ARTICLE

doi:10.1038/nature13772

Synaptic, transcriptional and chromatin genes disrupted in autism

A list of authors and their affiliations appears at the end of the paper

The genetic architecture of autism spectrum disorder involves the interplay of common and rare variants and their impact on hundreds of genes. Using exome sequencing, here we show that analysis of rare coding variation in 3,871 autism cases and 9,937 ancestry-matched or parental controls implicates 22 autosomal genes at a false discovery rate (FDR) < 0.05, plus a set of 107 autosomal genes strongly enriched for those likely to affect risk (FDR < 0.30). These 107 genes, which show unusual evolutionary constraint against mutations, incur *de novo* loss-of-function mutations in over 5% of autistic subjects. Many of the genes implicated encode proteins for synaptic formation, transcriptional regulation and chromatin-remodelling pathways. These include voltage-gated ion channels regulating the propagation of action potentials, pacemaking and excitability–transcription coupling, as well as histone-modifying enzymes and chromatin remodellers—most prominently those that mediate post-translational lysine methylation/demethylation modifications of histones.

Features of subjects with autism spectrum disorder (ASD) include compromised social communication and interaction. Because the bulk of risk arises from *de novo* and inherited genetic variation[1–10], characterizing which genes are involved informs ASD neurobiology and reveals part of what makes us social beings.

Whole-exome sequencing (WES) studies have proved fruitful in uncovering risk-conferring variation, especially by enumerating *de novo* variation, which is sufficiently rare that recurrent mutations in a gene provide strong evidence for a causal link to ASD. *De novo* loss-of-function (LoF) single-nucleotide variants (SNVs) or insertion/deletion (indel) variants[11–15] are found in 6.7% more ASD subjects than in matched controls and implicate nine genes from the first 1,000 ASD subjects analysed[11–16]. Moreover, because there are hundreds of genes involved in ASD risk, ongoing WES studies should identify additional ASD genes as an almost linear function of increasing sample size[11].

Here we conduct the largest ASD WES study so far, analysing 16 sample sets comprising 15,480 DNA samples (Supplementary Table 1 and Extended Data Fig. 1). Unlike earlier WES studies, we do not rely solely on counting *de novo* LoF variants, rather we use novel statistical methods to assess association for autosomal genes by integrating *de novo*, inherited and case-control LoF counts, as well as *de novo* missense variants predicted to be damaging. For many samples original data from sequencing performed on Illumina HiSeq 2000 systems were used to call SNVs and indels in a single large batch using GATK (v2.6)[17]. *De novo* mutations were called using enhancements of earlier methods[14] (Supplementary Information), with calls validating at extremely high rates.

After evaluation of data quality, high-quality alternative alleles with a frequency of <0.1% were identified, restricted to LoF (frameshifts, stop gains, donor/acceptor splice site mutations) or probably damaging missense (Mis3) variants (defined by PolyPhen-2 (ref. 18)). Variants were classified by type (*de novo*, case, control, transmitted, non-transmitted) and severity (LoF, Mis3), and counts tallied for each gene.

Some 13.8% of the 2,270 ASD trios (two parents and one affected child) carried a *de novo* LoF mutation—significantly in excess of both the expected value[19] (8.6%, $P < 10^{-14}$) and what was observed in 510 control trios (7.1%, $P = 1.6 \times 10^{-5}$) collected here and previously published[15]. Eighteen genes (Table 1) exhibited two or more *de novo* LoF mutations. These genes are all known or strong candidate ASD genes, but given the number of trios sequenced and gene mutability[14,19], we would expect to observe this in approximately two such genes by chance. While we expect only two *de novo* Mis3 events in these 18 genes, we observe 16 ($P = 9.2 \times 10^{-11}$, Poisson test). Because most of our data exist in cases and controls and because we observed an additional excess of transmitted LoF events in the 18 genes, it is evident that the optimal analytical framework must involve an integration of *de novo* mutation with variants observed in cases and controls and transmitted or untransmitted from carrier parents. Investigating beyond *de novo* LoFs is also critical given that many ASD risk genes and loci have mutations that are not completely penetrant.

Transmission and *de novo* association

We adopted TADA (transmission and *de novo* association), a weighted, statistical model integrating *de novo*, transmitted and case-control variation[20]. TADA uses a Bayesian gene-based likelihood model including per-gene mutation rates, allele frequencies, and relative risks of particular classes of sequence changes. We modelled both LoF and Mis3 sequence variants. Because no aggregate association signal was detected for inherited Mis3 variants, they were not included in the analysis. For each gene, variants of each class were assigned the same effect on relative risk. Using a prior probability distribution of relative risk across genes for each class of variants, the model effectively weighted different classes of variants in this order: *de novo* LoF > *de novo* Mis3 > transmitted LoF, and allowed for a distribution of relative risks across genes for each class. The strength of association was assimilated across classes to produce a gene-level Bayes factor with a corresponding FDR *q* value. This framework increases the power compared to the use of *de novo* LoF variants alone (Extended Data Fig. 2).

TADA identified 33 autosomal genes with an FDR < 0.1 (Table 1) and 107 with an FDR < 0.3 (Supplementary Tables 2 and 3 and Extended Data Fig. 3). Of the 33 genes, 15 (45.5%) are known ASD risk genes[9]; 11 have been reported previously with mutations in ASD patients but were not classed as true risk genes owing to insufficient evidence (*SUV420H1* (refs 11, 15), *ADNP*[12], *BCL11A*[15], *CACNA2D3* (refs 15, 21), *CTTNBP2* (ref. 15), *GABRB3* (ref. 21), *CDC42BPB*[13], *APH1A*[14], *NR3C2* (ref. 15), *SETD5* (refs 14, 22) and *TRIO*[11]) and 7 are completely novel (*ASH1L*, *MLL3* (also known as *KMT2C*), *ETFB*, *NAA15*, *MYO9B*, *MIB1* and *VIL1*). *ADNP* mutations have recently been identified in 10 patients with ASD and other shared clinical features[23]. Two of the newly discovered genes,

Table 1 | ASD risk genes

dnLoF count	FDR ≤ 0.01	0.01 < FDR ≤ 0.05	0.05 < FDR ≤ 0.1
≥2	*ADNP, ANK2, ARID1B, CHD8, CUL3, DYRK1A, GRIN2B, KATNAL2, POGZ, SCN2A, SUV420H1, SYNGAP1, TBR1*	*ASXL3, BCL11A, CACNA2D3, MLL3*	*ASH1L*
1		*CTTNBP2, GABRB3, PTEN, RELN*	*APH1A, CD42BPB, ETFB, NAA15, MYO9B, MYT1L, NR3C2, SETD5, TRIO*
0		*MIB1*	*VIL1*

TADA analysis of LoF and damaging missense variants found to be *de novo* in ASD subjects, inherited by ASD subjects, or present in ASD subjects (versus control subjects). dnLoF, de *novo* LoF events.

ASH1L and *MLL3*, converge on chromatin remodelling. *MYO9B* plays a key role in dendritic arborization[24]. *MIB1* encodes an E3 ubiquitin ligase critical for neurogenesis[25] and is regulated by miR-137 (ref. 26), a microRNA that regulates neuronal maturation and is implicated in schizophrenia risk[27].

When the WES data from genes with an FDR < 0.3 were evaluated for the presence of deletion copy number variants (CNVs) (such CNVs are functionally equivalent to LoF mutations), 34 CNVs meeting quality and frequency constraints (Supplementary Information) were detected in 5,781 samples (Extended Data Fig. 1). Of the 33 genes with an FDR < 0.1, 3 contained deletion CNVs mapping to 3 ASD subjects and one parent. Of the 74 genes meeting the criterion 0.1 ≤ FDR < 0.3, about one-third could be false positives. Deletion CNVs were found in 14 of these genes and the data supported risk status for 10 of them (Extended Data Table 1 and Extended Data Fig. 4). Two of these ten, *NRXN1* and *SHANK3*, were previously implicated in ASD[2,3,10]. The risk from deletion CNVs, as measured by the odds ratio, is comparable to that from LoF SNVs in cases versus controls or transmission of LoF variants from parents to offspring.

Estimated odds ratios of top genes

Inherent in our conception of the biology of ASD is the notion that there is variation between genes in their impact on risk; for a given class of variants (for example, LoF) some genes have a large impact, others smaller, and still others have no effect at all. In addition, misannotation of variants, among other confounds, can yield false variant calls in subjects (Supplementary Information). These confounds can often be overcome by examining the data in a manner orthogonal to gene discovery. For example, females have greatly reduced rates of ASD relative to males (a 'female protective effect'). Consequently, and regardless of whether this is diagnostic bias or biological protection, females have a higher liability threshold, requiring a larger genetic burden before being diagnosed[22,28,29]. A corollary is that if a variant has the same effect on autism liability in males as it does in females, that variant will be present at a higher frequency in female ASD cases compared to males. Importantly, the magnitude of the difference is proportional to risk as measured by the odds ratio; hence, the effect on risk for a class of variants can be estimated from the difference in frequency between males and females.

Genes with an FDR < 0.1 show profound female enrichment for *de novo* events ($P = 0.005$ for LoF, $P = 0.004$ for Mis3), consistent with *de novo* events having large impacts on liability (odds ratio ≥ 20; Extended Data Fig. 5). However, genes with an FDR between 0.1 and 0.3 show substantially less enrichment for female events, consistent with a modest impact for LoF variants (odds ratio range 2–4, whether transmitted or *de novo*) and little to no effect from Mis3 variants. The

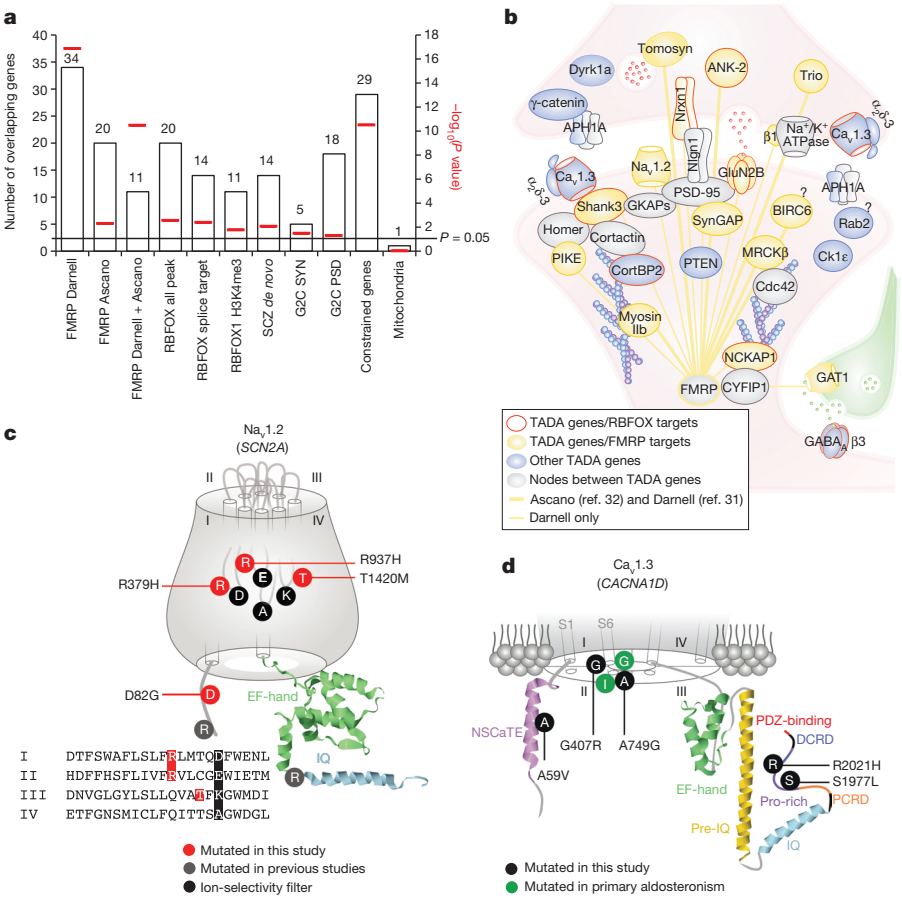


Figure 1 | ASD genes in synaptic networks. **a,** Enrichment of 107 TADA genes in: FMRP targets from two independent data sets[31,32] and their overlap; RBFOX targets; RBFOX targets with predicted alterations in splicing; RBFOX1 and H3K4me3 overlapping targets; genes with *de novo* mutations in schizophrenia (SCZ); human orthologues of Genes2Cognition (G2C) mouse synaptosome (SYN) or PSD genes; constrained genes; and genes encoding mitochondrial proteins (as a control). Red bars indicate empirical *P* values (Supplementary Information). **b,** Synaptic proteins encoded by TADA genes. **c,** *De novo* Mis3 variants in $Na_v1.2$ (*SCN2A*). The four repeats (I–IV) with P-loops, the EF-hand, and the IQ domain are shown, as are the four amino acids (DEKA) forming the inner ring of the ion-selectivity filter. **d,** Variants in $Ca_v1.3$ (*CACNA1D*). Part of the channel is shown, including helices one and six (S1 and S6) for domains I–IV, the NSCaTE motif, the EF-hand domain, the pre-IQ, IQ, proximal (PCRD) and distal (DCRD) C-terminal regulatory domains, the proline-rich region, and the PDZ domain-binding motif.

results are consistent with inheritance patterns: LoF mutations in FDR < 0.1 genes are rarely inherited from unaffected parents whereas those in the 0.1 ≤ FDR < 0.3 group are far more often inherited than they are *de novo* mutations.

By analysing the distribution of relative risk over inferred ASD genes[20], the number of ASD risk genes can be estimated. The estimate relies on the balance of genes with multiple *de novo* LoF mutations versus those with only one: the larger the number of ASD genes, the greater proportion that will show only one *de novo* LoF. This approach yields an estimate of 1,150 ASD genes (Supplementary Information). While there are many more genes to be discovered, many will have a modest impact on risk compared to the genes in Table 1.

Enrichment analyses

Gene sets with an FDR < 0.3 are strongly enriched for genes under evolutionary constraint[19] ($P = 3.0 \times 10^{-11}$; Fig. 1a and Supplementary Table 4), consistent with the hypothesis that heterozygous LoF mutations in these genes are ASD risk factors. Over 5% of ASD subjects carry *de novo* LoF mutations in our FDR < 0.3 list. We also observed that genes in the FDR < 0.3 list had a significant excess of *de novo* non-synonymous events detected by the largest schizophrenia WES study so far[30] ($P = 0.0085$; Fig. 1a), providing further evidence for overlapping risk loci between these disorders and independent confirmation of the signal in the gene sets presented here.

We found significant enrichment for genes encoding messenger RNAs targeted by two neuronal RNA-binding proteins: FMRP[31] (also known as FMR1), mutated or absent in fragile X syndrome ($P = 1.20 \times 10^{-17}$, 34 targets[31], of which 11 are corroborated by an independent data set[32]), and RBFOX (RBFOX1/2/3) ($P = 0.0024$, 20 targets, of which 12 overlap with FMRP), with RBFOX1 shown to be a splicing factor dysregulated in ASD[33,34] (Fig. 1a). These two pathways expand the complexity of ASD neurobiology to post-transcriptional events, including splicing and translation, both of which sculpt the neural proteome.

We found nominal enrichment for human orthologues of mouse genes encoding synaptic ($P = 0.031$) and post-synaptic density (PSD) proteins[35] ($P = 0.046$; Fig. 1a, b and Supplementary Tables 4–6). Enrichment analyses for InterPro, SMART or Pfam domains (FDR < 0.05 and a minimum of five genes per category) reveal an overrepresentation of DNA- or histone-related domains: eight genes encoding proteins with InterPro zinc-finger FYVE PHD domains (142 such annotated genes in the genome; FDR = 7.6×10^{-4}), and five with Pfam Su(var)3-9, enhancer-of-zeste, trithorax (SET) domains (39 annotated in the genome; FDR = 8.2×10^{-4}).

Integrating complementary data

To implicate additional genes in risk for ASD, we used a model called DAWN (detecting association with networks)[36]. DAWN evokes a hidden Markov random field framework to identify clusters of genes that show strong association signals and highly correlated co-expression in a key tissue and developmental context. Previous research suggests human mid-fetal prefrontal and motor-somatosensory neocortex is a critical nexus for risk[16], thus we evaluated gene co-expression data from that tissue together with TADA scores for genes with an FDR < 0.3. Because this list is enriched for genes under evolutionary constraint, we generalized DAWN to incorporate constraint scores (Supplementary Information). When TADA results, gene co-expression in mid-fetal neocortex and constraint scores are jointly modelled, DAWN identifies 160 genes that plausibly affect risk (Fig. 2), 91 of which are not in the 107 TADA genes with an FDR < 0.3. Moreover, the model parameter describing evolutionary constraint is an important predictor of clusters of putative risk genes ($P = 0.018$).

A subnetwork obtained by seeding the 160 DAWN genes within a high-confidence protein–protein interactome[14] confirmed that the putative genes are enriched for neuronal functions. We kept the largest connected component, containing 95 seed DAWN genes, 50 of which were in the FDR < 0.3 gene set. The DAWN gene products form four natural

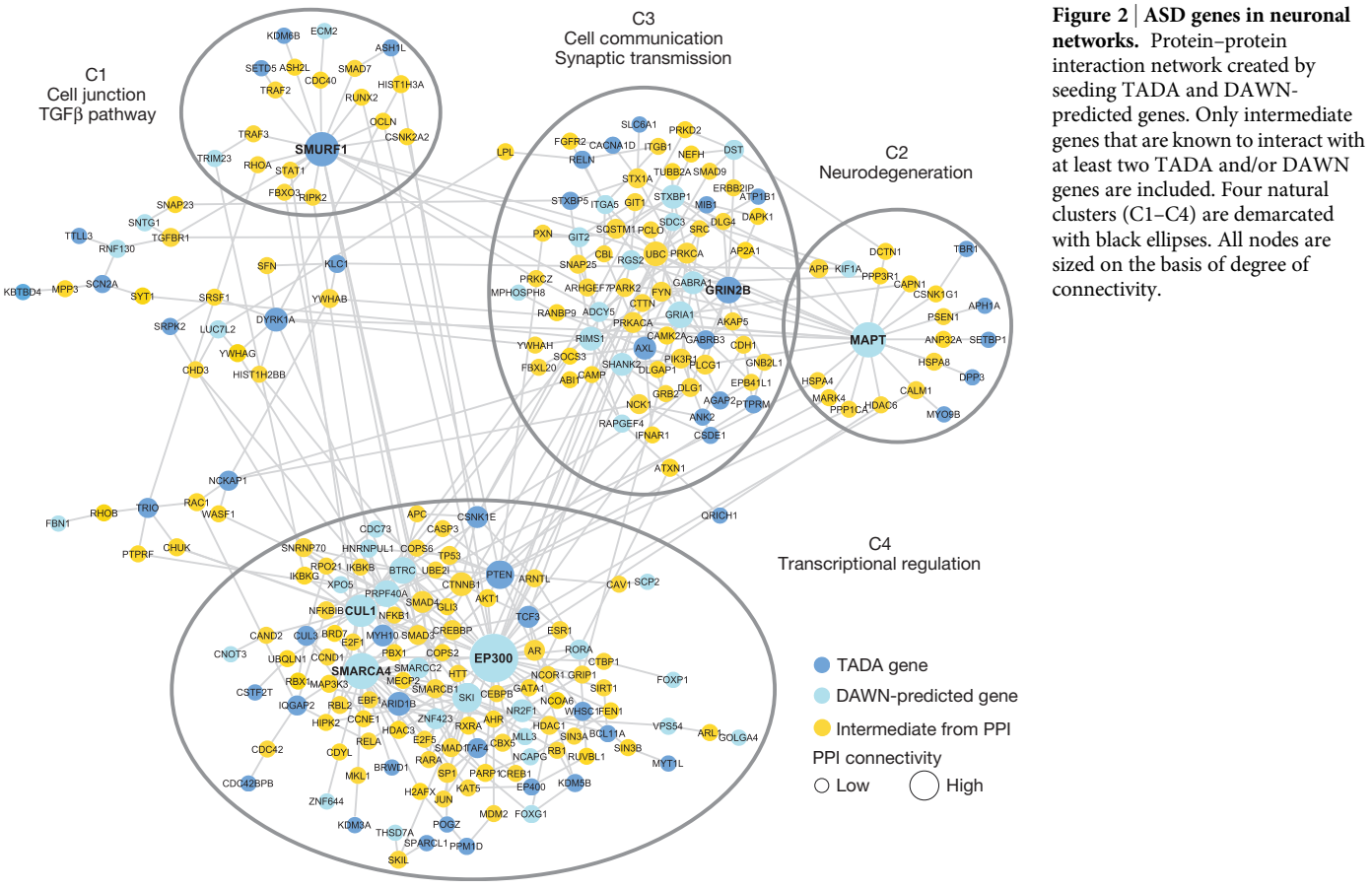


Figure 2 | ASD genes in neuronal networks. Protein–protein interaction network created by seeding TADA and DAWN-predicted genes. Only intermediate genes that are known to interact with at least two TADA and/or DAWN genes are included. Four natural clusters (C1–C4) are demarcated with black ellipses. All nodes are sized on the basis of degree of connectivity.

clusters on the basis of network connectivity (Fig. 2). We visualized the enriched pathways and biological functions for each of these clusters on 'canvases'[37] (Extended Data Fig. 6). Many of the previously known ASD risk genes fall in cluster C3, including genes involved in synaptic transmission and cell–cell communication. Cluster C4 is enriched for genes related to transcriptional and chromatin regulation. Many TADA and DAWN genes in this cluster interact tightly with other transcription factors, histone-modifying enzymes and DNA-binding proteins. Five TADA genes in the cluster C2 are bridged to the rest of the network through *MAPT*, as inferred by DAWN. The enrichment results for cluster C2 indicate that genes implicated in neurodegenerative disorders could also have a role in neurodevelopmental disorders.

Emergent results

Amongst the critical synaptic components found to be mutated in our study are voltage-gated ion channels involved in fundamental processes including the propagation of action potentials (for example, the $Na_v1.2$ channel), neuronal pacemaking and excitability–transcription coupling (for example, the $Ca_v1.3$ channel) (Fig. 1b). We identified four LoF and five Mis3 variants in *SCN2A* ($Na_v1.2$), three Mis3 variants in *CACNA1D* ($Ca_v1.3$) and two LoF variants in *CACNA2D3* ($\alpha_2\delta$-3 subunit). Remarkably, three *de novo* Mis3 variants in *SCN2A* affected residues mutated in homologous genes in patients with other syndromes, including Brugada syndrome (*SCN5A*) or epilepsy disorders (*SCN1A*) (Arg379His and Arg 937His). These arginines, as well as the threonine mutated in Thr1420Met, cluster to the P-loops forming the ion selectivity filter, located in proximity to the inner ring (DEKA motif) (Fig. 1c). Because homologous channels mutated in these arginines do not conduct inward Na^+ currents[38,39], Arg379His and Arg937His mutations might have similar effect.

Two *de novo CACNA1D* variants (Gly407Arg and Ala749Gly) emerged at positions proximal to residues mutated in patients with primary aldosteronism and neurological deficits (Fig. 1d). The reported mutations interfere with channel activation and inactivation[40]. Amongst variants found in cases, Ala59Val maps to the NSCaTE domain, also important for Ca^{2+}-dependent inactivation, and Ser1977Leu and Arg2021His co-cluster in the carboxy-terminal proline-rich domain, the site of interaction with SHANK3, a key PSD scaffolding protein. Mutations in RIMS1 and RIMBP2, which can associate with $Ca_v1.3$, were found in our cohort (but with an FDR > 0.3).

Chromatin remodelling involves histone-modifying enzymes (encoded by histone-modifier genes, HMGs) and chromatin remodellers (readers) that recognize specific histone post-translational modifications and orchestrate their effects on chromatin. Our gene set is enriched in HMGs (9 HMGs out of 152 annotated in HIstome[41], Fisher's exact test, $P = 2.2 \times 10^{-7}$). Enrichment in the gene ontology term 'histone-lysine *N*-methytransferase activity' (5 genes out of 41 so annotated; FDR $= 2.2 \times 10^{-2}$) highlights this as a prominent pathway.

Lysines on histones 3 and 4 can be mono-, di- or tri-methylated, providing a versatile mechanism for either activation or repression of transcription. Of 107 TADA genes, five are SET lysine methyltransferases, four are jumonji lysine demethylases, and two are readers (Fig. 3a). RBFOX1 co-isolates with histone H3 trimethyl Lys 4 (H3K4me3)[42], and our data set is enriched in targets shared by RBFOX1 and H3K4me3 ($P = 0.0166$; Fig. 1a and Supplementary Table 4). Some *de novo* missense variants targeting these genes map to functional domains (Extended Data Fig. 7).

For the H3K4me2 reader *CHD8*, we extended our analyses in search of additional *de novo* variation in the cases of the case-control sample. By sequencing complete parent–child trios for many *CHD8* variants, five variants were found to be *de novo*, two of which affect essential splice sites and cause LoF by exon skipping or activation of cryptic splice sites in lymphoblastoid cells (Fig. 3b).

Given the role of HMGs in transcription, we reasoned that TADA genes might be interconnected through transcription 'routes'. We searched for a connected network (seeded by 9 TADA HMGs) in a transcription factor interaction network (ChEA)[43]. We found that 46 TADA genes

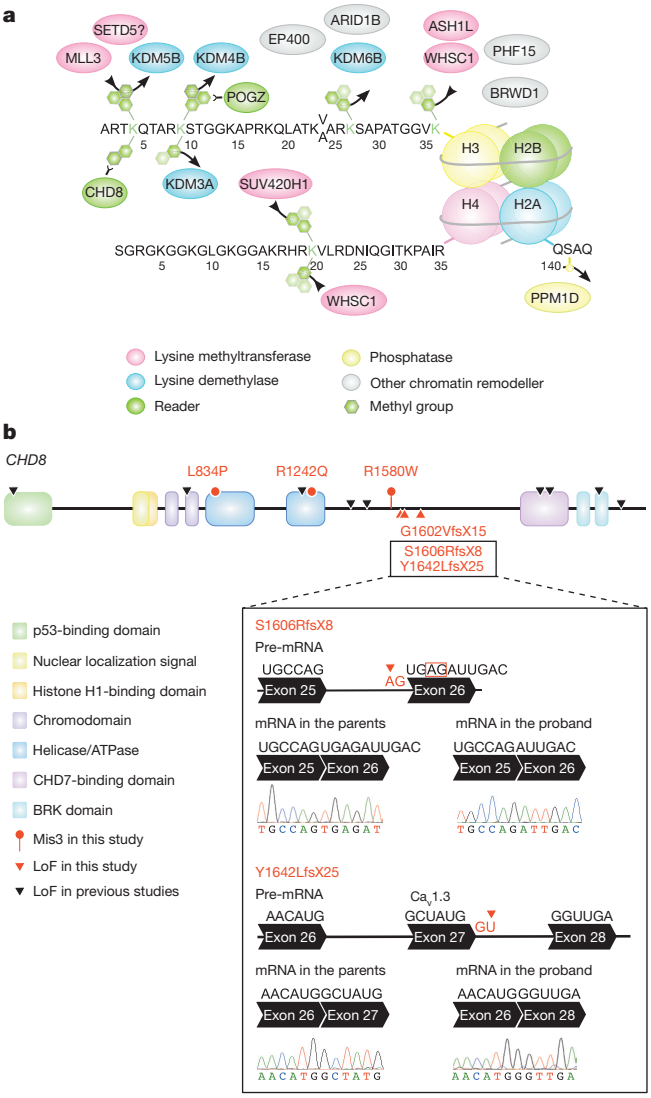


Figure 3 | ASD genes in chromatin remodelling. **a,** TADA genes cluster to chromatin-remodelling complexes. Amino-terminals of histones H3, H4 and part of H2A are shown. Lysine methyltransferases add methyl groups, whereas lysine demethylases remove them. **b,** *De novo* Mis3 and LoF variants in CHD8. The box shows the outcome of reverse transcription PCR and Sanger sequencing in lymphoblastoid cells for two newly identified *de novo* splice-site variants. The first mutation affects an acceptor splice site (red arrow), causing the activation of a cryptic splice site (red box), a four-nucleotide deletion, frame shift and a premature stop. The second mutation affects a donor splice site (red arrow), causing exon skipping, frame shift and a premature stop.

are directly interconnected in a 55-gene cluster (Extended Data Fig. 8) ($P = 0.002$; 1,000 random draws), for a total of 69 when including all known HMGs (Fig. 4) ($P = 0.001$; 1,000 random draws).

Examining the Human Gene Mutation Database we found that the 107 TADA genes included 21 candidate genes for intellectual disability, 3 for epilepsy, 17 for schizophrenia, 9 for congenital heart disease and 6 for metabolic disorders (Fig. 5).

Conclusions

Complementing earlier reports, ASD subjects show a clear excess of *de novo* LoF mutations above expectation, with a concentration of such events in a handful of genes. While this handful has a large effect on risk, most ASD genes have a much smaller impact. This gradient emerges most notably from the contrast of risk variation in male and female ASD subjects. Unlike some earlier studies, but consistent with expectation, the data also show clear evidence for effect of *de novo* missense SNVs

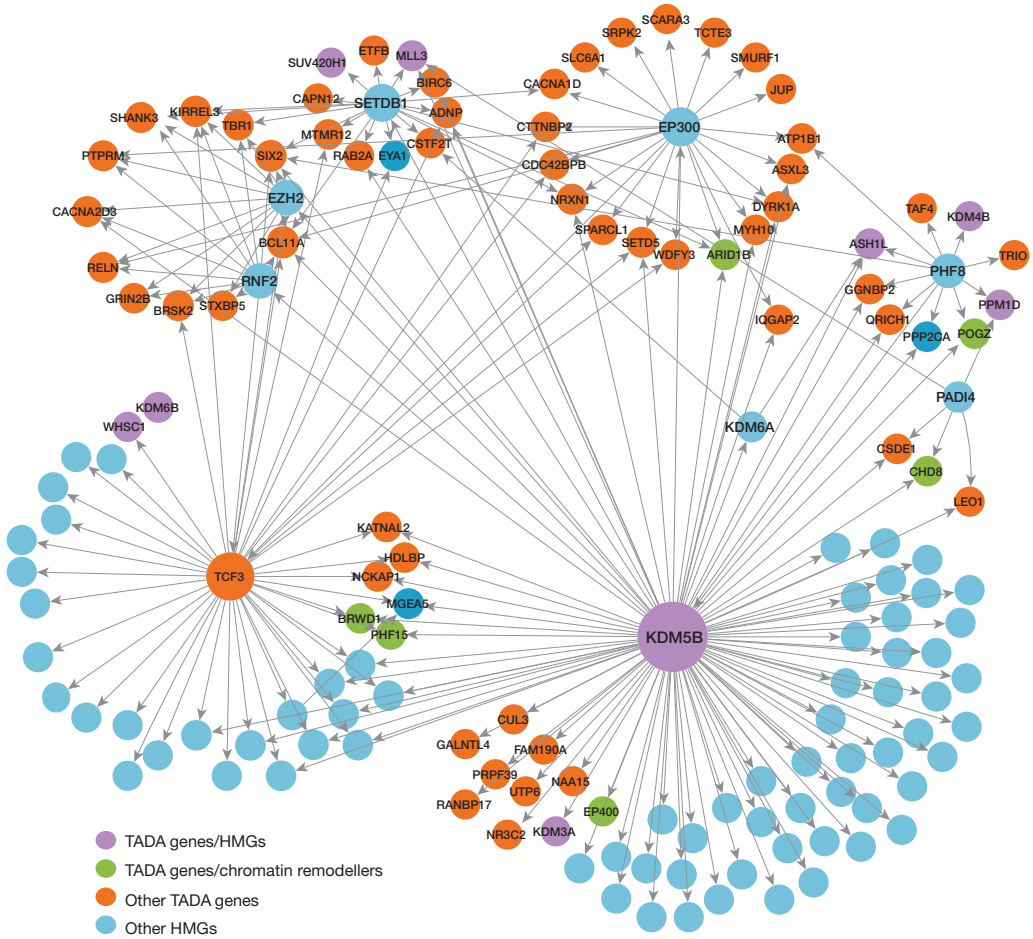


Figure 4 | Transcription regulation network of TADA genes. Edges indicate transcription regulators (source nodes) and their gene targets (target nodes) based on the ChEA network; interactions among only HMGs are ignored.

on risk; for risk generated by LoF variants transmitted from unaffected parents; and for the value of case-control design in gene discovery. By integrating data on *de novo*, inherited and case-control variation, the yield of ASD gene discoveries was doubled over what would be obtained from a count of *de novo* LoF variants alone. ASD genes almost uniformly show strong constraints against variation, a feature we exploit to implicate other genes in risk.

Three critical pathways for typical development are damaged by risk variation: chromatin remodelling, transcription and splicing, and synaptic function. Chromatin remodelling controls events underlying the formation of neural connections, including neurogenesis and neural differentiation[44], and relies on epigenetic marks as post-translational modifications of histones . Here we provide extensive evidence for HMGs and readers in sporadic ASD, implicating specifically lysine methylation and extending the mutational landscape of the emergent ASD gene *CHD8* to missense variants. Splicing is implicated by the enrichment of RBFOX targets in the top ASD candidates. Risk variation also affects multiple classes and components of synaptic networks, from receptors and ion channels to scaffolding proteins. Because a wide set of synaptic genes is disrupted in idiopathic ASD, it seems reasonable to suggest that altered chromatin dynamics and transcription, induced by disruption of relevant genes, leads to impaired synaptic function as well. *De novo* mutations in ASD[11–15], intellectual disability[45] and schizophrenia[30] cluster to synaptic genes, and synaptic defects have been reported in models of these disorders[46]. Integrity of synaptic function is essential for neural physiology, and its perturbation could represent the intersection between diverse neuropsychiatric disorders[47].

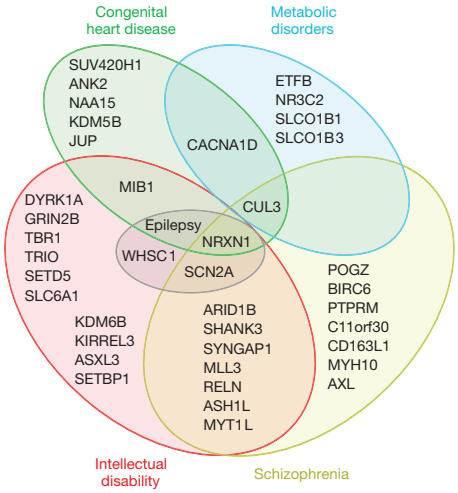


Figure 5 | Involvement in disease of ASD genes. The Venn diagram shows the overlap in disease involvement for the TADA genes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 May; accepted 18 August 2014.
Published online 29 October 2014.

1. Ronald, A. & Hoekstra, R. A. Autism spectrum disorders and autistic traits: a decade of new twin studies. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **156,** 255–274 (2011).
2. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316,** 445–449 (2007).
3. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466,** 368–372 (2010).
4. Klei, L. *et al.* Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism* **3,** 9 (2012).

5. Gaugler, T. *et al.* Most inherited risk for autism resides with common variation. *Nature Genet.* **46,** 881–885 (2014).
6. Yu, T. W. *et al.* Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77,** 259–273 (2013).
7. Lim, E. T. *et al.* Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* **77,** 235–242 (2013).
8. Poultney, C. S. *et al.* Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. *Am. J. Hum. Genet.* **93,** 607–619 (2013).
9. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380,** 42–77 (2011).
10. Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459,** 569–573 (2009).
11. Sanders, S. J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485,** 237–241 (2012).
12. O'Roak, B. J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338,** 1619–1622 (2012).
13. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485,** 246–250 (2012).
14. Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485,** 242–245 (2012).
15. Iossifov, I. *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74,** 285–299 (2012).
16. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155,** 997–1007 (2013).
17. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).
18. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7,** 248–249 (2010).
19. Samocha, K. E. *et al.* A framework for the interpretation of *de novo* mutation in human disease. *Nature Genet.* **46,** 944–950 (2014).
20. He, X. *et al.* Integrated model of *de novo* and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet.* **9,** e1003671 (2013).
21. Girirajan, S. *et al.* Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am. J. Hum. Genet.* **92,** 221–237 (2013).
22. Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* **94,** 677–694 (2014).
23. Helsmoortel, C. *et al.* A SWI/SNF-related autism syndrome caused by *de novo* mutations in *ADNP*. *Nature Genet.* **46,** 380–384 (2014).
24. Long, H. *et al.* Myo9b and RICS modulate dendritic morphology of cortical neurons. *Cereb. Cortex* **23,** 71–79 (2013).
25. Yoon, K. J. *et al.* Mind bomb 1-expressing intermediate progenitors generate Notch signaling to maintain radial glial cells. *Neuron* **58,** 519–531 (2008).
26. Smrt, R. D. *et al.* MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase Mind bomb-1. *Stem Cells* **28,** 1060–1070 (2010).
27. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet.* **45,** 1150–1159 (2013).
28. Robinson, E. B., Lichtenstein, P., Anckarsater, H., Happe, F. & Ronald, A. Examining and interpreting the female protective effect against autistic behavior. *Proc. Natl Acad. Sci. USA* **110,** 5258–5262 (2013).
29. Jacquemont, S. *et al.* A higher mutational burden in females supports a ''female protective model'' in neurodevelopmental disorders. *Am. J. Hum. Genet.* **94,** 415–425 (2014).
30. Fromer, M. *et al. De novo* mutations in schizophrenia implicate synaptic networks. *Nature* **506,** 179–184 (2014).
31. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146,** 247–261 (2011).
32. Ascano, M. Jr. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492,** 382–386 (2012).
33. Weyn-Vanhentenryck, S. M. *et al.* HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* **6,** 1139–1152 (2014).
34. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474,** 380–384 (2011).
35. Collins, M. O. *et al.* Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *J. Neurochem.* **97** (suppl. 1), 16–23 (2006).
36. Liu, L. *et al.* DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. *Mol. Autism* **5,** 22 (2014).
37. Tan, C. M., Chen, E. Y., Dannenfelser, R., Clark, N. R. & Ma'ayan, A. Network2Canvas: network visualization on a canvas with enrichment analysis. *Bioinformatics* **29,** 1872–1878 (2013).
38. Vatta, M. *et al.* Genetic and biophysical basis of sudden unexplained nocturnal death syndrome (SUNDS), a disease allelic to Brugada syndrome. *Hum. Mol. Genet.* **11,** 337–345 (2002).
39. Volkers, L. *et al.* Na_v 1.1 dysfunction in genetic epilepsy with febrile seizures-plus or Dravet syndrome. *Eur. J. Neurosci.* **34,** 1268–1275 (2011).
40. Scholl, U. I. *et al.* Somatic and germline *CACNA1D* calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. *Nature Genet.* **45,** 1050–1054 (2013).
41. Khare, S. P. *et al.* HIstome–a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.* **40,** D337–D342 (2012).
42. Feng, J. *et al.* Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol.* **15,** R65 (2014).
43. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26,** 2438–2444 (2010).
44. Ronan, J. L., Wu, W. & Crabtree, G. R. From neural development to cognition: unexpected roles for chromatin. *Nature Rev. Genet.* **14,** 347–359 (2013).
45. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380,** 1674–1682 (2012).
46. Penzes, P., Cahill, M. E., Jones, K. A., VanLeeuwen, J. E. & Woolfrey, K. M. Dendritic spine pathology in neuropsychiatric disorders. *Nature Neurosci.* **14,** 285–293 (2011).
47. Zoghbi, H. Y. Postnatal neurodevelopmental disorders: meeting at the synapse? *Science* **302,** 826–830 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health (NIH) grants U01MH100233, U01MH100209, U01MH100229 and U01MH100239 to the Autism Sequencing Consortium. Sequencing at Broad Institute was supported by NIH grants R01MH089208 (M.J.D.) and new sequencing by U54 HG003067 (S. Gabriel, E. Lander). Other funding includes NIH R01 MH089482, R37 MH057881 (B.D. and K.R.), R01 MH061009 (J.S.S.), UL1TR000445 (NCAT to VUMC); P50 HD055751 (E.H.C.); MH089482 (J.S.S.), NIH R01 MH083565 and RC2MH089952 (C.A.W.), NIMH MH095034 (P.S), MH077139 (P.F. Sullivan); 5UL1 RR024975 and P30 HD15052. The DDD Study is funded by HICF-1009-003 and WT098051. UK10K is funded by WT091310. We also acknowledge The National Children's Research Foundation, Our Lady's Children's Hospital, Crumlin; The Meath Foundation; AMNCH, Tallaght; The Health Research Board, Ireland and Autism Speaks, U.S.A. C.A.W. is an Investigator of the Howard Hughes Medical Institute. S.D.R., A.P.G., C.S.P., Y.K. and S.-C.F. are Seaver fellows, supported by the Seaver foundation. A.P.G. is also supported by the Charles and Ann Schlaifer Memorial Fund. P.F.B. is supported by a UK National Institute for Health Research (NIHR) Senior Investigator award and the NIHR Biomedical Research Centre in Mental Health at the South London & Maudsley Hospital. A.C. is supported by María José Jove Foundation and the grant FIS PI13/01136 of the Strategic Action from Health Carlos III Institute (FEDER). This work was supported in part through the computational resources and staff expertise provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai. We acknowledge the assistance of D. Hall and his team at National Database for Autism Research. We thank Jian Feng for providing a list of targets of both RBFOX1 and H3K4me3. We thank M. Potter for data coordination; K. Moore and J. Reichert for technical assistance; and, S. Lindsay for helping with molecular validation. We acknowledge the clinicians and organizations that contributed to samples used in this study. Finally, we are grateful to the many families whose participation made this study possible.

Author Contributions Study conception and design: J.D.B., D.J.C., M.J.D., S.D.R., B.D., M.F., A.P.G., X.H., T.L., C.S.P., K.Ro., M.W.S. and M.E.Z. Data analysis: J.C.B., P.F.B., J.D.B., J.C., A.E.C, D.J.C., M.J.D., S.D.R., B.D., M.F., S.-C.F., A.P.G., X.H., L.K., J.K., Y.K., L.L., A.M., C.S.P., S.P., K.Ro., K.S., C.S., T.S., C.St., S.W., L.W. and M.E.Z. Contribution of samples, WES data or analytical tools: B.A., J.C.B., M.B., P.F.B., J.D.B., J.C., N.G.C., A.C., M.H.C., A.G.C., A.E.C, H.C., E.L.C., L.C., S.R.C., D.J.C., M.J.D., G.D., S.D.R., B.D., E.D., B.A.F., C.M.F., M.F., L.G., E.G., M.G., A.P.G., S.J.G., X.H., R.H., C.M.H., I.I.-L., P.J.G., H.K., S.M.K., L.K., A.K., J.K., Y.K., I.L., J.L., T.Le., C.L., L.L., A.M., C.R.M., A.L.M., B.N., M.J.O., N.O., A.P., M.P., J.R.P., C.S.P., S.P., K.P., D.R., K.R., A.R., K.Ro., A.S., M.S., K.S., S.J.S., C.S., G.D.S., S.W.S., M.S.-R., T.S., P.S., D.S., M.W.S., C.St., J.S.S., P.Sz., K.T., O.V., A.V., S.W., C.A.W., L.W., L.A.W., J.A.W., T.W.Y., R.K.C.Y., M.E.Z. Writing of the paper: J.C.B., J.D.B., E.H.C., D.J.C., M.J.D., S.D.R., B.D., M.G., A.P.G., X.H., C.S.P., K.Ro., S.W.S., M.E.Z. Leads of ASC committees: J.D.B., E.H.C., M.J.D., B.D., M.G., K.Ro., M.W.S., J.S.S., M.E.Z. Administration of ASC: J.M.B.

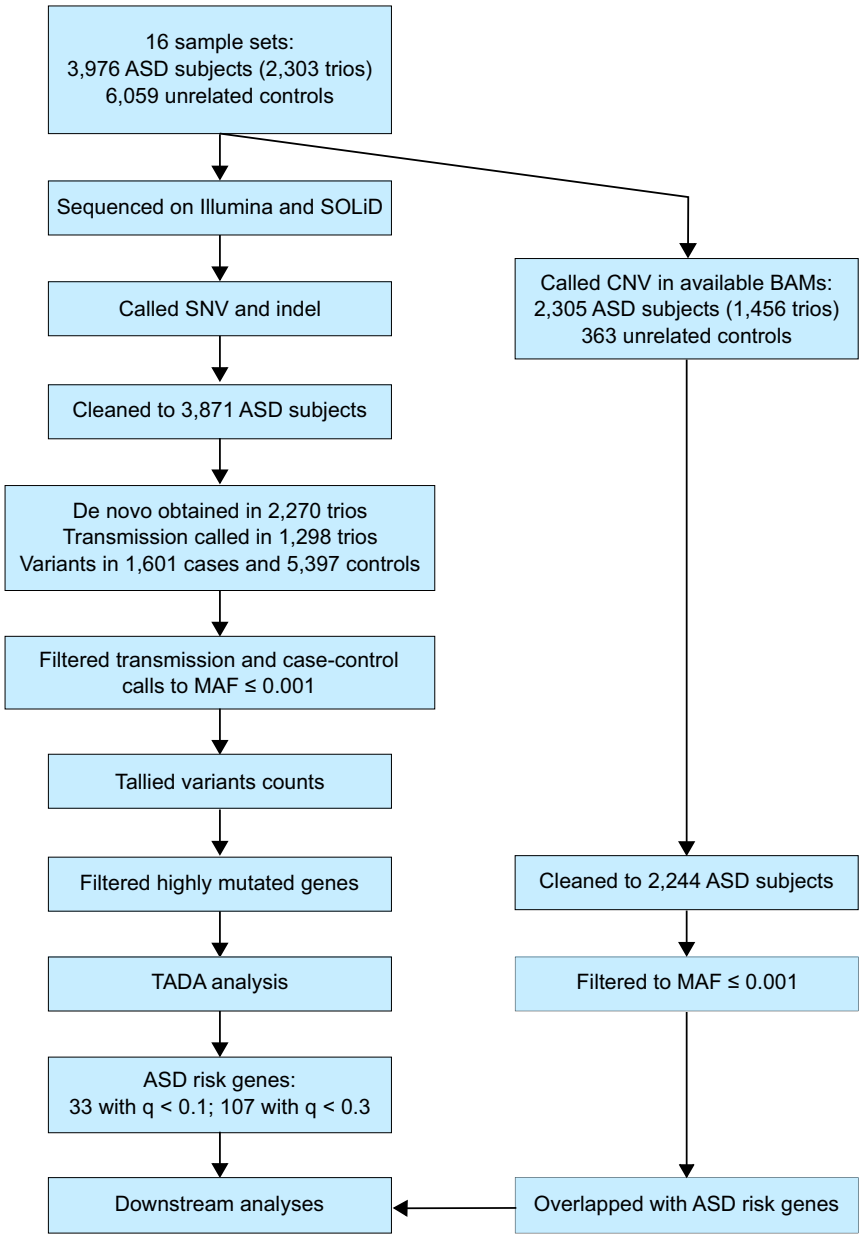
Author Information New data included in this manuscript have been deposited at dbGAP merged with our published data under accession number phs000298.v1.p1 and is available for download at (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000298.v1.p1). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.B. (joseph.buxbaum@mssm.edu) or M.J.D. (mjdaly@broadinstitute.org).

Silvia De Rubeis[1,2], Xin He[3], Arthur P. Goldberg[1,2,4], Christopher S. Poultney[1,2], Kaitlin Samocha[5], A. Ercument Cicek[3], Yan Kou[1,2], Li Liu[6], Menachem Fromer[2,4,5], Susan Walker[7], Tarjinder Singh[8], Lambertus Klei[9], Jack Kosmicki[5], Shih-Chen Fu[1,2], Branko Aleksic[10], Monica Biscaldi[11], Patrick F. Bolton[12], Jessica M. Brownfeld[1,2], Jinlu Cai[1,2], Nicholas G. Campbell[13,14], Angel Carracedo[15,16], Maria H. Chahrour[17,18], Andreas G. Chiocchetti[19], Hilary Coon[20,21], Emily L. Crawford[13,14], Lucy Crooks[8], Sarah R. Curran[12], Geraldine Dawson[22], Eftichia Duketis[19], Bridget A. Fernandez[23], Louise Gallagher[24], Evan Geller[25], Stephen J. Guter[26], R. Sean Hill[17,18], Iuliana Ionita-Laza[27], Patricia Jimenez Gonzalez[28], Helena Kilpinen[29], Sabine M. Klauck[30], Alexander Kolevzon[1,2,31], Irene Lee[32], Jing Lei[6], Terho Lehtimäki[33], Chiao-Feng Lin[25], Avi Ma'ayan[34], Christian R. Marshall[7], Alison L. McInnes[35], Benjamin Neale[36], Michael J. Owen[37], Norio Ozaki[10], Mara Parellada[38], Jeremy R. Parr[39], Shaun Purcell[2], Kaija Puura[40], Deepthi Rajagopalan[7], Karola Rehnström[8], Abraham Reichenberg[1,2,41], Aniko Sabo[42], Michael Sachse[19], Stephan J. Sanders[43], Chad Schafer[6], Martin Schulte-Rüther[44], David Skuse[32,45], Christine Stevens[36], Peter Szatmari[46], Kristiina Tammimies[7], Otto Valladares[25], Annette Voran[47], Li-San Wang[25], Lauren A. Weiss[43], A. Jeremy Willsey[43], Timothy W. Yu[17,18], Ryan K. C. Yuen[7], The DDD Study*, Homozygosity Mapping Collaborative for Autism*, UK10K Consortium*, The Autism Sequencing Consortium*, Edwin H. Cook[26], Christine M. Freitag[19], Michael Gill[24], Christina M. Hultman[48], Thomas Lehner[49], Aarno Palotie[5,50,51,52], Gerard D. Schellenberg[25], Pamela Sklar[2,4,53], Matthew W. State[43], James S. Sutcliffe[13,14], Christopher A. Walsh[17,18],

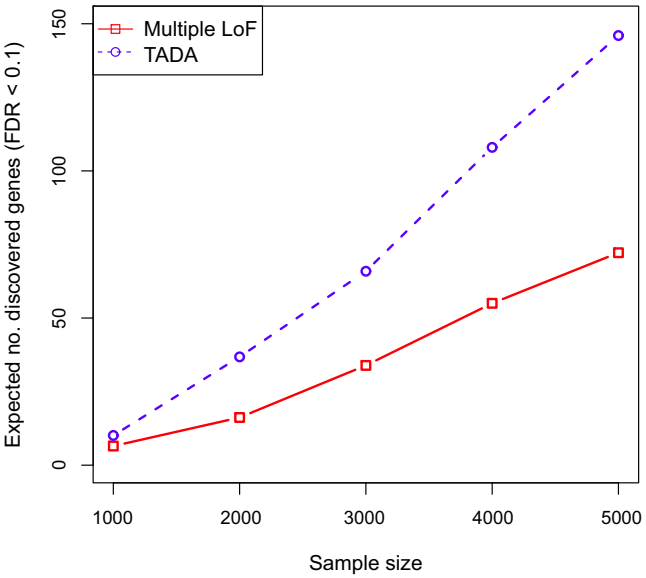
Stephen W. Scherer[7,54], Michael E. Zwick[55], Jeffrey C. Barrett[8], David J. Cutler[55], Kathryn Roeder[6,3], Bernie Devlin[9], Mark J. Daly[17,36,56] & Joseph D. Buxbaum[1,2,4,53,57,58]

[1]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York 10029, New York, USA. [3]Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. [4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [5]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [6]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. [7]Program in Genetics and Genome Biology, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. [8]The Wellcome Trust Sanger Institute, Cambridge, CB10 1SA, UK. [9]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA. [10]Department of Psychiatry, Graduate School of Medicine, Nagoya University, Nagoya 466-8550, Japan. [11]Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, University Medical Center Freiburg; Center for Mental Disorders, 79106 Freiburg, Germany. [12]Department of Child Psychiatry & SGDP Centre, King's College London Institute of Psychiatry, Psychology & Neuroscience, London, SE5 8AF, UK. [13]Vanderbilt Brain Institute, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [14]Department of Molecular Physiology and Biophysics and Psychiatry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. [15]Genomic Medicine Group, CIBERER, University of Santiago de Compostela and Galician Foundation of Genomic Medicine (SERGAS), 15706 Santiago de Compostela, Spain. [16]Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah 21589, Kingdom of Saudi Arabia. [17]Harvard Medical School, Boston, Massachusetts 02115, USA. [18]Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts 02115, USA. [19]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt, 60528 Frankfurt, Germany. [20]Department of Internal Medicine, University of Utah, Salt Lake City, Utah 84132, USA. [21]Department of Psychiatry, University of Utah, Salt Lake City, Utah 84108, USA. [22]Duke Institute for Brain Sciences, Duke University, Durham, North Carolina 27708, USA. [23]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John's, Newfoundland A1B 3V6, Canada. [24]Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin 8, Ireland. [25]Geisinger Health System, Danville, Pennsylvania 17822, USA. [26]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois 60608, USA. [27]Department of Biostatistics, Columbia University, New York, New York 10032, USA. [28] Hospital Nacional de Niños Dr Saenz Herrera, CCSS, Child Developmental and Behavioral Unit, San José, Costa Rica. [29]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. [30]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [31]Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [32]Institute of Child Health, University College London, London, WC1N 1EH, UK. [33]Department of Clinical Chemistry, Fimlab Laboratories, SF-33100 Tampere, Finland. [34]Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [35]Department of Psychiatry Kaiser Permanente, San Francisco, California 94118, USA. [36]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [37]MRC Centre for Neuropsychiatric Genetics and Genomics, and the Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, CF24 4HQ, UK. [38]Child and Adolescent Psychiatry Department, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, Universidad Complutense, 28040 Madrid, Spain. [39]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK. [40]Department of Child Psychiatry, University of Tampere and Tampere University Hospital, 33521 Tampere, Finland SF-33101. [41]Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [42]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. [43]Department of Psychiatry, University of California at San Francisco, San Francisco, California 94143–0984, USA. [44]Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Translational Brain Medicine in Psychiatry and Neurology, University Hospital RWTH Aachen / JARA Brain Translational Medicine, 52056 Aachen, Germany. [45]Department of Child and Adolescent Mental Health, Great Ormond Street Hospital for Children, National Health Service Foundation Trust, London, WC1N 3JH, UK. [46]Department of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies, McMaster University, Hamilton, Ontario L8S 4K1, Canada. [47]Department of Child and Adolescent Psychiatry, Saarland University Hospital, D-66424 Homburg, Germany. [48]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-171 77 Stockholm, Sweden. [49]National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892-9663, USA. [50]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [51]Institute for Molecular Medicine Finland, University of Helsinki, FI-00014 Helsinki, Finland. [52]Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [53]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [54]McLaughlin Centre, University of Toronto, Toronto, Ontario M5S 1A1, Canada. [55]Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, USA. [56]Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [57]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [58]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

*Lists of participants appear in the Supplementary Information.

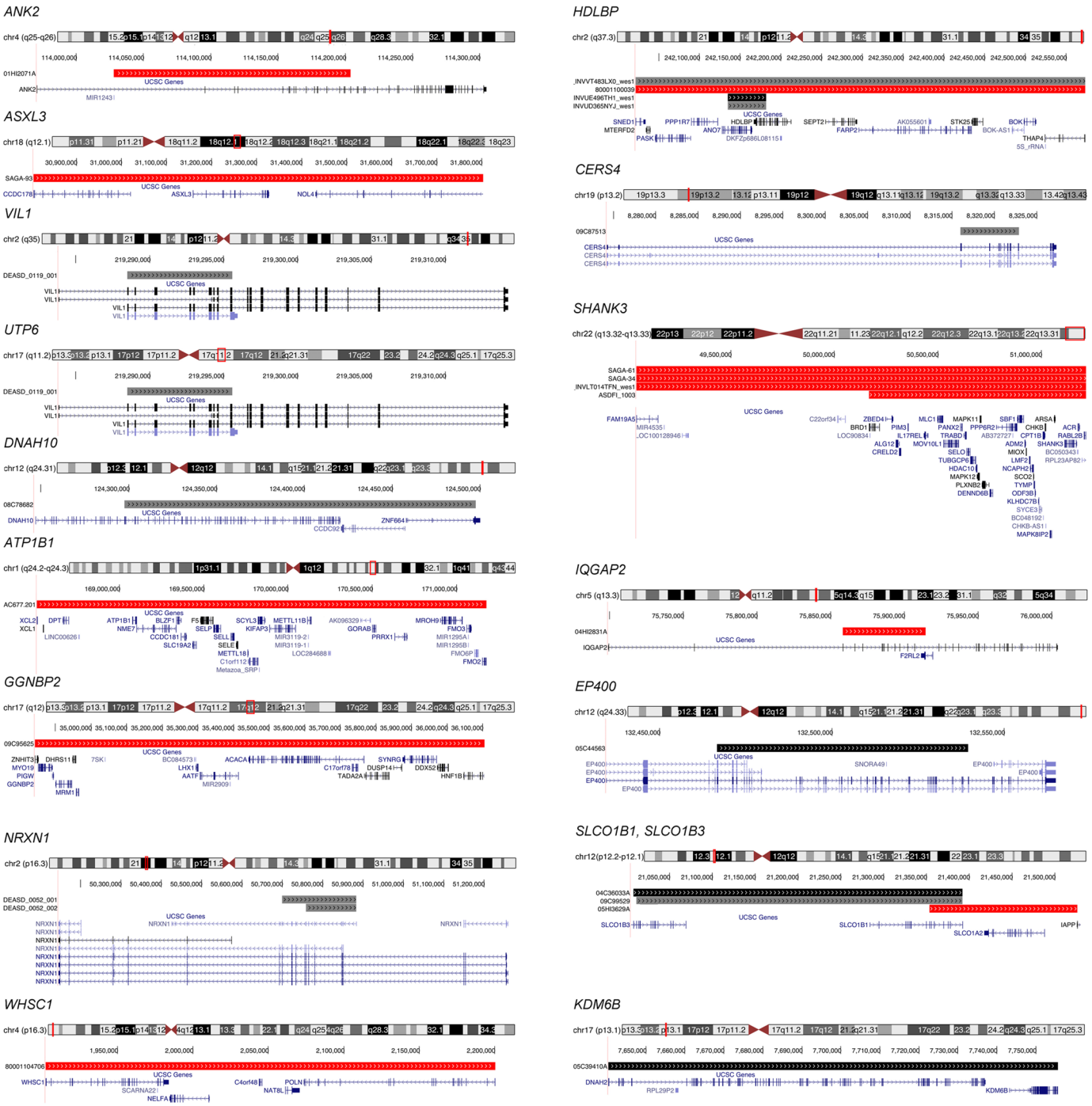


Extended Data Figure 1 | Workflow of the study. The workflow began with 16 sample sets, as listed in Supplementary Table 1. DNA was obtained, and exomes were captured and sequenced. After variant calling, quality control was performed: duplicate subjects and incomplete families were removed and subjects with extreme genotyping, *de novo*, or variant rates were removed. Following cleaning, 3,871 subjects with ASD remained. Analysis proceeded separately for SNVs and indels, and CNVs. *De novo* and transmission/non-transmission variants were obtained for trio data (published *de novo* variants from 825 trios[11,13–15] were incorporated). This led to the TADA analysis, which found 33 ASD risk genes with an FDR < 0.1; and 107 with an FDR < 0.3. CNVs were called in 2,305 ASD subjects. BAM, binary alignment/map; MAF, minor allele frequency.

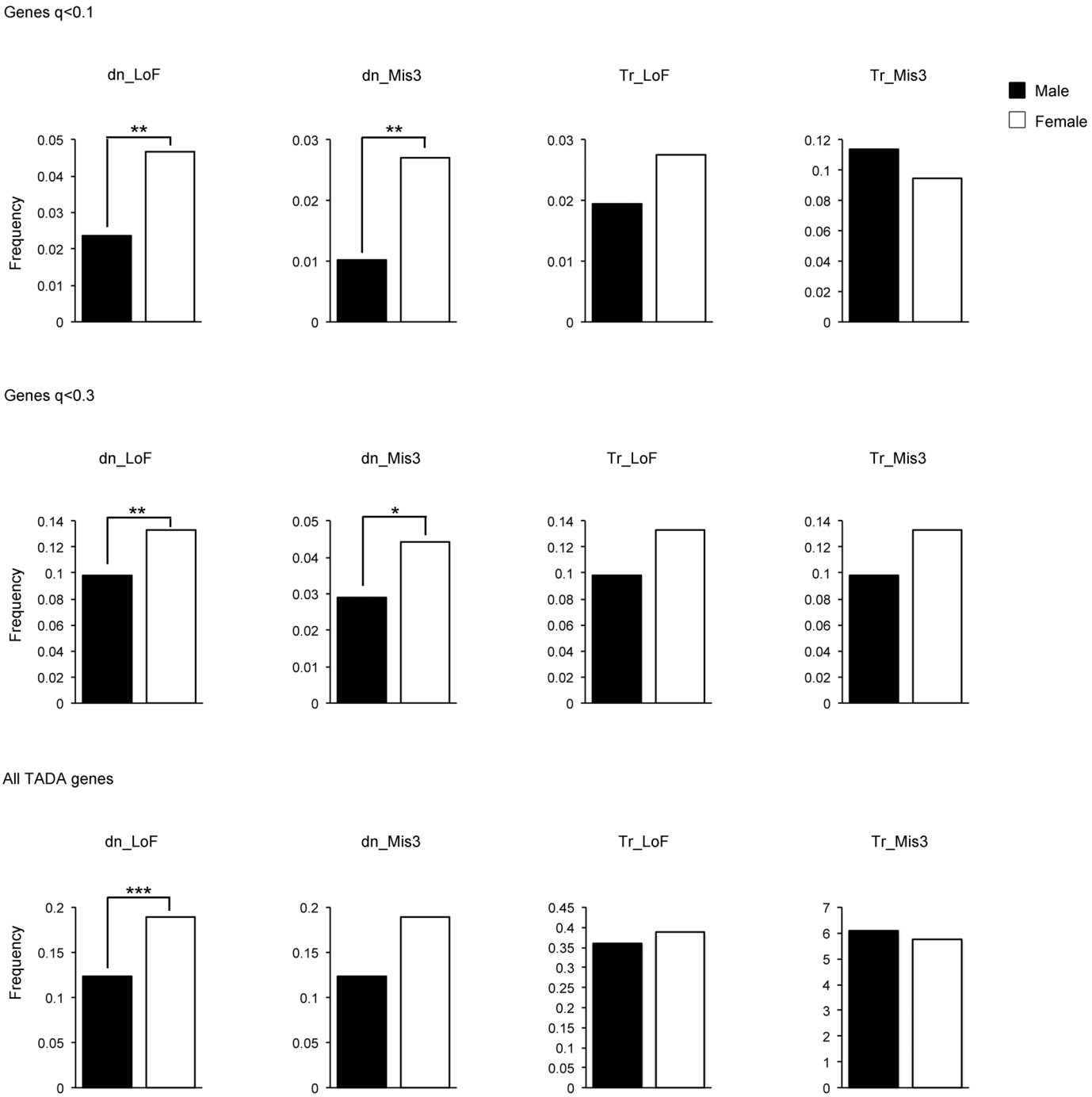


Extended Data Figure 2 | Expected number of ASD genes discovered as a function of sample size. The multiple LoF test (red) is a restricted version of TADA that uses only the *de novo* LoF data. TADA (blue) models *de novo* LoF, *de novo* Mis3, LoF variants transmitted/not transmitted and LoF variants observed in case-control samples. The sample size (n) indicates either n trios for which we record *de novo* and transmitted variation (TADA), or n trios for which we record only *de novo* events (multiple LoF), plus n cases and n controls.

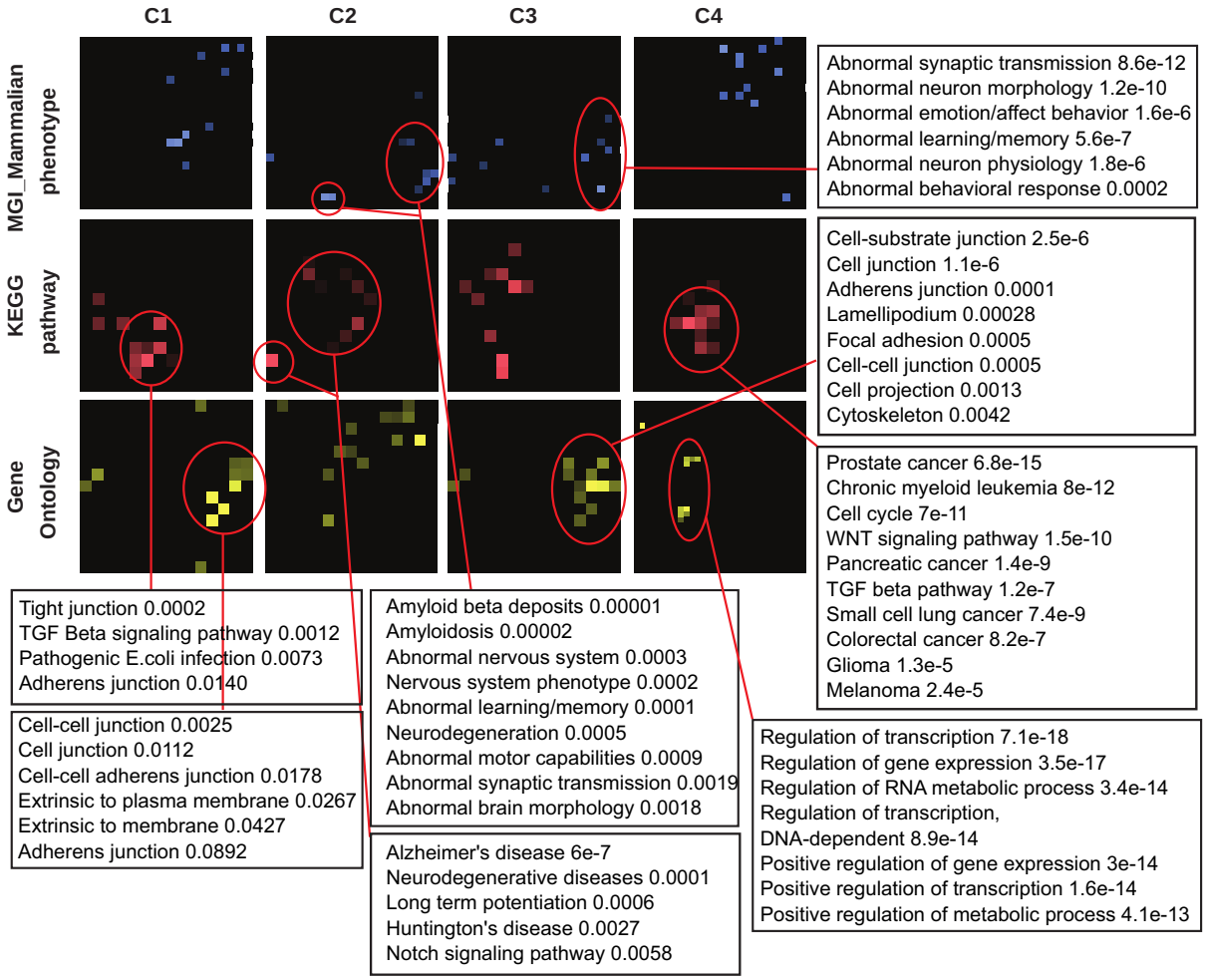
Extended Data Figure 3 | Heat map of the numbers of variants used in TADA analysis from each data set in genes with an FDR < 0.3. Left, variants in affected subjects; right, unaffected subjects. For the counts, we only included *de novo* LoF and Mis3 variants, transmitted/untransmitted and case-control LoF variants. These variant counts are normalized by the length of coding regions of each gene and sample size of each data set (|trio| + |case| for the left, |trio| + |control| for the right). Description of the samples can be found in Supplementary Table 1.



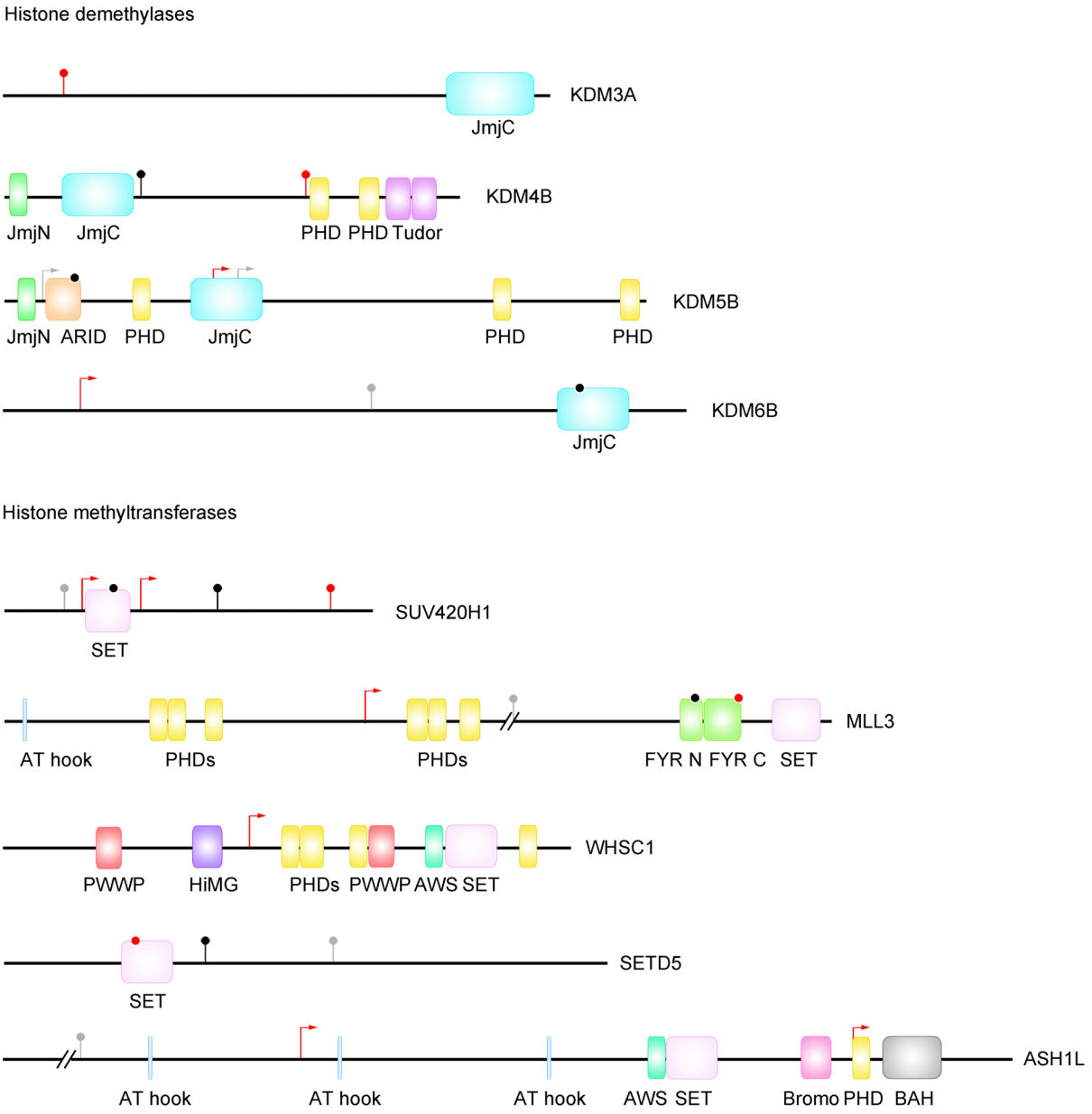
Extended Data Figure 4 | Genome browser view of the CNV deletions identified in ASD-affected subjects. The deletions are displayed in red if with unknown inheritance, in grey if inherited, and in black in unaffected subjects. Deletions in parents are not shown. For deletions within a single gene, all splicing isoforms are shown.



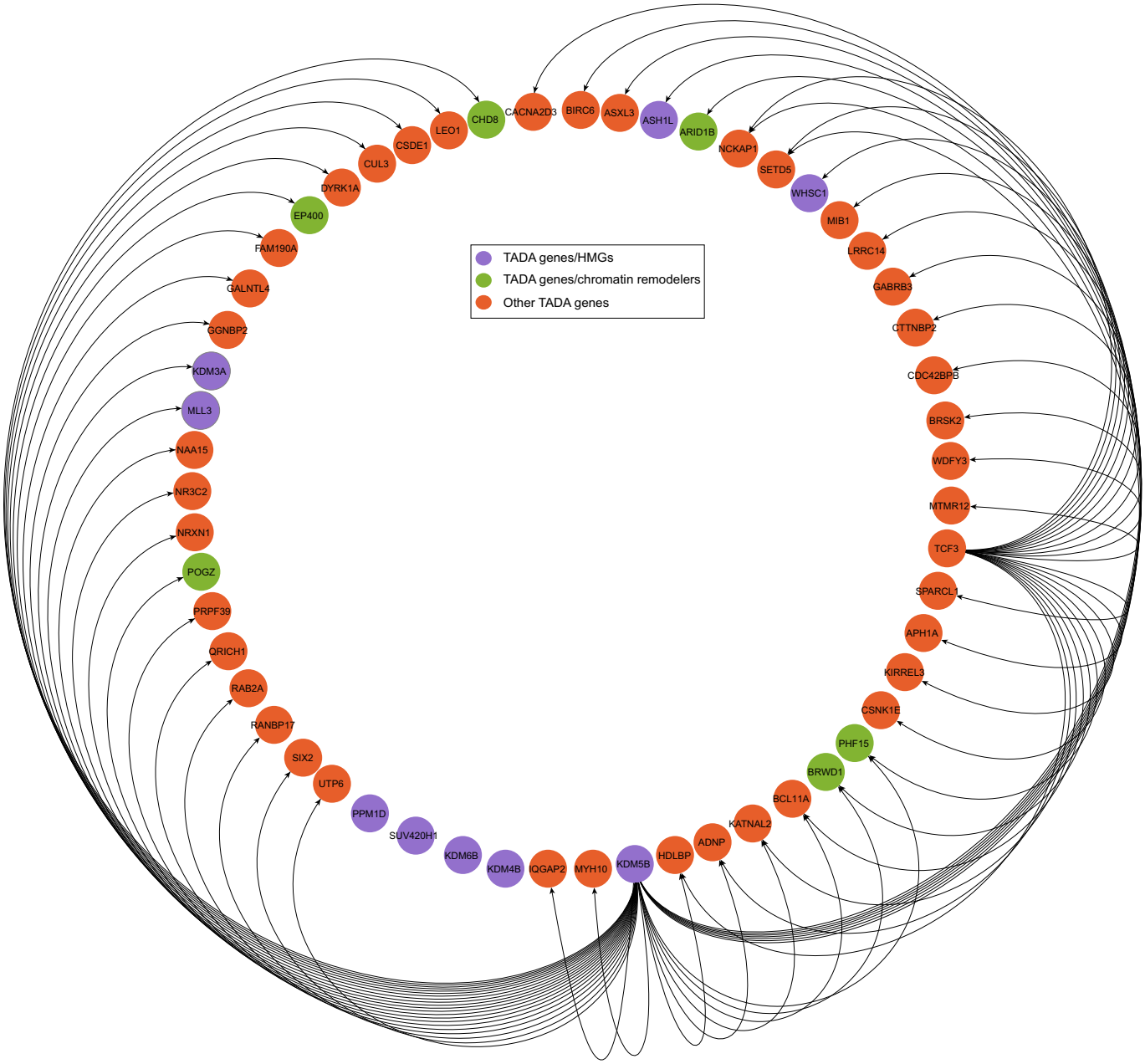
Extended Data Figure 5 | Frequency of variants by gender. Frequency of *de novo* (dn) and transmitted (Tr) variants per sample in males (black) and females (white) for genes with an FDR < 0.1 (top row), FDR < 0.3 (middle row), or all TADA genes (bottom row). The *P* values were determined by one-tailed permutation tests (*$P < 0.05$; **$P < 0.01$; ***$P < 0.01$).



Extended Data Figure 6 | Enrichment terms for the four clusters identified by protein–protein interaction networks. *P* values calculated using mouse-genome-informatics–mammalian-phenotype (MGI_Mammalian phenotype, blue), Kyoto encyclopaedia of genes and genomes (KEGG) pathways (red), and gene ontology biological processes (yellow) are indicated.



Extended Data Figure 7 | *De novo* variants in SET lysine methyltransferases and jumonji lysine demethylases. Mis3 variants are in black, LoF in red, and variants identified in other disorders in grey (Fig. 5). ARID, AT-rich interacting domain; AWS, associated with SET domain; BAH, bromo adjacent homology; bromo, bromodomain; FYR C, FY-rich C-terminal domain; FYR N, FY-rich N-terminal domain; HiMG, high mobility group box; JmjC, jumonji C domain; JmjN, jumonji N domain; PHD, plant homeodomain; PWWP, Pro-Trp-Trp-Pro domain; SET, Su(var)3-9, enhancer-of-zeste, trithorax domain.



Extended Data Figure 8 | Transcription regulation network of TADA genes only. Edges indicate transcription regulators (source nodes) and their gene targets (target nodes) based on the ChEA network.

Extended Data Table 1 | CNVs hitting TADA genes

Gene	ASD subject		Unaffected parent*			Unaffected	Odds Ratio†
	Unknown Inheritance	Inherited	Tr-ASD	NT	Tr-not-ASD		
q-value < 0.1							
ANK2	1						∞
ASXL3	1						∞
VIL1		1	1				1.49
0.1 ≤ q-value < 0.3: Evidence for role in ASD							
UTP6	1						∞
DNAH10		1	1				1.49
ATP1B1	1						∞
GGNBP2	1						∞
NRXN1		2	1				2.99
WHSC1	1						∞
HDLBP‡	1	2	1		1	1	2.24
CERS4		1	1				1.49
SHANK3	4						∞
IQGAP2	1						∞
0.1 ≤ q-value < 0.3: Evidence against role in ASD							
EP400						1	0
SLCO1B1 ‡ §	1	1	1	1		1	0.996
SLCO1B3 §		1	1	2		1	0.37
KDM6B						1	0

Count of deletion CNVs inferred from sequence for ASD subjects and those unaffected by ASD. Number of subjects and family status: 849 ASD subjects without family information; 1,467 ASD subjects in families; 2,766 unaffected parents; 319 unaffected siblings of ASD subjects; 373 unaffected subjects without family information. NT, parent a carrier but CNV not transmitted to affected child; Tr-ASD, transmitted to ASD subject from carrier parent; Tr-not-ASD, parent transmits a CNV to an unaffected child.

* No parents in this count were affected; seven parents in the study were affected, none carried a CNV reported in the table and these subjects did not enter the calculation.

† To compute the odds ratio we count the number of affected carriers (*a*), unaffected carriers (including parents) (*b*), affected subjects who do not have the CNV (*c*), and unaffected non-carriers (*d*). The odds ratio = $(ad)/(bc)$.

‡ One parent transmits the CNV to an affected and unaffected offspring; to obtain the total count of controls with a CNV, subtract one.

§ Genes are adjacent in the genome (see Extended Data Fig. 4). For three subjects both genes are affected by the same CNV (1 ASD and 2 unaffected subjects).