nature neuroscience

Genotype to phenotype relationships in autism spectrum disorders

Jonathan Chang[1,2,5], Sarah R Gilman[1,2,5], Andrew H Chiang[1,2], Stephan J Sanders[3,4] & Dennis Vitkup[1,2]

Autism spectrum disorders (ASDs) are characterized by phenotypic and genetic heterogeneity. Our analysis of functional networks perturbed in ASD suggests that both truncating and nontruncating *de novo* mutations contribute to autism, with a bias against truncating mutations in early embryonic development. We find that functional mutations are preferentially observed in genes likely to be haploinsufficient. Multiple cell types and brain areas are affected, but the impact of ASD mutations appears to be strongest in cortical interneurons, pyramidal neurons and the medium spiny neurons of the striatum, implicating cortical and corticostriatal brain circuits. In females, truncating ASD mutations on average affect genes with 50–100% higher brain expression than in males. Our results also suggest that truncating *de novo* mutations play a smaller role in the etiology of high-functioning ASD cases. Overall, we find that stronger functional insults usually lead to more severe intellectual, social and behavioral ASD phenotypes.

ASDs are associated with a wide range of cognitive and behavioral abnormalities[1,2]. It is estimated that many hundreds of genes may contribute to autism and related phenotypes[2,3]. Functionally important *de novo* mutations associated with ASDs are individually rare, but their collective contribution to sporadic ASD cases is likely to be substantial as a result of a large number of target genes[3–10]. Several recent studies have identified a large collection of *de novo* mutations associated with ASDs, including copy number variations (CNVs)[6] and single nucleotide variations (SNVs)[3,7,8,11]. These studies demonstrated that truncating SNVs (such as nonsense, splice site and frameshift mutations) and large CNVs are likely to have a causal role in ASDs.

To explore the underlying biological pathways, we previously developed a computational approach (NETBAG+) that searches for cohesive biological networks using a diverse collection of disease-associated genetic variants[12,13]. Using network-based approaches, we and others recently found that genetic variations associated with ASDs and other psychiatric disorders converge on several biological networks that are involved in neurogenesis and synaptic function[12–15]. In parallel with the identification of disease-associated genetic variations, complementary data sets of brain-related functional and phenotypic resources are rapidly being accumulated. These include a comprehensive database of gene expression across different cell types, distinct anatomical brain regions and developmental stages[16,17]. In addition, resources such as the Simons Simplex Collection (SSC)[18] have assembled a large compendium of ASD-related phenotypic data, including intelligence and social phenotypic scores. We focused our analyses on a set of genes that were implicated by our network-based computational approach and on all *de novo* truncating mutations from several recent studies. These two independent approaches provided us with complementary genes sets enriched in causal ASD mutations. We investigated the temporal, spatial and cell-specific expression profiles of implicated genes. We also explored how expression and functional properties of autism-associated genes affect ASD phenotypes.

RESULTS

Functional gene networks affected by *de novo* mutations

To elucidate functional networks perturbed in ASD, we applied NETBAG+ to a set of genes affected by *de novo* CNVs and SNVs observed in autistic patients from the SSC[3,6–8]. All of the mutations used as the input for our analyses were obtained using genome-wide methodologies and therefore are not biased by any preexisting hypotheses of ASD etiology. The combined input data contained a total of 991 unique genes from 624 independent genomic loci, including 580 unique genes with *de novo* SNVs and 434 genes within *de novo* CNVs (**Supplementary Table 1**); we note that the number of genomic loci used was considerably larger than the 47 loci considered in our previous analysis of *de novo* CNV events in autism[13]. We used NETBAG+ to identify a subset of the input genes that are strongly connected in the underlying phenotypic network (Online Methods). The NETBAG+ search revealed a functional network containing 159 genes ($P = 0.036$; **Fig. 1** and **Supplementary Table 2**), of which 131 genes were affected by *de novo* SNVs and 31 by *de novo* CNVs. The network's significance was estimated using random input sets that matched the real data in terms of protein length and network connectivity. Notably, no significant networks were detected using genes associated with the 368 non-synonymous *de novo* mutations identified in siblings.

To explore the biological functions associated with the implicated network, we used DAVID[19] to identify Gene Ontology (GO) terms that were significantly enriched among network gene

[1]Department of Biomedical Informatics, Columbia University, New York, New York, USA. [2]Department of Systems Biology, Center for Computational Biology and Bioinformatics, Columbia University, New York, New York, USA. [3]Department of Psychiatry, University of California, San Francisco, California, USA. [4]Department of Psychiatry, Department of Genetics, Yale University, New Haven, Connecticut, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to D.V. (dv2121@columbia.edu).

Received 12 August; accepted 26 November; published online 22 December 2014; doi:10.1038/nn.3907

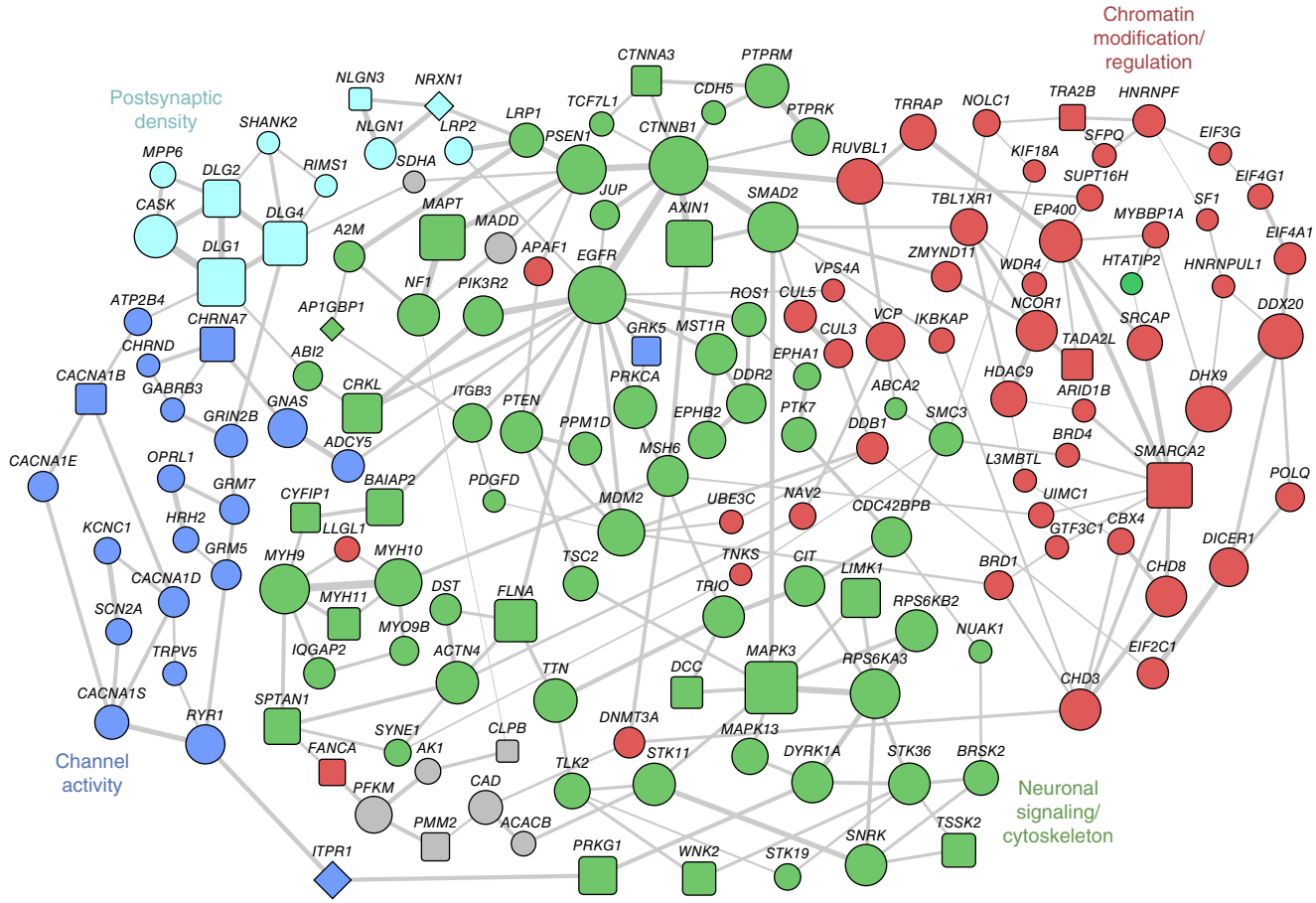

Figure 1 The network implicated by NETBAG+ based on ASD-associated *de novo* SNVs and CNVs from recent studies (network is comprised of 159 genes, $P = 0.036$). Node sizes are proportional to the contributions of each gene to the overall network score and edge widths are proportional to the likelihood that the corresponding gene pair contributes to the same genetic phenotype (Online Methods). For clarity, only the two strongest edges for each gene are shown. Node shapes indicate types of the corresponding mutations: circles represent genes from SNVs, squares represent genes from CNVs and diamonds represent genes affected by both mutation types. The network was divided into four cohesive functional clusters (indicated by node colors) using hierarchical clustering (**Supplementary Fig. 1**); general biological functions of these clusters, determined using DAVID, are shown (see **Supplementary Table 3** for a complete list of GO terms associated with each cluster). Gray nodes represent genes that are not members of the network clusters.

annotations (**Table 1**). This analysis identified a diverse set of functions associated with the implicated network, including synaptic function, chromatin modification and calcium channel activity. To better understand the functional relationships between genes in the network, we performed hierarchical clustering of the network genes using the strength of interactions in the phenotypic network as a metric (**Supplementary Fig. 1**). This analysis identified four major network clusters, each associated with distinct and non-overlapping biological functions (**Supplementary Table 3**).

One cluster in our network (**Fig. 1**) contained genes responsible for synapse formation and function. This cluster contained neurexin (*NRXN*) and neuroligin (*NRLG*), as well as important components (*SHANK2* and *DLG2/DLG4*) of the postsynaptic density (PSD) of excitatory synapses. A related cluster included a diverse set of ion channels and receptors. This cluster contained genes (*CACNA1B*, *CACNA1D*, *CACNA1E*, *CACNA1S*) for subunits of several voltage-dependent calcium channels, which are important in learning and memory[20]. The largest cluster of the implicated network contained genes that are associated with neuronal signaling and migration (*NF1, DCC, EPHA1/B2*), intracellular signaling (*MAPK3, EGFR, PTEN, MDM2, CTNNB1, PRKCA, LIMK1*), and the development of neuronal projections and actin cytoskeleton (*CYFIP1, TRIO, SPTAN1, FLNA, ACTN4*). As we have previously demonstrated, ASD-associated mutations often perturb multiple signaling pathways that converge on the regulation of actin cytoskeleton and other structural processes that are required for neuron migration, cell-cell adhesion and the development of neuronal projections[12,21,22]. Many genes in the clusters described above encode proteins related to the growth and function of dendritic spines, which have been implicated in our previous analyses of genetic insults in ASDs and schizophrenia[12,13]. The final cluster was primarily related to functions associated with chromatin modifications, chromatin remodeling and transcriptional regulation. Notably, there is growing evidence that chromatin regulatory mechanisms crucially affect various stages of neural development, neuroplasticity and learning[23]. This cluster also contained genes involved in other processes that have been implicated in ASDs: RNA interference (*DICER1, AGO1/EIF2C1*), translation (*EIF4A1, EIF4G1, EIF3G*) and splicing (*SFPQ, TRA2B*)[24].

Association of the implicated genes with specific functional subsets

In addition to the molecular and biological GO categories discussed above, we investigated whether there was an overlap between the network and specific gene subsets that may highlight the phenotypic and functional properties of the network genes. Recent exome-sequencing studies have revealed recurrent truncating mutations in several genes (*ADNP, ANK2, ARID1B, CHD8, CUL3, DYRK1A, GRIN2B,*

Table 1 GO terms associated with the implicated network

GO ID	Ontology	Term	*P* value
GO:0045202	CC	Synapse	3×10^{-7}
GO:0016568	BP	Chromatin modification	6×10^{-6}
GO:0051015	MF	Actin filament binding	7×10^{-6}
GO:0005262	MF	Calcium channel activity	1×10^{-5}
GO:0031252	CC	Cell leading edge	2×10^{-5}
GO:0004386	MF	Helicase activity	3×10^{-5}
GO:0015629	CC	Actin cytoskeleton	6×10^{-5}
GO:0030036	BP	Actin cytoskeleton organization	7×10^{-5}
GO:0016887	MF	ATPase activity	7×10^{-5}
GO:0005913	CC	Cell-cell adherens junction	8×10^{-5}
GO:0045211	CC	Postsynaptic membrane	8×10^{-5}
GO:0044456	CC	Synapse part	0.0001
GO:0007611	BP	Learning or memory	0.0001
GO:0030029	BP	Actin filament-based process	0.0001
GO:0005938	CC	Cell cortex	0.0001
GO:0005911	CC	Cell-cell junction	0.0002
GO:0014069	CC	Postsynaptic density	0.002
GO:0043005	CC	Neuron projection	0.004
GO:0030425	CC	Dendrite	0.004

GO terms enriched in the implicated network (**Fig. 1**), as identified by DAVID david.abcc.ncifcrf.gov. *P* values shown in the table were corrected for multiple hypotheses testing using the Benjamini-Hochberg procedure in DAVID. The ontology column indicates GO domain: BP for biological process, MF for molecular function and CC for cellular component. Nonspecific terms, that is, terms associated with more than 400 human genes, and redundant terms are not shown. For a list of GO terms associated with each network cluster, see **Supplementary Table 3**. The enrichment calculations were performed using the default background set of all human genes; similar enrichment results were also obtained using brain-expressed genes as background (**Supplementary Table 4**).

KATNAL2, POGZ, SCN2A and *TBR1*)[3,7,8,11,25]. Although only 131 of 580 (~23%) of all genes harboring missense SNVs form the implicated network, 6 of 11 genes with recurrent truncating mutations are in the network (Fisher's exact one-tail test, $P = 0.02$). This suggests that there is a substantial enrichment of causal genes in the implicated network. Because ASD *de novo* mutations are predominantly heterozygous, it is likely that true causal mutations would preferentially target haploinsufficient genes. Indeed, using haploinsufficiency probabilities from a recent study[26], we found that genes in the implicated network were significantly more likely to be haploinsufficient than genes that were not selected by NETBAG+ (median probability for network and non-network genes are 0.57 and 0.32, respectively; Wilcoxon rank-sum two-tail test, $P < 10^{-14}$); a similar result was also observed for genes with recurrent truncating mutations (median probability for genes with recurrent truncating mutations and unaffected sibling SNV genes are 0.69 and 0.39, respectively; Wilcoxon rank-sum two tail, $P = 0.016$). We also considered the Genomic Evolutionary Rate Profiling (GERP) scores to characterize the severity of SNV mutations (see Online Methods). Notably, the average GERP score of network genes was significantly higher than the average GERP score of genes not selected into the network (average network and non-network GERP scores are 4.0 and 3.3, respectively; Wilcoxon rank-sum one tail, $P = 0.001$). As higher GERP scores correspond to more damaging mutations, this analysis demonstrates that NETBAG predominantly selects genes harboring SNVs with potentially stronger functional effect.

Notably, a recent exome study showed a significant overlap between genes harboring truncating ASD *de novo* SNVs and targets of the fragile X mental retardation protein (FMRP)[3,27]. FMRP is an RNA-binding protein that is essential for a wide array of cognitive functions including synaptic plasticity and learning[28,29]. Failure to properly express FMRP can cause fragile X syndrome, a genetic disorder resulting in a spectrum of cognitive disabilities often accompanied by autistic symptoms. Following a previous analysis[3] of truncating *de novo* SNVs, we calculated the expected number of FMRP targets for ASD genes by inferring gene mutabilities from a large exome-sequencing study[30] (Online Methods); the expected number of FMRP targets was then compared with the observed number in various ASD gene sets (**Table 2**). This analysis revealed that, similar to truncating SNV genes, genes in the implicated network were significantly enriched for FMRP targets (truncating SNV gene enrichment 1:2.13, two-tail binomial test, $P = 3 \times 10^{-4}$; network gene enrichment 1:2.78, $P = 3 \times 10^{-11}$). The enrichment remained significant when considering only network genes harboring non-truncating SNVs (enrichment = 1:2.67, $P = 4 \times 10^{-9}$). Significant enrichment was also observed for FMRP targets from another recent study (**Supplementary Table 5**)[31]. In contrast, there was no significant enrichment for proband SNV genes that were not selected to the network ($P = 0.2$) nor for *de novo* SNV genes from unaffected siblings ($P = 0.15$). Furthermore, enrichment for FMRP targets remained significant when we separately considered each of the four functional clusters of the implicated network (**Table 2**). Overall, these analyses confirm that FMRP targets are important for autism etiology and suggest a causal role for a substantial fraction of genes with non-truncating *de-novo* SNVs.

Genes forming the PSD are also likely to be important in ASDs and other psychiatric disorders[12,32]. The PSD is localized at the postsynaptic membrane of excitatory synapses and is crucial to synaptic communication and plasticity[25,33,34]. There was a significant enrichment of PSD-associated genes in the implicated network (enrichment 1:3.4, two-tail binomial test, $P = 7 \times 10^{-9}$; **Supplementary Table 5**) and marginal enrichment for truncating *de novo* SNVs in probands (enrichment 1:1.9, $P = 0.05$). No enrichment was observed for genes harboring SNVs in unaffected siblings (enrichment 1:1.05, $P = 0.8$). For genes not selected to the network, PSD genes were significantly under-represented (enrichment 1:0.58, $P = 0.04$).

Temporal and spatial brain expression patterns of implicated genes

Gene expression patterns provide important clues for understanding physiological and developmental contexts of gene function. Thus, we next investigated brain expression of the ASD-implicated genes

Table 2 Overlap between ASD gene sets and FMRP targets

Gene set	Number of genes	Expected: observed	*P* value
Network genes	159	17.3:48 (1:2.78)	3×10^{-11}
Truncating SNV genes	108	11.7:25 (1:2.13)	0.0003
Network non-truncating SNV genes	138	15.0:40 (1:2.67)	4×10^{-9}
Neuronal signaling/cytoskeleton cluster genes	69	7.49:15 (1:2.00)	0.01
Chromatin modification/regulation cluster genes	50	5.43:17 (1:3.13)	1×10^{-5}
Postsynaptic density cluster genes	11	1.19:5 (1:4.19)	0.004
Channel activity cluster genes	21	2.28:9 (1:3.95)	0.0002
Non-network SNV genes	449	48.8:40 (1:0.82)	0.2
Sibling SNV genes	355	38.6:47 (1:1.22)	0.15

The expected and observed number of FMRP targets[27] among different sets of implicated genes. The expected numbers of FMRP targets were obtained on the basis of the proportion of rare synonymous variants observed in a recent large-scale survey of human genetic variation[30]. The observed and expected overlaps based on another recent study of FMRP targets[31] are given in **Supplementary Table 5**. The significances of the overlaps were established using a two-tail binomial test.

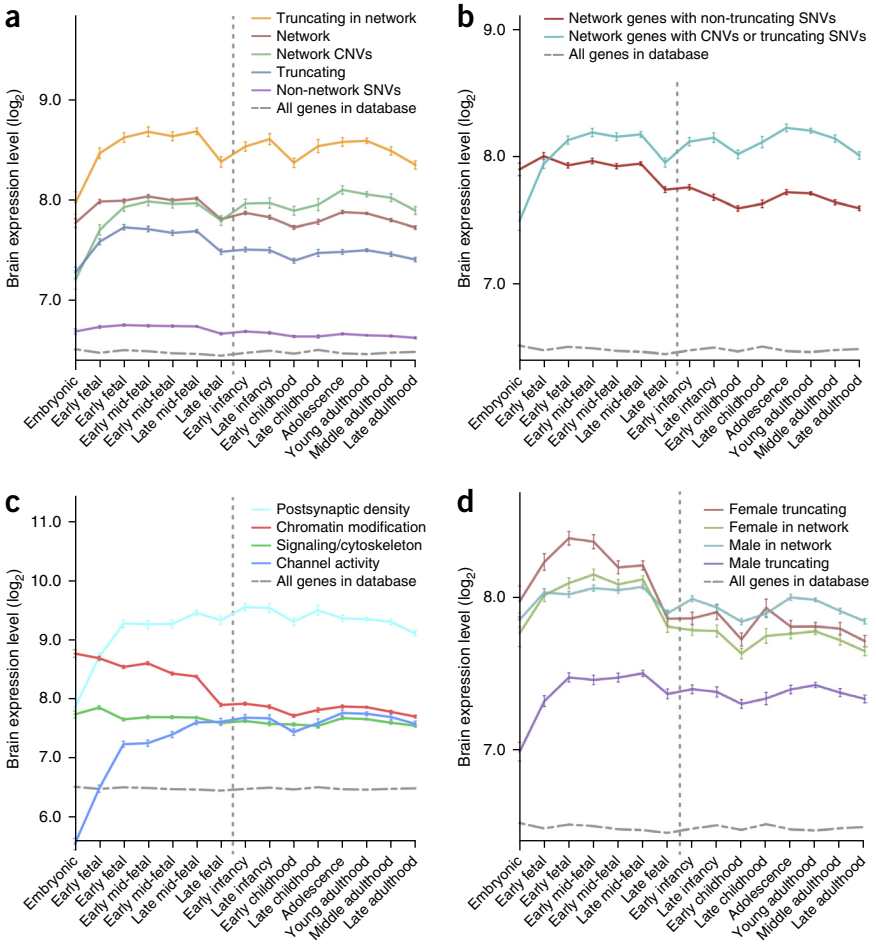

Figure 2 Temporal gene expression profiles in the human brain across developmental stages for implicated gene sets. Expression data (log_2) were obtained from the HBT database; average expression levels at each developmental stage were calculated using all genes in a given set. (**a**) Expression profiles for truncating SNV genes in the network (orange), all network genes (red), network genes from CNVs (green), all truncating SNV genes (blue) and non-network SNV genes (purple). (**b**) Expression profiles for network genes with truncating (cyan) and non-truncating (red) mutations observed in probands. (**c**) Expression profiles for the four functional clusters (**Fig. 1**) of the implicated network: postsynaptic density genes (cyan), chromatin modification/regulation genes (red), signaling/cytoskeleton genes (green) and channel activity genes (blue). (**d**) Expression profiles for network and truncating female/male SNV genes: female truncating SNV genes (red), female network genes (green), male network genes (blue) and male truncating SNV genes (purple). Vertical dashed lines separate prenatal and postnatal developmental stages. Error bars represent s.e.m.

using the Human Brain Transcriptome database (HBT, http://hbatlas.org/)[16]. Based on postmortem transcriptional analysis of tissue samples from healthy individuals, the HBT database provides a comprehensive map of mRNA expression across multiple brain regions and developmental stages. We determined the average expression levels across developmental periods for the various gene sets that formed the implicated network, as well as for genes with truncating *de novo* SNVs. We used a Wilcoxon rank-sum test (P_{WT}) to estimate the significance of the expression difference between sets of biological samples and a permutation-based test (P_{PT}) to estimate the probability of observing a greater bias in random probe sets of equal size (Online Methods). The brain expression of network genes (**Fig. 2a**) was significantly higher than the expression of SNV-containing genes that were not selected to the network ($P_{WT} < 10^{-15}$; $P_{PT} < 10^{-4}$). This result confirms that the implicated network is significantly enriched in brain-related genes. Notably, high brain expression of network genes is not simply a consequence of their length or connectivity in the phenotypic network; randomly selected human genes with equivalent length or network connectivity are expressed in the brain at significantly lower levels ($P < 10^{-4}$) across all developmental periods (**Supplementary Fig. 2**).

Expression of the network genes (**Fig. 2a**) was, on average, highest during the early fetal to early mid-fetal periods (8–19 post-conception weeks), possibly as a result of higher activity of ASD-related genes in brain cells or changes in brain cellular composition. To quantify the prenatal expression bias, we calculated the difference between the expression during prenatal and postnatal developmental periods for every gene harboring *de novo* mutations. The prenatal bias was highly significant for network genes (bias 0.16, one-tail test, $P_{WT} < 10^{-15}/P_{PT} < 10^{-4}$) and for genes with truncating SNVs (bias 0.18, $P_{WT} < 10^{-15}/P_{PT} < 10^{-4}$). However, for ASD network genes with truncating *de novo* SNVs or *de novo* CNVs, embryonic expression (less than 8 post-conception weeks) was substantially lower as compared with that in later developmental periods (two-tail test, $P_{WT} = 6 \times 10^{-7}/P_{PT} < 10^{-4}$ and $P_{WT} = 10^{-14}/P_{PT} < 10^{-4}$, respectively; **Fig. 2a**). This effect is apparent when comparing the average expression profiles for network genes with truncating (CNVs and truncating SNVs) and non-truncating mutations (**Fig. 2b**). The bias against mutations in the embryonic period was absent for non-synonymous *de novo* SNV genes not selected to the network ($P_{WT} = 0.4/P_{PT} = 0.3$) and for genes with *de novo* SNVs in siblings ($P_{WT} = 0.2/P_{PT} = 0.2$). The observed bias against *de novo* truncating SNVs and CNVs suggests that there is strong selection against harmful mutations during embryonic development, as these mutations lead to more severe developmental consequences than mutations typically associated with autism.

We also examined the temporal expression profiles for each of the four clusters forming the implicated network (**Fig. 2c**). The cluster that was primarily associated with the postsynaptic density (**Fig. 2c**) had the highest overall expression in the brain. Both the PSD cluster and the cluster associated with various channel activities showed a distinct rise during early fetal development, consistent with the start of synaptogenesis at the early mid-fetal developmental stage. In contrast, the cluster associated with chromatin modification and regulation showed high embryonic and fetal expression, consistent with a developmental peak of neuronal proliferation and differentiation, and then gradually decreased to a postnatal plateau. Sustained expression of chromatin modification genes after neurodevelopmental stages may be attributed to their involvement in synaptic plasticity[35]. Finally, the cluster associated with neuronal signaling and cytoskeleton was relatively constant across all developmental stages. This is consistent with the important roles of signaling and structural genes across

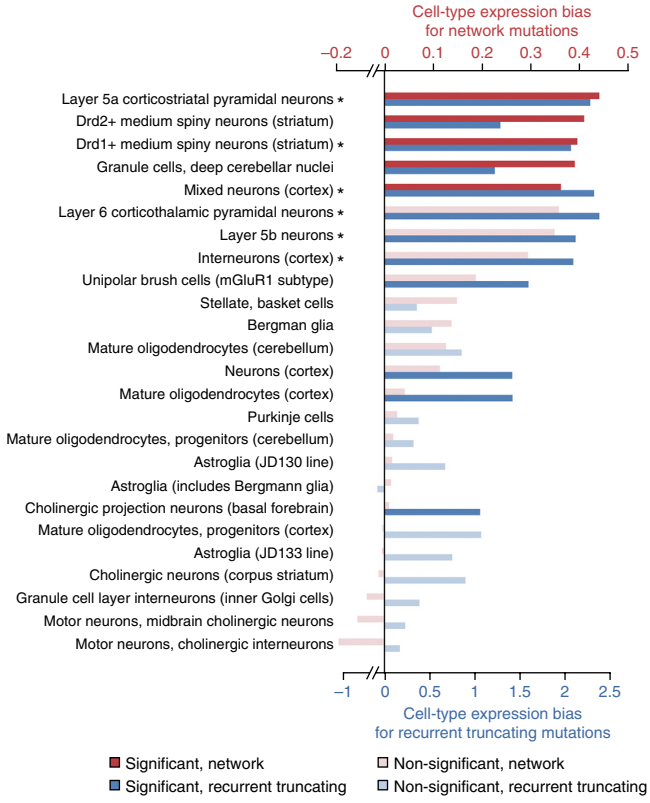

Figure 3 Cell-type expression biases for network mutations and recurrent truncating mutations. Proband versus unaffected sibling expression biases were computed across 25 cell types of the CNS for implicated network genes (shown in red) and for 11 genes with recurrent truncating SNVs (blue). The biases were calculated using previously described *Mus musculus* expression data[17]. To quantify the expression biases, we calculated for each cell type the difference between the average $\log_2$ expression of mouse orthologs for implicated human genes and the average $\log_2$ expression of mouse ortholog for human genes with *de novo* SNVs in unaffected siblings. Cell types are ordered by the magnitude of the cell type expression bias for network genes (red). The significance of the expression biases was evaluated using the Wilcoxon rank-sum one-tail test and corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure with a false discovery rate of 10%; significant cell types are shown in dark red and blue, and non-significant types are shown in light red and blue. In addition, the *P* values obtained from the two independent approaches, one based on network genes (red) and the other based on genes affected by recurrent truncating mutations (blue), were combined using Fisher's and Stouffer's meta-analysis methods. The combined *P* values were corrected using the Bonferroni method, and the cell types passing the significance cutoff for both meta-analyses are indicated with an asterisk (*). All *P* values associated with the figure are presented in **Supplementary Table 8**.

all developmental stages. Notably, there was substantial variability in the expression of the individual genes around the average cluster expression profiles (**Supplementary Fig. 3**). However, this variability is likely to be at least partially explained by different dosage requirements of individual genes. Normalized expression trajectories (**Supplementary Fig. 3**) revealed that the temporal profiles of individual genes were generally consistent with the average expression profiles of corresponding functional clusters.

A high male-to-female incidence ratio, estimated at more than 4:1 for high-functioning individuals, is one of the most consistent and notable findings in ASDs[36,37]. Evidence suggests that this bias may be a result of a female protective effect that requires a higher threshold of genetic insults to trigger ASDs in females than in males[38]. We also observed that females in the SSC collection had a greater burden of truncating mutations than males (Fisher's exact one-tail test, $P = 0.07$), which is consistent with this hypothesis. This gender dimorphism was even stronger for the genes in the implicated network (30% of SNV mutations in females are truncating compared with 13% in males; Fisher's exact one-tail test, $P = 0.03$). Notably, the average brain expression was significantly higher for genes harboring truncating *de novo* mutations in females than in males (average $\log_2$ expression for genes: females, 7.98; males, 7.40; one-tail test, $P_{WT} < 10^{-15}/P_{PT} = 2 \times 10^{-3}$; **Fig. 2d**). These log-scale differences translate to absolute expression level differences of 50–100% across developmental stages in the average expression of genes harboring truncating SNVs in females and males. Thus, the expression data also suggest that relatively stronger genetic perturbations, that is, truncating mutations in genes with higher brain expression, are preferentially associated with ASDs in females. Notably, the observed expression patterns cannot be explained by normal expression differences between males and females. For genes with truncating SNV mutations in female probands the average brain expression levels in females and males were 8.02 and 7.95 (<5% difference), respectively; for genes with truncating SNVs in male probands the average brain expression levels in females and males were 7.42 and 7.39 (<3% difference), respectively. We also note that the relative difference in brain expression for genes harboring mutations in females versus males was larger for genes with truncating SNVs than for implicated network genes. This result is likely a consequence of the severe effect of truncating mutations on protein function, which makes the expression of corresponding genes an important factor in explaining the variability of ASD phenotypes across patients.

To investigate whether ASD mutations preferentially affect specific brain regions, we analyzed spatial expression data available in the HBT database. For each brain region, we compared the average expression levels of network genes to the expression levels of genes with SNVs in unaffected siblings. Compared with controls, network genes showed significantly higher expression across all brain regions (**Supplementary Table 7**), which is consistent with the wide spectrum of phenotypic abnormalities observed in ASDs. Furthermore, we investigated the

Table 3 Prenatal versus postnatal brain expression biases across brain regions

	Network genes			Truncating SNVs			Sibling SNVs		
Brain region	Bias	S.e.m.	*P* value	Bias	S.e.m.	*P* value	Bias	S.e.m.	*P* value
Amygdala	0.25	0.08	1×10^{-12}	0.26	0.10	2×10^{-6}	0.02	0.05	1.0
Striatum/BG	0.24	0.08	2×10^{-12}	0.22	0.10	3×10^{-5}	0.04	0.05	1.0
Cerebellum	0.23	0.08	3×10^{-9}	0.17	0.10	4×10^{-3}	0.07	0.05	0.1
Occipital	0.17	0.08	3×10^{-7}	0.15	0.10	3×10^{-2}	0.04	0.05	1.0
Parietal	0.16	0.08	3×10^{-11}	0.17	0.10	3×10^{-4}	0.05	0.05	0.7
Temporal	0.16	0.07	2×10^{-17}	0.19	0.09	2×10^{-8}	0.05	0.05	1.0
Hippocampus	0.15	0.07	1×10^{-5}	0.18	0.09	4×10^{-4}	0.02	0.05	1.0
Thalamus	0.14	0.08	1×10^{-4}	0.22	0.09	1×10^{-5}	0.03	0.05	1.0
Frontal	0.12	0.08	1×10^{-16}	0.17	0.09	4×10^{-10}	0.03	0.05	1.0

The biases were calculated using human expression data obtained from the HBT database. To quantify the expression bias for each brain region, we calculated the difference between the average $\log_2$ prenatal expression and the average $\log_2$ postnatal expression, such that positive values in the table indicate higher expression levels in the prenatal periods. The significances of the expression biases were evaluated using the Wilcoxon rank-sum one-tail test and corrected using the Bonferroni procedure. Bias s.e.m. was calculated separately for each brain region. BG, basal ganglia.

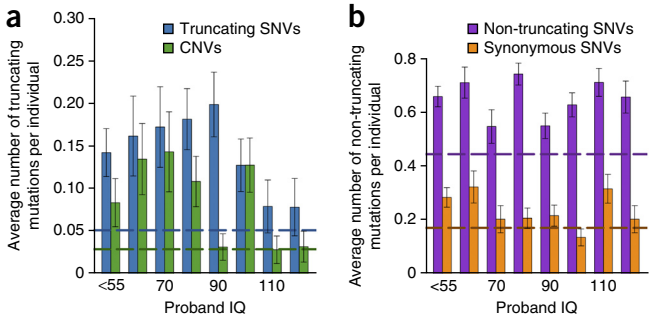

Figure 4 Average numbers of *de novo* mutations per individual for probands with different IQs. (**a**) The average number of truncating SNVs (blue) and CNVs (green) per individual. (**b**) The average number of non-truncating SNVs (purple) and synonymous SNVs (orange) per individual. Horizontal dashed lines represent the corresponding average numbers of mutations per individual for all unaffected siblings. Error bars represent s.e.m.

prenatal expression bias for network and truncating genes and compared these with the prenatal bias for genes with SNVs in siblings. We found that the prenatal bias was generally similar and statistically significant for all brain regions (**Table 3**). We note, however, that the amygdala, which is known to have a crucial role in processing emotional stimuli[39], as well as cerebellum and striatum, showed a numerically, although not significantly, larger prenatal bias than other regions. Overall, however, the lack of regional specificity underscores the sharing of gene functions across brain areas.

The human brain contains a variety of neuronal and non-neuronal cell types, and we sought to investigate possible biases of implicated genes toward specific cell types. To explore this question, we used an independent data set of cell-specific gene expression generated using translating ribosome affinity purification (TRAP)[17]; the data set contains gene expression profiles for 25 distinct cell types from the mouse CNS. To assess cell-specific expression biases, we compared, for each cell type, the expression of mouse orthologs for the implicated network genes with the expression of orthologs for genes harboring mutations in unaffected siblings. This analysis revealed that multiple neuronal and non-neuronal cell types were likely affected by *de novo* ASD mutations in the network genes (**Fig. 3** and **Supplementary Table 8**). The diversity of affected cells could be explained, at least partially, by shared usage of common signaling, structural and neural pathways across diverse cell types. Although multiple cell types were affected, particularly strong and significant expression biases for implicated network genes were observed in cortical neurons, especially cortical interneurons, pyramidal neurons and medium spiny neurons of the striatum (**Fig. 3**). Notably, similar cell types were also independently implicated by considering expression biases for 11 genes with recurrent truncating *de novo* mutations (**Fig. 3**). On the other hand, some other cell types, such as motor neurons and astroglial cells, were markedly less affected. Notably, deep layer cortical glutamatergic projection neurons were also identified recently as a point of spatio-temporal convergence in coexpression networks built around high confidence ASD genes[40].

Functional properties of implicated genes and disease phenotypes

Because ASDs manifest substantial pathophysiological differences across probands, it is important to understand how functional properties of implicated genes and associated mutations affect phenotypic characteristics of the disease. To investigate the effect of truncating *de novo* mutations on IQ, we calculated, separately for CNVs and truncating SNVs, the average number of truncating mutations per individual across the IQ spectrum (**Fig. 4a**). Notably, for high-functioning ASD probands, the average number of truncating mutations (CNVs and truncating SNVs) decreased and became similar to the average number in unaffected siblings. The average number of truncating SNVs for probands with IQ less than 100 was about twofold higher than probands with IQ greater than or equal to 100 (0.17 for IQ < 100, 0.09 for IQ ≥ 100; Fisher's exact two-tail test, $P = 0.02$). In contrast, similar analyses for non-truncating mutations (synonymous and non-truncating SNVs) in probands showed that the average number of non-truncating mutations per individual was

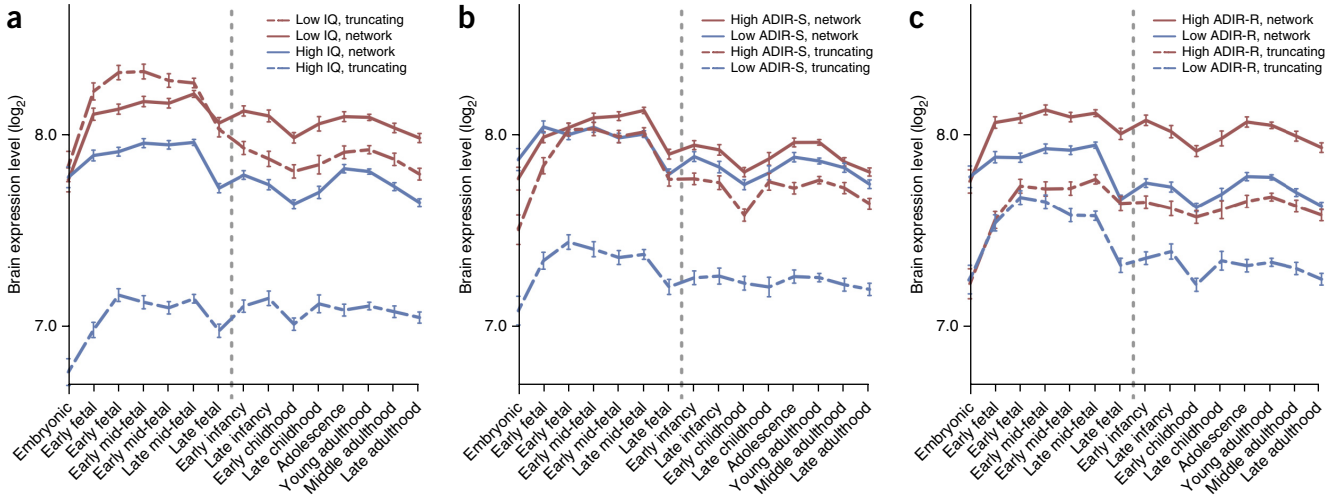

Figure 5 Temporal gene expression profiles in the human brain across developmental stages for genes affected in subsets of probands with different phenotypic scores. Probands were divided into two groups (high and low) relative to the corresponding median phenotypic scores. Expression data ($\log_2$) were obtained from the HBT database; average expression levels at each developmental stage were calculated using all genes in a given subset and error bars represent s.e.m. Expression profiles for network genes are displayed as solid lines and profiles for truncating SNVs as dashed lines. Profiles for genes affected in probands with more severe phenotypes are shown in red and less severe phenotypes in blue. Vertical dashed lines separate prenatal and postnatal developmental stages. (**a**) Profiles for probands with high and low IQ. (**b**) Profiles for probands with low and high ADIR-S scores. (**c**) Profiles for probands with low and high ADIR-R scores.

relatively constant across IQs (**Fig. 4b**), without a prominent decrease for high-functioning ASD probands (0.43 for IQ < 100, 0.46 for IQ ≥ 100; Fisher's exact two-tail test, $P = 0.7$). Relative to synonymous mutations, non-synonymous non-truncating were significantly enriched in probands with IQs above 70 compared with those below 70: 160 non-truncating and 72 synonymous mutations (2.2:1 ratio) were observed in probands with IQ ≤ 70, whereas 330 non-truncating and 110 synonymous mutations (3.3:1 ratio) in probands with IQ > 70 (Fisher's exact one-tail test, $P = 0.04$). Overall, these analyses suggest that truncating mutations are likely to play a less prominent role in high-functioning ASD cases. Another recent study[41] also demonstrated no excess of *de novo* loss-of-function mutations for high-functioning ASD probands.

Notably, there was a significant difference in IQs between probands with network genes affected by CNV deletions and duplications. The average IQs for probands with CNV deletions and duplications were 64.8 and 83.9, respectively (Wilcoxon rank-sum one-tail test, $P = 6 \times 10^{-3}$). This result suggests that a decrease in gene dosage has a substantially stronger functional effect than an increase in dosage.

Given that genes exhibited diverse expression patterns, we asked whether the level of brain expression for affected genes is associated, on average, with different phenotypic outcomes. For this analysis, we considered the full-scale IQ and Autism Diagnostic Interview-Revised (ADIR) social interaction and repetitive behavior scores. The ADIR scores are based on structured interviews with proband parents and reflect patterns in reciprocal social interactions (ADIR-S) and repetitive/restrictive behaviors (ADIR-R)[42]. To explore the relationship between the average expression level of affected genes and corresponding phenotypes, we divided the ASD cases into low- and high-scoring phenotypic subsets relative to the corresponding median phenotype scores. We then calculated the average expression levels of implicated genes identified in each phenotype subset (**Fig. 5**). This analysis revealed that affected genes associated with lower IQ scores or higher ADIR scores (both indicating more severe phenotypes) usually had significantly higher brain expression (network genes: one-tail test, $P_{WT} < 10^{-15}/P_{PT} = 0.01$ for IQ, $P_{WT} < 10^{-15}/P_{PT} = 0.3$ for ADIR-S, $P_{WT} < 10^{-15}/P_{PT} = 0.024$ for ADIR-R; genes with truncating mutations: one-tail test, $P_{WT} < 10^{-15}/P_{PT} < 10^{-4}$ for IQ, $P_{WT} < 10^{-15}/P_{PT} = 1.5 \times 10^{-3}$ for ADIR-S, $P_{WT} < 10^{-15}/P_{PT} = 0.08$ for ADIR-R; **Fig. 5**).

DISCUSSION

Our results suggest that the pathophysiological heterogeneity of ASD is matched by the diversity of genetic and functional insults associated with the disorder. We found that affected genes had diverse developmental expression profiles and are therefore likely to be important in multiple stages of neurogenesis, neuron mobility, synaptogenesis and brain function. Although the implicated genes and processes are active across multiple cell types, some types, such as cortical interneurons, pyramidal neurons and medium spiny neuron of the striatum, seemed to be more strongly affected. Notably, layer 5 cortical pyramidal neurons often project to the striatum and layer 6 often project to the thalamus; the corresponding cortico-striatal-thalamic circuits are known to mediate diverse motor, emotional, cognitive and habit-forming behaviors that are often perturbed in ASD[43–46]. Consequently, our unbiased genome-wide analysis implicates specific functional neural circuits that may mediate stereotypical and repetitive behaviors in ASD.

Although previous studies have primarily emphasized the role of truncating *de novo* mutations in ASD[3,7,8], our analysis suggests that a substantial fraction of non-truncating *de novo* missense mutations observed in probands also contributes to the disorder. Notably, we found that functional mutations were preferentially observed in haploinsufficient genes, which confirms that dosage effects are important for the disease mechanisms. We found that functional characteristics of affected genes, such as brain expression levels, are likely to influence the observed phenotypic consequences of *de novo* ASD mutations. Stronger functional insults lead, on average, to more severe ASD phenotypes. Thus, the distinction between intellectual disability and autism may lie primarily in the degree of overall functional effect rather than specific genes and pathways affected in the two disorders.

Our analysis of brain expression provides further evidence for the hypothesis that stronger functional insults, such as perturbation of genes with substantially higher brain expression, are associated with female autistic phenotypes. Stronger functional perturbations in females were previously demonstrated through analyses of autism CNVs sizes[6,47] and gene network properties[13]. Thus, multiple independent sources of evidence suggest a protective effect in females, although the mechanisms of this effect remain to be elucidated.

We also find, in agreement with recently published studies[41,48], that truncating mutations play a smaller role in high-functioning ASD cases. Because truncating mutations are usually associated with a loss of function, this result suggests that high-functioning autism phenotypes are less likely to be mediated by a loss of normal gene function in the brain. Functional gain and other types of genetic variations, such as non-truncating *de novo* mutations, common polymorphisms or mutations in non-coding regulatory regions, may be the primary contributors to high-functioning ASD cases.

Taken together, our results suggest that various functional properties of mutations and target genes may be useful in predicting ASD phenotypic severity. In the future, it will be important to investigate the extent to which individual ASD patients can be stratified on the basis of affected pathways, biological functions and cell types. Such patient stratification, now a common practice in cancer, may lead to individualized diagnostic and prognostic predictions and, ultimately, to targeted ASD therapies.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to sincerely thank M. Wigler, M. State, D. Geschwind, A. Packer, G. Fischbach and all of the members of the Vitkup laboratory for helpful discussions. This work was supported by a grant from the Simons Foundation SFARI# 308962 to D.V. and US National Centers for Biomedical Computing (MAGNet) grant U54CA121852 to Columbia University. S.R.G. was supported in part by US NIGMS training grant T32 GM082797. S.J.S. was supported by a Howard Hughes Medical Institute International Student Research Fellowship.

AUTHOR CONTRIBUTIONS

J.C., S.R.G. and A.H.C. performed computational analysis and interpreted the results. S.J.S. contributed data, interpreted the results and contributed to the functional analysis. D.V. designed the study, supervised the project and interpreted the results. J.C., S.R.G., A.H.C. and D.V. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Abrahams, B.S. & Geschwind, D.H. Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* **9**, 341–355 (2008).
2. Berg, J.M. & Geschwind, D.H. Autism genetics: searching for specificity and convergence. *Genome Biol.* **13**, 247 (2012).

3. Iossifov, I. *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
4. Freitag, C.M. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry* **12**, 2–22 (2007).
5. Geschwind, D.H. Genetics of autism spectrum disorders. *Trends Cogn. Sci.* **15**, 409–416 (2011).
6. Levy, D. *et al.* Rare *de novo* and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**, 886–897 (2011).
7. O'Roak, B.J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
8. Sanders, S.J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
9. Veltman, J.A. & Brunner, H.G. *De novo* mutations in human genetic disease. *Nat. Rev. Genet.* **13**, 565–575 (2012).
10. McClellan, J. & King, M.C. Genetic heterogeneity in human disease. *Cell* **141**, 210–217 (2010).
11. O'Roak, B.J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
12. Gilman, S.R. *et al.* Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat. Neurosci.* **15**, 1723–1728 (2012).
13. Gilman, S.R. *et al.* 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 (2011).
14. Kelleher, R.J. III & Bear, M.F. The autistic neuron: troubled translation? *Cell* **135**, 401–406 (2008).
15. Zoghbi, H.Y. & Bear, M.F. Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. *Cold Spring Harb. Perspect. Biol.* **4**, a009886 (2012).
16. Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
17. Doyle, J.P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).
18. Fischbach, G.D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).
19. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
20. Krey, J.F. & Dolmetsch, R.E. Molecular mechanisms of autism: a possible role for Ca^{2+} signaling. *Curr. Opin. Neurobiol.* **17**, 112–119 (2007).
21. Betancur, C., Sakurai, T. & Buxbaum, J.D. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci.* **32**, 402–412 (2009).
22. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
23. Ronan, J.L., Wu, W. & Crabtree, G.R. From neural development to cognition: unexpected roles for chromatin. *Nat. Rev. Genet.* **14**, 347–359 (2013).
24. Santini, E. *et al.* Exaggerated translation causes synaptic and behavioural aberrations associated with autism. *Nature* **493**, 411–415 (2013).
25. Ziff, E.B. Enlightening the postsynaptic density. *Neuron* **19**, 1163–1174 (1997).
26. Huang, N., Lee, I., Marcotte, E.M. & Hurles, M.E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6**, e1001154 (2010).
27. Darnell, J.C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
28. Huber, K.M., Gallagher, S.M., Warren, S.T. & Bear, M.F. Altered synaptic plasticity in a mouse model of fragile X mental retardation. *Proc. Natl. Acad. Sci. USA* **99**, 7746–7750 (2002).
29. Edbauer, D. *et al.* Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. *Neuron* **65**, 373–384 (2010).
30. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
31. Ascano, M. Jr. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382–386 (2012).
32. Bayés, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14**, 19–21 (2011).
33. Kennedy, M.B. The postsynaptic density at glutamatergic synapses. *Trends Neurosci.* **20**, 264–268 (1997).
34. Kennedy, M.B. Signal-processing machines at the postsynaptic density. *Science* **290**, 750–754 (2000).
35. Vogel-Ciernia, A. *et al.* The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nat. Neurosci.* **16**, 552–561 (2013).
36. Newschaffer, C.J. *et al.* The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health* **28**, 235–258 (2007).
37. Fombonne, E. Epidemiology of pervasive developmental disorders. *Pediatr. Res.* **65**, 591–598 (2009).
38. Zhao, X. *et al.* A unified genetic theory for sporadic and inherited autism. *Proc. Natl. Acad. Sci. USA* **104**, 12831–12836 (2007).
39. Adolphs, R. What does the amygdala contribute to social cognition? *Ann. NY Acad. Sci.* **1191**, 42–61 (2010).
40. Willsey, A.J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
41. Samocha, K.E. *et al.* A framework for the interpretation of *de novo* mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
42. Lord, C., Rutter, M. & Le Couteur, A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **24**, 659–685 (1994).
43. Shepherd, G.M. Corticostriatal connectivity and its role in disease. *Nat. Rev. Neurosci.* **14**, 278–291 (2013).
44. Langen, M., Durston, S., Kas, M.J., van Engeland, H. & Staal, W.G. The neurobiology of repetitive behavior: … and men. *Neurosci. Biobehav. Rev.* **35**, 356–365 (2011).
45. Burguière, E., Monteiro, P., Feng, G. & Graybiel, A.M. Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. *Science* **340**, 1243–1246 (2013).
46. Rothwell, P.E. *et al.* Autism-associated neuroligin-3 mutations commonly impair striatal circuits to boost repetitive behaviors. *Cell* **158**, 198–212 (2014).
47. Sanders, S.J. *et al.* Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).
48. Iossifov, I. *et al.* The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

ONLINE METHODS

ASD associated *de novo* variants. Variants were obtained from several recent studies of families in the SSC and included *de novo* CNVs[6] and *de novo* SNVs[3,7,8]. Overlapping CNVs were combined into a single event and CNVs larger than 5 Mb were ignored. When a gene affected by an SNV was also contained in a CNV event, the SNV gene was considered individually and the remaining genes in the CNV were considered as a separate event. Combining overlaps and removing duplicate genes resulted in 991 genes at 624 distinct genomic loci that were used as input in our analysis. 11 probands considered in our study had multiple SNV mutations; these probands were counted multiple times (once for each SNV) in various tests comparing mutations and corresponding functional properties. GERP scores used to assess the evolutionary severity of SNV mutations were obtained from the original studies[49].

Phenotypic network and the NETBAG+ algorithm. The NETBAG+ algorithm relies on the previously described phenotypic network in which all pairs of human genes are assigned a score proportional to the likelihood ratio that genes contribute to the same genetic phenotype[12,13]. This ratio was calculated using a naive Bayesian integration of various descriptors of protein function: shared GO annotations, shared pathways in Kyoto Encyclopedia of Genes and Genomes (KEGG), protein domains from the InterPro database, tissue expression from the TiGER database, direct protein-protein interactions, shared interaction partners in a number of databases (BIND, BioGRID, DIP, HPRD, InNetDB, IntAct, BiGG, MINT and MIPS), phylogenetic profiles and chromosomal co-clustering across genomes[50]. The likelihood network was constructed using a carefully curated set of human genes compiled by a previous study[51] that contains 476 human genes associated with 132 different genetic phenotypes.

The 991 genes affected by *de novo* mutations were mapped to the corresponding nodes of the phenotypic network. A greedy search algorithm was then used to find strongly interconnected networks among the input genes; starting from each input gene, the greedy algorithm consecutively adds the genes most strongly connected to the growing network. Networks are scored based on a weighted sum of all pairwise likelihood scores between the network genes[12,13]. Each CNV event was constrained to contribute at most one gene to the network. Small networks with five or fewer genes were ignored during the network searches.

Network significance was determined by applying the same greedy search algorithm to 5,000 random gene sets and comparing the score of the network obtained using real input data to the distribution of network scores obtained using random gene sets. The random gene sets used in the calculations were chosen to match the input genes in terms of network connectivity (based on the average of the top five edge scores) and protein lengths to ensure that network significance was not driven by highly connected or long genes; genes affected by *de novo* ASD mutations were allowed to participate in the randomization sets. The final network *P* value reflects the probability of finding a higher scoring network using random gene sets while also accounting for multiple hypothesis testing as a result of various network sizes[12,13]. The final network of 159 genes was selected as the largest network before a considerable decrease in network significance.

FMRP and PSD enrichment among implicated gene sets. Following a previously described approach[3] and other similar studies, we estimated the enrichments using the overlap between FMRP targets[27,31] or PSD genes[32] and a set of rare synonymous mutations from a large exome sequencing study; we used for this purpose the NHLBI ESP exome-sequencing study[30]. Using the exome-sequencing data of about 6,500 individuals, we calculated the fractions of all rare synonymous mutations, that is, synonymous mutations observed in the study only once, that affect FMRP target genes and genes associated with PSD. NHLBI ESP variant data was downloaded from http://evs.gs.washington.edu and unique entries were determined using the variant chromosomal position and allele. The calculated fractions represent the background (expected) probabilities for a random SNV mutation to occur in FMRP and PSD genes. We then compared the expected and the observed fractions of SNVs events in various ASD disease gene sets using a two-tail binomial test (**Table 2** and **Supplementary Table 5**). We also confirmed the FMRP enrichment results by comparing the overlap of FMRP target genes with various ASD gene sets to the overlap of FMRP target genes with random sets of human genes matched by protein length and network connectivity to the genes affected by ASD mutations (**Supplementary Table 6**).

Hierarchical clustering of the implicated network. We used the average linkage hierarchical clustering to divide the implicated network into functional clusters (**Supplementary Fig. 1**). The inverses of the phenotypic network likelihood scores were used as the clustering metric.

Spatial and temporal analysis of human brain expression. Expression data across developmental stages and brain regions was obtained from the Human Brain Transcriptome database (GEO accession ID, GSE25219)[16]; the HBT data represents quantile-normalized and $\log_2$-transformed expressions values from the post-mortem brain samples from healthy individuals.

To quantify the expression biases for each brain region, we calculated the difference between the average $\log_2$ expression of implicated genes and the average $\log_2$ expression of genes harboring SNVs in siblings. The prenatal expression bias for each brain region was quantified by calculating the difference between the average $\log_2$ of prenatal expression (embryonic to late fetal stages) and the average $\log_2$ postnatal expression (early infancy to late adulthood stages). The significance of regional and prenatal biases was evaluated using the Wilcoxon rank-sum test with a Bonferroni correction to account for multiple hypothesis testing.

Complementary to the Wilcoxon rank-sum test (P_{WT}), we also estimated the significance of various expression differences using a method based on randomly generated expression probe sets (**Supplementary Table 9**). The permutation test *P* values (P_{PT}) obtained with the random trials reflect the probability to obtain a bias greater or equal to the bias in the original data. The numbers of probe sets sampled in each of 10,000 random trials were equal to the numbers of probes sets in corresponding original gene sets. Analogous to the procedure used to process the original HBT data, we computed transcript-level expression by taking the median expression value of probe sets; expression biases were then calculated by taking the difference between the corresponding expression averages.

Cell type–specific expression analysis. To evaluate expression bias of the implicated genes toward specific cell types, we considered mouse expression data obtained previously[17] using ribosome affinity purification (GEO accession ID GSE13379). The considered data set contains expression information for 25 different cell types from the *Mus musculus* CNS. To connect the mouse expression data and the human genetic data, we used the list of human-mouse orthologs provided by Affymetrix for the Mouse Genome 430 Array. The expression biases in probands compared to siblings were quantified for each cell type by calculating the difference between the average $\log_2$ expression of mouse orthologs for implicated human genes and the average $\log_2$ expression of mouse ortholog for human genes with *de novo* SNVs in unaffected siblings. We evaluated significances of the expression biases using the Wilcoxon rank-sum one-tail test and corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure. *P* values obtained from the two independent approaches, one based on network genes and the other based on genes affected by recurrent truncating mutations, were combined using Fisher's and Stouffer's meta-analysis methods. The combined *P* values were corrected using the Bonferroni method.

A **Supplementary Methods Checklist** is available.

49. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
50. Chen, L. & Vitkup, D. Predicting genes for orphan metabolic activities using phylogenetic profiles. *Genome Biol.* **7**, R17 (2006).
51. Feldman, I., Rzhetsky, A. & Vitkup, D. Network properties of genes harboring inherited disease mutations. *Proc. Natl. Acad. Sci. USA* **105**, 4323–4328 (2008).