

## GENETICS

# Systematic reconstruction of autism biology from massive genetic mutation profiles

Weijun Luo,<sup>1,2\*</sup> Chaolin Zhang,<sup>3</sup> Yong-hui Jiang,<sup>4</sup> Cory R. Brouwer<sup>1,2</sup>

Autism spectrum disorder (ASD) affects 1% of world population and has become a pressing medical and social problem worldwide. As a paradigmatic complex genetic disease, ASD has been intensively studied and thousands of gene mutations have been reported. Because these mutations rarely recur, it is difficult to (i) pinpoint the fewer disease-causing versus majority random events and (ii) replicate or verify independent studies. A coherent and systematic understanding of autism biology has not been achieved. We analyzed 3392 and 4792 autism-related mutations from two large-scale whole-exome studies across multiple resolution levels, that is, variants (single-nucleotide), genes (protein-coding unit), and pathways (molecular module). These mutations do not recur or replicate at the variant level, but significantly and increasingly do so at gene and pathway levels. Genetic association reveals a novel gene + pathway dual-hit model, where the mutation burden becomes less relevant. In multiple independent analyses, hundreds of variants or genes repeatedly converge to several canonical pathways, either novel or literature-supported. These pathways define recurrent and systematic ASD biology, distinct from previously reported gene groups or networks. They also present a catalog of novel ASD risk factors including 118 variants and 72 genes. At a subpathway level, most variants disrupt the pathway-related gene functions, and in the same gene, they tend to hit residues extremely close to each other and in the same domain. Multiple interacting variants spotlight key modules, including the cAMP (adenosine 3',5'-monophosphate) second-messenger system and mGluR (metabotropic glutamate receptor) signaling regulation by GRKs (G protein-coupled receptor kinases). At a superpathway level, distinct pathways further interconnect and converge to three biology themes: synaptic function, morphology, and plasticity.

## INTRODUCTION

Autism spectrum disorder (ASD) is a complex genetic disorder. Genome-wide molecular profiling is a proven strategy for complex disease studies. Thousands of genomic variants or loci have been identified as potential ASD causes (1–12). These results confirm the genetic complexity of ASD and provide valuable biological insights. However, greater challenge remains: How to turn these enormous data sets into solid and systematic understanding of the disease mechanism, or biologically relevant molecular pathways, beyond a list of associated genes, their groups, or networks (13–15)? This is a common challenge remaining for genome-wide studies or complex diseases in general, not just ASD.

Recently, two whole-exome studies under two consortia, the Simons Simplex Collection (SSC) (10, 16) and the Autism Sequencing Consortium (ASC) (11, 17), analyzed thousands of ASD families or cases-controls, producing a vast amount of genetic data. These efforts identified thousands of rare mutations and firmly established their roles in ASD (18). Because these variants rarely recur, major challenges remain when we need to (i) evaluate the disease association of individual variants, (ii) pinpoint most pathogenic or driver events from a huge pool of passengers (random mutations), (iii) replicate independent studies, and (iv) verify their results systematically. Important findings have been made, yet a systematic and precise understanding of autism biology has not been achieved with these studies (14, 19).

To address these challenges, we used a holographic (or hologram-like) approach and analyzed these data in the hierarchical context of mul-

tipole annotation levels, that is, variant, gene, and pathways. Previously, mutations/variants were divided into one-dimensional groups according to their molecular effects (loss of function, missense, synonymous, etc.) (2, 4, 10), their target gene (CHD8, DYRK1A, GRIN2B, SCN2A, etc.) (10, 20), or function annotation (FMRP targets, synaptic or schizophrenia-related genes, etc.) (10, 11). Compared to this classical one-level approach, the multilevel approach provides more sophisticated and relevant classification and prioritization of de novo (DN) mutations, which makes new and more powerful analyses possible with these rare events. Note that the multilevel approach is a generic strategy, which is used in different types of analysis in this study, including variant prioritization, association analysis, and functional analysis. Although the multilevel approach can be generally useful, this study mainly focuses on solving autism biology through these analyses rather than the methodology.

First, we inferred potential ASD genetic causes using a novel sequential prioritization procedure, which incorporates experimental data, metadata, and annotation data at different levels (Materials and Methods). Second, we did a multilevel recurrence analysis. Certain ASD genes or functional groups have been repeatedly identified in literature. However, this is the first statistical analysis revealing that ASD genetics may be systematically replicated across independent cohorts, although the mutations themselves rarely recur. We dissected ASD genetic association across multiple levels and built a more inclusive gene + pathway dual-hit model. This model better explains ASD risk than the well-accepted mutation burden model. Finally, we did multilevel function analysis on selected variants, genes, and pathways. We reconstructed replicable, systematic, and multiscale molecular mechanisms for ASD, which are either novel or consistent with literature (1–12). Unlike previously reported gene groups or networks, they reveal more interpretable disease biology with relevant context and details. They may also be instrumental for the development of effective ASD diagnostics and therapeutics.

<sup>1</sup>Department of Bioinformatics and Genomics, University of North Carolina (UNC) at Charlotte, Charlotte, NC 28223, USA. <sup>2</sup>UNC Charlotte Bioinformatics Service Division, North Carolina Research Campus, Kannapolis, NC 28081, USA. <sup>3</sup>Department of Systems Biology, Department of Biochemistry and Molecular Biophysics, Center for Motor Neuron Biology and Disease, Columbia University, New York, NY 10032, USA. <sup>4</sup>Department of Pediatrics, Department of Neurobiology, Program in Genetics and Genomics, School of Medicine, Duke University, Durham, NC 27710, USA.

\*Corresponding author. Email: weijun.luo@uncc.edu

## RESULTS

## Sequential convergence from variant to gene to pathway level

We selected ASD-related DN mutations, genes, and pathways in probands or cases in SSC or ASC studies using a sequential prioritization procedure, as described in Materials and Methods. The results are listed in Table 1 and table S4. We applied the same selection procedure to siblings in SSC for control. We then assessed the recurrence or reproducibility of the prioritized features across multiple levels.

ASD mutations from different cohorts do not replicate at the variant level, but increasingly do so at gene and pathway levels (Fig. 1A). At the variant level (Fig. 1A), 2600 of 4792 entries were selected in ASC probands (or cases), and 1213 were selected out of the total 3392 considered entries for SSC probands. Among these SSC variants, one selected versus three considered entries replicate the ASC variants (both recurrent ratios equal 0.001). The replicate entries were not enriched in the selected list versus considered list ( $P = 0.74$ ). Similar comparisons were conducted at gene and pathway levels. At these levels (Fig. 1A), not only do the recurrent ratios increase but also replicate entries are increasingly enriched in the selected list ( $P = 9.3 \times 10^{-4}$  and  $9.7 \times 10^{-6}$ , respectively). Therefore, ASD mutations show multi-level sequential convergence. Conversely, mutations from probands and siblings in the same SSC cohort do not or rarely replicate at all three levels and do not converge (Fig. 1B).

Besides direct replication, pathway-level analysis shows additional reproducibility (Fig. 1, C and D). Pathway analysis statistics ( $-\log_{10} P$ )

are highly correlated between SSC and ASC ( $R^2 = 0.570$ ), but not so ( $R^2 = 0.030$ ) between probands and siblings in SSC.

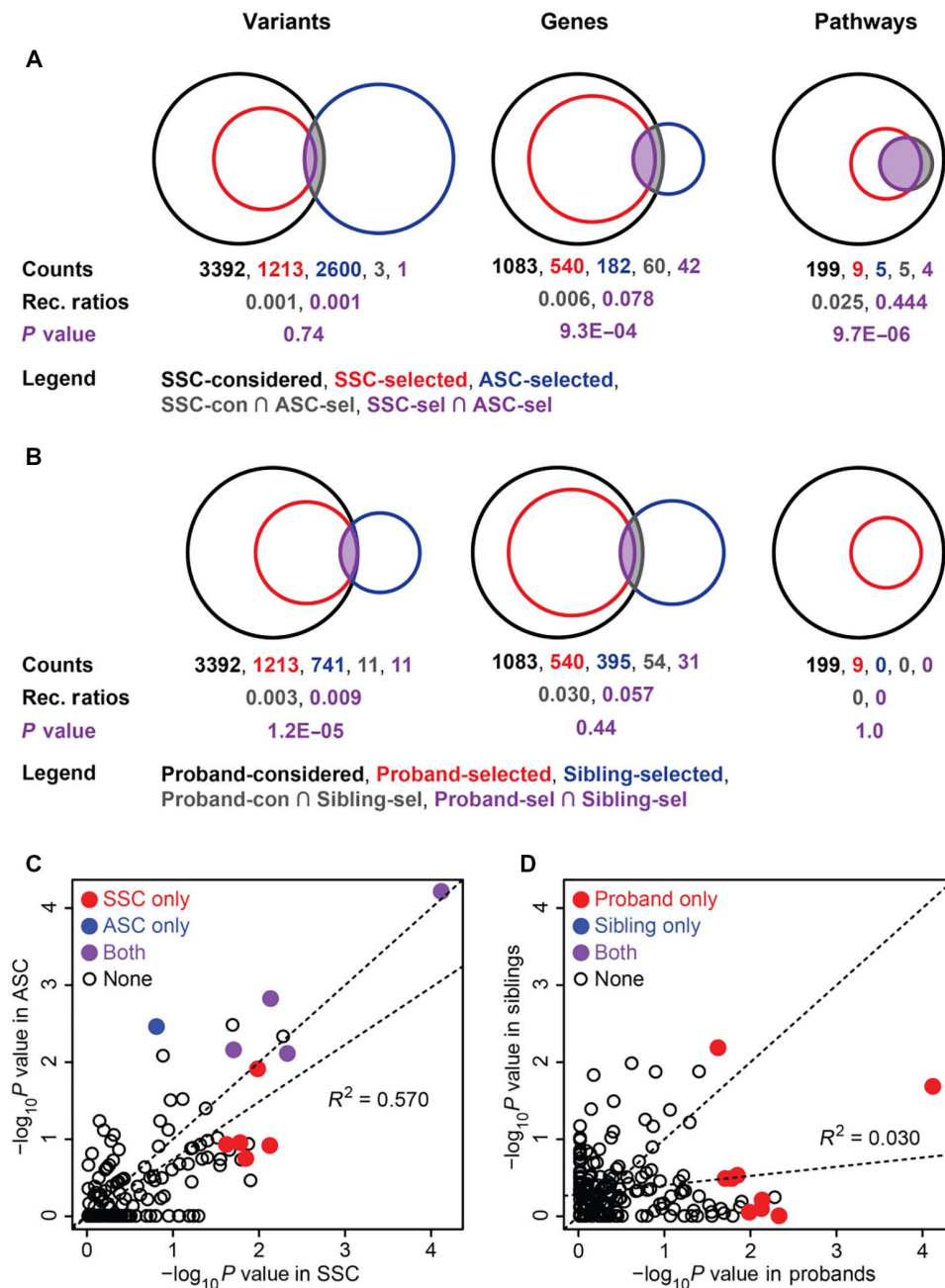
The actual replicability between studies should be even higher given that (i) the genetic background is much more divergent between different cohorts than between probands and siblings in the same cohort, and (ii) the exome-sequencing (Exome-Seq) assay and raw data processing procedures differ between the two studies.

ASD mutations within the same cohort do not recur at the variant level, but highly and increasingly do so at gene and pathway levels (fig. S1A). The analysis was done with available data from SSC (10). At the variant level, 0 of 1213 selected versus 5 of the 3392 considered variants are recurrent ( $P = 1.0$ ). At the gene level, 107 of 540 selected versus 110 of the 1083 considered genes are recurrent ( $P = 5.0 \times 10^{-31}$ ). At the pathway level, 9 of 9 selected versus 22 of the 199 considered pathways are recurrent ( $P = 8.9 \times 10^{-9}$ ). Conversely, mutations in siblings in the same SSC cohort do not recur or less so at all three levels (fig. S1B).

Gene- and pathway-level analysis shows additional evidence of recurrence. As described above, no variants are recurrent per se, but 240 variants come from recurrent genes in probands. These gene-level recurrent variants are enriched in both the selected genes and pathways (fig. S1C). In selected genes versus in other genes and in siblings, such recurrent events are 31.9- and 2.53-fold enriched ( $P < 0.001$ ; fig. S1C, Gene level). In selected pathways versus outside the pathways and in siblings, recurrent events are 2.08 and 3.21 times enriched ( $P < 0.001$ ; fig. S1C, Pathway level). For siblings, the selected genes, but not the selected pathways, are enriched for recurrent variants.

**Table 1. Significant pathways selected from SSC and ASC exome data.** Columns  $P$  and  $q$  are the  $P$ - and  $q$ -values from marginal tests, and  $P_{\text{con}}$  and  $q_{\text{con}}$  are from the conditional tests. Columns  $j$  and  $k$  are the marginal and conditional counts of selected genes. These pathways are likely drivers or disease-causing pathways because of the special analysis procedure (Materials and Methods). See table S4 for full lists of selected variants and genes.

	$P$	$q$	$P_{\text{con}}$	$q_{\text{con}}$	$j$	$k$	Reference
SSC pathways							
hsa00310 Lysine degradation	$8.0 \times 10^{-5}$	0.02	$8.0 \times 10^{-5}$	0.02	8	8	(11, 44)
hsa04727 GABAergic synapse	$4.9 \times 10^{-3}$	0.31	$3.9 \times 10^{-3}$	0.03	8	8	(45, 46)
hsa04974 Protein digestion and absorption	$1.5 \times 10^{-2}$	0.35	$9.7 \times 10^{-3}$	0.07	7	7	(35)
hsa04310 Wnt signaling pathway	$7.9 \times 10^{-3}$	0.31	$8.5 \times 10^{-3}$	0.08	10	9	(47, 48)
hsa04810 Regulation of actin cytoskeleton	$2.5 \times 10^{-2}$	0.35	$6.3 \times 10^{-3}$	0.06	12	12	(49, 50)
hsa04010 MAPK signaling pathway	$2.1 \times 10^{-2}$	0.35	$9.1 \times 10^{-3}$	0.07	14	10	(51, 52)
hsa04530 Tight junction	$1.7 \times 10^{-2}$	0.35	$6.6 \times 10^{-2}$	0.17	9	5	(53, 54)
hsa04713 Circadian entrainment	$7.7 \times 10^{-3}$	0.31	$6.6 \times 10^{-2}$	0.14	8	3	(34, 55)
hsa04724 Glutamatergic synapse	$2.1 \times 10^{-2}$	0.35	$1.9 \times 10^{-1}$	0.19	8	2	(56, 57)
ASC pathways							
hsa00310 Lysine degradation	$6.2 \times 10^{-5}$	0.01	$6.2 \times 10^{-5}$	0.01	5	5	(11, 44)
hsa04713 Circadian entrainment	$1.5 \times 10^{-3}$	0.15	$9.9 \times 10^{-4}$	0.01	5	5	(34, 55)
hsa04976 Bile secretion	$3.5 \times 10^{-3}$	0.18	$8.4 \times 10^{-3}$	0.06	4	3	(35)
hsa04727 GABAergic synapse	$7.8 \times 10^{-3}$	0.21	$4.4 \times 10^{-2}$	0.12	4	2	(45, 46)
hsa04010 MAPK signaling pathway	$7.1 \times 10^{-3}$	0.21	$6.7 \times 10^{-2}$	0.07	7	4	(51, 52)



**Fig. 1. Multilevel comparison of DN mutations between or within ASD whole-exome studies.** Venn diagrams and test statistics on overlap (A) between SSC and ASC ASD cohorts, and (B) between probands and siblings of SSC, at different levels. Correlation of pathway analysis statistics (C) between SSC and ASC ASD cohorts, and (D) between probands and siblings of SSC. The term “considered” or “selected” refers to items before or after the selection process at each level (Materials and Methods). See table S4 for full lists of selected variants and genes used in the analysis.

Three potential factors contribute to the increasing recurrence rate from variant to gene to pathway level: (i) the background space (that is, the size of the considered list) shrinks, (ii) high local mutation rate due to genomic and epigenomic features, and (iii) true and recurrent disease mechanisms. All potential factors are identical in the siblings/control group and proband/case group except factor (iii). Therefore, the recurrence observed in proband/case mutations, but not in sibling/control mutations, is ASD/disease-related. The higher-level recurrence and replication between SSC and ASC probands, but not for SSC siblings, (i) indicate that

our multilevel prioritization procedure is both sensitive and selective and (ii) suggest that our results are likely true and generalizable.

### Autism genetic association dissected across multiple levels

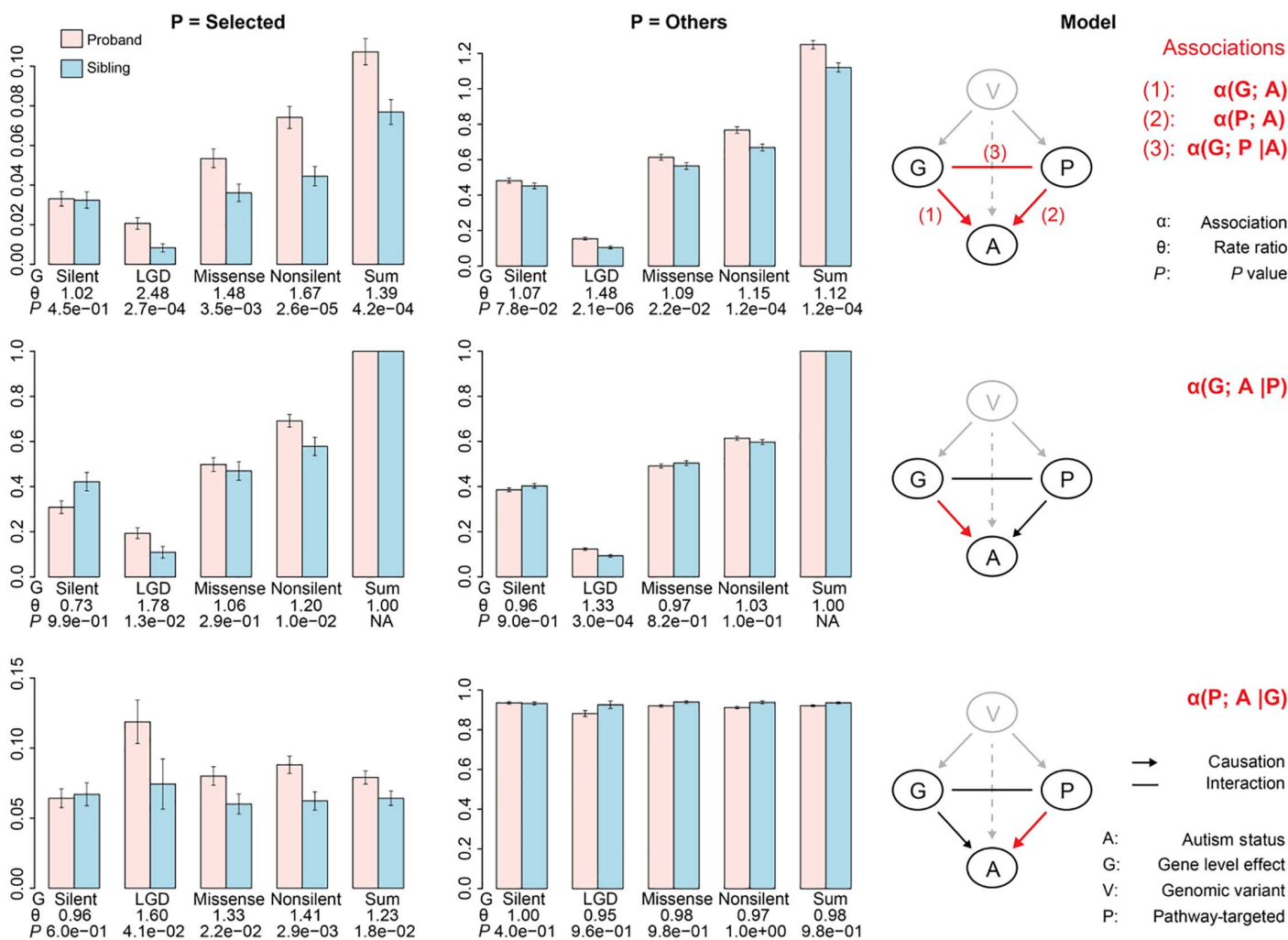
In this and following sections, we work only with the SSC data unless noted otherwise. The study is well controlled with simple data structure (10), ideal for association and function analyses. For association analysis, we use all DN variants for testing power. For pathway and function analyses, we focus on validated variants only (Materials and Methods).

Without recurrence or annotation, the DN events tell little about ASD genetics at the variant level alone. To fully dissect the ASD genetic association, we take their gene- and pathway-level effects into account. Variant effects at these two levels largely determine its association with ASD: (i) whether (and how much) the variant disrupts genes and (ii) whether (and potentially how much) it hits the selected pathways.

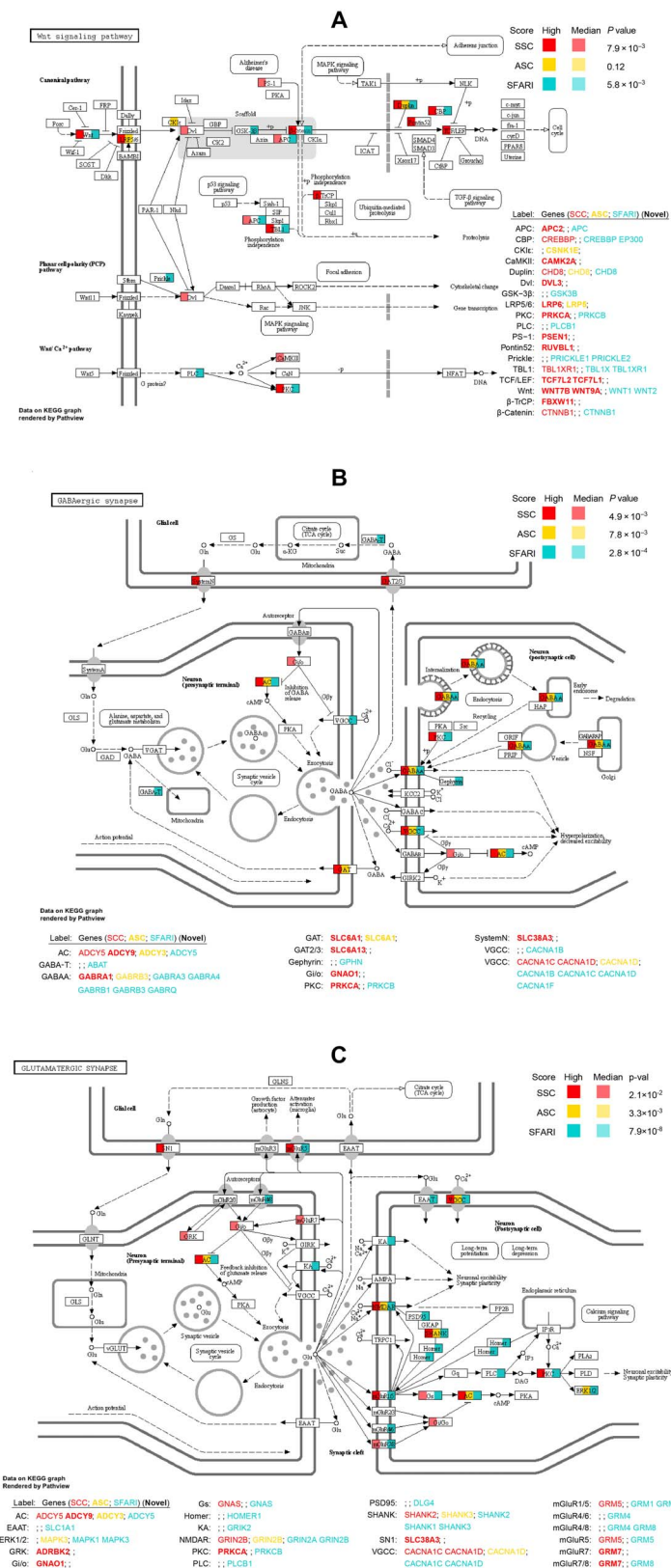
Proband have more variants in general, particularly variants that disrupt genes and hit selected pathways. Proband have 55% more LGD (likely gene-disrupting, or LoF) ( $0.175$  versus  $0.113$ ,  $P = 2.1 \times 10^{-6}$ ) and 11% more missense variants ( $0.667$  versus  $0.601$ ,  $P = 2.2 \times 10^{-2}$ ) than siblings, but only 6% more silent variants ( $0.515$  versus  $0.484$ ,  $P = 7.8 \times 10^{-2}$ ) (fig. S2, row 1). This is consistent with the original analysis (10). In addition, they have 39% more variants within selected pathways, but only 12% more outside (Fig. 2, row 1). That is, most of the differences between probands and siblings fall in LGD and missense

categories within selected pathways. This difference is well exhibited by variant distributions in the Wnt and synapse pathways (figs. S5 to S7, with variant types, and Fig. 3, without variant types).

Proband variants are more frequently gene-disrupting. As described above, probands have more events in LGD and missense categories and in general. After adjusting for the event numbers per pathway assignment or in total (row 2 in Fig. 2 and fig. S2), probands still have 78%, 33%, and 37% higher LGD within selected pathways, outside and together ( $P = 1.3 \times 10^{-2}$ ,  $3.0 \times 10^{-4}$ , and  $2.8 \times 10^{-5}$ , respectively). This difference is well exhibited by variant distributions in the Wnt and synapse pathways (figs. S5 to S7). The missense variant ratios are similar in probands and siblings (row 2 in Fig. 2 and fig. S2). However, greater portions of missense variants are gene-damaging in probands versus siblings ( $0.58$  versus  $0.43$  and  $0.50$  versus  $0.47$  within and outside selected pathways,  $P < 0.05$  and  $0.1$ , respectively), as predicted by SIFT



**Fig. 2. Autism genetic association analysis across variant, gene, and pathway levels with the SSC exome mutation data.** Columns 1 and 2: DN event rates and association test by gene- and pathway-level effects. Column 3: A descriptive model for autism genetic association. Rows are marginal (row 1) and conditional (rows 2 and 3) association tests and statistics, and corresponding model representations. Variants are grouped based on gene-level effects [G] [silent, missense, LGD, and nonsilent (missense + LGD)] and pathway-level effects (P) (hitting Selected pathways or Others). The associations marked in column 3 should be taken as the theme for corresponding row. Notations  $\alpha(G; A)$  and  $\alpha(G; A | P)$  are read as marginal association between G and A, and their conditional association given P. The association can be measured by rate difference (over noise), rate ratio ( $\theta$ ), or log  $\theta$ . P values come from the rate difference tests. Error bars represent SEM. Note raw counts are summarized in table S1. For technical details, see the "Multilevel association analysis" section under Materials and Methods. NA, not applicable.

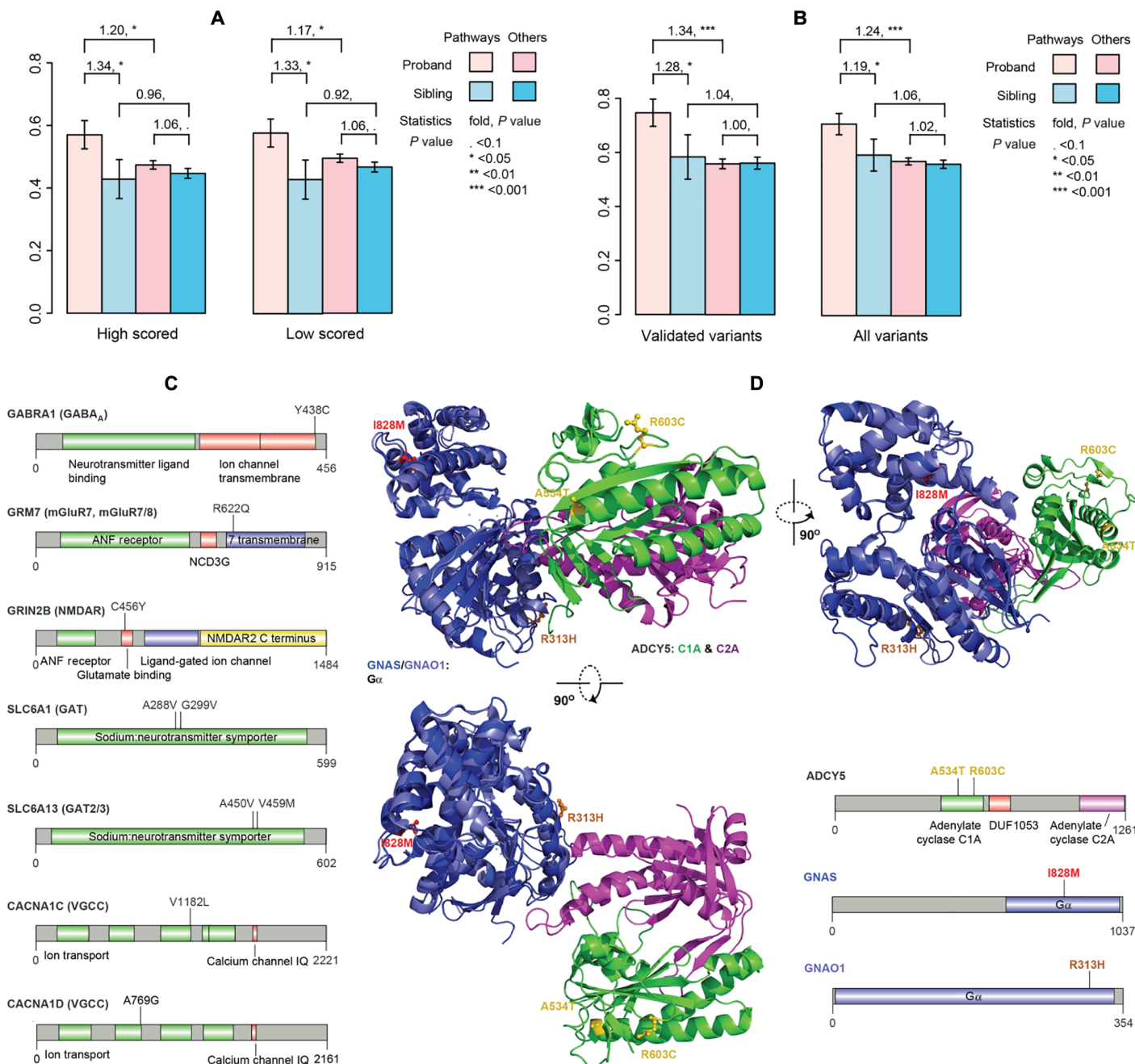


**Fig. 3. An integrated view of autism-associated DN variants or genes from multiple sources in selected KEGG pathways. (A)** hsa04310 Wnt signaling pathway, **(B)** hsa04727 GABAergic synapse, and **(C)** hsa04724 Glutamatergic synapse (next page). DN variant data come from SSC and ASC studies, and reported autism genes come from the SFARI Gene database. Gene-level scores (Materials and Methods) are marked by color. P values are from pathway analysis (Table 1). Data are integrated and visualized on KEGG pathway graphs using Pathview (23).

(Fig. 4A) (21). In addition, more missense variants in selected pathways hit a functional domain in probands versus siblings (0.747 versus 0.558,  $P < 0.05$ ) (Fig. 4B), especially in Wnt and synapse pathways (figs. S8 and S9; more details in section S4).

Proband variants more frequently hit the selected pathways. Probands have higher absolute event rates than siblings in selected pathways (0.11 versus 0.08,  $P = 4.2 \times 10^{-4}$ ), especially in LGD (0.021 versus 0.008,

$P = 2.7 \times 10^{-4}$ ) and missense (0.053 versus 0.036,  $P = 3.5 \times 10^{-3}$ ) categories, but not in the silent category (0.033 versus 0.032,  $P = 0.45$ ). After being adjusted for event numbers within each category or in total, probands still have consistently higher pathway event rates than siblings for both LGD (0.12 versus 0.07,  $P = 4.1 \times 10^{-2}$ ) and missense categories (0.08 versus 0.06,  $P = 2.2 \times 10^{-2}$ ), but not for the silent category (0.06 versus 0.07,  $P = 0.60$ ) (row 3 in Fig. 2).



**Fig. 4. Functional consequences of autism-associated missense mutations.** (A) Ratios of damaging events as predicted by SIFT. (B) Ratios of events hitting a function domain defined in Pfam. (C) 1D protein domain structure and missense variants of all neurotransmitter receptors, transporters, and ion channel genes in synapse pathways (the bold black box nodes in pathway graphs in figs. S6 and S7). (D) 1D and 3D protein structures and missense variants hitting the AC, that is, ADCY5, and interacting G proteins, GNAS (G $\alpha$ ) and GNAO1 (G $\beta\gamma$ ). The pathway context is shown in the green dashed box in fig. S7. AC controls the production of cAMP second-messenger in synapses (section S4). Error bars represent SEM.

We propose a gene + pathway dual-hit (or two-factor) model for ASD genetic association based on our results above (Fig. 2, column 3): the level of disrupting effect on target genes (G) and hitting the relevant pathways or not (P). These two factors have significant association with ASD both marginally and conditionally, as described above. In our model, variant load/burden per person (V) becomes less relevant and is marked as hidden. The extra variants in probands (versus in siblings) mostly fall into the gene-disrupting and pathway-hitting categories (Fig. 2, row 1, described above).

There is also a significant interaction between gene and pathway factors (Fig. 2, row 1). Probands and siblings have the biggest differences in variants that are both gene-disrupting and pathway-hitting. The differences diminish outside the pathways or disappear completely in the silent category. This interaction is significant, as indicated by significant overrepresentation in LGD hitting the pathways in probands (52 occurred versus 34.6 expected events,  $P = 0.001$ , table S1).

What we propose here is essentially a Noisy-AND model in that risk genetic variant tends to be both gene-disrupting AND pathway-hitting. The model is Noisy because our knowledge of gene-disrupting and pathway assignment is incomplete or because of the incomplete penetrance.

We also estimate the prevalence of DN events with different gene- and pathway-level effects (fig. S3). These statistics are similar to per-patient variant burden statistics (Fig. 2, row 1) and consistent with our two-factor genetic model for ASD (Fig. 2, row 1). With the DN variants alone, this model explains at least 5% (2.8% within and 2.2% outside selected pathways) of all ASD cases (fig. S3). Although consistent with the LoF mutation contribution in the ASC study (11), this is likely a substantial underestimation, because not all variants are called and not all genes in the relevant pathways are known. In addition, when other types of variants [copy number variants (CNVs), common variants, or transmitted/inherited variants] are considered, this model can be generic and more descriptive (section S2).

### Pathways of DN events, integrated molecular mechanism

Selected by our special pathway-level testing procedure, these pathways form a coherent yet nonredundant set of ASD disease mechanisms (Table 1). These pathways are novel in various aspects: (i) For two pathways (actin and T junction), this is the first time to report that they are involved in ASD; (ii) for other pathways (lysine, GABA, Wnt, MAPK, Circ, and Glut), there are some evidence on their potential roles in ASD in literature (as in Table 1), but this is the first report that is backed by statistical significance from whole-exome/whole-genome analysis; (iii) for all selected pathways, this is the first report that plots the pathway graphs to show the detailed molecular mechanisms of ASD (Fig. 3 and fig. S4).

These pathways present a catalog of genetic risk factors for ASD (table S5, Fig. 3, and fig. S4). At the variant level, 43 LGD and 75 missense mutations (118 in total) are mapped to the selected pathways (table S5). These mutations have not been identified previously as ASD events. At the gene level, 92 genes fall into these pathways (table S6). Among them, 20 have been reported previously as ASD genes based on the SFARI Gene database (22). The other 72 genes are novel candidates for ASD genes, among them 9 genes are replicated in both SSC and ASC cohorts. Being part of the selected pathways, these mutations and genes are likely pathogenic or disease-causing.

Note that we primarily focus on the selected pathways. A small portion of the events outside these pathways ( $P = \text{Others}$  in Fig. 2) can also be pathogenic because the pathway definitions are incomplete and some

relevant pathways may not be significant because of the limited sample size. Nonetheless, known ASD genes are greatly enriched in the selected pathways compared to all selected genes (20 of 92 versus 67 of 540,  $P = 0.004$ ). The enrichment of known ASD genes also serves as a systematic validation of our findings.

We generate pathway graphs integrating disease-associated variants and genes from multiple data sets: SSC (10), ASC (11), or SFARI Gene database (Fig. 3 and fig. S4) (22). These pathways are likely true positives and the primary molecular mechanisms for ASD because they are consistently selected in these independent analyses. These analyses agree in details, too: They frequently converge to the same genes, the same gene groups (nodes), or the same signaling branch in a pathway. They also complement each other. For instance, SSC and ASC data provide numerous ASD-associated genes besides those collected in SFARI Gene. Here, SFARI Gene is used mainly for a confirmatory analysis. This database collects ASD genes reported from the literature works, most of which are independent from the SSC and ASC studies.

The pathway list and data-integrated pathway graphs provide coherent and systematic insights into the ASD mechanism. We focus on three pathways for example. These pathways are not entirely new based on literature (more in Table 1) (13–15). However, in our analysis (Fig. 3 and fig. S4), they are visualized as novel molecular maps integrating numerous new and known disease genes or events.

1) Wnt signaling pathway: the canonical branch only (Fig. 3A). All DN events from SSC and ASC and all SFARI genes converge to the canonical Wnt pathway. However, just a few events/genes hit noncanonical Wnt pathways, and these genes are mostly shared by the canonical branch or other selected pathways. In addition, all aspects or steps of canonical Wnt signaling are involved in ASD (Fig. 3A). These include Wnt and co-receptor LRP5/6, messenger Dvl, key components of the destruction complex (APC, GSK3, and CK1 $\epsilon$ ), and other key players in  $\beta$ -catenin phosphorylation/ubiquitination/degradation [TBL1 in p53-induced Skp1/cullin/F-box protein (SCF)-like complex,  $\beta$ -TrCP in SCF E3 ubiquitin ligase complex, PS-1 (presenilin)], and repressors or activators (chromatin remodelers) in  $\beta$ -catenin-directed transcription [CHD8 (Duplin), RUVBL1 (Pontin52), and CREBBP (CBP)].

2) The whole GABAergic synapse pathway is involved in ASD, particularly the following parts (Fig. 3B): (i) GABA<sub>A</sub> receptors and their signal in the postsynaptic neurons, the negative feedback loops [G<sub>i/o</sub> and adenylate cyclase (AC)] in pre- and postsynaptic neurons, and the clearance channel through GABA transporters (GATs) on the presynaptic terminal or neighboring glial cells and (ii) glial cells besides the pre- and postsynaptic neurons.

3) The whole glutamatergic synapse pathway is involved in ASD, particularly the following parts (Fig. 3C): (i) the ionotropic glutamate receptor [iGluRs and NMDARs (*N*-methyl-D-aspartate receptors)] signal, the postsynaptic density scaffold proteins (SHANKs, etc.), the consequent synapse formation, and plasticity; (ii) metabotropic GluRs (mGluRs; mGluR1, 5, 7, and 8), the coupled G proteins (G<sub>s</sub>, G<sub>i</sub>, and G<sub>o</sub>), and the second-messenger systems downstream [Ca<sup>2+</sup>, cAMP, diacylglycerol (DAG), and inositol 1,4,5-trisphosphate (IP<sub>3</sub>)]; (iii) the inhibitory autoreceptor mechanism that suppresses excess glutamate release in presynaptic neurons [mGluR7, G<sub>i/o</sub>, GRK (G protein-coupled receptor kinase), and AC]; and (iv) glial cells besides the pre- and postsynaptic neurons, especially in the clearance and recycling of glutamate.

Other pathways and graphs are equally informative, and many of them are also supported by literature (Table 1). For details, see section S3 and fig. S4.

### Subpathway biology, coherent fine details

We analyze the functional consequences of DN variants in selected pathways. Here, we focus on missense, but not LGD, variants. Whereas the latter are highly destructive on overall protein structure and function no matter which part they hit (position-insensitive), the former are subtle and precisely tell what functions are perturbed in ASD (position-sensitive).

In probands, missense variants hit the relevant functions or domains in selected pathways. In Wnt signaling pathway, missense hit the histone acetylation domain KAT11 twice in the CREBBP (CBP) gene and TIP49 domain in RUVBL1, the scaffolding domain WD40 in TBL1XR1, and the CTNBN1-binding domain in TCF7L1 (fig. S8). In synapse pathways, the most essential players, including neurotransmitter receptors, transporters, and ion channels on cell membrane, are heavily targeted (Fig. 4, C and D, and figs. S6 and S7). Missense variants hit the neurotransmitter glutamate-binding domain in the GRIN2B (NMDAR) gene, the seven-transmembrane region of GRM7 (mGluR7), the ion channel domains in GABRA1 and CACNA1C, the sodium/neurotransmitter symporter domains in SLC6A1 (GAT) and SLC6A13 (GAT2/3).

Conversely, missense variants in siblings often hit the nonfunctional regions or the less relevant regions or genes (figs. S8 to S10). This proband-sibling difference is significant overall (Fig. 4B) and extremely so in the example pathways (fig. S8 and S9 and section S4).

Autistic missense events on the same genes tend to hit residues extremely close to each other and in the same domain. This occurs to all cases that we observed in Wnt and synapse pathways (Fig. 4, C and D, or figs. S8 and S9: ADCY5, CREBBP, SLC6A1, and SLC6A13). These data strongly suggest that missense events do not occur in random, but precisely and repeatedly target specific risky loci for ASD ( $P = 0.002$  to  $0.03$ , section S4).

We identified subpathway clusters of missense events in probands. Each event cluster hits multiple interacting genes along the pathway. They reveal recurrent molecular modules in ASD biology.

One cluster hit the cAMP second-messenger system (23) in the glutamatergic synapse pathway (Fig. 4D and fig. S7). Two types of G proteins bind and control adenylate cyclase (AC):  $G_s$  activates it, whereas  $G_{i/o}$  ( $G_i/G_o$ ) inhibits it (green dashed box in fig. S7). As shown in Fig. 4D, the  $G\alpha$  domains of GNAS ( $G_s$ ) and GNAO1 ( $G_{i/o}$ ) are similar and align seamlessly in 3D (24). They compete to bind to the AC C2A domain in a similar way. Missense variants in these two genes both hit the  $G\alpha$  domain, which affects their binding to AC and hence affects AC's catalytic activity on cAMP production and downstream signal. In parallel, the two missense events on AC (ADCY5) hit its C1A domain, which perturb AC's catalytic function, too (Fig. 4D). In the direct upstream (fig. S7), GRM5 (mGluR5) was hit by a destructive in-frame deletion (K679) (table S5). GRM7 (mGluR7) was hit by a missense at the seven-transmembrane region (Fig. 4C), which produces an unbounded cytosol form of the protein and likely a strong antagonist of the original transmembrane signal.

In another cluster, GRKs inhibit mGluR signaling by sequestering heterotrimeric G proteins. See section S4 and fig. S11 for details.

All these subpathway-level biological stories we present above reveal coherent fine details on ASD mechanism. These details are consistent with, but complement to, the integrated pathway graphs (Fig. 3 and figs. S6 and S7).

### Superpathway biology, emergent big picture

The selected pathways are distinct yet highly interconnected. For example, MAPKs feed into canonical Wnt pathway and inhibit T cell factor

(TCF)/lymphoid enhancer factor (LEF)-dependent transcription (Fig. 3A and fig. S4C). In addition, they also share numerous other connections. For example, Wnt and MAPK are both involved in adherens junctions and focal adhesion (fig. S4, G and H). These commonly connected pathways are also perturbed in ASD except that they are only marginally significant ( $P = 0.01$  to  $0.10$ , table S2).

Two distinct biological themes or modules emerge from the network of selected and connected pathways (Fig. 5). Module I includes Wnt signaling, cell adhesion, junction, and cytoskeleton-related pathways. They are involved in synapse morphology, that is, synapse assembly and stability. Module II includes glutamatergic synapse, GABAergic synapse, and related processes. They are involved in synapse functions, that is, chemical and electrical signal transmissions, regulations, and patterns. Module I concerns neuronal wiring or the "hardware," whereas module II concerns synaptic transmission or the "software." These modules are distinct in network topology, too. Connections are dense within each module but none between them. The MAPK pathway is the only bridge node and is highly connected in both modules. There is one less prominent theme: transcription (not shown in Fig. 5). Both Wnt and MAPK pathways end at target gene transcription, which involves chromatin modification, especially the histone lysine methylation branch of lysine degradation (Table 1). We also did a parallel GO term analysis, which converges to the same set of biological themes (section S5 and fig. S12).

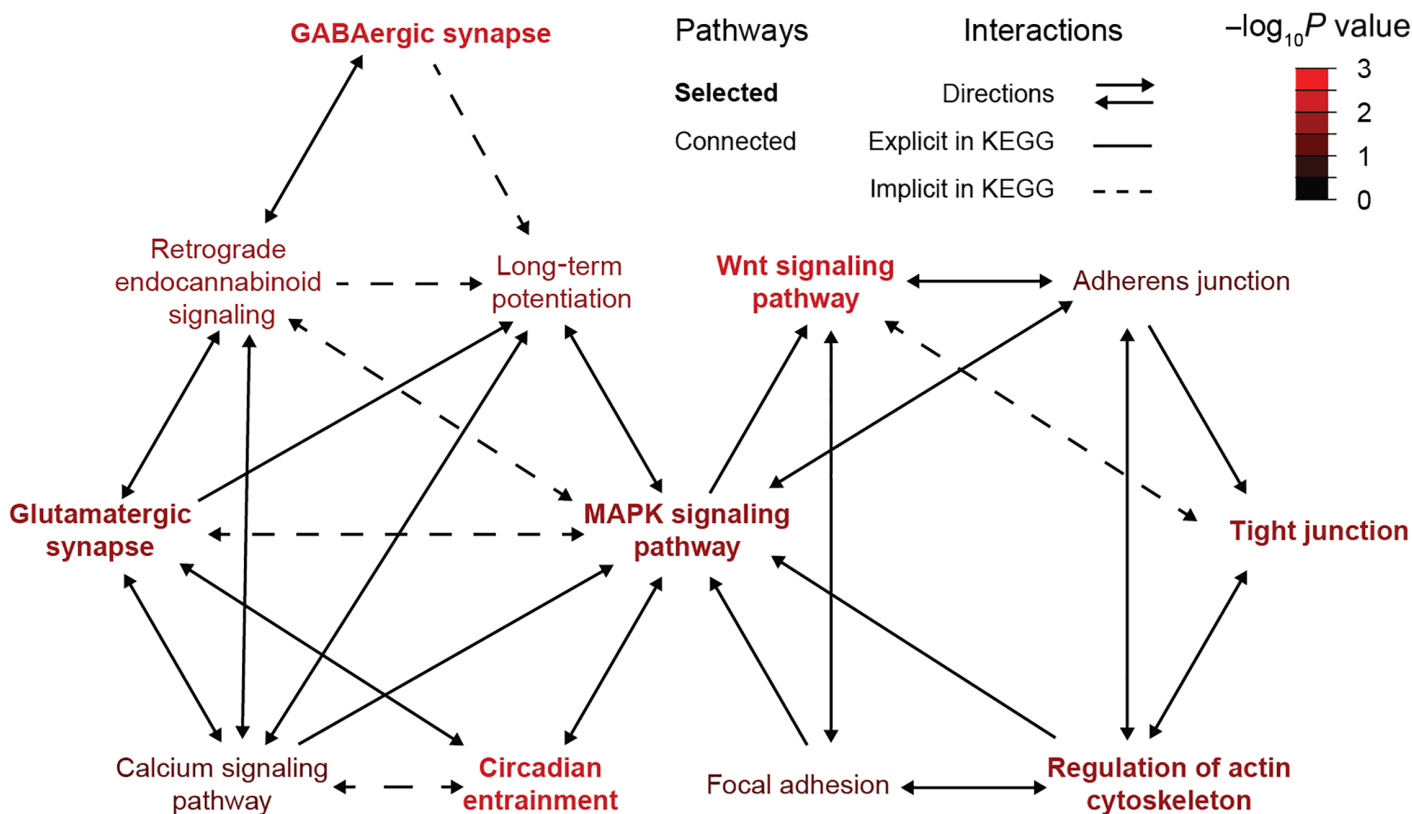
### DISCUSSION

We analyzed thousands of ASD exome mutations across multiple levels using a novel sequential prioritization procedure. Our results suggest that isolated, rare mutation events are actually connected and recurrent at higher (gene and pathway) levels (Fig. 1). In addition, the otherwise random and divergent results from independent data sets become reproducible. This cross-validation not only confirms the prioritization results but also justifies the multilevel prioritization procedure. As discussed in the literature, valid method and reproducible results are critical for genome-scale studies (25) and autism genetics (26). This same multiple-level analysis procedure is applicable to other complex diseases or problems as well.

We also did a multilevel association analysis and proposed a gene + pathway dual-hit model for ASD risk (Fig. 2). The disease variants need to both (i) disrupt the target genes and (ii) hit the relevant pathways, show reduced or no correlation with ASD risk. Variants missing either factor, including silent variants in the selected pathways or variants outside the pathways, show reduced or no correlation with ASD risk. Previous studies have established the disease association of DN mutation burden with autism (4, 10, 27). In this new model, the contribution of variant load/burden can be explained away by the gene + pathway factors and hence becomes less relevant. This model likely applies to other types of genetic variants including CNV and single-nucleotide polymorphisms (SNP).

We reconstructed a set of coherent molecular mechanisms for ASD (Figs. 3 to 5 and Table 1). We identified several canonical pathways that may cause the disease, as supported by multiple independent data sets. These disease pathways present a catalog of ASD genetic associations (table S5) and also connect hundreds of interacting genes and variants into a whole multiscale system (Figs. 3 to 5). All mutated pathways or functions converge to synapse biology. Synaptic function, morphology, or plasticity (as indicated by transcription) (28, 29) is disrupted in these cases. These results are consistent with previous research on synaptic morphology (30), function, and transcription (11, 13–15). Therefore, ASD is





## Modules II: Synaptic transmission, function

## I: Neuronal wiring, morphology

**Fig. 5. The superpathway clusters emerged from the pathway-level analysis of SSC exome variant data.** Seven selected pathways are highly interconnected and frequently connected to five additional pathways. All these pathways form a superpathway-level network. Two clusters or modules emerged with distinct topology and function.

not only a multigenic but also a multipathway disease, and it is ultimately a synapse disease.

Our results are consistent with previous studies (1–12). For example, synapse function and chromatin remodeling–related genes have been repeatedly reported as possible factors in ASD (10, 11, 31), and cell projection/morphogenesis and cytoskeleton/actin-related genes have also been suggested (30–32). In addition to these results, our analyses also present some novel and coherent insights into ASD biology (Figs. 3 to 5), which are more interpretable and informative than the gene networks or Gene Ontology (GO) groups described in the literature. Although most identified pathways are potentially causal, they may also explain associated symptoms of ASD, including intellectual disability (glutamate and GABA synapses) (33), sleeping (circadian entrainment) (34), and digestive (protein digestion and absorption) problems (35).

These results were systematically reproduced or validated in independent cohorts/analyses (Figs. 1 and 3, figs. S1 and S4, and Table 1). With the specific contexts (Fig. 3 and figs. S4 to S7) and multilevel details (Fig. 4 and figs. S8 to S11), these findings are experimentally testable if needed. Although our analysis has taken a systematic approach, our results are far from complete. For example, temporal and spatial information could eventually be incorporated into the gene- and pathway-level analyses to further refine the disease mechanism (9, 36). New insights can be expected with bigger sample size, more data, and better methodology coming in the near future.

## MATERIALS AND METHODS

### Data collection and integration

The Exome-Seq DN variants from the SSC cohort (10) and ASC cohort (11) were used for this study. Please see the original publications for details of the experimental design, quality control, and raw data processing. The final SSC data include 2517 families, with 2508 affected children, 1911 unaffected siblings, and the parents of each family. The ASC data we used consist of two cohorts: one includes 1445 trios and the other includes 1601 cases and 5397 ancestry-matched controls. The ASC paper originally included 825 trios from the SSC cohort. This overlap was intentionally excluded to create two completely independent data sets for downstream analysis and comparison.

### Multilevel genetic prioritization

To search for the genetic causes of disease (ASD), we devised a novel sequential prioritization procedure. The whole procedure is similar to a multistage refinement process. We started with a few thousand rare variants, of which only a small portion are disease-causing mutations (drivers) and most are irrelevant (passengers). Driver events were then selected against passengers through a sequential set of criteria (filters), and in the end, a minor fraction (a few selected pathways) remains greatly enriched in drivers. The procedure thus has three consecutive steps, that is, variant-, gene-, and pathway-level selection; each

incorporates experimental data, metadata, and annotation data at corresponding levels. Details of individual selection steps are as follows.

### Variant-level selection

Variants were divided into three major categories based on their effects on the target genes. The silent group includes all synonymous variants and variants which fall in the 3' untranslated region (3'UTR), 5'UTR, intergenic, intron, and noncoding regions. The missense group includes missense variants; the LGD or LoF group includes exon indels (both frameshift and no frameshift), nonsense, and splice-site variants. Variants are selected for gene- and pathway-level analyses based on a few criteria: (i) LGD (or LoF) and missense only, because silent variants are usually not damaging and have little disease association as a group (Fig. 2). (ii) For the SSC study (10), we only considered validated variants, which included those experimentally verified or cross-validated or called in at least two of the three laboratories (Cold Spring Harbor Laboratory, Yale, or University of Washington). (iii) For ASC study (11), we only considered DN variants in the trio families or case-only variants from the case-control cohorts.

### Gene-level selection

All selected variants were mapped to genes based on the chromosome coordinates. We selected genes using the following scoring function, which essentially sums up the weighted evidence for each gene.

$$s_i = \sum_j I_{ij} \cdot w_j$$

$$G_i : s_i \geq s_0$$

where  $i$  is the gene index,  $j$  is the patient index,  $I$  is the indicator on whether a selected variant occurs to the gene-patient pair,  $w$  is the weight, and  $s$  is the score

Corresponding to their different study designs and data quality, we used different parameters for the two cohorts. In SSC, we take  $w_j = 1/n_j$  (number of selected variants occurred to patient  $j$ ) and  $s_0 = 0.5$ , whereas in ASC,  $w_j = 1$  and  $s_0 = 2$ . Both criteria control the amount of evidence from mapped variants. The former is more inclusive and the latter is more restrictive because SSC has simpler data structure and better data quality compared to ASC.

Note that for rare DN mutations, we did not consider linkage disequilibrium. Unlike SNPs, which are transmitted in blocks, rare DN mutations are random and independent events. Therefore, linkage disequilibrium does not apply to these events. Gene length itself is an important factor contributing to ASD risk (37, 38); hence, we decided not to explicitly control it. On the other hand, most genes were called significant in gene length-based test because these DN events are rare. All possible genomic features are shared by probands and siblings; hence, the best control to use is the sibling group as we have done (Fig. 1 and fig. S1).

### Pathway-level selection

We identified pathways enriched in the selected genes. Different from the regular gene set analysis, we test for both marginal and conditional overrepresentation given the previously selected pathways. This procedure ensures that pathways selected are drivers instead of passengers, which share genes with the former.

The analysis is an application of the set theory. Below is a list of basic and relevant notations following the standard set theory:

{: set definition or description

$A \cap B$ : intersection between set A and B

$A \cup B$ : union between set A and B

$A \setminus B$ : relative complement, subset that belongs to A but not to B

$|A|$ : cardinality, or the number of elements of set A

The list of sets and variables defined for our analyses:

$G = \{\text{selected genes above}\}$

$P_i = \{\text{pathway or gene set under testing}\}$

$P_s = \{\text{selected pathways or gene sets}\}$

$P_0 = \{\text{all pathways or gene sets}\}$

$U = |G \cap P_0|$

$V = |G \cap P_0 \setminus P_s|$

$X = |G \cap P_i|$

$Y = |G \cap P_i \setminus P_s|$

For marginal significance test:

$X = j \sim \text{hyperG}(j; |P_i|, |P_0 \setminus P_i|, U)$

$P(X \geq j) = \sum_l \text{PhyperG}(j; |P_i|, |P_0 \setminus P_i|, U)$ , where  $l = \{j, j+1, \dots, \min(|P_i|, U)\}$

For conditional significance test:

$X = k | P_s \sim Y = k \sim \text{hyperG}(k; |P_i \setminus P_s|, |P_0 \setminus P_i \setminus P_s|, V)$

$P(X \geq k | P_s) = P(Y \geq k) = \sum_l \text{PhyperG}(k; |P_i \setminus P_s|, |P_0 \setminus P_i \setminus P_s|, V)$ , where  $l = \{k, k+1, \dots, \min(|P_i \setminus P_s|, V)\}$

Here, hyperG is the hypergeometric distribution and PhyperG is the standard probability mass function of the hypergeometric distribution.

The same analysis procedure was applied to KEGG pathways and GO terms. The metabolic and signaling pathways from KEGG were tested and analyzed together, and the three branches of GO, that is, biological process, cellular component, molecular function, were analyzed separately. We did multiple-testing correction on  $P$  values using false discovery rate (FDR; or  $q$  value).

### Multilevel recurrence analysis

As shown in Fig. 1, in each level (step) of the sequential prioritization procedure above, we have an input list (considered) and an output list (selected) of entries (variants, genes, or pathways). We test whether the output list (compared to the input list) was enriched in recurrent entries. Recurrent entries were defined relative to the reference list. To analyze the recurrence in SSC proband variants (either input or output), the reference list can be variants from ASC probands, SSC siblings, or SSC probands (themselves) corresponding to recurrent entries in an independent cohort, in the control group of the same cohort, or in the same proband group, respectively. Entries recurring between independent cohorts measure reproducibility of studies. Entries recurring in normal condition were likely irrelevant to the disease, and those recurring in the same disease group were likely relevant to the disease. These different types of recurrence can be used for validation of prioritization procedures. Enrichment of recurrent changes between different cohorts or within the same disease group (likely disease-causing or true positives) suggests sensitivity, whereas depletion of recurrent changes from the control groups (likely disease-irrelevant or false positives) suggests selectivity.

As in Fig. 1 (A and B) and fig. S1 (A and B), enrichment of recurrent entries was tested using hypergeometric test. Input lists were marked as considered (black circle), output lists were marked as selected (red circle), and selected lists from the contrast condition were used as a reference list (blue circle) for recurrence (overlaps in gray or purple). The recurrence rate difference between probands and siblings was analyzed using two-sample  $t$  tests (fig. S1C).

### Multilevel association analysis

To analyze the ASD association of DN mutations, we take their gene- and pathway-level effects into account. DN variants can be divided into groups based on these effects. At the gene level, they were assigned to

silent, LGD, missense, or nonsilent (LGD + missense) groups, as described above. At the pathway level, they belong to either the Selected pathways (Table 1) or Others. That is, DN variants can be divided into 1D or 2D groups based on two variables: gene-level effect (G) and pathway assignment (P), as shown in table S1 (raw counts) and Fig. 2 and fig. S2 (event rates).

The ASD association was then tested for each DN variant group (Fig. 2 and fig. S2). The ASD association can be measured by rate difference (over noise), rate ratio ( $\theta$ ), or  $\log \theta$  between probands and siblings. The data suggested a strong interaction between the two factors G and P (Fig. 2, row 1, and table S1). Therefore, we tested the marginal ASD associations (Fig. 2, row 1) and the conditional ASD associations by fixing factor P or G (Fig. 2, rows 2 and 3). Marginal association test was conducted on nominal event rate [raw count per cell per number of individuals ( $n$ ) in table S1] for each variant group, whereas conditional test was conducted on normalized event rate, that is, the nominal rate of individual variant group normalized by the total event rate of each P or G level (raw count per cell/row or column sum in table S1). To test the rate difference between probands and siblings, we conducted two-proportion  $z$  test for conditional rates and two-sample  $t$  test for marginal rates. Odd ratio tests gave similar results as in our conditional rate tests but were not suitable for marginal tests on absolute variant rates.

### Multilevel function analysis

We analyzed the perturbed biological processes and functions in ASD within the hierarchical context of variant, gene, and pathways. In particular, we focused on the few selected pathways (Table 1) and the mapped variants and genes. Details of the function analyses are covered by levels below.

#### Pathway level—Data integration and visualization

We visualized all selected pathways with ASD variants/genes mapped and integrated (Fig. 3 and figs. S4 to S7). These graphs reveal integrated molecular mechanisms at the pathway level. Pathview package (39) was used for pathway-based data integration and visualization. Variants were first mapped to the target genes, which were then mapped and visualized onto the selected KEGG pathway graphs. In the disease gene view (Fig. 3 and fig. S4), variant-targeted genes from SSC and ASC and SFARI genes are collected, integrated, and shown in the relevant pathways. Different data sources were marked by colors, gene-level scores were marked by brightness, and corresponding pathway analysis  $P$  values are also shown. In variant type views (figs. S5 to S7), DN variants from SSC were projected and visualized on the target pathways. Variant types or effects (LGD, missense, or silent) were marked by different colors, and their corresponding event counts were also shown.

#### Subpathway level—Protein structure and function analysis

We analyzed the functional consequences of DN variants in selected pathways. Here, we focused on missense variants, which are subtle and precisely tell what functions are perturbed in ASD (position-sensitive). The overall functional impact of these variants was assessed using SIFT tool (similar results with PolyPhen and not shown) or based on Pfam protein domains (as domain hitting or not). Variants from probands or siblings were divided into groups by pathway assignment (Selected pathways or Others). The damaging or domain hitting ratios of each variant group were then compared using two-sample  $t$  tests assuming binomial variances (Fig. 4, A and B).

Exome variants were mapped to amino acid changes in the target protein using Bioconductor VariantAnnotation package (40). 1D linear protein domain structures were visualized using cBioPortal MutationMapper (41). Protein domain data were retrieved from the

Pfam database (42) and updated the protein domain locations. 3D protein structure data were retrieved from the Protein Data Bank (PDB) (43). The mapped exome variants coded into amino acid changes were then visualized with the 3D protein structure using PyMOL ([www.pymol.org](http://www.pymol.org)).

#### Superpathway level—Networks of KEGG pathways or GO terms

KEGG pathway and GO definition and gene annotation data were downloaded from the corresponding database as of March 2015. We analyzed the functional interactions between pathways. First, we built a network of pathways (Fig. 5), where nodes are pathways and edges are pathway interactions explicitly documented in the KEGG LinkDB database or implicitly suggested by KEGG pathway diagrams or definitions. We focused on a subnetwork where nodes are either selected pathways or other pathways but directly connected by at least three selected pathways. Network on selected GO terms was done similarly (fig. S12). However, we used a bipartite graph. This graph includes GO terms as one node type and biology themes as the other. Edges were directed from GO terms to biology themes only, as suggested by GO term definition or at least one literature work in table S3.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/4/e1701799/DC1>

- section S1. Sequential convergence from variant to gene to pathway level
- section S2. Autism genetic association dissected across multiple levels
- section S3. Pathways of DN events, integrated molecular mechanism
- section S4. Subpathway biology, coherent fine details
- section S5. Superpathway biology, emergent big picture
- fig. S1. Multilevel recurrence of de novo mutations in SSC study.
- fig. S2. Autism genetic association analysis across variant, gene, and pathway levels with the SSC exome DN mutation data.
- fig. S3. Prevalence of DN variant types by gene- and pathway-level effects in the SSC exome study.
- fig. S4. An integrated view of autism associated DN variants or genes from multiple sources.
- fig. S5. An integrated view of DN variants by gene level effects in the selected KEGG pathway, hsa04310 Wnt signaling pathway.
- fig. S6. An integrated view of DN variants by gene level effects in the selected KEGG pathways, hsa04727 GABAergic synapse.
- fig. S7. An integrated view of DN variants by gene level effects in the selected KEGG pathways, hsa04724 Glutamatergic synapse.
- fig. S8. 1D protein domain structure and missense variants of genes in Wnt signaling pathway for both probands and siblings.
- fig. S9. 1D protein domain structure and missense variants of genes in synapse pathways for both probands and siblings.
- fig. S10. 1D protein domain structure and DN mutations of gene pairs in Wnt signaling and synapse pathways.
- fig. S11. 1D and 3D protein structures and missense variants hitting mGluR inhibitor GRK (ADRBK2) and interacting G proteins.
- fig. S12. Selected GO terms and emerging biological themes from the SSC exome variant data.
- table S1. The actual (and expected) counts of DN events by gene-level (columns) and pathway-level (rows) effects.
- table S2. Other pathways connected by three or more selected pathways in Table 1.
- table S3. Significant GO groups selected from SSC, their test statistics, and references.
- table S4. (selectVariant.xlsx file). Validated and selected variants in the multilevel integrated analyses of SSC and ASC data.
- table S5. (variantAnnot.xlsx file). Validated variants and functional annotations in the selected pathways of SSC data.
- table S6. (geneAnnot.xlsx file). Genes hit by LGD or missense events in probands in selected pathways.
- References (58–89)

### REFERENCES AND NOTES

1. B. J. O'Roak, P. Deriziotis, C. Lee, L. Vives, J. J. Schwartz, S. Girirajan, E. Karakoc, A. P. MacKenzie, S. B. Ng, C. Baker, M. J. Rieder, D. A. Nickerson, R. Bernier, S. E. Fisher,

- J. Shendure, E. E. Eichler, Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* **43**, 585–589 (2011).
2. B. M. Neale, Y. Kou, L. Liu, A. Ma'ayan, K. E. Samocha, A. Sabo, C.-F. Lin, C. Stevens, L.-S. Wang, V. Makarov, P. Polak, S. Yoon, J. Maguire, E. L. Crawford, N. G. Campbell, E. T. Geller, O. Valladares, C. Schafer, H. Liu, T. Zhao, G. Cai, J. Lihm, R. Dannenfeller, O. Jabado, Z. Peralta, U. Nagaswamy, D. Muzny, J. G. Reid, I. Newsham, Y. Wu, L. Lewis, Y. Han, B. F. Voight, E. Lim, E. Rossin, A. Kirby, J. Flannick, M. Fromer, K. Shakir, T. Fennell, K. Garimella, E. Banks, R. Poplin, S. Gabriel, M. DePristo, J. R. Wimbish, B. E. Boone, S. E. Levy, C. Betancur, S. Sunyaev, E. Boerwinkle, J. D. Buxbaum, E. H. Cook Jr., B. Devlin, R. A. Gibbs, K. Roeder, G. D. Schellenberg, J. S. Sutcliffe, M. J. Daly, Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
  3. B. J. O'Roak, L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D. Smith, E. H. Turner, I. B. Stanaway, B. Vernot, M. Malig, C. Baker, B. Reilly, J. M. Akey, E. Borenstein, M. J. Rieder, R. A. Nickerson, R. Bernier, J. Shendure, E. E. Eichler, Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250 (2012).
  4. S. J. Sanders, M. T. Murtha, A. R. Gupta, J. D. Murdoch, M. J. Raubeson, A. J. Willsey, A. G. Ercan-Sencicek, N. M. DiLullo, N. N. Parikshak, J. L. Stein, M. F. Walker, G. T. Ober, N. A. Teran, Y. Song, P. El-Fishawy, R. C. Murtha, M. Choi, J. D. Overton, R. D. Bjornson, N. J. Carriero, K. A. Meyer, K. Bilguvar, S. M. Mane, N. Šestan, R. P. Lifton, M. Günel, K. Roeder, D. H. Geschwind, B. Devlin, M. W. State, De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
  5. I. Iossifov, M. Ronemus, D. Levy, Z. Wang, I. Hakker, J. Rosenbaum, B. Yamrom, Y.-h. Lee, G. Narzisi, A. Leotta, J. Kendall, E. Grabowska, B. Ma, S. Marks, L. Rodgers, A. Stepansky, J. Troge, P. Andrews, M. Bekris, K. Pradhan, E. Ghiban, M. Kramer, J. Parla, R. Demeter, L. L. Fulton, R. S. Fulton, V. J. Magrini, K. Ye, J. C. Darnell, R. B. Darnell, E. R. Mardis, R. K. Wilson, M. C. Schatz, W. R. McCombie, M. Wigler, De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
  6. J. J. Michaelson, Y. Shi, M. Gujral, H. Zheng, D. Malhotra, X. Jin, M. Jian, G. Liu, D. Greer, A. Bhandari, W. Wu, R. Corominas, Á. Peoples, A. Koren, A. Gore, S. Kang, G. N. Lin, J. Estabillo, T. Gadowski, B. Singh, K. Zhang, N. Akshoomoff, C. Corsello, S. McCarroll, L. M. Iakoucheva, Y. Li, J. Wang, J. Sebat, Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* **151**, 1431–1442 (2012).
  7. T. W. Yu, M. H. Chahrouh, M. E. Coulter, S. Jiralerspong, K. Okamura-Ikeda, B. Ataman, K. Schmitz-Abe, D. A. Harmin, M. Adli, A. M. Malik, A. M. D'Gama, E. T. Lim, S. J. Sanders, G. H. Mochida, J. N. Partlow, C. M. Sunu, J. M. Felie, J. Rodriguez, R. H. Nasir, J. Ware, R. M. Joseph, R. S. Hill, B. Y. Kwan, M. Al-Saffar, N. M. Mukaddes, A. Hashmi, S. Balkhy, G. G. Gascon, F. M. Hisama, E. LeClair, A. Poduri, O. Oner, S. Al-Saad, S. A. Al-Awadi, L. Bastaki, T. Ben-Omran, A. S. Teebi, L. Al-Gazali, V. Eapen, C. R. Stevens, L. Rappaport, S. B. Gabriel, K. Markianos, M. W. State, M. E. Greenberg, H. Taniguchi, N. E. Braverman, E. M. Morrow, C. A. Walsh, Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77**, 259–273 (2013).
  8. Y.-h. Jiang, R. K. C. Yuen, X. Jin, M. Wang, N. Chen, X. Wu, J. Ju, J. Mei, Y. Shi, M. He, G. Wang, J. Liang, Z. Wang, D. Cao, M. T. Carter, C. Chrysler, I. E. Drmic, J. L. Howe, L. Lau, C. R. Marshall, D. Merico, T. Nalpathamkalam, B. Thiruvahindrapuram, A. Thompson, M. Uddin, S. Walker, J. Luo, E. Anagnostou, L. Zwaigenbaum, R. H. Ring, J. Wang, C. Lajonchere, J. Wang, A. Shih, P. Szatmari, H. Yang, G. Dawson, Y. Li, S. W. Scherer, Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* **93**, 249–263 (2013).
  9. N. N. Parikshak, R. Luo, A. Zhang, H. Won, J. K. Lowe, V. Chandran, S. Horvath, D. H. Geschwind, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
  10. I. Iossifov, B. J. O'Roak, S. J. Sanders, M. Ronemus, N. Krumm, D. Levy, H. A. Stessman, K. T. Witherspoon, L. Vives, K. E. Patterson, J. D. Smith, B. Paepier, D. A. Nickerson, J. Dea, S. Dong, L. E. Gonzalez, J. D. Mandell, S. M. Mane, M. T. Murtha, C. A. Sullivan, M. F. Walker, Z. Waqar, L. Wei, A. J. Willsey, B. Yamrom, Y.-h. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M. C. Schatz, K. Ye, W. R. McCombie, J. Shendure, E. E. Eichler, M. W. State, M. Wigler, The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
  11. S. De Rubeis, X. He, A. P. Goldberg, C. S. Poultney, K. Samocha, A. E. Cicek, Y. Kou, L. Liu, M. Fromer, S. Walker, T. Singh, L. Klei, J. Kosmicki, F. Shih-Chen, B. Alekic, M. Biscaldi, P. F. Bolton, J. M. Brownfeld, J. Cai, N. G. Campbell, A. Carracedo, M. H. Chahrouh, A. G. Chiochetti, H. Coon, E. L. Crawford, S. R. Curran, G. Dawson, E. Duketis, B. A. Fernandez, L. Gallagher, E. Geller, S. J. Guter, R. S. Hill, J. Ionita-Laza, P. Jimenez Gonzalez, H. Kilpinen, S. M. Klauk, A. Kolevzon, I. Lee, I. Lei, J. Lei, T. Lehtimäki, C. F. Lin, A. Ma'ayan, C. R. Marshall, A. L. McInnes, B. Neale, M. J. Owen, N. Ozaki, M. Parellada, J. R. Parr, S. Purcell, K. Puura, D. Rajagopalan, K. Rehnström, A. Reichenberg, A. Sabo, M. Sachse, S. J. Sanders, C. Schafer, M. Schulte-Rüther, D. Skuse, C. Stevens, P. Szatmari, K. Tammimies, O. Valladares, A. Voran, W. Li-San, L. A. Weiss, A. J. Willsey, T. W. Yu, R. K. Yuen; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium, E. H. Cook, C. M. Freitag, M. Gill, C. M. Hultman, T. Lehner, A. Palotie, G. D. Schellenberg, P. Sklar, M. W. State, J. S. Sutcliffe, C. A. Walsh, S. W. Scherer, M. E. Zwick, J. C. Baret, D. J. Cutler, K. Roeder, B. Devlin, M. J. Daly, J. D. Buxbaum, Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
  12. R. K. C. Yuen, B. Thiruvahindrapuram, D. Merico, S. Walker, K. Tammimies, N. Hoang, C. Chrysler, T. Nalpathamkalam, G. Pellicchia, Y. Liu, M. J. Gazzellone, L. D'Abate, E. Deneault, J. L. Howe, R. S. C. Liu, A. Thompson, M. Zarrei, M. Uddin, C. R. Marshall, R. H. Ring, L. Zwaigenbaum, P. N. Ray, R. Weksberg, M. T. Carter, B. A. Fernandez, W. Roberts, P. Szatmari, S. W. Scherer, Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.* **21**, 185–191 (2015).
  13. N. Krumm, B. J. O'Roak, J. Shendure, E. E. Eichler, A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci.* **37**, 95–105 (2014).
  14. A. J. Willsey, M. W. State, Autism spectrum disorders: From genes to neurobiology. *Curr. Opin. Neurobiol.* **30**, 92–99 (2015).
  15. L. de la Torre-Ubieta, H. Won, J. L. Stein, D. H. Geschwind, Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016).
  16. G. D. Fischbach, C. Lord, The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).
  17. J. D. Buxbaum, M. J. Daly, B. Devlin, T. Lehner, K. Roeder, M. W. State; Autism Sequencing Consortium, The Autism Sequencing Consortium: Large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron* **76**, 1052–1056 (2012).
  18. M. Ronemus, I. Iossifov, D. Levy, M. Wigler, The role of de novo mutations in the genetics of autism spectrum disorders. *Nat. Rev. Genet.* **15**, 133–141 (2014).
  19. D. H. Geschwind, M. W. State, Gene hunting in autism spectrum disorder: On the path to precision medicine. *Lancet Neurol.* **14**, 1109–1120 (2015).
  20. B. J. O'Roak, L. Vives, W. Fu, J. D. Egerton, I. B. Stanaway, I. G. Phelps, G. Carvill, A. Kumar, C. Lee, K. Ankenman, J. Munson, J. B. Hiatt, E. H. Turner, R. Levy, D. R. O'Day, N. Krumm, B. P. Coe, B. K. Martin, E. Borenstein, D. A. Nickerson, H. C. Mefford, D. Doherty, J. M. Akey, R. Bernier, E. E. Eichler, J. Shendure, Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
  21. P. C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
  22. B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, I. Menashe, T. Wadkins, S. Banerjee-Basu, A. Packer, SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).
  23. P. A. Salin, R. K. Malenka, R. A. Nicoll, Cyclic AMP mediates a presynaptic form of LTP at cerebellar parallel fiber synapses. *Neuron* **16**, 797–803 (1996).
  24. J. J. G. Tesmer, R. K. Sunahara, A. G. Gilman, S. R. Sprang, Crystal structure of the catalytic domains of adenylyl cyclase in a complex with  $G_{\text{ss}}\text{-GTP}\gamma\text{S}$ . *Science* **278**, 1907–1916 (1997).
  25. D. Venet, J. E. Dumont, V. Detours, Most random gene expression signatures are significantly associated with breast cancer outcome. *PLOS Comput. Biol.* **7**, e1002240 (2011).
  26. J. C. Barrett, J. Buxbaum, D. Cutler, M. Daly, B. Devlin, J. Gratten, M. E. Hurler, J. A. Kosmicki, E. S. Lander, D. G. MacArthur, B. M. Neale, K. Roeder, P. M. Visscher, N. R. Wray, New mutations, old statistical challenges. *bioRxiv* **2017**, 115964 (2017).
  27. J. Sebat, B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, B. Yamrom, A. S. Yoon, A. Krasnitz, J. Kendall, A. Leotta, D. Pai, R. Zhang, Y.-H. Lee, J. Hicks, S. J. Spence, A. T. Lee, K. Puura, T. Lehtimäki, D. Ledbetter, P. K. Gregersen, J. Bregman, J. S. Sutcliffe, V. Jobanputra, W. Chung, D. Warburton, M.-C. King, D. Skuse, D. H. Geschwind, T. C. Gilliam, K. Ye, M. Wigler, Strong association of de novo copy number mutations with autism. *Science* **316**, 445–449 (2007).
  28. C. M. Alberini, Transcription factors in long-term memory and synaptic plasticity. *Physiol. Rev.* **89**, 121–145 (2009).
  29. J. Cortés-Mendoza, S. Díaz de León-Guerrero, G. Pedraza-Alva, L. Pérez-Martínez, Shaping synaptic plasticity: The role of activity-mediated epigenetic regulation on gene transcription. *Int. J. Dev. Neurosci.* **31**, 359–369 (2013).
  30. C. Correia, G. Oliveira, A. M. Vicente, Protein interaction networks reveal novel autism risk genes within GWAS statistical noise. *PLOS ONE* **9**, e112399 (2014).
  31. S. J. Sanders, X. He, A. J. Willsey, A. G. Ercan-Sencicek, K. E. Samocha, A. E. Cicek, M. T. Murtha, V. H. Bal, S. L. Bishop, S. Dong, A. P. Goldberg, C. Jinlu, J. F. Keaney III, L. Klei, J. D. Mandell, D. Moreno-De-Luca, C. S. Poultney, E. B. Robinson, L. Smith, T. Solli-Nowlan, M. Y. Su, N. A. Teran, M. F. Walker, D. M. Werling, A. L. Beaudet, R. M. Cantor, E. Fombonne, D. H. Geschwind, D. E. Grice, C. Lord, J. K. Lowe, S. M. Mane, D. M. Martin, E. M. Morrow, M. E. Talkowski, J. S. Sutcliffe, C. A. Walsh, T. W. Yu; Autism Sequencing Consortium, D. H. Ledbetter, C. L. Martin, E. H. Cook, J. D. Buxbaum, M. J. Daly, B. Devlin, K. Roeder, M. W. State, Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
  32. D. Pinto, E. Delaby, D. Merico, M. Barbosa, A. Merikangas, L. Klei, B. Thiruvahindrapuram, X. Xu, R. Ziman, Z. Wang, J. A. Vorstman, A. Thompson, R. Regan, M. Pilorge, G. Pellicchia, A. T. Pagnamenta, B. Oliveira, C. R. Marshall, T. R. Magalhaes, J. K. Lowe, J. L. Howe,

- A. J. Griswold, J. Gilbert, E. Duketis, B. A. Dombroski, M. V. De Jonge, M. Cuccaro, E. L. Crawford, C. T. Correia, J. Conroy, I. C. Conceição, A. G. Chiochetti, J. P. Casey, G. Cai, C. Cabrol, N. Bolshakova, E. Bacchelli, R. Anney, S. Gallinger, M. Cotterchio, G. Casey, L. Zwaigenbaum, K. Wittemeyer, K. Wing, S. Wallace, H. van Engeland, A. Tryfon, S. Thomson, L. Soorya, B. Rogé, W. Roberts, F. Poustka, S. Mougá, N. Minshew, L. A. McInnes, S. G. McGrew, C. Lord, M. Leboyer, A. S. Le Couteur, A. Kolevzon, P. Jiménez González, S. Jacob, R. Holt, S. Guter, J. Green, A. Green, C. Gillberg, B. A. Fernandez, F. Duque, R. Delorme, G. Dawson, P. Chaste, C. Café, S. Brennan, T. Bourgeron, P. F. Bolton, S. Bölte, R. Bernier, G. Baird, A. J. Bailey, E. Anagnostou, J. Almeida, E. M. Wijsman, V. J. Vieland, A. M. Vicente, G. D. Schellenberg, M. Pericak-Vance, A. D. Paterson, J. R. Parr, G. Oliveira, J. I. Nurnberger, A. P. Monaco, E. Maestrini, S. M. Klauck, H. Hakonarson, J. L. Haines, D. H. Geschwind, C. M. Freitag, S. E. Folstein, S. Ennis, H. Coon, A. Battaglia, P. Szatmari, J. S. Sutcliffe, J. Hallmayer, M. Gill, E. H. Cook, J. D. Buxbaum, B. Devlin, L. Gallagher, C. Betancur, S. W. Scherer, Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* **94**, 677–694 (2014).
33. H. Y. Zoghbi, M. F. Bear, Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. *Cold Spring Harb. Perspect. Biol.* **4**, a009886 (2012).
34. G. Glickman, Circadian rhythms and sleep in children with autism. *Neurosci. Biobehav. Rev.* **34**, 755–768 (2010).
35. K. Horvath, J. C. Papadimitriou, A. Rabsztyń, C. Drachenberg, J. T. Tildon, Gastrointestinal abnormalities in children with autistic disorder. *J. Pediatr.* **135**, 559–563 (1999).
36. A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan, M. W. State, Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
37. I. F. King, C. N. Yandava, A. M. Mabb, J. S. Hsiao, H.-S. Huang, B. L. Pearson, J. M. Calabrese, J. Starmer, J. S. Parker, T. Magnuson, S. J. Chamberlain, B. D. Philpot, M. J. Zylka, Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58–62 (2013).
38. K. Jabbari, P. Nürnberg, A genomic view on epilepsy and autism candidate genes. *Genomics* **108**, 31–36 (2016).
39. W. Luo, C. Brouwer, Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
40. V. Obenchain, M. Lawrence, V. Carey, S. Gogarten, P. Shannon, M. Morgan, VariantAnnotation: A Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* **30**, 2076–2078 (2014).
41. J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
42. R. D. Finn, A. Bateman, J. Clements, P. Coghill, R. Y. Eberhardt, S. R. Eddy, A. Heeger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, M. Punta, Pfam: The protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
43. P. W. Rose, C. Bi, W. F. Bluhm, C. H. Christie, D. Dimitropoulos, S. Dutta, R. K. Green, D. S. Goodsell, A. Pilić, M. Quesada, G. B. Quinn, A. G. Ramos, J. D. Westbrook, J. Young, C. Zardecki, H. M. Berman, P. E. Bourne, The RCSB Protein Data Bank: New resources for research and education. *Nucleic Acids Res.* **41**, D475–D482 (2013).
44. S. Akbarian, H.-S. Huang, Epigenetic regulation in human brain—Focus on histone lysine methylation. *Biol. Psychiatry* **65**, 198–203 (2009).
45. S. Coghlan, J. Horder, B. Inkster, M. A. Mendez, D. G. Murphy, D. J. Nutt, GABA system dysfunction in autism and related disorders: From synapse to symptoms. *Neurosci. Biobehav. Rev.* **36**, 2044–2055 (2012).
46. H.-T. Chao, H. Chen, R. C. Samaco, M. Xue, M. Chahrouh, J. Yoo, J. L. Neul, S. Gong, H.-C. Lu, N. Heintz, M. Ekker, J. L. R. Rubenstein, J. L. Noebels, C. Rosenmund, H. Y. Zoghbi, Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468**, 263–269 (2010).
47. H. O. Kalkman, A review of the evidence for the canonical Wnt pathway in autism spectrum disorders. *Mol. Autism* **3**, 10 (2012).
48. N. D. Okerlund, B. N. R. Cheyette, Synaptic Wnt signaling—A contributor to major psychiatric disorders? *J. Neurodev. Disord.* **3**, 162–174 (2011).
49. D. Pinto, A. T. Pagnamenta, L. Klei, R. Anney, D. Merico, R. Regan, J. Conroy, T. R. Magalhaes, C. Correia, B. S. Abrahams, J. Almeida, E. Bacchelli, G. D. Bader, A. J. Bailey, G. Baird, A. Battaglia, T. Berney, N. Bolshakova, S. Bölte, P. F. Bolton, T. Bourgeron, S. Brennan, J. Brian, S. E. Bryson, A. R. Carson, G. Casallo, J. Casey, B. H. Chung, L. Cochrane, C. Corsello, E. L. Crawford, A. Crosslet, C. Cytrynbaum, G. Dawson, M. de Jonge, R. Delorme, I. Drmic, E. Duketis, F. Duque, A. Estes, P. Farrar, B. A. Fernandez, S. E. Folstein, E. Fombonne, C. M. Freitag, J. Gilbert, C. Gillberg, J. T. Glessner, J. Goldberg, A. Green, J. Green, S. J. Guter, H. Hakonarson, E. A. Heron, M. Hill, R. Holt, J. L. Howe, G. Hughes, V. Hus, R. Iglizzo, C. Kim, S. M. Klauck, A. Kolevzon, O. Korvatska, V. Kustanovich, C. M. Lajonchere, J. A. Lamb, M. Laskawiec, M. Leboyer, A. Le Couteur, B. L. Leventhal, A. C. Lionel, X. Q. Liu, C. Lord, L. Lotspeich, S. C. Lund, E. Maestrini, W. Mahoney, C. Mantoulan, C. R. Marshall, H. McConachie, C. J. McDougale, J. McGrath, W. M. McMahon, A. Merikangas, O. Migita, N. J. Minshew, G. K. Mirza, J. Munson, S. F. Nelson, C. Noakes, A. Noor, G. Nygren, G. Oliveira, K. Papanikolaou, J. R. Parr, B. Parrini, T. Paton, A. Pickles, M. Pilorge, J. Piven, C. P. Ponting, D. J. Posey, A. Poustka, F. Poustka, A. Prasad, J. Ragoussis, K. Renshaw, J. Rickaby, W. Roberts, K. Roeder, B. Rogé, M. L. Rutter, L. J. Bierut, J. P. Rice, J. Salt, K. Sansom, D. Sato, R. Segurado, A. F. Sequeira, L. Senman, N. Shah, V. C. Sheffield, L. Soorya, I. Sousa, O. Stein, N. Sykes, V. Stoppioni, C. Strawbridge, R. Tancredi, K. Tansey, B. Thiruvahindrapuram, A. P. Thompson, S. Thomson, A. Tryfon, J. Tsiantis, H. Van Engeland, J. B. Vincent, F. Volkmar, S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Wittemeyer, S. Wood, J. Wu, B. L. Yaspan, D. Zurawiecki, L. Zwaigenbaum, J. D. Buxbaum, R. M. Cantor, E. H. Cook, H. Coon, M. L. Cuccaro, B. Devlin, S. Ennis, L. Gallagher, D. H. Geschwind, M. Gill, J. L. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger Jr., A. D. Paterson, M. A. Pericak-Vance, G. D. Schellenberg, P. Szatmari, A. M. Vicente, V. J. Vieland, E. M. Wijsman, S. W. Scherer, J. S. Sutcliffe, C. Betancur, Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
50. L. Luo, Actin cytoskeleton regulation in neuronal morphogenesis and structural plasticity. *Annu. Rev. Cell Dev. Biol.* **18**, 601–635 (2002).
51. C. N. G. Giachello, F. Fiumara, C. Giacomini, A. Corradi, C. Milanese, M. Ghirardi, F. Benfenati, P. G. Montarolo, MAPK/Erk-dependent phosphorylation of synapsin mediates formation of functional synapses and short-term homosynaptic plasticity. *J. Cell Sci.* **123**, 881–893 (2010).
52. G. M. Thomas, R. L. Huganir, MAPK cascade signalling and synaptic plasticity. *Nat. Rev. Neurosci.* **5**, 173–183 (2004).
53. Z. Liu, N. Li, J. Neu, Tight junctions, leaky intestines, and pediatric diseases. *Acta Paediatr.* **94**, 386–393 (2005).
54. V. W. Tang, Proteomic and bioinformatic analysis of epithelial tight junction reveals an unexpected cluster of synaptic molecules. *Biol. Direct* **1**, 37 (2006).
55. T. Bourgeron, The possible interplay of synaptic and clock genes in autism spectrum disorders. *Cold Spring Harb. Symp. Quant. Biol.* **72**, 645–654 (2007).
56. R. Delorme, E. Ey, R. Toro, M. Leboyer, C. Gillberg, T. Bourgeron, Progress toward treatments for synaptic defects in autism. *Nat. Med.* **19**, 685–694 (2013).
57. I. Cusco, A. Medrano, B. Gener, M. Vilardell, F. Gallastegui, O. Villa, E. González, B. Rodríguez-Santiago, E. Vilella, M. Del Campo, L. A. Pérez-Jurado, Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Hum. Mol. Genet.* **18**, 1795–1804 (2009).
58. J. T. Glessner, K. Wang, G. Cai, O. Korvatska, C. E. Kim, S. Wood, H. Zhang, A. Estes, D. W. Brune, J. P. Bradfield, M. Imielinski, E. C. Frackelton, J. Reichert, E. L. Crawford, J. Munson, P. M. Sleiman, R. Chiavacci, K. Annaiah, K. Thomas, C. Hou, W. Glaberson, J. Flory, F. Otieno, M. Garris, L. Soorya, L. Klei, J. Piven, K. J. Meyer, E. Anagnostou, T. Sakurai, R. M. Game, D. S. Rudd, D. Zurawiecki, C. J. McDougale, L. K. Davis, J. Miller, D. J. Posey, S. Michaels, A. Kolevzon, J. M. Silverman, R. Bernier, S. E. Levy, R. T. Schultz, G. Dawson, T. O'Leary, W. M. McMahon, T. H. Wassink, J. A. Sweeney, J. I. Nurnberger, H. Coon, J. S. Sutcliffe, N. J. Minshew, S. F. Grant, M. Bucan, E. H. Cook, J. D. Buxbaum, B. Devlin, G. D. Schellenberg, H. Hakonarson, Autism genome-wide copy number variation reveals ubiquitous and neuronal genes. *Nature* **459**, 569–573 (2009).
59. L. A. Weiss, D. E. Arking; Gene Discovery Project of Johns Hopkins & the Autism Consortium, M. J. Daly, A. Chakravarti A genome-wide linkage and association scan reveals novel loci for autism. *Nature* **461**, 802–808 (2009).
60. K. Wang, H. Zhang, D. Ma, M. Bucan, J. T. Glessner, B. S. Abrahams, D. Salyakina, M. Imielinski, J. P. Bradfield, P. M. Sleiman, C. E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Takahashi, T. Sakurai, E. Rappaport, C. M. Lajonchere, J. Munson, A. Estes, O. Korvatska, J. Piven, L. I. Sonnenblick, A. I. Alvarez Retuerto, E. I. Herman, H. Dong, T. Hutman, M. Sigman, S. Ozonoff, A. Klin, T. O'Leary, J. A. Sweeney, W. Brune, R. M. Cantor, R. Bernier, J. R. Gilbert, M. L. Cuccaro, W. M. McMahon, J. Miller, M. W. State, T. H. Wassink, H. Coon, S. E. Levy, R. T. Schultz, J. I. Nurnberger, J. L. Haines, J. S. Sutcliffe, E. H. Cook, N. J. Minshew, J. D. Buxbaum, G. Dawson, S. F. Grant, D. H. Geschwind, M. A. Pericak-Vance, G. D. Schellenberg, H. Hakonarson, Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528–533 (2009).
61. C. Nava, F. Lamari, D. Héron, C. Mignot, A. Rastetter, B. Keren, D. Cohen, A. Faudet, D. Bouteiller, M. Gilleron, A. Jacqueline, S. Whalen, A. Afenjar, D. Périse, C. Laurent, C. Dupuits, C. Gautier, M. Gérard, G. Huguet, S. Caillet, B. Leheup, M. Leboyer, C. Gillberg, R. Delorme, T. Bourgeron, A. Brice, C. Depienne, Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including *TMLHE*. *Transl. Psychiatry* **2**, e179 (2012).
62. P. B. C. Celestino-Soper, C. A. Shaw, S. J. Sanders, J. Li, M. T. Murtha, A. G. Ercan-Sencicek, L. Davis, S. Thomson, T. Gambin, A. C. Chinault, Z. Ou, J. R. German, A. Milosavljevic, J. S. Sutcliffe, E. H. Cook Jr., P. Stankiewicz, M. W. State, A. L. Beaudet, Use of array CGH to

- detect exonic copy number variants throughout the genome in autism families detects a novel deletion in *TMLHE*. *Hum. Mol. Genet.* **20**, 4360–4370 (2011).
63. M. N. Ziats, M. S. Comeaux, Y. Yang, F. Scaglia, S. H. Elsea, Q. Sun, A. L. Beaudet, C. P. Schaaf, Improvement of regressive autism symptoms in a child with *TMLHE* deficiency following carnitine supplementation. *Am. J. Med. Genet. A* **167**, 2162–2167 (2015).
  64. D. Grafodatskaya, B. Chung, P. Szatmari, R. Weksberg, Autism spectrum disorders and epigenetics. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 794–809 (2010).
  65. M. K. Belmonte, G. Allen, A. Beckel-Mitchener, L. M. Boulanger, R. A. Carper, S. J. Webb, Autism and abnormal development of brain connectivity. *J. Neurosci.* **24**, 9228–9231 (2004).
  66. L. A. Cingolani, Y. Goda, Actin in action: The interplay between the actin cytoskeleton and synaptic efficacy. *Nat. Rev. Neurosci.* **9**, 344–356 (2008).
  67. T. Bourgeron A synaptic trek to autism. *Curr. Opin. Neurobiol.* **19**, 231–234 (2009).
  68. M. Missler, T. C. Südhof, T. Biederer, Synaptic cell adhesion. *Cold Spring Harb. Perspect. Biol.* **4**, a005694 (2012).
  69. P. D'Eufemia, M. Celli, R. Finocchiaro, L. Pacifico, L. Viozzi, M. Zaccagnini, E. Cardi, O. Giardini, Abnormal intestinal permeability in children with autism. *Acta Paediatr.* **85**, 1076–1079 (1996).
  70. E. R. Kandel, R. D. Hawkins, The biological basis of learning and individuality. *Sci. Am.* **267**, 78–86 (1992).
  71. C. H. Kim, J. Lee, J.-Y. Lee, K. W. Roche, Metabotropic glutamate receptors: Phosphorylation and receptor signaling. *J. Neurosci. Res.* **86**, 1–10 (2008).
  72. V. M. Tesmer, T. Kawano, A. Shankaranarayanan, T. Kozasa, J. J. Tesmer, Snapshot of activated G proteins at the membrane: The  $G_{\alpha_q}$ -GRK2-G $\beta\gamma$  complex. *Science* **310**, 1686–1690 (2005).
  73. J. J. Zhu, Y. Qin, M. Zhao, L. Van Aelst, R. Malinow, Ras and Rap control AMPA receptor trafficking during synaptic plasticity. *Cell* **110**, 443–455 (2002).
  74. H.-J. Chen, M. Rojas-Soto, A. Oguni, M. B. Kennedy, A synaptic Ras-GTPase activating protein (p135 SynGAP) inhibited by CaM kinase II. *Neuron* **20**, 895–904 (1998).
  75. V. Rangaraju, N. Calloway, T. A. Ryan, Activity-driven local ATP synthesis is required for synaptic function. *Cell* **156**, 825–835 (2014).
  76. J. J. Harris, R. Jolivet, D. Attwell, Synaptic energy use and supply. *Neuron* **75**, 762–777 (2012).
  77. J. Rohrbough, K. Broadie, Lipid regulation of the synaptic vesicle cycle. *Nat. Rev. Neurosci.* **6**, 139–150 (2005).
  78. N. H. Jung, W. G. Janzarik, I. Delvendahl, A. Münchau, M. Biscaldi, F. Mainberger, T. Bäumer, R. Rauh, V. Mall, Impaired induction of long-term potentiation-like plasticity in patients with high-functioning autism and Asperger syndrome. *Dev. Med. Child Neurol.* **55**, 83–89 (2013).
  79. L. Zhang, B. E. Alger, Enhanced endocannabinoid signaling elevates neuronal excitability in fragile X syndrome. *J. Neurosci.* **30**, 5724–5729 (2010).
  80. D. M. Kerr, L. Downey, M. Conboy, D. P. Finn, M. Roche, Alterations in the endocannabinoid system in the rat valproic acid model of autism. *Behav. Brain Res.* **249**, 124–132 (2013).
  81. C. Betancur, T. Sakurai, J. D. Buxbaum, The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci.* **32**, 402–412 (2009).
  82. H. Wei, M. Malik, A. M. Sheikh, G. Merz, W. Ted Brown, X. Li, Abnormal cell properties and down-regulated FAK-Src complex signaling in B lymphoblasts of autistic subjects. *Am. J. Pathol.* **179**, 66–74 (2011).
  83. E. L. Williams, M. F. Casanova Above genetics: Lessons from cerebral development in autism. *Transl. Neurosci.* **2**, 106–120 (2011).
  84. J. J. Gargus Genetic calcium signaling abnormalities in the central nervous system: Seizures, migraine, and autism. *Ann. N. Y. Acad. Sci.* **1151**, 133–156 (2009).
  85. A. T.-H. Lu, X. Dai, J. A. Martinez-Agosto, R. M. Cantor, Support for calcium channel gene defects in autism spectrum disorders. *Mol. Autism* **3**, 18 (2012).
  86. J. M. Lasalle, Autism genes keep turning up chromatin. *OA Autism* **1**, 14 (2013).
  87. J. J. Hutsler, H. Zhang, Increased dendritic spine densities on cortical projection neurons in autism spectrum disorders. *Brain Res.* **1309**, 83–94 (2010).
  88. A. El-Ansary, L. Al-Ayadhi, Lipid mediators in plasma of autism spectrum disorders. *Lipids Health Dis.* **11**, 160 (2012).
  89. H. Ye, J. Liu, J. Y. Wu, Cell adhesion molecules and their involvement in autism spectrum disorder. *Neurosignals* **18**, 62–71 (2010).

#### Acknowledgments

**Funding:** This work was supported by the NSF (ABI-1565030 to W.L.). C.Z. was supported by grants from Simons Foundation (307711) and NIH (R01GM124486 and R01NS89676). Y.-h.J. was supported by NIH grants MH098114, MH104316, and HD087795.

**Author contributions:** W.L. designed the study, developed the methods, collected the data, and performed the study. W.L., C.Z., Y.-h.J., and C.R.B. wrote the paper, and all authors have read and approved the final version. W.L. claims the responsibility for all figures in the main text and the Supplementary Materials. **Competing interests:** The authors declare that they have no conflict of interest. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 29 May 2017

Accepted 22 February 2018

Published 11 April 2018

10.1126/sciadv.1701799

**Citation:** Luo, C. Zhang, Y.-h. Jiang, C. R. Brouwer, Systematic reconstruction of autism biology from massive genetic mutation profiles. *Sci. Adv.* **4**, e1701799 (2018).

## Systematic reconstruction of autism biology from massive genetic mutation profiles

Weijun Luo, Chaolin Zhang, Yong-hui Jiang and Cory R. Brouwer

*Sci Adv* 4 (4), e1701799.

DOI: 10.1126/sciadv.1701799

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/4/4/e1701799>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2018/04/09/4.4.e1701799.DC1>

### REFERENCES

This article cites 88 articles, 10 of which you can access for free  
<http://advances.sciencemag.org/content/4/4/e1701799#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Advances* is a registered trademark of AAAS.