Synaptic, transcriptional, and chromatin genes disrupted in autism

HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2015 May 13.

Published in final edited form as:

Nature. 2014 November 13; 515(7526): 209–215. doi:10.1038/nature13772.

Synaptic, transcriptional, and chromatin genes disrupted in autism

A full list of authors and affiliations appears at the end of the article.

Summary

The genetic architecture of autism spectrum disorder involves the interplay of common and rare variation and their impact on hundreds of genes. Using exome sequencing, 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, and 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, transcriptional, and chromatin remodeling pathways. These include voltage-gated ion channels regulating propagation of action potentials, pacemaking, and excitability-transcription coupling, as well as histone-modifying enzymes and chromatin remodelers, prominently histone post-translational modifications involving lysine methylation/demethylation.

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 on ASD neurobiology and on 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 causal evidence. *De novo* loss-of-function (LoF) single-nucleotide variants (SNV) 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

*Correspondence should be addressed to J.D.B. (joseph.buxbaum@mssm.edu) or M.J.D. (mjdaly@broadinstitute.org)..
§Full list of members in the Supplementary Notes.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

The authors have no competing interests

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., AE.C, D.J.C., M.J.D., S.D.R., B.D., M.F., SC.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.J.C., A.C., M.H.C., A.G.C., AE.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.McI., 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.KC.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

1000 ASD subjects[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 to date, analyzing 16 sample sets comprising 15,480 DNA samples (Supplementary Table 1; 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). *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 alternate alleles with a frequency of < 0.1% were identified, restricting to LoF (frameshifts, stop gains, donors/acceptor splice site mutations) or probably damaging missense (Mis3) variants (defined by PolyPhen-2[17]). 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 2270 autism trios (two parents and one affected child) carried a *de novo* LoF mutation – significantly in excess of expectation[18] (8.6%, $P<10^{-14}$) or what is observed in 510 control trios (7.1%, $P=1.6\times10^{-5}$) collected here and previously published[15]. Eighteen genes (Table 1) were hit by 2 or more *de novo* LoF mutations. These genes are all known or strong candidate ASD genes, but given the number of trios sequenced, we expect approximately two such genes by chance given gene mutability[14,18]. While we expect only 2 *de novo* Mis3 events in these 18 genes, we observe 16 ($P=9.2\times10^{-11}$, Poisson test). Because much 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 analysis framework must involve an integration of *de novo* mutation with variants observed in cases and controls and transmitted or untransmitted from carrier parents. Going 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 (for 'Transmission and *De novo* Association'), a weighted, statistical model integrating *de novo,* transmitted and case-control variation[19]. 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 modeled 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 (BF) with a corresponding False Discovery Rate or FDR q-value. This framework increases the power compared to use of *de novo* LoF alone (Extended Data Fig. 2).

TADA identified 33 autosomal genes with an FDR < 0.1 (Table 1) and 107 genes 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*[11,15], *ADNP*[12], *BCL11A*[15], and *CACNA2D3*[15,20], *CTTNBP2*[15], *GABRB3*[20], *CDC42BPB*[13], *APH1A*[14], *NR3C2*[15], *SETD5*[14,21], *TRIO*[11]); and 7 are completely novel (*ASH1L*, *MLL3*, *ETFB*, *NAA15, MYO9B, MIB1*, *VIL1*). ADNP mutations have recently been identified in 10 patients with ASD and other shared clinical features[22]. Two of the newly discovered genes, *ASH1L* and *MLL3*, converge on chromatin remodeling. *MYO9B* plays a key role in dendritic arborization[23]. *MIB1* encodes an E3 ubiquitin ligase critical for neurogenesis[24] and is regulated by miR-137[25], a microRNA that regulates neuronal maturation and is implicated in risk for schizophrenia[26].

When the WES data from genes with FDR < 0.3 were evaluated for the presence of deletion copy number variants (such CNVs are functionally equivalent to LoF mutations), 34 CNVs meeting quality and frequency constraints (Supplementary Information) were detected in 5781 samples (Extended Data Fig. 1). Of the 33 genes with FDR < 0.1, three contained deletion CNVs mapping to three ASD subjects and one parent. Of the 74 genes meeting the criterion 0.1 ≤FDR < 0.3, about a 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, Extended Data Fig. 4). Two of the 10, *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 SNV in cases versus controls or transmission of LoF 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 (e.g., LoF), some genes have large impact, others smaller, and still others have no effect at all. Yet mis-annotation of variants, among other confounds, can produce 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 so-called ‘female protective effect’). Consequentially, 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[21,27,28]. 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 at 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 (OR); hence, the effect on risk for a class of variants can be estimated from the difference in frequency between males and females.

Genes with 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 impact on liability (OR ≥20; Extended Data Fig. 5). Genes with FDR between 0.1 and 0.3, however, show substantially less enrichment for female events, consistent with a modest impact for LoF variants (OR range 2-4, whether transmitted or *de novo*) and little to no effect from Mis3 variants. The results are consistent with inheritance patterns, LoF mutations in FDR < 0.1 genes are rarely inherited from unaffected parents while those in the 0.1 < FDR < 0.3 group are far more often inherited than *de novo*.

By analyzing the distribution of relative risk over inferred ASD genes[19], 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

FDR < 0.3 gene sets are strongly enriched for genes under evolutionary constraint[18] ($P=3.0\times10^{-11}$, Fig. 1a, Supplementary Table 4), consistent with the hypothesis that heterozygous LoF mutations in these genes are ASD risk factors. Indeed 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* LoF events detected by the largest schizophrenia WES study to date[29] (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 mRNAs targeted by two neuronal RNA-binding proteins: FMRP[30] (also known as FMR1), mutated or absent in fragile X syndrome ($P=1.20\times10^{-17}$, 34 targets[30], of which 11 are corroborated by an independent data set[31]), 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[32,33] (Fig. 1a). These two pathways expand the complexity of the ASD neurobiology to post-transcriptional events, including splicing and translation, both of which would sculpt the neural proteome.

We found nominal enrichment for human orthologs of mouse genes encoding synaptic (P=0.031) and postsynaptic density (PSD) proteins[34] (P=0.046, Fig. 1a, 1b, Supplementary Tables 4, 5 and 6). Enrichment analyses for InterPro, SMART, or Pfam domains (FDR < 0.05 and a minimum of 5 genes per category) reveal an overrepresentation of DNA/histone-related domains: 8 genes encoding proteins with InterPro zinc finger (Znf) FYVE PHD domains (142 such annotated genes in the genome; $FDR=7.6\times10^{-4}$), and five with Pfam Su(var)3-9, Enhancer-of-zeste (SET) domains (39 annotated in the genome; $FDR=8.2\times10^{-4}$).

Integrating complementary data

To implicate additional genes in risk for ASD, we use a model called DAWN[35]. DAWN evokes a hidden Markov random field framework to identify clusters of genes that show strong association signal and highly correlated co-expression in a key tissue and developmental context. Previous research suggests human mid-fetal prefrontal and motor-somatosensory neocortex is such a critical nexus for risk[16], thus we evaluated gene co-expression data from that tissue together with TADA scores for genes with FDR < 0.3. Because this list is enriched for genes under evolutionary constraint, we generalized DAWN to incorporate constraint scores (Supplementary Information). When (a) TADA results, (b) gene co-expression in mid-fetal neocortex, and (c) constraint scores are jointly modeled, DAWN identifies 160 genes that plausibly affect risk (Fig. 2), 91 of which are not in the top 107 TADA genes. 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 clusters based on network connectivity (Fig. 2). We visualized the enriched pathways and biological functions for each of these clusters on canvases[36] (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*, inferred by DAWN. The enrichment results for C2 indicate that genes implicated in neurodegenerative disorders could also play a role in neurodevelopmental disorders.

Emergent results

Amongst critical synaptic components found mutated in our study are voltage-gated ion channels involved in fundamental processes including propagation of action potentials (e.g., $Na_v1.2$ channel), neuronal pacemaking, and excitability-transcription coupling (e.g., $Ca_v1.3$ channel) (Fig. 1b). We identified, 4 LoF and 5 Mis3 variants in *SCN2A* ($Na_v1.2$), 3 Mis3 in *CACNA1D* ($Ca_v1.3$), 2 LoF in *CACNA2D3* (α-δauxiliary subunits of L-type voltage-gated Ca^{2+} channels, including $Ca_v1.3$). Remarkably, three *de novo* Mis3 variants in *SCN2A* hit residues mutated in homologous genes in patients with other syndromes, including Brugada syndrome (*SCN5A*) or epilepsy disorders (*SCN1A*) (p.R379H and p.R937H). These arginines, as well as the threonine mutated in p.T1420M, cluster to the P-loops forming the ion selectivity filter, in proximity of the inner ring (DEKA motif) (Fig. 1c). Because homologous channels mutated in these arginines do not conduct inwards Na^+ currents[37,38], p.R379H and p.R937H might have similar effect.

Two *de novo CACNA1D* variants (p.G407R and p.A749G) hit positions proximal to residues mutated in patients with primary aldosteronism and neurological deficits (Fig. 1d). The

reported mutations interfere with channel activation and inactivation[39]. Amongst variants found in cases, p.A59V maps to the NSCaTE domain, also important for Ca^{2+}-dependent inactivation, while p.S1977L and p.R2021H co-cluster in the C-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 remodeling involves histone-modifying enzymes (encoded by histone modifier genes, HMGs) and chromatin remodelers ('readers') that recognize specific histone post-translational modifications (PTMs) and orchestrate their effects on chromatin. Our gene set is enriched in HMGs (9 HMGs out of 152 annotated in *HI*stome[40], Fisher's exact test, $P=2.2\times10^{-7}$). Enrichment in the GO term 'histone-lysine N-methytransferase activity' (5 genes out of 41 so annotated; $FDR=2.2\times10^{-2}$) 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 (JmjC) lysine demethylases, and two are readers (Fig. 3a). RBFOX1 co-isolates with H3K4me3[41], and our dataset is enriched in targets shared by RBFOX1 and H3K4me3 (P=0.0166, Fig. 1a, 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 loss of function 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)[42]. We found that 46 TADA genes 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).

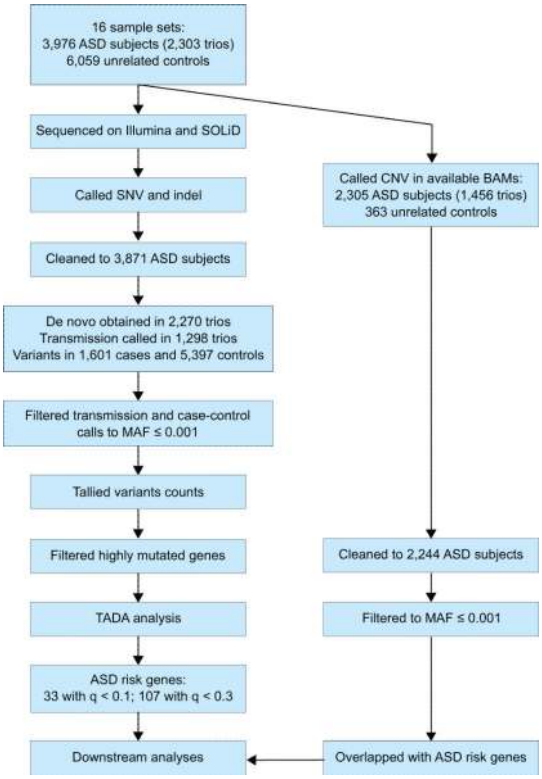
Conclusions

Complementing earlier reports, ASD subjects show a clear excess of *de novo* LoF mutations over expectation, with a pile-up of such events in a handful of genes. While this handful has a large effect on risk, most ASD genes have much smaller impact. This gradient emerges most strikingly 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 SNV on risk; for risk generated by LoF variants transmitted from

unaffected parents; and for the value of case-control design in gene discovery. Indeed, 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 alone. Almost uniformly ASD genes show large constraint against variation, a feature we exploit to implicate other genes in risk.

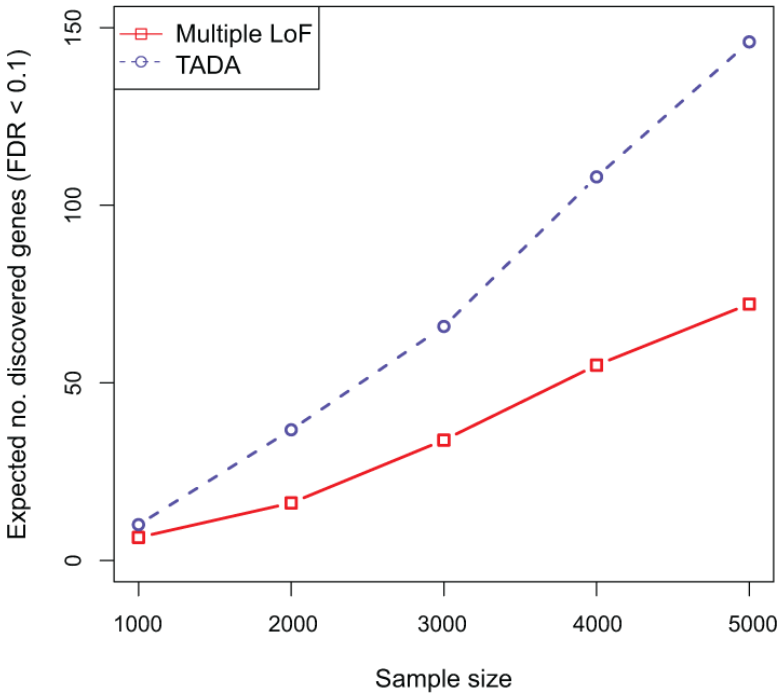
Three critical pathways for typical development are damaged by risk variation: (1) chromatin remodeling, (2) transcription and splicing, and, (3) synaptic function. Chromatin remodeling controls events underlying the formation of neural connections, including neural neurogenesis and neural differentiation[43], and relies on epigenetic marks as histone PTMs. 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 hits 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 conjecture 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[44] and schizophrenia[29] cluster to synaptic genes, and synaptic defects have been reported in models of these disorders[45]. Integrity of synaptic function is essential for neural physiology, and its perturbation could represent the intersection between diverse neuropsychiatric disorders[46].

Extended Data



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 QC 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 were obtained for trio data (published *de novo* from 825 trios[11,13-15] were incorporated). This path led to the TADA analysis, which found 33 ASD risk genes with $q < 0.1$; and 107 with $q < 0.3$. CNV were called in 2,305 ASD subjects.



Extended Data Figure 2. Expected number of ASD genes discovered as a function of sample size

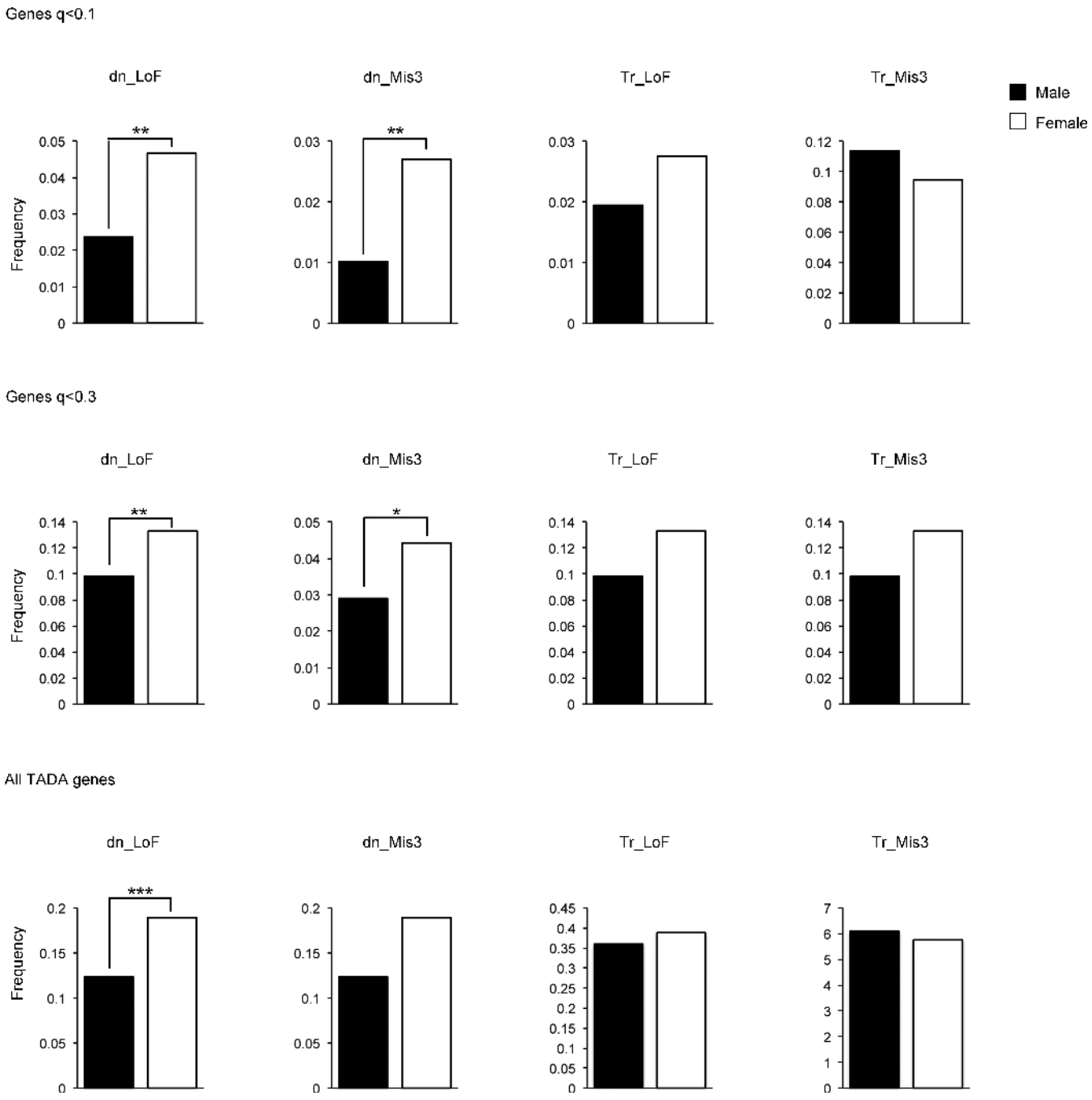
The Multiple LoF test (red) is a restricted version of TADA that uses only the *de novo* LoF data. TADA (blue) models *de novo* LoF, *de novo* Mis3, LoF variants transmitted/not transmitted and LoF variants observed in case/control samples. The sample size (N) indicates either (i) N trios, for which we record *de novo* and transmitted variation, or (ii) N trios, for which we record only *de novo* events, plus N cases and N controls.

Extended Data Figure 3. Heat map of the numbers of variants used in TADA analysis from each dataset in genes with q < 0.3

Left panel, variants in affected subjects; right panel, unaffected subjects. For the counts, we only focus on *de novo* LoF and Mis3 variants, transmitted/un-transmitted and case/control LoF variants. These variant counts are normalized by the length of coding regions of each gene and sample size of each dataset (|trio|+|case| for left panel, |trio|+|control| for the right panel).

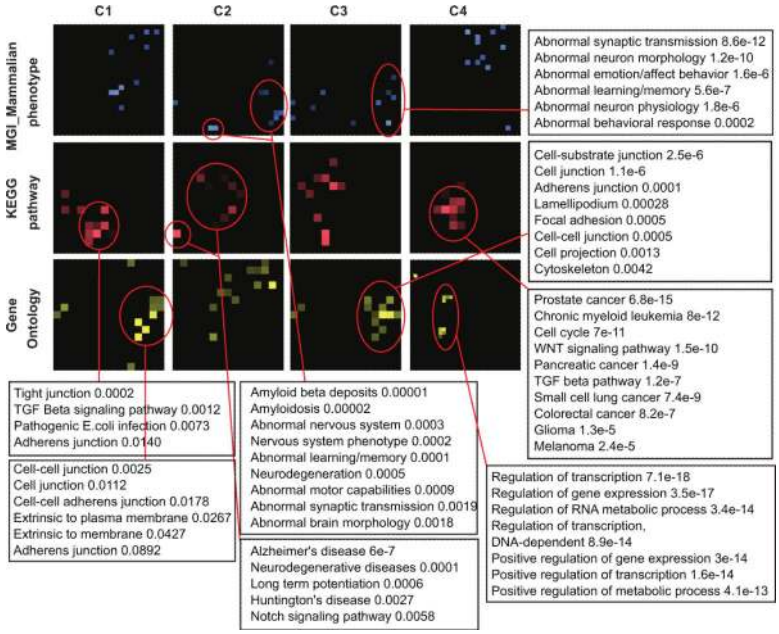
Extended Data Figure 4. Genome browser view of the CNV deletions identified in ASD affected subjects

The deletions are displayed in red if with unknown inheritance, in grey if inherited, and in black in un unaffected subjects. Deletions in parents are not shown. For deletions within a single gene, all splicing isoforms are shown.



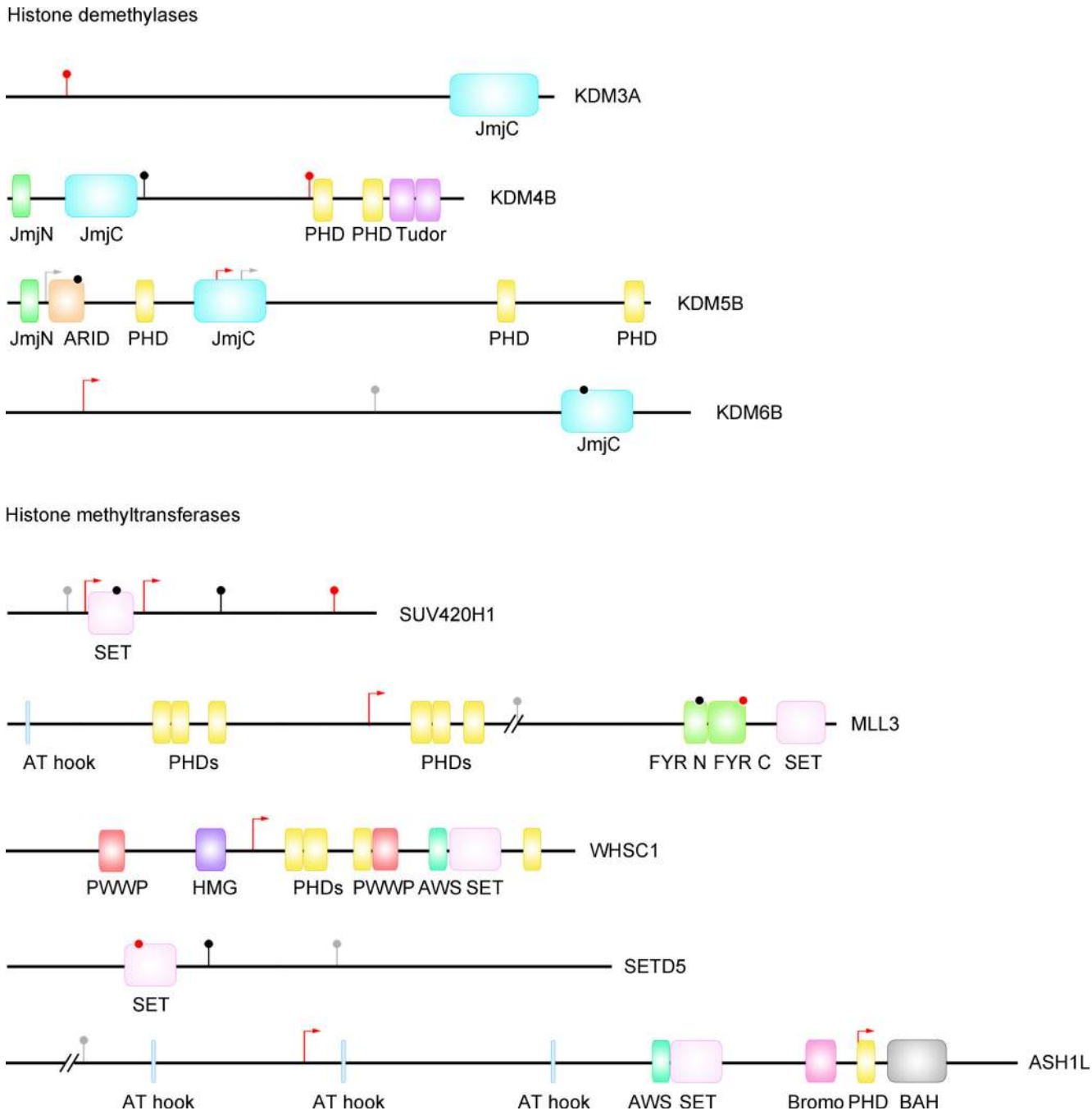
Extended Data Figure 5. Frequency of variants by gender

Frequency of *de novo* (DN) and transmitted (TR) variants per sample in males (black) and females (white) for genes with q < 0.1 (upper panel), q < 0.3 (central panel), or all TADA genes (lower panel). The P values were determined by a one-tailed permutation test (*P < 0.5; **P < 0.01; ***P < 0.01).



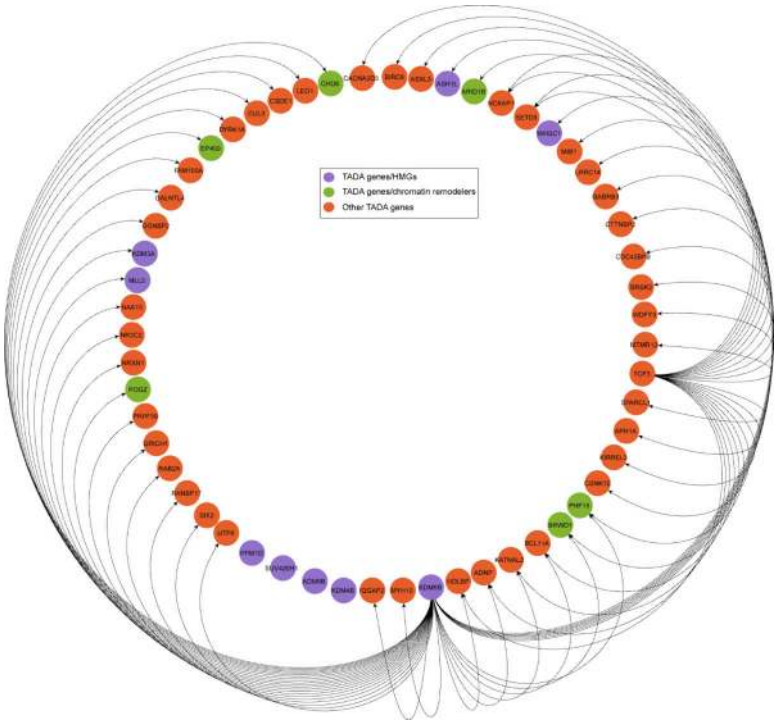
Extended Data Figure 6. Enrichment terms for the four clusters identified by protein-protein interaction network

P-values using Mouse-Genome-Informatics/Mammalian-Phenotype (MGI-MP, blue), Kyoto Encyclopedia of Genes and Genomes pathways (KEGG, red), and Gene Ontology biological processes (GO, yellow) are indicated.



Extended Data Figure 7. *De novo* variants in SET lysine methyltransferases and JmjC lysine demethylases

Mis3 are in black, LoF in red, and variants identified in other disorders in grey (Fig. 5). JmjC, Jumonji C domain; JmjN, Jumonji N domain; JmjC, PHD, plant homeodomain; ARID, AT-rich interacting domain; SET, Su(var)3-9, Enhancer-of-zeste, Trithorax domain; FYR N, FY-rich N-terminal domain; FYR C, FY-rich C-terminal domain; PWWP, Pro-Trp-Trp-Pro domain; HMG, high mobility group box; AWS, associated with SET domain; Bromo, bromodomain; BAH, bromo adjacent homology.



Extended Data Figure 8. Transcription regulation network of TADA genes only

Edges indicate transcription regulator (source node) and its gene targets (target node) based on ChEA network.

Extended Data Table 1

CNVs hitting TADA genes.

Gene	ASD subject		Unaffected parent[2]			Unaffected	Odds[4] Ratio
	Unknown Inheritance	Inherited	Tr-ASD[3]	NT[3]	Tr-not-ASD[3]		
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[5]	1	2	1		1	1	2.24
CERS4		1	1				1.49
SHANK3	4						∞

Gene	ASD subject		Unaffected parent[2]			Unaffected	Odds[4] Ratio
	Unknown Inheritance	Inherited	Tr-ASD[3]	NT[3]	Tr-not-ASD[3]		
IQGAP2	1						∞
0.1 ≤q-value < 0.3: Evidence against role in ASD							
EP400						1	0
SLCO1B1[5,6]	1	1	1	1		1	0.996
SLCO1B3[6]		1	1	2		1	0.37
KDM6B						1	0

Count of deletion copy number variants, inferred from sequence, for ASD subjects and those unaffected by ASD. Number of subjects and family status: 849 ASD without family information; 1467 ASD subjects in families; 2766 unaffected parents; 319 unaffected siblings of ASD subjects; 373 unaffected subjects without family information.

[2]No parents in this count were affected; 7 parents in the study were affected, none carried a CNV reported in the table and these subjects did not enter the calculation.

[3]Tr-ASD = transmitted to ASD subject from carrier parent; NT=parent a carrier but CNV not transmitted to affected child; Tr-not-ASD = parent transmits a CNV to an unaffected child.

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

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

[6]Genes are adjacent in the genome (see Extended Data Fig. 4). For 3 subjects both genes are hit by the same CNV (1 ASD and 2 unaffected subjects).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

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 J. 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,*], and Joseph D. Buxbaum[1,2,4,53,57,58,*]

Affiliations

[1]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [3]Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [5]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [6]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, 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, United Kingdom. [9]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. [10]Department of Psychiatry, Graduate School of Medicine, Nagoya University, Nagoya, Japan. [11]Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, University Medical Center Freiburg; Center for Mental Disorders, Freiburg, Germany. [12]Department of Child Psychiatry & SGDP Centre, King's College London Institute of Psychiatry, Psychology & Neuroscience, London, United Kingdom. [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, USA. [15]Genomic Medicine Group. University of Santiago de Compostela and Galician Foundation of Genomic Medicine (SERGAS). Santiago de Compostela, Spain. [16]Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, KSA. [17]Harvard Medical School, Boston, Massachusetts, USA. [18]Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, USA. [19]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt, Frankfurt, Germany. [20]Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA. [21]Department of Psychiatry, University of Utah, Salt Lake City, Utah, USA. [22]Duke Institute for Brain Sciences, Duke University, Durham, North Carolina, 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, Ireland. [25]University of Pennsylvania Perelman School of Medicine, Department of Pathology and Laboratory Medicine, Philadelphia, Pennsylvania 19104, USA [26]Institute for Juvenile Research, Department of Psychiatry, University

of Illinois at Chicago, Chicago, Illinois, USA. [27]Department of Biostatistics, Columbia University, New York, New York, 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, United Kingdom. [30]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany. [31]Department of Pediatrics, Icahn School of Medicine at Mount Sinai , New York, New York, USA. [32]Institute of Child Health, University College London, London WC1N 1EH, United Kingdom. [33]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland SF-33100. [34]Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [35]Department of Psychiatry Kaiser Permanente, San Francisco, USA. [36]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [37]MRC Centre for Neuropsychiatric Genetics and Genomics, and the Neuroscience and Mental Health Research Institute, Cardiff University. [38]Child and Adolescent Psychiatry Department, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, Universidad Complutense, Madrid, Spain. [39]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK. [40]Department of Child Psychiatry, University of Tampere and Tampere University Hospital, Tampere, Finland SF-33101. [41]Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [42]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA. [43]Department of Psychiatry, University of California at San Francisco, San Francisco, California, 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, Aachen, Germany. [45]Department of Child and Adolescent Mental Health, Great Ormond Street Hospital for Children, National Health Service Foundation Trust, London, United Kingdom. [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, 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, USA. [50]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [51]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [52]Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [53]Department of Neuroscience, Icahn School of Medicine at Mount Sinai , New York, New York, USA. [54]McLaughlin Centre, University of Toronto, Toronto, Ontario M5S 1A1, Canada. [55]Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA. [56]Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [57]Friedman Brain Institute, Icahn School of

Medicine at Mount Sinai, New York, New York, USA. [58]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

Acknowledgements

This work was supported by 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 RO1MH083565 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 Hospital, Crumlin, The Meath Foundation, AMNCH, Tallaght The Health Research Board, Ireland Autism Speaks, US. 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 an UK 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 timely assistance of Dan 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, without whose participation this study would not have been possible.

References

1. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: a decade of new twin studies. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2011; 156B:255–274.
2. Sebat J, et al. Strong association of de novo copy number mutations with autism. Science. 2007; 316:445–449. [PubMed: 17363630]
3. Pinto D, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010; 466:368–372. [PubMed: 20531469]
4. Klei L, et al. Common genetic variants, acting additively, are a major source of risk for autism. Molecular autism. 2012; 3:9. [PubMed: 23067556]
5. Gaugler T, et al. Most inherited risk for autism resides with common variation. Nature genetics. (in press).
6. Yu TW, et al. Using whole-exome sequencing to identify inherited causes of autism. Neuron. 2013; 77:259–273. [PubMed: 23352163]
7. Lim ET, et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron. 2013; 77:235–242. [PubMed: 23352160]
8. Poultney CS, et al. Identification of Small Exonic CNV from Whole-Exome Sequence Data and Application to Autism Spectrum Disorder. American journal of human genetics. 2013; 93:607–619. [PubMed: 24094742]
9. Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain research. 2011; 1380:42–77. [PubMed: 21129364]
10. Glessner JT, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature. 2009; 459:569–573. [PubMed: 19404257]

11. Sanders SJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012; 485:237–241. [PubMed: 22495306]
12. O'Roak BJ, et al. Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders. Science. 2012
13. O'Roak BJ, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012; 485:246–250. [PubMed: 22495309]
14. Neale BM, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012; 485:242–245. [PubMed: 22495311]
15. Iossifov I, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. 2012; 74:285–299. [PubMed: 22542183]
16. Willsey AJ, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013; 155:997–1007. [PubMed: 24267886]
17. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010; 7:248–249. [PubMed: 20354512]
18. Samocha KE, et al. A framework for the interpretation of de novo mutation in human disease. Nature genetics. 2014
19. He X, et al. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS genetics. 2013; 9:e1003671. [PubMed: 23966865]
20. Girirajan S, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. American journal of human genetics. 2013; 92:221–237. [PubMed: 23375656]
21. Pinto D, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. American journal of human genetics. 2014; 94:677–694. [PubMed: 24768552]
22. Helsmoortel C, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nature genetics. 2014; 46:380–384. [PubMed: 24531329]
23. Long H, et al. Myo9b and RICS modulate dendritic morphology of cortical neurons. Cerebral cortex. 2013; 23:71–79. [PubMed: 22250289]
24. Yoon KJ, et al. Mind bomb 1-expressing intermediate progenitors generate notch signaling to maintain radial glial cells. Neuron. 2008; 58:519–531. [PubMed: 18498734]
25. Smrt RD, et al. MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. Stem cells. 2010; 28:1060–1070. [PubMed: 20506192]
26. Ripke S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nature genetics. 2013; 45:1150–1159. [PubMed: 23974872]
27. Robinson EB, Lichtenstein P, Anckarsater H, Happe F, Ronald A. Examining and interpreting the female protective effect against autistic behavior. Proc Natl Acad Sci U S A. 2013; 110:5258–5262. [PubMed: 23431162]
28. Jacquemont S, et al. A higher mutational burden in females supports a “female protective model” in neurodevelopmental disorders. American journal of human genetics. 2014; 94:415–425. [PubMed: 24581740]
29. Fromer M, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature. 2014; 506:179–184. [PubMed: 24463507]
30. Darnell JC, et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell. 2011; 146:247–261. [PubMed: 21784246]
31. Ascano M Jr. et al. FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature. 2012; 492:382–386. [PubMed: 23235829]
32. Weyn-Vanhentenryck SM, et al. HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. Cell reports. 2014; 6:1139–1152. [PubMed: 24613350]
33. Voineagu I, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011; 474:380–384. [PubMed: 21614001]
34. Collins MO, et al. Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. Journal of neurochemistry. 2006; 97(Suppl 1):16–23. [PubMed: 16635246]

35. Liu L, et al. DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. Molecular autism. 2014; 5:22. [PubMed: 24602502]
36. Tan CM, Chen EY, Dannenfelser R, Clark NR, Ma'ayan A. Network2Canvas: network visualization on a canvas with enrichment analysis. Bioinformatics. 2013; 29:1872–1878. [PubMed: 23749960]
37. Vatta M, et al. Genetic and biophysical basis of sudden unexplained nocturnal death syndrome (SUNDS), a disease allelic to Brugada syndrome. Human molecular genetics. 2002; 11:337–345. [PubMed: 11823453]
38. Volkers L, et al. Nav 1.1 dysfunction in genetic epilepsy with febrile seizures-plus or Dravet syndrome. The European journal of neuroscience. 2011; 34:1268–1275. [PubMed: 21864321]
39. Scholl UI, et al. Somatic and germline CACNA1D calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. Nature genetics. 2013; 45:1050–1054. [PubMed: 23913001]
40. Khare SP, et al. HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic acids research. 2012; 40:D337–342. [PubMed: 22140112]
41. Feng J, et al. Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. Genome biology. 2014; 15:R65. [PubMed: 24758366]
42. Lachmann A, et al. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics. 2010; 26:2438–2444. [PubMed: 20709693]
43. Ronan JL, Wu W, Crabtree GR. From neural development to cognition: unexpected roles for chromatin. Nat Rev Genet. 2013; 14:347–359. [PubMed: 23568486]
44. Rauch A, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012; 380:1674–1682. [PubMed: 23020937]
45. Penzes P, Cahill ME, Jones KA, VanLeeuwen JE, Woolfrey KM. Dendritic spine pathology in neuropsychiatric disorders. Nat Neurosci. 2011; 14:285–293. [PubMed: 21346746]
46. Zoghbi HY. Postnatal neurodevelopmental disorders: meeting at the synapse? Science. 2003; 302:826–830. [PubMed: 14593168]

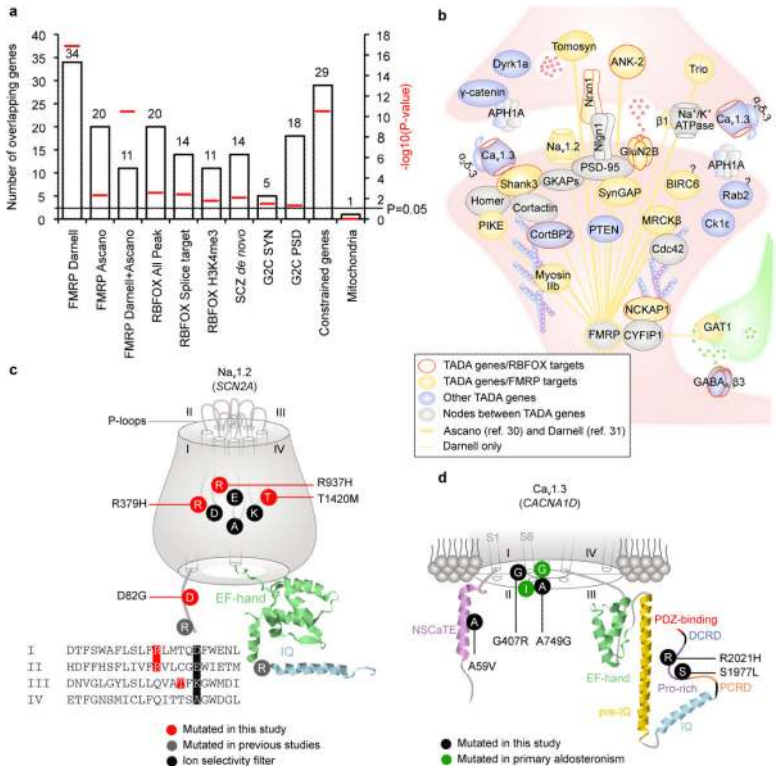


Figure 1. ASD genes in synaptic network

a. Enrichment of 107 TADA genes in: FMRP targets from two independent datasets and their overlap; RBFOX targets; RBFOX targets with predicted alterations in splicing; RBFOX and H3K4me3 overlapping targets; genes with *de novo* mutations in schizophrenia; human orthologues of Genes2Cognition mouse synaptosome or PSD genes; constrained genes; and, genes encoding mitochondrial proteins (as a control). Red bars indicate empirical P-values. **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.** Relevant variants in $Ca_V1.3$ (*CACNA1D*). Part of the channel is shown, including helices one and six (S1 and S6) for the I-IV domains, NSCaTE motif, EF-hand domain, pre-IQ, IQ, PCRD, DCRD, proline-rich region, and PDZ-binding motif.

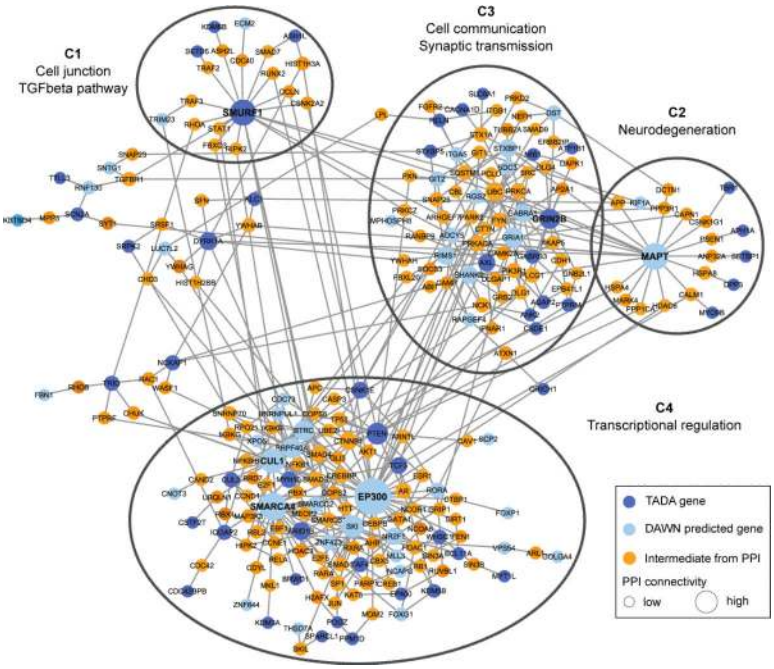


Figure 2. ASD genes in neuronal networks

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 based on degree of connectivity.

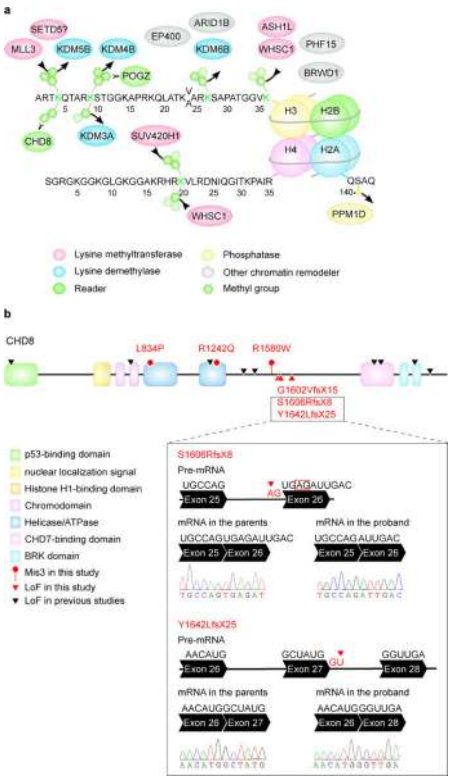


Figure 3. ASD genes in chromatin remodeling

a. TADA genes cluster to chromatin remodeling complexes. Amino terminals of histones H3, H4 and part of H2A, are shown. Lysine methyltransferases add methyl groups, while lysine demethylases remove them. **b.** *De novo* Mis3 and LoF variants in CHD8. The box shows the outcome of RT-PCR and Sanger sequencing in lymphoblastoid cells for two newly identified *de novo* splice-site variants. The first mutation hits 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 hits a donor splice site (red arrow), causing exon skipping, frame shift and a premature stop.

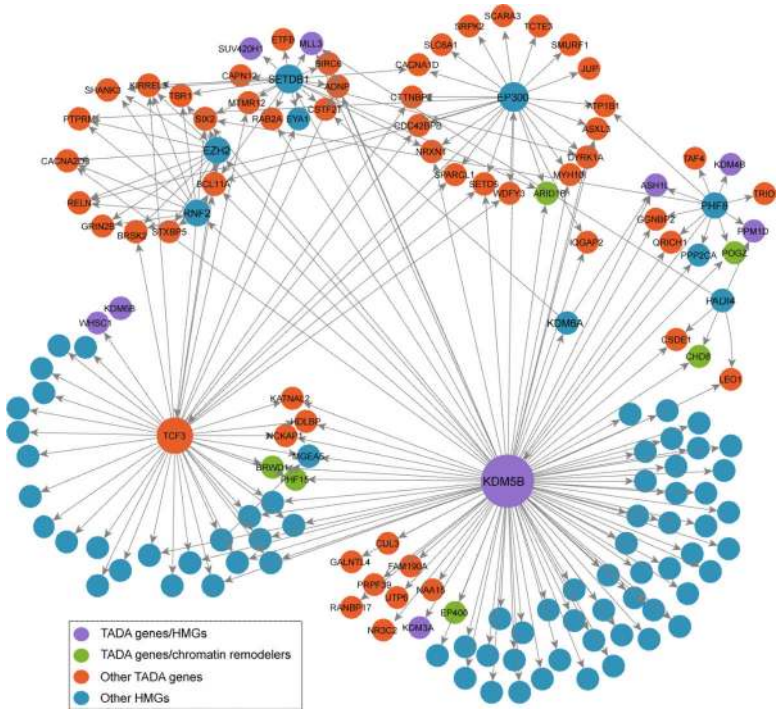


Figure 4. Transcription regulation network of TADA genes

Edges indicate transcription regulator (source node) and its gene targets (target node) based on ChEA network; interactions among only HMGs are ignored.

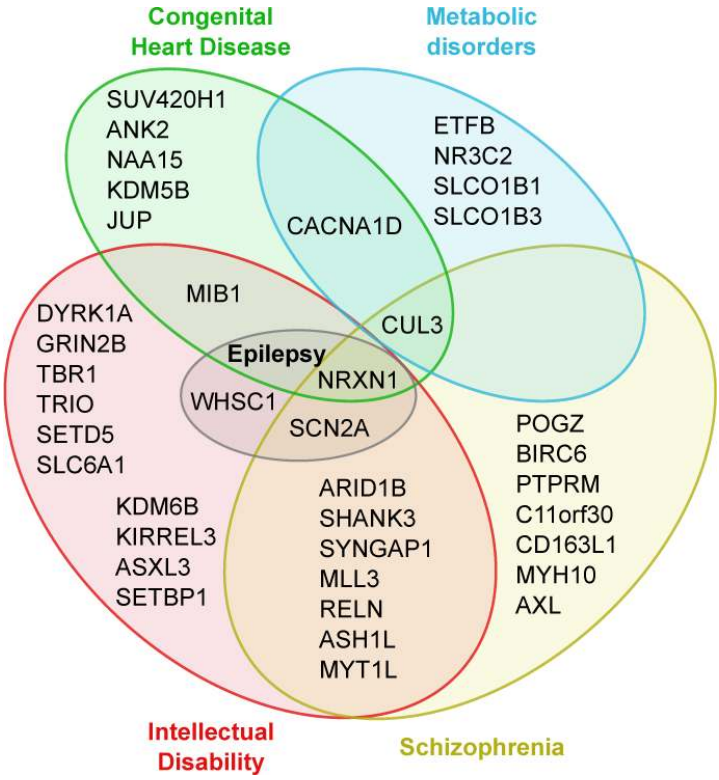


Figure 5. Involvement in disease of ASD genes

Venn diagram to visualize the overlap in disease involvement for the TADA genes.

Table 1

ASD risk genes[1].

dnLoF[2] count	q ≤0.01	0.01<q ≤0.05	0.05<q ≤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*

[1] TADA analysis of loss-of-function (LoF) and damaging missense variants found to be *de novo* in ASD subjects, inherited by ASD subjects, or in ASD subjects (versus control subjects).

[2] *De novo* LoF events.