Supplementary Materials for

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

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

Published 11 April 2018, *Sci. Adv.* 4, e1701799 (2018)

DOI: 10.1126/sciadv.1701799

The PDF file includes:

- 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.
• Legends for tables S4 to S6
• References (58–89)

Other Supplementary Material for this manuscript includes the following:
(available at advances.sciencemag.org/cgi/content/full/4/4/e1701799/DC1)

• table S4 (Microsoft Excel format). Validated and selected variants in the multilevel integrated analyses of SSC and ASC data.
• table S5 (Microsoft Excel format). Validated variants and functional annotations in the selected pathways of SSC data.
• table S6 (Microsoft Excel format). Genes hit by LGD or missense events in probands in selected pathways.
section S1. Sequential convergence from variant to gene to pathway level

Compare to the overlap between SSC and ASC cohorts at variant level, the overlap between probands and siblings in SSC cohort is actually higher (Fig. 1b). They are expected to have more similar genomic background, because they came from the same family in the same cohort.

section S2. Autism genetic association dissected across multiple levels

Previous studies report higher variant or CNV burden in autism (4, 10, 27, 49). In our analysis, these extra variants most comes from the gene disrupting and pathway hitting categories (Fig. 2 row 1, section 2 in main text). In other words, our two-factor model largely explains away the association of extra variant burden with autism, i.e. it is not a direct causal factor of autism. In addition, to affect autism disease status, such extra variants need to occur to the effective variant categories (gene disrupting and pathway hitting), but not other categories (like the silent variants within selected pathways and most variants outside) (Fig. 2 row 1).

As described in Fig. 2 and main text, what we proposed is essentially a Noisy-AND model for autism genetic risk factors, i.e. risk genetic variant tend to be both gene disrupting AND pathway hitting. The model is Noisy because our knowledge on gene disrupting and pathway assignment is incomplete. For instance, there are still some level of association between autism and gene disrupting variants outside selected pathways (Fig. 2 row 1, section 2 in main text), presumably because some relevant pathways remain unknown hence unselected and more genes can be assigned to the selected pathways than currently known. Similarly, error/noise may arise from the inaccurate assignment of variants to silent or gene disrupting categories or their incomplete penetrance.

This model is built using DN rare mutations (local, SNV or indel). However, we expect other types of data, i.e. CNV(27, 49, 58) or common variants (59, 60), would converge to the same two-factor model and the same set of molecular mechanisms (pathways or genes). Although they are different types of variants, they should follow the same set of rules to be associated with autism: they should disrupt the target genes, either the structure/function or the expression, and at the same time they should hit the same set of pathways as these local variants. We would like to explore multilevel integrated analysis on CNV or GWAS data in our future work.
section S3. Pathways of DN events, integrated molecular mechanism

Besides the three pathways described in Section 3 of main text, other pathways and graphs are equally informative (Table 1 and fig. S4). All pathway identified are potentially causal, except Protein digestion and absorption, which is likely just association (35). They even explain associated symptoms of ASD, including intellectual disability (33) (Glut, GABA), sleeping (34) (Circ) and digestive (35) (digestion) problems. Here are the mechanisms revealed by these pathways with respect to ASD.

**hsa00310 Lysine degradation (fig. S4a)**

All data converge to the protein lysine methylation branch, these genes are histone-lysine N-methyltransferase for histone modification and transcription regulation. In addition, the hydroxylation may also be important for autism. The other two genes, TMLHE (61–63) and PLOD3, likely affect histone methylation by change histone markers through hydroxylation at up or down stream steps. In Gene Ontology, most of these genes also fall to GO:0018024 histone-lysine N-methyltransferase activity. It is known that histone modifications and other epigenetic changes play a key role in autism etiology (11, 64).

**hsa04810 Regulation of actin cytoskeleton (fig. S4b)**

Both the inputs (chemotaxis and immediate responses/effectors) and outputs (Actin polymerization, stabilization, focal adhesion and adhesion junction) of this pathway are relevant.

The actin cytoskeleton plays a major role in neuronal morphogenesis and structural plasticity (50). Example processes include axon initiation, growth, guidance and branching, in morphogenesis of dendrites and dendritic spines, in synapse formation and stability, and in axon and dendrite retraction (50). Indeed, autism involves abnormal neuronal development and connectivity (65). In addition, actin remodeling in neurons are the key to synapse formation and function (66), which is a primary factor in autism pathology (67). Interestingly, an independent protein-protein interaction (PPI) analyses of AGP and AGRE GWAS datasets also implicated cytoskeleton/actin and axon guidance in autism (30).

**hsa04010 MAPK signaling pathway (fig. S4c)**

Both the classical MAPK and Jnk-p38 pathways are relevant. Similar to the actin cytoskeleton pathway, MAPK signaling also plays an important role in synaptic formation, function and plasticity (51, 52). Both ERK and p38 MAPKs have defined roles (52). Consistently, our results showed that both perturbations in both classical and JNK/p38 MAPK pathways lead to autism. In known autism genes in SFARI gene
database (labeled in blue in fig. S4c), there are plenty evidences on the classical MAPK pathway, but not on JNK/p38 pathway. In other words, the latter or the most of its hit genes are novel autism associations.

**hsa04530 Tight junction (fig. S4d)**

Neuron communication depends on synapses. Synapses are essentially a type of cell junction. Cell junctions and cell adhesion molecules not only connect pre- and postsynaptic compartments, but also mediate trans-synaptic recognition and signaling processes that are essential for the establishment, specification, and plasticity of synapses (68). It is known that brain function depends on cell adhesion molecules (68). Tight junction may be part of synaptic structure and function (54).

**hsa04713 Circadian entrainment (fig. S4e)**

Circadian entrainment is also a synapse related process. Like other synapse pathways, most part of circadian entrainment is relevant to autism. Our results showed that circadian entrainment pathway is hit by autism genes. In other words, autism shares causal genes or molecular pathways with circadian disorders. In addition, autism or synaptic genes interact with clock genes. In other words, these two conditions may affect each other directly (55). Indeed, Circadian disorder is a primary comorbidity of autism. Circadian disorder and sleep problem are prevalent (up to 83%) in autism patients (34).

**hsa04974 Protein digestion and absorption (fig. S4f)**

Recent research shows that more than 50% of children with autism have GI symptoms, food allergies, and maldigestion or malabsorption issues (35). In addition, altered intestinal permeability was found in 43% of autistic patients, but not found in any of the controls (69). Intestinal permeability, commonly called "leaky gut", may allow undigested food and other toxins to enter the blood stream, which could cause health problems.

**section S4. Subpathway biology, coherent fine details**

In general, missense variants hitting selected pathways in probands are more relevant biologically than in siblings or other missense (Fig. 3). The probands-sibling difference in missense variants is well exhibited by the example pathways too. In probands (vs in sibling), 6 out of 8 or 75% (vs 1 out of 4 or 25%, p= 4.8×10^{-2}) missense events in Wnt signaling, and 15 out of 15 or 100% (vs 1 out of 3 or 33%, p= 4.0×10^{-4}) missense events in the synapses pathways hit the relevant function domains. Our data consistently shows,
for probands vs siblings, there are not only more DN events in selected pathways, but these events are also more disruptive (LGD and missense vs silent, Fig. 2 and fig. S10, figs. S5-7), and more relevant in function (Fig. 4a-b and figs. S8-9).

Autistic missense events on the same genes tend to hit residues extremely close and in the same domain. This occurs to all cases we observed in Wnt and synapse pathways (Fig. 5, figs. S8-9), including ADCY5: A534T and R603C on AC domain, SLC6A1 or SLC6A13: A288V and G299V or A450V and V459M on Sodium:neurotransmitter symporter domain, CREBBP: Y1539C and T1569A on KAT11 domain. Assuming missense events occur randomly, the chance to observe such closely occurred events in the same domain are very small ($p=0.01, 0.002, 0.03$ and $0.02$ respectively). These data strongly suggest that these missense events do not occur in random, but precisely and consistently targeting specific risky loci for autism.

Missense event cluster on cAMP second-messenger system (biology)
cAMP second-messenger system is key component of the synapse function, neurotransmission, learning and memory (cognitive) processes (23, 70). In post synapses, glutamate receptors (mGluRs) activate the coupled G proteins upon glutamate binding. G proteins then binds and control Adenylate cyclases, which produce cAMP, a primary second messenger for downstream signaling.

Another missense event cluster in the glutamatergic synapse pathway
GRK (G protein-coupled receptor kinases) can inhibit mGluR or GPCR signaling by phosphorylating activated the receptors and by sequestering heterotrimeric G proteins (71) (figs. S7 and 11). Figure S11 shows the phosphorylation independent mechanism, including sequestration of G-alpha (Gi/o here) subunits with its RGS homology (RH) domain and G-beta/gamma with its pleckstrin homology (PH) domain [protein structure (72)]. Indeed, missense variant hit GRK (ADRBK2) in its PH domain likely perturb its interaction with G-beta and G-gamma. The missense variant in GNAO1 hit the G-alpha domain close to the RGS binding site. Notice the same GNAO1 hit the G-alpha domain also interacts with AC (ADCY5) in the cAMP system (fig. S7 and Fig. 4d).

section S5. Superpathway biology, emergent big picture
The selected pathways are distinct yet highly interconnected. For example, MAPK feed into canonical Wnt pathway and inhibit TCF/LEF dependent transcription (Fig. 3a and fig. S4c). MAPK (PKC/PKA-Erk signaling), Glutamatergic synapse and Circadian entrainment are tightly intertwined by definition (Fig. 3c, fig. S4c and 4e). In addition, they also share numerous other connections. For example, Wnt and MAPK are both involved in adherens junctions and focal adhesion (fig. S4g-4h). Calcium signal is part of
MAPK pathway and Glutamatergic synapse (Fig. 3c and fig. S4c). These commonly connected pathways are also perturbed in ASD yet marginally significant (p.val=0.01-0.10, table S2). They may hit the significance level with a larger sample size.

Independent GO term analysis converges to the same set of biological themes. We analyzed all 3 branches of GO tree, but only describe Biological Process (BP) and Molecular Function (MF) in details here. Cellular Component (CC) results are less informative and specific for functional interpretation. Nonetheless, CC are still consistent and biologically relevant (table S3).

Three major biological themes emerged from the selected BP and MF terms (fig. S12). These are synapse function, morphology and transcription, highly consistent with KEGG pathway analysis (Fig. 5).

Each theme is supported by multiple GO terms. Each GO term associates with at least 1 theme, some with multiple themes. Two small GTPase related GO terms contribute to all 3 themes. Ras GTPase family regulate actin cytoskeleton hence synapse morphology, and gene transcription through MAPK (fig. S4b, Regulation of actin cytoskeleton). They also control synapse transmission directly (73, 74). Another GO term associates with to all 3 themes, i.e. Adenyl ribonucleotide binding (mainly ATP binding 79/81). Indeed, neurotransmission, synapse assembly (actin cytoskeleton) and activity dependent transcription are all energy intensive processes (75, 76). ATP binding and energy supply is critical for all of them.

The remaining associations between GO terms and biological themes are 1-to-1 and straightforward. The Transcription theme is supported by chromatin modification, histone-lysine N-methyltransferase activity, RNA polymerase II core binding. The Morphology theme is supported by neuron projection morphogenesis, cell adhesion molecule binding. The Synapse function theme is supported by calcium ion transmembrane transporter activity and intracellular lipid transport. The latter is involved in synaptic vesicle and protein delivery (77).

In addition to agreement on big themes, the selected GO terms are complimentary yet consistent with selected KEGG pathways. GO:0018024 histone-lysine N-methyltransferase activity is essentially the same gene set as hsa00310 Lysine degradation, both are subsets of chromatin modification. GO:0050839 cell adhesion molecule binding overlap with hsa04520 Adherens junction, and GO:0048812 neuron projection morphogenesis with hsa04530 Tight junction.
**fig. S1. Multilevel recurrence of de novo mutations in SSC study.** Venn diagrams and test statistics on recurrence at different levels for (a) probands and (b) siblings. (c) recurrence rates in selected vs other genes or pathways for probands and siblings. Term “considered” or “selected” refers to items before or after selection process at each level (Methods). See table S4 for full lists of selected variants and genes used in the analysis. Error bars represent standard error of the mean (SEM).
**fig. S2. Autism genetic association analysis across variant, gene, and pathway levels with the SSC exome DN mutation data.** Column 1: DN event rates and association test by gene level effect only without considering pathway level effect, i.e. this is the marginal sum of column 1 and 2 from Fig. 2; column 2: A descriptive model for autism genetic association. Rows are marginal (Row 1) and conditional (Row 2-3) association tests and statistics, and corresponding model representations. Error bars represent standard error of the mean (SEM). This figure should be interpreted together with Fig. 2, table S1 and *Multilevel association analysis* subsection in Methods.
**Fig. S3.** Prevalence of DN variant types by gene- and pathway-level effects in the SSC exome study. This figure is comparable to row 1 and column 1-2 of Fig. 2, and legend is the same as there too. Error bars represent standard error of the mean (SEM).
fig. S4. An integrated view of autism associated DN variants or genes from multiple sources. (a-f) selected KEGG pathways besides those shown in Fig. 4, and (g-h) commonly connected pathways. DN variants data come from SSC and ASC studies, and reported autism genes from SFARI Gene Database. P-values are from pathway analysis (Table 1). Data are integrated and visualized on KEGG pathway graphs using Pathview (23).
fig. S5. An integrated view of DN variants by gene level effects in selected KEGG pathway, hsa04310 Wnt signaling pathway. (a) probands, (b) siblings. DN variants data come from SSC study. Data are integrated and visualized on KEGG pathway graphs using Pathview (23).
fig. S6. An integrated view of DN variants by gene level effects in selected KEGG pathways, hsa04727 GABAergic synapse. (a) probands, (b) siblings. DN variants data come from SSC study. Data are integrated and visualized on KEGG pathway graphs using Pathview (23). Nodes in bold black box are synapse membrane proteins (neurotransmitter receptors, transporters and ion channels) hit by missense events. Effects of these events on protein functions are exhibited in Fig. 4c.
fig. S7. An integrated view of DN variants by gene level effects in selected KEGG pathways, hsa04724 Glutamatergic synapse. (a) probands, (b) siblings. DN variants data come from SSC study. Data are integrated and visualized on KEGG pathway graphs using Pathview (23). Nodes in bold black box are synapse membrane proteins (neurotransmitter receptors, transporters and ion channels) hit by missense events. Effects of these events on protein functions are exhibited in Fig. 4c. We also identified two clusters of interacting missense events as shown in blue and green dashed boxes. Detailed biology of these events is interpreted in Fig. 4d and fig. S11.
fig. S8. 1D protein domain structure and missense variants of genes in Wnt signaling pathway for both probands and siblings. It is recommended to view and interpret these events in the pathway context in fig. S6.
**Probands**

| Gene            | Variant(s)          | Domain/Function                      |
|-----------------|---------------------|--------------------------------------|
| ADCY5 (AC)      | A534T R603C         | Adenylate/Guanylate cyclase          |
| GABRA1 (GABAA)  | Y438C               | Neurotransmitter ligand binding      |
| GABA1 (GAT)     | A286V G299V         | Sodium:neurotransmitter symporter    |
| SLC6A1 (GAT)    | A450V V459M         | Sodium:neurotransmitter symporter    |
| GNAO1 (Gi/o, Gi/Go) | R313H          | G-alpha                              |
| ADRBK2 (GRK)    | K544N               | RGS Protein kinase PH                |
| GNAS (Gs)       |                     | G-alpha                              |
| GRM7 (mGluR7, mGluR7/8) | R822Q       | ANF Receptor ligand-gated ion channel |
| GRIN2B (NMDAR)  | C456Y R414C         | Ligand-gated ion channel C-terminus  |
| PRKCA (PKC)     | R238Q               | Protein kinase C-terminus            |
| CACNA1C (VGCC)  | V1182L              | Calcium channel IQ                   |
| CACNA1D (VGCC)  | A769G               | Calcium channel IQ                   |

**siblings**

| Gene            | Variant(s) | Domain/Function                      |
|-----------------|------------|--------------------------------------|
| GRIK1 (KA)      | I555T      | ANF Receptor Glutamate binding       |
| PRKCB (PKC)     | D235N      | Protein kinase C-terminus            |
| SHANK2 (SHANK)  | Y34D       | Variant PDZ SH3 SAM                  |

**fig. S9.** 1D protein domain structure and missense variants of genes in synapse pathways for both probands and siblings. It is recommended to view and interpret these events in the pathway context in figs. S7-8.
fig. S10. 1D protein domain structure and DN mutations of gene pairs in Wnt signaling and synapse pathways. These gene pairs are the same or similar genes hit by DN variants in both probands and siblings. Frame shift variants are labeled by suffix \(fs\) with position. It is recommended to view and interpret these events in the pathway context in figs. S6-8.
fig. S11. 1D and 3D protein structures and missense variants hitting mGluR inhibitor GRK (ADRBK2) and interacting G proteins. Pathway context is shown in the blue dashed box in fig. S7. Biology background and function impact of these missense variants are described in section S4.
fig. S12. Selected GO terms and emerging biological themes from the SSC exome variant data. GO analysis was done the same way as KEGG pathway analysis and results are listed in table S3. Significant terms from Biological Process or Molecular Function branches were considered here. These GO terms converge to 3 higher level biological themes within the synapse context, i.e. synapse function (transmission), morphology (wiring) and transcription (plasticity or dynamics). Each theme is supported by multiple GO terms. Each GO term associates with at least 1 theme, some with multiple themes.
table S1. The actual (and expected) counts of DN events by gene-level (columns) and pathway-level (rows) effects. Independent categories by gene level effects are Silent, LGD (likely gene disrupting), and Missense. Nonsilent group is a composite of LGD and missense, Sum includes all. Pathway level effects or targets include the Selected pathways, Others and All. We also did overrepresentation tests on each cell, and cell with significant p-values are marked (**p= 0.001; * p= 0.006).

|                  | Silent     | LGD (expected) | Missense (expected) | Nonsilent (expected) | Sum (expected) |
|------------------|------------|----------------|---------------------|---------------------|----------------|
| **Probands**     |            |                |                     |                     |                |
| (n=2508)         |            |                |                     |                     |                |
| Selected         | 83 (102.1) | 52 (34.6)**    | 134 (132.2)         | 186 (166.9)*        | 269            |
| Others           | 1209 (1189.9) | 386 (403.4) | 1539 (1540.8)       | 1925 (1944.1)       | 3134           |
| All              | 1292       | 438            | 1673                | 2111                | 3403           |

|                  | Silent     | LGD     | Missense | Nonsilent | Sum |
|------------------|------------|---------|----------|-----------|-----|
| **Siblings**     |            |         |          |           |     |
| (n=1911)         |            |         |          |           |     |
| Selected         | 62 (59.4)  | 16 (13.8)| 69 (73.8)| 85 (87.6) | 147 |
| Others           | 863 (865.6)| 199 (201.2)| 1079 (1074.2)| 1278 (1275.4)| 2141 |
| All              | 925        | 215     | 1148     | 1363      | 2288 |

table S2. Other pathways connected by three or more selected pathways in Table 1. Columns $j$ and $k$ are the marginal and conditional counts of selected genes. These pathways are marginally significant (p.val) in the pathway-level analysis, and likely relevant to ASD based on literature.

| SSC Pathways                                  | p.val          | q.val | p.con | q.con | j   | k   | Reference |
|-----------------------------------------------|----------------|-------|-------|-------|-----|-----|-----------|
| hsa04720 Long-term potentiation               | $1.38 \times 10^{-02}$ | 0.35  | 0.13  | 0.19  | 6   | 1   | (78)      |
| hsa04723 Retrograde endocannabinoid signaling | $3.09 \times 10^{-02}$ | 0.35  | 0.43  | 0.43  | 7   | 1   | (79, 80)  |
| hsa04510 Focal adhesion                       | $4.13 \times 10^{-02}$ | 0.41  | 0.04  | 0.41  | 11  | 11  | (81, 82)  |
| hsa04520 Adherens junction                    | $6.42 \times 10^{-02}$ | 0.46  | 0.06  | 0.46  | 5   | 5   | (83)      |
| hsa04020 Calcium signaling pathway            | $8.75 \times 10^{-02}$ | 0.52  | 0.09  | 0.52  | 9   | 9   | (84, 85)  |
**table S3. Significant GO groups selected from SSC, their test statistics, and references.** Columns \( j \) and \( k \) are the marginal and conditional counts of selected genes. These groups are likely drivers or disease causing groups due to the special analysis procedure (see Methods). Analysis was done the same way as for pathways in Table 1.

| Sub | SSC Pathways | p.val | q.val | p.con | q.con | \( j \) | \( k \) | Ref. |
|-----|---------------|-------|-------|-------|-------|--------|--------|------|
| BP  | GO:0016568 chromatin modification | 6.2×10^{-8} | 7.4×10^{-4} | 6.2×10^{-8} | 7.4×10^{-4} | 8 | 8 | (11, 86) |
| BP  | GO:0048812 neuron projection morphogenesis | 2.0×10^{-6} | 5.5×10^{-3} | 3.7×10^{-7} | 2.0×10^{-5} | 8 | 8 | (87) |
| BP  | GO:0032365 intracellular lipid transport | 2.7×10^{-4} | 9.4×10^{-2} | 1.6×10^{-4} | 6.9×10^{-3} | 7 | 7 | (77, 88) |
| BP  | GO:0051056 regulation of small GTPase mediated signal transduction | 1.6×10^{-4} | 6.6×10^{-2} | 1.9×10^{-4} | 4.8×10^{-3} | 10 | 9 | (49, 73, 74) |
| CC  | GO:0030054 cell junction | 4.2×10^{-6} | 5.6×10^{-3} | 4.2×10^{-6} | 5.6×10^{-3} | 12 | 12 | (53, 54, 83) |
| CC  | GO:0035097 histone methyltransferase complex | 1.7×10^{-3} | 1.4×10^{-1} | 1.2×10^{-3} | 1.8×10^{-2} | 14 | 10 | (11, 44) |
| CC  | GO:0031672 A band | 8.8×10^{-6} | 5.8×10^{-3} | 1.8×10^{-3} | 1.7×10^{-2} | 9 | 5 | |
| CC  | GO:0044459 plasma membrane part | 2.3×10^{-4} | 6.2×10^{-2} | 4.5×10^{-3} | 2.7×10^{-2} | 8 | 3 | |
| CC  | GO:0005694 chromosome | 3.6×10^{-3} | 1.8×10^{-1} | 1.2×10^{-2} | 8.8×10^{-2} | 8 | 2 | |
| CC  | GO:0005578 proteinaceous extracellular matrix | 4.5×10^{-3} | 1.9×10^{-1} | 1.4×10^{-2} | 4.2×10^{-2} | 8 | 8 | |
| MF  | GO:0005083 small GTPase regulator activity | 7.7×10^{-9} | 3.0×10^{-5} | 7.7×10^{-9} | 3.0×10^{-5} | 8 | 8 | (49, 73, 74) |
| MF  | GO:0032559 adenyl ribonucleotide binding | 1.3×10^{-9} | 9.6×10^{-5} | 1.4×10^{-9} | 3.1×10^{-5} | 7 | 7 | (75, 76) |
| MF  | GO:0018024 histone-lysine N-methyltransferase activity | 1.7×10^{-4} | 2.3×10^{-2} | 7.2×10^{-5} | 1.7×10^{-3} | 10 | 9 | (11, 44) |
| MF  | GO:0050839 cell adhesion molecule binding | 1.1×10^{-3} | 1.2×10^{-1} | 3.1×10^{-4} | 2.5×10^{-3} | 12 | 12 | (81, 89) |
| MF  | GO:0000993 RNA polymerase II core binding | 2.0×10^{-3} | 1.9×10^{-1} | 8.2×10^{-4} | 2.5×10^{-3} | 14 | 10 | (11) |
| MF  | GO:0015085 calcium ion transmembrane transporter activity | 1.9×10^{-3} | 1.8×10^{-1} | 1.8×10^{-3} | 3.5×10^{-3} | 9 | 5 | (84, 85) |
| MF  | GO:0030234 enzyme regulator activity | 1.1×10^{-9} | 9.6×10^{-5} | 2.0×10^{-9} | 2.0×10^{-9} | 8 | 3 | |
table S4. Validated and selected variants in the multilevel integrated analyses of SSC and ASC data. Events are selected at variant and gene level as described in Methods. For SSC, all validated variants are listed, including the silent group. Variants are defined based on hg19 reference genome.

table S5. Validated variants and functional annotations in the selected pathways of SSC data.

table S6. Genes hit by LGD or missense events in probands in selected pathways.