Building Connections in Autism

Autism is a collection of neurodevelopmental disorders characterized by impaired social skills, delayed language development, and repetitive behaviors. In classic (idiopathic) autism, these three characteristics predominate, whereas in syndromic autism disorders, these characteristics are part of a much larger set of symptoms. Clearly, common molecular pathways should underlie these similar disorders but trying to identify these pathways has proved difficult. Genetic studies have implicated genes mutated in syndromic autism disorders such as fragile X syndrome, Angelman syndrome, and Rett syndrome. Meanwhile, recent genome-wide association studies have pinpointed a large number of susceptibility loci that may contribute to autism spectrum disorders. Now, Zoghbi and her team turn their attention from genes associated with autism to the proteins they encode. These authors have compiled a network of protein-protein interactions based on autism-associated genes with the aim of finding common molecular pathways that underpin both classic and syndromic autism.

The authors divided autism-associated genes into three groups and then used a yeast two-hybrid screen of a human complementary DNA library to obtain interacting partners for the products of 26 autism-associated genes. They then built a protein-protein interaction network and from it pulled out very highly connected proteins such as those known to be mutated in fragile X syndrome. Their network showed that two proteins, SHANK3 and PSD95, located in the postsynaptic region of neurons and known to be associated with classic autism interact with each other and share nine common protein binding partners. SHANK3 also interacts with the TSC1 protein mutated in a form of syndromic autism called tuberous sclerosis complex, and these two proteins share 21 binding partners. SHANK3 and TSC1 also associate with two other postsynaptic proteins, ACTN1 and HOMER3. With their protein interaction network as a guide, the authors then performed microarray-based comparative genome hybridization for genes encoding 627 proteins in the network using DNA from 288 patients with classic autism. They discovered four new gene mutations including a mutation in the FLNA gene encoding a protein that turns out to bind to SHANK3. This new protein-protein interaction network provides a starting point for elucidating the common molecular pathways that may underlie both classic and syndromic autism. It also provides a framework on which to build a much bigger autism interactome that could be used to develop drugs with potential for treating several different forms of autism.
Protein Interactome Reveals Converging Molecular Pathways Among Autism Disorders

Yasunari Sakai,1,2 Chad A. Shaw,2*, Brian C. Dawson,1,2 Diana V. Dugas,2 Zaina Al-Mohtaseb,2 David E. Hill,3 Huda Y. Zoghbi1,2,4,5,6*

To uncover shared pathogenic mechanisms among the highly heterogeneous autism spectrum disorders (ASDs), we developed a protein interaction network that identified hundreds of new interactions among proteins encoded by ASD-associated genes. We discovered unexpectedly high connectivity between SHANK and TSC1, previously implicated in syndromic autism, suggesting that common molecular pathways underlie autistic phenotypes in distinct syndromes. ASD patients were more likely to harbor copy number variations that encompass network genes than were control subjects. We also identified, in patients with idiopathic ASD, three de novo lesions (deletions in 16q23.3 and 15q22 and one duplication in Xq28) that involve three network genes (NECAB2, PKM2, and FLNA). The protein interaction network thus provides a framework for identifying causes of idiopathic autism and for understanding molecular pathways that underpin both syndromic and idiopathic ASDs.

INTRODUCTION

Autism spectrum disorders (ASDs) are a heterogeneous group of neurodevelopmental disorders with three core features: impaired social skills (for example, gaze avoidance), delayed language development, and repetitive or stereotyped behaviors (1–4). “Classic” or idiopathic autism principally involves these three features; “syndromic” ASDs are those disorders in which the autistic phenotype is one aspect of a much broader clinical syndrome. In tuberous sclerosis complex (TSC), for example, which is caused by a single gene mutation in either the TSC1 or TSC2 gene, the core autistic phenotype is accompanied by seizures, developmental delay, cortical tubers, facial angiofibromas, and other skin lesions (5, 6). Similarly, Phelan-McDermid syndrome (PMS) is caused by microdeletions of chromosome 22q13.3 encompassing the SHANK3 gene (encoding SH3 and ankyrin domain containing protein) and is characterized by general hypotonia, seizures, intellectual disability, and ASD (7). Autistic features are commonly observed in fragile X syndrome, Angelman syndrome, phosphatase and tensin homolog (PTEN) hamartoma, Rett syndrome, and Timothy syndrome and in individuals with neuroligin mutations (table S1) (8). Most ASD cases (85 to 90%) do not show such clinically distinct phenotypes, and the genetic causes remain largely unknown; in these cases, the ASD is considered “nonsyndromic.”

Although ~50 genes or genomic variants that either cause or predispose individuals to ASDs have been identified (1–4), each accounts for no more than 0.5 to 2% of total ASD cases (9–11), and many of these correspond to the syndromic ASDs. Together, these genes and genomic variants account for at most 30% of all ASD cases (1, 12, 13). For nonsyndromic ASDs, copy number variations (CNVs) of submicroscopic DNA segments may prove more relevant; growing numbers of ASD-susceptibility loci have been reported based on genome-wide studies.

To make the challenge of understanding the pathogenesis of ASDs even more daunting, the known ASD-related proteins (including those mapping within CNVs) span diverse categories, from transcription factors (4, 14) and enzymes involved in protein modification (15) to cell adhesion molecules (16) and RNA binding proteins (17). Given the clinical heterogeneity of ASDs, it would not be surprising if mutations in hundreds or even thousands of genes cause ASD phenotypes—but might these genes converge on a few pathways?

We hypothesized that, if this were actually the case, it would best be revealed by a protein-protein interaction analysis of the existing autism-associated genes. Several years ago, when confronting a slightly different challenge with another group of neurological disorders, the inherited ataxias, we created an “ataxia interactome.” Using this interactome, we identified interactors of ataxia-associated proteins (18), mapped their interconnections, and uncovered unexpected functional relationships (19–21). Lacking a unifying neuropathology for the ASDs, we decided for the present study to rely on key phenotypic features to create a protein-protein interaction network that would reveal functional relationships between the gene products. Uncovering functional relationships among diverse ASD-related proteins is a first step toward the ultimate goal of developing therapies that might benefit multiple functionally or mechanistically related ASDs.

RESULTS

We began by identifying protein partners of ASD-associated proteins (1–3) and determining whether any of them interacted with other ASD proteins. We selected genes from three groups: those whose mutation results in syndromic ASDs (“syndromic ASD proteins,” table S1), those whose mutation causes severe language delay, and those whose products are paralogs, known binding partners, or functionally related to syndromic ASD proteins. We refer to the second and third groups as “ASD-associated” genes to distinguish them from the first group, syndromic ASD proteins (table S1). Although language delay can occur without accompanying autistic features, and the reasons for language delay may be as heterogeneous as the ASDs themselves,
we reasoned that language development might depend on the same pathways as those involved in social communication deficits observed in ASDs.

We performed a yeast two-hybrid (Y2H) screen of a human complementary DNA (cDNA) library using 192 bait fragments for 35 gene products, each of which encoded either full-length or partial segments of coding sequences. After a series of stringent tests (18, 22), we obtained 7933 interacting prey clones, which belonged to 783 unique proteins; 539 passed a second round of testing in yeast to demonstrate that the interactions held up in an independent reconstitution system (22). We considered only these 539 proteins as candidate binary interacting partners of the bait proteins. These 539 proteins had 848 interactions with 26 syndromic ASD proteins or ASD-associated proteins (tables S1 and S2). Among them, only 32 interactions (4%) were previously reported (tables S2 and S3). Baits for nine proteins failed to identify definite binding partners in this stringent Y2H screen (table S1).

To validate the interaction data in a mammalian system, we performed glutathione-Sepharose affinity copurifications in human embryonic kidney (HEK)293T cells for 52 randomly selected interactions (6% of the total) (fig. S1). The mammalian cells recapitulated 44 of 52 interactions (85%) (fig. S1 and table S2). In general, bona fide binary interactions do not validate at more than 50% (23–25) when different assay systems are used; our unusually high validation rate thus supports the reliability of our stringent screening methods compared to previous interactomes (18, 22, 26). We did not remove candidate pairs from the screen data (table S2) when they failed in the validation assays because the Y2H screen is excellent at detecting transient and biologically relevant interactions that are often difficult to recapitate in coimmunoprecipitation and affinity purification systems (18, 22–25).

Using the interaction data from the Y2H screen (table S2), we generated an ASD protein interaction network (Fig. 1A). Twenty-four of the 26 syndromic ASD or associated proteins were interconnected in one major component (CDKL5 and NFI1 were located outside) (Fig. 1A). The fragile X–related proteins 1 and 2 (FMR1 and FMR2) showed the greatest connectivity and were also connected to the fragile X mental retardation protein (Fig. 1B). The ASD network thus recapitulates and expands previously described in vivo interactions and functional relationships among the fragile X–related proteins (27, 28). Similarly, we expanded previously established relationships between TSC1 and TSC2 (29) by identifying four new partners (Fig. 1A and table S2). We also confirmed that the postsynaptic proteins SHANK3 and PSD95 interact in vivo (30, 31) and identified nine shared partners between them (ACTN2, CLU, DLGAP1, DLGAP3, DLGAP4, HNRNPC, LZTS2, PICK1, and SYNGAP1) (Fig. 1C).

Notably, our ASD interactome revealed previously unsuspected connectivity between two syndromic ASD proteins, SHANK3 and TSC1, which share at least 21 partners (Fig. 2A). SHANK1, the paralog of SHANK3, arose as a potential partner of both TSC1 and SHANK3. This suggested that these syndromic ASD proteins interact in a complex in the neuronal postsynaptic compartment (the microenvironment of dendritic spines beneath the specialized membrane structure of the synapse) (3, 32), a suggestion confirmed in vivo by coimmunoprecipitation using mouse brain extracts (Fig. 2, C and D). In vivo studies further confirmed that ACTN1, a postsynaptic scaffold protein identified as a partner of SHANK3 and TSC1 (table S2 and Fig. 2A), interacts with TSC1 as well as with the postsynaptic scaffolding proteins SHANK3 and HOMER3 (Fig. 2E). Further, the Y2H screen recapitulated 11 previously reported in vivo interactions (tables S2 and S3), which we obtained from the Human Protein Reference Database (HPRD) (33) and the Biological General Repository for Interaction Datasets (BioGRID, http://thebiogrid.org/). All except TSC1-AXIN1 were reported in mouse brain tissue. We also identified an additional 21 interactions, previously demonstrated by in vitro copurifications or Y2H screens (table S3).

To verify that the syndromic ASD proteins are coexpressed with their binding partners in vivo, we analyzed microarray data from our previous studies of brain tissue from wild-type mice (34, 35). We observed strong correlations among expression profiles for genes in the network, but not for similarly sized sets of randomly selected probes on the microarrays (P < 0.001, Fig. 3A). Further, most of the genes that encode the proteins annotated in the network coclustered into a dominant coexpression group in the hypothalamus (78%), cerebellum (78%), and amygdala (55%) (Fig. 3B). In conjunction with the physical interaction data (Fig. 2, B to D), SHANK3, TSC1, ACTN1, and HOMER3 showed highly correlated expression in these brain regions (table S4). Averaged correlation matrices of the genes in the three brain regions did not show such strong correlation, suggesting that subsets of the ASD-associated proteins and their binding partners may enjoy unique relationships in different brain regions, rather than be ubiquitously coexpressed (Fig. 3B). Ninety-six percent of the proteins identified in our primary screen were found to be expressed in brain in the mouse studies, a substantially greater proportion than the expected 59% for randomly sampled genes (P < 1 × 10−19).

To understand the unique topology (that is, the pattern of interconnections) of the protein interaction network and systematically assess the connectivity of syndromic ASD proteins, we incorporated literature-curated interaction data for both bait and prey proteins from the HPRD and the BioGRID. This allowed us to produce an extended network consisting of one component with 3507 proteins connected through 6881 interactions (Fig. 4A). Of the 35 bait proteins that were used in the Y2H screen, 34 were directly or indirectly connected inside this network; only one protein (SLEC6A8) was not connected. Next, we calculated the mean path length in the extended network for 8 syndromic ASD proteins (tables S1 and S2) that were in the experimental network (Fig. 1A), and we compared it with the distribution of mean path lengths for 8 randomly sampled proteins selected from the remaining 18 of 26 bait proteins that had at least one binding protein in the primary screen (table S2). We performed 10,000 random draws of eight proteins. The eight syndromic ASD proteins showed a significantly shorter mean path length (2.14) than the random samples from the remaining baits (mean of 2.78, P = 0.004, fig. S2). The close connectivity of the eight syndromic ASD proteins led us to investigate whether different ASD proteins might share common molecular pathways that relate to the pathogenesis of ASD. Indeed, Gene Ontology (GO) analysis of the network (excluding all of the baits) revealed marked enrichment for proteins associated with synapse, postsynaptic density, and cytoskeleton under the “Cellular Component” branch of the GO (fig. S3A), and for small guanosine triphosphatase (GTPase)–mediated signaling and metabotropic glutamate receptor signaling under the “Biological Process” GO branch (fig. S3B) (a biological process describes a series of molecular events or functions). Such coherence between the cellular compartments and biological processes of the ASD baits and their interactome partners underscores the biological value of the network. It also points to key molecular pathways responsible for autistic phenotypes in distinct genetic syndromes.
**Fig. 1.** Landscape of protein-protein interactions for syndromic ASDs and associated disorders. (A) Protein interaction network derived from the Y2H screens. Inset indicates color codes for baits, prey, and interactions. Twenty-six bait proteins are included. (B) Fragile X mental retardation protein FMRP and its paralogs (FXR1 and FXR2) are highly connected via shared interactors. (C) The synaptic proteins SHANK3 and PSD95 share nine interactors.
It remained to be determined whether a protein interaction network built on syndromic ASD proteins would prove relevant to the pathogenesis of nonsyndromic or idiopathic ASDs. To address this question, we collected information from published studies on CNVs that were observed in normal populations or in nonsyndromic ASD patients (36, 37). We then searched for genes that were annotated both in our network and in the intervals of CNVs found in normal individuals or ASD patients. Individuals from the ASD group showed an increased rate of CNVs spanning genes in the Y2H interactome compared to the control group by a factor of 2.4 (incidence, 0.43 versus 0.18; P = 6.16 × 10^{-8}; two-sided Fisher’s exact test; “Core,” Fig. 4B) with an odds ratio of 3.3. Conversely, there was a lower rate of individuals in the ASD group whose CNVs failed to encompass genes in the network, that is, fewer ASD CNVs mapped to loci encoding non-network proteins (0.25 versus 0.39 for the control group, P = 0.70 versus 0.565 in frequency, “Extended,” Fig. 4B).

To consider both the connectivity of the network and the mult-CN V load in each individual, we computed an additional measure that we defined as the network connectivity score. This score represents the sum of the number of connections in the network of all genes present in CNVs of each individual; the score therefore takes into consideration the contribution of genes by their network relevance. This score was also significantly higher in ASDs versus controls.

**Fig. 2.** In vivo validation of the network’s connectivity. (A) The protein interaction data show high connectivity between SHANK3 and TSC1. The shared binding partners (purple nodes) between SHANK3 and TSC1 are annotated with a magnified view. (B) ACTN1, SHANK1, and HOMER3 were identified as common binding partners for the two proteins. (C to E) In vivo interactions among endogenous TSC1, SHANK, HOMER, and ACTN1 in mouse brain extracts. The reciprocal coimmunoprecipitation was performed with anti–pan-SHANK (C) or anti-TSC1 (D) antibodies. Proteins in the immunoprecipitated (IP) complex were labeled with the indicated antibodies. Antibodies against SHANK3 and HOMER3 were used to monitor stable protein complex formation in the immunoprecipitated samples with anti–pan-SHANK antibody. Normal rabbit IgG was used for the negative control (IgG). (E) Coimmunoprecipitation with anti-ACTN1 shows that it interacts with TSC1, SHANK3, and HOMER3 in vivo. WB, Western blot.
To explore the role of genes in the interaction network in idiopathic ASDs, we performed microarray-based comparative genome hybridization (CGH) for 627 genes in our network using genomic DNA from 288 relatively high-functioning individuals (average IQ, 80.94) with a diagnosis of idiopathic ASD (that is, nonsyndromic autism) from the Simons Foundation Simplex Collection (39). These probands do not show any signs of syndromic disorders (systemic malformation, abnormal facies, or severe intellectual disability) on physical examination or brain imaging. We focused on events with large segmental duplications or deletions spanning more than 10 kb. This analysis revealed a segmental duplication in chromosome Xq28 involving FLNA and three segmental deletions in chromosomes 15q13.3, 16q23.3-q24.1, and 14q13.3, which involved the PKM2, NECAB2, and MIPOL1 genes, respectively (Fig. 5, A to D). CGH analyses of the DNA from the parents of the probands confirmed that the duplication of Xq28 (FLNA) and deletions of 15q13.3 (PKM2) and 16q23.3 (NECAB2) were all de novo, whereas the deletion of 14q13.3 (MIPOL1) was maternally inherited (Fig. S5, A to D, and table S5). Duplications and point mutations of FLNA cause various degrees of intellectual disability, periventricular heterotopia (a disorder caused by abnormal migration of neurons), and dysmorphic features but, to our knowledge, have not been associated with an autistic phenotype in the absence of intellectual deficits (40). By identifying a de novo duplication of FLNA in a patient with an IQ of 109 and autism, we broaden the clinical spectrum of phenotypes associated with such duplications. Furthermore, our discovery that FLNA binds to SHANK3 (mutation of which causes syndromic ASD) using both Y2H assays (table S2) and coimmunoprecipitation (Fig. S6) in mouse brain extracts validated the physical interaction of these proteins and shows that both syndromic and nonsyndromic ASDs are functionally linked.

None of the autosomal deletion events were observed in the clinical database of the Molecular Genetics Laboratory at Baylor College of Medicine (BCM), which houses samples from more than 15,000 patients with dysmorphology (abnormal form or anatomy) or intellectual and developmental disability screened with a high-density oligonucleotide clinical chromosomal microarray. The BCM clinical array has much greater genome-wide coverage and was able to delineate precise boundaries of the initial CNV findings in our experimental cohort for NECAB2, PKM2, and FLNA. In the clinical database, only three males and one carrier mother had the duplication event of Xq28 that involves FLNA but not MECP2, the gene involved in the neurological disorder Rett syndrome (14, 41). All male patients had developmental disabilities and cognitive deficits, but the female patient was asymptomatic (as expected for an X-linked defect). Thus, all four segmental CNVs confirmed in this study were extremely rare structural variations rather than polymorphic events. We turned to an additional study (42) to examine a set of cognitively normal control individuals for these events using an extremely high-density tiling array. There were...
no deletion events covering the three genes (PKM2, NECAB2, and MIPOL1) or the FLNA duplication event in these control individuals. Furthermore, there were no deletion events overlapping these three deleted genes nor duplication events overlapping with FLNA loci among the CNV data from 1500 controls (36) that we used for Fig. 4B.

**DISCUSSION**

We have developed a protein interaction network for ASD-causing and ASD-associated proteins (together referred to as ASD proteins). We began by identifying protein partners of ASD proteins (1, 2, 8, 13) and determining whether any of them interact with other ASD proteins. Although most ASD-related proteins identified cause syndromic ASDs, we reasoned that converging pathways might provide insight into nonsyndromic or idiopathic ASDs. Indeed, our interactome revealed an unexpected convergence around several proteins. Although the interactome validated several interactions that had been previously identified in the literature, it did not identify others. There are several reasons for this: The screen was not done to saturation; the interactome identifies direct interactions only, unlike coimmunoprecipitation, which identifies direct and indirect interactions; protein interactions vary according to tissue; and our system has relatively low expression. Thus, we are more likely to have false negatives (that is, missed valid interactions) than false positives. The large number of validated interactions, however, demonstrates that this interactome is a robust framework for future studies exploring relationships among ASD proteins. Further, the significant overlap between CNVs from a previously published ASD cohort and our experimental network supports its utility as a platform for ASD gene discovery. Finally, the network can also shed light on other questions in ASD. For example, we found only one overlapping gene (CBS) on chromosome 21, despite evidence that Down syndrome (caused by trisomy of chromosome 21) shows some comorbidity with ASDs. This finding suggests that the ASD phenotype in Down syndrome may differ from the phenotypes of other nonsyndromic ASDs.

An important aspect of the interactome is the connections it reveals between

---

Fig. 4. Overlap of the protein interaction network with CNV regions identified in nonsyndromic ASD patients. (A) Extended ASD protein interaction network. Nodes and lines indicate the 11 syndromic ASD proteins (red), the 23 ASD-associated proteins used as bait in Y2H (pink), new binding partners identified in this study (purple), known binding partners (orange), new interactions (green), and previously known interactions (light blue). (B) Bar plots show the relative frequency (y axis) of ASD or control individuals harboring CNV overlap with genes in the interaction network. Bait, bait proteins used for the Y2H screen; Core, network proteins identified in this study; Extended, network proteins identified by literature searches; Non-network proteins, proteins not present in the network. The values represented by each bar do not sum up to one because an individual may have more than one CNV overlapping the core and extended network. (C) The curves display density estimates (y axis) of the total network connectivity score (x axis) for the genes encompassed by all CNVs in each individual from the ASD or control groups. The total network connectivity score is the log2 of the sum of network connectivity of all genes harbored in CNVs of each individual in each cohort. The ASD and control groups are significantly different (P < 2.2 x 10^-16, Wilcoxon’s rank sum test).
syndromic ASD-causing and ASD-associated proteins in one network. Higher connectivity between two ASD-causing proteins suggests that they might interact in a protein complex or that they might function in a common molecular pathway. Here, we uncovered in vivo interactions between the TSC1 and SHANK proteins. Further exploration of such interactions should shed light on the pathogenic mechanisms underlying the autistic features observed in TSC (8, 43) and PMS (7, 44–47). The TSC1 and TSC2 proteins regulate mammalian target of rapamycin (mTOR), a promoter of protein synthesis in response to growth factors and stress (48); up-regulation of protein synthesis due to functional loss of either of the TSC proteins is a likely mechanism for the ASD phenotypes of both disorders (8). The molecular mechanism that specifies the location of abnormal protein synthesis, however, remains to be determined. Our interactome suggests that these two distinct pathways are brought together in one protein complex built upon the postsynaptic scaffold protein SHANK3.

Links between SHANK3 and other ASD proteins lend support to the idea that common pathways lead to the broader ASD phenotypes (3, 4, 8). In vivo studies of mice carrying mutant alleles of Shank3 and displaying autism-like phenotypes highlight the importance of the SHANK3 protein for maintaining the normal levels of many synaptic proteins that are critical for glutamatergic signaling (44, 45, 47). Recently reported pathogenic mutations in the SHANK2 gene in sporadic ASD cases (9, 49) further support our notion that various ASD-associated proteins, including FLNA, functionally interact with the SHANK proteins. These findings have therapeutic implications. Benefits reported from preclinical trials using rapamycin in Pten or Tsc2 mutant animals (50, 51) raise the possibility that such therapies might be beneficial in broader groups of patients with syndromic and nonsyndromic ASDs.

In summary, we developed an ASD interactome that facilitates classification of ASDs according to functional pathways and interacting proteins. The short-term utility of this interactome will be to increase our molecular diagnostic capabilities with a raft of new genes. The mid- and long-term benefits will come from advancing pathogenesis studies to promote development of rationally designed therapeutics that could be used to treat more than one ASD.

Fig. 5. New CNVs in idiopathic ASD that overlap genes encoding proteins in the interactome. (A) The proband (11233) shows 5-Mb segmental deletion in chromosome 15q23, spanning the PKM2 gene. (A to D) The chromosomal location of events (top), plots of log2 ratios (middle), and the gene structure for the regions (bottom) are from the UCSC browser. Red (A to C) and green (D) dots in the hybridization plots and the colored lines of each panel indicate the intervals of segmental deletion and duplication. Arrows denote genes in the network. (B) The proband (11327) shows 5.9-Mb segmental deletion in 16q23.3-q24.1 spanning the NECAB2 gene. (C) The proband (11505) shows partial deletion (43 kb) of the MIPOL1 gene at 14q13.3. (D) The proband (11092) shows a 200-kb segmental duplication in Xq28 spanning the FLNA gene.
MATERIALS AND METHODS

Y2H screen
All cDNAs encoding full-length or partial domains of bait proteins (table S1) were cloned into DB-dest vectors with the Gateway system (Invitrogen) as described (18, 22). The constructs encoding each DB-autism fusion protein were transfected into yeast cells, MaV203 (Mat-alpha), and screened for positive interactions from 1 × 10^5 to 2 × 10^6 independent clones in human brain cDNA libraries (Invitrogen).

Coimmunoprecipitation for mouse brain extracts
Whole mouse brains were freshly homogenized in buffer containing 320 mM sucrose, 5 mM Hepes (pH 7.4), and 1 mM EDTA as described (52). Total protein (1 mg) from cytosolic (S2) fractions was incubated at 4°C for 2 hours with either anti–pan-SHANK (5 μg of antisera), anti-TSC1 (1 μl), and anti-ACTN1 (5 μl) antibodies or normal rabbit immunoglobulin G (IgG) (5 μg) in the TS buffer (150 mM NaCl, 10 mM tris-HCl, pH 7.4) supplemented with 1% hemoglobin. Immune complexes were precipitated with prewashed protein A agarose beads (Millipore) by incubating at 4°C for another 1 hour and were washed in the TS buffer five times and extracted with SDS-loading buffer.

CGH microarray analysis
We determined the exon coordinates of the genes that were annotated in the protein interaction network and then selected a total of ~42,000 oligonucleotide probes from Agilent’s open resource library to design a custom microarray for CGH studies. Using the custom 4×44K chips, we performed DNA digestion, labeling, and hybridization as previously described (53, 54). Because the CNV regions for NECAB2, PIK3M, and FLNA exceeded the targeted gene loci, we used the BCM Medical Genetics Laboratories (MGL) clinical array (CMA BAC V8.1, Baylor MGL) to determine boundaries of the events (53). All the intervals of CNV events and other information referred to in this study were defined with the assembly of NCBI36/hg18 at UCSC (University of California, Santa Cruz) genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

Gene expression analysis
We used HomoloGene (http://www.ncbi.nlm.nih.gov/homologene) release 6.1 mapping human genes to their mouse orthologs. Of the 565 in the network, we were able to map 478 (85%) to unique mouse genes. We then obtained gene expression data from wild-type mouse hypothalamus, cerebellum, and amygdala samples using data from previous studies (34, 35). We computed Spearman’s rank correlation coefficient for all network gene pairs and determined the median. To compare our results with a random set of genes, we sampled 1000 iterations of an equal number of random genes from the array and computed the same median Spearman correlation measure. The distributions for the median correlation values for random genes are plotted for each brain region. The median values for our mapped network proteins are shown with three vertical lines. In addition, we computed the correlation heat map between all pairs of genes, and we sorted the genes according to cluster analysis based on pairwise correlations.

CNV database analysis
We obtained all variant data from dbVAR (http://www.ncbi.nlm.nih.gov/dbvar/) for ASD (37) and controls (36), which contained CNV data from 419 and 1552 samples, respectively. We calculated the overlap with the protein interaction network for each CNV in terms of the network genes, the network node degree, and network components (Bait, Core, or Extended). Data were processed to show over-lapping events per individual. We computed two summaries: overlap indicators for each person with the components of the interaction network (Bait, Core, and Extend) and network connectivity score defined as log2 of the sum of network degree for overlapping genes. Individuals with no network overlap were treated as missing values for the connectivity score.

Further details for data analysis are provided in the Supplementary Material.

SUPPLEMENTARY MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/3/86/86ra49/DC1

Materials and Methods
Fig. S1. Validation of the interaction data from the Y2H study by coaffinity purification experiments in HEK293T cell.
Fig. S2. Syndromic ASD proteins are highly connected to each other and have the shortest path length in the protein interaction network.
Fig. S3. The GO analysis for the extended ASD network.
Fig. S4. Chromosomal location of the overlapping genes between the protein interaction network and CNV intervals of ASD group.
Fig. S5. Confirmation of de novo or inherited CNVs.
Fig. S6. Physical interaction of FLNA with SHANK3 in vivo.

Table S1. Summary of yeast two-hybrid screening for ASD and related disorders.
Table S2. Summary of protein interaction data obtained from Y2H screening and Co-AP study.
Table S3. Summary of known interaction data in the literature and the overlap with this study.
Table S4. Coexpression pattern of the genes in the mouse hypothalamus, cerebellum, and amygdala.
Table S5. Summary of the probands with novel CNVs.

REFERENCES AND NOTES

Received Article
H. Y. Zoghbi, Opposing effects of polyglutamine expansion on native protein complexes con-
Drosophila
H. Y. Zoghbi, J. Botas, dAtaxin-2 mediates expanded Ataxin-1-induced neurodegeneration in a
development and function.
J. M. Silverman, R. Bernier, S. E. Levy, R. T. Schultz, G. Dawson, T. Owley, W. M. McMahon,
L. Klei, J. Piven, K. J. Meyer, E. Anagnostou, T. Sakurai, R. M. Game, D. S. Rudd,
S. Naisbitt, E. Kim, J. C. Tu, B. Xiao, C. Sala, J. Valtschanoff, R. J. Weinberg, M. Sheng,
A. Itsara, H. Wu, J. D. Smith, D. A. Nickerson, I. Romieu, S. J. London, E. E. Eichler, De novo
mutations in X-linked methyl-CpG-binding protein 2.
S. Ben-Shachar, M. Chahrour, C. A. Shaw, H. Y. Zoghbi, Mouse models of MeCP2
phosphorylation of the fragile X mental retardation protein (FMRP) and displaying selective
interacting with the fragile X mental retardation protein (FMRP) and displaying selective
interaction on January 4, 2012 www.sciencemag.org
Downloaded from www.sciencemag.org on January 4, 2012

J. P. Bradford, P. M. Sleiman, C. E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Li, A. Barabasi, M. Vidal,
A. Chesneau, T. Hao, H. D. Goldberg, N. I. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari,
S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Witternike, S. Wood,
J. J. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger Jr., A. D. Paterson, M. A. Pericak-Vance,
G. J. Bassell, S. T. Warren, Fragile X syndrome: Loss of local mRNA regulation alters synaptic
development and function.
K. Wang, H. Zhang, D. Ma, M. Bucan, J. T. Glessner, B. S. Abrahams, D. S. Rudd, A. Imieliński,
J. P. Bradford, P. M. Sleiman, C. E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Li, A. Barabasi, M. Vidal,
A. Chesneau, T. Hao, H. D. Goldberg, N. I. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari,
S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Witternike, S. Wood,
J. J. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger Jr., A. D. Paterson, M. A. Pericak-Vance,
J. P. Bradford, P. M. Sleiman, C. E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Li, A. Barabasi, M. Vidal,
A. Chesneau, T. Hao, H. D. Goldberg, N. I. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari,
S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Witternike, S. Wood,
J. J. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger Jr., A. D. Paterson, M. A. Pericak-Vance,
J. P. Bradford, P. M. Sleiman, C. E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Li, A. Barabasi, M. Vidal,
A. Chesneau, T. Hao, H. D. Goldberg, N. I. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari,
S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Witternike, S. Wood,
J. J. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger Jr., A. D. Paterson, M. A. Pericak-Vance,
J. P. Bradford, P. M. Sleiman, C. E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Li, A. Barabasi, M. Vidal,
A. Chesneau, T. Hao, H. D. Goldberg, N. I. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari,
S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Witternike, S. Wood,
J. J. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger Jr., A. D. Paterson, M. A. Pericak-Vance,
J. P. Bradford, P. M. Sleiman, C. E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Li, A. Barabasi, M. Vidal,
A. Chesneau, T. Hao, H. D. Goldberg, N. I. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari,
S. Wallace, K. Wang, Z. Wang, T. H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Witternike, S. Wood,
J. J. Haines, J. Hallmayer, J. Miller, A. P. Monaco, J. I. Nurnberger Jr., A. D. Paterson, M. A. Pericak-Vance,
J. P. Bradford, P. M. Sleiman, C. E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Li, A. Barabasi, M. Vidal,
A. Chesneau, T. Hao, H. D. Goldberg, N. I. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari,


57. Acknowledgments: We thank L. White, L. Liles, and M. Hoang at BCM microarray core; L. Lewis at BCM Graduate Student Council; BCM Medical Genetics Laboratories for the array-based CGH study; A. McCall, D. Walker, M. Strivens, and M. Rao for technical assistance; M. Sheng, W. Dobyns, D. Picketts, R. Gibbons, S. Dindot, A. Beaudet, D. Nelson, S.-K. Lee, B. Franco, and I. Bezpoxovany for antisera and expression constructs; M. Sardellos, C. Skaaf, M. Costa-Mattioli, and J. Neul for critical comments; C. Schaaf for computing IQ averages of the Simons Simplex Collection patients used in this study; V. Brandt for comments on the manuscript; and H.Y.Z. lab members for helpful discussions. Funding: Supported by the Howard Hughes Medical Institute (H.Y.Z.), the Simons Foundation Autism Research Initiative (SFARI award to H.Y.Z.), and the Ellison Foundation (D.E.H.; awarded to M. Vidal). We are grateful to all of the families at the participating SFARI Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goix-Kochel, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, B. Peterson, J. Piggott, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, and E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC population data set described in this study by applying at https://base.sfari.org. The array CGH data for the probes analyzed in the study have been deposited in National Center for Biotechnology Information’s Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE29576 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29576). Author contributions: Y.S., C.A.S., and H.Y.Z. designed the study, evaluated the data, and wrote the manuscript. Y.S. performed all the experiments with technical support for the Y2H screen by Z.A.M. C.A.S., B.C.D., and D.V.D. conducted the bioinformatic analysis. D.E.H. provided the OffFemre clone and edited the manuscript. Submitted 20 January 2011 Accepted 19 May 2011 Published 8 June 2011 10.1126/scitranslmed.3002166