# Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism

Neelroop N. Parikshak,[1,2] Rui Luo,[3,4] Alice Zhang,[2] Hyejung Won,[1] Jennifer K. Lowe,[1,4] Vijayendran Chandran,[5] Steve Horvath,[3,6] and Daniel H. Geschwind[1,2,3,4,5,*]

[1]Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2]Interdepartmental Program in Neuroscience, University of California, Los Angeles, Los Angeles, CA 90095, USA
[3]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[4]Center for Autism Treatment and Research, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[5]Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA
[6]Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90095, USA
*Correspondence: dhg@ucla.edu
http://dx.doi.org/10.1016/j.cell.2013.10.031

## SUMMARY

Genetic studies have identified dozens of autism spectrum disorder (ASD) susceptibility genes, raising two critical questions: (1) do these genetic loci converge on specific biological processes, and (2) where does the phenotypic specificity of ASD arise, given its genetic overlap with intellectual disability (ID)? To address this, we mapped ASD and ID risk genes onto coexpression networks representing developmental trajectories and transcriptional profiles representing fetal and adult cortical laminae. ASD genes tightly coalesce in modules that implicate distinct biological functions during human cortical development, including early transcriptional regulation and synaptic development. Bioinformatic analyses suggest that translational regulation by FMRP and transcriptional coregulation by common transcription factors connect these processes. At a circuit level, ASD genes are enriched in superficial cortical layers and glutamatergic projection neurons. Furthermore, we show that the patterns of ASD and ID risk genes are distinct, providing a biological framework for further investigating the pathophysiology of ASD.

## INTRODUCTION

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder in which hundreds of genes have been implicated (Berg and Geschwind, 2012; Geschwind and Levitt, 2007). Analysis of copy number variation (CNV) and exome sequencing have identified rare variants that alter dozens of protein-coding genes in ASD, none of which account for more than 1% of ASD cases (Devlin and Scherer, 2012). This and the fact that a significant fraction (40%–60%) of ASD is explained by common variation (Klei et al., 2012) point to a heterogeneous genetic architecture.

These findings raise several issues. Based on the background human mutation rate (MacArthur et al., 2012), most genes affected by only one observed rare variant to date are likely false positives that do not increase risk for ASD (Gratten et al., 2013). It is therefore essential to develop approaches that prioritize singleton variants, especially missense mutations. Furthermore, given the heterogeneity of ASD, it would be valuable to identify common pathways, cell types, or circuits disrupted within ASD itself. Recent studies combining gene expression, protein-protein interactions (PPIs), and other systematic gene annotation resources suggest some molecular convergence in subsets of ASD risk genes (Ben-David and Shifman, 2013; Gilman et al., 2011; Sakai et al., 2011; Voineagu et al., 2011). Yet, it remains unclear how the large number of genes implicated through different methods may converge to affect human brain development, which is critical to a mechanistic understanding of ASD (Berg and Geschwind, 2012). Additionally, ASD has considerable overlap with ID at the genetic level, so identifying molecular pathways and circuits that confer the phenotypic specificity of ASD would be of considerable utility (Geschwind, 2011; Matson and Shoemaker, 2009).

Here, we took a stepwise approach to determine whether genes implicated in ASD affect convergent pathways during in vivo human neural development and whether they are enriched in specific cells or circuits (Figure 1A). First, we constructed transcriptional networks representing genome-wide functional relationships during fetal and early postnatal brain development and mapped genes from multiple ASD and ID resources to these networks. We then assessed shared neurobiological function among these genes, including coregulatory relationships and enrichment in layer-specific patterns from
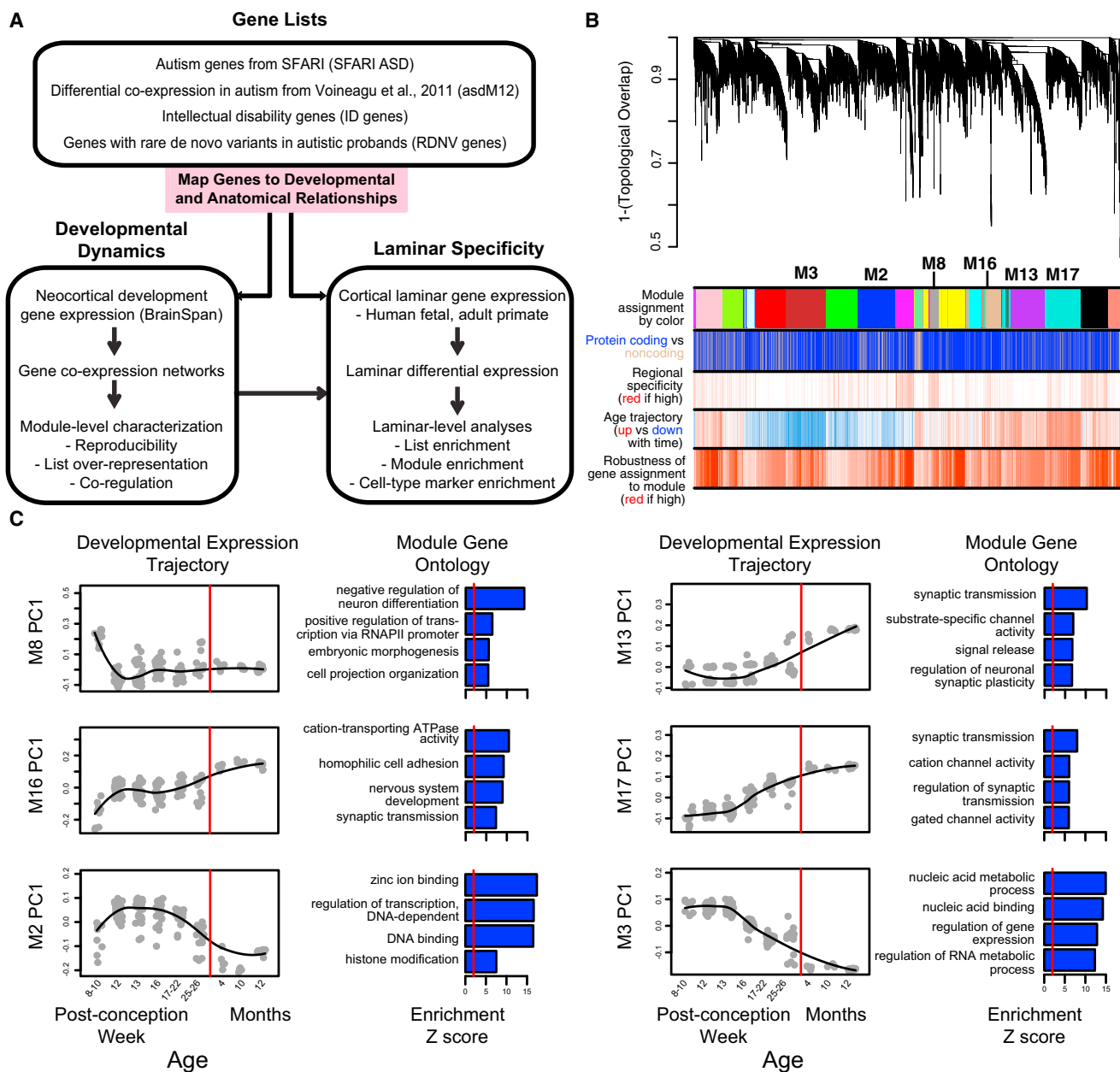
**Figure 1. Methodological Overview and Coexpression Network Analysis**

(A) Flowchart of the overall approach.

(B) Network analysis dendrogram showing modules based on the coexpression topological overlap of genes throughout development. Color bars below give information on module membership, gene biotype, cortical region specificity, age trajectory, and robustness of module assignment.

(C) Module characterization, including GO enrichment and trajectory throughout development. The fit line represents locally weighted scatterplot smoothing (Extended Experimental Procedures). GO enrichments are adjusted for multiple comparisons (FDR < 0.05), and reported Z scores represent relative enrichment in the module compared to all cortex-expressed genes, with the red line at Z = 2.

See also Table S1 and Figure S1.

microdissected human fetal and adult primate cortical laminae. We used validation in independent in vivo and in vitro expression data and additional functional evidence (shared annotated pathways and PPIs) to confirm shared coexpression and function among genes, and we replicated the enrichment analyses in independent data to ensure robustness. Our integration of an un-

supervised network analysis with large gene sets from multiple resources permits rigorous interrogation of biological convergence in ASD that takes its heterogeneity into consideration and enables comparison of ASD with ID. In addition, we have made these data accessible for biologists by creating an interactive network web browser (Experimental Procedures).

## RESULTS

### Genome-wide Coexpression Networks Reflect Biological Processes Essential to Human Neocortical Development

We reasoned that transcriptomic data from human neocortex would inform our understanding of ASD pathophysiology, as the cerebral cortex has been consistently implicated in ASD pathophysiology by multiple modalities (Amaral et al., 2008; Ecker et al., 2012; Geschwind, 2011; Rubenstein, 2011; Voineagu et al., 2011). We focused on gene expression from cortical development spanning postconception week (PCW) 8 to month 12 after birth, as this time period reflects many critical molecular processes that orchestrate brain circuit formation that could be disrupted by genetic hits in ASD (Andersen, 2003; Courchesne et al., 2011).

We constructed networks of gene relationships agnostic to ASD candidate genes based on BrainSpan whole-genome transcriptomic data collected by RNA-seq (BrainSpan, 2013). We applied signed weighted gene coexpression network analysis (WGCNA; Experimental Procedures; Zhang and Horvath, 2005) and identified 17 coexpression modules (labeled numerically, e.g., M8, and by color, e.g., magenta, see Table S1B [available online] for module details). These modules represent genes that share highly similar expression patterns during cortical development (Figure 1B). Several additional analyses show that these modules identify highly significant shared expression patterns that are replicated in independent data from both in vivo and in vitro human neural development (Figures S1A–S1C and Extended Experimental Procedures).

First, we investigated each module's developmental trajectory by calculating the module eigengene (ME, the first principal component of the module) and assessed shared function among genes within the module by enrichment for Gene Ontology (GO) annotation terms. Representative examples for up- and downregulated modules are shown in Figure 1C. MEs for M13, M16, and M17 increase during early cortical development and are each enriched for the GO term synaptic transmission (Figure 1C). M16 is upregulated the earliest, starting at PCW 10 and its hubs (most interconnected genes based on correlation to the ME, kME) include genes coding for the structural synaptic proteins *SV2A* and *NRXN1*. M16 GO terms include cation transporter activity, homophilic cell adhesion, and nervous system development, which is consistent with early development of synaptic ultrastructure. M17 represents a later phase of synaptic maturation, as it is upregulated after PCW 13, and its hubs include *CAMK2B* and *CACNA1C*, which are important for calcium-dependent regulation of synaptic activity. M13 increases last, after PCW 16, and its hubs include the NMDA and GABA receptor subunits *GRIN2A* and *GABRA1*, whereas GO terms include substrate-specific channel activity and regulation of neuronal synaptic plasticity. These three modules have closely aligned but distinct developmental trajectories that likely reflect sequential phases of synaptic development, maturation, and function, all of which are essential to the development of the cerebral cortex.

In contrast, M2 and M3 have anticorrelated trajectories to M13, M16, and M17 ($r = -0.46$ to $-0.96$; Table S1B) and are enriched in GO terms associated with DNA binding and transcriptional regulation (Figure 1C). Expression in M3 is initially upregulated and then decreases after PCW 12, suggesting that its functions may be most important prior to M2, which is upregulated after PCW 10 and peaks later (PCW 12 to PCW 22). Given the GO enrichment and anticorrelation to the synaptic module MEs, genes in these modules may be critical to orchestrating processes such as progenitor proliferation and cell fate specification via initial repression followed by derepression of neuronal genes (Srinivasan et al., 2012). Furthermore, many of the genes found in M2 and M3 are part of well-studied chromatin remodeling complexes, most notably the BAF complex (*ARID1A* and *SMARCA4* in M2; *ARID1B*, *SMARCB1*, *SMARCC1*, *SMARCC2*, *SMARCD1*, *ARID2*, *DPF2*, *BCL11A*, *BCL11B*, and *ACTL6A* in M3), which has recently been linked to neural differentiation and neurodevelopmental disorders (Ronan et al., 2013; Yoo et al., 2009).

Because positive correlations among genes also reflect pairwise interactions between proteins (Ramani et al., 2008), enrichment for protein-protein interactions within modules provides an independent line of validation for shared function in these modules at the protein level. We combined all known PPIs from InWeb (Rossin et al., 2011) and BioGRID (Stark et al., 2006) into one network, comprising 251,881 interactions among 18,384 proteins, and observed that 12/17 of all coexpression modules, including all the modules in Figure 1C, are enriched for PPI after stringent multiple testing correction ($p < 0.003$, Table S1B). Overall, 10/17 coexpression modules are preserved in independent gene expression data sets, enriched for GO terms, and enriched for PPI. An additional 2/17 modules are enriched for two of these three criteria. These results demonstrate the utility of a systems biology approach—instead of analyzing lists of thousands of genes regulated during development, we focused on this set of 12 reproducible and biologically meaningful modules sharing distinct expression patterns and biological functions. An interactive network is available at our website for graphical exploration of individual genes in these modules, as well as their relationships with each other (Experimental Procedures).

### Genes Implicated in ASD Are Highly Coexpressed during Human Cortical Development

We next asked whether genes associated with risk for ASD converge on common biological processes. We compiled a set of 155 ASD genetic risk candidates from the Simons Foundation Autism Research Initiative (SFARI) AutDB database (Basu et al., 2009), which we refer to as SFARI ASD. The SFARI ASD list is a manually curated set of candidate genes implicated by common variant association, candidate gene studies, genes within ASD-associated CNV, and, to a lesser extent, syndromic forms of ASD (Experimental Procedures). We mapped this gene set to the protein-coding genes in the developmental coexpression network and observed that SFARI ASD genes are most overrepresented in M16 ($p = 0.0024$; odds ratio [OR] = 2.9; 95% confidence interval = [1.4–5.5]; false discovery rate [FDR] < 0.05) and less so in M13 and M17 (Figure 2A).

We also examined a set of ASD genes previously shown to be dysregulated in postmortem ASD temporal and frontal cortex
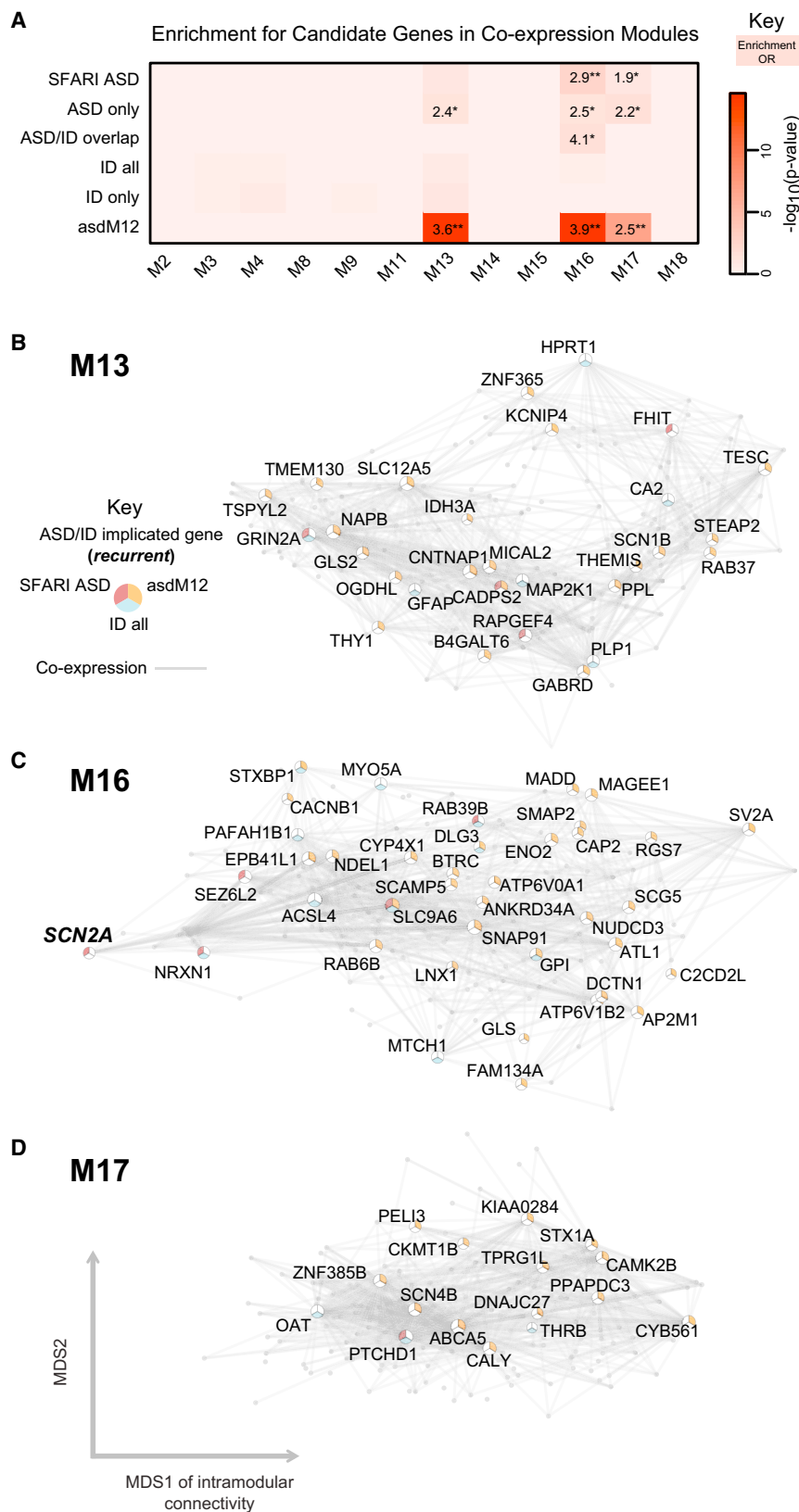
**Figure 2. Enrichment of SFARI ASD, asdM12, and ID Genes in Developmental Networks**

(A) Module-level enrichment for gene sets from a curated set of ASD risk genes (SFARI ASD), a curated set of ID genes ("ID all"), and an unbiased set of ASD risk genes (asdM12). Overlapping (ASD/ID overlap) and nonoverlapping sets ("ASD only" and "ID only") are also shown. All enrichment values for overrepresented lists with p < 0.05, OR > 1 are shown to demonstrate enrichment trends (*p < 0.05 and **FDR < 0.05). Heatmap colors for p values reflect enrichment trends; p values for gene sets with OR < 1 can be seen in Table S2B.

(B–D) These panels show network plots for M13, M16, and M17, respectively. Most hub genes overlapping with SFARI ASD and asdM12 enrichment are not the same, showing that enrichment of these two sets is not driven by a narrow shared subset of genes. Network plots comprise the top 200 connected genes (based on kME, a measure of intramodular connectivity) and their top 1,000 connections in the subnetwork. By definition, all edges in the network reflect positive correlations. Genes with membership in SFARI ASD, asdM12, or the "ID all" list are labeled and plotted according to multidimensional scaling of gene expression correlations, which graph genes with similar expression patterns closer to each other.

See also Table S2.

(asdM12; Voineagu et al., 2011), which represents a shared molecular pathology in ASD brain identified in an unbiased, genome-wide manner. The asdM12 gene set was strongly enriched in the same three modules as SFARI ASD genes, M13, M16, and M17 (asdM12-M13; $p = 3.0 \times 10^{-15}$; OR 3.6 [2.7–4.8]; asdM12-M16; $p = 3.5 \times 10^{-15}$; OR 3.9 [2.8–5.3]; asdM12-M17; $p = 1.0 \times 10^{-7}$; OR 2.5 [1.8–3.4]; each at FDR < 0.05). A remarkable 42% of asdM12 and 25% of the SFARI ASD sets are found in one of these three modules. Our analysis, which uses gene sets identified based on different methods (only 15 genes overlap between SFARI ASD and asdM12), converges onto three modules involved in prenatal and early postnatal synaptic development.

We next hypothesized that mapping ID genes to this network would enable us to assess whether ASD susceptibility genes show any specificity in their developmental expression patterns. We compiled an extensive set of high confidence genes implicated in monogenic forms of ID from multiple publications (Inlow and Restifo, 2004; Lubs et al., 2012; Ropers, 2008; van Bokhoven, 2011), referred to as "ID all" (see Experimental Procedures). Remarkably, this set of 401 genes (of which 364 are expressed in human neocortex) is not enriched in any of the 12 coexpression modules. Importantly, this lack of enrichment is at a relaxed threshold that reduces the risk of false negatives (uncorrected p > 0.05). Removing the small set of 37 genes (<10%) that overlap between ASD and ID to establish exclusive sets ("ASD only" and "ID only") further confirms that ASD genes exhibit enrichment, whereas ID genes do not (Figure 2A and Table S2B). Thus, it is genes connected with the ASD phenotype that are enriched in three specific transcriptional modules related to synaptic function during development, but not those that have been related solely to ID.

### Rare De Novo Variants Are Highly Enriched in Two Coexpression Modules in Early Fetal Development

Additional evidence implicating specific genes in ASD comes from whole-exome sequencing in families (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012b; Sanders et al., 2012), which has identified many rare protein-disrupting variants (nonsense, splice site, and frameshift) overrepresented in individuals with ASD compared to their unaffected siblings (OR > 2). This evidence is largely distinct from the evidence implicating genes in SFARI ASD and asdM12, as it is from purely noninherited, rare variation discovered in an unbiased, genome-wide manner. We therefore asked whether genes affected by protein altering rare de novo variation (RDNV) in ASD probands shared biological function. We also tested silent RDNVs because they should not exhibit a similar pattern of functional enrichment, providing a key control for gene size, GC content, and other features affecting mutability (Michaelson et al., 2012).

We first tested for enrichment using RDNVs from three studies sharing similar coverage criteria and variant calling methodology (Neale et al., 2012; O'Roak et al., 2012b; Sanders et al., 2012), representing 622 ASD probands and 222 unaffected siblings. Strikingly, genes expressed during development and affected by protein-disrupting RDNVs in probands (60 genes, Table S2A, Discovery Set) are significantly enriched in two modules,

M2 and M3, which exhibit highly similar developmental trajectories and functional enrichment indicative of remarkable biological specificity. Eight genes harboring protein-disrupting RDNVs are enriched in M2 (p = 0.006; OR = 3.2 [1.3–6.8]; FDR < 0.05), and ten are enriched in M3 (p = 0.0011; OR = 3.6 [1.6–7.2]; FDR < 0.05). A trend for enrichment is observed for M16 as well, but this does not pass the FDR threshold. For comparison, genes affected by RDNVs in unaffected siblings or affected by silent mutations are not enriched in any modules (Table S2B, p > 0.05). Because missense RDNVs are only weakly overrepresented in ASD (Sanders et al., 2012), we reasoned that overlap with network modules might prioritize specific subsets of this RDNV class. We find that a subset of missense RDNV affected genes is overrepresented in the same pathways as the more deleterious protein-disrupting RDNVs (M2 and M3; Table S2B). Taken together, out of 385 protein-disrupting or missense RDNV-affected genes expressed in brain, 34 are found in M2 ($p = 2.9 \times 10^{-4}$; OR = 2.1 [1.4–3.0]; FDR < 0.05) and 41 in M3 ($p = 2.3 \times 10^{-5}$; OR = 2.2 [1.5–3.1]; FDR < 0.05). Furthermore, the combined set of protein-disrupting and missense RDNVs from unaffected siblings was not found enriched in any modules (p > 0.05).

We further validated the observed RDNV enrichment pattern in M2 and M3 in an independent set of patients from a study with more stringent RDNV calling criteria (Iossifov et al., 2012). In this additional set of 343 ASD probands and unaffected siblings, we found that the patterns of RDNV enrichment replicated, with the combined protein-disrupting and missense RDNV sets from ASD probands enriched specifically in M2 and M3 (p < 0.05) and RDNVs from siblings and silent RDNVs not enriched in any set (Table S2B; Replication Set). Combining all four studies, we find that, out of 598 protein-disrupting or missense RDNV-affected genes expressed in cortex, 52 are in M2 ($p = 9.6 \times 10^{-6}$; OR = 2.0 [1.5–2.8]), and 61 in M3 ($p = 8.5 \times 10^{-7}$; OR = 2.1 [1.6–2.8]). Importantly, the enrichment pattern across modules is not only replicated in the independent set but is stronger in the combined set, is robust to perturbations in module composition (Figure S3A), and is not driven by variants from any one study (Tables S2C and S2D). We show the enrichment pattern of this combined set across 965 ASD probands and 565 unaffected siblings in Figure 3A and use this combined set for the remainder of our analyses. Furthermore, modules showing weak enrichment in some mutation categories of the discovery set (M11, M16 at p < 0.05, but not FDR < 0.05) did not replicate at p < 0.05 in the replication set and are not enriched when considering all four studies together (Figure 3A).

We next asked whether M2 and M3 prioritized functional subsets of genes with RDNVs. We confirmed that RDNV-affected genes in M2 and M3 are significantly enriched for interactions at a protein level (Figures S2A–S2D) and highlight genes that are both PPI hubs and coexpression hubs in Figures 3B and S3C. Furthermore, M2 and M3 genes harboring protein disrupting or missense RDNVs are also more dosage sensitive, as evidenced by the significant increase in the probability of haploinsufficiency (P[HI], Extended Experimental Procedures) among genes affected by these mutation classes (Huang et al., 2010; Luo et al., 2012). This is consistent with the heterozygous state of variants observed in ASD probands. Overall, a remarkable
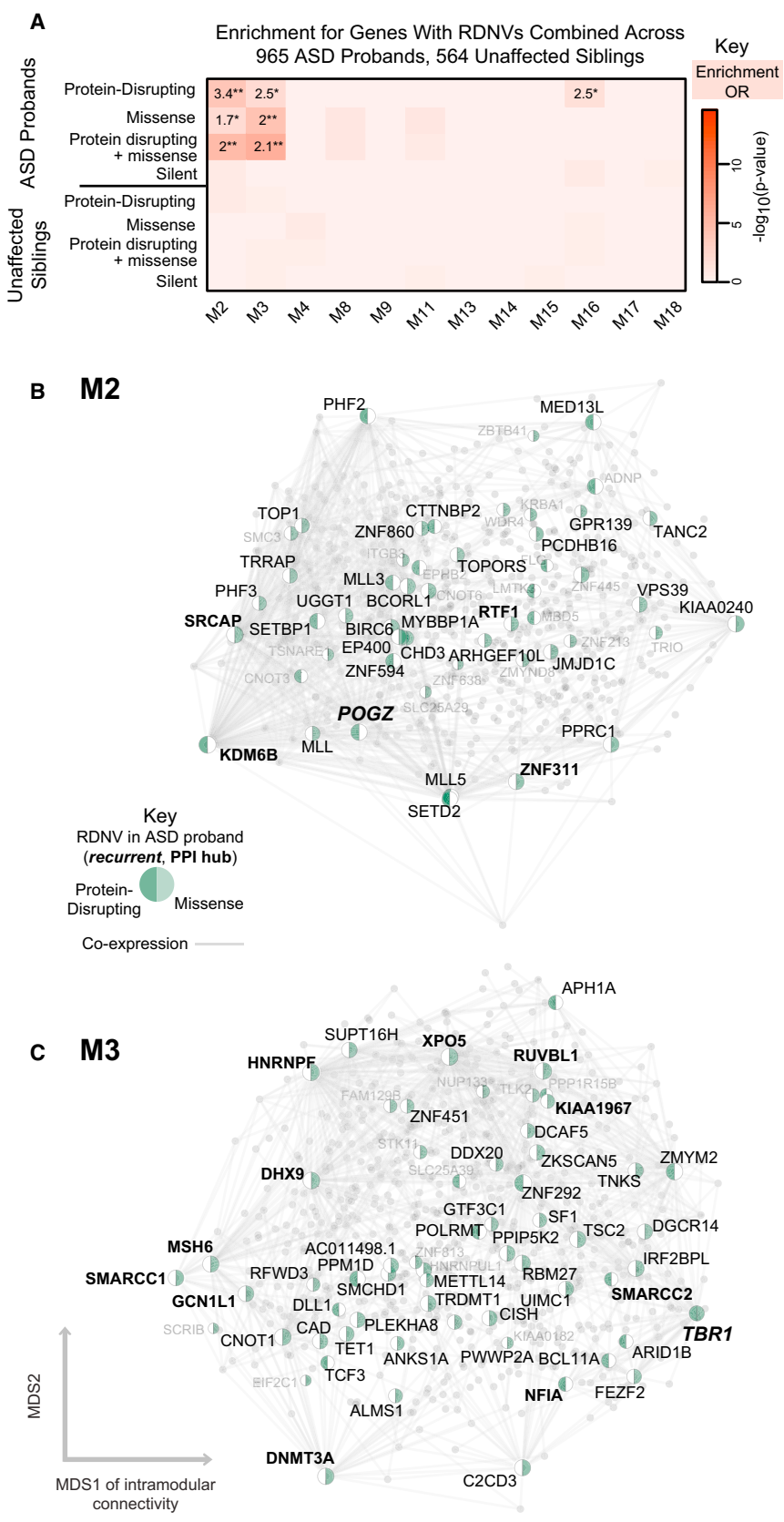
**A**



Figure 3. Enrichment of Genes Affected by RDNVs in Developmental Networks

(A) Module-level enrichment for multiple categories of RDNV in ASD affected probands and unaffected siblings combined across four studies. M2 and M3 are strongly enriched for protein disrupting and missense RDNV-affected genes in probands. Enrichment for genes affected by silent RDNVs in probands and RDNV gene sets affected in siblings represent control gene sets and do not show enrichment. All enrichment values for over-represented lists with p < 0.05, OR > 1 are shown to demonstrate enrichment trends (*p < 0.05, **validated in replication set). Heatmap colors for p values reflect enrichment trends; p values for gene sets with OR < 1 can be seen in Table S2B.

(B and C) (B) and (C) show network plots for M2 and M3, with all genes plotted and all genes carrying RDNVs displayed.

Network plots show all genes in the module with protein disrupting or missense RDNV-affected genes highlighted. For visualization, genes with high intramodular connectivity (kME > 0.75) are labeled in black, and the rest are labeled in gray. By definition, all edges in the network reflect positive correlations. The top 1,000 connections are shown, and genes are plotted according to the multidimensional scaling of coexpression as in Figure 2. See also Figures S2 and S3 and Table S2.

**B  M2**



**C  M3**

proportion (113/598 [19%]) of genes affected by known RDNVs are coexpressed in two modules reflecting similar temporal trends of high expression in cortex during the neurodevelopmental period of early neuronal fate determination, migration, and cortical lamination. Of note, as with M13, M16, and M17, which were enriched for asdM12 and SFARI ASD, ID genes did not show enrichment in M2 or M3 (p > 0.05).

We also observed that the SFARI ASD genes and asdM12 genes, which are enriched for inherited common variants in ASD (small average effect size), affect the synaptic modules, M13, M16, and M17. In contrast, the noninherited (larger average effect size) RDNVs preferentially affect the early transcriptional regulation modules (Extended Experimental Procedures). We emphasize that this is not absolute, as M16 includes some genes harboring RDNVs (e.g., in SCN2A, SHANK2, and NRXN1; Figure 2A). To formally assess common variant enrichment using independent data, we compared ASD GWA signals across these modules (Extended Experimental Procedures). Genes in M13 and M16 were more strongly affected by common variation than M2 or M3 in at least one of two ASD GWA studies (Anney et al., 2012; Wang et al., 2009) (Figure S3E). This is consistent with susceptibility of distinct biological processes for different mutational classes. In general, we predict that more severe neurodevelopmental consequences would result from disrupting early transcriptional dysregulation during neuronal proliferation and differentiation, as compared with later disruption of synaptic development and neuronal function.

### ASD Gene-Enriched Modules Are Linked by Translational and Transcriptional Regulation

Upregulated and downregulated modules are highly anticorrelated throughout development, so we hypothesized that common molecular regulatory relationships could potentially link genes within these modules. We first used a set of FMRP-RNA interactors from a crosslinking and immunoprecipitation (CLIP) experiment (Darnell et al., 2011) because Iossifov et al. (2012) had previously shown that RDNVs identified in their exome sequencing study were enriched in this class of genes. Remarkably, FMRP targets are specifically enriched in modules that also contain ASD-related genes M2, M16, and M17 (FMRP-M2 p = $1.6 \times 10^{-13}$; OR = 3.0 [2.3–3.9]; FMRP-M16 p = $2.4 \times 10^{-29}$; OR = 5.7 [4.3–7.6]; FMRP-M17 p = $9.3 \times 10^{-10}$I OR = 2.4 [1.8–3.1]; all at FDR < 0.05; Figure 4A). This provides a strong, independent line of evidence that translational regulation by FMRP not only affects genes harboring RDNVs but also links different molecular pathways that are coexpressed during early fetal cortical development and are susceptible to diverse classes of ASD genetic mutation.

We next tested whether ASD-associated modules are also linked at the transcriptional level (Experimental Procedures). We found 17 TFs that are predicted to link at least one upregulated and one downregulated module based on binding site enrichment (Figure 4B and Tables S3A and S3B). Many of the genes encoding these TFs are expressed during fetal development (Table S1A), have been previously implicated in relevant neuronal functions, and have DNA binding targets that have been experimentally characterized (Table S3B). For example, MEF2A and MEF2C, both members of a TF family regulating syn-

aptic plasticity and glutamatergic synapse number (Ebert and Greenberg, 2013), are enriched for binding targets in M2 and M17, which are anticorrelated across development (Figures 4C and 4D). SATB1, which is required for the development of cortical interneurons (Close et al., 2012), ELF1, which is involved in axonal guidance, and FOXO1, which regulates neuronal polarity (de la Torre-Ubieta and Bonni, 2011) also link these two modules (Figures 4E and 4F). To provide further evidence that these are experimentally plausible binding sites, we overlaid TF gene bioinformatic predictions with chromatin immunoprecipitation (ChIP) data where available, supporting many of these predicted interactions, including 39% of MEF2A, 23% of MEF2C, and 87% of ELF1 interactions (Figure 4C, 4D, and 4G and Extended Experimental Procedures). These results implicate existing and novel TFs as putative coregulators of ASD-associated gene networks during neocortical development.

### ASD-Associated Genes Exhibit Laminar and Cellular Enrichment

Deficits in cortical patterning have been observed in ASD (Voineagu et al., 2011), so we tested whether ASD-affected genes are enriched in the developing laminae of fetal cortex and the terminally differentiated laminae of adult cortex (Experimental Procedures). We compared multiple ASD gene lists with the ID gene sets for enrichment in laminae of the developing and adult cortex and found a sharp contrast in laminar enrichment between ASD and ID genes (Figures 5A and 5B). Additionally, in adult, asdM12 exhibits strongly significant enrichment in L3 (Z > 2.7, FDR < 0.01), whereas other ASD lists follow a similar trend of superficial layer enrichment (Z > 2, p < 0.05). In contrast, the "ID all" and "ID only" gene sets follow a trend of lower layer enrichment (Figure 5B), an across-layer pattern that is significantly different from all of the ASD lists (Figures 5C and 5D and Extended Experimental Procedures).

We also observed a similar trend in superficial layer (L2–L4) enrichment for the modules that are enriched in asdM12 genes (M13, M16, and M17; Figure 5F). M13 and M16 also exhibit weaker enrichment in deeper layers (L5 and L6). Module-level analysis in fetal brain also highlighted a difference between the RDNV-enriched modules, M2 and M3. Although both M2 and M3 are most highly expressed in early human fetal development (prior to PCW 17), M2 reaches its peak later and is enriched in the cortical plate (CPi/CPo), whereas M3 peaks earlier, which is consistent with its enrichment in the germinal zone (VZ, SZi, and SZo; Figure 5E). In adult, this distinction is no longer present (Figure 5F), with both M2 and M3 showing enrichment in superficial layers (L2 and L4). We also asked whether any of these gene sets or modules were enriched for cell-type-specific marker expression patterns paralleling the observed laminar enrichment. We observed enrichment for a set of well-curated upper-layer glutamatergic neuron markers among asdM12, M2, and M3 genes (Extended Experimental Procedures and Figures S4C and S4D), which agrees with the L2-4 enrichment of asdM12 and ASD risk gene modules.

Figure 6A highlights adult layer-level expression patterns of several strong ASD candidate genes with enriched expression in superficial layers (e.g., SHANK2 and CNTNAP2) and shows that many genes recurrently affected by protein-disrupting
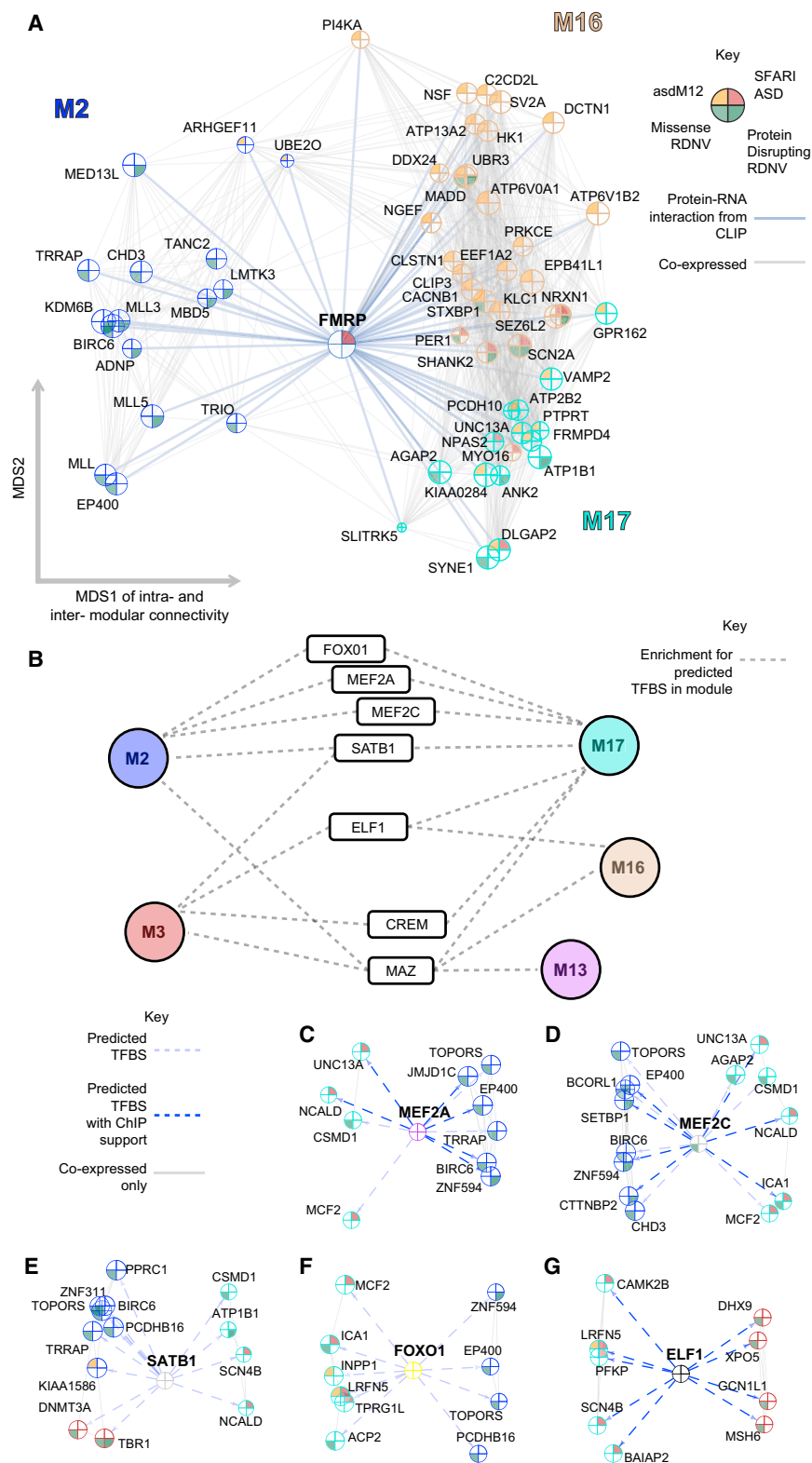
**Figure 4. Translational and Transcriptional Coregulation Connect Developmentally Distinct ASD-Affected Modules**

(A) Coexpression-based network plot of FMRP interactions with genes in M2, M16, and M17 that are either affected by RDNVs or are in an ASD candidate list. Genes are plotted as in Figures 2 and 3 but now across modules, with FMRP placed at the center.

(B) Summary of TF binding site (TFBS) enrichment in modules for TFs that have evidence for function in a neurodevelopmental context and link anti-correlated modules. Dashed lines indicate enrichment in the module for predicted binding sites.

(C–G) MEF2A, MEF2C, SATB1, FOXO1, and ELF1 are all enriched for their binding motifs in the upstream regions of ASD gene-enriched modules following anticorrelated developmental patterns. Network plots highlight genes with a predicted binding site (light dashed arrow) for the TF (placed at the center) contributing to this enrichment that are also affected by RDNVs or in an ASD candidate list. Arrows representing a TFBS found in a ChIP experiment are marked in dark blue.

For network plots, the top 1,000 positive connections between genes are plotted, and node size is proportional to connectivity within the genes' assigned module; therefore, larger nodes are more central hubs. The outer color of each node reflects its module membership, and coexpression edges in the network reflect positive correlations. See also Tables S2 and S3.

*POGZ*; Figure 6B). We use these laminae for cell-marker enrichment analyses because adult laminar expression patterns are more clearly delineated relative to PCW 15–21 (Figures 5A, 5E, S4A, and S4B). Furthermore, neuronal migration in humans persists into the third trimester, and upper-layer neuronal identity is not finalized until after PCW 28 (Bystron et al., 2008). Out of the six genes with recurrent RDNVs in probands in which we can detect layer preference, five are predominantly expressed in superficial layers in adult. Some of the genes in Figure 6 also show expression in a lower layer (*NLGN1*, *SCN2A*, *ITPR1*, and *MLL3*), though superficial layer enrichment is stronger (larger differential expression t value in Table S1A).

## DISCUSSION

Our analyses offer a genome-wide neurobiological context to begin to unify the genetics of ASD, providing robust evidence of both molecular pathway and circuit-level convergence (Figures 7A and 7B). Integration of ASD genes with developmental coexpression

RDNVs in the 965 ASD probands and an additional set of patients assessed by targeted sequencing (O'Roak et al., 2012a) also show superficial layer enrichment (e.g., *SCN2A* and

netics of ASD, providing robust evidence of both molecular pathway and circuit-level convergence (Figures 7A and 7B). Integration of ASD genes with developmental coexpression
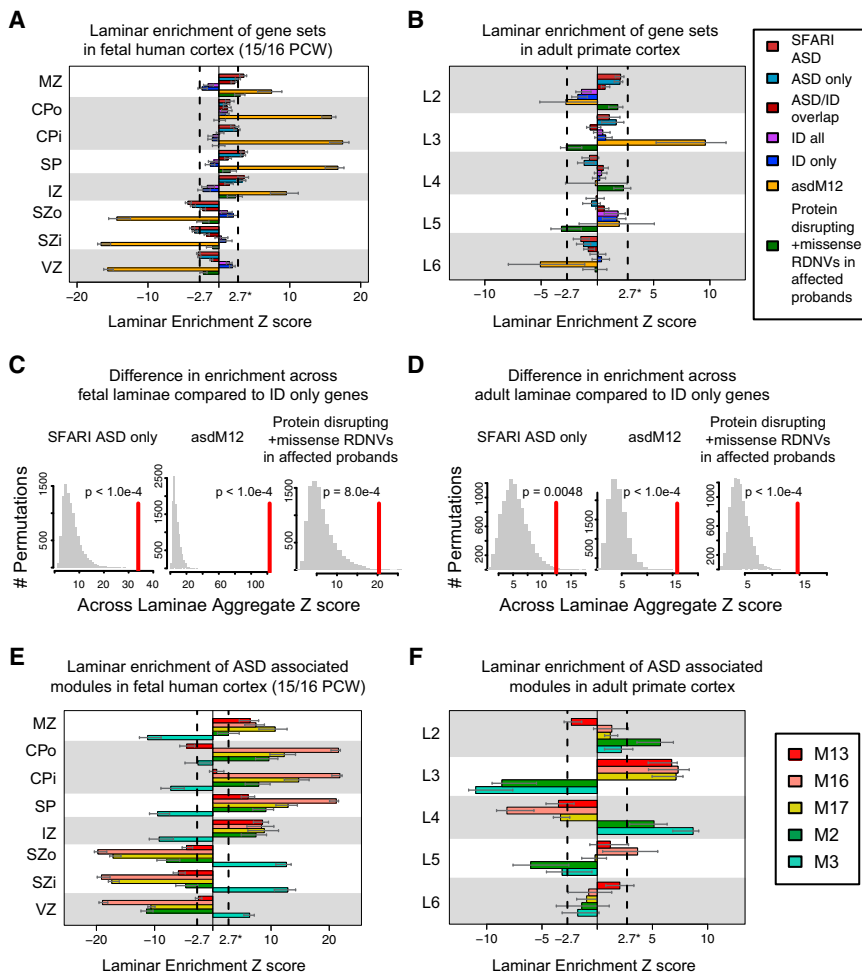
**Figure 5. Enrichment for Laminar Differential Expression of Gene Sets and Associated Developmental Coexpression Modules in Fetal Human and Adult Primate Cortex**

(A) In fetal cortex, ASD sets (SFARI, asdM12, and RDNV affected) are enriched for differential expression in laminae containing postmitotic neurons, whereas genes implicated in ID are weakly enriched in germinal layers. A high $Z$ score for a gene set in a layer corresponds to differential expression across the gene set in that layer.

(B) In adult cortex, asdM12 sets show strong enrichment in layer 3, whereas ID genes are weakly enriched in layer 5.

(C and D) Summing the $Z$ score across layers in (A) and (B) and comparing to randomly permuted sets of genes of similar size demonstrates that, in both fetal and adult cortex, the laminar distribution of multiple ASD implicated gene sets is significantly distinct from that of genes implicated only in ID.

(E) SFARI/asdM12-associated developmental coexpression modules M13, M16, and M17 follow enrichment trends similar to the SFARI/asdM12 gene set in fetal brain. However, the modules strongly associated with the RDNV affected genes, M2 and M3, show distinct enrichment patterns.

(F) ASD-associated modules are predominantly enriched in superficial layers 2–4 of adult cortex. Additionally, M16 shows weak enrichment in L5. In contrast to fetal cortex, M2 and M3 are in enriched in the same laminae in adult, suggesting that they serve distinct functions during cortical development that contribute to superficial cortical layers 2–4.

Dashed lines in bar plots indicate Z = 2.7 (equivalent to FDR = 0.01); error bars indicate 95% bootstrapped CIs. Laminae: marginal zone (MZ), outer/inner cortical plate (CPo/CPi), subplate (SP), intermediate zone (IZ), outer/inner subventricular zone (SZo/SZi), ventricular zone (VZ), and adult cortical layers 2–6 (L2–6). See also Figure S4.

networks and laminar expression data connects multiple ASD risk-enriched modules to glutamatergic neurons in upper cortical layers, tying ASD risk genes to specific brain circuitry (Figure 7C). The observation of convergent biology in ASD stands in striking contrast with ID, which does not show the same level of developmental or anatomical specificity. Laminar enrichment in the "ASD/ID overlap" genes shows a similar pattern as the "ASD only" genes (in L2, Figure 5B). Therefore, disruption in ID genes that also cause ASD affects superficial layers compared to disruption in genes causing ID only. Our analyses lead to the prediction that specific disruption of cortical-cortical connectivity—by targeting upper layer glutamatergic neurons that predominantly comprise inter- and intrahemispheric projections, for example—is more likely to affect core ASD phenotypes such as social behavior, rather than general intellectual ability alone.

Our analysis further links specific molecules and pathways to the cortical-cortical intra- and interhemispheric disconnection that has been hypothesized as a shared circuit-level deficit unifying diverse ASD etiologies (Belmonte et al., 2004; Geschwind and Levitt, 2007). An illustrative example is the disruption

of *ARID1B*, a BAF complex member that harbors a RDNV and is a hub of M3. Severe mutations in *ARID1B* cause corpus callosum abnormalities, ID, and ASD (Halgren et al., 2012; Santen et al., 2012). Another BAF complex member, *SMARCC2*, implicated by RDNVs in probands, controls cortical thickness by repressing the pool of intermediate progenitors, which preferentially contribute to forming cortical layers 2–4 (Tuoc et al., 2013), providing another molecular link to inter- and intrahemispheric connectivity. These analyses make the first systematic connection between genes disrupted in ASD and this circuit-level disruption. As additional genes in the early fetal coexpression modules are found to harbor recurrent RDNVs, cortical-cortical connectivity will be a valuable phenotype to assess in both animal models and human patients.

Translational regulation by FMRP during fetal cortical development and transcriptional coregulation of ASD candidate genes provide another level of convergent biology in ASD and a rich starting point for further experimental investigation. Notable also are TFs that are predicted to drive the transcriptional coregulation of molecular and circuit-level processes, including
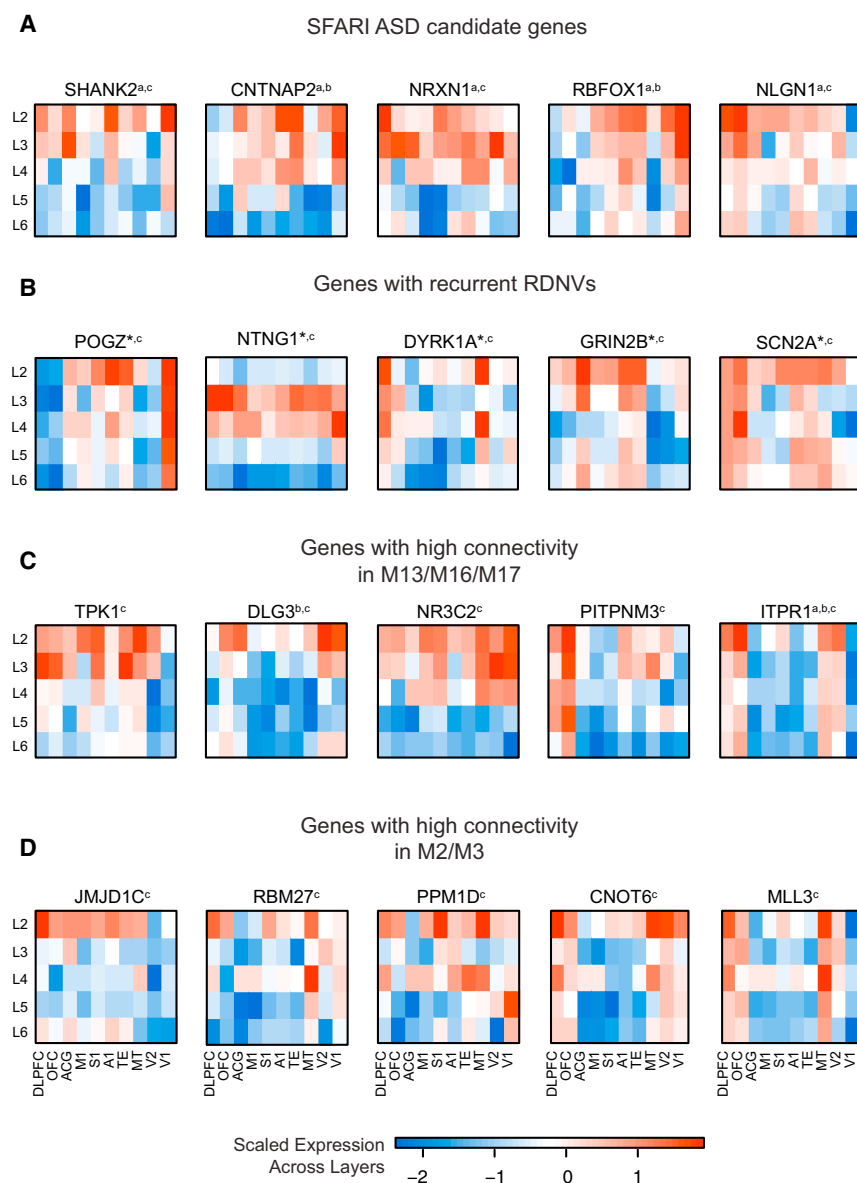
**Figure 6. Laminar Patterns for Genes Implicated in ASD**

(A) SFARI candidate genes for ASD.

(B) Genes with recurrent RDNV evidence across studies. Genes not displayed include *TBR1* (lower layer enriched), *CHD8* (no layer enrichment detected), *CUL3* (no layer enrichment detected), and *KATNAL2* (not detected in these data).

(C) Genes with high connectivity in M13, M16, and M17.

(D) RDNV genes with high connectivity in M2 and M3.

[a] indicates membership in SFARI ASD, [b] indicates membership in asdM12, [c] indicates the gene is affected by a RDNV, and the asterisk indicates recurrent RDNVs.

Color bar values represent scaled expression (SDs from the mean-centered expression value across layers). All genes shown have t > 2 for enrichment in an upper layer (L2, L3, or L4) over background and t < 2 for lower layers (L5 or L6). Regions: dorsolateral prefrontal (DLPFC), orbitofrontal (OFC), anterior central gyrus (ACG), primary motor (M1), primary somatosensory (S1), primary auditory (A1), higher-order visual area TE (TE), higher-order visual area MT/5 (MT), secondary visual cortex (V2), and primary visual cortex (V1).

MEF2A, MEF2C, and SATB1, which have binding site enrichment in M2 and M17. This is intriguing in light of decreased *PVALB* expression in ASD brain (Voineagu et al., 2011), the hypothesized convergent mechanism of a shift in the excitation-inhibition balance in ASD (Rubenstein and Merzenich, 2003), and the observation that SATB1 plays a key role in regulating cortical PV+ and SST+ interneuron development (Close et al., 2012; Denaxa et al., 2012). We speculate that M2 and M17 reflect processes involved in the migration and differentiation of inhibitory and excitatory cell populations whose balanced coregulation may be essential to proper cortical development. These analyses underscore the notion that understanding the structure of the transcriptional and chromatin regulatory networks underlying cortical development and their relationship to translational control will better inform the genetic risk architecture of ASD.

In addition to demonstrating biological convergence, network analysis further allowed us to stratify the full set of 684 RDNV-affected genes to a narrower list of 113 genes (Table S1A) that we hypothesize are more likely to confer increased ASD risk based on their enrichment in M2 and M3 and an elevated probability of conferring a phenotype when haploinsufficient. Furthermore, we demonstrate that the observed enrichment is specific by comparison to silent RDNVs and unaffected siblings' RDNVs. As an example of how to prioritize these candidates further based on the functional relationships summarized in Figure 7, we constructed a list of candidates using Table S1A, filtering by expression during development, membership in M2 or M3, high predicted haploinsufficiency (P[HI] > 0.5), protein disrupting or missense mutation in probands, and either a layer preference (t > 2 for a particular layer) or a cell-type preference (r > 0.2 for a cell type) in Table S4. This yields a set of 24 candidates with a hypothesized layer or cell-type phenotype for investigation. Among these, *TBR1* is known to harbor recurrent mutations, whereas *CHD3* is a member of the same gene family as *CHD8*, a gene with strong recurrent de novo mutation evidence (O'Roak et al., 2012a). Additionally, *SMARCC1* and *SMARCC2* are members of the BAF complex, which is of particular interest because it is statistically associated with ASD—6/28 BAF complex genes are affected by RDNVs (p = 0.0015; OR = 5.7 [1.9–14.5]). Remarkably, one of the genes in M2 implicated by our
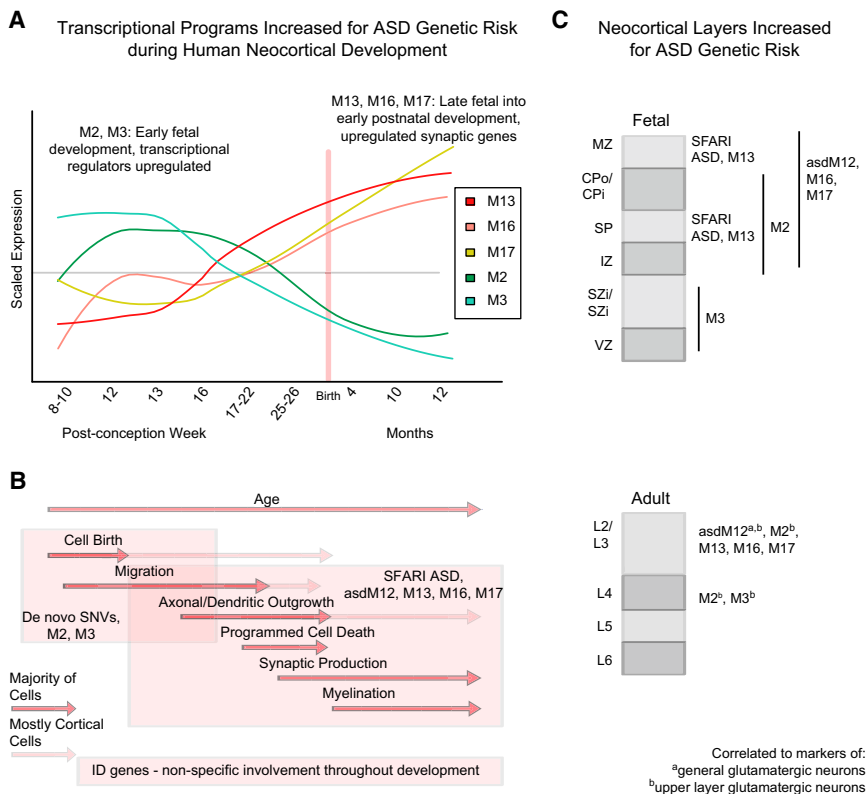
**A** Transcriptional Programs Increased for ASD Genetic Risk during Human Neocortical Development

**B**

**C** Neocortical Layers Increased for ASD Genetic Risk

**Figure 7. Summary of Findings and Model for Effects of ASD Implicated Gene Sets**

(A) ASD risk genes from multiple sources were enriched in five coexpression modules throughout development—M2, M3, M13, M16, and M17.

(B) Early transcriptional regulators in M2/M3 are enriched for RDNVs, whereas the later expressed synaptic genes are associated with previously studied ASD genes (biological process time periods adopted from Andersen [2003]).

(C) ASD genes are most consistently associated with laminae containing postmitotic neurons during early fetal development (broadly in IZ, SP, CPo/CPi, and MZ) and superficial layers in adult (L2–L4). Multiple modules are also strongly associated with markers of upper-layer glutamatergic neurons in adult cortex, suggesting many ASD genes preferentially affect these cell types.

(B) and (C) also summarize that ID genes are largely distinct from ASD genes in both developmental trajectory and neocortical layer enrichment.

See also Table S4. Both (A) and (B) correspond to the same timescale as marked by the axis on the plot in (A). We summarize the strongly enriched findings but note that weaker enrichment for other patterns exists that may be important for subsets of ASD. Individual genes can be prioritized for biological validation using a combination of network position, bioinformatic scores, and the biological context highlighted here, as described in the Discussion and as shown in Table S4.

prioritization is *TOP1* (also highlighted in Figures 3B, and S2A, and S2B), as it contains a missense RDNV, has a P(HI) of 0.99, and is correlated with upper-layer glutamatergic neuronal markers. *TOP1* has been shown to regulate the transcription of long transcripts preferentially implicated in ASD (King et al., 2013). Therefore, M2 provides many potential interactions to investigate at a mechanistic level, as it links *TOP1* with other regulators of chromatin structure expressed during cortical development that include members of CCR4-NOT complex (CNOT family) and chromodomain helicase DNA-binding proteins (CHD), which have previously been linked to the regulation of neuronal proliferation and differentiation (Feng et al., 2013; Potts et al., 2011; Ronan et al., 2013; Zheng et al., 2012).

In parallel work in this issue of *Cell*, Willsey et al. (2013) find strongest convergence on fetal developmental coexpression networks in frontal lobe by seeding with a subset of high-confidence ASD genes identified by exome sequencing. Despite the different analytical approaches, there is remarkable overlap between the developmental processes implicated by the gene networks identified in our studies. Although we see the strongest cell type and layer enrichment in adult L2–L4, we also see a signal in CPi during fetal development and a weaker signal in L5–L6 of adult, which is consistent with a subset of genes affecting lower-layer glutamatergic neurons. Together, our studies highlight the importance of understanding the spatial and temporal context of specific genes for future mechanistic investigation.

We also acknowledge several issues that challenged our approach. Many of the genes we identified as putatively involved in ASD do not have complete PPI data, P(HI) scores, TF binding site information, or are not well studied in brain. This is one reason why we rely most heavily on RNA-seq-based transcriptome data, as they comprehensively represent relationships present in the developing human brain in an unbiased manner. We did not assess enrichment of genetic hits in other brain regions across development, as sample size and cell-type heterogeneity make it difficult to interpret coexpression across cytoarchitecturally diverse brain regions such as cerebellum and amygdala, which may also be involved in ASD (Amaral et al., 2008). We also focused on single gene disruption in ASD and did not include CNVs affecting multiple genes to improve signal to noise.

Additionally, current genetic approaches favor de novo mutation detection; as different classes of mutations (e.g., inherited rare coding or noncoding regulatory variants) are identified, we speculate that heritable variants will have less severe phenotypic consequences and will affect genes in the modules related to synaptic development and function, rather than earlier transcriptional regulation. Likewise, it will also be useful to investigate rare, inherited recessive ASD risk variants (Lim et al., 2013; Yu et al., 2013) when sufficient data are available, so as to compare it with other forms of genetic variation. Importantly, as larger sets of individuals are sequenced, it will be essential to look at how mutational effect relates to biological effect, as recent work (Yu et al., 2013) has shown that ASD can result from milder hits to the same proteins affected in more severe disorders associated

with severe ID. Here, we investigated a large composite list of known ID genes that reflects multiple mutational mechanisms. Certain subgroups of ID genes such as those implicated in X-linked ID (Lubs et al., 2012) or those from de novo disruptions in individuals with severe ID (de Ligt et al., 2012; Rauch et al., 2012) overlap with genes in M2 and M3 that also have RDNVs found in ASD. This is consistent with the observation that hits in M2 and M3 are highly deleterious to brain development. Although more specificity for these subgroups of ID genes may arise as additional individuals are sequenced, our analyses indicate that the degree of stage or regional specificity for ID genes is far less than that observed in ASD.

The conclusions summarized in Figure 7 pass a stringent multiple comparisons cut-off; weaker enrichment patterns may become more salient with higher-resolution tiling of gene expression during development and increased sample sizes in sequencing studies. To facilitate future studies, we have shared the code used in this analysis (Extended Experimental Procedures) and have provided a network browser for user-friendly interactive exploration of specific genes, including links to other public data (Experimental Procedures). We have shown how an integrative approach, which is not driven by any small set of samples, candidate genes, or candidate hypotheses, can place heterogeneous genetic etiologies into a unifying structure. These analyses provide a working framework for mechanistic investigation and hypothesis testing, which points to interactions between genes in specific cell types and circuits, as well as the general biological processes in which these genes are implicated.

## EXPERIMENTAL PROCEDURES

### Developmental Expression Data

BrainSpan developmental RNA-seq data (BrainSpan, 2013) summarized to Gencode 10 (Harrow et al., 2006) gene-level reads per kilobase million mapped reads (RPKM) values were used (Extended Experimental Procedures for data preprocessing; see Table S1D for sample details). Only neocortical regions were used in our analysis, and only genes with a normalized RPKM value of 1 in at least one region at one time point for 80% of the available samples were considered expressed.

### Weighted Gene Coexpression Network Analysis

We used the R package WGCNA (Langfelder et al., 2008) to construct coexpression networks, as previously done (Voineagu et al., 2011) and as described in detail in the Extended Experimental Procedures. The modules were characterized using GO Elite to control the network-wide FDR, with all enriched pathways comprising at least ten genes at Z > 2 and FDR < 0.05 (Zambon et al., 2012). All network plots were constructed using the igraph package in R (Csárdi and Nepusz, 2006).

### Protein-Protein Interaction Enrichment

When assessing PPI enrichment in modules, a degree-matched permutation analysis was applied in order to control for biological and methodological biases in PPI data (see Extended Experimental Procedures for details).

### Gene Sets

The SFARI ASD set was compiled using the online SFARI gene database, AutDB. We used the "Gene Score," which classifies evidence levels, to restrict our set to those categorized as Syndromic (S) and evidence levels 1–4 (high-confidence—minimal evidence). We obtained asdM12 and adsM16 from a prior, independent gene expression study that profiled expression changes in ASD cortex and applied WGCNA to identify modules of dysregulated genes ASD (Voineagu et al., 2011). We curated ID genes from four reviews cataloging

genes causing "ID all" (Inlow and Restifo, 2004; Lubs et al., 2012; Ropers, 2008; van Bokhoven, 2011) resulting in 401 genes. For candidate lists, we used the HUGO gene nomenclature to find updated gene symbols. We obtained RDNVs from four publications (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012b; Sanders et al., 2012) and split them into discovery and validation sets as discussed in the results (see Extended Experimental Procedures for further details about gene sets).

### Gene Set Overrepresentation

All enrichments of gene sets were performed using a two-sided Fisher's exact test with 95% confidence calculated according to the R function fisher.test. The FDR was controlled across candidate ASD gene set enrichments, the discovery RDNV set enrichment, and FMRP target enrichment (Table S2B). For RDNV enrichment, we required an OR > 1 and an FDR-adjusted p value < 0.05 for enrichment in the discovery set and OR > 1 with p < 0.05 for validation in the replication set. When claiming a lack of enrichment, we require an uncorrected p > 0.05 to reduce false negatives, as future studies that add expression time points for networks and genes for enrichment may find enrichment in pathways not significantly enriched here.

### Transcription Factor Binding Site Enrichment

The top 200 genes in each module (ranked by kME) were used for TF motif enrichment analysis. Enrichment for each TF motif in TRANSFAC (Matys et al., 2003) was compared to three background data sets to ensure robustness: 1,000 bp sequences upstream of all human genes, human CpG islands, and the sequence of human chromosome 20 (Extended Experimental Procedures). Only TFs with p < 0.05 across all backgrounds are considered enriched. ChIP data were obtained from ENCODE (ENCODE Project Consortium, 2011) and the ChIP enrichment analysis (Lachmann et al., 2010) resource.

### Layer-Specific and Cell-Type Marker Enrichment

We utilized human fetal neocortical laminar gene expression data sets from BrainSpan at PCW 15/16 and PCW 21 and primate neocortical laminar gene expression data from a published study (Bernard et al., 2012). For laminar specificity, differential expression of each gene in each layer was calculated against background, resulting in t values for each gene in each layer (Table S1A). We quantified the skew of differential expression t values of each gene set in each layer, applied a FDR cutoff across all enrichments in all layers (Z = 2.7 and FDR = 0.01), and computed bootstrapped confidence intervals to assess enrichment of gene sets in layers. To quantify cell-marker relationships, we used an analogous method, replacing the t value 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). Statistical comparison of enrichment trends across layers between ASD and ID gene sets set was performed by comparing the distribution of scores across layers using a permutation analysis (Extended Experimental Procedures).

### Interactive Network Plot

We have made the coexpression network and associated gene-level data available for the top 500 connections in each module in an interactive browser at the following website (http://geschwindlab.neurology.ucla.edu/sites/all/files/networkplot/ParikshakDevelopmentalCortexNetwork.html).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.10.031.

## REFERENCES

Amaral, D.G., Schumann, C.M., and Nordahl, C.W. (2008). Neuroanatomy of autism. Trends Neurosci. *31*, 137–145.

Andersen, S.L. (2003). Trajectories of brain development: point of vulnerability or window of opportunity? Neurosci. Biobehav. Rev. *27*, 3–18.

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. Hum. Mol. Genet. *21*, 4781–4792.

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

Belmonte, M.K., Allen, G., Beckel-Mitchener, A., Boulanger, L.M., Carper, R.A., and Webb, S.J. (2004). Autism and abnormal development of brain connectivity. J. Neurosci. *24*, 9228–9231.

Ben-David, E., and Shifman, S. (2013). Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. Mol. Psychiatry *18*, 1054–1056.

Berg, J.M., and Geschwind, D.H. (2012). Autism genetics: searching for specificity and convergence. Genome Biol. *13*, 247.

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.

BrainSpan, 2013. BrainSpan: Atlas of the Developing Human Brain [Internet]. http://www.brainspan.org.

Bystron, I., Blakemore, C., and Rakic, P. (2008). Development of the human cerebral cortex: Boulder Committee revisited. Nat. Rev. Neurosci. *9*, 110–122.

Close, J., Xu, H., De Marco García, N., Batista-Brito, R., Rossignol, E., Rudy, B., and Fishell, G. (2012). Satb1 is an activity-modulated transcription factor required for the terminal differentiation and connectivity of medial ganglionic eminence-derived cortical interneurons. J. Neurosci. *32*, 17690–17705.

Courchesne, E., Campbell, K., and Solso, S. (2011). Brain growth across the life span in autism: age-specific changes in anatomical pathology. Brain Res. *1380*, 138–145.

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

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247–261.

de la Torre-Ubieta, L., and Bonni, A. (2011). Transcriptional regulation of neuronal polarity and morphogenesis in the mammalian brain. Neuron *72*, 22–40.

de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C.,

et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. *367*, 1921–1929.

Denaxa, M., Kalaitzidou, M., Garefalaki, A., Achimastou, A., Lasrado, R., Maes, T., and Pachnis, V. (2012). Maturation-promoting activity of SATB1 in MGE-derived cortical interneurons. Cell Rep. *2*, 1351–1362.

Devlin, B., and Scherer, S.W. (2012). Genetic architecture in autism spectrum disorder. Curr. Opin. Genet. Dev. *22*, 229–237.

Ebert, D.H., and Greenberg, M.E. (2013). Activity-dependent neuronal signalling and autism spectrum disorder. Nature *493*, 327–337.

Ecker, C.C., Suckling, J.J., Deoni, S.C.S., Lombardo, M.V.M., Bullmore, E.T.E., Baron-Cohen, S.S., Catani, M.M., Jezzard, P.P., Barnes, A.A., Bailey, A.J.A., et al.; MRC AIMS Consortium. (2012). Brain anatomy and its relationship to behavior in adults with autism spectrum disorder: a multicenter magnetic resonance imaging study. Arch. Gen. Psychiatry *69*, 195–209.

ENCODE Project Consortium. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. *9*, e1001046.

Feng, W., Khan, M.A., Bellvis, P., Zhu, Z., Bernhardt, O., Herold-Mende, C., and Liu, H.-K. (2013). The chromatin remodeler CHD7 regulates adult neurogenesis via activation of SoxC transcription factors. Cell Stem Cell *13*, 62–72.

Geschwind, D.H. (2011). Genetics of autism spectrum disorders. Trends Cogn. Sci. *15*, 409–416.

Geschwind, D.H., and Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. Curr. Opin. Neurobiol. *17*, 103–111.

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.

Halgren, C., Kjaergaard, S., Bak, M., Hansen, C., El-Schich, Z., Anderson, C.M., Henriksen, K.F., Hjalgrim, H., Kirchhoff, M., Bijlsma, E.K., et al. (2012). Corpus callosum abnormalities, intellectual disability, speech impairment, and autism in patients with haploinsufficiency of ARID1B. Clin. Genet. *82*, 248–255.

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. Genome Biol. *7* (Suppl 1), 1–9.

Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet. *6*, e1001154.

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.

King, I.F., Yandava, C.N., Mabb, A.M., Hsiao, J.S., Huang, H.-S., Pearson, B.L., Calabrese, J.M., Starmer, J., Parker, J.S., Magnuson, T., et al. (2013). Topoisomerases facilitate transcription of long genes linked to autism. Nature *501*, 58–62.

Klei, L., Sanders, S.J., Murtha, M.T., Hus, V., Lowe, J.K., Willsey, A.J., Moreno-De-Luca, D., Yu, T.W., Fombonne, E., Geschwind, D., et al. (2012). Common genetic variants, acting additively, are a major source of risk for autism. Mol. Autism *3*, 9.

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.

Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., et al.; NHLBI Exome Sequencing Project. (2013). Rare complete knockouts in humans: population

distribution and significant role in autism spectrum disorders. Neuron 77, 235–242.

Lubs, H.A., Stevenson, R.E., and Schwartz, C.E. (2012). Fragile X and X-linked intellectual disability: four decades of discovery. Am. J. Hum. Genet. 90, 579–590.

Luo, R., Sanders, S.J., Tian, Y., Voineagu, I., Huang, N., Chu, S.H., Klei, L., Cai, C., Ou, J., Lowe, J.K., et al. (2012). Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. Am. J. Hum. Genet. 91, 38–55.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al.; 1000 Genomes Project Consortium. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science 335, 823–828.

Matson, J.L., and Shoemaker, M. (2009). Intellectual disability and its relation-ship to autism spectrum disorders. Res. Dev. Disabil. 30, 1107–1114.

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.

Michaelson, J.J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., et al. (2012). Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell 151, 1431–1442.

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 485, 242–245.

O'Roak, B.J., Vives, L.L., Fu, W.W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G.G., Kumar, A.A., Lee, C.C., Ankenman, K.K., et al. (2012a). Multi-plex targeted sequencing identifies recurrently mutated genes in autism spec-trum disorders. Science 338, 1619–1622.

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. (2012b). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485, 246–250.

Potts, R.C., Zhang, P., Wurster, A.L., Precht, P., Mughal, M.R., Wood, W.H., 3rd, Zhang, Y., Becker, K.G., Mattson, M.P., and Pazin, M.J. (2011). CHD5, a brain-specific paralog of Mi2 chromatin remodeling enzymes, regulates expression of neuronal genes. PLoS ONE 6, e24515.

Ramani, A.K., Li, Z., Hart, G.T., Carlson, M.W., Boutz, D.R., and Marcotte, E.M. (2008). A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. Mol. Syst. Biol. 4, 180.

Rauch, A.A., Wieczorek, D.D., Graf, E.E., Wieland, T.T., Endele, S.S., Schwarz-mayr, T.T., Albrecht, B.B., Bartholdi, D.D., Beygo, J.J., Di Donato, N.N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380, 1674–1682.

Ronan, J.L., Wu, W., and Crabtree, G.R. (2013). From neural development to cognition: unexpected roles for chromatin. Nat. Rev. Genet. 14, 347–359.

Ropers, H.H. (2008). Genetics of intellectual disability. Curr. Opin. Genet. Dev. 18, 241–250.

Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C., Daly, M.J., and Daly, M.J.; International Inflammatory Bowel Disease Genetics Consortium. (2011). Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 7, e1001273.

Rubenstein, J.L.R. (2011). Annual Research Review: Development of the cere-bral cortex: implications for neurodevelopmental disorders. J. Child Psychol. Psychiatry 52, 339–355.

Rubenstein, J.L.R., and Merzenich, M.M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. Genes Brain Behav. 2, 255–267.

Sakai, Y., Shaw, C.A., Dawson, B.C., Dugas, D.V., Al-Mohtaseb, Z., Hill, D.E., and Zoghbi, H.Y. (2011). Protein interactome reveals converging molecular pathways among autism disorders. Sci. Transl. Med. 3, 86ra49.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Will-sey, 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.

Santen, G.W.E., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., Kant, S.G., Snoeck, I.N., Peeters, E.A.J., Hilhorst-Hofstee, Y., et al. (2012). Muta-tions in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. Nat. Genet. 44, 379–380.

Srinivasan, K., Leone, D.P., Bateson, R.K., Dobreva, G., Kohwi, Y., Kohwi-Shigematsu, T., Grosschedl, R., and McConnell, S.K. (2012). A network of genetic repression and derepression specifies projection fates in the devel-oping neocortex. Proc. Natl. Acad. Sci. USA 109, 19071–19078.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 34 (Database issue), D535–D539.

Tuoc, T.C., Boretius, S., Sansom, S.N., Pitulescu, M.-E., Frahm, J., Livesey, F.J., and Stoykova, A. (2013). Chromatin regulation by BAF170 controls cere-bral cortical size and thickness. Dev. Cell 25, 256–269.

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., 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 disor-ders. Nature 459, 528–533.

Willsey, J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155, this issue, 997–1007.

Yoo, A.S., Staahl, B.T., Chen, L., and Crabtree, G.R. (2009). MicroRNA-medi-ated switching of chromatin-remodelling complexes in neural development. Nature 460, 642–646.

Yu, T.W., Chahrour, M.H., Coulter, M.E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., Schmitz-Abe, K., Harmin, D.A., Adli, M., Malik, A.N., et al. (2013). Using whole-exome sequencing to identify inherited causes of autism. Neuron 77, 259–273.

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 28, 2209–2210.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4, e17.

Zheng, X., Dumitru, R., Lackford, B.L., Freudenberg, J.M., Singh, A.P., Archer, T.K., Jothi, R., and Hu, G. (2012). Cnot1, Cnot2, and Cnot3 maintain mouse and human ESC identity and inhibit extraembryonic differentiation. Stem Cells 30, 910–922.