Molecular Psychiatry (2013), 1–7
© 2013 Macmillan Publishers Limited All rights reserved 1359-4184/13
www.nature.com/mp

ORIGINAL ARTICLE

Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations

C Toma[1,2], B Torrico[1,2], A Hervás[3], R Valdés-Mas[4], A Tristán-Noguero[1], V Padillo[5], M Maristany[5], M Salgado[3], C Arenas[6], XS Puente[4], M Bayés[7] and B Cormand[1,2,8]

Autism is a severe neurodevelopmental disorder, the aetiology of which remains mainly unknown. Family and twin studies provide strong evidence that genetic factors have a major role in the aetiology of this disease. Recently, whole exome sequencing (WES) efforts have focused mainly on rare *de novo* variants in singleton families. Although these studies have provided pioneering insights, *de novo* variants probably explain only a small proportion of the autism risk variance. In this study, we performed exome sequencing of 10 autism multiplex families with the aim of investigating the role of rare variants that are coinherited in the affected sibs. The pool of variants selected in our study is enriched with genes involved in neuronal functions or previously reported in psychiatric disorders, as shown by Gene Ontology analysis and by browsing the Neurocarta database. Our data suggest that rare truncating heterozygous variants have a predominant role in the aetiology of autism. Using a multiple linear regression model, we found that the burden of truncating mutations correlates with a lower non-verbal intelligence quotient (NVIQ). Also, the number of truncating mutations that were transmitted to the affected sibs was significantly higher (twofold) than those not transmitted. Protein–protein interaction analysis performed with our list of mutated genes revealed that the postsynaptic YWHAZ is the most interconnected node of the network. Among the genes found disrupted in our study, there is evidence suggesting that *YWHAZ* and also the X-linked *DRP2* may be considered as novel autism candidate genes.

Molecular Psychiatry advance online publication, 3 September 2013; doi:10.1038/mp.2013.106

Keywords: autism spectrum disorder; exome sequencing; multiplex families; novel candidate genes; rare genetic variants; truncating mutations

INTRODUCTION

Autism spectrum disorders (ASDs) represent a group of neurodevelopmental disorders characterized by impairments in reciprocal social interaction, verbal and non-verbal communication and stereotyped behavioural patterns.[1] The prevalence of ASD is estimated to be around 0.6%, making it one of the most prevalent disorders in childhood.[2] Family and twin studies have provided strong evidence that genetic factors have a major role in the aetiology of ASD: the sibling recurrence risk is ~20-fold higher than in the general population,[3] and the concordance rate is ~80–90% in monozygotic twins versus ~10–30% in dizygotic twins.[4]

Although in recent years several high-throughput genetic studies have identified novel ASD genes in idiopathic autism, the vast majority of the underlying genetic factors remain still unknown. In this regard, no mutation or copy number variant (CNV) found to be associated with autism accounts for more than 1% of cases, which makes the estimation of the number of independent risk loci extremely high.[5] Nowadays, whole exome sequencing (WES) represents a powerful technology for identifying rare single-nucleotide variants (SNVs) that may help to depict the complex genetic architecture of autism and unravel some of the missing heritability. Recently, several studies performed WES in parent–child trios, focusing on *de novo* variants.[6–10] These studies show that *de novo* mutations have an increased mutation rate in autism and correlate with increased paternal age.[7,9] However, even the high mutation rate observed (approximately one *de novo* single-nucleotide variant in the exome per autistic child) cannot fully explain the aetiology of the disorder, in which additive effects of several gene mutations are expected under a multihit model of inheritance.[11,12] The exome studies performed to date have focused mainly on *de novo* variants. Although this approach is useful, it is not sufficient to uncover the complete pool of aetiological variants, the majority of which would be expected to be in the inherited fraction. Several studies in singleton and multiplex families suggest a differential genetic landscape, with families with multiple affected individuals being more likely to bear inherited events than *de novo* variants.[13–15] According to the multihit model, we would expect mutations in several genes to be at the basis of the disorder. Current data show that most aetiological variants would be heterozygous, whereas homozygous or compound heterozygous mutations would be much less frequent.[16–18]

The aim of this study was to uncover new ASD candidate genes by analysing the pool of inherited rare variants in multiplex families. We performed WES on 10 families, considering those rare variants not present in public or in-house databases that were predicted to be damaging and coinherited in the affected siblings.

[1]Department of Genetics, University of Barcelona, Barcelona, Spain; [2]Biomedical Network Research Centre on Rare Diseases (CIBERER), Barcelona, Spain; [3]Child and Adolescent Mental Health Unit, Hospital Universitari Mútua de Terrassa, Barcelona, Spain; [4]Department of Biochemistry and Molecular Biology, University of Oviedo-IUOPA, Barcelona, Spain; [5]Developmental Disorders Unit (UETD), Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain; [6]Department of Statistics, University of Barcelona, Barcelona, Spain; [7]National Centre for Genomic Analysis (CNAG), Barcelona, Spain and [8]Institute of Biomedicine, University of Barcelona, Barcelona, Spain. Correspondence: Professor B Cormand, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Avinguda Diagonal 643, Edifici Prevosti, 3ª planta, 08028 Barcelona, Spain.
E-mail: bcormand@ub.edu
Received 13 February 2013; revised 23 May 2013; accepted 22 July 2013

MATERIALS AND METHODS

ASD patients and phenotype assessment

In our study, we considered 10 multiplex families with ASD: nine with two affected probands and one with three affected probands. Clinical data are summarized in Table 1, and more details are provided in the Supplementary Information.

CNVs analysis with CytoScan HD Array

Genome-wide DNA copy number analysis was performed in all ASD patients in our study (21 individuals), using the CytoScan HD Array (Affymetrix, Santa Clara, CA, USA). This study was carried out to rule out families with chromosome abnormalities previously reported in autism. Details of this study are summarized in the Supplementary Information.

Exome capture and sequencing

A total of 41 individuals including 21 probands and their parents were processed for exome enrichment using 3 μg of blood genomic DNA. The TruSeq DNA Sample Preparation kit (Illumina, San Diego, CA, USA) and the NimbleGen SeqCap EZ Human Exome Library v.1.0 kit (Roche, Basel, Switzerland), which targets approximately 20 000 genes corresponding to 34 Mb, were used according to the manufacturer's instructions. The resulting exome libraries were applied to an Illumina flowcell for cluster generation. Paired-end sequencing was performed on a HiSeq2000 instrument (Illumina) using 76-base reads.

Exome data analysis

Reads were aligned to the reference genome (GRCh37) using Burrows–Wheeler analysis with the same option,[19] and a BAM file was generated using SAMtools.[20] Polymerase chain reaction duplicates were removed using SAMtools and custom scripts, and initial single-nucleotide polymorphisms calling was performed using a combination of SAMtools and Sidrón as described previously.[21] Statistics for the callability (percentage of the exome covered by at least 10 reads with mapping quality $\geqslant 30$ and base quality $\geqslant 20$) of each sample are shown in Supplementary Table 1. On average, individuals had 83.5% of the target covered. Common variants, defined as those present in dbSNP135 with a minor allele frequency $\geqslant 1\%$, were filtered out.

Variants selection and validation

After the common single-nucleotide polymorphisms were filtered out, we selected all variants transmitted from parents to both affected probands, with the exception of family SJD_50, in which we selected variants transmitted to two out of three or to all three affected probands. All untransmitted variants or variants transmitted to only one affected sib were excluded from the study. In-house and WES false-positive databases were also used to filter the pool of variants.[22] Our selection included all truncating mutations (nonsense and indels producing a frame shift), splice sites and missense variants. Missense variants were included in our selection only if they were predicted to be pathogenic by SIFT (Sorting Intolerant From Tolerant) or PolyPhen (Polymorphism Phenotyping).[23,24] Those variants that passed all filtering steps were validated by performing standard polymerase chain reactions and Sanger sequencing. On average, 90.1% of the selected variants were successfully validated. The workflow followed in this study is summarized in Supplementary Figure 1.

Statistical and bioinformatic analyses

We performed Gene Ontology enrichment analysis of our set of selected variants using the FatiGO software available at the Babelomics suite (http://babelomics.bioinfo.cipf.es).[25] The database Neurocarta (http://gemma-doc.chibi.ubc.ca/neurocarta/) was used to identify potential gene–phenotype relationships in our gene list. The SFARI database (https://gene.sfari.org/autdb/Welcome.do) was used to obtain an annotated list of genes involved in autism: genes predisposing to autism in the context of syndromic disorder (category S), strong ASD candidate genes (category 2) and ASD genes with suggestive evidence (category 3).

The pool of genes selected in this study was investigated for networking and gene function relationships using Ingenuity Pathway Analysis (IPA) software (www.ingenuity.com). IPA computes a score based on the probability of finding a given pool of genes in a network from the Ingenuity's Knowledge Database by chance (score $= -\log$ (Fisher exact test probability)). IPA was also used to assess 'disease and disorder' associations emerging from our pool of genes using the 'Top Bio Functions' category.

A multiple linear regression model was considered to assess the relation between the non-verbal intelligence quotient (NVIQ) and the different categories of genetic variants (truncating, non-synonymous damaging, non-synonymous benign, synonymous). The basic regression assumptions (errors are normally and independently distributed random variables) were checked to ensure that the model was appropriate. Following a conservative approach, to detect the regressors (the different variant categories) significantly related to the response NVIQ, the Bonferroni-adjusted procedure was considered. For all regressor variables, the sample Pearson's correlation coefficient with the response variable and its significance was calculated. Furthermore, a simulation study was performed. By Boostrap resampling, 50 new simulated analyses with a sample size of 100 individuals each were computed and the corresponding multiple linear regression model was checked.

A multiple logistic regression model was used to establish the relation between variants status (the non-synonymous variants transmitted to both affected siblings in a family (TR) and those variants not transmitted from parents (NT)) and SIFT, PolyPhen scores.

For details about the multiple linear regression, model adequacy, Bonferroni correction, simulation study and multiple logistic regression see Supplementary Information.

Furthermore, the $[0,1] \times [0,1]$ square determined by the SIFT and PolyPhen scores of TR and NT variants was divided into four quadrants according to the pathogenicity thresholds of 0.05 and 0.5 used for SIFT and PolyPhen, respectively. We compared the proportion of TR and NT variants between the benign (values from 0.05 to 1 for SIFT and 0 to 0.5 for PolyPhen) and pathogenic (values from 0 to 0.05 for SIFT and 0.5 to 1 for PolyPhen) quadrant using the exact binomial test and Fisher's exact test. The analyses were performed using the R package (http://CRAN.R-project.org).

RESULTS

In this study, we sequenced the exomes of 10 autism multiplex families and validated 220 rare variants transmitted from each parent to at least two affected siblings. All selected variants were filtered for common variants and using in-house databases and

Table 1. Description of the multiplex ASD families included in our study

Family ID (sib pair ID)	*Sex (years)*	*Phenotype*	*NVIQ*	*Language delay*
MT_69 (sib.3–sib.4)	M (17)/M (20)	Aut/Aut	50/56	Yes/yes
MT_76 (sib.3–sib.4)	M (22)/M (29)	PDD-NOS/Asp	81/85	Not/not
MT_109 (sib.3–sib.4)	M (13)/M (25)	Aut/Aut	64/35	Yes/yes
MT_160 (sib.3–sib.4)	M (14)/M (11)	PDD-NOS/PDD-NOS	68/105	Not/not
MT_28 (sib.3–sib.4)	M (16)/M (14)	PDD-NOS/Asp	105/107	Not/not
MT_151 (sib.3–sib.4)	M (20)/M (11)	Asp/Asp	98/92	Not/not
SJD_49 (sib.3–sib.4)	M (6)/M (4)	Asp/PDD-NOS	129/112	Not/not
SJD_10 (sib.3–sib.4)	M (13)/M (10)	PDD-NOS/PDD-NOS	105/95	Not/not
SJD_34 (sib.3–sib.4)	M (13)/M (9)	Asp/Asp	107/139	Not/not
SJD_50 (sib.3–sib.4–sib.5)	M (16)/F (15)/M (8)	Asp/Asp/PDD-NOS	97/98/95	Not/not/not

Abbreviations: Asp, Asperger syndrome; Aut, autism; F, female; M, male; NVIQ, non-verbal intelligence quotient; PDD-NOS, pervasive developmental disorder not otherwise specified.
For each family, data are provided for the two or three affected siblings.

© 2013 Macmillan Publishers Limited

Table 2. Gene-disrupting rare variants shared by the affected sibs in each multiplex family

Family_ID	*Chromosome*	*Nucleotide position*[a]	*Gene*	*AA position*[b]	*Effect*
MT_109	5	180 477 255	*BTNL9*	208/535	fs (del-C)
MT_109	6	28 962 882–886	*ZNF311*	632/666	fs (del-TTCTT)
MT_109	11	133 712 408	*SPATA19*	137/167	Nonsense
MT_109	19	11 943 225[c]	*ZNF440*	412/595	Nonsense
MT_151	1	152 883 252	*IVL*	327/585	fs (del-G)
MT_151	3	113 652 385[c]	*GRAMD1C*	413/662	Nonsense
MT_151	4	17 606 229–232	*LAP3*	400/519	fs (del-TGGG)
MT_151	7	99 489 859	*TRIM4*	477/500	Nonsense
MT_151	8	144 642 836	*GSDMD*	158/484	Nonsense
MT_151	X	100 503 119	*DRP2*	432/957	Nonsense
MT_160	8	10 557 940–941	*C8orf74*	282/294	fs (ins-C)
MT_160	8	101 936 203–204	*YWHAZ*	220/245	fs (ins-T)
MT_160	12	75 892 479	*GLIPR1*	209/266	Nonsense
MT_28	2	114 392 641[c]	*RABL2A*	78/238	fs (del-G)
MT_28	10	91 162 622–623	*IFIT1*	197/478	fs (ins-A)
MT_69	3	49 274 079	*CCDC36*	52/594	Nonsense
MT_69	8	67 813 550	*C8orf45*	579/681	fs (del-G)
MT_69	11	2 339 146–147	*TSPAN32*	319/320	fs (del-TC)
MT_69	11	114 577 559[c]	*FAM55B*	529/559	Nonsense
MT_69	15	42 154 996	*SPTBN5*	2419/3674	Nonsense
MT_69	22	26 890 166–169	*TFIP11*	700/837	fs (del-AGAT)
MT_76	5	38 412 639–642	*EGFLAM*	462/1017	fs (del-AAGT)
MT_76	5	74 882 867	*POLK*	415/870	fs (del-A)
MT_76	10	88 703 106	*MMRN2*	479/949	Nonsense
MT_76	19	7 965 639–640	*LRRC8E*	745/796	fs (del-CA)
SJD_10	8	131 921 998	*ADCY8*	532/1251	Nonsense
SJD_10	11	34 167 721–722	*NAT10*	949/953	fs (ins-A)
SJD_10	12	54 756 741	*GPR84*	299/396	fs (del-G)
SJD_10	22	18 609 783	*TUBA8*	346/449	Nonsense
SJD_34	1	160 143 953	*ATP1A4*	682/1029	Nonsense
SJD_34	11	121 028 796	*TECTA*	1518/2155	Nonsense
SJD_34	19	53 644 208–209	*ZNF347*	626/840	fs (del-AA)
SJD_49	13	20 796 930–931	*GJB6*	231/261	fs (ins-T)
SJD_49	19	10 610 629–633	*KEAP1*	27/624	fs (del-TGCCC)
SJD_50	15	76 225 440–443	*FBXO22*	402/403	fs (del-ATAA)
SJD_50[d]	17	71 205 688	*FAM104A*	202/207	Nonsense

Abbreviations: AA, amino acid; del or ins, deletion or insertion followed by the nucleotides involved; fs, frame shift.
[a]Positions are indicated according to the GRCh37/hg19 assembly of the UCSC Genome Browser (genome.ucsc.edu). [b]Amino acid position altered by the mutation/full polypeptide encoded by the transcript. [c]Rare variants already described in dbSNP 137. [d]Multiplex family with three probands: variant found in two out of three affected.

were predicted to be pathogenic. The pool of validated variants is listed in Supplementary Table 2. Most (83%) are missense changes, whereas variants predicted to cause protein truncation (nonsense mutations and frameshift indels) accounted for 16% of the total. Variants altering canonical splice sites or start codons were less represented, at 0.5% each (Supplementary Figure 2). All these variants were heterozygous, none of them was homozygous or compound heterozygous, and no gene was found mutated in more than one family. Truncating mutations represent remarkable events in the exome and may pinpoint potential candidate genes for a disease. Interestingly, in our study this kind of mutation represented a substantial proportion of the whole pool of variants, with 20 indels and 16 nonsense mutations (Table 2).

Chromosomal rearrangements and previously described fully penetrant deletions were ruled out in the affected probands, by karyotyping and through a CNV study. This analysis, followed by experimental validation, revealed six inherited CNVs present in five sib-pairs identifying the following genes: *COL4A3-MFF*, *FHIT*, *MRPL36-NDUFS6*, *CTNND2*, *GRM1* and *ASAH1* (Table 3). Structural variants spanning the genes *FHIT* and *CTNND2* had previously been described in autism.[14,26]

To establish whether the pool of variants selected in our exome study was enriched in potential ASD susceptibility genes, we conducted a Gene Ontology study. Cellular component analysis showed 'Cell junction' to be a significantly enriched category after applying multiple testing corrections ($P = 0.043$) (Supplementary Table 3). Several genes in this category such as *CYFIP1*, *SCN1A*, *DNM2*, *FLNB*, *GABRA4*, *P2RX4*, *PJA2* and *SV2B* have been described in autism or are involved in neuronal functions. Several GO biological processes were over-represented in our study and some of them may be relevant for autism, such as 'Neuronal tube development' ($P = 0.044$) and 'Regulation of action potential' ($P = 0.044$) (Supplementary Table 4). In addition, we used IPA to analyze the 'Top Bio function' category of 'Diseases and disorders' and found that 'Developmental disorders' and 'Neurological diseases', among others, were the most significant groups related to the genes in our study (Supplementary Table 5). The Neurocarta database was used to build a map of psychiatric phenotypes previously linked to these genes (Figure 1). The majority of genes were associated with autism, but also with epilepsy, dyslexia, intellectual disability, attention-deficit hyperactivity disorder and schizophrenia.

We also investigated the possible correlation between the severity of the mutations and intellectual disability, by using the NVIQ. NVIQ was assessed in our sample of 21 affected siblings, with scores ranging from 35 to 139 and a mean of 91 (Supplementary Figure 3). A multiple linear regression model was applied with NVIQ as a response variable and truncating, non-synonymous (damaging and benign changes according to SIFT and PolyPhen) and synonymous variants as regressor variables. The results showed that only truncating variants contribute significantly to NVIQ ($P = 0.007$). Furthermore, truncating variants

© 2013 Macmillan Publishers Limited

Table 3. Validated CNVs shared by the affected sibs in each multiplex family

Ind_ID	*Chromosome*	*Min*[a]	*Max*[a]	*Cytoband*	*Type*	*Size (kb)*	*Genes*
SJD_34.3–34.4	2	228 149 569	228 193 389	q36.3	Gain	43.82	*COL4A3*, *MFF*
MT_109.3–109.4	3	60 478 959	60 572 752	p14.2	Loss	93.793	*FHIT*
MT_76.3–76.4	5	1 742 845	1 849 924	p15.33	Gain	107.079	*MRPL36*, *NDUFS6*
SJD_34.3–34.4	5	11 619 568	12 183 983	p15.2	Gain	564.415	*CTNND2*
SJD_49.3–49.4	6	146 309 359	146 373 644	q24.3	Loss	64.285	*GRM1*
SJD_50.3–50.4	8	17 908 916	17 946 695	p22	Loss	37.779	*ASAH1*

Abbreviation: CNV, copy number variant.
[a]Positions are indicated according to the GRCh37/hg19 assembly of the UCSC Genome Browser (www.genome.ucsc.edu).

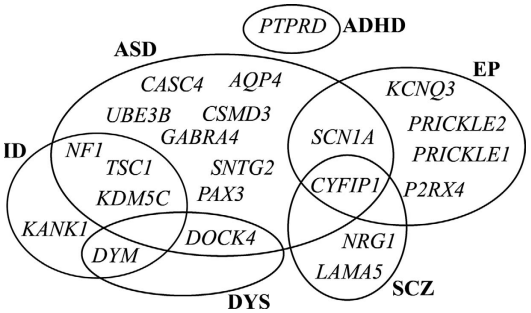

Figure 1. Genes identified in our study that have already been associated with autism or with other psychiatric conditions according to Neurocarta (http://gemma-doc.chibi.ubc.ca/neurocarta/). ADHD, attention-deficit hyperactivity disorder; ASD, autism spectrum disorder; EP, epilepsy; SCZ, schizophrenia; DYS, dyslexia; ID, intellectual disability.

were correlated with lower NVIQ scores (correlation coefficient $r = -0.517$, $P = 0.016$; Figure 2) and explained 26% of NVIQ variance in our ASD sample ($r^2 = 0.267$). A simulation study including 50 new simulated analyses of 100 individuals each obtained the same results as in the original sample (for more details see Supplementary Information, Supplementary Table 7 and Supplementary Figure 8).

We subsequently investigated whether truncating mutations might have a major role in autism aetiology by comparing the number of those that were cotransmitted (Table 1) with those that were not transmitted (Supplementary Table 6). We found a significant difference, considering the total number of rare variants, between disrupting mutations that were transmitted to affected probands (16 nonsense variants and 20 frameshift indels) and those not transmitted (9 nonsense variants and 9 frameshift indels) (Fisher's exact test, $P = 0.015$; Supplementary Figure 4). Subsequently, we investigated the potential contribution to ASD of rare non-synonymous variants predicted to be pathogenic, by comparing the group of transmitted (TR) single-nucleotide variants with those that were not transmitted (NT). A multiple logistic regression analysis (fitted logit model) was performed to investigate whether the scores given by SIFT and PolyPhen were able to discriminate the single-nucleotide variants in relation to their TR or NT status, but the results were not significant (SIFT $P = 0.122$; PolyPhen $P = 0.811$). To further test the possible relation of the non-synonymous variants with the disease, we plotted all missense changes against the PolyPhen scores (X axis) and SIFT scores (Y axis) (Supplementary Figure 5) to determine whether there was any difference in the proportion of variants predicted to be pathogenic (PolyPhen scores > 0.5 and SIFT score < 0.05) between the TR and NT groups. No significant differences were detected (two-sided exact binomial test, $P = 0.47$).

The list of 220 genes harbouring rare predicted pathogenic mutations (Supplementary Table 2) and genes found in the CNV study (Table 2) were considered together and analysed by IPA to identify canonical networks that may suggest interactions of potential ASD candidate genes. The best scoring network was obtained after considering only direct interactions (score = 89) (Figure 3). It is noteworthy that this network included genes previously involved in autism or other psychiatric disorders such as *SCN1A*, *PAX3*, *KDM5C*, *TSC1*, *NF1*, *CYFIP1*, *KCNQ3* and *LAMA5*. The *YWHAZ* gene was the most interconnected node in this network. Furthermore, when the protein–protein interaction analysis was performed with an additional 43 ASD genes from the SFARI Gene database (Categories S, 2 and 3), the score of this network improved (score = 106). *YWHAZ* was again the main node and interacted directly with *TSC1* and *CYFIP1*, two ASD genes found to be mutated in our study, and also with other ASD genes from the SFARI list such as *UBE3A*, *DISC1*, *MET* and *TSC2* (Supplementary Figure 6).

DISCUSSION

Several exome sequencing reports of autism trios have been published in the past few years. These studies have enabled the identification of novel candidate genes for ASD by focusing on *de novo* variants. Despite these encouraging results, *de novo* variants represent probably $< 5\%$ of autism risk variance,[8] and hence inherited rare variants may account for a considerable proportion of the missing heritability in autism. Here we present findings from the exome sequencing of 10 multiplex families, in which only the inherited rare variants shared by the affected siblings in a family and predicted to be pathogenic were considered. The resulting list of about 220 identified genetic variants was assessed for GO enrichment analysis and networks of gene interactions (IPA). These approaches identified interesting categories related to developmental disorders or involving neuronal functions, and revealed plausible interactions with previously reported ASD genes. In our study, we identified a substantial number of genes already associated with autism or with other psychiatric conditions, suggesting that there is a common genetic background for psychiatric disorders.[27,28] The most interesting finding emerging from our study suggests a major role for truncating mutations in autism. We found that those probands with a higher number of heterozygous disrupting mutations are those with lower NVIQ scores. In addition, we found more truncating mutations that were transmitted and shared between the affected sibs (36 mutations) than those that were not transmitted (18 mutations). Interestingly, Iossifov *et al.*[9] analysed exome data and described a twofold higher rate of disrupting *de novo* mutations in affected probands compared with unaffected siblings. Also, a very recent WES study on inherited homozygous or compound heterozygous loss-of-function mutations found a twofold enrichment in autism compared to a control group.[18] Such data suggest a genetic model for autism based on the cumulative contribution of truncating alleles and other rare

© 2013 Macmillan Publishers Limited

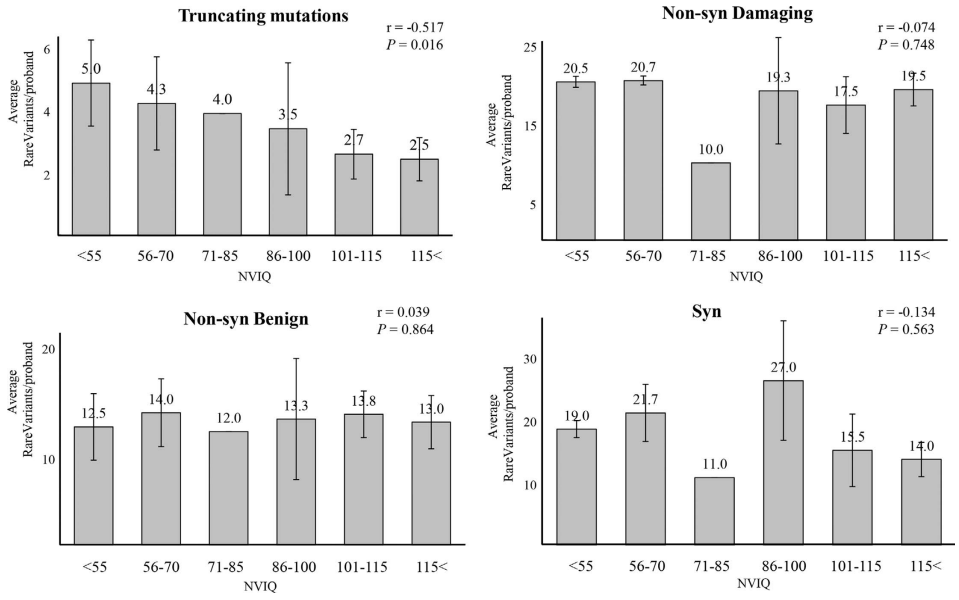

Figure 2. Distribution of the average number of rare variants per proband for each interval of non-verbal intelligence quotient (NVIQ) in four variant categories: (1) truncating mutations, which include nonsense and frameshift mutations; (2) non-syn damaging, non-synonymous mutations predicted to be damaging by SIFT or PolyPhen; (3) non-syn benign, non-synonymous mutations predicted to be benign by SIFT and PolyPhen; and (4) Syn, rare synonymous variants. The pool of rare variants considered are those inherited by two or three affected sibs in a family. Multiple linear regression analysis of these data showed that the number of truncating variants contributes significantly to NVIQ ($P = 0.007$). In contrast, the contribution of the other mutation types was not significant ($P > 0.385$). The correlation coefficients between NVIQ and each mutation type with their significances are also presented.

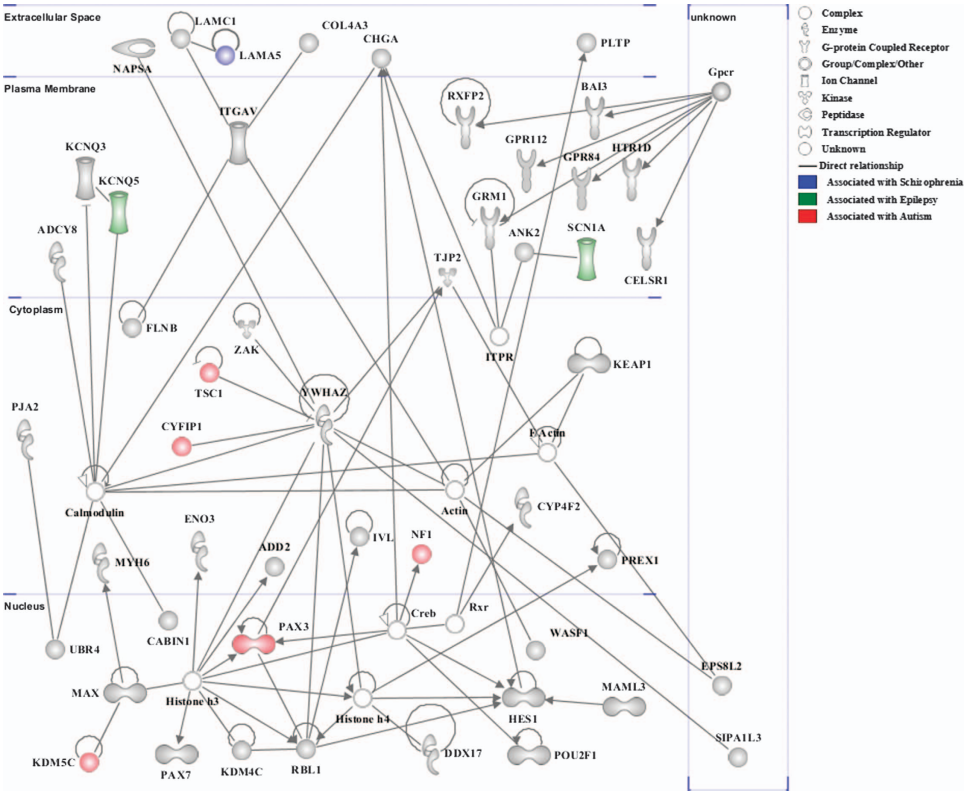

Figure 3. Protein–protein interaction analysis (Ingenuity Pathway Analysis, IPA) including all genes found to be mutated in our study. Only direct interactions among proteins were considered. Proteins in grey or colour represent genes identified in our study, with red indicating a previous association with autism, green with epilepsy and blue with schizophrenia. Proteins depicted in white are those not present in our study. Upregulatory effects are represented by outward pointing arrows, downregulatory effects are represented by outward ticks, and circular arrows indicate homotypic interactions.

© 2013 Macmillan Publishers Limited

pathogenic variants, coupled with rare structural variants, in which the impact of common variants may be less important than previously thought.[29] Most genes with truncating mutations found in our study have an unknown function and none have previously been described in autism. Among these, *YWHAZ* and *DRP2* may be considered as strong novel ASD candidate genes. The NHLBI Exome Sequencing Project (ESP) database (http://evs.gs.washington.edu/EVS/) was used to verify the frequency of disrupting mutations in these genes in about 6500 individuals. No truncating mutations were listed in either *YWHAZ* or *DRP2*.

The *YWHAZ* gene, encoding a postsynaptic protein, is the most intriguing candidate in our study. The protein physically interacts with numerous ASD gene products such as *TSC1*, *TSC2*, *DISC1*, *UBE3A* and *CYFIP1* (Supplementary Figure 6). YWHAZ belongs to the highly conserved 14-3-3 protein family that comprises seven isoforms (β, γ, ε, η, ζ, σ, θ) involved in signal transduction by binding specific pSer/pThr motifs. These proteins are involved in a wide range of processes including cell cycle, transcription, neuronal development, migration and neurite outgrowth. Although ubiquitously expressed, expression levels are highest in the brain.[30,31] The members of this family have been associated with several neurodevelopmental disorders, syndromes or psychiatric diseases. Deletions of the contiguous *YWHAE* and *PAFAH1B1* genes are responsible for two distinct Mendelian disorders depending on the size of the deletion: isolated lissencephaly and Miller–Dieker syndrome.[32] Duplications encompassing only the *YWHAE* gene are associated with a distinct phenotype involving autism and other behavioural symptoms.[33,34] This gene has also been strongly associated with schizophrenia and found to be downregulated together with other 14-3-3 isoforms in the prefrontal cortex of schizophrenic patients.[35,36] Other 14-3-3 members have also been associated with psychiatric phenotypes: heterozygous microdeletions encompassing *YWHAG* and *HIP1* were associated with epilepsy, learning difficulties and intellectual disability,[37] whereas *YWHAH* was associated with bipolar disorder.[38,39] Recently, Cheah *et al.*[40] reported that *YWHAZ* knockout mice show neurobehavioural and cognitive deficiencies, and aberrant development of the hippocampus with migratory defects of pyramidal cells and granular neurons. Furthermore, neuroproteomic studies showed decreased expression of *YWHAZ* in the brain of schizophrenic patients.[41] YWHAZ forms a molecular complex with DISC1 (Disrupted In Schizophrenia 1), Ndel1 and LIS1 to control the development of the hippocampus by coordinating neuronal migration, axonal pathfinding and synapse formation.[40] In our study, the 1-bp insertion found in *YWHAZ* causes a frame shift leading to a premature stop codon after 18 amino acids (Supplementary Figure 7). If nonsense-mediated mRNA decay does not prevent the degradation of this transcript, it is possible that this abnormal protein may act in a dominant-negative manner altering the normal protein interactions or preventing homo- and heterodimerization of the protein with 14-3-3 members.

The X-linked dystrophin-related protein 2 gene (*DRP2*), expressed mainly in the brain and spinal cord, is also a good ASD candidate. DRP2 forms a complex with periaxin and dystroglycan, regulating the myelination of Schwann cells. Sherman *et al.*[42] showed that loss of DRP2 affects the organization of the Schwann cell cytoplasm in the Cajal bands, although other members of the dystrophin family may partially supply the absence of DRP2. No human disease has been associated with the *DRP2* gene, although deletions spanning *DRP2* were described in individuals with X-linked agammaglobulinaemia and Mohr–Tranebjaerg syndrome caused by *BTK* and *TIMM8A* mutations, respectively. The four patients described with the microdeletion spanning *BTK*, *TIMM8A*, *TAF7L* and *DRP2* presented the typical X-linked agammaglobulinaemia and Mohr–Tranebjaerg syndrome phenotypes, but also autism and language delay.[43–45] Interestingly, a recent WES study reported the same mutation (E432*) that we found in *DRP2* in one autistic family, which was absent from a large cohort of controls.[18]

This work represents one of the first comprehensive studies of multiplex families with ASD by exome sequencing. Similar to other studies using WES technologies, known limitations should be considered. First, the coverage of the exome fraction was not complete (83.5% of target sequence covered on average); thus, it is possible that we missed disease-causing mutations. Second, we used stringent criteria to filter false-positives, and as a consequence we may have missed true aetiological variants that were not considered as being pathogenic. Third, we considered only variants shared by affected probands, although aetiological variants transmitted to only one child are also likely to contribute to the disease. Fourth, repetitive elements or variants located on non-coding sequences were not explored in our study. Finally, the sample under study comprises only 10 families.

In conclusion, our data suggest that inherited disrupting mutations in multiplex families may have a major role in the aetiology of ASD. We highlight novel potential ASD candidate genes such as *YWHAZ* and *DRP2*. Further WES studies of inherited rare variants in larger ASD samples are warranted to corroborate these results and to gain more insight into the missing heritability of ASD.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to all families for their participation in our study. We thank Patricia Romarís (Hospital Universitari Mútua de Terrassa) for contributing to clinical delineation of patients and Lara Nonell and Eulàlia Puigdecanet (Servei d'Anàlisi de Micorarrays, IMIM-Hospital del Mar, Parc de Recerca Biomèdica de Barcelona) for their contribution to the CNV studies. Exome sequencing services were provided by the National Centre for Genomic Analysis (CNAG). CT was supported by the European Union (Marie Curie, PIEF-GA-2009-254930) and BT by AGAUR (FI grant). Financial support was received from 'Fundació La Marató de TV3' (092010), 'Fundación Alicia Koplowitz', AGAUR (2009SGR00971) and 'Ministerio de Economía y Competitividad, Spain' (SAF2012-33484, SAF2010-21165).

REFERENCES

1 Lord C, Jones RM. Annual research review: re-thinking the classification of autism spectrum disorders. *J Child Psychol Psychiatry* 2012; **53**: 490–509.
2 Elsabbagh M, Divan G, Koh YJ, Kim YS, Kauchali S, Marcin C *et al.* Global prevalence of autism and other pervasive developmental disorders. *Autism Res* 2012; **5**: 160–179.
3 Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L *et al.* Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 2011; **128**: e488–e495.
4 Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: a decade of new twin studies. *Am J Med Genet B* 2011; **156B**: 255–274.
5 Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D *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.
6 O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012; **338**: 1619–1622.
7 O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* 2012; **485**: 246–250.
8 Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
9 Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**: 285–299.
10 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.

© 2013 Macmillan Publishers Limited

11 Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A *et al.* Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum Mol Genet* 2011; **20**: 3366–3375.
12 Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G *et al.* Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet* 2012; **8**: e1002521.
13 Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995; **25**: 63–77.
14 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al.* Strong association of *de novo* copy number mutations with autism. *Science* 2007; **316**: 445–449.
15 Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 2008; **82**: 477–488.
16 Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS *et al.* Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet* 2012; **8**: e1002635.
17 Puffenberger EG, Jinks RN, Wang H, Xin B, Fiorentini C, Sherman EA *et al.* A homozygous missense mutation in HERC2 associated with global developmental delay and autism spectrum disorder. *Hum Mutat* 2012; **33**: 1639–1646.
18 Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, Macarthur DG *et al.* Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 2013; **77**: 235–242.
19 Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
20 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
21 Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2011; **475**: 101–105.
22 Fuentes Fajardo KV, Adams D, Mason CE, Sincan M, Tifft C, Toro C *et al.* Detecting false-positive signals in exome sequencing. *Hum Mutat* 2012; **33**: 609–613.
23 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**: 1073–1081.
24 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
25 Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, Conesa A *et al.* Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res* 2010; **38**: W210–W213.
26 He WZ, Liu WQ, Zhong XQ, Chen XL, Li SY, Zhang HM *et al.* Analysis of *de novo* copy number variations in a family affected with autism spectrum disorders using high-resolution array-based comparative genomic hybridization. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2012; **29**: 266–269.
27 Guilmatre A, Dubourg C, Mosca AL, Legallic S, Goldenberg A, Drouin-Garraud V *et al.* Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry* 2009; **66**: 947–956.
28 Brooks-Kayal A. Epilepsy and autism spectrum disorders: are there common developmental mechanisms? *Brain Dev* 2010; **32**: 731–738.
29 Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G *et al.* Individual common variants exert weak effects on the risk for autism spectrum disorderspi. *Hum Mol Genet* 2012; **21**: 4781–4792.
30 Kleppe R, Martinez A, Doskeland SO, Haavik J. The 14-3-3 proteins in regulation of cellular metabolism. *Semin Cell Dev Biol* 2011; **22**: 713–719.
31 Foote M, Zhou Y. 14-3-3 Proteins in neurological disorders. *Int J Biochem Mol Biol* 2012; **3**: 152–164.
32 Cardoso C, Leventer RJ, Ward HL, Toyo-Oka K, Chung J, Gross A *et al.* Refinement of a 400-kb critical region allows genotypic differentiation between isolated lissencephaly, Miller–Dieker syndrome, and other phenotypes secondary to deletions of 17p13.3. *Am J Hum Genet* 2003; **72**: 918–930.
33 Bruno DL, Anderlid BM, Lindstrand A, van Ravenswaaij-Arts C, Ganesamoorthy D, Lundin J *et al.* Further molecular and clinical delineation of co-locating 17p13.3 microdeletions and microduplications that show distinctive phenotypes. *J Med Genet* 2010; **47**: 299–311.
34 Capra V, Mirabelli-Badenier M, Stagnaro M, Rossi A, Tassano E, Gimelli S *et al.* Identification of a rare 17p13.3 duplication including the BHLHA9 and YWHAE genes in a family with developmental delay and behavioural problems. *BMC Med Genet* 2012; **13**: 93.
35 Ikeda M, Hikita T, Taya S, Uraguchi-Asaki J, Toyo-oka K, Wynshaw-Boris A *et al.* Identification of YWHAE, a gene encoding 14-3-3epsilon, as a possible susceptibility gene for schizophrenia. *Hum Mol Genet* 2008; **17**: 3212–3222.
36 Middleton FA, Peng L, Lewis DA, Levitt P, Mirnics K. Altered expression of 14-3-3 genes in the prefrontal cortex of subjects with schizophrenia. *Neuropsychopharmacology* 2005; **30**: 974–983.
37 Ramocki MB, Bartnik M, Szafranski P, Kolodziejska KE, Xia Z, Bravo J *et al.* Recurrent distal 7q11.23 deletion including HIP1 and YWHAG identified in patients with intellectual disabilities, epilepsy, and neurobehavioral problems. *Am J Hum Genet* 2010; **87**: 857–865.
38 Grover D, Verma R, Goes FS, Mahon PL, Gershon ES, McMahon FJ *et al.* Family-based association of YWHAH in psychotic bipolar disorder. *Am J Med Genet B* 2009; **150B**: 977–983.
39 Pers TH, Hansen NT, Lage K, Koefoed P, Dworzynski P, Miller ML *et al.* Meta-analysis of heterogeneous data sources for genome-scale identification of risk genes in complex phenotypes. *Genet Epidemiol* 2011; **35**: 318–332.
40 Cheah PS, Ramshaw HS, Thomas PQ, Toyo-Oka K, Xu X, Martin S *et al.* Neurodevelopmental and neuropsychiatric behaviour defects arise from 14-3-3zeta deficiency. *Mol Psychiatry* 2012; **17**: 451–466.
41 English JA, Pennington K, Dunn MJ, Cotter DR. The neuroproteomics of schizophrenia. *Biol Psychiatry* 2011; **69**: 163–172.
42 Sherman DL, Wu LM, Grove M, Gillespie CS, Brophy PJ. Drp2 and periaxin form Cajal bands with dystroglycan but have distinct roles in Schwann cell growth. *J Neurosci* 2012; **32**: 9419–9428.
43 Sediva A, Smith CI, Asplund AC, Hadac J, Janda A, Zeman J *et al.* Contiguous X-chromosome deletion syndrome encompassing the BTK, TIMM8A, TAF7L, and DRP2 genes. *J Clin Immunol* 2007; **27**: 640–646.
44 Jyonouchi H, Geng L, Toruner GA, Vinekar K, Feng D, Fitzgerald-Bocarsly P. Monozygous twins with a microdeletion syndrome involving BTK, DDP1, and two other genes; evidence of intact dendritic cell development and TLR responses. *Eur J Pediatr* 2008; **167**: 317–321.
45 Arai T, Zhao M, Kanegane H, van Zelm MC, Futatani T, Yamada M *et al.* Genetic analysis of contiguous X-chromosome deletion syndrome encompassing the BTK and TIMM8A genes. *J Hum Genet* 2011; **56**: 577–582.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)

© 2013 Macmillan Publishers Limited