RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART B
Neuropsychiatric Genetics

The Contribution of Protein Intrinsic Disorder to Understand the Role of Genetic Variants Uncovered by Autism Spectrum Disorders Exome Studies

Jaqueline Bohrer Schuch,[1] Vanessa Rodrigues Paixão-Côrtes,[2] Deise C. Friedrich,[3] and Luciana Tovo-Rodrigues[4]*

[1]Department of Genetics, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

[2]Postgraduate Program in Genetics and Biodiversity, Federal University of Bahia (UFBA), Salvador, Bahia, Brazil

[3]Department of Cellular and Molecular Biology, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Porto Alegre, Rio Grande do Sul, Brazil

[4]Postgraduate Program in Epidemiology, Federal University of Pelotas (UFPel), Pelotas, Rio Grande do Sul, Brazil

Manuscript Received: 24 August 2015; Manuscript Accepted: 4 February 2016

Several autism spectrum disorders (ASD) exome studies suggest that coding single nucleotide variants (SNVs) play an important role on ASD etiology. Usually, the pathogenic effect of missense mutations is estimated through predictors that lose accuracy for those SNVs placed in intrinsically disordered regions of protein. Here, we used bioinformatics tools to investigate the effect of mutations described in ASD published exome studies (549 mutations) in protein disorder, considering post-translational modification, PEST and Molecular Recognition Features (MoRFs) motifs. Schizophrenia and type 2 diabetes (T2D) datasets were created for comparison purposes. The frequency of mutations predicted as disordered was comparable among the three datasets (38.1% in ASD, 35.7% in schizophrenia, 46.4% in T2D). However, the frequency of SNVs predicted to lead a gain or loss of functional sites or change intrinsic disorder tendencies was higher in ASD and schizophrenia than T2D (46.9%, 36.4%, and 23.1%, respectively). The results obtained by SIFT and PolyPhen-2 indicated that 38.9% and 34.4% of the mutations predicted, respectively, as tolerated and benign showed functional alterations in disorder properties. Given the frequency of mutations placed in IDRs and their functional impact, this study suggests that alterations in intrinsic disorder properties might play a role in ASD and schizophrenia etiologies. They should be taken into consideration when researching the pathogenicity of mutations in neurodevelopmental and psychiatric diseases. Finally, mutations with functional alterations in disorder properties must be potential targets for in vitro and in vivo functional studies. © 2016 Wiley Periodicals, Inc.

Key words: exome; protein intrinsic disorder; autism spectrum disorders; post-translational modification

How to Cite this Article:
Schuch JB, Paixão-Côrtes VR, Friedrich DC, Tovo-Rodrigues L. 2016. The Contribution of Protein Intrinsic Disorder to Understand the Role of Genetic Variants Uncovered by Autism Spectrum Disorders Exome Studies.
Am J Med Genet Part B 171B:479–491.

INTRODUCTION

Autism spectrum disorders (ASD) comprise a group of neurodevelopmental conditions characterized by impairments in communication and social interaction as well as repetitive and stereotyped behaviors. This disease has a strong genetic component: in approximately 15% of cases, several genetic syndromes or alterations (e.g., duplications, inversions) in chromosomal regions are associated with their etiology [Betancur, 2011]. Although its heritability is one of the highest among all psychiatric disorders (around 80%, Ronald and Hoekstra, 2011]), the genetic architecture of ASD is still not completely understood. A considerable number of molecular studies have identified both common and rare variants, demonstrating the complexity of this disorder. One of the most promising approaches consists of exome analysis, which involves the sequencing of all protein-coding regions of the

Conflicts of interest: The authors declare that they have no conflicts of interest.
Grant sponsor: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil).
*Correspondence to:
Luciana Tovo-Rodrigues, Ph.D., Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, Caixa Postal 464-96020-220, Pelotas, Rio Grande do Sul, Brazil.
E-mail: luciana.tovo@gmail.com
Article first published online in Wiley Online Library (wileyonlinelibrary.com): 19 February 2016
DOI 10.1002/ajmg.b.32431

genome. Attempting to explain the "missing heritability" for complex diseases, the underlying premise of exome analysis is based on the "common disease-rare variant" hypothesis, which proposes that rare variants with large effects could contribute to the disorder's development [Manolio et al., 2009; Ionita-Laza et al., 2011].

Using both case–control and family designs, ASD exome studies have revealed inherited and de novo mutations associated with different biological mechanisms, such as chromatin remodeling, neurodevelopment, serotonin and glutamate neurotransmission, cell migration and organization, and ubiquitination [O'Roak et al., 2011; Bi et al., 2012; Chahrour et al., 2012; Neale et al., 2012; O'Roak et al., 2012a, b; Sanders et al., 2012; Shi et al., 2013; Toma et al., 2013; Yu et al., 2013; Cukier et al., 2014]. Although missense mutations are suggested to play a role in ASD etiology, the functionality of these mutations was predicted regardless of the protein region to which the mutations are mapped (i.e., structured or disordered regions) [O'Roak et al., 2011; Bi et al., 2012; Chahrour et al., 2012; Neale et al., 2012; Sanders et al., 2012; Shi et al., 2013; Toma et al., 2013; Cukier et al., 2014].

Intrinsically disordered regions (IDRs) are very flexible protein segments that lack secondary structure in solution. They are characterized by a biased amino acid composition (i.e., enriched in hydrophilic, charged, and structure-breaking residues and depleted in bulky hydrophobic and aromatic ones) and may be present in different lengths in the protein, including the whole protein. Proteins containing long IDRs are involved in important functional roles in the cell, such as signaling and molecular regulation [Iakoucheva et al., 2002; Uversky et al., 2005], usually serving as hubs (i.e., highly connected proteins) in protein interaction networks [Dunker et al., 2005]. Their conformational flexibility allows IDRs to expose interaction-prone linear motifs (Molecular Recognition Features; MoRFs), as well as post-translational modification (PTM) motifs [Torkamani et al., 2008; Vacic et al., 2012]. PEST motifs (i.e., classical acidic protein degradation targeting signals enriched in proline [P], glutamic acid [E], serine [S], and threonine [T]) are another important feature related to protein stability and interaction frequently observed inside disordered regions [Singh et al., 2006]. Considering the functional relevance, factors that alter intrinsically disordered proteins' regulation or function may lead to disease. An excess of intrinsically disordered proteins (IDPs) has already been observed in many diseases, such as cancer, diabetes, and cardiovascular and neurodegenerative disorders [Iakoucheva et al., 2002; Uversky et al., 2008].

It has been suggested that genetic variants inside disordered regions differ in frequency, function, and other features when compared to those found in ordered regions (OR) [Vacic et al., 2012]. In contrast to mutations inside ORs, pathogenic mutations inside IDRs are more prone to cause disorder-to-order (D-O) changes and to disrupt/create post-translational modification sites [Radivojac et al., 2008; Li et al., 2010; Vacic et al., 2012]. Most of the common methods used to predict the function of non–synonymous mutations are based on solved or modeled protein structures [Chasman and Adams, 2001] and/or on evolutionary conservation [Saunders and Baker, 2002]. It was shown that they do not have the same accuracy when the protein region is very malleable and has low conservation scores (i.e., the case of intrinsic disordered protein) [Brown et al., 2011; Vacic et al., 2012]. Missense mutations in IDRs may impact those features of the protein directly related to intrinsic disorder, altering proper IDP interaction, regulation, signaling pathways, and thus cell environment.

Exome studies have emerged as an appropriate approach to estimating the role of mutations in IDRs as they are able to identify rare variants with a possible high risk of disease susceptibility. Given the supportive evidence for the role of rare mutations in ASD etiology, this study aims to characterize missense mutations reported in published ASD exome studies, by (i) estimating the rate of mutations found in IDRs, (ii) estimating the putative functional impact of each mutation, (ii) comparing the IDR estimates with brain and non-brain diseases, and (iv) characterizing mutations using SIFT and PolyPhen-2 predictors, and comparing with intrinsic disorder results. The effect of SNVs on D-O transitions, PEST motifs, PTM sites, and MoRF motifs was evaluated. The results suggest that mutations in disordered regions play an important role in ASD etiology. This study provides relevant information to further our understanding of the molecular causation of neurodevelopmental diseases and it reinforce the relevance of considering disordered regions to estimate the functional impact of mutations.

METHODS

Datasets

A comprehensive search in the literature for whole exome studies in autism samples was conducted, including all studies published until January 2014, and revealed a total of 11 studies [O'Roak et al., 2011; Bi et al., 2012; Chahrour et al., 2012; Neale et al., 2012; O'Roak et al., 2012a, b; Sanders et al., 2012; Shi et al., 2013; Toma et al., 2013; Yu et al., 2013; Cukier et al., 2014]. The most relevant and reliable missense mutations described in each article (as discussed by the authors) were included in the present study, comprising a total of 549 missense mutations located in 482 genes.

For further comparison with psychiatric and non-psychiatric diseases, SNVs described in exome studies of schizophrenia [Need et al., 2012; Xu et al., 2012a, b; Timms et al., 2013; Fromer et al., 2014; Guipponi et al., 2014; McCarthy et al., 2014; Ambalavanan et al., 2015; Kranz et al., 2015] and type 2 diabetes (T2D; [Lohmueller et al., 2013; Estrada et al., 2014; Steinthorsdottir et al., 2014]) were also included in the analyses. A total of 154 SNVs in 136 proteins were retrieved from schizophrenia exome studies. For T2D, only mutations with MAF$<$5% were included, comprising 28 SNVs in 26 proteins.

Protein sequences encoded by these genes were retrieved from the UniProtKB/Swiss-Prot database (http://www.uniprot.org/). Studies and proteins included in this investigation are described in Supplementary Tables SI and SII. Wild protein sequences were manually mutated using the MEGA vs6 program [Tamura et al., 2013]. In order to estimate the functional effect of each single nucleotide variant (SNV), both wild and mutated sequences were analyzed for the parameters evaluated in this study and compared, as described below.

Protein Intrinsic Disorder Predictions

The intrinsic disorder of full length wild-type and mutated proteins was predicted using IUPRED [Dosztanyi et al., 2005], PONDR-VSL2B [Obradovic et al., 2005], PONDR-VL3 [Obradovic et al., 2003], PONDR-VLXT [Romero et al., 2001], and PONDR-FIT [Xue et al., 2010], accessed through the DisProt database [Sickmeier et al., 2007].

These programs, except PONDR-FIT, have already been used to estimate the impact of mutations in D-O and order-to-disorder (O-D) transitions [Vacic et al., 2012]. We included PONDR-FIT in the analyses because it is a consensus artificial neural network (ANN) metapredictor developed by combining the outputs of several individual disorder predictors. All predictors estimate a per residue disorder score. Scores <0.5 were indicative of order, whereas those ≥ 0.5 indicated disorder. D-O and O-D transitions were estimated using these cut-offs.

Additional analysis was performed to estimate the random chance of D-O and O-D transitions considering the codon distribution. We predicted the probability of a given amino acid to be changed due to a random mutation. For this analysis, all the multiples codons for all 20 amino acids were considered. All nucleotide positions of the codons were individually mutated taking into account all possible nucleotides (A, C, T, or G). The codons resulting from mutations were translated using the universal genetic code. Only missense mutations were taken into account. Subsequently, we classified the amino acids in three categories based on previous report concerning their enrichment in disorder or order regions [Radivojac et al., 2007]: (I) disorder promoting (Arg, Glu, Gln, Lys, Met, Pro, and Ser); (II) order promoting (Cys, Ile, Leu, Phe, Trp, Tyr, and Val); and (III) neutral (Ala, Asn, Asp, Gly, His, and Thr,). The rate of D-D, D-O, O-O, and O-D transitions was then calculated.

Long regions of disorder (IDRs) are defined as 30 or more consecutive residues of protein predicted as disordered. In this study, the term intrinsically disordered proteins (IDPs) refers to proteins having IDRs.

Linear Binding Sites Prediction

The Molecular Recognition Feature Predictor (MoRFpred; [Disfani et al., 2012]) was used to predict the binding regions in proteins (http://biomine-ws.ece.ualberta.ca/MoRFpred/index.html). This software is able to identify all types of MoRFs: α, β, γ, and complex (which are mixtures of these features). Only MoRFs that were predicted to be five or more tandem residues in length were considered.

PEST Motif Prediction

PEST motifs were predicted through the ePESTfind server (http://emboss.bioinformatics.nl/cgi-bin/emboss/epestfind; [Rechsteiner and Rogers, 1996]). The score prediction is based on the local enrichment of critical amino acids and on the motif's hydrophobicity. Only valid predicted PEST motifs were included in this study (i.e., PEST motif scores ≥ 5).

Post-Translational Modification Prediction

Three different PTM types were evaluated: phosphorylation, ubiquitination, and palmitoylation. The prediction of phosphorylation of serine, threonine, and tyrosine residues was performed using Disphos {Iakoucheva et al., 2004] and Nephos 2.0 [Blom et al., 1999] programs. We considered scores within a window of four amino acids around the described mutation. The Disphos (http://www.dabi.temple.edu/disphos/) uses data about disordered regions to differentiate phosphorylation and non-phosphorylation sites. The Nephos program (http://www.cbs.dtu.dk/services/NetPhos/) predicts phosphorylation sites using a neural network method. Phosphorylation indicated by at least one of the predictors was considered as a phosphorylation site here.

Ubiquitination prediction was estimated using the UbiPred webserver [Tung and Ho, 2008]. UbiPred is a Support Vector Machine (SVM)-based prediction server using that detects the presence/absence of ubiquitylation site with a prediction accuracy of 84.44% using leave-one-out cross-validation.

Palmitoylation prediction was estimated using the CSS-Palm 4.0 webserver [Ren et al., 2008], which uses a clustering and scoring strategy for cysteine residue palmitoylation prediction.

SIFT and Polyphen-2 Prediction

In order to compare the results of function prediction considering intrinsic disorder properties with usual missense mutation predictors, the SNVs analyzed here were submitted to SIFT [Kumar et al., 2009] and PolyPhen-2 [Adzhubei et al., 2010] predictors. The SIFT algorithm is a multistep, sequence-homology based algorithm that uses evolutionary conservation of the amino acids within protein families to classify amino acid substitutions. In contrast, the PolyPhen-2 algorithm uses a naïve Bayes classifier to predict damaging effects of nonsynonymous variants based on sequence-based and structure-based predictive features. We obtained the predicted functional scores of all missense SNVs by means of the online versions of the SIFT algorithm (http://sift.jcvi.org/index.html) and the PolyPhen-2 algorithm (http://genetics.bwh.harvard.edu/pph2/).

Functional Annotation of Proteins

Functional enrichment analysis was performed using the WebGestaltc Program [Zhang et al., 2005; Wang et al., 2013]. Proteins were investigated for overrepresentation of Gene Ontology (GO) biological process (GOBP), molecular function (GOMF) and cellular component (GOCC) terms. Enrichment was evaluated with a conditional hypergeometric test that considers the dependence structure of GO terms [Alexa et al., 2006]. The universe of Human Entrez genes with at least one GO term annotation was used as background in the statistical test. *P*-values were adjusted for multiple tests with the Benjamini–Hochberg correction [Benjamini and Hochberg, 1995] and considered significant at $P < 0.05$. In the last step of the analysis, the enriched GO terms were clustered and filtered for redundancy using REVIGO [Supek et al., 2011], adopting the "SimRel" semantic similarity measure with a similarity cut-off of 0.5.

RESULTS

Protein Intrinsic Disorder Estimates

The results of 11 published ASD exome studies were included in this investigation. We analyzed 549 missense variants located in 482 proteins. Two independent datasets were further evaluated for comparison purposes. The mutations described in nine schizophrenia (154 mutations; 136 proteins) and three T2D (28 mutations; 26 proteins) published exome studies (Supplementary Tables SI and SII) were included in the analyses.

Five different predictors (PONDR-VLXT, PONDR-VSL2, PONDR-VSL3, PONDR-FIT, and IUPred) were used to estimate the disorder score for the alleles in the causative loci. Considering the wild protein sequences, the percentage of loci predicted as disordered varied among predictors (Table I). In agreement with Vacic et al. [2012], PONDR VLXT was the most sensitive program for estimating the differences in disorder prediction between wild and mutated sequences, as indicated by the largest delta range (Table I). It was chosen, then, to perform the remaining analyses. According to PONDR VLXT, 78.4% of the proteins of ASD dataset presented long regions of disorder. Considering the same dataset, 38.1% (n = 209) of the wild variants were predicted as intrinsically disordered and 19.1% were located in long regions of disorder (Fig. 1). The proportion of SNVs predicted as disordered observed in schizophrenia and T2D datasets were very similar to ASD: 35.7% and 46.4%, respectively (Table I).

Functional Sites and Disorder-to-Order Transition Estimates

Functional in silico estimates, considering disorder properties, showed that 54.0% of SNVs predicted as disordered in ASD dataset were located in putative functional sites. The rates observed for schizophrenia and T2D datasets were 61.2% and 46.2%, respectively. The most observed functional sites for ASD, schizophrenia, and T2D datasets was phosphorylation (37.3%, 56.9%, and 46.2%, respectively). Ubiquitination site was the second most observed for ASD and schizophrenia groups (17.2% and 9.1%, respectively) (Table II).

Even though the enrichment of sites predicted as disordered and the number of sites placed in functional sites were similar among the three diseases datasets, important differences between them emerged when we considered disorder prediction tendency and loss/gain of sites.

In order to evaluate the impact of SNVs on disorder prediction, we first estimated the probability of D-O or O-D transitions to occur by chance. Considering codon distribution, and given a random mutation, the probability of a disorder-promoting amino acid to be replaced for an order-promoting corresponded to 48% (e.g., lead to D-O transition). The O-D transition probability was estimated in 31.3%. Thus, it is expect that D-O transitions occur 1.55 times more often than O-D transitions. Interestingly, SNVs of both ASD and schizophrenia datasets preferentially induced D-O transitions rather than O-D transitions compared to the expected

TABLE I. Proportion of Wild and Mutated Alleles Predicted as Disordered for Each Program, the Range of the Difference Between Wild and Mutated Alleles Disorder Score, as Well as the Frequency of Disorder-to-Order Mutations (D-O) and Order to Disorder Mutations (O-D)

Predictor	Wild allele (%)	Mutant allele (%)	Delta (Δ)—wild-mutant median (minimum, maximum)	D-O[a] (%)	O-D[b] (%)
Autism spectrum disorders (N = 549 mutations; 482 proteins)					
PONDR VLXT	38.1	36.3	0 (−0.4 to 0.6)	12.4	4.7
PONDR VSL2	46.5	42.6	0 (−0.2 to 0.2)	13.7	4.8
PONDR VSL3	35.9	34.1	0 (−0.2 to 0.3)	9.1	2.3
PONDR FIT	26.7	26.4	0 (−0.3 to 0.4)	8.9	3.0
IUPred	28.8	27.5	0 (−0.3 to 0.1)	10.1	2.3
Schizophrenia (N = 154 mutations; 136 proteins)					
PONDR VLXT	35.7	31.8	0 (−0.4 to 0.5)	18.2	4.0
PONDR VSL2	47.4	42.2	0 (−0.1 to 0.3)	12.3	1.2
PONDR VSL3	41.6	39.6	0 (−0.1 to 0.3)	6.3	1.1
PONDR FIT	24.0	25.3	0 (−0.2 to 0.3)	2.7	2.6
IUPred	24.0	24.0	0 (−0.1 to 0.1)	5.4	0.9
Type 2 diabetes (N = 28 mutations; 26 proteins)					
PONDR VLXT	46.4	46.4	0 (−0.1 to 0.2)	0	0
PONDR VSL2	53.6	53.6	0 (−0.1 to 0.2)	0	0
PONDR VSL3	46.4	46.4	0 (−0.1 to 0.1)	7.7	6.7
PONDR FIT	28.6	32.1	0 (−0.1 to 0.1)	12.5	10.0
IUPred	35.7	35.7	0 (−0.1 to 0.1)	0	0

[a]Proportion considering SNVs predicted as disordered.
[b]Proportion considering SNVs predicted as ordered.

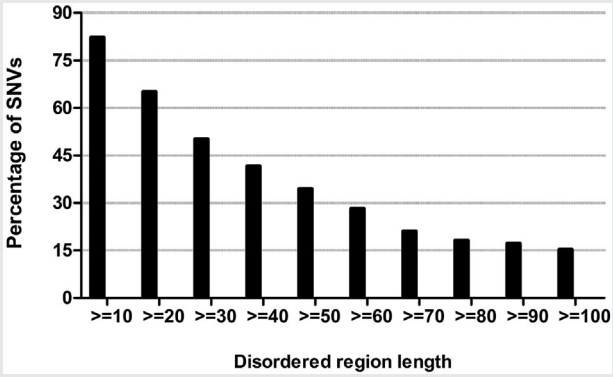

FIG. 1. Characterization of SNVs regarding the length of the protein stretch predicted as disordered.

ratio. The estimated D-O/O-D ratio for ASD was 2.6 (D-O: 12.4% vs. O-D: 4.7%), and for schizophrenia it was 4.6 (D-O: 18.2% vs. O-D: 4.0%). This tendency was consistent across different predictors. On the other hand, T2D dataset presented a different pattern. Although it presented the highest frequency of SNVs predicted as disordered, D-O and O-D transitions were not observed in this dataset (Table I).

Including the disruption/gain of sites, a total of 46.9% (n = 98) of the ASD loci predicted as disordered led to functional changes in at least one of the properties evaluated (disruption or gain of functional sites, or D-O transitions). This frequency was 1.3-fold the observed for schizophrenia (36.4%) and twice the observed for T2D (23.1%).

In ASD and schizophrenia datasets, except for ubiquitination sites (and PEST motifs for schizophrenia dataset), the results showed a consistent trend, suggesting that mutations in disordered regions often lead to a disruption of putative functional sites, mainly phosphorylation sites. Disruption of PEST and MoRF motifs in long regions of disorder for ASD dataset was also an important feature. For PEST motifs, the mutations led to a disruption three times more frequently than a gain, and for MoRF motifs, all mutations lead to disruption of functional sites (Supplementary Table SIII). In opposite, the trend observed in T2D dataset suggests gaining of sites (Table II).

Disruption of phosphorylation sites was the most relevant alteration for both ASD and schizophrenia datasets (Table II). In order to compare the observed rates of gain and loss of phosphorylation sites with those expected by chance, we estimated a reference ratio of gain/loss of sites using the putative neutral

TABLE II. Characterization of Exome Variants Regarding Putative Functional Effect

	SNVs predicted as ordered	SNVs predicted as disordered			
Functional site	**Wild[a] (%)**	**Wild[a] (%)**	**Mutant (%)**	**Disruption of sites (%)**	**Gain of sites (%)**
Autism spectrum disorders (SNVs ordered N = 340, disordered N = 209)					
Ubiquitination	12.1	17.2	19.1	4.3	6.2
Palmitoylation	2.9	1.0	0.5	0.9	0.5
Phosphorylation	22.1	37.3	30.1	18.8	8.7
PEST motif	0.3	4.3	2.9	2.8	1.4
MoRF	2.6	4.3	1.9	2.4	0
Schizophrenia (SNVs ordered N = 99, disordered N = 55)					
Ubiquitination	11.1	9.1	14.5	0	5.4
Palmitoylation	1.0	0	0	0	0
Phosphorylation	36.4	56.9[b]	52.9[b]	12.7	1.8
PEST motif	0	7.3	9.1	0	1.8
MoRF	3.1[c]	5.5	3.6	1.8	0
Type 2 diabetes (SNVs ordered N = 15, disordered N = 13)					
Ubiquitination	13.3	0	0	0	0
Palmitoylation	0	0	7.7	0	7.7
Phosphorylation	46.7	46.2	46.2	7.7	7.7
PEST motif	0	0	0	0	0
MoRF	0	7.7	7.7	0	0

SNVs, single nucleotide variants; PEST, acidic proteolytic cleavage sites; MoRF, molecular recognition features.
[a]Long regions of disorder are defined as 30 or more consecutive residues of protein sequence predicted as disordered.
[b]Four sequences were excluded from analyses.
[c]Two sequences were excluded from analyses.

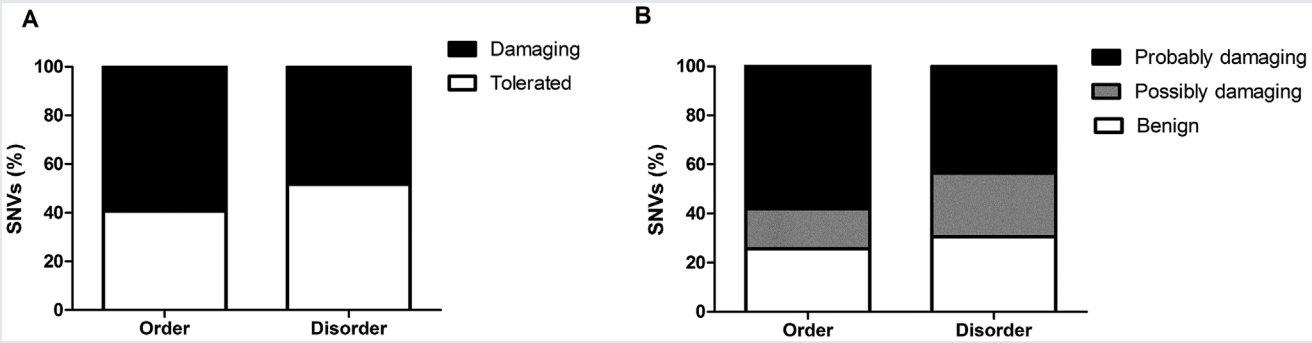

FIG. 2. Putative effect of SNVs according to SIFT and PolyPhen-2 prediction. (A) SIFT prediction (518 sites evaluated; Fisher's exact test *P*-value: 0.015). (B) PolyPhen-2 prediction (548 sites evaluated; Fisher's exact test *P*-value: 0.0015).

variants data published by Radivojac et al. [2008]. The mean ratio of loss/gain of phosphorylation sites using three putative neutral datasets of Radivojac et al. [2008] (i.e., neutral data from Ensembl, data from SeattleSNPs, and data from Swiss Prot Random mutation datasets) was estimated in 0.9. The results of loss/gain ratio observed for ASD and schizophrenia were 2.2- and 7.9-fold the expected ratio for neutral variation, respectively, whereas T2D dataset presented similar values: 1.1-fold.

Functional Effect Estimates Comparing Intrinsic Disorder Properties and Usual Predictors for ASD SNVs

The comparison between SNVs predicted as ordered and disordered suggests that the SIFT and PolyPhen-2 results are different with regard to protein structure regions (Fig. 2, SIFT Fisher's exact test $P = 0.015$; PolyPhen2 $P = 0.0015$). Both programs predicted damaging or probably damaging mutations more often in ordered regions than in disordered regions, suggesting that the prediction of damaging mutations is not homogeneous along the protein sequence. Specifically for mutations inside disordered regions, we observed an association between alteration in disorder properties and prediction of damaging mutations (SIFT, Fisher's exact test $P = 0.019$) and probably/possibly damaging categories considering the PolyPhen2 results (Fisher's exact test $P = 0.045$, Fig. 3). We emphasize that on the one hand we observed a correlation between putative alteration in disorder properties and damaging prediction, but on the other hand a great proportion of SNVs that showed functional alterations in disordered properties are predicted as tolerated (38.9%) and benign (34.4%), as shown in Figure 3. Some examples of SNVs that were predicted as both tolerated and benign/possibly damaging are shown in Table III.

Functional Characterization of Proteins Presenting Alteration in Disorder Properties

Functional enrichment (i.e., GOBP, GOMF, GOCC terms) was analyzed to functionally characterize the proteins associated with ASD, which SNVs induced a putative modification regarding intrinsic disorder properties. The GO terms with significant enrichment ($P < 0.05$) are shown in Table IV. In general, the proteins are predicted to be involved in maintenance of protein location in the cell, protein ubiquitination, and cytoskeletal protein binding and are located in several parts of the cell, such as cell–cell

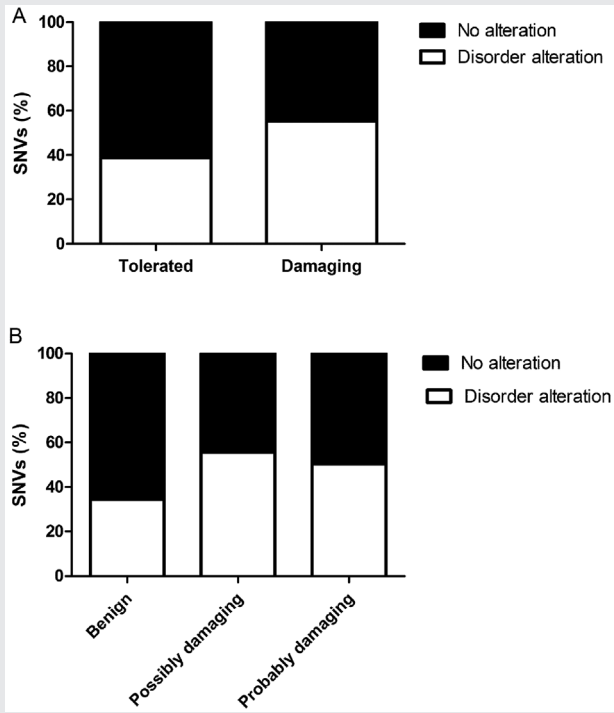

FIG. 3. Frequency of SNVs that lead to functional alteration in protein intrinsic disorder properties considering the different categories of functionality provided by SIFT and PolyPhen-2 predictors. (A) SIFT prediction (209 sites evaluated; Fisher's exact test *P*-value: 0.019). (B) Polyphen2 prediction (209 sites evaluated; Fisher's exact test *P*-value: 0.045).

TABLE III. Examples of SNVs Predicted as Tolerated by SIFT and Benign/Possibly Damaging by Polyphen-2 and With Putative Effect on Protein Function Considering Intrinsic Disorder Properties

Gene (protein)	SNV	IDR length	Reference	SIFT/PolyPhen-2 prediction	Disorder effect	Gene function	Other associations with ASD and/or other diseases
FAM63B (FA63B)	K234E	234	Neale et al. [2012]	Tolerated/benign	Disruption of ubiquitination site K234. Functional evidence for ubiquitination at position K234 (http://www.phosphosite.org/siteAction.do?id=21174471).	Part of the networks regulated by miRNA that can be linked to neuronal differentiation and dopaminergic gene expression [Aberg et al., 2014].	ASD: Genome wide linkage study with positive findings on chr15q21.1–q22-2 [Allen-Brady et al., 2010]. CNV detected on chr15q21.3 [Christian et al., 2008]. SCZ: *FAM63B* was a top find in a methylome association study [Aberg et al., 2014].
GTF2IRD1 (GT2D1)	K648E	50	Sanders et al. [2012]	Tolerated/benign	Disruption of putative ubiquitination site K648. Gain of PEST site of 32 amino acids between position 639 and 672.	It may have the ability to interact with other HLH-proteins and function as a transcription factor or as a positive transcriptional regulator under the control of retinoblastoma protein. This gene plays a role in craniofacial and cognitive development.	ASD and/or Williams-Beuren Syndrome: CNV detected on chr7q11.23 [Depienne et al., 2007; Levy et al., 2011; Sanders et al., 2011; Malenfant et al., 2012].
SCN1A (SCN1A)	Q1166R	68	Toma et al. [2013]	Tolerated/possibly damaging	Disruption of PEST motif of 43 aa between positions 1144 and 1188.	Mediates the voltage-dependent sodium ion permeability of excitable membranes.	ASD: association in family design studies [Weiss et al., 2003].
ANK3 (ANK3)	M2271T	60	Sanders et al. [2012]	Tolerated/possibly damaging	Gain of a putative phosphorylation site T2271.	Membrane-cytoskeleton linker. May participate in the maintenance/targeting of ion channels and cell adhesion molecules at the nodes of Ranvier and axonal initial segments.	ASD: CNV detected [Sebat et al., 2007; Pinto et al., 2010; Sanders et al., 2011]. Association with psychiatric disorders [Baum et al., 2008; Schizophrenia Psychiatric Genome-Wide Association Study, 2011; Iqbal et al., 2013; Judy et al., 2013; Hayashi et al., 2015].

TABLE IV. Gene Ontology Terms (GO) Enrichment for Proteins in Which SNVs Induced a Putative Functional Modification

GO biological process terms		***P*-value**
GO:0032507	Maintenance of protein location in cell	0.0456
GO:0042787	Protein ubiquitination involved in ubiquitin-dependent protein catabolic process	0.0456
GO molecular function terms		
GO:0030507	Spectrin binding	0.0048
GO:0008092	Cytoskeletal protein binding	0.0065
GO cellular component terms		
GO:0005911	Cell–cell junction	0.0020
GO:0033268	Node of Ranvier	0.0032
GO:0030054	Cell junction	0.0035
GO:0044449	Contractile fiber part	0.0136
GO:0043292	Contractile fiber	0.0142
GO:0043226	Organelle	0.0168
GO:0045211	Postsynaptic membrane	0.0168
GO:0005737	Cytoplasm	0.0168
GO:0043227	Membrane-bounded organelle	0.0168
GO:0044424	Intracellular part	0.0259
GO:0005622	Intracellular	0.0312
GO:0005856	Cytoskeleton	0.0380
GO:0031981	Nuclear lumen	0.0380
GO:0005634	Nucleus	0.0405
GO:0031252	Cell leading edge	0.0440
GO:0005604	Basement membrane	0.0446
GO:0031974	Membrane-enclosed lumen	0.0448
GO:0005813	Centrosome	0.0448

junctions, nodes of Ranvier, postsynaptic membranes, and the cytoskeleton, among others.

DISCUSSION

Autism spectrum disorders comprise a wide range of neurodevelopmental diseases that result from a combination of environmental and genetic factors. Recently, ASD has been explored under the "common disease-rare variant" hypothesis and many potential causative variants have been identified in whole-exome studies. Collectively, the findings observed by exome studies represent solid evidence for the contribution of de novo protein-coding mutations in the etiology of ASD [O'Roak et al., 2011; Neale et al., 2012; O'Roak et al., 2012b; Sanders et al., 2012]. Although the pathogenic effect of missense mutations observed in exome analyses has been already explored, protein disorder properties were never considered in these predictions. In this study, the functional roles of variants observed in exome studies were reevaluated, taking intrinsic disorder protein properties into account. A comparison with psychiatric and non-psychiatric multifactorial diseases was also performed.

In this study, we observed that proteins involved in ASD etiology presented a high content of intrinsic disorder. Approximately 80% of the proteins exhibited long regions of disorder, evidencing the relevance of considering intrinsic disorder to understand the role of SNVs. About 40% of the SNVs were located in disordered regions. This is higher than the reported by Vacic et al. [2012; 22%] for of the mendelian diseases' causative mutations, probably due to the polygenic characteristic of ASD genetic architecture. Approximately half of them (46.9%) putatively affect protein function through disruption or gain of post-translational modifications and functional sites, or lead to D-O transition. It is noteworthy that although the proportion of mutations predicted as disorder in schizophrenia and T2D datasets was similar to the estimated for ASD, the impact of mutations on disorder properties was less frequent in schizophrenia and T2D datasets when compared to ASD (46.9%, 36.4%, and 23.1%, respectively).

It is reported that mutations in disordered regions often cause D-O transition and influence predicted MoRFs [Hu et al., 2011; Vacic and Iakoucheva, 2012]. Also, proteome-wide analyses of disease-related mutations have shown that gain or loss of post-translational modification sites, which are generally found in IDRs, contributes to human disease [Radivojac et al., 2008; Li et al., 2010]. Moreover, Uyar et al. [2014], studying cancer causative mutations, showed that disease-related mutations were frequently observed inside short linear motifs in IDRs. Many examples of mutations in disordered regions mentioned in literature can be cited. For instance, mutations leading to D-O in transcription factor resulted in lower efficiency of dissociation from DNA [Dembinski et al., 2014]. In addition, mutations in disordered regions in the catenin gene (CTNNB1) are implicated in various cancers [Wang et al., 2008]. In our study, mutations associated to ASD preferentially induced D-O transitions. Vacic et al. [2012] observed that the frequency of D-O transitions is higher in disease mutations, compared to polymorphisms and neutral evolutionary substitutions. Although lower than the rate observed for schizophrenia,

the D-O/O-D ratio observed for ASD was 1.7-fold that expected by chance and that observed for T2D. Moreover, similarly to the observed for schizophrenia, the pattern of alterations in ASD dataset indicated a trend toward increased disruption of sites compared to gain of sites, which could lead to signaling dysregulation.

Disordered regions are more prone to genetic variation and have higher evolutionary rates when compared to ordered regions due to relaxed protein structural constraint. Vacic et al. [2012] observed that mutations located in disordered regions were less accurately predicted than mutations in ordered one, due to the specific properties of predictors. The SNVs identified in ASD exome studies presented a tendency to disrupt MoRFs, PEST, and post-translational motifs, suggesting that alteration of protein signaling, degradation, location, and interaction may be impaired in these proteins, instead of just alterations in protein structure. The comparisons with commonly used SNP effect predictors reinforce the relevance of these findings. Although there was an association between SIFT, Polyphen-2, and effect in disorder properties prediction, more than 34% of mutations that affect intrinsic disorder properties are misclassified as benign or tolerated in ASD dataset. One example of alteration on PTM sites was observed in the ANK3 protein, a member of the ankyrin family, which gained a putative phosphorylation site at 2271 position (M2271T). This SNV was predicted as tolerated (SIFT) and possibly damaging (PolyPhen-2), contrasting with results considering protein disorder. Phosphorylation modification of ankyrin was previously reported in regulating affinity for spectrin tetramers [Lu et al., 1985]. This PTM alteration may be relevant as ANK3 binds to diverse proteins connected with cell adhesion molecules and voltage-gated sodium channels, playing an important role on the organization of the cytoskeleton in neuronal cells. Other examples of proteins impaired by changes in disorder tendencies observed for ASD exome data analyzed are shown in Table III.

Biologically, the functional sites considered in this study are involved in a number of activities such as regulation of gene expression, protein signaling, protein location, activation/deactivation of enzymatic activity, protein stability or degradation, and mediation of protein–protein interactions. Functional enrichment analysis suggested that these functions are probably impaired in proteins involved in processes associated mainly with cytoskeleton and scaffolding related functions, such as cell–cell junctions, and maintenance of protein location in the cell. Such processes are associated with neuron morphology and synapse assembly and have already been associated with ASD pathophysiology [Banerjee et al., 2014]. Moreover, the result suggested that the machinery involved in post-translational modifications may also be impaired due to alterations in disorder properties in ASD etiology. Protein ubiquitination emerged as one of the enriched functions. Ubiquitination and proteasome-mediated degradation have been suggested as universal mechanisms in the control of synaptic protein homeostasis and have already been associated with ASD neurobiology [Glessner et al., 2009; Lin and Man, 2013; Crider et al., 2014]. The overall scenario suggests that the biology of ASD is linked to basal cell processes, also underlying synaptic function. The heterogeneity of ASD phenotype possibly could be associated to general and not specific processes.

This study has some limitations but addresses important issues since it rescues usually neglected mutation in psychiatric studies. First, it is possible that other functional sites not evaluated here, such as other PTMs, may be involved in the pathophysiology of ASD. It might have led to a possible underestimate of functional mutations predicted as disordered. Second, only a subset of the variants (missense mutations) described as relevant in the exome studies were included in this study. This severely limited the sample size of the datasets (mainly T2D) what resulted in a lack of statistic power to find significant differences among them. However, we assumed that exome studies aim to capture rare variants with high effects, identifying plausible variants that are directly involved in ASD, schizophrenia, and T2D etiology. Identification of reliable mutations through further exome studies is desirable to confirm our findings. Finally, the predictors used, although very accurate, are not definitive to define the impact of SNVs in the protein and the estimates should be interpreted with caution as pointed by Radivojac et al. [2008]. Nonetheless, the results are encouraging clues and suggest good targets for further functional in vitro experiments.

To our knowledge, this is the first study that considered intrinsic protein disorder in the evaluation of the role of mutations identified in exomes of neurodevelopmental and psychiatric diseases. Almost 40% of the mutations analyzed for ASD lie in disordered region, and more than 34% of the mutations predicted as tolerated (SIFT) and benign (PolyPhen-2) showed functional alterations in disorder properties. The present results reinforce the importance of properly analyzing SNVs in disordered regions, supporting the relevance of considering protein regions in order to estimate the functional role of mutations. Moreover, considering the similarities between ASD and schizophrenia and the differences with type 2 diabetes, it is possible that alterations in intrinsic disorder are common and important features in psychiatric diseases. Finally, further studies evaluating the impact of mutations on IDP regulatory and coding regions would be interesting to better understand the role of those proteins in ASD and other psychiatric diseases.

ACKNOWLEDGMENT

The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) for financial support.

REFERENCES

Aberg KA, McClay JL, Nerella S, Clark S, Kumar G, Chen W, Khachane AN, Xie L, Hudson A, Gao G, Harada A, Hultman CM, Sullivan PF, Magnusson PK, van den Oord EJ. 2014. Methylome-wide association study of schizophrenia: Identifying blood biomarker signatures of environmental insults. JAMA Psychiatry 71(3):255–264.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7(4):248–249.

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22(13):1600–1607.

Allen-Brady K, Robison R, Cannon D, Varvil T, Villalobos M, Pingree C, Leppert MF, Miller J, McMahon WM, Coon H. 2010. Genome-wide linkage in Utah autism pedigrees. Mol Psychiatry 15(10):1006–1015.

Ambalavanan A, Girard SL, Ahn K, Zhou S, Dionne-Laporte A, Spiegelman D, Bourassa CV, Gauthier J, Hamdan FF, Xiong L, Dion PA, Joober R, Rapoport J, Rouleau GA. 2015. De novo variants in sporadic cases of childhood onset schizophrenia. Eur J Hum Genet 1–5. doi: 10.1038/ejhg.2015.218 [Epub ahead of print].

Banerjee S, Riordan M, Bhat MA. 2014. Genetic aspects of autism spectrum disorders: Insights from animal models. Front Cell Neurosci 8:58.

Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nothen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Hofels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. 2008. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry 13(2):197–207.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 57:289–300.

Betancur C. 2011. Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain Res 1380:42–77.

Bi C, Wu J, Jiang T, Liu Q, Cai W, Yu P, Cai T, Zhao M, Jiang YH, Sun ZS. 2012. Mutations of ANK3 identified by exome sequencing are associated with autism susceptibility. Hum Mutat 33(12):1635–1638.

Blom N, Gammeltoft S, Brunak S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294(5):1351–1362.

Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. 2011. Evolution and disorder. Curr Opin Struct Biol 21(3):441–446.

Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS, Stevens CR, Schubert CR, Collaboration AAS, Greenberg ME, Gabriel SB, Walsh CA. 2012. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. PLoS Genet 8(4):e1002635.

Chasman D, Adams RM. 2001. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation. J Mol Biol 307(2):683–706.

Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, Karamohamed S, Badner JA, Matsui S, Conroy J, McQuaid D, Gergel J, Hatchwell E, Gilliam TC, Gershon ES, Nowak NJ, Dobyns WB, Cook EH Jr. 2008. Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biol Psychiatry 63(12):1111–1117.

Crider A, Pandya CD, Peter D, Ahmed AO, Pillai A. 2014. Ubiquitin-proteasome dependent degradation of GABAAalpha1 in autism spectrum disorder. Mol Autism 5:45.

Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, Leyva N, Konidari I, Gentry RC, Hulme WF, Booven DV, Mayo V, Hofmann NK, Schmidt MA, Martin ER, Haines JL, Cuccaro ML, Gilbert JR, Pericak-Vance MA. 2014. Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. Mol Autism 5(1):1.

Dembinski H, Wismer K, Balasubramaniam D, Gonzalez HA, Alverdi V, Iakoucheva LM, Komives EA. 2014. Predicted disorder-to-order transition mutations in IkappaBalpha disrupt function. Phys Chem Chem Phys 16(14):6480–6485.

Depienne C, Heron D, Betancur C, Benyahia B, Trouillard O, Bouteiller D, Verloes A, LeGuern E, Leboyer M, Brice A. 2007. Autism, language delay and mental retardation in a patient with 7q11 duplication. J Med Genet 44(7):452–458.

Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L. 2012. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics 28(12): 75–83.

Dosztanyi Z, Csizmok V, Tompa P, Simon I. 2005. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21(16):3433–3434.

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 272(20):5129–5148.

Estrada K, Aukrust I, Bjørkhaug L, Burtt NP, Mercader JM, García-Ortiz H, Huerta-Chagoya A, Moreno-Macías H, Walford G, Flannick J, Williams AL, Gómez-Vázquez MJ, Fernandez-Lopez JC, Martínez-Hernández A, Jiménez-Morales S, Centeno-Cruz F, Mendoza-Caamal E, Revilla-Monsalve C, Islas-Andrade S, Córdova EJ, Soberón X, González-Villalpando ME, Henderson E, Wilkens LR, Le Marchand L, Arellano-Campos O, Ordóñez-Sánchez ML, Rodríguez-Torres M, Rodríguez-Guillén R, Riba L, Najmi LA, Jacobs SBR, Fennell T, Gabriel S, Fontanillas P, Hanis CL, Lehman DM, Jenkinson CP, Abboud HE, Bell GI, Cortes ML, Boehnke M, González-Villalpando C, Orozco L, Haiman CA, Tusié-Luna T, Aguilar-Salinas CA, Altshuler D, Njølstad PR, Florez JC, MacArthur DG. 2014. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA 311:2305–2314.

Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, Carrera N, Humphreys I, Johnson JS, Roussos P, Barker DD, Banks E, Milanova V, Grant SG, Hannon E, Rose SA, Chambert K, Mahajan M, Scolnick EM, Moran JL, Kirov G, Palotie A, McCarroll SA, Holmans P, Sklar P, Owen MJ, Purcell SM, O'Donovan MC. 2014. De novo mutations in schizophrenia implicate synaptic networks. Nature 506:179–184.

Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, Imielinski M, Frackelton EC, Reichert J, Crawford EL, Munson J, Sleiman PM, Chiavacci R, Annaiah K, Thomas K, Hou C, Glaberson W, Flory J, Otieno F, Garris M, Soorya L, Klei L, Piven J, Meyer KJ, Anagnostou E, Sakurai T, Game RM, Rudd DS, Zurawiecki D, McDougle CJ, Davis LK, Miller J, Posey DJ, Michaels S, Kolevzon A, Silverman JM, Bernier R, Levy SE, Schultz RT, Dawson G, Owley T, McMahon WM, Wassink TH, Sweeney JA, Nurnberger JI, Coon H, Sutcliffe JS, Minshew NJ, Grant SF, Bucan M, Cook EH, Buxbaum JD, Devlin B, Schellenberg GD, Hakonarson H. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 459(7246):569–573.

Guipponi M, Santoni FA, Setola V, Gehrig C, Rotharmel M, Cuenca M, Guillin O, Dikeos D, Georgantopoulos G, Papadimitriou G, Curtis L, Méary A, Schürhoff F, Jamain S, Avramopoulos D, Leboyer M, Rujescu D, Pulver A, Campion D, Siderovski DP, Antonarakis SE. 2014. Exome sequencing in 53 sporadic cases of schizophrenia identifies 18 putative candidate genes. PLoS ONE 9:e112745.

Hayashi A, Le Gal K, Sodersten K, Vizlin-Hodzic D, Agren H, Funa K. 2015. Calcium-dependent intracellular signal pathways in primary cultured adipocytes and ANK3 gene variation in patients with bipolar disorder and healthy controls. Mol Psychiatry 20(8):931–940.

Hu Y, Liu Y, Jung J, Dunker AK, Wang Y. 2011. Changes in predicted protein disorder tendency may contribute to disease risk. BMC Genomics 12(Suppl 5):S2.

Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323(3):573–584.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. 2004. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 32(3):1037–1049.

Ionita-Laza I, Buxbaum JD, Laird NM, Lange C. 2011. A new testing strategy to identify rare variants with either risk or protective effect on disease. PLoS Genet 7(2):e1001289.

Iqbal Z, Vandeweyer G, van der Voet M, Waryah AM, Zahoor MY, Besseling JA, Roca LT, Vulto-van Silfhout AT, Nijhof B, Kramer JM, Van der Aa N, Ansar M, Peeters H, Helsmoortel C, Gilissen C, Vissers LE, Veltman JA, de Brouwer AP, Frank Kooy R, Riazuddin S, Schenck A, van Bokhoven H, Rooms L. 2013. Homozygous and heterozygous disruptions of AN K3: At the crossroads of neurodevelopmental and psychiatric disorders. Hum Mol Genet 22(10):1960–1970.

Judy JT, Seifuddin F, Pirooznia M, Mahon PB, Bipolar Genome Study C, Jancic D, Goes FS, Schulze T, Cichon S, Noethen M, Rietschel M, Depaulo JR Jr., Potash JB, Zandi PP. 2013. Converging evidence for epistasis between ANK3 and potassium channel gene KCNQ2 in bipolar disorder. Front Genet 4:87.

Kranz TM, Harroch S, Manor O, Lichtenberg P, Friedlander Y, Harkavy-friedman J, Walsh-messinger J, Dolgalev I, Heguy A, Moses V. 2015. De novo mutations from sporadic schizophrenia cases highlight important signaling genes in an independent sample. Schizophr Res 166:119–124.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4(7):1073–1081.

Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, Buja A, Krieger A, Yoon S, Troge J, Rodgers L, Iossifov I, Wigler M. 2011. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron 70(5):886–897.

Li S, Iakoucheva LM, Mooney SD, Radivojac P. 2010. Loss of post-translational modification sites in disease. Pac Symp Biocomput 8:337–347.

Lin AW, Man HY. 2013. Ubiquitination of neurotransmitter receptors and postsynaptic scaffolding proteins. Neural Plast 2013:432057.

Lohmueller KE, Sparsø T, Li Q, Andersson E, Korneliussen T, Albrechtsen A, Banasik K, Grarup N, Hallgrimsdottir I, Kiil K, Kilpeläinen TO, Krarup NT, Pers TH, Sanchez G, Hu Y, DeGiorgio M, Jørgensen T, Sandbæk A, Lauritzen T, Brunak S, Kristiansen K, Li Y, Hansen T, Wang J, Nielsen R, Pedersen O. 2013. Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am J Hum Genet 93:1072–1086.

Lu PW, Soong CJ, Tao M. 1985. Phosphorylation of ankyrin decreases its affinity for spectrin tetramer. J Biol Chem 260(28):14958–14964.

Malenfant P, Liu X, Hudson ML, Qiao Y, Hrynchak M, Riendeau N, Hildebrand MJ, Cohen IL, Chudley AE, Forster-Gibson C, Mickelson EC, Rajcan-Separovic E, Lewis ME, Holden JJ. 2012. Association of GTF2i in the Williams-Beuren syndrome critical region with autism spectrum disorders. J Autism Dev Disord 42(7):1459–1469.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. 2009. Finding the missing heritability of complex diseases. Nature 461(7265):747–753.

McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y, Mistry M, Pavlidis P, Solomon R, Ghiban E, Antoniou E, Kelleher E, O'Brien C, Donohoe G, Gill M, Morris DW, McCombie WR, Corvin A. 2014. De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol Psychiatry 19:652–658.

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfelser R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, Lewis L, Han Y, Voight BF, Lim E, Rossin E, Kirby A, Flannick J, Fromer M, Shakir K, Fennell T, Garimella K, Banks E, Poplin R, Gabriel S, DePristo M, Wimbish JR, Boone BE, Levy SE, Betancur C, Sunyaev S, Boerwinkle E, Buxbaum JD, Cook EH Jr., Devlin B, Gibbs RA, Roeder K, Schellenberg GD, Sutcliffe JS, Daly MJ. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485(7397):242–245.

Need AC, McEvoy JP, Gennarelli M, Heinzen EL, Ge D, Maia JM, Shianna KV, He M, Cirulli ET, Gumbs CE, Zhao Q, Campbell CR, Hong L, Rosenquist P, Putkonen A, Hallikainen T, Repo-Tiihonen E, Tiihonen J, Levy DL, Meltzer HY, Goldstein DB. 2012. Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. Am J Hum Genet 91:303–312.

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43(6):585–589.

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, Munson J, Hiatt JB, Turner EH, Levy R, O'Day DR, Krumm N, Coe BP, Martin BK, Borenstein E, Nickerson DA, Mefford HC, Doherty D, Akey JM, Bernier R, Eichler EE, Shendure J. 2012a. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 338(6114):1619–1622.

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE. 2012b. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485(7397):246–250.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. 2003. Predicting intrinsic disorder from amino acid sequence. Proteins 53(Suppl 6):566–572.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. 2005. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 61(Suppl 7):176–182.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bolte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI Jr., Paterson AD,

Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466(7304):368–372.

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. 2007. Intrinsic disorder and functional proteomics. Biophys J 92:1439–1456.

Radivojac P, Baenziger PH, Kann MG, Mort ME, Hahn MW, Mooney SD. 2008. Gain and loss of phosphorylation sites in human cancer. Bioinformatics 24(16):241–247.

Rechsteiner M, Rogers SW. 1996. PEST sequences and regulation by proteolysis. Trends Biochem Sci 21(7):267–271.

Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. 2008. CSS-Palm 2.0: An updated software for palmitoylation sites prediction. Protein Eng Des Sel 21(11):639–644.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. 2001. Sequence complexity of disordered protein. Proteins 42(1):38–48.

Ronald A, Hoekstra RA. 2011. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am J Med Gen B Neuropsychiatr Genet 156B(3):255–274.

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PB, Choi M, Crawford EL, Davis L, Wright NR, Dhodapkar RM, DiCola M, DiLullo NM, Fernandez TV, Fielding-Singh V, Fishman DO, Frahm S, Garagaloyan R, Goh GS, Kammela S, Klei L, Lowe JK, Lund SC, McGrew AD, Meyer KA, Moffat WJ, Murdoch JD, O'Roak BJ, Ober GT, Pottenger RS, Raubeson MJ, Song Y, Wang Q, Yaspan BL, Yu TW, Yurkiewicz IR, Beaudet AL, Cantor RM, Curland M, Grice DE, Gunel M, Lifton RP, Mane SM, Martin DM, Shaw CA, Sheldon M, Tischfield JA, Walsh CA, Morrow EM, Ledbetter DH, Fombonne E, Lord C, Martin CL, Brooks AI, Sutcliffe JS, Cook EH Jr., Geschwind D, Roeder K, Devlin B, State MW. 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70(5):863–885.

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Gunel M, Roeder K, Geschwind DH, Devlin B, State MW. 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485(7397):237–241.

Saunders CT, Baker D. 2002. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. J Mol Biol 322(4):891–901.

Schizophrenia Psychiatric Genome-Wide Association Study Consortium. 2011. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 43(10):969–976.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimaki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M. 2007. Strong association of de novo copy number mutations with autism. Science 316(5823):445–449.

Shi L, Zhang X, Golhar R, Otieno FG, He M, Hou C, Kim C, Keating B, Lyon GJ, Wang K, Hakonarson H. 2013. Whole-genome sequencing in an autism multiplex family. Mol Autism 4(1):8.

Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK. 2007. DisProt: The database of disordered proteins. Nucleic Acids Res 35(Database issue):D786–D793.

Singh GP, Ganapathi M, Sandhu KS, Dash D. 2006. Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. Proteins 62(2):309–315.

Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, Helgadottir HT, Johannsdottir H, Magnusson OT, Gudjonsson SA, Justesen JM, Harder MN, Jørgensen ME, Christensen C, Brandslund I, Sandbæk A, Lauritzen T, Vestergaard H, Linneberg A, Jørgensen T, Hansen T, Daneshpour MS, Fallah M-S, Hreidarsson AB, Sigurdsson G, Azizi F, Benediktsson R, Masson G, Helgason A, Kong A, Gudbjartsson DF, Pedersen O, Thorsteinsdottir U, Stefansson K. 2014. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet 46:294–298.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE 6(7):e21800.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEG A6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725–2729.

Timms AE, Dorschner MO, Wechsler J, Choi KY, Kirkwood R, Girirajan S, Baker C, Eichler EE, Korvatska O, Roche KW, Horwitz MS, Tsuang DW. 2013. Support for the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. JAMA Psychiatry 70:582–590.

Toma C, Torrico B, Hervas A, Valdes-Mas R, Tristan-Noguero A, Padillo V, Maristany M, Salgado M, Arenas C, Puente XS, Bayes M, Cormand B. 2013. Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. Mol Psychiatry 19(7):784–790.

Torkamani A, Kannan N, Taylor SS, Schork NJ. 2008. Congenital disease SNPs target lineage specific structural elements in protein kinases. Proc Natl Acad Sci USA 105(26):9011–9016.

Tung C-W, Ho S-Y. 2008. Computational identification of ubiquitylation sites from protein sequences. BMC Bioinformatics 9:310.

Uversky VN, Oldfield CJ, Dunker AK. 2005. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recogn 18(5):343–384.

Uversky VN, Oldfield CJ, Dunker AK. 2008. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. Ann Rev Biophys 37:215–246.

Uyar B, Weatheritt RJ, Dinkel H, Davey NE, Gibson TJ. 2014. Proteome-wide analysis of human disease mutations in short linear motifs: Neglected players in cancer? Mol Biosyst 10(10):2626–2642.

Vacic V, Iakoucheva LM. 2012. Disease mutations in disordered regions-exception to the rule? Mol BioSyst 8(1):27–32.

Vacic V, Markwick PR, Oldfield CJ, Zhao X, Haynes C, Uversky VN, Iakoucheva LM. 2012. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. PLoS Comput Biol 8(10):e1002709.

Wang J, Duncan D, Shi Z, Zhang B. 2013. Web-based gene set analysis toolkit (WebGestalt): Update 2013. Nucleic Acids Res 41(W1):W77–W83.

Wang X, Goode EL, Fredericksen ZS, Vierkant RA, Pankratz VS, Liu-Mares W, Rider DN, Vachon CM, Cerhan JR, Olson JE, Couch FJ. 2008. Association of genetic variation in genes implicated in the beta-catenin destruction complex with risk of breast cancer. Cancer Epidemiol Biomarkers Prev 17(8):2101–2108.

Weiss LA, Escayg A, Kearney JA, Trudeau M, MacDonald BT, Mori M, Reichert J, Buxbaum JD, Meisler MH. 2003. Sodium channels SCN1A, SCN2A and SCN3A in familial autism. Mol Psychiatry 8(2):186–194.

Xu B, Ionita-laza I, Roos JL, Boone B, Woodrick S, Levy S, Gogos JA, Karayiorgou M. 2012a. De novo gene mutations highlight patterns of

genetic and neural complexity in schizophrenia. Nat Genet 44:1365–1369.

Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Gogos J a, Karayiorgou M. 2012b. Exomae sequencing supports a de novo mutational paradigm for schizophrenia. Nat Genet 43:864–868.

Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. 2010. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 1804(4):996–1010.

Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, Schmitz-Abe K, Harmin DA, Adli M, Malik AN, D'Gama AM, Lim ET, Sanders SJ, Mochida GH, Partlow JN, Sunu CM, Felie JM, Rodriguez J, Nasir RH, Ware J, Joseph RM, Hill RS, Kwan BY, Al-Saffar M, Mukaddes NM, Hashmi A, Balkhy S, Gascon GG, Hisama FM, LeClair E, Poduri A, Oner O, Al-Saad S, Al-Awadi SA, Bastaki L, Ben-Omran T, Teebi AS, Al-Gazali L, Eapen V, Stevens CR, Rappaport L, Gabriel SB, Markianos K, State MW, Greenberg ME, Taniguchi H, Braverman NE, Morrow EM, Walsh CA. 2013. Using whole-exome sequencing to identify inherited causes of autism. Neuron 77(2):259–273.

Zhang B, Kirov S, Snoddy J. 2005. WebGestalt: An integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33(Suppl 2):W741–W748.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.