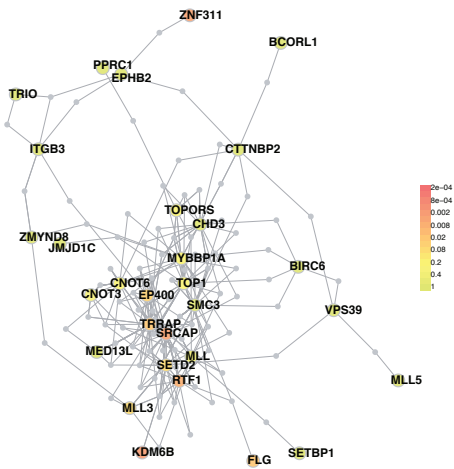
Direct Edges:	8	Direct Degree Mean:	1.6
Expected:	2.8	Expected:	1.1
p-value =	0.011	p-value =	0.04

C RDNV-affected Genes in M3 Direct PPIs From DAPPLE



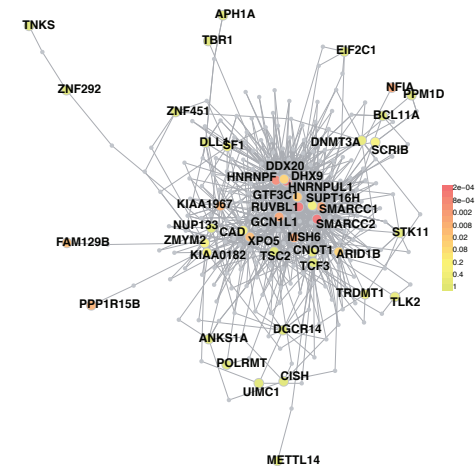
Direct Edges:	21	Direct Degree Mean:	2.8
Expected:	10.2	Expected:	1.4
p-value =	0.0009	p-value =	0.0002

B RDNV-affected Genes in M2 Indirect PPIs From DAPPLE



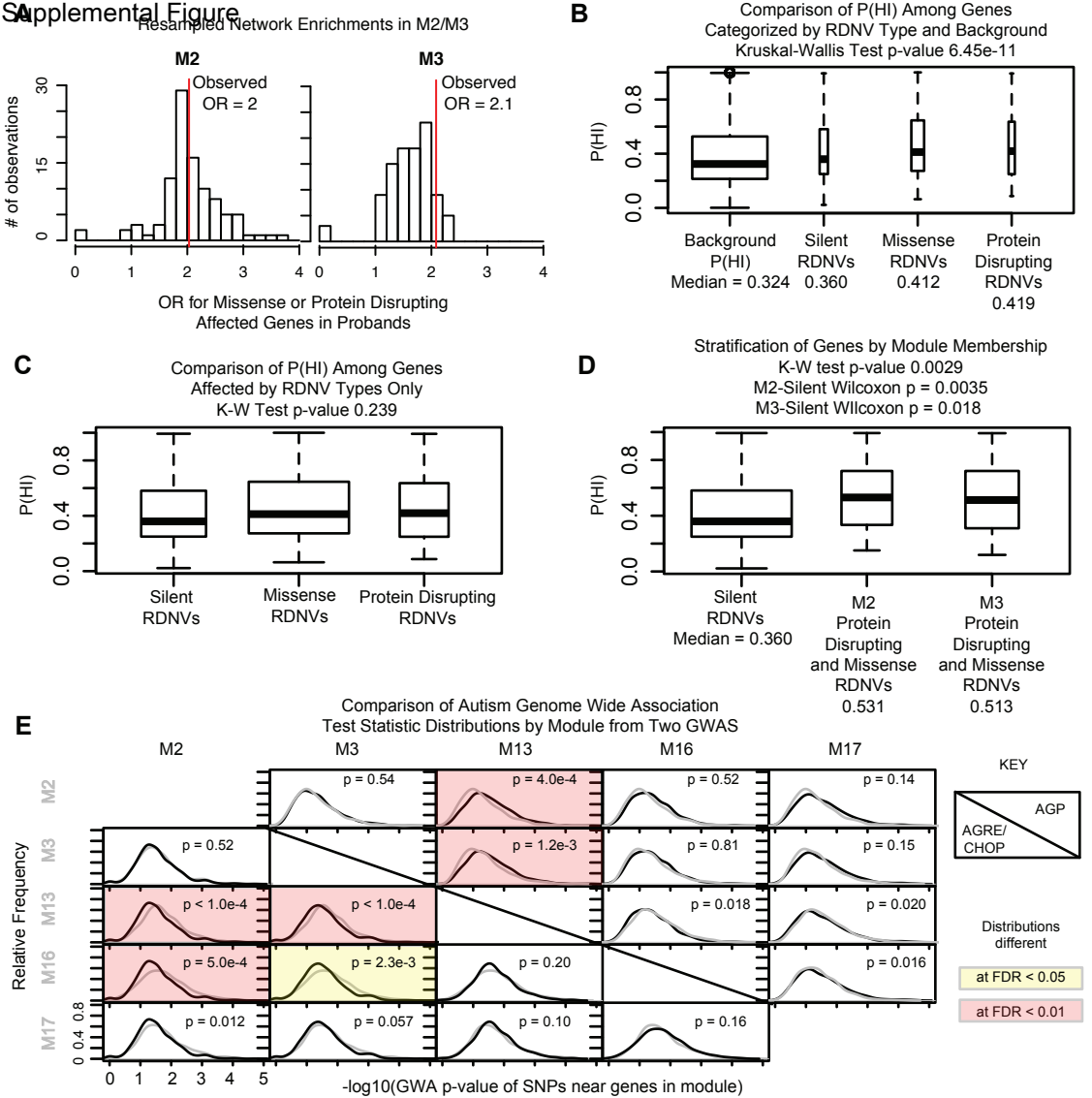
Indirect Degree Mean:	28.9
Expected:	16.4
p-value =	0.0023

D RDNV-affected Genes in M3 Indirect PPIs From DAPPLE



Indirect Degree Mean:	100.3
Expected:	62.8
p-value =	0.0048

Supplemental Figure



Supplemental Figure