Increased burden of deleterious variants in essential genes in autism spectrum disorder

Xiao Ji^{a,b}, Rachel L. Kemner^b, Christopher D. Brown^{b,1}, and Maja Bućan^{b,c,1}

^aGenomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; ^bDepartment of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; and ^cDepartment of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Edited by Eugene V. Koonin, National Institutes of Health, Bethesda, MD, and approved November 7, 2016 (received for review August 9, 2016)

Autism spectrum disorder (ASD) is a heterogeneous, highly heritable neurodevelopmental syndrome characterized by impaired social interaction, communication, and repetitive behavior. It is estimated that hundreds of genes contribute to ASD. We asked if genes with a strong effect on survival and fitness contribute to ASD risk. Human orthologs of genes with an essential role in pre- and postnatal development in the mouse [essential genes (EGs)] are enriched for disease genes and under strong purifying selection relative to human orthologs of mouse genes with a known nonlethal phenotype [nonessential genes (NEGs)]. This intolerance to deleterious mutations, commonly observed haploinsufficiency, and the importance of EGs in development suggest a possible cumulative effect of deleterious variants in EGs on complex neurodevelopmental disorders. With a comprehensive catalog of 3,915 mammalian EGs, we provide compelling evidence for a stronger contribution of EGs to ASD risk compared with NEGs. By examining the exonic de novo and inherited variants from 1,781 ASD quartet families, we show a significantly higher burden of damaging mutations in EGs in ASD probands compared with their non-ASD siblings. The analysis of EGs in the developing brain identified clusters of coexpressed EGs implicated in ASD. Finally, we suggest a high-priority list of 29 EGs with potential ASD risk as targets for future functional and behavioral studies. Overall, we show that large-scale studies of gene function in model organisms provide a powerful approach for prioritization of genes and pathogenic variants identified by sequencing studies of human disease.

essential genes | mouse knockouts | mutational burden | autism spectrum disorder | coexpression modules

Autism spectrum disorder (ASD) is a heterogeneous, heritable neurodevelopmental syndrome characterized by impaired social interaction, communication, and repetitive behavior (1, 2). The highly polygenic nature of ASD (3–5) suggests that the analysis of the full spectrum of sequence variants in hundreds of genes will be necessary for deeper understanding of disrupted neuronal function. Prioritization of ASD risk genes initially focused on known pathways with recognized relevance to pathogenesis of ASD, such as synaptic function and neuronal development (6). However, combined analyses of de novo, inherited, and case–control variation in over 2,500 ASD parent–child nuclear families identified around 100 genes contributing to ASD risk (7–9), converging on pathways implicated in transcriptional regulation and chromatin modeling in addition to synaptic function.

The main challenge in the current understanding of genetic architecture of ASD comes from a need to study the interplay between variants with a high effect (for example, recurrent de novo variants) and a background of variants with an intermediate effect but that nevertheless still disrupt proper neuronal development. Essential genes (EGs) or genes that are necessary for successful completion of pre- and postnatal development are prime candidates for the source of this background or load of variants with a cumulative intermediate effect. EGs are highly enriched for human disease genes and under strong purifying selection (10–14). In addition to intolerance to loss-of-function

and deleterious mutations, the functional impact of EGs is reflected by haploinsufficiency that is commonly observed in heterozygous mutations (11, 15). In addition to their role in defining a “minimal gene set” (16, 17), EGs tend to play important roles in protein interaction networks (18). Therefore, one may consider that EGs are involved in rate-limiting steps that affect a range of disease pathways (19).

Recently, three large-scale screens (gene trap and CRISPR-Cas9) have been performed to assess the effect of single-gene mutations on cell viability or survival of haploid human cancer cell lines (“cell-based essentiality”) (20–22). These studies identified an overlapping core set of genes that were essential in the majority of cell lines tested ($n = 956$), although a subset of genes were essential in specific cell lines. In an alternative and complementary approach, we assembled a catalog of human orthologs of EGs in the mouse ($n = 3,326$) (14) based on the organismal-level phenotypes of loss-of-function mouse mutants from the Mouse Genome Informatics (MGI) database (23) and the International Mouse Phenotyping Consortium (IMPC) web portal (24). Based on these data, homozygous loss-of-function mutations in 3,326 genes lead to prenatal or preweaning lethality, with a significant overlap between the core set of human cell EGs and human orthologs of EGs in the mouse (14). These studies are consistent with 30% (or ~6,000) of protein-coding genes to be essential for pre- and postnatal survival (14, 25).

A deeper understanding of the mutational spectrum of EGs in a neurodevelopmental disorder, such as ASD, is important, because EGs are less likely to be redundant, are more likely to have functional consequences when mutated, and may produce a gradation of phenotypes (25). Our previous work reported an enrichment of EGs among genes with de novo mutations in ASD

Significance

Essential genes (EGs) are necessary for survival and the development of an organism. Our study is focused on investigating the role of EGs in autism spectrum disorder (ASD). With a comprehensive catalog of 3,915 mammalian EGs, we show that there is both an elevated burden of damaging mutations in EGs in ASD probands and also, an enrichment of EGs in known ASD risk genes. Moreover, the analysis of EGs in the developing brain identified clusters of coexpressed EGs implicated in ASD. Overall, we provide evidence that genes that are essential for survival and fitness also contribute to ASD risk and lead to the disruption of normal social behavior.

Author contributions: X.J., R.L.K., C.D.B., and M.B. designed research; X.J. performed research; X.J. analyzed data; and X.J., R.L.K., C.D.B., and M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence may be addressed. Email: chrbro@mail.med.upenn.edu or bucan@mail.med.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613195113/-DCSupplemental.

patients (11). Several groups reported an enrichment of de novo and rare inherited single-nucleotide loss-of-function variants in ASD probands (8, 26), although there is a depletion of damaging mutations in ASD risk genes in population controls (12, 27, 28). In this report, we compiled, to our knowledge, the most comprehensive list of human EGs and extended the analysis to both de novo and inherited damaging variants in 1,781 ASD families. In addition to disease status, we further showed the effect of damaging variants in EGs on ASD-related traits, such as the social skill measurement in 2,348 ASD probands. Finally, we performed coexpression analysis of EGs in the developing human brain to identify clusters of interacting EGs that contribute to ASD risk and suggest ASD candidate genes.

Results

To identify the most comprehensive set of EGs in mammals, we combined the set of human orthologs of EGs in the mouse ($n = 3,326$) (14) with a set of human “core EGs” ($n = 956$) that were found to be essential in cell-based assays (20–22). Based on a significant overlap between tested mouse and human EGs (14), we expanded our original set of 3,326 EGs with the addition of nonoverlapping 589 EGs identified only in human cell lines for a total of 3,915 EGs (*SI Materials and Methods* and *Dataset S1*). In our subsequent analyses, we compared features of and genetic variation in these EGs with 4,919 human orthologs of genes with reported nonlethal phenotypes in the mouse [nonessential genes (NEGs)].

Homozygous loss-of-function mutations in EGs lead to lethality (or miscarriages in humans) and as such, cannot contribute to disease. Although we and others reported a depletion of loss-of-function mutations in EGs in humans (11, 12, 14), heterozygosity for a loss-of-function mutation or other “milder” alleles in EGs may contribute to both dominant and recessive diseases. We illustrated this point using a catalog of disease-linked genes in Online Mendelian Inheritance in Man (29) (*SI Materials and Methods*); EGs were enriched relative to NEGs in 1,000 genes underlying dominant diseases (odds ratio = 1.95, P value = 3.17×10^{-19} ; two-sided Fisher’s exact test) and 1,645 genes underlying recessive disease (odds ratio = 1.52, P value = 4.94×10^{-11} ; two-sided Fisher’s exact test) (Fig. 1*A*). A stronger enrichment of EGs among genes underlying dominant disease implies that dominant negative alleles and haploinsufficiency play an important role. We provide multiple lines of evidence for higher probability of haploinsufficiency of EGs (Fig. 1*A* and *SI Materials and Methods*). First, using the systematically rated dosage-sensitive genes from ClinGen (30), we found that EGs were significantly enriched compared with NEGs and that the levels of EG enrichment positively correlated with levels of evidence supporting dosage sensitivity of rated genes (odds ratio = 3.94, P value = 5.07×10^{-20} for “sufficient evidence”; odds ratio = 5.26, P value = 7.08×10^{-5} for “some evidence”; odds ratio = 2.52, P value = 0.0106 for “little evidence”; odds ratio = 1.14, P value = 0.608 for “not dosage sensitive”; two-sided Fisher’s exact test). Second, as an extension of the earlier findings from the work by Georgi et al. (11), we confirmed the enrichment of EG relative to NEG for 262 human haploinsufficient genes (31) with the updated EG and NEG list (183 EGs vs. 62 NEGs; P value = 1.64×10^{-22} , odds ratio = 3.84; two-sided Fisher’s exact test). Third, EGs are significantly overrepresented among 313 human orthologs of mouse genes with heterozygous alleles associated with mutant phenotypes from the MGI (23) (odds ratio = 3.43, P value = 2.74×10^{-23} ; two-sided Fisher’s exact test). Fourth, with two genome-wide prediction models of haploinsufficient genes in the human genome (32, 33), we observed that EGs have significantly higher probability of exhibiting haploinsufficiency compared with NEGs (P value < 2.2×10^{-16} for both models; two-sided Wilcoxon rank sum test) (Fig. 1*B* and *C* and *SI Materials and Methods*). Based on our findings that EGs linked to Mendelian disease are overwhelmingly

Fig. 1. Haploinsufficiency of EGs. (*A*) For each class of genes with different essentiality status (EG in red, NEG in turquoise, and unknown in gray), the proportion of genes among each gene set of interest is plotted in *Left*. Dosage-sensitive genes from ClinGen (30) were classified into five categories (1, sufficient evidence; 2, some evidence; 3, little evidence; 4, no evidence; and 5, not sensitive/recessive). Two-sided Fisher’s exact test was performed to assess the enrichment of EGs vs. NEGs, and the P values were indicated. The odds ratios for enrichment of EGs compared with NEGs and the 95% confidence intervals of odds ratios are plotted in *Right*. OMIM, Online Mendelian Inheritance in Man (29). (*B* and *C*) Histograms and estimated density curves indicating the distribution of (*B*) the Haploinsufficiency Score (HIS) (32) and (*C*) the Genome-Wide Haploinsufficiency Score (GHIS) (33) across three gene sets, including EGs (red), NEGs (turquoise), and all protein-coding genes (56) (gray). EGs have significantly higher probability of exhibiting haploinsufficiency compared with NEGs (P value < 2.2×10^{-16} for both models; two-sided Wilcoxon rank sum test).

dosage-sensitive, we explored the possibility that a cumulative effect of pathogenic variants in multiple EGs may underlie the genetic basis of a complex disease with early postnatal onset, such as ASD.

To address a possible cumulative effect of variants in EGs in ASD in a larger cohort of 1,781 ASD quartet families (with 1,781 probands and 1,781 siblings) from the Simons Simplex Collection (34), we acquired de novo and rare inherited mutations from the exome sequencing data of these families (8, 26). We examined the individual mutational burden defined by the number of de novo loss-of-function (dnLoF), de novo nonsynonymous damaging (dnNSD), and inherited rare damaging (inhRD) mutations per individual (Fig. S1, *SI Materials and Methods*, and *Datasets S2–S4*). On average, an ASD proband carried 0.06 dnLoF, 0.21 dnNSD, and 10.74 inhRD mutations in EGs. The mutational burden in EGs was significantly elevated in ASD probands compared with unaffected siblings for the three classes of variants considered (P value = 4.75×10^{-7} for dnLoF, P value = 3.41×10^{-4} for dnNSD, and P value = 0.017 for inhRD; one-sided Wilcoxon signed ranked test) (Fig. 2*A* and *Table S1*). In contrast, no significant difference in mutational burden in NEGs was observed (P value = 0.10 for dnLoF, P value = 0.069 for dnNSD, and P value = 0.75 for inhRD) (*Table S1*). Interestingly, 10,823 genes that are currently not assigned as EG or NEG (i.e., phenotypically uncharacterized in mouse knockouts and human cell-based assays) have a moderately elevated burden of dnLoF but

Fig. 2. Assessment of the contribution of EGs to ASD risk. (A) Individual mutational burden analysis in 1,781 pairs of ASD probands and unaffected siblings (Table S1). The analyses were performed separately for 3,915 EGs (red) and 4,919 NEGs (turquoise). The individual mutational burden is defined by the number of dnLoF, dnNSD, and inhRD mutations per individual. Effect sizes were measured by Cohen's d , which is defined as the difference between both means divided by the SD of the paired differences. The estimated 95% confidence intervals of effect sizes were plotted (SI Materials and Methods). P values were obtained from one-sided Wilcoxon signed ranked test. * P value < 0.05. (B) ASD candidate genes categorized by SFARI genes scores (S, syndromic; 1, high confidence; 2, strong candidate; 3, suggestive evidence; 4, minimal evidence; 5, hypothesized; and 6, not supported) (37) and their essentiality status (EG in red, NEG in turquoise, and unknown in gray). ***The P value from two-sided Fisher's exact test (EG vs. NEG) is less than 0.001. (C) The distribution of TADA FDR q values of EGs and NEGs. The FDR q value of the TADA test evaluates ASD association based on combined evidence from de novo SNVs and small deletions, rare inherited variants, and variants (9). The observed negative $\log_{10}(q)$ values of 3,915 EGs (red) and 4,919 NEGs (turquoise) are compared with the expected counterparts under the null hypothesis. The dashed lines indicate the FDR thresholds (FDR = 0.1 in red and FDR = 0.5 in blue) for identification of ASD risk genes. The 95% confidence intervals of the expected negative $\log_{10}(q)$ values are shaded in gray.

not dnNSD and inhRD variants in ASD probands (P value = 0.0042) (Table S1). Notably, the effect sizes of EG burden in each variant type correspond to our understanding of the severity of the variant type; de novo mutations, which are expected to have a larger functional impact, also display the strongest difference between ASD probands and unaffected siblings (effect size = 0.117 for dnLoF; effect size = 0.079 for dnNSD; Cohen's d). In contrast, inherited mutations are expected to have a moderate functional impact, and a smaller difference is observed between probands and siblings (effect size = 0.042 for inhRD). Although we observed marginally increased burden of dnLoF and dnNSD mutations in EGs in female ($n = 325$) compared with male ($n = 2,043$) probands (Table S2), the analysis of families divided by gender of proband-sibling pairs (female-female, male-female, female-male, and male-male) showed that gender bias does not underlie the observed differences in mutational burden between probands and siblings (Table S3).

To evaluate the effect of rare damaging mutations in EGs on ASD-associated traits, we used the available quantitative phenotype data on social and cognitive impairments in ~2,500 ASD families from Simons Simplex Collection (8, 26) ([Dataset S2](#)). As a measure of sociability, we used the total raw score from the Social Responsiveness Scale (SRS) (35), and as cognitive measures, we used three different intelligence quotient (IQ) scores (full-scale IQ, verbal IQ, and nonverbal IQ). As previously reported (36), SRS scores were unrelated to IQ, especially in subjects with IQ higher than 50 ([Fig. S2](#)). In male probands, we observed that the mutational burden in EGs was positively correlated with the SRS total raw score (P value = 1.08×10^{-6} ; Poisson regression) (Table 1). The effect was not significant in NEGs ($P = 0.21$). In female probands, mutational burden in NEGs but not EGs was negatively correlated with SRS total raw score ($P = 0.085$ for EG and $P = 6.06e-06$ for NEG). In addition, we found that mutational burden in both EGs and NEGs had a significant effect (P value $< 2.2 \times 10^{-16}$) on verbal and nonverbal IQ scores and that the effect sizes of mutational burden in EGs and NEGs were comparable ([Table S4](#)). These results suggest that, in ASD probands, deleterious variants in EGs contribute to

decreased social skills in males, whereas deleterious variants in both EGs and NEGs lead to decreased IQ.

To initially explore the overlap between EGs and known ASD genes, we examined the essentiality status of ~500 ASD candidate genes from the Simons Foundation Autism Research Initiative (SFARI) AutDB database (updated December of 2015) (37) (Fig. 2B). Compared with NEGs, EGs were enriched among ASD candidates categorized as “syndromic” (category S: odds ratio = 3.95, P value = 0.0003; two-sided Fisher’s exact test), candidates with “high confidence” (category 1: odds ratio = 15.12, P value = 0.0004), and candidates with “suggestive evidence” (category 3: odds ratio = 2.14, P value = 0.0006). Trends of enrichment of EGs were also observed for “strong candidates” (category 2: odds ratio = 1.62, P value = 0.21). We did not observe enrichment of EGs among candidate genes with less supportive evidence (categories 4–6).

To further address whether EGs contribute to ASD risk, we compared the strength of ASD association signals between EGs and NEGs in data from a recent comprehensive analysis of ASD genomic architecture (9), where the transmission and de novo association (TADA) test (38) was used to evaluate ASD

Table 1. Relationship between individual mutational burden and SRS in ASD probands

Group and gene set	Estimate	Standard error	<i>P</i> value
2,031 Male probands			
EG (3,915 genes)	0.001860	0.000381	$1.08 \times 10^{-6*}$
NEG (4,919 genes)	0.000407	0.000324	0.209
317 Female probands			
EG (3,915 genes)	-0.001511	0.000877	0.085
NEG (4,919 genes)	-0.003084	0.000682	6.04×10^{-6}

Coefficients for Poisson regression are shown, which model the relationship between SRS total raw score and individual burden of all rare damaging mutations (including dnLOF, dnNSD, and inhRD mutations).

*The *P* value with statistical significance with positive estimated effects (*P* value < 0.05; estimate > 0).

Moreover, the analysis of EGs in the developing brain identified clusters of coexpressed EGs implicated in ASD, including 29 EGs functionally related to previously identified ASD risk genes.

We find that ASD individuals have a higher burden of mutations in EGs compared with their unaffected siblings. It is notable but not surprising that this effect is particularly pronounced when considering *de novo* mutations, because this class of mutations is only subject to selection pressure after originating in the individual and has exhibited some of the most prominent associations with the risk of ASD (8, 43–45). Similarly, a moderately increased burden of dnLoF variants in ASD probands was detected with a group of 10,823 phenotypically uncharacterized genes. Based on current estimates, one-fifth of these uncharacterized genes (~2,000) are expected to be EGs, which may explain the higher mutational burden of dnLoF variants in ASD probands. Recent studies have begun to show that additional genetic factors, such as rare and common inherited variations, also contribute to ASD (26, 46). Our result supports this finding, showing that inherited, rare, damaging mutations in EGs also have a significant effect on ASD risk. Furthermore, we show an EG-specific effect on social responsiveness, a measure of the social aspects of ASD. In contrast, mutational burden in both EGs and NEGs has an effect on IQ measures. Complex social behaviors result from a range of different cognitive processes; however, in ASD subjects, there is a striking dissociation in the level of impairment in social interaction or communication and general cognitive abilities (as measured by IQ) (36) (Fig. S2). Moreover, studies in model organisms clearly show a fetal origin for social behavior deficits (47). Our results are in line with these findings and suggest that, although a higher mutational burden over all genes may have consequences on IQ, mutational burden in a set of genes with a role at critical early developmental stages influences the development of social behavior. Moreover, our findings are also further supported by the recent report that genomic regions that are under accelerated evolution have essential functions in the human brain development and when mutated, may cause increased risk for autism (48). Therefore, understanding the regulatory landscape of dosage-sensitive EGs expressed at critical stages of brain development may reveal risk alleles for many neurodevelopmental and psychiatric disorders.

The analysis of the overlapping set of Simons Simplex Collection ASD families by several groups using complementary approaches led to the identification of around 100 ASD risk genes and the finding of a depletion of damaging mutations in ASD risk genes (12, 27, 28). We show that a significant number of reported ASD risk genes are essential for survival and fitness and therefore, have a distinctive mutational spectrum, providing a biological foundation for this intolerance to damaging mutations. Of the spectrum of existing alleles, homozygosity or compound heterozygosity for loss-of-function alleles will never be observed. Also, because of synthetic lethality, some combinations of mutations in EGs are eliminated. Therefore, individuals will have only a subset of “milder” coding or regulatory alleles. The current list of candidate genes consists of 100 (high-confidence ASD genes) to 400 genes (potential ASD genes) (9). It is striking that our study provides strong statistical evidence for the aggregate effect across 3,915 EGs impacting risk for this neurodevelopmental disorder. A recent SNP-based heritability study reported the extreme polygenicity of schizophrenia, with 70% of 1-Mb genomic regions harboring schizophrenia risk alleles (49). Assuming a similar genetic architecture in ASD and schizophrenia, genomic maps of EGs with “surviving” deleterious and regulatory variants in ASD probands represent a complementary approach for the analysis of combinations of culprit genes or alleles.

Because of the fundamental functional role of EGs in an organism, genetic variants in these genes are likely to contribute to many traits and diseases as reflected by the previous finding that EGs are enriched for human disease genes (11, 13, 14). Our

study is focused on a specific neurodevelopmental disorder—ASD—because it has been suggested that ASD has its roots in abnormalities in prenatal brain development (50–52). Specifically, our analysis of the temporal expression patterns of coexpressed gene modules in the developing brain shows that genes in three EG-enriched coexpression modules implicated in ASD are expressed at a high level at the earliest stages of brain development, as early as 8 weeks after conception. In contrast, at later stages of brain development, the expression levels of genes in these EG-enriched modules decrease, whereas the expression levels of genes in NEG-enriched modules increase. This finding suggests that EGs have a distinctive influence at some of the earliest brain developmental stages as previously reported for constrained genes (53) and genes in functional networks perturbed in ASD (54). However, it is not clear whether the contribution of EGs is specific to ASD or widespread across disorders with various underlying mechanisms. A comparison of the burden of deleterious variants in EGs across other complex disorders, including those with a later onset, is warranted.

Each individual can carry a number of deleterious mutations, each of which can have a small effect. Because brain function may be particularly sensitive to mutation accumulation, identifying a specific set of genes in which mutations have a behavioral effect will assist us in understanding how mutation accumulation within an individual can result in a phenotype, such as ASD. Hallmarks of ASD are phenotypic heterogeneity, frequent comorbidities, and that no specific brain region or cell type is uniquely implicated (5), further supporting the role of genes with a global effect on embryonic and fetal development. Here, we provide evidence that genes that are essential for survival and fitness also contribute to ASD risk and lead to the disruption of normal social behavior.

Materials and Methods

Identification of EGs. Mouse Phenotype (MP) terms for the annotation of EGs are listed in Table S8. More details on identification of the catalog of EGs are in *SI Materials and Methods*.

Analysis of Haploinsufficiency of EGs. Details on collection of genes sets for the analysis of haploinsufficiency of EGs are in *SI Materials and Methods*.

Burden Analysis of Mutations in EGs in ASD Families. Details on collection of genetic and phenotypic data of ASD families and variant filtering process are in *SI Materials and Methods*.

Comparison Between Observed and Expected TADA FDR q Values. To compare the strength of association signals to ASD between EGs and NEGs, FDR q values for the TADA test of 18,665 genes were obtained from the work by Sanders et al. (9). For each gene set of interest (i.e., 3,915 EGs or 4,919 NEGs), the null distribution of TADA FDR q values was generated by randomly resampling with replacement. Within one iteration of the resampling procedure, the TADA FDR q value of a random gene from the tested 18,665 genes was obtained for each gene in the gene set of interest. The resampled TADA FDR q values were then ranked from low to high. The resampling procedure was repeated for 100,000 iterations. For each observed TADA FDR q value ranked from low to high, the median of 100,000 resampled q values with the same rank was considered the expected TADA FDR q value. The 2.5th and 97.5th percentiles of 100,000 resampled q values were considered the estimated 95% confidence intervals of each expected TADA FDR q value. The observed FDR q values were then compared with the expected FDR q values.

Construction of Coexpression Modules and Coexpression Network in Brain. Details on construction of coexpression modules and coexpression network in the developing human brain are in *SI Materials and Methods*.

Pathway Enrichment Analysis. We performed pathway enrichment analysis in the Reactome database (42) using Enrichr (55) for three EG-enriched modules (M01, M02, and M16) that were also enriched for potential ASD genes (Table S6). The enriched pathways were ranked by P values with Benjamini–Hochberg adjustment (FDR q values) from the Fisher’s exact test.

Code Availability. Details on availability of code used to generate reported results are in Table S9.

ACKNOWLEDGMENTS. We thank Steve Murray and the International Mouse Phenotyping Consortium (IMPC) for help with generation of gene lists, and Benjamin Georgi, Benjamin Voight, Hakon Hakonarson, Steve Brown, Judith Miller, Edward Brodtkin, and Lu Chen for discussions. X.J. was supported by a

fellowship from Biomedical Graduate Studies at the University of Pennsylvania. This work was supported by the Pennsylvania Commonwealth Grant and NIH Grants R01MH101822 (to C.D.B.) and R01MH093415 (to M.B. and Steven M. Paul; multiple principal investigators).

- State MW, Levitt P (2011) The conundrums of understanding genetic risks for autism spectrum disorders. *Nat Neurosci* 14(12):1499–1506.
- Huguet G, Ey E, Bourgeron T (2013) The genetic landscapes of autism spectrum disorders. *Annu Rev Genomics Hum Genet* 14:191–213.
- Willsey AJ, State MW (2015) Autism spectrum disorders: From genes to neurobiology. *Curr Opin Neurobiol* 30:92–99.
- De Rubeis S, Buxbaum JD (2015) Recent advances in the genetics of autism spectrum disorder. *Curr Neurol Neurosci Rep* 15(6):36.
- de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH (2016) Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* 22(4):345–361.
- Geschwind DH, Levitt P (2007) Autism spectrum disorders: Developmental disconnection syndromes. *Curr Opin Neurobiol* 17(1):103–111.
- De Rubeis S, et al.; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515(7526):209–215.
- Iossifov I, et al. (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515(7526):216–221.
- Sanders SJ, et al.; Autism Sequencing Consortium (2015) Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* 87(6):1215–1233.
- Zhang M, Zhu C, Jacomy A, Lu LJ, Jegga AG (2011) The orphan disease networks. *Am J Hum Genet* 88(6):755–766.
- Georgi B, Voight BF, Bućan M (2013) From mouse to human: Evolutionary genomics analysis of human orthologs of essential genes. *PLoS Genet* 9(5):e1003484.
- Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB (2013) Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* 9(8):e1003709.
- Dickerson JE, Zhu A, Robertson DL, Hentges KE (2011) Defining the role of essential genes in human disease. *PLoS One* 6(11):e27368.
- Dickinson ME, et al.; International Mouse Phenotyping Consortium; Jackson Laboratory; Infrastructure Nationale PHENOMIN, Institut Clinique de la Souris (ICS); Charles River Laboratories; MRC Harwell; Toronto Centre for Phenogenomics; Wellcome Trust Sanger Institute; RIKEN BioResource Center (2016) High-throughput discovery of novel developmental phenotypes. *Nature* 537(7621):508–514.
- Deutschbauer AM, et al. (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169(4):1915–1925.
- Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* 93(19):10268–10273.
- Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* 1(2):127–136.
- Hwang YC, et al. (2009) Predicting essential genes based on network and sequence analysis. *Mol Biosyst* 5(12):1672–1678.
- Chakravarti A, Turner TN (2016) Revealing rate-limiting steps in complex disease biology: The crucial importance of studying rare, extreme-phenotype families. *BioEssays* 38(6):578–586.
- Blomen VA, et al. (2015) Gene essentiality and synthetic lethality in haploid human cells. *Science* 350(6264):1092–1096.
- Wang T, et al. (2015) Identification and characterization of essential genes in the human genome. *Science* 350(6264):1096–1101.
- Hart T, et al. (2015) High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163(6):1515–1526.
- Eppig JT, et al.; Mouse Genome Database Group (2005) The Mouse Genome Database (MGD): From genes to mice—a community resource for mouse biology. *Nucleic Acids Res* 33(Database issue):D471–D475.
- Koscielny G, et al. (2014) The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res* 42(Database issue):D802–D809.
- White JK, et al.; Sanger Institute Mouse Genetics Project (2013) Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* 154(2):452–464.
- Krumm N, et al. (2015) Excess of rare, inherited truncating mutations in autism. *Nat Genet* 47(6):582–588.
- Samocha KE, et al. (2014) A framework for the interpretation of de novo mutation in human disease. *Nat Genet* 46(9):944–950.
- Iossifov I, et al. (2015) Low load for disruptive mutations in autism genes and their biased transmission. *Proc Natl Acad Sci USA* 112(41):E5600–E5607.
- Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33(Database issue):D514–D517.
- Rehm HL, et al.; ClinGen (2015) ClinGen—the Clinical Genome Resource. *N Engl J Med* 372(23):2235–2242.
- Dang VT, Kassahn KS, Marcos AE, Ragan MA (2008) Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur J Hum Genet* 16(11):1350–1357.
- Huang N, Lee I, Marcotte EM, Hurles ME (2010) Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 6(10):e1001154.
- Steinberg J, Honti F, Meader S, Webber C (2015) Haploinsufficiency predictions without study bias. *Nucleic Acids Res* 43(15):e101.
- Fischbach GD, Lord C (2010) The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron* 68(2):192–195.
- Constantino J, Gruber C (2005) *The Social Responsiveness Scale Manual* (Western Psychological Services, Los Angeles).
- Constantino JN, et al. (2003) Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the autism diagnostic interview-revised. *J Autism Dev Disord* 33(4):427–433.
- Abrahams BS, et al. (2013) SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* 4(1):36.
- He X, et al. (2013) Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet* 9(8):e1003671.
- Buxbaum JD, et al.; Autism Sequencing Consortium (2012) The autism sequencing consortium: Large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron* 76(6):1052–1056.
- BrainSpan (2011) *BrainSpan: Atlas of the Developing Human Brain*. Available at brainspan.org. Accessed October 4, 2013.
- Croft D, et al. (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res* 42(Database issue):D472–D477.
- Fabregat A, et al. (2016) The Reactome pathway Knowledgebase. *Nucleic Acids Res* 44(D1):D481–D487.
- Sanders SJ, et al. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485(7397):237–241.
- O’Roak BJ, et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485(7397):246–250.
- Iossifov I, et al. (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron* 74(2):285–299.
- Gaugler T, et al. (2014) Most genetic risk for autism resides with common variation. *Nat Genet* 46(8):881–885.
- Belinson H, et al. (2016) Prenatal β -catenin/Brn2/Tbr2 transcriptional cascade regulates adult social and stereotypic behaviors. *Mol Psychiatry* 21(10):1417–1433.
- Doan RN, et al. (2016) Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* 167(2):341–354.e12.
- Loh PR, et al.; Schizophrenia Working Group of Psychiatric Genomics Consortium (2015) Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* 47(12):1385–1392.
- Willsey AJ, et al. (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155(5):997–1007.
- Parikhshak NN, et al. (2013) Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155(5):1008–1021.
- Stoner R, et al. (2014) Patches of disorganization in the neocortex of children with autism. *N Engl J Med* 370(13):1209–1219.
- Choi J, Shoshitani P, Samocha KE, Daly MJ, Cotsapas C (2016) Network analysis of genome-wide selective constraint reveals a gene network active in early fetal brain intolerant of mutation. *PLoS Genet* 12(6):e1006121.
- Chang J, Gilman SR, Chiang AH, Sanders SJ, Vitkup D (2015) Genotype to phenotype relationships in autism spectrum disorders. *Nat Neurosci* 18(2):191–198.
- Chen EY, et al. (2013) Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14:128.
- Flicek P, et al. (2014) Ensembl 2014. *Nucleic Acids Res* 42(Database issue):D749–D755.
- Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24(5):659–685.
- Lord C, et al. (2000) The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 30(3):205–223.
- Kircher M, et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46(3):310–315.
- McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.
- Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907*.
- NHLBI Exome Sequencing Project (ESP) Exome Variant Server. Available at evs.gs.washington.edu/EVS/. Accessed November 11, 2015.
- Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.
- Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q (2008) GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 9(Suppl 1):S4.
- Shannon P, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504.