AUTISM

# A method to delineate de novo missense variants across pathways prioritizes genes linked to autism

**Amanda Koire[1,2,3,4], Panagiotis Katsonis[5], Young Won Kim[6], Christie Buchovecky[5,7], Stephen J. Wilson[8], Olivier Lichtarge[1,5,8]***



**Genotype-phenotype relationships shape health and population fitness but remain difficult to predict and interpret. Here, we apply an evolutionary action method to de novo missense variants in whole-exome sequences of individuals with autism spectrum disorder (ASD) to unravel genes and pathways connected to ASD. Evolutionary action predicts the impact of missense variants on protein function by measuring the fitness effect based on phylogenetic distances and substitution odds in homologous gene sequences. By examining de novo missense variants in 2384 individuals with ASD (probands) compared to matched siblings without ASD, we found missense variants in 398 genes representing 23 pathways that were biased toward higher evolutionary action scores than expected by random chance; these pathways were involved in axonogenesis, synaptic transmission, and neurodevelopment. The predicted fitness impact of de novo and inherited missense variants in candidate genes correlated with the IQ of individuals with ASD, even for new gene candidates. Taking an evolutionary action method, we detected those missense variants most likely to contribute to ASD pathogenesis and elucidated their phenotypic impact. This approach could be applied to integrate missense variants across a patient cohort to identify genes contributing to a shared phenotype in other complex diseases.**

## INTRODUCTION

The relationship between genotype and phenotype can be difficult to predict and interpret. This presents particular challenges when interpreting mutations of complex diseases like autism spectrum disorder (ASD), which is both phenotypically and genetically heterogeneous. Some predictions place the number of genes involved in ASD pathogenesis in the hundreds (*1*, *2*) to thousands (*3*–*5*), and the highly multigenic nature of the disorder means that few causative genes can be identified through an excess of mutations. In the absence of any single gene responsible for the majority of ASD cases, the most commonly mutated genes each account for about 2% of the cases (*6*, *7*). To explain additional cases, it is critical to expand analysis to interpret the collectively large number of variants in rarely mutated genes.

Although ASD has many implicated contributing factors, including environment (*8*, *9*), common polymorphisms (*10*), and inherited rare variants (*11*), de novo variants in particular are suspected to be enriched as a class for causative mutations because they have not been subjected to generations of evolutionary selection. Analysis of de novo mutations in ASD has largely focused on copy number variants (CNVs) (*12*–*14*), single-nucleotide variants (SNVs) resulting in an obvious loss of function (LOF) (*15*–*17*), and genes with a detectably elevated mutation rate (*18*, *19*). Far less attention has been paid to the role of missense variants on genes with low mutation rates, with such analyses limited to genes already implicated in ASD or in ASD-related pathways (*20*). The overall role of missense variants in driving phenotype severity has also remained unclear. Whereas strong links between mutation and lower patient intelligence quotient (IQ) have been detected for LOF de novo variants (*16*, *21*)—defined by the combined class of nonsense, frameshift, and splice-site mutations—studies have not yet been able to link missense mutations to the same patient presentations on a large scale and without prior knowledge of ASD-associated genes (*16*). However, individuals with ASD are more likely to carry a de novo missense variant than either a de novo LOF or a de novo CNV (*16*), so the prioritization and interpretation of these variants are paramount, especially if they are revealed to be an important and understudied source of driver events.

Here, we prioritized rarely mutated, potentially causative ASD genes by their de novo missense variants alone. Without making any a priori assumptions of which genes or pathways drive ASD, we tested whether groups of functionally related genes were biased toward high-impact variants. Interpreting the variant effects on protein function is challenging (*22*, *23*) and subject to disagreement between different methods of variant impact prediction (*24*). To estimate the impact of each variant, we first used the evolutionary action equation (*25*), a state-of-the-art prediction method that was consistently assessed to be one of the best methods in the objective, blind contests of the Critical Assessment of Genome Interpretation community (*26*, *27*). Briefly, the evolutionary action equation models the genotype-to-phenotype relationship to first-order approximation with an equation that equates the functional impact of a mutation on fitness to the product of the functional importance of the mutated residue and the amino acid dissimilarity of the substitution. It is well suited for considering variants from groups of genes in aggregate because its scoring system for variants has built-in normalization for gene selection pressure. To quantify mutational bias in pathways, we considered the evolutionary action scores over the de novo missense mutations of functionally related genes across all individuals with ASD in our cohort. This integrative approach

[1]Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX, USA. [2]Medical Scientist Training Program, Baylor College of Medicine, Houston, TX, USA. [3]Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, USA. [4]Harvard Medical School, Boston, MA, USA. [5]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. [6]Program in Integrative Molecular and Biomedical Sciences, Baylor College of Medicine, Houston, TX, USA. [7]Division of Carrier Screening and Prenatal Testing, SEMA4, Stamford, CT, USA. [8]Department of Biochemistry, Baylor College of Medicine, Houston, TX, USA.
*Corresponding author. Email: lichtarge@bcm.edu

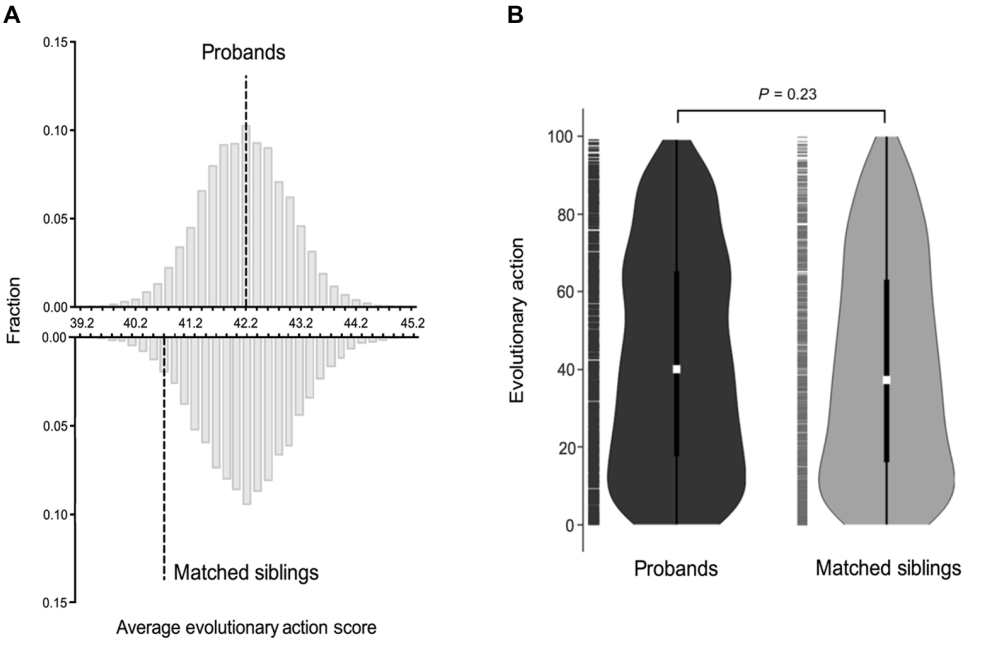


**Fig. 1. The impact of missense variants in probands and matched siblings.** (**A**) The impact of de novo missense variants in probands and matched siblings without ASD is compared to random nucleotide changes. The average evolutionary action score of all de novo missense variants is superimposed upon the distribution of averages produced by 10,000 simulations of 1418 randomly selected coding missense variants for probands (top plot) and 976 randomly selected coding missense variants for matched siblings (bottom plot). (**B**) The impact of missense variants is compared between patients with ASD and matched siblings. The distribution of evolutionary action scores for missense variants in patients with ASD (black) compared to matched siblings (gray) is represented by violin plots with the center dot indicating the median and the center bar indicating the 25th to 75th percentiles of the data. The plots were compared statistically using a two-sample Kolmogorov-Smirnoff test.

detected nonrandom mutational patterns indicative of proband-specific selection of missense variants associated with axonogenesis, synaptic transmission, and other neurodevelopmental pathways. In the genes prioritized by this approach, both missense de novo variants and rare inherited missense variants correlated with patient IQ, demonstrating a direct relationship to patient phenotype. We concluded that evolutionary action integration in pathways detected some missense variants that could contribute to ASD pathogenesis, with implications for prioritizing genes and variants in ASD.

## RESULTS

### Characterization of the de novo missense variant class in ASD probands

We first assessed whether de novo missense variants in 2384 individuals with ASD have, as a class, a distinct and more impactful variant profile compared to random expectation or those in matched siblings without ASD. Across 2384 individuals with ASD, we identified 1418 missense variants affecting 1269 unique genes and annotated the impact of the variants using evolutionary action scores (table S1). Close to half of the probands (43.9%) carried a de novo missense variant, and the observed de novo missense mutation prevalence was 0.59 per proband, similar to the rates reported by Neale *et al.* (*28*) (0.58 per proband) and Sanders *et al.* (*18*) (0.55 per proband). Across 1792 unaffected siblings, we identified 976 missense variants affecting 911 unique genes (table S2). Compared to their matched unaffected siblings, probands had more de novo missense variants, consistent with previous studies (*16*), and even after filtering to exclude genes with more than three variants, this difference remained significant (difference of 6.25%, $P = 0.016$). The average predicted impact of all missense variants in probands was not significantly different from what would be expected by random mutagenesis ($z$ score, +0.13; Fig. 1A), indicating no evidence of selection. This finding is consistent with the very low fraction of expected ASD driver variants, as indicated by the fraction of excessive variants in probands compared to their healthy siblings and also suggested in previous studies (*17*). For the healthy siblings, this difference was significant ($P = 0.03$), indicating selection against pathogenic variants, which may be due to the lack of disease driver variants in these individuals. Comparing evolutionary action distributions of de novo missense variants in probands with their healthy matched siblings showed a small difference that did not reach significance ($P = 0.23$; Fig. 1B), suggesting a broad spectrum of fitness effects for the variants that drive ASD, which agrees with previous conclusions that these drivers involve mild effects (*17*). These results suggest that the landscape of de novo missense variants over all individual genes in patients with ASD is similar to that of the matched siblings and dominated by mutations with relatively mild impact on protein fitness.

However, network analysis of the 1269 genes in which de novo missense variants occurred in the patient group exposed an underlying nonrandom signal within this class of variants. Affected genes had significantly more protein-protein interactions in the STRING (Search Tool for Retrieval of Interacting Genes/Proteins) database (*29*) than would be expected by chance ($P = 7.3 \times 10^{-12}$), and hundreds of Gene Ontology (GO) biological processes were significantly enriched. For the matched siblings, the protein-protein interaction enrichment was barely significant ($P = 0.045$). However, the vast majority of genes (1195 of 1269 genes) under consideration exhibited these network features (fig. S1A). A gene-centric interaction or enrichment approach is fundamentally limited in its ability to isolate the detected signal or stratify candidate genes; of the 1269 genes affected by de novo missense variants, 86% interacted with another in the set compared to 79% expected by chance, and there was no way to identify which genes were the excess driving the significance (fig. S1B). For these reasons, a complementary approach to evaluating events within the missense variant class was necessary to extricate a causative subset of genes and variants.

### Prioritization of de novo missense variants using variant-centric pathway analysis

To pinpoint the source of the interaction signal within the de novo missense class and meaningfully prioritize a subset of the missense de novo variants and their associated genes, we therefore pursued a variant-centric approach in which we examined patterns of variant impact across functionally related groups of genes. Genes were grouped by ontology using the software tool GO2MSIG (*30*), yielding 368 pathways encompassing 15,310 total genes (table S3); variant impact was annotated with the evolutionary action method, producing impact scores on a continuous scale between 0 (minimum predicted impact) and 100 (maximum predicted impact). For the 1792 individuals with ASD with matched siblings without ASD, 1037 de novo missense variants across 960 genes in probands and 976 de novo missense variants across 911 genes in the healthy siblings were considered (after we focused on genes with three or fewer missense mutations to avoid false discovery of pathways due to a single important gene). From these, 860 de novo missense variants across 796 genes in probands and 776 de novo missense variants across 725 genes in healthy siblings were assigned to the pathways. For each pathway, the evolutionary action score distribution of the de novo variants within the pathway was compared to the evolutionary action distribution of all other de novo variants in those with ASD. Pathways that displayed a bias toward high-impact variants and remained significant after multiple hypothesis testing were considered to be of interest, and genes that were affected by de novo variants and present in a significant pathway were considered prioritized genes. This approach revealed 23 significant pathways in the probands, with functions that demonstrated clear ties to nervous system development, including axonogenesis and synaptic transmission (Fig. 2A and table S4). For example, in the synaptic transmission pathway, 49 mutations contributed from 43 individual genes produced a variant impact distribution statistically ($P = 6.95 \times 10^{-4}$, $q = 0.037$) and visibly biased to higher evolutionary action scores (Fig. 2B). As a control, the same process was repeated using all 976 de novo missense variants from the matched siblings; no pathways exhibited significant bias toward high functional impact (Fig. 2C). For subsequent analysis, genes falling into pathways with significant evolutionary action bias toward high-impact mutations were grouped together into a single set of 398 prioritized genes, and all other 562 genes with de novo missense variants were considered deprioritized (table S5). An independent cohort of 5134 individuals with ASD [DB6 release of MSSNG (*31*)] had 933 de novo missense variant calls in any of the 23 prioritized pathways, and these variants were biased to higher evolutionary action impact compared to the 1389 de novo missense variants in the rest of the genes ($P < 0.0001$). This result validated our conclusion that individuals with ASD have de novo missense variants that affected the function of the prioritized gene pathways.

### Evolutionary action burden of de novo missense variants in prioritized genes correlates with phenotypic severity

To determine whether prioritizing genes according to their evolutionary action distributions in pathways provides a meaningful stratification between causative and noncausative genes, the variants in the prioritized genes were tested for their relationship to patient phenotypic presentation, defined here by IQ. This analysis was performed only for the probands because of the lack of IQ information for the unaffected siblings. The capacity of evolutionary action scores alone to predict patient phenotypic presentation within this prioritized gene set was tested by comparing the clinical presentation of male patients included in the initial analysis who were affected



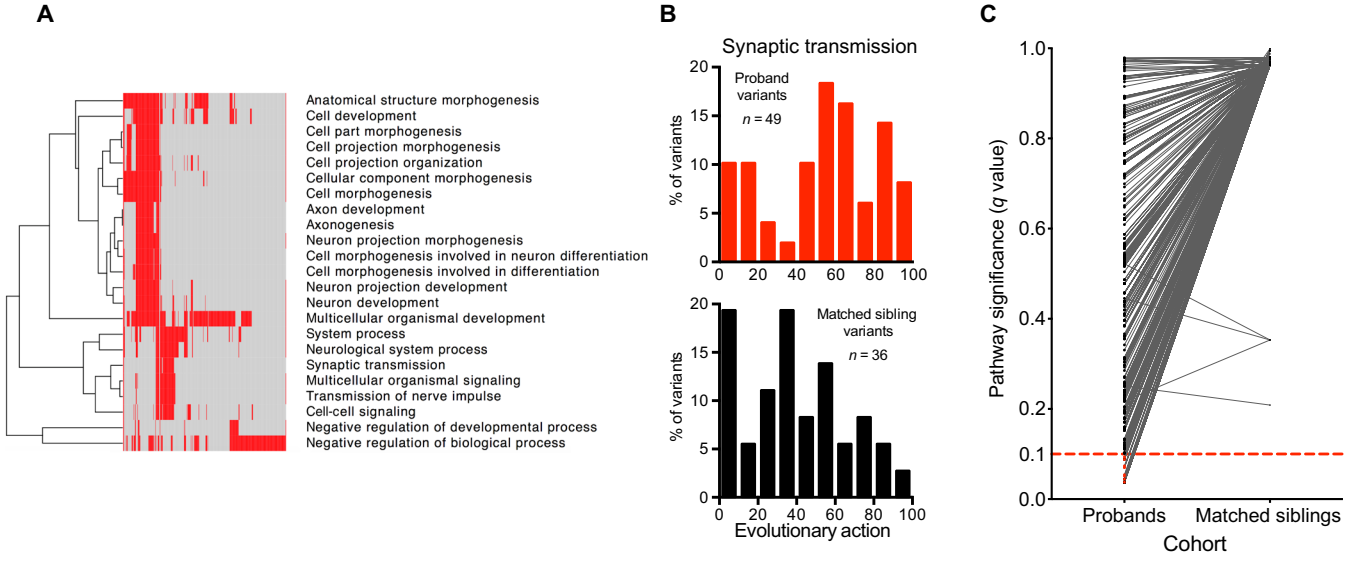


**Fig. 2. Prioritization of de novo missense variants using variant impact on pathways.** (**A**) Hierarchical clustering of the 23 prioritized pathways is shown. For the 398 genes with missense variants that were associated with at least one relevant pathway, a matrix was created to denote whether the gene was (red) or was not (gray) a component of the pathway. Pathways were then grouped according to their patterns of affected genes via hierarchical clustering performed by GENE-E. (**B**) Evolutionary action score distribution for the synaptic transmission pathway is shown. Evolutionary action scores for the proband variants (red) and matched sibling variants (black) in this pathway were binned in deciles and represented as histograms. (**C**) Significance of all tested pathways in cohorts of patients with ASD and their matched siblings is presented. Each point represents 1 of the 368 tested pathways and is connected with a line to the same pathway in the matched cohort. The $q = 0.1$ significance threshold after false discovery rate (FDR) correction is represented as a dashed red line.

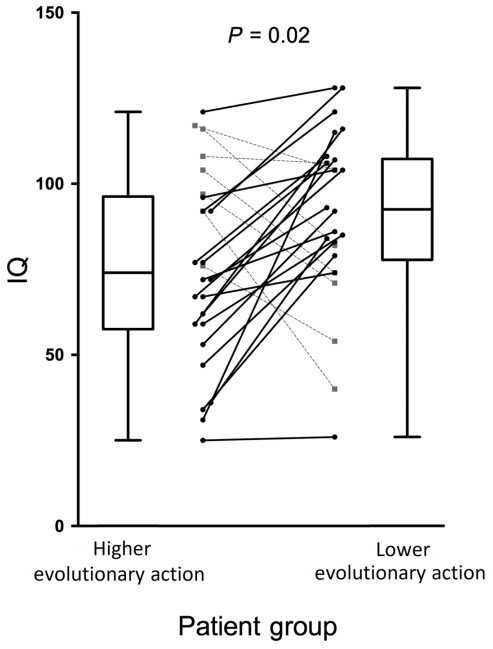


**Fig. 3. Predicted variant impact on IQ of ASD patient pairs with different de novo missense variants in the same candidate gene.** Pairs of patients with ASD affected by different de novo missense variants in the same prioritized gene were identified across the 398 prioritized genes ($n = 28$). Within each pair, the patient with the higher variant evolutionary action score was determined. Full-scale IQ scores were compared between the higher and lower evolutionary action groups using a paired *t* test. Correctly prioritized pairs are shown linked by a solid black line; incorrectly prioritized pairs are shown linked by a dashed gray line.

by different de novo missense variants in the same candidate gene. Although female probands with de novo missense mutations in prioritized genes contributed a minority of the data, they were highly disproportionately represented at low IQ and were analyzed separately from male patients to prevent confounding based on gender. When more than one phenotyped patient had a de novo missense variant in a given prioritized gene, the higher evolutionary action variant within the gene correctly predicted the patient with the lower IQ in 71.4% of paired comparisons ($n = 28$). Across all such cases, patients harboring the higher evolutionary action variant demonstrated significantly lower IQ overall, corresponding to a 15.2-point drop in IQ on average between the two groups ($P = 0.023$, paired *t* test) (Fig. 3).

To further explore the relationship between these variants and patient presentation, all male patients with ASD were divided into three groups corresponding to phenotypic severity: high IQ of 100 or greater (i.e., greater than or equal to population average), low IQ of less than 70 (i.e., more than 2 SD below population average and consistent with a diagnosis of intellectual disability), and intermediate IQ. Prioritized genes were grouped together into a single set of candidate causative ASD genes, and the evolutionary action score burden (sum of evolutionary action scores) of mutations in these genes was calculated for each patient and considered across the three groups. Significant differences in total variant impact were found between the three IQ groups, with the lowest-IQ patient group having the highest impact mutations in the prioritized genes ($P = 0.048$; Kruskal-Wallis test) (fig. S2A). This relationship between IQ and mutation evolutionary action scores was not seen when applied to all genes affected by de novo mutations ($P = 0.89$) or to genes that were not prioritized by the method ($P = 0.58$) (fig. S2A). This correlation can be explained by comparing the distributions of evolutionary action scores of prioritized genes between the three IQ groups: Patients with the lowest IQ had more variants with high scores than expected by chance, in contrast to patients with the highest IQ, who had fewer variants with high scores than expected by chance (fig. S2B). These data showed that the impact of the variant on the protein (as estimated by evolutionary action score) correlated with patient phenotype (IQ).

We next investigated the impact of the protein itself on human health, estimated here using RVIS (Residual Variation Intolerance Score) calculations of genic tolerance to mutation (*32*). Although genic tolerance to mutation was not on its own a significant predictor of phenotypic severity (Fig. S3A and Fig. S3B), weighting the evolutionary action impact score to account for differences in genic tolerance to mutation (weighted evolutionary action) further improved the ability of the evolutionary action score burden in prioritized genes to predict patient phenotype when binned ($P = 0.0028$; Fig. 4A and Fig. 4B), and this relationship also became significant when unbinned ($R = -0.14$, $P = 0.013$, linear regression) (tables S6 and S7). In addition, this correlation was stronger when RVIS was used to measure genic tolerance to mutation compared to pLI (probability of being LOF intolerant) or the Missense Constraint Metric (table S8); the correlation was significant also when verbal or nonverbal IQ was defined as the primary outcome (table S9). The correlation was generally reproducible with other variant impact prediction methods, such as Polyphen2 (Polymorphism Phenotyping v2), CADD (Combined Annotation Dependent Depletion), SIFT (Sorting Intolerant From Tolerant), MPC (Missense badness, PolyPhen-2, and Constraint), and BLOSUM62 (BLOcks SUbstitution Matrix using sequences with less than 62% similarity), but it was stronger and more robust to changes in the analysis when evolutionary action was used instead (table S10). The correlation between evolutionary action score burden and IQ also could not be explained by confounding due to the presence of de novo nonsense variants or CNVs, which are known to affect IQ. There was no significant correlation between de novo CNV deletion size and evolutionary action score burden (Pearson $R = -0.0001$, $P = 0.98$), and patients with a concurrent nonsense variant did not have a higher evolutionary action burden in prioritized genes ($P = 0.97$). No significant relationship between patient IQ and evolutionary action score burden was found when the relevant gene set was instead considered to be all genes affected by de novo mutations ($P = 0.21$), genes that were not prioritized by the method ($P = 0.74$) (Fig. 4A and Fig. 4B), or genes belonging to independent gene sets of interest a priori, such as those enriched for expression in the brain (*33*), proposed by orthogonal methods (*3*, *34*–*36*), or connected to other candidates in a protein interaction network (table S11). Whereas the correlation strength between missense variant burden and IQ is more modest than comparable continuous correlations that have been published connecting CNV deletion length to IQ (*37*), this is likely due to sample size and subsetting; the correlation that we detect is at least as strong as the established connection between CNVs and IQ after restriction to the same patients used in our analysis ($R = -0.09$, $P = 0.03$).

Furthermore, whereas the evolutionary action score burden accounted for cases in which more than one variant of interest was detected in a patient's exome, the results could not be explained by an uneven distribution of patients affected by multiple de novo

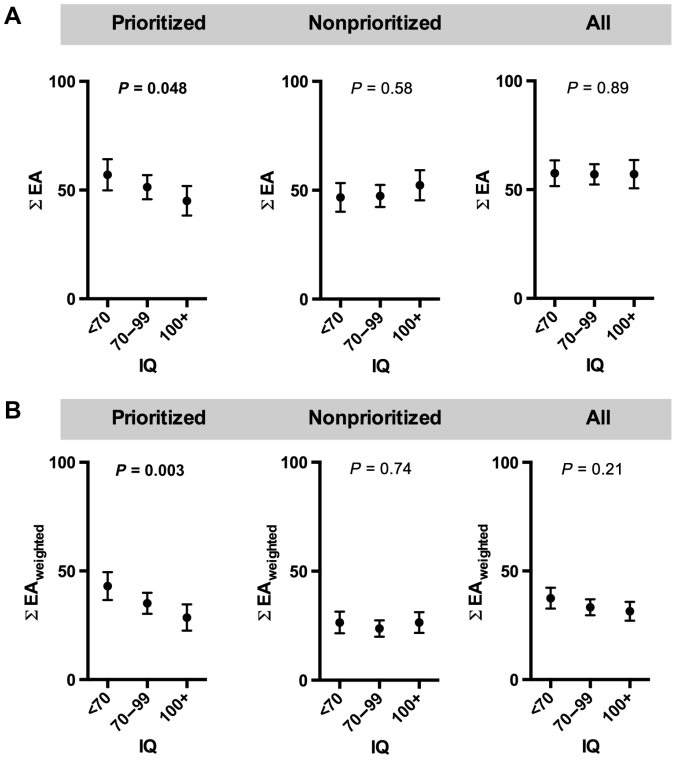


**Fig. 4. De novo evolutionary action score burden and ASD patient IQ for prioritized and deprioritized gene groups.** Prioritized genes, deprioritized genes, and all genes with de novo missense variants were assessed for their relationship to the IQ of patients with ASD. For each male patient, the summed evolutionary action burden of de novo missense variants was calculated for each category with evolutionary action burden (A) unweighted (ΣEA), and (B) weighted for genic intolerance to mutation ($\Sigma EA_{weighted}$). The patients were then split into three groups by their full-scale IQ scores, and the scores were compared using Kruskal-Wallis tests. Error bars reflect the 95% CI of the mean.

variants in prioritized genes ($P$ = 0.51; chi-square test) (fig. S4A), and the genotype-phenotype relationship remained significant ($P$ = 0.014 for unweighted evolutionary action burden and P = 0.003 for evolutionary action burden weighted for genic tolerance) when considering only patients affected by a single variant of interest (Fig. S4B and Fig. S4C). Female patients were assessed separately, and although their variant impact profile across prioritized genes was equally biased to high action (fig. S5A), the genotype-phenotype analysis was underpowered to detect a relationship of the magnitude present in male patients (fig. S5B), and the correlation between IQ and evolutionary action burden was not significant ($P$ = 0.40, linear regression).

## Prioritized gene set demonstrates enrichment for genes linked to ASD

To determine whether prioritization using evolutionary action pathway distributions captured established knowledge, we next compared our prioritized gene set to the 2017 version of the genes for ASD manually curated by the Simons Foundation Autism Research Initiative (SFARI). We considered SFARI categories 1 to 3 (high confidence, strong candidate, and suggestive evidence) to be appropriate for comparison. We then quantified the overlap of our gene lists to the SFARI list and found that the prioritized genes were highly enriched for genes in the SFARI gene set compared to deprioritized genes (35 of 398 versus 14 of 562; $P = 10^{-5}$, Fisher's exact test). Even better enrichment was obtained for the genes with the lowest RVIS scores (38 of 398 versus 11 of 494; $P = 10^{-6}$, Fisher's exact test), which is orthogonal information to our gene prioritization. The prioritized genes were also significantly enriched in genes with higher pLI scores ($P = 1.5 \times 10^{-6}$; fig. S6A), in brain-expressed genes ($P = 10^{-7}$; fig. S6B), and in 102 genes ($P = 9 \times 10^{-5}$) implicated in ASD risk according to a recent study (*38*), when compared to the nonprioritized genes. These positive control data showed that prioritization using evolutionary action pathway distributions preferentially captured current knowledge. Furthermore, of our prioritized genes, 28 of 363 (7.71%) that were not recognized by SFARI as high-confidence genes in 2017 were recognized as such in the most recent release of SFARI, whereas for the nonprioritized genes, only 11 of 548 (2%) went from unrecognized to recognized over the same time frame. This difference was highly significant in a chi-square analysis ($P < 0.0001$) and demonstrated the utility of the evolutionary action–based prioritization approach. To determine whether prioritized genes without a known link to ASD were also contributing to the relationship between genotype and phenotype, we next tested the ability of evolutionary action burden in prioritized genes to predict patient phenotype when the gene was either supported by the high-confidence curated SFARI gene set or unscored by SFARI. For each subset of the prioritized genes, patients with de novo variants in these genes were split into two groups based on whether their burden was above or below the mean of all such patients. Across all prioritized genes, the patient group with above-average evolutionary action burden demonstrated significantly lower IQ scores corresponding to an ~8-point drop in IQ ($P$ = 0.006) (Fig. 5). The difference became more pronounced when restricting to prioritized genes also in the SFARI gene set, with the average IQ a full 30 points lower in the patient group with a higher evolutionary action burden (90.3 versus 60.3, $P$ = 0.0015), 5 points more than would be found when considering the SFARI gene set without the aid of prioritization (fig. S7A). Moreover, considering prioritization status improved the unthresholded correlation value from $R = -0.33$ to $R = -0.37$ ($P$ = 0.02) (fig. S7B). However, the majority (84.5%) of prioritized genes were not placed into any category by SFARI curation, and a significant ~6.5–IQ point difference between the groups persisted when considering only these unannotated genes ($P$ = 0.03). For all tests, statistical significance was maintained across a wide range of thresholds (fig. S8A-E).

We next reperformed the analyses more stringently, comparing our prioritized gene set to an uncurated assessment of the current published literature. We defined genes with at least one association in PubMed between the gene name and the term "autism" as being supported by the literature and found that prioritized genes were significantly enriched for literature support compared to deprioritized genes ($P < 0.0001$; Fig. 6A). Simply considering the genes with the lowest RVIS scores also yielded a significant but weaker association ($P = 3 \times 10^{-12}$ compared to $P = 9 \times 10^{-21}$; fig. S9). Among all genes with literature support, those that were prioritized had a larger number of associations per gene ($P$ = 0.007; Fig. 6B), indicating more extensive support. Moreover, whereas prioritized genes with literature support exhibited a significant relationship between IQ and evolutionary action burden both when binned (Fig. 6C) and unbinned ($P$ = 0.019, linear regression), the deprioritized genes with literature support did not (Fig. 6D), suggesting that associations with deprioritized genes may have been false positives. We then tested the ability of evolutionary action burden in prioritized genes

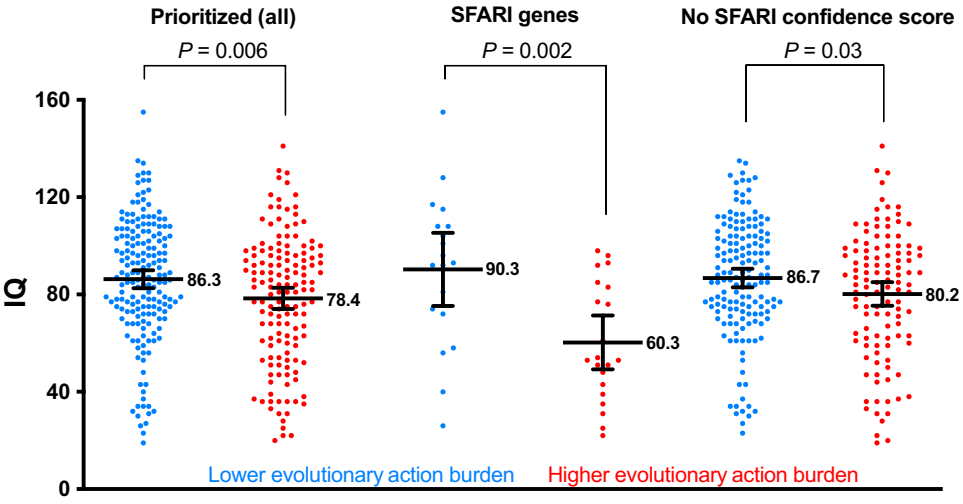


**Fig. 5. Effect of SFARI curation confidence and prioritization status on the relationship between genotype and phenotype.** There were three gene sets of interest: all prioritized genes, prioritized genes overlapping with the SFARI category 1 to 3 curated gene set, and prioritized genes without a SFARI confidence score. The evolutionary action burdens of all three gene sets for all male patients with ASD with at least one de novo missense variant within the gene set were averaged. Patients were split into higher and lower evolutionary action burden groups based on whether their score was above or below the average burden, respectively. Groups were compared statistically with an unpaired *t* test, and the mean and 95% CI interval of the mean for each group are displayed overlaying all IQ scores for patients in the group.

to predict patient phenotype when the gene was either supported or unsupported by the literature. When considering prioritized genes with literature support, patients with an above-average evolutionary action burden had IQ scores ~11 points lower than those with a below-average evolutionary action burden ($P = 0.01$; Fig. 6E). When considering prioritized genes with no literature associations with ASD, the same trend was seen with a significant decrease in IQ of more than seven points ($P = 0.04$; Fig. 6E). Again, statistical significance was maintained across a wide range of thresholds (fig. S8).

### Evolutionary action score burden of rare and low-frequency inherited variants in prioritized genes correlates with phenotype severity

Given that the impact of de novo mutations in the candidate causative genes correlated with patient presentation, we next tested whether rare inherited variations in these same genes exhibited a similar relationship with IQ. We considered rare and low-frequency inherited variants with a minor allele frequency (MAF) of less than 0.05 in male patients that were detected in at least one parent but were not inherited by the healthy sibling. Across the cohort, there were 25,042 variants in the candidate genes that met these criteria. For each patient, we calculated the inherited evolutionary action burden in the candidate genes as the summation of all evolutionary action scores in these variants after weighting for gene-specific tolerance to mutation. We found that there was a significant correlation between IQ and inherited variant evolutionary action burden as well, with patients with a high IQ having a lower inherited evolutionary action burden in the prioritized gene set ($P = 0.0005$; Fig. 7A). There was no relationship between evolutionary action burden and IQ when considering genes that were not prioritized ($P = 0.26$) or that were low-confidence SFARI genes (SFARI categories 4 to 6; $P = 0.83$); the same relationships could be found when limiting the MAF cutoff to more stringent definitions of rare variant status (Fig. 7B). Incorporation of the de novo variants into the evolutionary action burden increased significance further ($P = 0.0003$). These data show that within the prioritized gene set, rare inherited variants also linked genotype to phenotype.

## DISCUSSION

Our data show that the evolutionary action distributions of de novo missense variants can be used to elucidate causative pathways in a complex multigenic disease and prioritize variants that stratify disease severity. Here, using sequencing data from 2384 individuals diagnosed with ASD, we hypothesized that affected cohort-specific selection for large variant fitness effects within a group of functionally related genes implied an association of those pathways and genes with ASD. We observed significant impact signatures in 23 pathways, including those involved in axonogenesis, neuron development, and synaptic transmission. The mutated genes from these pathways were enriched for literature associations with ASD and are highly consistent with pathways of importance derived from analyses of CNV and LOF variant data (*12*, *36*, *39*, *40*), as well as pathways identified through recurrent missense mutations in patients with neurodevelopmental disorders (*41*). This suggested that the putative causative missense SNVs identified in this study may operate through mechanisms similar, rather than orthologous, to well-documented processes involved in ASD etiology.

Our study directly links the evolutionary impact of missense variants to a measure of ASD phenotypic severity without a priori knowledge of ASD-associated genes. Although IQ cannot reflect all aspects of the phenotypic severity of patients with ASD, it correlates well with behavior-based observer-rating scales that encompass diverse areas of autistic symptomatology (*42*) and repetitive behaviors in patients (*43*) and therefore can provide a relevant index of ASD severity. Past work relating de novo variants to IQ in patients with ASD has focused almost exclusively on CNV and LOF variants, with studies finding a significant relationship between IQ and the de novo mutation rate of LOF variants (*16*, *21*) as well as CNVs and truncating SNVs (*44*). However, when these same studies assessed missense variants, no correlation with IQ was found, even after restricting analyses to recurrent missense variants (*16*, *44*). Our initial assessments of the de novo missense variant class agreed with others who have reported that the overall impact of de novo missense variants in ASD does not differ substantially from expectations (*18*, *28*). However, we found that this collective profile did not preclude the detection of gene and variant subsets with mutational signatures indicating a significant genotype-phenotype relationship. We observed a 30-point IQ decrease in patients with above-average missense variant impact burdens across the

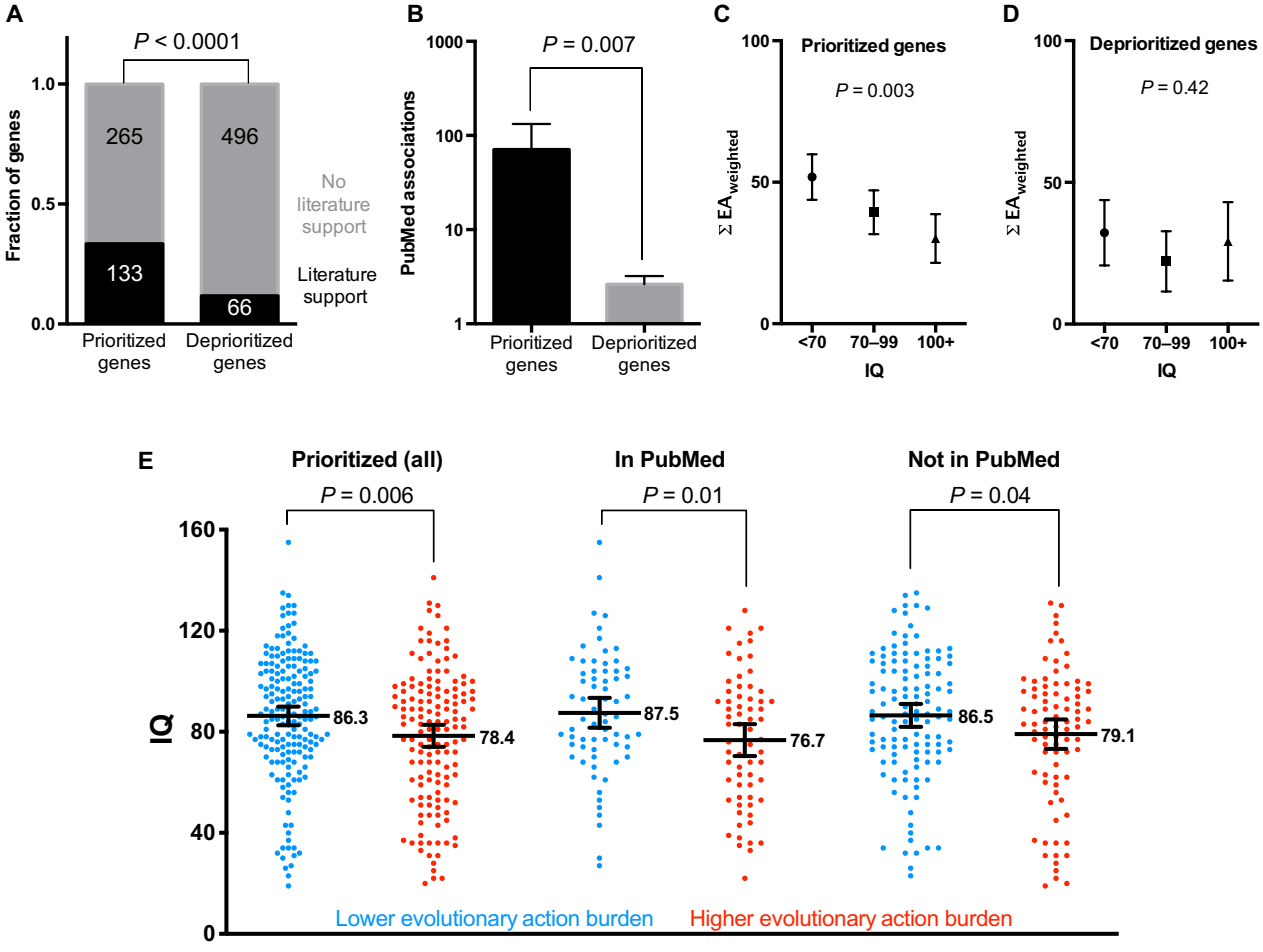


**Fig. 6. Uncurated literature associations with ASD and the effect of literature and prioritization status on the relationship between genotype and phenotype.** (**A**) Enrichment of a prioritized gene set for associations with ASD in PubMed is compared to a deprioritized gene set. (**B**) All genes with de novo missense variants and support in the literature were separated by prioritization status. Then, the numbers of PubMed associations with ASD for the genes in each category were compared using an unpaired *t* test. (**C**) De novo missense variants in prioritized genes with literature support ($n = 133$) were assessed for their relationship with ASD patient IQ. Male patients with ASD were then split into three groups according to their full-scale IQ scores, and then their weighted for genic intolerance to mutation evolutionary action burdens ($\Sigma EA_{weighted}$) were compared using a Kruskal-Wallis test. Error bars reflect the 95% CI of the mean. (**D**) De novo missense variants in deprioritized genes with literature support ($n = 66$) were assessed for their relationship with ASD patient IQ. Male patients with ASD were then split into three groups according to their full-scale IQ scores, and then their weighted for genic intolerance to mutation evolutionary action burdens ($\Sigma EA_{weighted}$) were compared using a Kruskal-Wallis test. Error bars reflect the 95% CI of the mean. (**E**) For each of the three gene sets of interest (all prioritized genes, prioritized genes with at least one PubMed association with ASD, and prioritized genes without a PubMed association with ASD), the gene set evolutionary action burdens of all male patients with ASD with at least one de novo missense variant within the gene set were averaged. Patients were split into higher and lower evolutionary action burden groups based on whether their score was above or below the average burden, respectively. Groups were compared statistically with an unpaired *t* test, and the mean and 95% CI interval of the mean for each group are displayed overlaying all IQ scores for patients in the group.

highest-confidence gene candidates and a 7.4-point IQ decrease in patients with above-average missense variant impact burdens across unexpected gene candidates. In addition, we demonstrated that different variants within the same candidate gene could be linked to phenotypic outcomes through their predicted evolutionary action impact on protein fitness. Furthermore, a modest but highly significant correlation between rare inherited missense variant burden and IQ when considering the prioritized genes indicated that these genes may contribute to ASD etiology through pathways beyond de novo variation.

Our results suggest that de novo missense variants, especially those with high impact affecting important genes in neurological pathways, have the potential to influence the phenotypic presentation of patients with ASD even if they or the genes in which they occur have not been previously linked to ASD in the literature. However, lower-impact missense variants in a gene should not be assumed to produce a similar effect even if the gene or pathway has been previously associated with ASD. These findings have implications for clinical interpretation of de novo missense variants of unknown significance in patients diagnosed with ASD, which, in turn, could improve estimations of recurrence risk in siblings by helping to clarify whether a patient's de novo missense variant influences their presentation or is merely incidental. In the future, larger cohorts and additional sequenced trios will enable refinement of the observed genotype-phenotype relationship into a clinically valuable outcome predictor and will clarify whether missense variants in

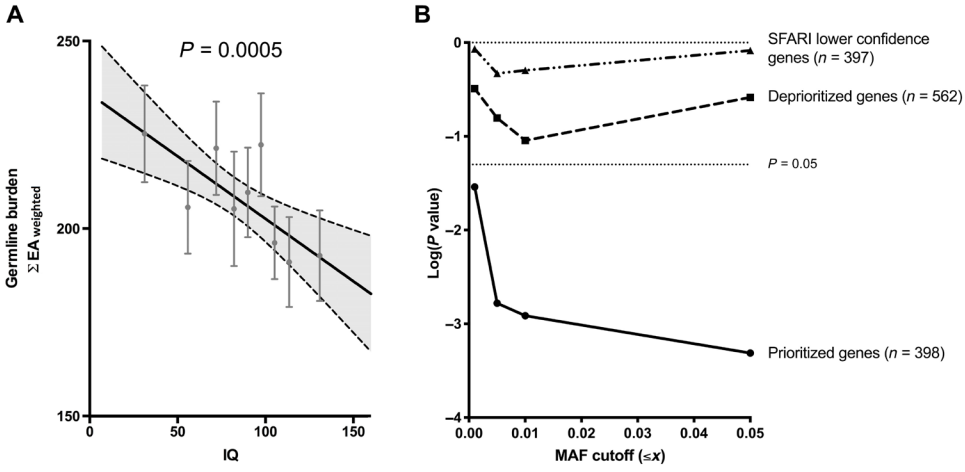


**Fig. 7. Relationship between inherited evolutionary action score burden and patient IQ for prioritized and deprioritized gene groups.** (**A**) For each male patient with ASD, rare and low-frequency inherited variants (MAF < 0.05) that were detected in at least one parent, but not inherited by the healthy sibling, were identified across prioritized genes. The inherited evolutionary action burden was calculated as the summation of all evolutionary action scores of these variants after weighting for gene tolerance to mutation ($\Sigma EA_{weighted}$). The line indicates the linear regression across all points, and the shaded gray area represents the 95% CI; the *P* value corresponds to the significance of the regression. For visualization purposes, the patients were also sorted by IQ and divided into nine equal groups; the average burden and IQ of each group are overlaid upon the regression, and error bars indicate the SEM. (**B**) Log(*P* values) of the linear regression of IQ and inherited evolutionary action burden as the MAF threshold is increasingly restricted to lower frequencies for prioritized genes, deprioritized genes, and lower-confidence SFARI genes.

female patients with ASD demonstrate the same relationship to clinical presentation.

Our results also have implications for laboratory testing by suggesting which genes and variants to prioritize for experimental validation and inclusion in the SFARI gene set. One gene with a single missense variant in the cohort, *CAMK2A*, was not included in the SFARI gene study when it was completed and had minimal literature support for an association with ASD but was prioritized by the pathway-evolutionary action integration as part of the synaptic transmission pathway. The detected variant in this gene has recently been shown to decrease excitatory synaptic transmission in cultured neurons and produce aberrant behavior including social deficits and increased repetitive behavior in mice with a knock-in of the variant (*45*). Now, *CAMK2A* has been linked to intellectual disability (*46*) and incorporated into SFARI. Although this is a single example, pathway-evolutionary action can prospectively aid ongoing large-scale experimental efforts to test the functional effect of de novo missense mutations detected in major trio studies. In addition, as statistical power grows along with cohort sizes for ASD, genes suggested by pathway-evolutionary action can be further prioritized using results of frequency-based analyses to create short lists for testing. Already 10 candidate genes, including several not in SFARI, overlap with a recent list of 35 genes identified through recurrent missense mutations in patients with neurodevelopmental disorders (*41*). Independent studies have reported four of our prioritized genes as new candidate ASD-associated genes (*19*, *20*).

Several limitations may affect the sensitivity and specificity of our evolutionary action method. First, our approach to identify de novo missense variants was stringent and biased toward high specificity rather than toward finding more variants of lower confidence. Consequently, some genes with relevant de novo variants may not have been prioritized for ASD association. Another limitation concerns the gene groups defined with GO terms, which are far from a complete accounting of molecular functions and pathways. More genes are likely to be prioritized using alternative groupings. Also, given the polygenic character of ASD and the finite number of probands, small pathways may lack enough variants to achieve statistical significance. On the other hand, very large pathways that mostly include unrelated genes to ASD may be dominated by nondriver variants, leaving the very few driver genes deprioritized. An evolutionary action limitation is that 5% of the human genes were not scored because of an insufficient number of homologous sequences. Although most of these genes are pseudogenes or functionally unimportant genes, it is possible that ASD driver genes were also included at a lower fraction. All of these limitations decrease the sensitivity of our approach, but our prioritization could also include false-positive genes. This is because any association with ASD is at the level of the gene groups and not at the level of single genes, given that most genes have only one de novo variant. Last, in the future, other phenotypes besides IQ should also be used to account for ASD phenotype severity as well as larger independent cohorts.

More broadly, the elucidation of the genotype-phenotype relationship through the integration of mutation impact and gene importance scores is an approach with implications for evolutionary theory and biology. The mathematics underpinning the use of evolutionary action distributions to identify pathways and genes of interest is founded on the assumption of an evolutionary fitness function that maps genotypes to phenotypes in the fitness landscape but which is not directly calculable. Differentiation of this fitness function yields the evolutionary action equation to predict variant impact, in which the perturbation of the fitness landscape is equal to the product of the evolutionary fitness gradient, estimated by Evolutionary Trace (*47*), and the substitution log odds of the amino acid change (*25*). These values are calculable from sequence data, and predictions have been shown to correlate well with experimental assessments of protein fitness (*25*, *48*) to consistently outperform machine learning methods (*26*, *27*) and to enable stratification of patient morbidity (*25*) and mortality (*49*) in other disease contexts. This evolutionary action theory is extended in our study by considering the distribution of variant evolutionary action scores over a pathway. Such distributions are akin to integrating the evolutionary action equation across the pathway to recover the original genotype-phenotype relationship. Significant distributions indicate a nonrandom genotype-phenotype relationship. As we show here, this new evolutionary calculus in fitness landscapes can, in practice, identify candidate phenotypic driver genes and the relationship between variant impact and patient clinical outcome. The pathway-evolutionary action approach should be generalizable beyond ASD to other multigenic diseases and phenotypes and can be applied to germline and de novo mutations alike.

## MATERIALS AND METHODS

### Study design

In this study, we used whole-exome sequencing data and associated phenotypic data from family trios and quartets of the Simons Simplex Collection (SSC) to suggest pathways and genes that may be associated with ASD and clarify the genotype to phenotype relationship of de novo missense variants in ASD. De novo missense variants in patients with ASD and their matched healthy siblings were annotated with a computationally predicted impact on fitness (evolutionary action), and functionally related groups of genes were defined by GO hierarchy and gene association data (GO2MSIG). Groups with a collective variant bias toward high impact were identified by examining the evolutionary action score distributions of their de novo missense variants. Genes that contributed variants to a biased score distribution were suggested as potentially involved in ASD. These genes were further examined for plausibility via genotype-phenotype correlations to IQ, comparisons to established knowledge, and time-stamped analyses of experimental verification.

### Data acquisition

Variant call files (.vcfs) produced by the SSC were downloaded from the National Database for Autism Research (NDAR; Study 349) (*50*); this exome sequencing data encompassed 2392 families and used FreeBayes SNV calling performed by Krumm *et al.* (*11*) at the University of Washington. Phenotype data for the associated patients were obtained from the same source. CNV data for the correlation between evolutionary action score burden and IQ were downloaded from NDAR (Study 361) (*50*) and restricted to patient de novo deletions. Researchers can obtain the underlying SSC population dataset described in this study (https://sfari.org/resources/autism-cohorts/simons-simplex-collection) by request to https://base.sfari.org.

### De novo variant calling and quality assessment

Variants were called as de novo if the proband call was heterozygous with a depth higher than 10, alternate allele fraction of 0.3 or higher, and average alternate allele quality of 15 or higher; the same position was required in both parents to have a depth of at least 30, at least 95% of reads supporting a reference call, and no more than 5 reads supporting a nonreference call. These thresholds produced a set of de novo variants indicating high quality (Ti/Tv = 2.64) and an absence of negative selection ($\lambda$ is 0.009, when 0.01 indicates no selection pressure and 0.038 indicates the negative selection pressure seen in inherited human variants) (*51*). Together, the Ti/Tv ratio and the lambda value suggest that our set of de novo variants is consistent to de novo variants reported in other studies and distinct from inherited variants, somatic variants, and random sequencing errors (fig. S10). More than 98% of the de novo variants were autosomal, whereas the rest variants were on the X chromosome. Using this procedure, we identified de novo variants in both patients and siblings. Eight families were excluded from downstream analysis because of specific technical errors that resulted in an excessive number of apparent de novo sequence events in either the patient or sibling. To focus on genes that are infrequently mutated, we did not consider genes with more than three missense mutations, which notably included the well-documented ASD driver *SCN2A* and six more genes (*HLA-B*, *MAGEC1*, *MUC4*, *MUC5B*, *PABPC1*, and *RBMX*).

### Network/gene set enrichment analysis of genes affected by de novo variants in patients

Protein-protein interactions were defined by the *Homo sapiens* STRING v.10.0 network (*29*) using the aggregate score of all evidence types and were considered as interactions if they had "medium confidence" or higher (interaction score $\geq 0.4$). Enrichment tests for protein-protein interactions, as well as gene set enrichment analysis for GO biological processes, were performed through the STRING graphical user interface. Gene sets were considered significantly enriched at the default $q < 0.05$ threshold reported by STRING. STRING analysis accounts for gene length when assessing significance of enrichments.

### Annotation of missense variants with evolutionary action

The impact of missense variants on protein fitness was computed with the evolutionary action equation, which has won multiple challenges of the Critical Assessment of Genome Interpretation community in 2017, 2015, 2013, and 2011 (*26*, *27*). Briefly, this equation follows from viewing evolution as a differentiable mapping, $f$, of genotypes ($\gamma$) onto the fitness landscape ($\varphi$), so that

$$f(\gamma) = \varphi \tag{1}$$

Differentiation then leads to the evolutionary action equation

$$\nabla f . d\gamma = d\varphi \tag{2}$$

where $\nabla f$ is the evolutionary gradient in the fitness landscape, $d\gamma$ is a genotype perturbation such as a mutation, and $d\varphi$ is the fitness effect. In practice, Eq. 2 is approximated to first order. For a substitution from amino acid type $X$ to type $Y$ at a protein residue, $r_i$, the evolutionary gradient $\nabla f$ reduces to $\partial f/\partial r_i$, which is the mutational sensitivity at $r_i$ and equivalent to its evolutionary importance defined by the Evolutionary Trace method (*47*, *52*). To estimate $d\gamma$, we use odds of amino acid substitution from $X$ to $Y$. This approach produced scores on a continuous scale between 0 and 100, where a higher value indicated a larger predicted impact on protein fitness resulting from the amino acid substitution. When a variant affected multiple isoforms of a protein, the impact score was averaged across all affected isoforms. Evolutionary Action calculations are described at greater length in the original publication of the method (*25*).

### Annotation of genes for genic tolerance to mutation

We used RVIS (*32*) as our main measure of genic sensitivity to mutation; RVIS scores were converted with the equation of mutation intolerance score = ((100 − RVIS%)/100) to lie on a scale from 0 to 1 with 1 indicating maximum intolerance to mutation. RVIS is based on mathematical determinations of mutation population frequencies and is unbiased by the state of scientific knowledge regarding ASD. For a small fraction of genes, an RVIS score did not yet exist. If a variant without an RVIS score was the only variant in a prioritized gene in the patient, the patient was not included in the IQ correlation analysis because their burden could not be accurately assessed.

### Identification of gene groups with bias toward impactful missense variants

Gene groups were defined using GO terms customized by GO2MSIG (*30*); customization was specific to *H. sapiens* and ensured at least

500 genes in each group. This approach produced 368 pathways encompassing 15,310 total genes (table S3). Gene groups with a collective variant bias toward high impact were identified by examining the evolutionary action score distributions of their missense de novo variants. For each pathway, the evolutionary action score distribution of the de novo variants within the pathway was compared to the evolutionary action distribution of all other de novo variants using a one-sided Kolmogorov-Smirnov test. Note that, to account for properties that are unique to de novo variants in patients with ASD, these comparisons were performed only within the relevant variant class. To account for multiple hypothesis testing across a large number of gene groups while maximizing discovery by limiting false negatives, groups that were significant after false discovery rate (FDR) with $q < 0.1$ were considered significant (table S4). This analysis was performed using missense de novo variants from 1792 patients with matched siblings and then repeated using missense de novo variants from the 1792 matched siblings.

### Relating evolutionary action scores in prioritized genes to patient phenotype

Although female probands with de novo missense mutations in prioritized genes contributed a very small fraction of the data (<1/7), they were highly disproportionately represented at low full-scale IQ scores (42% with IQ < 70 versus 26% for male probands) and were analyzed separately from male patients to prevent confounding based on gender. Patients with ASD were divided into three groups by phenotype severity as defined by high full-scale IQ (greater than or equal to population average), low full-scale IQ (more than 2 SD below population average and consistent with a diagnosis of intellectual disability), and intermediate full-scale IQ. Genes falling into pathways with significant bias toward high-impact mutations were grouped together into a single set of prioritized candidate ASD genes, and we considered the evolutionary action scores of mutations in these genes across the three groups for all binned analyses. For each patient, the sum of the evolutionary action scores of de novo variants in their affected candidate genes was calculated. For each gene, variant evolutionary action scores (EA) were then weighted by the RVIS score such that weighted EA = EA*mutation intolerance score, and the total patient burden was recalculated with weighted evolutionary action scores in place of the raw evolutionary action scores. For comparison, we also substituted raw ExAC LOF Constraint Metric (pLI) and ExAC Missense Constraint Metric scores as alternative measures of genic intolerance to mutation (table S8) (*53*), nonverbal and verbal IQ scores as alternate measures of phenotypic severity (table S9), other prioritization approaches (table S11), and CADD (*54*), SIFT (*55*), BLOSUM62 (*56*), MPC (*57*), and Polyphen-2 (*58*) scores as alternate measures of variant severity (table S10).

For analysis of inherited germline variants, we considered low-frequency inherited variants (MAF < 0.05) in prioritized genes that were observed in at least one parent but were not inherited by the healthy sibling (using the same thresholds to confirm the absence of the variant as were used for parent calls when determining de novo variants). The inherited evolutionary action burden for each male patient was calculated as the summation of all evolutionary action scores of these variants after weighting for genic tolerance to mutation. The MAFs of variants were obtained from ExAC (*53*) using the entire population.

### Comparison of prioritized genes to published knowledge

Genes with at least one association in PubMed between the gene name and the term autism were defined as being supported by the literature, whereas genes with no search results returned were defined as lacking literature support. These values were obtained automatically using a Biopython script on 4 October 2016. An updated PubMed search was performed on 13 January 2020, which resulted in 65 (33%) more genes associated to ASD, and this update reassured our previous conclusion (fig. S9). SFARI gene annotations were obtained from SFARI Gene and the SFARI Gene Scoring Module (*4*) on 13 January 2017. An updated version of the SFARI gene annotations was obtained on 15 September 2020, and this version was used to perform a time stamp analysis. All tests for enrichment of prioritized genes in the literature, either as defined by PubMed or SFARI, were performed using Fisher's exact tests, and in all cases, the prioritized genes were compared against the other genes affected by de novo missense mutations in the patient cohort studied, such that all genes are derived from the same source dataset.

### Statistical analysis

Collective bias of a gene set toward variants with high computationally predicted fitness impact was assessed using one-sided Kolmogorov-Smirnov tests that compared these gene set variants to variants outside that gene set. FDR correction for multiple hypothesis testing was applied, and gene sets with FDR < 0.1 were considered for further analysis. Pearson's $R$ value was calculated for correlation between continuous datasets. Data were additionally analyzed using paired $t$ tests, unpaired $t$ tests, Kruskal-Wallis tests, and two-sample Kolmogorov-Smirnoff tests as appropriate, with all tests being two-sided. $P$ values of <0.05 were considered statistically significant. Data are displayed as means with error bars representing the 95% CI (confidence interval) of the mean; statistical tests are provided in the figure legends. Statistical analyses were conducted using SciPy statistical packages for Python 2.7, and graphs were plotted using GraphPad Prism 6.0.

## SUPPLEMENTARY MATERIALS

stm.sciencemag.org/cgi/content/full/13/594/eabc1739/DC1

Fig. S1. Stratification of de novo missense variants using gene-centric network analysis.
Fig. S2. Relationship between de novo evolutionary action score burden and patient IQ for prioritized and deprioritized gene groups.
Fig. S3. Relationship between patient IQ and genic tolerance to mutation (RVIS).
Fig. S4. Relationship between evolutionary action burden and phenotype cannot be explained by the number of de novo missense variants of interest in a patient.
Fig. S5. Comparison of prioritized de novo missense variants and genotype-phenotype analysis in male versus female probands.
Fig. S6. Enrichment of prioritization status in genes with high pLI scores and in brain-expressed genes.
Fig. S7. Effect of prioritization status on the relationship between genotype and phenotype in SFARI autism gene list.
Fig. S8. Effect of threshold variation on high and low weighted evolutionary action burden group comparisons.
Fig. S9. Updated PubMed search for associating our prioritization to ASD.
Fig. S10. Quality control of de novo variant calls.
Table S1. Proband de novo missense variants.
Table S2. Healthy sibling de novo missense variants.
Table S3. GO pathway definitions.
Table S4. GO pathway bias in patients and matched siblings.
Table S5. Gene annotation with prioritization status.
Table S6. IQ and total weighted evolutionary action burden for male patients with at least one missense variant in a prioritized gene.
Table S7. Variant annotation with evolutionary action, RVIS, IQ, and gene prioritization status for variants in male patients with available phenotypic data.

Table S8. Genotype-phenotype relationship of de novo missense variants within gene sets, separated by prioritization group and metric used to estimate genic tolerance to mutation.
Table S9. Genotype-phenotype relationship of de novo missense variants within gene sets, separated by prioritization group and metric used to define phenotype.
Table S10. IQ correlation to mutation burden as measured by six prediction methods.
Table S11. IQ correlation to mutation burden as measured by evolutionary action for different gene sets.
References (*59*–*61*)

View/request a protocol for this paper from *Bio-protocol*.

## REFERENCES AND NOTES


1. I. Iossifov, M. Ronemus, D. Levy, Z. Wang, I. Hakker, J. Rosenbaum, B. Yamrom, Y. H. Lee, G. Narzisi, A. Leotta, J. Kendall, E. Grabowska, B. Ma, S. Marks, L. Rodgers, A. Stepansky, J. Troge, P. Andrews, M. Bekritsky, K. Pradhan, E. Ghiban, M. Kramer, J. Parla, R. Demeter, L. L. Fulton, R. S. Fulton, V. J. Magrini, K. Ye, J. C. Darnell, R. B. Darnell, E. R. Mardis, R. K. Wilson, M. C. Schatz, W. R. McCombie, M. Wigler, De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
2. C. Betancur, Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).
3. L. Liu, J. Lei, S. J. Sanders, A. Willsey, Y. Kou, A. Cicek, L. Klei, C. Lu, X. He, M. Li, R. A. Muhle, A. Ma'ayan, J. P. Noonan, N. Šestan, K. A. McFadden, M. W. State, J. D. Buxbaum, B. Devlin, K. Roeder, DAWN: A framework to identify autism genes and subnetworks using gene expression and genetics. *Mol. Autism.* **5**, 22 (2014).
4. B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, I. Menashe, T. Wadkins, S. Banerjee-Basu, A. Packer, SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism.* **4**, 36 (2013).
5. J. Yin, C. P. Schaaf, Autism genetics - an overview. *Prenat. Diagn.* **37**, 14–30 (2017).
6. B. S. Abrahams, D. H. Geschwind, Advances in autism genetics: On the threshold of a new neurobiology. *Nat. Rev. Genet.* **9**, 341–355 (2008).
7. J. Y. An, C. Claudianos, Genetic heterogeneity in autism: From single gene to a pathway perspective. *Neurosci. Biobehav. Rev.* **68**, 442–453 (2016).
8. M. S. Durkin, M. J. Maenner, C. J. Newschaffer, L.-C. Lee, C. M. Cunniff, J. L. Daniels, R. S. Kirby, L. Leavitt, L. Miller, W. Zahorodny, L. A. Schieve, Advanced parental age and the risk of autism spectrum disorder. *Am. J. Epidemiol.* **168**, 1268–1276 (2008).
9. K. Cheslack-Postava, K. Liu, P. S. Bearman, Closely spaced pregnancies are associated with increased odds of autism in California sibling births. *Pediatrics* **127**, 246–253 (2011).
10. T. Gaugler, L. Klei, S. J. Sanders, C. A. Bodea, A. P. Goldberg, A. B. Lee, M. Mahajan, D. Manaa, Y. Pawitan, J. Reichert, S. Ripke, S. Sandin, P. Sklar, O. Svantesson, A. Reichenberg, C. M. Hultman, B. Devlin, K. Roeder, J. D. Buxbaum, Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
11. N. Krumm, T. N. Turner, C. Baker, L. Vives, K. Mohajeri, K. Witherspoon, A. Raja, B. P. Coe, H. A. Stessman, Z. X. He, S. M. Leal, R. Bernier, E. E. Eichler, Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* **47**, 582–588 (2015).
12. D. Pinto, E. Delaby, D. Merico, M. Barbosa, A. Merikangas, L. Klei, B. Thiruvahindrapuram, X. Xu, R. Ziman, Z. Wang, J. A. S. Vorstman, A. Thompson, R. Regan, M. Pilorge, G. Pellecchia, A. T. Pagnamenta, B. Oliveira, C. R. Marshall, T. R. Magalhaes, J. K. Lowe, J. L. Howe, A. J. Griswold, J. Gilbert, E. Duketis, B. A. Dombroski, M. V. de Jonge, M. Cuccaro, E. L. Crawford, C. T. Correia, J. Conroy, I. C. Conceição, A. G. Chiocchetti, J. P. Casey, G. Cai, C. Cabrol, N. Bolshakova, E. Bacchelli, R. Anney, S. Gallinger, M. Cotterchio, G. Casey, L. Zwaigenbaum, K. Wittemeyer, K. Wing, S. Wallace, H. van Engeland, A. Tryfon, S. Thomson, L. Soorya, B. Rogé, W. Roberts, F. Poustka, S. Mouga, N. Minshew, L. A. McInnes, S. G. McGrew, C. Lord, M. Leboyer, A. S. le Couteur, A. Kolevzon, P. Jiménez González, S. Jacob, R. Holt, S. Guter, J. Green, A. Green, C. Gillberg, B. A. Fernandez, F. Duque, R. Delorme, G. Dawson, P. Chaste, C. Café, S. Brennan, T. Bourgeron, P. F. Bolton, S. Bölte, R. Bernier, G. Baird, A. J. Bailey, E. Anagnostou, J. Almeida, E. M. Wijsman, V. J. Vieland, A. M. Vicente, G. D. Schellenberg, M. Pericak-Vance, A. D. Paterson, J. R. Parr, G. Oliveira, J. I. Nurnberger, A. P. Monaco, E. Maestrini, S. M. Klauck, H. Hakonarson, J. L. Haines, D. H. Geschwind, C. M. Freitag, S. E. Folstein, S. Ennis, H. Coon, A. Battaglia, P. Szatmari, J. S. Sutcliffe, J. Hallmayer, M. Gill, E. H. Cook, J. D. Buxbaum, B. Devlin, L. Gallagher, C. Betancur, S. W. Scherer, Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* **94**, 677–694 (2014).
13. S. J. Sanders, A. G. Ercan-Sencicek, V. Hus, R. Luo, M. T. Murtha, D. Moreno-De-Luca, S. H. Chu, M. P. Moreau, A. R. Gupta, S. A. Thomson, C. E. Mason, K. Bilguvar, P. B. S. Celestino-Soper, M. Choi, E. L. Crawford, L. Davis, N. R. D. Wright, R. M. Dhodapkar, M. DiCola, N. M. DiLullo, T. V. Fernandez, V. Fielding-Singh, D. O. Fishman, S. Frahm, R. Garagaloyan, G. S. Goh, S. Kammela, L. Klei, J. K. Lowe, S. C. Lund, A. D. McGrew, K. A. Meyer, W. J. Moffat, J. D. Murdoch, B. J. O'Roak, G. T. Ober, R. S. Pottenger, M. J. Raubeson, Y. Song, Q. Wang, B. L. Yaspan, T. W. Yu, I. R. Yurkiewicz, A. L. Beaudet, R. M. Cantor, M. Curland, D. E. Grice, M. Günel, R. P. Lifton, S. M. Mane, D. M. Martin, C. A. Shaw, M. Sheldon, J. A. Tischfield, C. A. Walsh, E. M. Morrow, D. H. Ledbetter, E. Fombonne, C. Lord, C. L. Martin, A. I. Brooks, J. S. Sutcliffe, E. H. Cook Jr., D. Geschwind, K. Roeder, B. Devlin, M. W. State, Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).
14. V. M. Leppa, S. N. Kravitz, C. L. Martin, J. Andrieux, C. le Caignec, D. Martin-Coignard, C. DyBuncio, S. J. Sanders, J. K. Lowe, R. M. Cantor, D. H. Geschwind, Rare inherited and de novo CNVs reveal complex contributions to ASD risk in multiplex families. *Am. J. Hum. Genet.* **99**, 540–554 (2016).
15. T. Wang, H. Guo, B. Xiong, H. A. F. Stessman, H. Wu, B. P. Coe, T. N. Turner, Y. Liu, W. Zhao, K. Hoekzema, L. Vives, L. Xia, M. Tang, J. Ou, B. Chen, Y. Shen, G. Xun, M. Long, J. Lin, Z. N. Kronenberg, Y. Peng, T. Bai, H. Li, X. Ke, Z. Hu, J. Zhao, X. Zou, K. Xia, E. E. Eichler, De novo genic mutations among a Chinese autism spectrum disorder cohort. *Nat. Commun.* **7**, 13316 (2016).
16. I. Iossifov, B. J. O'Roak, S. J. Sanders, M. Ronemus, N. Krumm, D. Levy, H. A. Stessman, K. T. Witherspoon, L. Vives, K. E. Patterson, J. D. Smith, B. Paeper, D. A. Nickerson, J. Dea, S. Dong, L. E. Gonzalez, J. D. Mandell, S. M. Mane, M. T. Murtha, C. A. Sullivan, M. F. Walker, Z. Waqar, L. Wei, A. J. Willsey, B. Yamrom, Y. H. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M. C. Schatz, K. Ye, W. R. McCombie, J. Shendure, E. E. Eichler, M. W. State, M. Wigler, The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
17. A. Buja, N. Volfovsky, A. M. Krieger, C. Lord, A. E. Lash, M. Wigler, I. Iossifov, Damaging de novo mutations diminish motor skills in children on the autism spectrum. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1859–E1866 (2018).
18. S. J. Sanders, M. T. Murtha, A. R. Gupta, J. D. Murdoch, M. J. Raubeson, A. J. Willsey, A. G. Ercan-Sencicek, N. M. DiLullo, N. N. Parikshak, J. L. Stein, M. F. Walker, G. T. Ober, N. A. Teran, Y. Song, P. el-Fishawy, R. C. Murtha, M. Choi, J. D. Overton, R. D. Bjornson, N. J. Carriero, K. A. Meyer, K. Bilguvar, S. M. Mane, N. Šestan, R. P. Lifton, M. Günel, K. Roeder, D. H. Geschwind, B. Devlin, M. W. State, De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
19. S. H. Lelieveld, L. Wiel, H. Venselaar, R. Pfundt, G. Vriend, J. A. Veltman, H. G. Brunner, L. E. L. M. Vissers, C. Gilissen, Spatial clustering of de novo missense mutations identifies candidate neurodevelopmental disorder-associated genes. *Am. J. Hum. Genet.* **101**, 478–484 (2017).
20. S. Chen, R. Fragoza, L. Klei, Y. Liu, J. Wang, K. Roeder, B. Devlin, H. Yu, An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nat. Genet.* **50**, 1032–1040 (2018).
21. E. B. Robinson, K. E. Samocha, J. A. Kosmicki, L. McGrath, B. M. Neale, R. H. Perlis, M. J. Daly, Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15161–15165 (2014).
22. L. A. Miosge, M. A. Field, Y. Sontani, V. Cho, S. Johnson, A. Palkova, B. Balakishnan, R. Liang, Y. Zhang, S. Lyon, B. Beutler, B. Whittle, E. M. Bertram, A. Enders, C. C. Goodnow, T. D. Andrews, Comparison of predicted and actual consequences of missense mutations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5189–E5198 (2015).
23. P. Katsonis, A. Koire, S. J. Wilson, T. K. Hsu, R. C. Lua, A. D. Wilkins, O. Lichtarge, Single nucleotide variations: Biological impact and theoretical interpretation. *Protein Sci.* **23**, 1650–1666 (2014).
24. S. Hicks, S. E. Plon, M. Kimmel, Statistical analysis of missense mutation classifiers. *Hum. Mutat.* **34**, 405–406 (2013).
25. P. Katsonis, O. Lichtarge, A formal perturbation equation between genotype and phenotype determines the evolutionary action of protein-coding variations on fitness. *Genome Res.* **24**, 2050–2058 (2014).
26. P. Katsonis, O. Lichtarge, Objective assessment of the evolutionary action equation for the fitness effect of missense mutations across CAGI blinded contests. *Hum. Mutat.* **38**, 1072–1084 (2017).
27. P. Katsonis, O. Lichtarge, CAGI5: Objective performance assessments of predictions based on the Evolutionary Action equation. *Hum. Mutat.* **40**, 1436–1454 (2019).
28. B. M. Neale, Y. Kou, L. Liu, A. Ma'ayan, K. E. Samocha, A. Sabo, C. F. Lin, C. Stevens, L. S. Wang, V. Makarov, P. Polak, S. Yoon, J. Maguire, E. L. Crawford, N. G. Campbell, E. T. Geller, O. Valladares, C. Schafer, H. Liu, T. Zhao, G. Cai, J. Lihm, R. Dannenfelser, O. Jabado, Z. Peralta, U. Nagaswamy, D. Muzny, J. G. Reid, I. Newsham, Y. Wu, L. Lewis, Y. Han, B. F. Voight, E. Lim, E. Rossin, A. Kirby, J. Flannick, M. Fromer, K. Shakir, T. Fennell, K. Garimella, E. Banks, R. Poplin, S. Gabriel, M. DePristo, J. R. Wimbish, B. E. Boone, S. E. Levy, C. Betancur, S. Sunyaev, E. Boerwinkle, J. D. Buxbaum, E. H. Cook Jr, B. Devlin, R. A. Gibbs, K. Roeder, G. D. Schellenberg, J. S. Sutcliffe, M. J. Daly, Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
29. D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, C. von Mering, STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).

30. J. A. Powell, GO2MSIG, an automated GO based multi-species gene set generator for gene set enrichment analysis. *BMC Bioinformatics* **15**, 146 (2014).
31. R. K. C. Yuen, D. Merico, M. Bookman, J. L. Howe, B. Thiruvahindrapuram, R. V. Patel, J. Whitney, N. Deflaux, J. Bingham, Z. Wang, G. Pellecchia, J. A. Buchanan, S. Walker, C. R. Marshall, M. Uddin, M. Zarrei, E. Deneault, L. D'Abate, A. J. S. Chan, S. Koyanagi, T. Paton, S. L. Pereira, N. Hoang, W. Engchuan, E. J. Higginbotham, K. Ho, S. Lamoureux, W. Li, J. R. MacDonald, T. Nalpathamkalam, W. W. L. Sung, F. J. Tsoi, J. Wei, L. Xu, A.-M. Tasse, E. Kirby, W. Van Etten, S. Twigger, W. Roberts, I. Drmic, S. Jilderda, B. MacKinnon Modi, B. Kellam, M. Szego, C. Cytrynbaum, R. Weksberg, L. Zwaigenbaum, M. Woodbury-Smith, J. Brian, L. Senman, A. Iaboni, K. Doyle-Thomas, A. Thompson, C. Chrysler, J. Leef, T. Savion-Lemieux, I. M. Smith, X. Liu, R. Nicolson, V. Seifer, A. Fedele, E. H. Cook, S. Dager, A. Estes, L. Gallagher, B. A. Malow, J. R. Parr, S. J. Spence, J. Vorstman, B. J. Frey, J. T. Robinson, L. J. Strug, B. A. Fernandez, M. Elsabbagh, M. T. Carter, J. Hallmayer, B. M. Knoppers, E. Anagnostou, P. Szatmari, R. H. Ring, D. Glazer, M. T. Pletcher, S. W. Scherer, Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).
32. S. Petrovski, Q. Wang, E. L. Heinzen, A. S. Allen, D. B. Goldstein, Genic intolerance to functional variation and the interpretation of personal genomes. *PLOS Genet.* **9**, e1003709 (2013).
33. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. M. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
34. J. C. Darnell, S. J. van Driesche, C. Zhang, K. Y. S. Hung, A. Mele, C. E. Fraser, E. F. Stone, C. Chen, J. J. Fak, S. W. Chi, D. D. Licatalosi, J. D. Richter, R. B. Darnell, FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
35. N. N. Parikshak, R. Luo, A. Zhang, H. Won, J. K. Lowe, V. Chandran, S. Horvath, D. H. Geschwind, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
36. S. R. Gilman, I. Iossifov, D. Levy, M. Ronemus, M. Wigler, D. Vitkup, Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* **70**, 898–907 (2011).
37. S. Girirajan, M. Y. Dennis, C. Baker, M. Malig, B. P. Coe, C. D. Campbell, K. Mark, T. H. Vu, C. Alkan, Z. Cheng, L. G. Biesecker, R. Bernier, E. E. Eichler, Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am. J. Hum. Genet.* **92**, 221–237 (2013).
38. F. K. Satterstrom, J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J.-Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias; Autism Sequencing Consortium; iPSYCH-Broad Consortium, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, J. D. Buxbaum, Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584.e23 (2020).
39. J. T. Glessner, K. Wang, G. Cai, O. Korvatska, C. E. Kim, S. Wood, H. Zhang, A. Estes, C. W. Brune, J. P. Bradfield, M. Imielinski, E. C. Frackelton, J. Reichert, E. L. Crawford, J. Munson, P. M. A. Sleiman, R. Chiavacci, K. Annaiah, K. Thomas, C. Hou, W. Glaberson, J. Flory, F. Otieno, M. Garris, L. Soorya, L. Klei, J. Piven, K. J. Meyer, E. Anagnostou, T. Sakurai, R. M. Game, D. S. Rudd, D. Zurawiecki, C. J. McDougle, L. K. Davis, J. Miller, D. J. Posey, S. Michaels, A. Kolevzon, J. M. Silverman, R. Bernier, S. E. Levy, R. T. Schultz, G. Dawson, T. Owley, W. M. McMahon, T. H. Wassink, J. A. Sweeney, J. I. Nurnberger, H. Coon, J. S. Sutcliffe, N. J. Minshew, S. F. A. Grant, M. Bucan, E. H. Cook, J. D. Buxbaum, B. Devlin, G. D. Schellenberg, H. Hakonarson, Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569–573 (2009).
40. S. Shohat, E. Ben-David, S. Shifman, Varying intolerance of gene pathways to mutational classes explain genetic convergence across neuropsychiatric disorders. *Cell Rep.* **18**, 2217–2227 (2017).
41. M. R. Geisheker, G. Heymann, T. Wang, B. P. Coe, T. N. Turner, H. A. F. Stessman, K. Hoekzema, M. Kvarnung, M. Shaw, K. Friend, J. Liebelt, C. Barnett, E. M. Thompson, E. Haan, H. Guo, B. M. Anderlid, A. Nordgren, A. Lindstrand, G. Vandeweyer, A. Alberti, E. Avola, M. Vinci, S. Giusto, T. Pramparo, K. Pierce, S. Nalabolu, J. J. Michaelson, Z. Sedlacek, G. W. E. Santen, H. Peeters, H. Hakonarson, E. Courchesne, C. Romano, R. F. Kooy, R. A. Bernier, M. Nordenskjöld, J. Gecz, K. Xia, L. S. Zweifel, E. E. Eichler, Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. *Nat. Neurosci.* **20**, 1043–1051 (2017).
42. T. Nishiyama, H. Taniai, T. Miyachi, K. Ozaki, M. Tomita, S. Sumi, Genetic correlation between autistic traits and IQ in a population-based sample of twins with autism spectrum disorders (ASDs). *J. Hum. Genet.* **54**, 56–61 (2009).
43. J. Richler, S. L. Bishop, J. R. Kleinke, C. Lord, Restricted and repetitive behaviors in young children with autism spectrum disorders. *J. Autism Dev. Disord.* **37**, 73–85 (2007).
44. B. J. O'Roak, L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D. Smith, E. H. Turner, I. B. Stanaway, B. Vernot, M. Malig, C. Baker, B. Reilly, J. M. Akey, E. Borenstein, M. J. Rieder, D. A. Nickerson, R. Bernier, J. Shendure, E. E. Eichler, Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
45. J. R. Stephenson, X. Wang, T. L. Perfitt, W. P. Parrish, B. C. Shonesy, C. R. Marks, D. P. Mortlock, T. Nakagawa, J. S. Sutcliffe, R. J. Colbran, A novel human *CAMK2A* mutation disrupts dendritic morphology and synaptic transmission, and causes ASD-related behaviors. *J. Neurosci.* **37**, 2216–2233 (2017).
46. S. Küry, G. M. van Woerden, T. Besnard, M. P. Onori, X. Latypova, M. C. Towne, M. T. Cho, T. E. Prescott, M. A. Ploeg, S. Sanders, H. A. F. Stessman, A. Pujol, B. Distel, L. A. Robak, J. A. Bernstein, A.-S. Denommé-Pichon, G. Lesca, E. A. Sellars, J. Berg, W. Carré, Ø. L. Busk, B. W. M. van Bon, J. L. Waugh, M. Deardorff, G. E. Hoganson, K. B. Bosanko, D. S. Johnson, T. Dabir, Ø. L. Holla, A. Sarkar, K. Tveten, J. de Bellescize, G. J. Braathen, P. A. Terhal, D. K. Grange, A. van Haeringen, C. Lam, G. Mirzaa, J. Burton, E. J. Bhoj, J. Douglas, A. B. Santani, A. I. Nesbitt, K. L. Helbig, M. V. Andrews, A. Begtrup, S. Tang, K. L. I. van Gassen, J. Juusola, K. Foss, G. M. Enns, U. Moog, K. Hinderhofer, N. Paramasivam, S. Lincoln, B. H. Kusako, P. Lindenbaum, E. Charpentier, C. B. Nowak, E. Cherot, T. Simonet, C. A. L. Ruivenkamp, S. Hahn, C. A. Brownstein, F. Xia, S. Schmitt, W. Deb, D. Bonneau, M. Nizon, D. Quinquis, J. Chelly, G. Rudolf, D. Sanlaville, P. Parent, B. Gilbert-Dussardier, A. Toutain, V. R. Sutton, J. Thies, L. E. L. M. Peart-Vissers, P. Boisseau, M. Vincent, A. M. Grabrucker, C. Dubourg; Undiagnosed Diseases Network, W.-H. Tan, N. E. Verbeek, M. Granzow, G. W. E. Santen, J. Shendure, B. Isidor, L. Pasquier, R. Redon, Y. Yang, M. W. State, T. Kleefstra, B. Cogné; GEM HUGO; Deciphering Developmental Disorders Study, S. Petrovski, K. Retterer, E. E. Eichler, J. A. Rosenfeld, P. B. Agrawal, S. Bézieau, S. Odent, Y. Elgersma, S. Mercier, De novo mutations in protein kinase genes *CAMK2A* and *CAMK2B* cause intellectual disability. *Am. J. Hum. Genet.* **101**, 768–788 (2017).
47. O. Lichtarge, H. R. Bourne, F. E. Cohen, An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* **257**, 342–358 (1996).
48. J. Gallion, A. Koire, P. Katsonis, A. M. Schoenegge, M. Bouvier, O. Lichtarge, Predicting phenotype from genotype: Improving accuracy through more robust experimental and computational modeling. *Hum. Mutat.* **38**, 569–580 (2017).
49. D. M. Neskey, A. A. Osman, T. J. Ow, P. Katsonis, T. McDonald, S. C. Hicks, T. K. Hsu, C. R. Pickering, A. Ward, A. Patel, J. S. Yordy, H. D. Skinner, U. Giri, D. Sano, M. D. Story, B. M. Beadle, A. K. el-Naggar, M. S. Kies, W. N. William, C. Caulin, M. Frederick, M. Kimmel, J. N. Myers, O. Lichtarge, Evolutionary action score of *TP53* identifies high-risk mutations associated with decreased survival and increased distant metastases in head and neck cancer. *Cancer Res.* **75**, 1527–1536 (2015).
50. N. Payakachat, J. M. Tilford, W. J. Ungar, National Database for Autism Research (NDAR): Big data opportunities for health services research and health technology assessment. *Pharmacoeconomics* **34**, 127–138 (2016).
51. A. Koire, P. Katsonis, O. Lichtarge, Repurposing germline exomes of the cancer genome atlas demands a cautious approach and sample-specific variant filtering. *Pac. Symp. Biocomput.* **21**, 207–218 (2016).
52. I. Mihalek, I. Res, O. Lichtarge, A family of evolution-entropy hybrid methods for ranking protein residues by importance. *J. Mol. Biol.* **336**, 1265–1282 (2004).
53. M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. M. Carroll, M. I. McCarthy, D. M. Govern, R. M. Pherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur; Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
54. M. Kircher, D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper, J. Shendure, A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
55. P. C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).

56. S. Henikoff, J. G. Henikoff, Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10915–10919 (1992).
57. K. E. Samocha, J. A. Kosmicki, K. J. Karczewski, A. H. O'Donnell-Luria, E. Pierce-Hoffman, D. G. MacArthur, B. M. Neale, M. J. Daly, Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* **2017**, 148353 (2017).
58. I. A. Adzhubei, S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, S. R. Sunyaev, A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
59. B. Xu, I. Ionita-Laza, J. L. Roos, B. Boone, S. Woodrick, Y. Sun, S. Levy, J. A. Gogos, M. Karayiorgou, De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* **44**, 1365–1369 (2012).
60. L. C. Francioli, P. P. Polak, A. Koren, A. Menelaou, S. Chun, I. Renkens; Genome of the Netherlands Consortium, C. M. van Duijn, M. Swertz, C. Wijmenga, G. van Ommen, P. E. Slagboom, D. I. Boomsma, K. Ye, V. Guryev, P. F. Arndt, W. P. Kloosterman, P. I. W. de Bakker, S. R. Sunyaev, Genome-wide patterns and properties of de novo mutations in humans. *Nat. Genet.* **47**, 822–826 (2015).
61. EuroEPINOMICS-RES Consortium; Epilepsy Phenome/Genome Project; Epi4K Consortium, De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am. J. Hum. Genet.* **95**, 360–370 (2014).


**Acknowledgments:** We are grateful to all of the families at the participating Simons Simplex Collection sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, and E. Wijsman). We appreciate obtaining access to phenotypic data from SFARI. **Funding:** This work is supported by the NIH (grant numbers GM079656-8, DE025181, GM066099, and AG061105 to O.L.), the Oskar Fischer Foundation (to O.L.), the NSF (grant number DBI1356569 to O.L.), and the Defense Advance Research Project Agency (grant number N66001-15-C-4042 to O.L). A.K. is supported by the Baylor College of Medicine Comprehensive Cancer Training Program grant number RP160283, the Baylor Research Advocates for Student Scientists, and the McNair MD/PhD Scholars program. **Author contributions:** A.K. conceptualized and planned the project with advice from Y.W.K., S.J.W., C.B., P.K., and O.L. Data acquisition was performed by A.K., C.B., and O.L. De novo variant calling and quality assessment were performed by A.K. and C.B., and processing of data to score variants with the evolutionary action method was performed by P.K. Network and gene set enrichment analysis of genes affected by de novo variants was performed by A.K. and S.J.W. Time-stamped analysis and comparisons to other variant prediction methods were performed by Y.W.K. and A.K. Remaining analyses including pathway-based prioritization of de novo missense variants, genotype-phenotype correlations, and additional validation analyses were performed by A.K. The manuscript was written by A.K., C.B., P.K., and O.L., and all authors commented on the manuscript. **Competing interests:** O.L., P.K., and A.K. are coinventors on U.S. patent no. 10886005B2 entitled "Identifying genes associated with a phenotype." **Data and materials availability:** All data associated with this study are present in the paper or the Supplementary Materials. Data for generating the figures are in the Supplementary Tables. Researchers can obtain the underlying SSC population dataset described in this study (https://sfari.org/resources/autism-cohorts/simons-simplex-collection) by request to https://base.sfari.org.

Submitted 10 April 2020
Accepted 1 March 2021
Published 19 May 2021
10.1126/scitranslmed.abc1739

**Citation:** A. Koire, P. Katsonis, Y. W. Kim, C. Buchovecky, S. J. Wilson, O. Lichtarge, A method to delineate de novo missense variants across pathways prioritizes genes linked to autism. *Sci. Transl. Med.* **13**, eabc1739 (2021).