Molecular Autism

RESEARCH **Open Access**

Identification of rare X-linked neuroligin variants by massively parallel sequencing in males with autism spectrum disorder

Karyn Meltz Steinberg[1,2,3†], Dhanya Ramachandran[1†], Viren C Patel[1], Amol C Shetty[1], David J Cutler[1] and Michael E Zwick[1*]

Abstract

Background: Autism spectrum disorder (ASD) is highly heritable, but the genetic risk factors for it remain largely unknown. Although structural variants with large effect sizes may explain up to 15% ASD, genome-wide association studies have failed to uncover common single nucleotide variants with large effects on phenotype. The focus within ASD genetics is now shifting to the examination of rare sequence variants of modest effect, which is most often achieved via exome selection and sequencing. This strategy has indeed identified some rare candidate variants; however, the approach does not capture the full spectrum of genetic variation that might contribute to the phenotype.

Methods: We surveyed two loci with known rare variants that contribute to ASD, the X-linked neuroligin genes by performing massively parallel Illumina sequencing of the coding and noncoding regions from these genes in males from families with multiplex autism. We annotated all variant sites and functionally tested a subset to identify other rare mutations contributing to ASD susceptibility.

Results: We found seven rare variants at evolutionary conserved sites in our study population. Functional analyses of the three 3' UTR variants did not show statistically significant effects on the expression of *NLGN3* and *NLGN4X*. In addition, we identified two *NLGN3* intronic variants located within conserved transcription factor binding sites that could potentially affect gene regulation.

Conclusions: These data demonstrate the power of massively parallel, targeted sequencing studies of affected individuals for identifying rare, potentially disease-contributing variation. However, they also point out the challenges and limitations of current methods of direct functional testing of rare variants and the difficulties of identifying alleles with modest effects.

Keywords: Autism spectrum disorder, Massively parallel DNA sequencing, Rare variation, Evolutionary conservation

Background

The rapid development of better methods of targeted enrichment and genome sequencing has made it possible to detect a more complete spectrum of genetic variation [1-3]. These approaches hold out the hope of uncovering the genetic basis of polygenic complex human diseases, including autism (OMIM 209850), a childhood-onset disorder characterized by impaired social interactions, abnormal verbal communication, restricted interests, and repetitive behaviors. Autism has an estimated prevalence of one percent [4,5], and one of its most striking epidemiological features is a four-fold excess of affected male individuals.

Autism, or the broader autism spectrum disorder (ASD) phenotype, is an example of a highly heterogenous, multifactorial disorder with substantial heritability [6-13], (see reviews in [14,15]). Recent reports, in which X-chromosome coding exons in individuals with ASD were sequenced, identified an excess of rare mutations predicted to be damaging

* Correspondence: mzwick@emory.edu
†Equal contributors
[1]Department of Human Genetics, Emory University School of Medicine, Whitehead Biomedical Research Building, Suite 301, Atlanta 30322GA, USA
Full list of author information is available at the end of the article

in a variety of genes related to synaptic function [16,17]. To date, more than 100 different genes and genomic regions have been linked to this complex trait (see reviews in [18,19]). Despite these findings, however, most of the genetic risk for ASD remains unexplained. Recent studies of ASD genetics generally adopt one of three study designs. The first employs genome-wide association studies, which have identified a few loci of interest, but largely failed to replicate findings between studies [7,20,21]; a meta-analysis of these studies, with a total of over 2500 study subjects, reveals it is extremely unlikely that there is any common variant influencing autism susceptibility, with an odds ratio of greater than 1.5 [22]. The second design focuses on large but very rare (frequency usually less than one in a thousand in the general population) *de novo* and inherited copy number variants (CNVs). Numerous studies have now shown convincingly that this class of rare variation makes a significant contribution to autism susceptibility [23-34], explaining up to 15% of all ASD cases. Unfortunately, these studies also point to a highly heterogenous allelic architecture, as no single risk variant is found in more than 1% of surveyed cases. The third has applied exome sequencing to identify *de novo* or inherited variants that contribute to ASD [35-37]. Overall, although genetic studies have uncovered many candidate loci, much ASD heritability remains unexplained [15].

Neuroligin pathway genes, including the neuroligins, neurexins, and *SHANK* genes, are critical to synapse development and function [38-40]. Several rare mutations in neuroligin genes, including single nucleotide variants (SNVs), insertions, splice variants, and deletions of whole exons, have been implicated in the pathogenesis of ASD [41-49]. These mutations often segregate with ASD in families [41,42]; however, they are also associated with variable cognitive phenotypes, including intellectual disability (ID) [42,50], Tourette syndrome [46], and language disability [43]. Neuroligin's binding partners, the neurexins, are also critical to synaptic function, and point mutations and copy number variants in neurexin genes have been linked to ASD [27,51-54]. In addition, variants in the *SHANK* genes that anchor the neuroligins in the post-synaptic density, are also implicated in ASD [29,55-58]. There is therefore substantial evidence that perturbations of genes from the neuroligin pathway contribute to ASD susceptibility.

Most prior studies of neuroligin pathway genes have focused primarily on either discovering variants located within exons or large CNVs that disrupt the locus (but see [59]). Here we sought to test the hypothesis that rare variants found at evolutionarily conserved sites within noncoding regions might act as ASD susceptibility alleles. Although the statistical power to test any single rare variant is low [60], direct functional testing and functional annotation of variant sites might reveal alleles with modest

effects on ASD susceptibility. Based on the substantial male bias in autism prevalence and the fact that two of the suspected neuroligin pathway genes are on the X-chromosome, we performed comprehensive sequencing of the *NLGN3* (Xp13.1) and *NLGN4X* (Xp22.3) loci, including the coding exons, 5'UTR, 3'UTR, and flanking intronic and intergenic sequences in 144 male individuals with ASD obtained from the Autism Genetic Resource Exchange (AGRE) repository. Our motivation for this design arises from the fact that male individuals are hemizygous for the X-chromosome, while female individuals are heterozygous. Thus, recessive acting alleles would be expressed exclusively in male individuals and could therefore increase the prevalence of male ASD. The results of analysis of our sequence data identified a set of rare noncoding alleles at highly evolutionarily conserved sites that were worthy of further evaluation for their role in ASD susceptibility.

Methods

Selection of male individuals with ASD

We sequenced 144 male individuals from the AGRE multiplex collection [61]. Detailed diagnostic criteria can be found on the AGRE website [62]. Male individuals were chosen from families with two or more male affected sibpairs (ASPs) that either shared identical X-chromosome markers, DXS9895 and DXS9902, or shared > 98% of 52 genotyped single nucleotide polymorphisms (SNPs) in the Xp22.3 region. A total of 152 families fit these criteria. One individual was randomly chosen for sequencing if both affected siblings were equally affected; if they were not equally affected, those with autism, those classified as not quite autism (NQA), or those classified as broad-spectrum were chosen, in that order, to maintain consistency. Among the 152 samples, two were unavailable from the AGRE repository at the time of this experiment and six had global PCR failure. Thus we had a total of 144 samples for processing. All human samples used in our study were de-identified and obtained from the AGRE repository, which obtains consent to participate in research studies and publish findings. Our study was reviewed and approved by the Emory University Institutional Review Board (IRB) because it met the criteria for exemption under 45 CFR 46.101(b).

Target DNA amplification for the *NLGN3* and *NLGN4X* loci

Long-range PCR (LPCR) primers for amplifying the target DNA sequence were designed using EmPrime [63]. All primers were obtained from Invitrogen (Carlsbad, CA, USA) list of all primers used in this experiment can be found in Additional file 1. 500 ng of sample DNA was added to 1X LA Taq buffer (TaKaRa Bio Inc., Otsu Shigh, JP), 250 μM dNTP Mix (TaKaRa Bio Inc., Otsu Shigh, JP), 400 nM of

both forward and reverse LPCR primers, and 0.1 U/µl of LA Taq (TaKaRa Bio Inc., Otsu Shigh, JP). If the amplicon had a high GC content, we used 1X GC Buffer (TaKaRa Bio Inc., Otsu Shigh, JP) in place of 1X LA Taq buffer. PCR was performed using the following parameters: initial denaturation at 94°C for 2 minutes, 29 cycles of 94°C for 10 seconds, 68°C for 1 minute per kilobase (of amplicon size), and a final extension time of 5 minutes.

Amplification was confirmed using 1% agarose 96-well E-Gels (Invitrogen, Carlsbad, CA, USA). We determined the concentration of each amplicon using PicoGreen dsDNA Quantitation Kits (Invitrogen, Carlsbad, CA, USA) and the Tecan Ultra Evolution plate reader. An equimolar concentration of each fragment was then pooled by sample for a total DNA concentration per sample of 10 ug. Pooled amplicons were then purified using the Invitrogen PureLink PCR Purification Kit with the HC buffer.

Sample preparation for Illumina sequencing
Pooled, purified samples were sheared to approximately 300 bp using the Covaris E210, and fragmentation was confirmed by running Agilent Bioanalyzer DNA 7500 chips (Agilent Technologies, Santa Clara, CA, USA). We performed end repair using the NEBNext DNA Sample Prep Reagent Set 1 (New England BioLabs, Ipswich, MA, USA) with 0.4 mM dNTP mix, 5 ul of T4 DNA Polymerase, 1 ul of DNA Polymerase I (Klenow) fragment, 5 ul of T4 Polynucleotide Kinase, and 1X T4 DNA ligase buffer. The reactions were incubated in a thermal cycler for 30 minutes at 20°C. Following incubation, the reactions were purified using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA). To the purified, blunt, phosphorylated DNA fragments, we added 1X NEB Buffer 2, 1 mM dATP (NEB, Ipswich, MA, USA), and 3 ul of Klenow fragment (NEBNext Set 1). Following a 30-minute incubation at 37°C, reactions were purified using a QIAquick MinElute Kit (Qiagen, Valencia, CA, USA). To the DNA we added 1X Quick Ligation Buffer (NEBNext Set 1), 10 ul of Index PE Adapter Oligo Mix (from the Multiplexing Sample Preparation Kit; Illumina), and 5 ul of Quick T4 DNA Ligase. The reactions were incubated for 15 minutes at room temperature, and then purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA). This protocol uses a 10:1 molar Adapter:DNA ratio based on the starting concentration of DNA. We used the Size Select 2% E-Gels (Invitrogen, Carlsbad, CA, USA) to remove all unligated adapters and to accurately select the 300-bp band. The 300-bp band was successfully removed, and then selectively enriched using PCR to amplify the amount of DNA in the library and attach the 6-base index tag into the adapter. To 10 ul of DNA we added 1X Phusion PCR Master Mix (Finnzymes, Thermo

Scientific, Lafayette, Co, USA; NEBNext Set 1), 1 ul each of PCR Primer lnPE 1.0 and PCR Primer lnPE 2.0, and 1 ul of PCR Primer Index (from Mulitplexing Sample Preparation Kit; Illumina). PCR parameters were as follows for 30 cycles: 98°C for 30 seconds, 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 30 seconds, with a final extension time of 5 minutes at 72°C. Following incubation, samples were purified using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA), and enrichment was confirmed using the Agilent BioAnalyzer 7500 DNA chip. Four pM of enriched DNA was used for cluster generation and paired-end sequencing on the Illumina Genome Analyzer II (IGA)

Analysis of Illumina sequence data
Raw base-calling data generated by IGA were used as input for mapping and alignment. Paired-end reads were mapped and variants were called relative to a reference sequence using PEMapper (Cutler DJ *et al.*, in revision). Briefly, the PEMapper is composed of four interconnected programs. The first program prepared a hashed index of the target sequence, the second program generated a list of potential mapping locations for each read. In the third stage, a Smith-Waterman alignment was performed at each potential location to determine the optimal position and alignment score. The output of the third stage, consisting of the pileup statistics of each base (number of reads where each nucleotide (A, C, G, T) was seen, together with the number of times that each base appeared deleted or an insertion immediately following the base) was used to make the genotype calls.

In total, 99.7% of target bases had at least 8X coverage, with a median depth of coverage of 452. SNVs and small insertions and deletions (indels) were annotated using the Sequence Annotator (SeqAnt) [64]. Functional annotation from hg18 included the genomic position, amino acid change, presence or absence in dbSNP132, and conservation scores (PhastCons, PhyloP) for each variant base. Additional filtering using dbSNP135 was carried out using the Feb. 2009, GRCh37(hg19) assembly from the UCSC Genome Browser [65]. The SNVs at highly conserved sites had coverages of 198 to 1,354, with the user base (non-reference allele) being called in > 92% in the sequence reads at the corresponding variant sites. A list of all SNVs and indels are contained in Additional files 2 and 3, respectively. As a comparison, we downloaded 3'UTR variants in *NLGN3* and *NLGN4X* from 1,094 individuals sequenced and deposited into the 1000 Genomes database [66]. A total of 49 3'UTR variants (38 SNVs, 11 indels) were identified in the *NLGN3* and *NLGN4X* genes (see Additional file 4).

We used popgen_fasta2.0.c code to perform population genetic analyses. This code calculated Watterson's estimator of the population mutation rate (Θw per site) as well

as a point estimate for Tajima's D as previously described [67]. Variants at highly conserved sites were validated independently by Sanger sequencing (Agencourt Bioscience, MA). PCR primers for validation were designed using the Primer 3 [68]. Additionally, we sequenced the mothers and affected and unaffected male siblings with the validated UTR variants to verify the segregation pattern with autism (see Additional file 5). We also sequenced the mothers and two affected male siblings for two rare novel intronic variants that fell within transcription factor binding sites to verify the segregation pattern with autism (see Additional file 6).

Control genotyping

Control samples used for genotyping were from male adults of European descent who had been screened to rule out psychiatric disorders, and were obtained from the National Institute of Mental Health (NIMH) Human Genetics Initiative [69]. Genotyping was performed by the iPLEX Gold Method (Sequenom, San Diego, CA, USA) per the manufacturer's instructions, using primers designed with the Sequenom Assay design 3.1 software (see Additional file 7). A positive control was included in each plate to confirm the sensitivity of the assay.

Functional testing of 3'UTR variants in a luciferase assay

Luciferase assays were performed to check whether the novel UTR variants we identified had altered gene expression relative to a construct containing the reference sequence. Full-length UTR sequences were amplified for three rare variants in *NLGN3* (70306922 (C > T), 70306764 (A > G), and 70306767 (C > G), and two rare variants in *NLGN4X* (5818136 (A > G), 5820149/50 (CT > −−)). The amplified sequences were cloned in to the multiple cloning site, downstream of the luciferase (*luc2* +) gene in the pmirGLO expression vector (Promega, Madison, WI, USA). A full-length 3'UTR sequence amplified from an unaffected normal control sample was cloned in to the same vector as the wild type. The *NLGN3* variant (70306764/67) served as the control for non-conserved UTR variant. The presence of the novel variant site was confirmed by Sanger sequencing.

Cell culture, transfection, and luciferase assays were performed on two different cell lines, mouse Neuro2a and human embryonic kidney 293 (HEK293), following the manufacturer's instructions as reported previously with minor modifications described below [70]. In short, HEK293 cells and Neuro2a cells were cultured at 37°C with 5% CO_2 in DMEM and RPMI 1640 (Cellgro Mediatech, Manassas, VA, USA) respectively, supplemented with 10% fetal bovine serum (Cellgro Mediatech, Manassas, VA, USA). Twenty-four hours before transfection, $0.2*10^6$ cells were plated in each well of 48-well cell culture dishes. Transfections were carried out using Lipofectamine™ 2000 in Opti-MEM (Invitrogen, Carlsbad, CA, USA) using 500 ng of plasmid. Just before each transfection, the old media were replaced with fresh media (DMEM or RPMI supplemented with 10% FBS). Twenty-four hours post transfection, cells were lysed with 250 ul of Passive Lysis Buffer (Promega, Madison, WI, USA), and cell debris were removed by centrifugation at 14,000 rpm for 5 minutes at 4°C. From each lysate, 20 µl of the cell extract were transferred into a luminometer tube, and 100 µl of Dual Luciferase Reporter Assay reagent (Promega, Madison, WI, USA) was added in each well. A manual luminometer (TD-20/20, Promega, Madison, WI, USA) was used to measure the luminescence over a 10-second period, with a delay time of 2 seconds. The luminometer reading was repeated after adding 100 ul of Stop and Glo reagent. For each lysate, the firefly luciferase activity was normalized to Renilla luciferase activity. We performed one independent transfection for each of the three 3'UTR alleles in two different cell lines (mouse Neuro2a and HEK293). Each transfection was replicated three times. A two-tailed, unequal variance Student's *t*-test was performed to determine whether constructs with 3'UTR variants showed altered gene expression compared to constructs with the reference sequence.

Functional annotation of intronic variants

Annotation of the variants was based on hg build 18 of the UCSC Genome Browser. Information regarding the Enhancer- and Promoter-Associated Histone Mark (H3K4me1 and H3K4me3) and the Transcription Factor Binding ChIP Seq were obtained from ENCODE Integrated Regulation tracks. Nuclease accessible site (NAS) information was obtained from the EIO/JCVI NAS Track, which annotates the location of NAS in the genome of human CD34+ and CD34- cells by NA-Seq technology. Conserved transcription factor binding sites (TFBS) were from human/mouse/rat (HMR) conserved TFBS track and were identified by searching within human-mouse-rat alignments using the position weight matrices (PWMs) from the TRANSFAC Matrix database (v7.0) [71]. The final z score can be interpreted as the number of standard deviations above the mean raw score for that binding matrix across the upstream regions of all RefSeq genes. The conserved transcription factor binding motif was displayed as a sequence logo [72] obtained at the Sequence Logo website [73].

Results

We sequenced the *NLGN3* and *NLGN4X* loci in a sample of 144 male individuals with a diagnosis of autism; all the patient samples were obtained from the multiplex AGRE repository. We identified a total of 208 sites of variation, with 176 SNVs (see Additional file 2), and 32 indels (see Additional file 3). Overall levels of variation

were estimated at 5.8×10^{-4} (Θ_w per site [74]), with an excess of rare variants as evidenced by a negative value for the Tajima's D test statistic (−0.27) [75]. For the SNVs, a total of 37 (21%), had not been reported before (18 in *NLGN3* and 19 in *NLGN4X*). For the indels, a total of 22 (69%), had not been reported before (5 in NLGN3 and 17 in *NLGN4X*). As summarized in Figure 1, almost all common variation (> 5% frequency in our sample) is contained in dbSNP, whereas most rare variants (< 5%) have not been cataloged in dbSNP.

Our study focused on previously undiscovered variation found at sites with elevated evolutionary conservation, so we did not follow up the 139 variants included in dbSNP. The only missense mutation we saw (*NLGN4X*, 5821532 G > A) had been reported before and, because of a nearby compensatory mutation, was not predicted to alter the primary structure of the protein [76]. Our data provide further evidence that coding sequence mutations at NLGN3 or NLGN4X that cause autism are very rare. Assuming the number of disease-causing coding mutations is Poisson distributed, we are 99% confident that the combined frequency of disease causing coding mutations at *NLGN3* and *NLGN4X* is less than 3% (no observations in 144 tries). Functional annotation of the remaining variant sites revealed that six SNVs and one indel were located at sites with elevated evolutionary conservation (PhastCons > 0.7, Table 1).

All of the rare variants were observed at a frequency of less than 1% in our ASD cases. To arrive at a better estimate of their population frequency, we genotyped six of the variants in a collection of 1,450 unaffected male controls obtained from the NIMH Human Genetics Initiative (Table 1). All of the variants genotyped had a frequency of less than 0.002. Thus, these data suggest that the variants we found are very rare in the general population.

Functional analysis of 3' UTR variants

Rare noncoding variants could act as autism susceptibility alleles by altering the level of expression of either *NLGN3* or *NLGN4X*. We sought to determine whether any of the three highly conserved 3'UTR variants of *NLGN3* (chrX:70306922) and *NLGN4X* (chrX:5818136, 5820149−50) could potentially lead to altered neuroligin expression in a luciferase reporter gene assay (Table 1). In addition to a construct containing the reference sequence, we also checked the expression of a construct containing two rare 3' UTR *NLGN3* variants from a single individual (chrX:70306764/67) that were not located at evolutionary conserved sites, as an internal control. Each construct was tested in both mouse Neuro2a and human embryonic kidney 293 (HEK293) cells. The construct bearing the 3'UTR *NLGN3* variant (chrX:70306922) showed a trend for reduced luciferase activity the Neuro2a ($P < 0.10$) cells compared to the construct with the reference sequence (see Additional file 8). However, this result was not statistically significant and the average reduction (approximately 9%) was modest. Furthermore, the control construct showed a similar trend in the Neuro2a cells ($P < 0.22$). Neither construct showed a significant difference in the HEK293 cells (see Additional file 8). Inheritance of the 3'UTR *NLGN3* variants did not segregate with
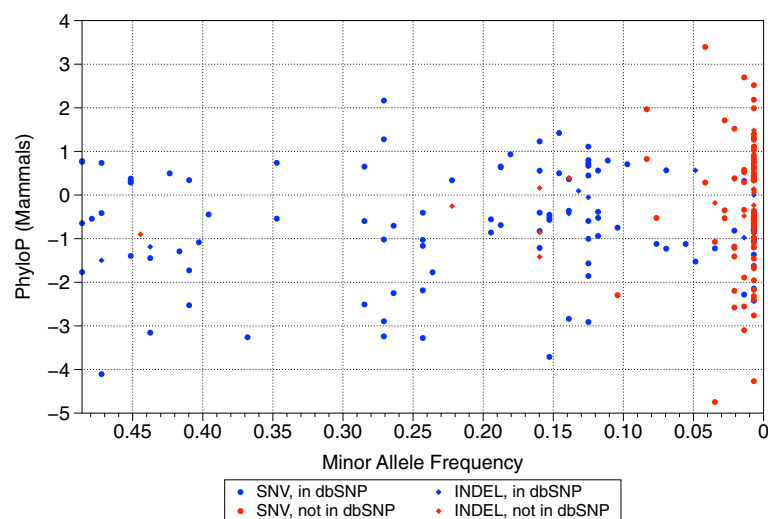

Figure 1 Summary of single nucleotide variants (SNV) and insertions or deletions (indel) variation discovered at the *NLGN3* and *NLGN4X* loci in male individuals with autism spectrum disorder (ASD). The frequency of SNVs and indels (minor alleles) in cases is plotted against their level of evolutionary conservation. Most common variation has already been discovered and exists in public databases (blue; circles and diamonds). Most of the rare variation at both loci was discovered in our study and not contained in public databases (red; circles and diamonds).

Table 1 Summary of variants found at sites of increased evolutionary conservation

Functional class	Position (hg18)	Variant type	PhastCons	PhyloP	Frequency in ASD cases	Frequency in NIMH control samples
3'UTR	5818136	SNV	1	2.19	1/144	0/1440
3'UTR	5820149-50	Indel	1	0.85/1.76	1/144	ND
Intron	70284973	SNV	1	1.99	1/144	1/1416
Intron	70285256	SNV	0.98	0.90	1/144	0/1441
Intron	70288838	SNV	0.69	1.03	1/144	2/1441
Intron	70291656	SNV	1	2.52	1/144	0/1440
3'UTR	70306922	SNV	0.73	−0.47	1/144	0/1440

ASD, autism spectrum disorder; NIMH, National Institute of Mental Health; ND, Not Determined.

autism as shown in Additional file 5. None arose as *de novo* events in the ASD cases.

The 3'UTR *NLGN4X* variants did segregate with autism as shown in Additional file 5. None arose as *de novo* events in the ASD cases. A construct with the 3'UTR *NLGN4X* SNV (chrX:5818136) suggested a modest trend for increased luciferase activity in both the Neuro2a ($P < 0.27$) and HEK293 ($P < 0.23$) cells. However, the difference in expression was not statistically significant in either case (Additional file 8). The construct with the 3'UTR *NLGN4X* INDEL was not significant in either cell type (Additional file 8).

Analysis of intronic variants

We next sought to determine whether any of the four rare, intronic variants in *NLGN3* could act as autism susceptibility alleles. If so, we would predict that these variants should fall in regions identified as functional by the ENCODE Project [77]. All of the intronic variants are located within regions of enriched H3K4Me1 markers in H1 ES, HMEC, and K562 cells (Figure 2A). Regions with the mono-methylation of histone H3 lysine 4 are suggestive of enhancer and/or promotor activity due to the epigenetic modification of histone proteins [78].

One of the variants (chrX:70291656) was found to be located within the NAS of CD34- cells, and there are no common intronic variants located nearby. NAS are loci that are free of nucleosomes and are therefore hypothesized to allow *cis*-acting DNA to interact with trans-acting factors [79]. The same variant (chrX:70291656) also falls within a HMR conserved Bach1 TFBS (*z*-score 2.68, $P < 0.004$) (Figure 2B). Bach1 is a member of the BTB-basic leucine zipper transcription factor family and is a mammalian repressor of heme oxygenase 1 (HO-1) [80]. The intronic *NLGN3* variant chrX:70291656 found at the most highly evolutionary conserved site did segregate with autism as shown in Additional file 6.

An additional variant (chrX:70284973) falls within an HMR conserved Roaz TFBS (*z*-score 2.86, $P < 0.003$; Figure 2C). Roaz is a zinc finger protein that impairs the ability of the Olf-1/EBF transcription factor family to activate olfactory neuron-specific promotors [81]. This variant was found in one case and one control. There are no SNPs or repetitive elements in either of the regions of TFBS; the closest SNP annotated in dbSNP135 is located >150 bp upstream or downstream, and repetitive elements are >1 kbp from either variant. The intronic *NLGN3* variant chrX:70284973 also segregates with autism as shown in Additional file 6.

Discussion

For the past 15 years, genomic studies of complex diseases have relied on a model in which common genetic variation contributes significantly to common diseases [82-84]. Based on this model, the systematic genotyping of common variants was perceived as the best way to begin characterizing the allelic architecture of complex human traits [85]. To make such experiments possible required the development of highly accurate, low-cost, high-throughput genotyping platforms and a catalog of common human genetic variation like the HapMap project [86,87]. Furthermore, because direct sequencing was not a viable strategy, assessing the role of common variation was really the only feasible genome-wide experiment. Thus, until recently the contribution of rare coding and noncoding variation to complex disorders like autism has gone largely unexplored.

While most quantitative traits, including human diseases, show substantial heritability in most populations, their allelic architecture remains poorly understood [88,89]. Haldane in the 1920s was the first to recognize that deleterious alleles of large effect will be maintained only at very low frequencies in the general population [90]. Copy number variation studies of ASD have identified variants with a large effect size, having odds ratios (ORs) often greater than 5.0 [23-34]. Much as Haldane would have predicted, these variants are quite rare, often occurring much less often than one in a thousand in the general population, a frequency generally consistent with a large effect locus at mutation selection balance.

At the same time, genome-wide association studies have shown that common variants with large effects are unlikely to exist in the human population for many disorders, although a large number of loci with alleles with
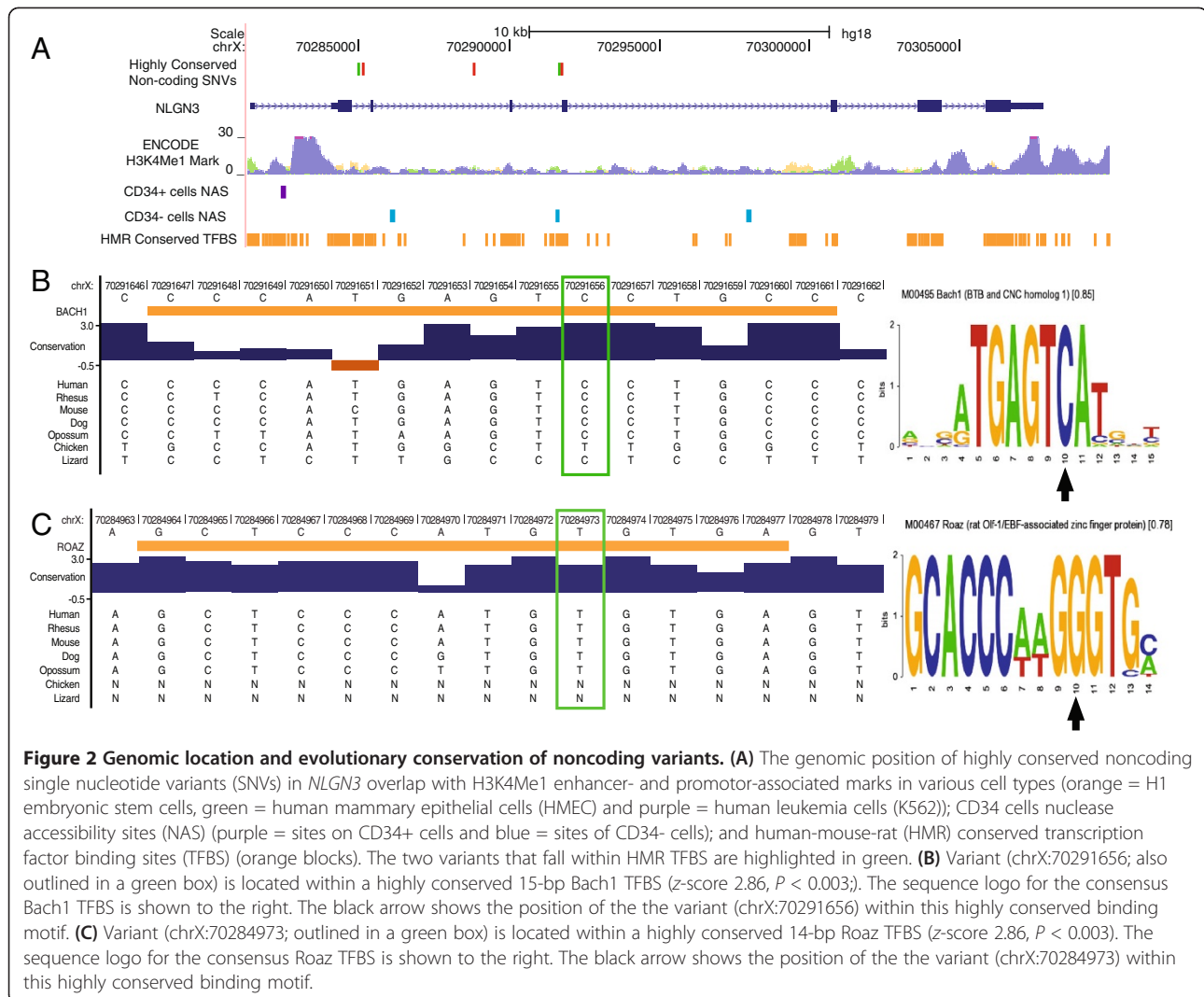
Figure 2 Genomic location and evolutionary conservation of noncoding variants. (A) The genomic position of highly conserved noncoding single nucleotide variants (SNVs) in *NLGN3* overlap with H3K4Me1 enhancer- and promotor-associated marks in various cell types (orange = H1 embryonic stem cells, green = human mammary epithelial cells (HMEC) and purple = human leukemia cells (K562)); CD34 cells nuclease accessibility sites (NAS) (purple = sites on CD34+ cells and blue = sites of CD34- cells); and human-mouse-rat (HMR) conserved transcription factor binding sites (TFBS) (orange blocks). The two variants that fall within HMR TFBS are highlighted in green. **(B)** Variant (chrX:70291656; also outlined in a green box) is located within a highly conserved 15-bp Bach1 TFBS (*z*-score 2.86, *P* < 0.003;). The sequence logo for the consensus Bach1 TFBS is shown to the right. The black arrow shows the position of the the variant (chrX:70291656) within this highly conserved binding motif. **(C)** Variant (chrX:70284973; outlined in a green box) is located within a highly conserved 14-bp Roaz TFBS (*z*-score 2.86, *P* < 0.003). The sequence logo for the consensus Roaz TFBS is shown to the right. The black arrow shows the position of the the variant (chrX:70284973) within this highly conserved binding motif.

much smaller ORs (< 1.2) remains plausible (see review in [91]). This is borne out in ASD, as genome-wide association studies have identified just a few loci of interest, which have largely failed to replicate findings between studies [7,20,21], whereas a meta-analysis suggests it is extremely unlikely that any common variant influences autism susceptibility with an OR of greater than 1.5 [15,22].

Here we used targeted, massively parallel sequencing of two X-linked genes, previously shown to harbor very rare point mutations causing ASD, to explore whether they might also have rare noncoding variants at evolutionary conserved sites that act as ASD susceptibility alleles. Using this approach we found a set of seven candidate variants, including three located in the 3'UTR, in the two genes examined among the 144 individuals sequenced (Table 1). As a comparison, a search for similar variants at highly conserved sites among 1,094 individuals sequenced and deposited into the 1,000 Genomes database identified a total of 49 3'UTR variants (38 SNVs, 11 indels) identified in *NLGN3* and *NLGN4X* genes (Additional file 4). None of the indels were found in highly conserved regions. A total of seven SNVs were found at highly conserved sites (PhastCons > 0.7), and two of the variants had an estimated minor allele frequency of 0.001. The remaining five variants did not have an estimated minor allele frequency. In considering this comparison, it is important to note that because we sequenced our samples to a far greater depth as compared to the 1,000 Genomes samples, our study had a greater probability of detecting rare variation.

Functional analysis of the 3' UTR variants in a luciferase assay did not show a statistically significant difference in their expression (Additional file 8). The most likely interpretation is that these variants do not influence the risk of autism in these probands. However, two points are worth noting. First, under a quantitative genetic model of autism,

we would not expect to find noncoding variants with large effects (that is, monogenic causes of autism), and instead might expect to find many alleles at many different loci, each with modest effects [15,92]. Second, our functional assays may be imperfect or insufficiently sensitive to reveal how these variants might act on their respective genes. Collectively these results point out the challenges of functional validation of alleles with modest effect sizes, even though the great heterogeneity of autism implies that such alleles should exist.

Our most promising intronic variant (chrX:70291656) is located in a highly conserved site in a TFBS that has been associated with neuronal dysfunction (Figure 2B). The Bach1 transcription factor protects cells from damage by activating HO-1. Bach1 dysregulation has been associated with Down syndrome (DS): Bach1 is significantly overexpressed in the fetal cortex of DS fetuses when compared to controls [93], whereas in another study, expression was significantly reduced in the frontal cortex of DS patients. In Bach1 knockout mice, expression of Bach1 mRNA was significantly higher in the olfactory bulb, but lower in the cortex versus wild-type mice, providing another link to olfaction [94]. It is possible the variant we found within the conserved TFBS influences olfactory neuron development and expression which could contribute to the sensory dysregulation phenotype of ASD. Interestingly, the affected individual harboring this variant in addition to being autistic, is intellectually disabled. He is diagnosed with sensory abnormalities including increased sensitivity towards acoustic and decreased sensitivity towards tactile senses. Still, our data do not demonstrate that this variant is functional through a direct experiment, but do predict that effects ought to be observed in such an experiment (for example, ChIP Seq).

Compared to children without neurodevelopmental disorders, children with ASD demonstrate olfactory and taste dysfunction [95,96]. Notably, in mice the *NLGN3* gene is expressed in all neurons of the olfactory bulb [40]. It is also interesting that we identified an intronic variant (chrX:70284973) that falls within a highly conserved TFBS related to olfactory neuron development (Figure 2C). Interestingly, this variant is predicted to increase binding efficiency at this TFBS. The Roaz transcription factor regulates both the temporal and spatial pattern of olfactory neuronal gene expression by binding to a consensus recognition sequence and modulating transcriptional activity [81,97]. Over 90% of children with ASD report sensory abnormalities, among them visual, auditory, tactile, and olfactory dysregulation (reviewed in [98]).

Our results highlight the importance of targeted sequencing of both coding and noncoding regions of candidate genes for complex, polygenic traits. Genetic studies of the X-chromosome have suggested that both rare and common X-linked variation may

contribute to ASD [16,17,31,99-101], but much remains to be discovered. Although exome sequencing studies are now identifying point mutations, small indels, and *de novo* variants that contribute to ASD [35-37], these studies are limited by the regions they include in their exome capture chips, as well as biases in the capture efficiency of paralogous genes. Due to these constraints, these kinds of studies would have completely missed the noncoding variants we identified here. A study such as ours is also an important follow up for exome studies to assess the complete spectrum of genetic variation in genes known to harbor ASD-contributing mutations. These genes are often in candidate pathways related to neuronal development and function, and identifying mutations in noncoding and regulatory regions will likely shed more light on the etiology of ASD pathogenesis. As ASD is a polygenic trait, noncoding mutations probably play a role in the genetic contribution to ASD, in combination with other forms of genetic variation, including CNVs, coding mutations, and gene-disruptive indels that affect pathways related to brain development [102]. Still, our study points out that functional testing of rare variants remains challenging and not sufficiently high-throughput to perform this experiment on a genome-wide scale, especially when the effect sizes are modest. Finally, as whole-genome sequencing becomes increasingly cost effective and a more feasible experimental paradigm, detailed analyses of both coding and noncoding variation, as we have carried out here, can be expected to uncover ever more genetic variants that contribute to complex disorders like autism. These studies, however, will face significant challenges in direct functional testing of large numbers of these rare variants at highly conserved evolutionary sites.

Conclusions

In conclusion, we used a highly targeted approach to identify rare variants that may contribute to ASD using massively parallel sequencing of the X-linked neuronal cell adhesion genes, *NLGN3* and *NLGN4X*. These data suggest that coding sequence variations in *NLGN3* and *NLGN4X* are rare. We identified three 3'UTR SNVs that did not show statistically significant effects in a luciferase assay. In addition, we uncovered intronic mutations that may affect regulatory regions, such as enhancer- and promotor-associated histone modification sites, NAS and TFBS. We suspect these variants may make modest contributions to ASD pathogenesis, as would be predicted by a quantitative genetic model of autism susceptibility. These data highlight one of the main challenges researchers face in the current era of next generation sequencing technology, namely establishing a direct link between the candidate variants identified and its contribution to the clinical phenotype of complex traits like autism.

Additional files

Additional file 1: Table showing long PCR primers. Forward and reverse primers are listed for each of the regions sequenced.

Additional file 2: Table showing single nucleotide variants detected in 144 males with diagnosis of autism from the Autism Genetic Resource Exchange. Contains variant position, dbSNP ID, functional annotation, and frequency in cases and controls.

Additional file 3: Table showing small insertion and deletion variants detected in 144 males with diagnosis of autism from the Autism Genetic Resource Exchange. Contains variant position, dbSNP ID, functional annotation, and frequency in cases and controls.

Additional file 4: Table showing *NLGN3* and *NLGN4X* 3'UTR genetic variation from the 1000 Genomes database. Contains variant position, dbSNP ID, PhastCons score, and estimated frequency.

Additional file 5: Figure showing segregation analysis of the highly conserved *NLGN3* and *NLGN4X* UTR variants. The segregation of the highly conserved UTR variants in *NLGN3* (chrX:70306922) and *NLGN4X* (chrX:5818136, chrX:5820149–50) with a diagnosis of autism were checked by sequencing the mother and affected and unaffected male siblings of the corresponding proband. The top base shown is the reference base, the bottom base shown is the variant base. Both the *NLGN4X* variants segregated with autism. We also tested segregation of the two control variants (chrX:70306764, chrX:70306767) with autism. DNA was not available for the unaffected sibling of the proband carrying the *NLGN4X* UTR variant (chrX:5820149–150).

Additional file 6: Figure showing segregation analysis of two highly conserved *NLGN3* intronic variants. The segregation of *NLGN3* intronic variants (chrX:70291656) and (chrX:70284973) with a diagnosis of autism. Both are located within transcription factor binding sites (TFBS).

Additional file 7: Table showing sequenom genotyping assay primers for single nucleotide variants discovered in *NLGN3* and *NLGN4X* loci. This table contains detailed information about each primer pair and extension primer.

Additional file 8: Figure showing the results of luciferase assays for one 3'UTR variant in *NLGN3* and two 3'UTR variants in *NLGN4X*. (A). Results of luciferase expression assay in Neuro2a cell lines. These constructs were tested against the reference allele constructs using a luciferase reporter assay. One of the *NLGN3* constructs (chrX:70306764/67) serves as a control for the variant at a non-conserved site. Error bars show 2 standard errors on either side of the mean value. (B). Results of luciferase expression assay in human embryonic kidney (HEK) cell lines. The experiment was similar to that described above.

Abbreviations

AGRE: Autism Genetic Resource Exchange; ASD: Autism Spectrum Disorder; ASP: Affected SibPair; CNV: Copy Number Variant; DMEM: Dulbecco's Modified Eagle's Medium; DS: Down Syndrome; FBS: Fetal Bovine Serum; HEK: Human Embryonic Kidney; HMEC: Human Mammary Epithelial Cells; HMR: Human/Mouse/Rat; HO-1: Heme Oxygenase 1; ID: Intellectual Disability; indels: insertions and deletions; LPCR: Long-range Polymerase Chain Reaction; NAS: Nuclease Accessibility Site; NIMH: National Institute of Mental Health; NQA: Not Quite Autism; OR: Odds Ratio; PWMs: Position Weight Matrices; SNP: Single Nucleotide Polymorphism; SNV: Single Nucleotide Variant; TFBS: Transcription Factor Binding Site.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KMS, DR, and MZ participated in the design of the study. KMS performed the target DNA amplification and Illumina sequencing. KMS and DR performed validation of variant sites. DR performed genotyping and luciferase functional assays. Bioinformatic and statistical analyses were conducted by KMS, VCP, ACS, DJC, and MEZ. KMS, DR, DC, and MZ drafted the manuscript. All authors read and approved the final manuscript.

Author's information

Karyn Meltz Steinberg and Dhanya Ramachandran are co-first authors.

Author details

[1]Department of Human Genetics, Emory University School of Medicine, Whitehead Biomedical Research Building, Suite 301, Atlanta 30322GA, USA. [2]Graduate Program in Population Biology, Ecology and Evolution, Emory University, 1510 Clifton Road, Atlanta 30322GA, USA. [3]Department of Genome Sciences, University of Washington, 3720 15th Avenue NE, Seattle 98195WA, USA.

References

1. Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ, Ji HP, Snyder M: **Performance comparison of whole-genome sequencing platforms.** *Nat Biotechnol* 2012, **30**:78–82.
2. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ: **Target-enrichment strategies for next-generation sequencing.** *Nat Methods* 2010, **7**:111–118.
3. Shendure J, Ji H: **Next-generation DNA sequencing.** *Nat Biotechnol* 2008, **26**:1135–1145.
4. Kogan MD, Blumberg SJ, Schieve LA, Boyle CA, Perrin JM, Ghandour RM, Singh GK, Strickland BB, Trevathan E, van Dyck PC: **Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007.** *Pediatrics* 2009, **124**:1395–1403.
5. Fombonne E: **The prevalence of autism.** *JAMA* 2003, **289**:87–89.
6. Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM: **Concordance for the syndrome of autism in 40 pairs of afflicted twins.** *Am J Psychiatry* 1985, **142**:74–77.
7. Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Sykes N, Pagnamenta AT, Almeida J, Bacchelli E, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Carson AR, Casallo G, Casey J, Chu SH, Cochrane L, Corsello C, Crawford EL, Crossett A, *et al*: **A genome-wide scan for common alleles affecting risk for autism.** *Hum Mol Genet* 2010, **19**: 4072–4082.
8. Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J, Kalaydjieva L, McCague P, Dimiceli S, Pitts T, Nguyen L, Yang J, Harper C, Thorpe D, Vermeer S, Young H, Hebert J, Lin A, Ferguson J, Chiotti C, Wiese-Slater S, Rogers T, Salmon B, Nicholas P, Petersen PB, Pingree C, McMahon W, Wong DL, Cavalli-Sforza LL, Kraemer HC, *et al*: **A genomic screen of autism: evidence for a multilocus etiology.** *Am J Hum Genet* 1999, **65**:493–507.
9. Taniai H, Nishiyama T, Miyachi T, Imaeda M, Sumi S: **Genetic influences on the broad spectrum of autism: study of proband-ascertained twins.** *Am J Med Genet Neuropsychiatr Genet* 2008, **147B**:844–849.
10. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA: **Characteristics and concordance of autism spectrum disorders among 277 twin pairs.** *Arch of Pediat Adol Med* 2009, **163**:907–914.
11. Constantino JN, Zhang Y, Frazier T, Abbacchi AM, Law P: **Sibling recurrence and the genetic epidemiology of autism.** *Am J Psychiatry* 2010, **167**:1349–1356.

12. Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H: **The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood.** *Am J Psychiatry* 2010, **167**:1357–1363.

13. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, Miller J, Fedele A, Collins J, Smith K, Lotspeich L, Croen LA, Ozonoff S, Lajonchere C, Grether JK, Risch N: **Genetic heritability and shared environmental factors among twin pairs with autism.** *Arch Gen Psychiatry* 2011, **68**:1095–1102.

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

15. Devlin B, Scherer SW: **Genetic architecture in autism spectrum disorder.** *Curr Opin Genet Dev* 2012, **22**:229–237.

16. Piton A, Gauthier J, Hamdan FF, Lafrenière RG, Yang Y, Henrion E, Laurent S, Noreau A, Thibodeau P, Karemera L, Spiegelman D, Kuku F, Duguay J, Destroismaisons L, Jolivet P, Côté M, Lachapelle K, Diallo O, Raymond A, Marineau C, Champagne N, Xiong L, Gaspar C, Rivière JB, Tarabeux J, Cossette P, Krebs MO, Rapoport JL, Addington A, Delisi LE, *et al*: **Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia.** *Mol Psychiatry* 2011, **16**:867–880.

17. Mondal K, Ramachandran D, Patel VC, Hagen KR, Bose P, Cutler DJ, Zwick ME: **Excess variants in AFF2 detected by massively parallel sequencing of males with autism spectrum disorder.** *Hum Mol Genet* 2012, **21**:4356–4364.

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

19. State MW, Levitt P: **The conundrums of understanding genetic risks for autism spectrum disorders.** *Nat Neurosci* 2011, **14**:1499–1506.

20. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Alvarez Retuerto AI, Herman EI, Dong H, Hutman T, Sigman M, Ozonoff S, Klin A, *et al*: **Common genetic variants on 5p14.1 associate with autism spectrum disorders.** *Nature* 2009, **459**:528–533.

21. Weiss LA, Arking DE, Daly MJ, Chakravarti A: **A genome-wide linkage and association scan reveals novel loci for autism.** *Nature* 2009, **461**:802–808.

22. Devlin B, Melhem N, Roeder K: **Do common variants play a role in risk for autism? evidence and theoretical musings.** *Brain Res* 2011, **1380**:78–84.

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

24. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, Gilliam TC, Nowak NJ, Cook EH, Dobyns WB, Christian SL: **Recurrent 16p11.2 microdeletions in autism.** *Hum Mol Genet* 2008, **17**:628–638.

25. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, Thiruvahindrapduram B, Fiebig A, Schreiber S, Friedman J, Ketelaars CE, Vos YJ, Ficicioglu C, Kirkpatrick S, Nicolson R, Sloman L, Summers A, Gibbons CA, Teebi A, Chitayat D, Weksberg R, Thompson A, Vardy C, Crosbie V, Luscombe S, Baatjes R, *et al*: **Structural variation of chromosomes in autism spectrum disorder.** *Am J Hum Genet* 2008, **82**:477–488.

26. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T, Platt OS, Ruderfer DM, Walsh CA, Altshuler D, Chakravarti A, Tanzi RE, Stefansson K, Santangelo SL, Gusella JF, Sklar P, Wu BL, Daly MJ: **Association between microdeletion and microduplication at 16p11.2 and autism.** *N Engl J Med* 2008, **358**:667–675.

27. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnenblick LI, Alvarez Retuerto AI, Imielinski M, Hadley D, Bradfield JP, Kim C, Gidaya NB, Lindquist I, Hutman T, Sigman M, Kustanovich V, Lajonchere CM, Singleton A, Kim J, Wassink TH, McMahon WM, Owley T, Sweeney JA, Coon H, Nurnberger JI, Li M, Cantor RM, Minshew NJ, Sutcliffe JS, Cook EH, *et al*: **Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes.** *PLoS Genet* 2009, **5**:e1000536.

28. 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, *et al*: **Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.** *Nature* 2009, **459**:569–573.

29. 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, Bölte 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, *et al*: **Functional impact of global rare copy number variation in autism spectrum disorders.** *Nature* 2010, **466**:368–372.

30. Moreno-De-Luca D, SGENE Consortium, Mulle JG, Simons Simplex Collection Genetics Consortium, Kaminsky EB, Sanders SJ, GeneSTAR, Myers SM, Adam MP, Pakula AT, Eisenhauer NJ, Uhas K, Weik L, Guy L, Care ME, Morel CF, Boni C, Salbert BA, Chandrareddy A, Demmer LA, Chow EW, Surti U, Aradhya S, Pickering DL, Golden DM, Sanger WG, Aston E, Brothman AR, Gliem TJ, Thorland EC, Ackley T, Iyer R, Huang S, *et al*: **Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia.** *Am J Hum Genet* 2010, **87**:618–630.

31. Noor A, Whibley A, Marshall CR, Gianakopoulos PJ, Piton A, Carson AR, Orlic-Milacic M, Lionel AC, Sato D, Pinto D, Drmic I, Noakes C, Senman L, Zhang X, Mo R, Gauthier J, Crosbie J, Pagnamenta AT, Munson J, Estes AM, Fiebig A, Franke A, Schreiber S, Stewart AF, Roberts R, McPherson R, Guter SJ, Cook EH Jr, Dawson G, Schellenberg GD, *et al*: **Disruption at the PTCHD1 locus on Xp22.11 in autism spectrum disorder and intellectual disability.** *Sci Transl Med* 2010, **2**:49ra68.

32. 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, *et al*: **Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism.** *Neuron* 2011, **70**:863–885.

33. Levy D, Ronemus M, Yamrom B, Lee Y-H, 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: **Rare de novo and transmitted copy-number variation in autistic spectrum disorders.** *Neuron* 2011, **70**:886–897.

34. Celestino-Soper PB, Shaw CA, Sanders SJ, Li J, Murtha MT, Ercan-Sencicek AG, Davis L, Thomson S, Gambin T, Chinault AC, Ou Z, German JR, Milosavljevic A, Sutcliffe JS, Cook EHJ, Stankiewicz P, State MW, Beaudet AL: **Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE.** *Hum Mol Genet* 2011, **20**:4360–4370.

35. 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, *et al*: **Patterns and rates of exonic de novo mutations in autism spectrum disorders.** *Nature* 2012, **485**:242–245.

36. 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: **Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations.** *Nature* 2012, **485**:246–250.

37. 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: **De novo mutations revealed by whole-exome sequencing are strongly associated with autism.** *Nature* 2012, **485**:237–241.

38. Prange O, Wong TP, Gerrow K, Wang YT, El-Husseini A: **A balance between excitatory and inhibitory synapses is controlled by PSD-95 and neuroligin.** *Proc Natl Acad Sci USA* 2004, **101**:13915–13920.

39. Dean C, Dresbach T: **Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function.** *Trends Neurosci* 2006, **29**:21–29.

40. Varoqueaux F, Aramuni G, Rawson RL, Mohrmann R, Missler M, Gottmann K, Zhang W, Südhof TC, Brose N: **Neuroligins determine synapse maturation and function.** *Neuron* 2006, **51**:741–754.

41. Jamain S, Quach H, Betancur C, Råstam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, Bourgeron T, Study PARIS: Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 2003, **34**:27–29.
42. Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, Moizard M-P, Raynaud M, Ronce N, Lemonnier E, Calvas P, Laudier B, Chelly J, Fryns J-P, Ropers H-H, Hamel BCJ, Andres C, Barthélémy C, Moraine C, Briault S: X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet* 2004, **74**:552–557.
43. Yan J, Oliveira G, Coutinho A, Yang C, Feng J, Katz C, Sram J, Bockholt A, Jones IR, Craddock N, Cook EH, Vicente A, Sommer SS: Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. *Mol Psychiatry* 2005, **10**:329–332.
44. Ylisaukko-oja T, Rehnström K, Auranen M, Vanhala R, Alen R, Kempas E, Ellonen P, Turunen JA, Makkonen I, Riikonen R, Nieminen-von Wendt T, von Wendt L, Peltonen L, Järvelä I: Analysis of four neuroligin genes as candidates for autism. *Eur J Hum Genet* 2005, **13**:1285–1292.
45. Talebizadeh Z, Lam DY, Theodoro MF, Bittel DC, Lushington GH, Butler MG: Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism. *J Med Genet* 2006, **43**:e21.
46. Lawson-Yuen A, Saldivar J-S, Sommer S, Picker J: Familial deletion within NLGN4 associated with autism and tourette syndrome. *Eur J Hum Genet* 2008, **16**:614–618.
47. Yan J, Feng J, Schroer R, Li W, Skinner C, Schwartz CE, Cook EH, Sommer SS: Analysis of the neuroligin 4Y gene in patients with autism. *Psychiatr Genet* 2008, **18**:204–207.
48. Pampanos A, Volaki K, Kanavakis E, Papandreou O, Youroukos S, Thomaidis L, Karkelis S, Tzetis M, Kitsiou-Tzeli S: A substitution involving the NLGN4 gene associated with autistic behavior in the greek population. *Genet Test Mol Biomarkers* 2009, **13**:611–615.
49. Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, Lapointe M, Spiegelman D, Noreau A, Lafrenière RG, Fathalli F, Joober R, Krebs M-O, Delisi LE, Mottron L, Fombonne E, Michaud JL, Drapeau P, Carbonetto S, Craig AM, Rouleau GA: Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Hum Genet* 2011, **130**:563–573.
50. Qi H, Xing L, Zhang K, Gao X, Zheng Z, Huang S, Guo Y, Zhang F: Positive association of neuroligin-4 gene with nonspecific mental retardation in the Qinba Mountains Region of China. *Psychiatr Genet* 2009, **19**:1–5.
51. Feng J, Schroer R, Yan J, Song W, Yang C, Bockholt A, Cook EH, Skinner C, Schwartz CE, Sommer SS: High frequency of neurexin 1beta signal peptide structural variants in patients with autism. *Neurosci Lett* 2006, **409**:10–13.
52. Autism Genome Project Consortium, Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L, Feuk L, Qian C, Bryson SE, Jones MB, Marshall CR, Scherer SW, Vieland VJ, Bartlett C, Mangin LV, Goedken R, Segre A, Pericak-Vance MA, Cuccaro ML, Gilbert JR, Wright HH, Abramson RK, Betancur C, Bourgeron T, Gillberg C, Leboyer M, *et al*: Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007, **39**:319–328.
53. Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, Hefferan MP, Van Gorp S, Nazor KL, Boscolo FS, Carson CT, Laurent LC, Marsala M, Gage FH, Remes AM, Koo EH, Goldstein LSB: Probing sporadic and familial alzheimer's disease using induced pluripotent stem cells. *Nature* 2012, **482**:216–220.
54. Wiśniowiecka-Kowalnik B, Nesteruk M, Peters SU, Xia Z, Cooper ML, Savage S, Amato RS, Bader P, Browning MF, Haun CL, Duda AW, Cheung SW, Stankiewicz P: Intragenic rearrangements in NRXN1 in three families with autism spectrum disorder, developmental delay, and speech delay. *Am J Med Genet B* 2010, **153B**:983–993.
55. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, Nygren G, Rastam M, Gillberg IC, Anckarsäter H, Sponheim E, Goubran-Botros H, Delorme R, Chabane N, Mouren-Simeoni M-C, de Mas P, Bieth E, Rogé B, Héron D, Burglen L, Gillberg C, Leboyer M, Bourgeron T: Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* 2007, **39**:25–27.
56. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, Zwaigenbaum L, Fernandez B, Roberts W, Szatmari P, Scherer SW: Contribution of SHANK3 mutations to autism spectrum disorder. *Am J Hum Genet* 2007, **81**:1289–1297.
57. Gauthier J, Spiegelman D, Piton A, Lafrenière RG, Laurent S, St-Onge J, Lapointe L, Hamdan FF, Cossette P, Mottron L, Fombonne E, Joober R, Marineau C, Drapeau P, Rouleau GA: Novel de novo SHANK3 mutation in autistic patients. *American Journal of Medical Genetics Part B, Neuropsychiatric Genetics* 2009, **150B**:421–424.
58. Qin J, Jia M, Wang L, Lu T, Ruan Y, Liu J, Guo Y, Zhang J, Yang X, Yue W, Zhang D: Association study of SHANK3 gene polymorphisms with autism in Chinese Han population. *BMC Medical Genetics* 2009, **10**:61.
59. Daoud H, Bonnet-Brilhault F, Védrine S, Demattéi M-V, Vourc'hP, Bayou N, Andres CR, Barthélémy C, Laumonnier F, Briault S: Autism and nonsyndromic mental retardation associated with a de novo mutation in the NLGN4X gene promoter causing an increased expression level. *Biol Psychiatry* 2009, **66**:906–910.
60. Mitchell AA, Chakravarti A, Cutler DJ: On the probability that a novel variant is a disease-causing mutation. *Genome Res* 2005, **15**:960–966.
61. Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P, Ducat L, Spence SJ, Committee AGRES: The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet* 2001, **69**:463–466.
62. http://research.agre.org.
63. EmPrime. *Primer Designer*, http://primer.genetics.emory.edu.
64. Shetty AC, Athri P, Mondal K, Horner VL, Steinberg KM, Patel V, Caspary T, Cutler DJ, Zwick ME: SeqAnt: a web service to rapidly identify and annotate DNA sequence variations. *BMC Bioinformatics* 2010, **11**:471.
65. Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, Kuhn RM, Meyer LR, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Pohl A, Malladi VS, Li CH, Learned K, Kirkup V, Hsu F, Harte RA, Guruvadoo L, Goldman M, Giardine BM, Fujita PA, Diekhans M, Cline MS, Clawson H, Barber GP, Haussler D, James Kent W: The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res* 2012, **40**:D918–D923.
66. Consortium GP: A map of human genome variation from population-scale sequencing. *Nature* 2010, **467**:1061–1073.
67. Zwick ME, Mcafee F, Cutler DJ, Read TD, Ravel J, Bowman GR, Galloway DR, Mateczun A: Microarray-based resequencing of multiple Bacillus anthracis isolates. *Genome Biol* 2005, **6**:R10.
68. Primer 3. http://frodo.wi.mit.edu.
69. NIMH human genetics initiative. https://www.nimhgenetics.org/available_data/controls/.
70. Collins SC, Bray SM, Suhl JA, Cutler DJ, Coffee B, Zwick ME, Warren ST: Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. *Am J Med Genet A* 2010, **152A**:2512–2520.
71. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E: TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 2003, **31**:374–378.
72. Schneider TD, Stephens RM: Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 1990, **18**:6097–6100.
73. Sequence Logo. http://genome.tugraz.at/Logo/.
74. Watterson GA: On the number of segregating sites in genetical models without recombination. *Theor Pop Biol* 1975, **7**:256–276.
75. Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989, **123**:585–595.
76. Gauthier J, Bonnet A, St-Onge J, Karemera L, Laurent S, Mottron L, Fombonne E, Joober R, Rouleau GA: NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population. *Am J Med Genet B Psychiatr Genet* 2005, **132B**:74–75.
77. Myers RM, Stamatoyannopoulos J, Snyder M, Dunham I, Hardison RC, Bernstein BE, Gingeras TR, Kent WJ, Birney E, Wold B: A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 2011, **9**:e1001046.
78. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B: Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 2009, **459**:108–112.
79. Gargiulo G, Levy S, Bucci G, Romanenghi M, Fornasari L, Beeson KY, Goldberg SM, Cesaroni M, Ballarini M, Santoro F, Bezman N, Frigè G, Gregory PD, Holmes MC, Strausberg RL, Pelicci PG, Urnov FD, Minucci S: NA-Seq: a discovery tool for the analysis of chromatin structure and dynamics during differentiation. *Dev Cell* 2009, **16**:466–481.
80. Tsiftsoglou AS, Tsamadou AI, Papadopoulou LC: Heme as key regulator of major mammalian cellular functions: molecular, cellular, and pharmacological aspects. *Pharmacol Ther* 2006, **111**:327–345.

81. Tsai RY, Reed RR: **Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz.** *Mol Cell Biol* 1998, **18**:6447–6456.
82. Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273**:1516–1517.
83. Lander ES: **The new genomics: global views of biology.** *Science* 1996, **274**:536–539.
84. Kruglyak L: **Prospects for whole-genome linkage disequilibrium mapping of common disease genes.** *Nat Genet* 1999, **22**:139–144.
85. Zwick ME, Cutler DJ, Chakravarti A: **Patterns of genetic variation in Mendelian and complex traits.** *Annu Rev Genomics Hum Genet* 2000, **1**:387–407.
86. Consortium IH: **A haplotype map of the human genome.** *Nature* 2005, **437**:1299–1320.
87. Consortium IH, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, *et al*: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 2007, **449**:851–861.
88. Barton NH, Turelli M: **Evolutionary quantitative genetics: how little do we know?** *Annu Rev Genet* 1989, **23**:337–370.
89. Barton NH, Keightley PD: **Understanding quantitative genetic variation.** *Nat Rev Genet* 2002, **3**:11–21.
90. Haldane JBS: **A mathematical theory of natural and artificial selection, Part V: selection and mutation.** *Math Proc Cambridge* 1927, **23**:838–844.
91. Visscher PM, Brown MA, McCarthy MI, Yang J: **Five years of GWAS discovery.** *Am J Hum Genet* 2012, **90**:7–24.
92. Falconer DS: **The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus.** *Ann Hum Genet* 1967, **31**:1–20.
93. Ferrando-Miguel R, Cheon MS, Yang JW, Lubec G: **Overexpression of transcription factor BACH1 in fetal Down syndrome brain.** *J Neural Transm Suppl* 2003, **67**:193–205.
94. Sakoda E, Igarashi K, Sun J, Kurisu K, Tashiro S: **Regulation of heme oxygenase-1 by transcription factor Bach1 in the mouse brain.** *Neurosci Lett* 2008, **440**:160–165.
95. Bennetto L, Kuschner ES, Hyman SL: **Olfaction and taste processing in autism.** *Biol Psychiatry* 2007, **62**:1015–1021.
96. Suzuki Y, Critchley HD, Rowe A, Howlin P, Murphy DG: **Impaired olfactory identification in Asperger's syndrome.** *J Neuropsychiatry Clin Neurosci* 2003, **15**:105–107.
97. Tsai RY, Reed RR: **Cloning and functional characterization of Roaz, a zinc finger protein that interacts with O/E-1 to regulate gene expression: implications for olfactory neuronal development.** *J Neurosci* 1997, **17**:4159–4169.
98. Baron-Cohen S, Ashwin E, Ashwin C, Tavassoli T, Chakrabarti B: **Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity.** *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2009, **364**:1377–1383.
99. Shinawi M, Patel A, Panichkul P, Zascavage R, Peters SU, Scaglia F: **The Xp contiguous deletion syndrome and autism.** *Am J Med Genet A* 2009, **149A**:1138–1148.
100. Allen-Brady K, Robison R, Cannon D, Varvil T, Villalobos M, Pingree C, Leppert MF, Miller J, Mcmahon WM, Coon H: **Genome-wide linkage in Utah autism pedigrees.** *Mol Psychiatry* 2010, **15**:1006–1015.
101. Chung R-H, Ma D, Wang K, Hedges DJ, Jaworski JM, Gilbert JR, Cuccaro ML, Wright HH, Abramson RK, Konidari I, Whitehead PL, Schellenberg GD, Hakonarson H, Haines JL, Pericak-Vance MA, Martin ER: **An X-chromosome-wide association study in autism families identifies TBL1X as a novel autism spectrum disorder candidate gene in males.** *Mol Autism* 2011, **2**:18.
102. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, Shafer N, Bernier R, Ferrero GB, Silengo M, Warren ST, Fichera M, Romano C, Raskind WH, Eichler EE: **Relative burden of large CNVs on a range of neurodevelopmental phenotypes.** *PLoS Genet* 2011, **7**:e1002334.