Mutation Research 784-785 (2016) 46–52

Contents lists available at ScienceDirect

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

journal homepage: www.elsevier.com/locate/molmut
Community address: www.elsevier.com/locate/mutres

Comprehensive molecular testing in patients with high functioning autism spectrum disorder

Maria Isabel Alvarez-Mora[a,b,c], Rosa Calvo Escalona[d], Olga Puig Navarro[d], Irene Madrigal[a,b,c], Ines Quintela[e], Jorge Amigo[e,f], Dei Martinez-Elurbe[a], Michaela Linder-Lucht[g], Gemma Aznar Lain[g,h], Angel Carracedo[e,f,i], Montserrat Mila[a,b,c], Laia Rodriguez-Revenga[a,b,c,*]

[a] Biochemistry and Molecular Genetics Department, Hospital Clinic, Villarroel 170, 08036 Barcelona, Spain
[b] CIBER of Rare Diseases (CIBERER), Villarroel 170, 08036 Barcelona, Spain
[c] IDIBAPS, Rosselló 149, 08036 Barcelona, Spain
[d] Child and Adolescent Psychiatry Department, Hospital Clinic, SGR-1119, CIBERSAM, Villarroel 170, 08036 Barcelona, Spain
[e] Genomic Medicine Group-Santiago de Compostela University, CEGEN–PRB2–ISCIII, Travesía da Choupana, s/n, 15706 Santiago de Compostela, A Coruña, Spain
[f] Genomic Medicine Group, CIBERER, Fundacion Publica Galega de Medicina Xenomica-SERGAS, Travesía da Choupana, s/n, 15706 Santiago de Compostela, A Coruña, Spain
[g] Pediatric Service, Hospital del Mar, Passeig Marítim, 25-29, 08003 Barcelona, Spain
[h] Autonoma University of Barcelona, Cerdanyola del Vallès, 08193 Barcelona, Spain
[i] King Abdulaziz University, Center of Excellence in Genomic Medicine Research, Abdullah Sulayman, Jeddah 22254, Saudi Arabia

ARTICLE INFO

Article history:
Received 20 August 2015
Received in revised form
28 December 2015
Accepted 30 December 2015
Available online 6 January 2016

Keywords:
Autism spectrum disorders
Candidate genes
Copy number variants
Next generation sequencing
Single nucleotide variants

ABSTRACT

Autism spectrum disorders (ASD) include a range of complex neurodevelopmental disorders with extreme genetic heterogeneity. Exome and target sequencing studies have shown to be an effective tool for the discovery of new ASD genes. The aim of this study was to design an ASD candidate gene panel that covers 44 of the top ASD candidate genes. As a pilot study we performed comprehensive molecular diagnostic testing, including the study of the *FMR1* and *FMR2* repeat regions, copy number variant analysis in a collection of 50 Spanish ASD cases and mutation screening using targeted next generation sequencing-based techniques in 44 out of the total cohort. We evaluated and clinically selected our cohort, with most of the cases having high functioning ASD without facial dysmorphic features. The results of the present study allowed the detection of copy number and single nucleotide variants not yet identified. In addition, our results underscore the difficulty of the molecular diagnosis of ASD and confirm its genetic heterogeneity. The information gained from this and other genetic screenings is necessary to unravel the clinical interpretation of novel variants.

Abbreviations: ADI-R, autism diagnosis interview-revised; ASD, autism spectrum disorder; ASSQ, autism spectrum screening questionnaire; BAP, broader autism phenotype; CGH-array, comparative genomic hybridization microarrays; CNV, copy number variant; DNA, deoxyribonucleic acid; ESP, exome sequencing project; GERP, genomic evolutionary rate profiling; ID, intellectual disability; IQ, intelligence quotient; LOH, loss of heterozygosity; MAPD, median absolute pairwise difference; NGS, next generation sequencing; PCR, polymerase chain reaction; QC, quality control; SNV, single nucleotide variant; SNP, single nucleotide polymorphism; UPD, uniparental disomy; UTR, untranslated region.

* Corresponding author at: Biochemistry and Molecular Genetics Department, Hospital Clínic, C/Villarroel, 170, 08036 Barcelona, Spain. Fax: +34 93 451 52 72.
E-mail addresses: misabel_alvarez@hotmail.com (M.I. Alvarez-Mora), rcalvo@clinic.ub.es (R. Calvo Escalona), opuig@clinic.cat (O. Puig Navarro), imadbajo@clinic.ub.es (I. Madrigal), ines.quintela@usc.es (I. Quintela), jorge.amigo@usc.es (J. Amigo), deielurbe@gmail.com (D. Martinez-Elurbe), MLinder@parcdesalutmar.cat (M. Linder-Lucht), GAznar@parcdesalutmar.cat (G. Aznar Lain), angel.carracedo@usc.es (A. Carracedo), mmila@clinic.ub.es (M. Mila), lbodi@clinic.ub.es (L. Rodriguez-Revenga).

http://dx.doi.org/10.1016/j.mrfmmm.2015.12.006
0027-5107/

1. Introduction

Autism spectrum disorders (ASD) include a range of complex neurodevelopmental disorders characterized by impaired communication and socialization, restricted interests, and stereotypical behavioral patterns. The term ASD is now commonly used to designate this group of highly heterogeneous and complex developmental diseases. ASD is one of the most common neurodevelopmental disabilities, with an average estimated global prevalence of 62 cases per 10,000 children and an approximate 4:1 male to female ratio [1]. The first signs of ASD usually appear by the age of 1–2 years, and it can be clearly detected by 2–4 years of age [2].

Although the causes of ASD have not been completely defined, the strong genetic component of ASD has been evident since early twin and family studies [reviewed in 3]. In fact over the past few years, the genetic basis of ASDs has been aggressively pursued using different types of high-throughput genomic analysis technologies (single nucleotide polymorphism (SNP-array) microarrays, comparative genomic hybridization microarrays (CGH-array), next generation sequencing based techniques (NGS) and genome-wide association studies) [e.g. 4–6]. Despite the progress in the identification of several candidate genes and causative genomic copy number variations (CNVs), the vast majority of ASD cases still remain unexplained. One of the obstacles of achieving the molecular diagnosis of ASD has been the clinical and genetic heterogeneity of patient cohorts in combination with a recently pointed out multiple-hit model of the disease [7–9]. This model is based on the observation that most of the abnormalities identified have been associated with highly variable phenotypes and seem insufficient to cause ASD on their own, supporting the hypothesis that CNVs contribute to ASD in association with other CNVs or point variants located elsewhere in the genome [10]. Therefore, genetic interactions between rare variants probably play an important role in the etiology of ASD. Indeed, nowadays it is currently accepted that autism most likely results from a combination of genetic, epigenetic, and environmental factors [11].

Hundreds of ASD candidate genes have been identified by NGS supporting the extreme locus heterogeneity underlying the genetic etiology of autism [5,8]. With the use of a large-scale resequencing approach, OíRoak et al. [8] recently identified a strong subset of candidate genes that are recurrently mutated in ASD [reviewed in 12]. In an attempt to evaluate the role of these genes, we have developed a gene capture panel including 44 of the top candidate genes. As a pilot study we performed comprehensive diagnostic testing in a collection of 50 Spanish ASD cases, including the study of the *FMR1* and *FMR2* repeat regions, CNV analysis and mutation screening by targeted NGS-based techniques.

2. Material and methods

2.1. Patients

A total of 50 unrelated ASD male patients were included in this study. All the subjects recruited fulfilled the DSM-5 criteria for ASD. Diagnosis was confirmed with the semi-structured interview autism diagnosis interview-revised (ADI-R) [13]. Nearly one third of the sample ($n = 15$) fulfilled criteria for the former Asperger Syndrome (DSM-IV) while the rest were categorized as autistic in the DSM-IV classification. No subjects were in the PDD-NOS category. Symptoms severity was assessed with the High Functioning Autism Spectrum Screening Questionnaire (ASSQ) [14], excluding subjects with severe comorbid conditions, such as schizophrenia and bipolar disorder. Comorbidity was evaluated with the administration of the schedule for affective disorders and schizophrenia for school-age children-present and lifetime versions (K-SADS-PL) semi-structured interview [15]. The mean age of the sample was 10.85 years old (SD ± 3.09), and the average total IQ was 95.70 (SD ± 8.02). Only three subjects had a total IQ below 70.

Written informed consent was obtained from all the patient's parents or legal guardians. The protocol was approved by the Committee for Ethical Issue at Hospital Clinic Barcelona.

DNA extraction was performed from peripheral blood using the Gentra Puregene blood kit (Qiagen Inc., Valencia, CA, USA).

2.2. FMR1 and FMR2 molecular analysis

Molecular analysis of the *FMR1* (FRAXA) and *FMR2* (FRAXE) repeat region of the 50 unrelated male ASD patients was performed by polymerase chain reaction (PCR) amplification using fluorescently labeled primers (upon request). The reaction product was analyzed on an ABI3100 for fragment analysis (Applied Biosystems, Foster City, CA, USA).

2.3. CGH using agilent microarray (180 K CGH + SNP)

500 ng of DNA sample of 44 ASD patients were fluorescently labeled using SureTag Complete DNA Labeling Kit (Agilent Technologies, Santa Clara, CA, USA). Sex-matched labeled DNAs were hybridized to a SurePrint G3Human Genome CGH + SNP Microarray Kit that allows simultaneous detection of CNVs and copy neutral aberrations, such as loss of heterozygosity (LOH) and uniparental disomy (UPD) (PN G4890A, Agilent Technologies, Santa Clara, CA, USA). This array targets ~500 ISCA regions (described in the International Standards for Cytogenomics Array consortium, http://www.ncbi.nlm.nih.gov/projects/dbvar/ISCA/), has a 25 KB backbone probe density and a 5–10 MB LOH/UPD resolution. Afterwards the slides were washed and scanned on an Agilent G2565CA Microarray Scanner System (Agilent Technologies, Santa Clara, CA, USA). Images were analyzed using Cytogenomics software (version 2.0, Agilent Technologies, Santa Clara, CA, USA) and the results were presented on the human genome assembly hg19. A minimum of 3 consecutive oligonucleotides exceeding an absolute log2-ratio threshold of 0.30 were required to identify a CNV.

2.4. CGH using affymetrix microarray (CytoScan HD)

DNA samples of 40 ASD patients were genotyped using the CytoScan High-Density SNP array (Affymetrix, Santa Clara, CA, US). This array contains more than 2.6 million markers across the entire genome, including approximately 750,000 SNPs to detect CNVs, copy-neutral loss of heterozygosity (LOH), uniparental disomy (UPD), regions identical-by-descent and low-level mosaicism. Microarray-based CNV analysis was performed using the Chromosome Analysis Suite software (version 1.2.2, Affymetrix, Santa Clara, CA, USA) and the results were presented on the human genome assembly hg19. The main quality control (QC) parameters were the Median Absolute Pairwise Difference (MAPD) and SNP-QC scores for copy number probes and SNP probes, respectively. Samples with MAPD > 0.27 and SNP-QC < 1.1 for the Cytogenetics array or MAPD > 0.25 and SNP-QC < 15 for the CytoScan array were excluded from the analysis. Only exonic CNVs detected in the array by at least 50 markers with a median intermarker distance of less than 2.5 KB were considered.

2.5. Gene panel

We designed a multiplex, PCR-based primer panel to amplify all the exons, flanking regions (25 bp average), and 5′-3′UTRs of 44 ASD-associated genes using the SureDesign tool (Agilent Technologies, Santa Clara, CA, USA). Genes that conferred more reliability to

risk variants in the recurrence of mutations in unrelated individuals were selected and analyzed in 44 ASD patients of our cohort. The gene names and accession numbers are shown in Supplementary Table 1. Capture of the target regions was performed with reagents from a custom design HaloPlex Target Enrichment kit 1–500 KB (Agilent, Santa Clara, CA, USA). Briefly, genomic DNA was digested into 8 different restriction reactions followed by hybridization to Haloplex probes for target enrichment and sample indexing. In this step, fragments are circularized and sequencing motifs including index sequences are incorporated. Secondly, circularized target DNA-HaloPlex probe hybrids, containing biotin, are captured on streptavidin beads. Finally, PCR amplification of the captured target libraries is performed (HaloPlex Target Enrichment System Protocol, version D.4, March 2013). In order to quantify the enriched target DNA, microfluidic analysis was performed using a Bioanalyzer High Sensitivity DNA Assay kit and the 2100 Bioanalyzer with 2100 Expert Software (Agilent Technologies, Santa Clara, CA, USA). The captured target libraries were pooled at a final concentration of 2000 pM, using concentrations determined for the 175–625 peak of each sample.

2.6. NGS and data analysis

Paired-end sequencing (150 bp reads) of each multiplex library pool was performed on an Illumina MiSeq for 150 cycles, plus a 6 base-index sequence read, using the MiSeq Reagent Kit (Illumina, San Diego, CA, USA). Sequences were aligned to the human genome reference (hg19); base quality score recalibration, local realignment, and variant calling were performed with the DNAnexus Software (DNAnexus, Mountain View, CA, USA) and the Genome Analysis Toolkit (GaTK version 2.1, Broad Institute, Cambridge, MA, USA). Annotation of the variants identified was performed using the ANNOVAR version 2012 March 08. Samples with an average depth of coverage within targets below 30 × were discarded. Variants present in dbSNPv137 (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi) and common variants in the NHLBI Exome sequencing project (ESP6500, http://evs.gs.washington.edu/EVS/) were also discarded. Three *in silico* tools were used to predict mutation pathogenicity: SIFT (http://sift.jcvi.org/), Polyphen2 (http://genetics.bwh.harvard.edu/pph2/), and MutationTaster (http://www.mutationtaster.org/). The SIFT algorithm is based on the evolutionary conservation of the amino acids, giving a SIFT score below 0.05 when the variant is predicted to be damaging and above this value when it is predicted to be tolerated. In contrast, the PolyPhen-2 algorithm uses physical and evolutionary comparative considerations to predict the possible impact on the structure and function of human proteins. The algorithm classifies as benign [0,0.2], possibly damaging [0.2, 0.85], or probably damaging [0.85, 1]; according to the calculated posterior probability of being damaging. The MutationTaster is designed to predict the functional consequences of not only amino acid substitutions but also intronic and synonymous alterations, short insertion and/or deletion (indel) mutations and variants spanning intron-exon borders. The MutationTaster integrates information from different biomedical databases and uses established analysis tools to analyze comprise evolutionary conservation, splice-site changes, loss of protein features and changes that might affect the amount of mRNA. Test results are then evaluated by a naive Bayes classifier 2, which predicts the disease potential.

Finally, genomic evolutionary rate profiling (GERP) was applied in order to quantify the position-specific constraint on each site in the genome based on an alignment of 35 mammals to hg19. Rejected substitution (RS) scores vary with alignment depth and the level of sequence conservation, ranging from a maximum of 6.18 to a minimum of −12.36. An RS score threshold of 2 provides high sensitivity while still strongly enriching truly constrained sites.

2.7. Sanger sequencing

Single nucleotide variants (SNVs) were confirmed by Sanger sequencing. PCR primers were designed using the Primer3 Input version 4.0.0 web tool (http://primer3.ut.ee/). PCR products were directly sequenced using the BigDye® Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). The reaction was run in an ABI Prism 3100XL automated sequencer (Applied Biosystems, Foster City, CA, USA) and the results were analyzed with SEQUENCE® Pilot version 4.0.1 software (JSI medical systems GmbH, USA).

3. Results

3.1. FRAXA and FRAXE locus analysis

FMR1 (FRAXA) CGG repeat size and *FMR2* (FRAXE) CCG repeats were determined in the 50 male ASD patients. The mean CGG repeat number for *FMR1* gene was 28.57 ± 4.13 (mean ± SD), and the mean CCG repeat number for the *FMR2* gene was 9 ± 3.71 (mean ± SD). A full mutation was not detected in either the FRAXA or the FRAXE locus. However, we did identify one intermediate *FMR1* allele of 53CGGs (with two AGG interruptions) (ASD-35).

3.2. CNVs detected

A total of 50 samples from male ASD patients were tested by chromosomal microarray. Thirty-four of these samples were analyzed by both the Agilent microarray (180 K CGH + SNP) and the Affymetrix microarray (CytoScan HD), whereas 16 were screened by one or the two platforms. Although no relevant LOH or UPD regions were identified, five different heterozygous CNVs were classified as 'clinically significant' or 'likely clinically significant' (Table 1) . CNV assessment was performed as described elsewhere [16].

The 15q11.2 duplication involving the BP1-BP2 region of chromosome 15 was detected in two cases (ASD-12 and ASD-58). Inheritance patterns showed the BP1-BP2CNV to be paternally inherited in both cases. Clinical evaluation was available for ASD-12 father, with traits meeting criteria for a broader autism phenotype (BAP).

Analysis of ASD-02 revealed a chromosome 10q25.1 duplication affecting the *SORCS3* gene. Familial study identified the same CNV in the father and in two siblings (Fig. 1a). While the father presented a BAP, the two siblings were not diagnosed with ASD but one did present dyslexia (learning disability) and the other has recently been diagnosed with moderate depression and social phobia.

In case ASD-26, an intragenic deletion involving exons 6–8 of the *DLGAP2* gene was detected. Finally, a *de novo* Xp22.33/Yp11.32p11.31 duplication was detected in a child presenting the lowest IQ (total IQ = 64), less language development and higher scores on the ASSQ and all dimensions of the ADI-R among all the patients in whom CNVs were detected.

3.3. Gene panel data output and variant calling

The targeted NGS panel was designed to amplify all the exons of 44 known ASD-associated genes (Supplementary Table 1). Variant calling identified a range of 12–48 SNVs per patient after filtering out the common variants present in dbSNP and the 1000 Genome project with MAF > 1%. A small number of variants were selected for validation by Sanger sequencing after excluding intronic and synonymous variants. This prioritization process significantly reduced

Table 1
Summary of abnormal microarray data for autism spectrum disorder.

Case ID	Genomic imbalance	Size (MB)	Origin	Selected gene(s)
ASD-02	arr[hg19] 10q25.1(106,545,364-107,763,181)x3	1.2	pat	SORCS3
ASD-12	arr[hg19]15q11.2(21,886,479-23,217,514)x3	1.33	pat	NIPA1, NIPA2, CYFIP1, TUBGCP5
ASD-26	arr[hg19] 8p23.3(1,603,885-1,625,406)x1	0.22	ND	DLGAP2
ASD-58	arr[hg19] 15q11.2 (22,770,421-23,288,350)x3	0.52	pat	NIPA1, NIPA2, CYFIP1, TUBGCP5
ASD-61	arr[hg19] Xp22.33/Yp11.32p11.31(168,546-2,703,633)x3	2.53	de novo	SHOX

ND: not determined; pat: paternal.

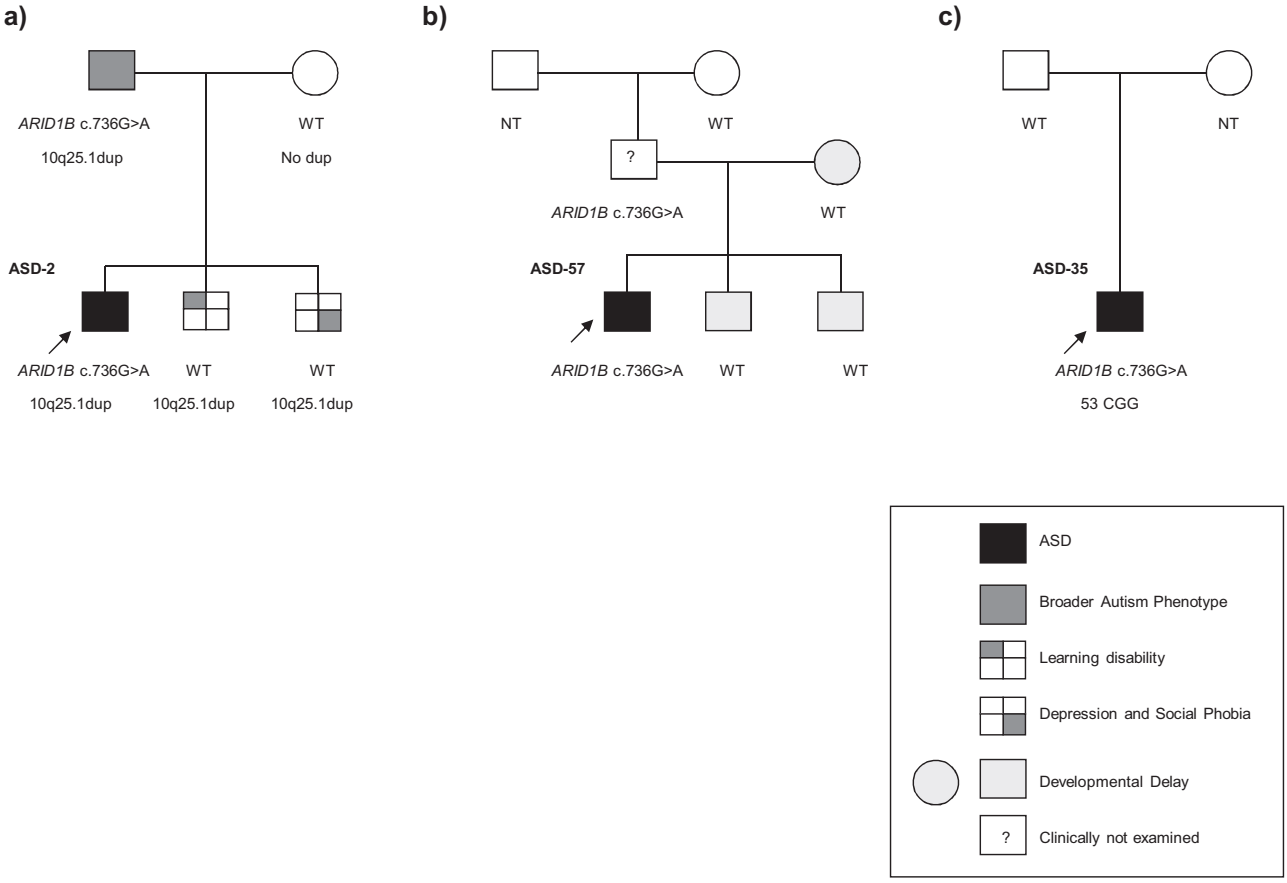

Fig. 1. Family pedigrees of probands with the *ARID1B* c.736G > A variant. Circles indicate females and squares indicate males. The arrow indicates the index case. WT = wild type, NT = not tested.

the number of candidate SNVs, to an average of nearly one variant per sample.

3.4. Candidate rare SNVs associated with ASD

We identified 22 rare heterozygous SNVs (21 missense and 1 indel) in 21 of the 44 ASD patients included in the study. All these variants have been confirmed by Sanger sequencing, and their inheritance was tested if parental samples were available (Table 2). Inheritance was determined in the majority of the cases (14/21). In 6 cases the variant was maternally inherited, 3 from a clinically affected mother and 3 from an unaffected mother. In 7 cases the SNV was paternally inherited with all of the fathers being clinically affected except for one who was not available for clinical evaluation. In one case (ASD-22), the variant identified, an in frame insertion predicted as a polymorphism by the Mutation Taster was determined to be *de novo*.

Among the 22 rare heterozygous SNVs identified, six (27%) were determined to be probably pathogenic by three or more *in silico* prediction algorithms (Table 2). After filtering by this stringent criteria *FOXG1* p.P486S, *CNTNAP2* p.D1073 V, *ADNP* p.T443A, *SCN1A* p.R604H, *RELN* p.Y1183C and *FOXP2* p.Q142H were the six SNVs found to likely contribute to the ASD phenotype (Table 2). Interestingly the *ARID1B* c.736G > A variant was detected in 3 out of the 21 patients (ASD-2, ASD-35 and ASD-57). Paternal inheritance was determined in ASD-2 and ASD-57 (Fig. 1a and b). Clinical evaluation was only available for the ASD-2 father, revealing a BAP. Remarkably, the ASD-2 patient and his father also presented a CNV affecting the *SORCS3* gene (Table 1). As mentioned previously, familial studies showed that the duplication, but not the *ARID1B* c.736G > A variant, was also inherited in two siblings presenting different clinical phenotypes (Fig. 1a). Regarding ASD-57, familial study revealed that only the index case inherited this rare variant. Despite presenting clinical involvement, the clinical phenotypes of the two other siblings as well as their mother differed, including developmental delay and learning disabilities (Fig. 1b). Finally, inheritance was not completely determined in case ASD-35 since maternal sample was not available. However, Sanger sequencing in the unaffected

Table 2
Rare SNVs identified with amino acid changes, origin and allele frequencies from control population. Computational predictions of pathogenicity from rare SNVs are shown. SNVs deemed probably pathogenic by three or more *in silico* prediction algorithms are highlighted in bold.

ID	Gene	RefSeq	AA substitution	Origin	rsIDsnp137	EA ESP data*	SIFT	POLYPHEN	Mutation taster	GERP score
ASD-2	*ARID1B*	NM_017519	G246S	pat[a]	rs375160616	No data	deleterious	unknown	polymorphism	1.31
ASD-7	*MECP2*	NM_001110792	A456T	ND	rs61753975	0.001	tolerated	benign	polymorphism	2.72
ASD-10	*FOXG1*	NM_005249	**P486S**	mat[a]	–	–	deleterious	probably damaging	disease causing	4.28
ASD-11	*KDM5C*	NM_004187	M1119L	ND	–	–	tolerated	benign	polymorphism	4.78
ASD-14	*EHMT1*	NM_001145527	A688T	ND	–	–	tolerated	benign	polymorphism	−9.31
ASD-15	*CYFIP1GABRB3*	NM_001033028NM_021912	A96V P11S	pat[a] pat[a]	– rs25409	– 0.005	– –	benign benign	polymorphism polymorphism	−0.759 2.18
ASD-17	*CNTNAP2*	NM_014141	**D1073V**	pat[a]	–	–	deleterious	probably damaging	disease causing	5.25
ASD-20	*SETD2*	NM_014159	S2060P	pat[a]	–	–	deleterious	possibly damaging	polymorphism	3.39
ASD-21	*SBF1 CACNA1C*	NM_002972NM_001129837	D1619E V2019I	mat[a] mat[a]	–rs193922615	– 0.00	toleratedtolerated	possibly damagingbenign	disease causingpolymorphism	−0.06713.18
ASD-22	*MED12*	NM_005120	Q2113delinsQQQHQ	de novo	–	–	–	–	polymorphism	−3.57
ASD-23	*ADNP*	NM_181442	**T443A**	ND	–	–	deleterious	probably damaging	disease causing	6.01
ASD-25	*SCN1A*	NM_001165963	**R604H**	mat	rs121918769	0.001	tolerated	probably damaging	disease causing	5.37
ASD-27	*LAMC3 EHMT1*	NM_006059 NM_024757	P1522L R1016Q	mat[a] mat[a]	– –	– –	toleratedtolerated	benignpossibly damaging	polymorphismdisease causing	−3.42 4.99
ASD-33	*CTNNB1*	NM_001098209	G752A	ND	rs373158451	$1.16e^{-4}$	tolerated	benign	disease causing	5.68
ASD-35	*ARID1B*	NM_017519	G246S	ND	rs375160616	No data	deleterious	unknown	polymorphism	1.31
ASD-40	*LAMC3*	NM_006059	A972T	ND	–	–	tolerated	benign	polymorphism	−3.24
ASD-44	*ARID1B*	NM_017519	S41F	pat[a]	–	–	deleterious	unknown	polymorphism	0.779
ASD-51	*LAMC3*	NM_006059	A1232T	mat	–	–	tolerated	possibly damaging	disease causing	3.93
ASD-55	*RELN*	NM_005045	**Y1183C**	mat	–	–	tolerated	probably damaging	disease causing	5.75
ASD-56	*FOXP2*	NM_001172767	**Q142H**	pat[a]	–	–	deleterious	probably damaging	disease causing	2
ASD-57	*ARID1B*	NM_017519	G246S	pat	rs375160616	No data	deleterious	unknown	polymorphism	1.31

ND: not determined; pat: paternally inherited; mat: maternally inherited.
Assembled to the human genome reference CRGh37/hg19.

* EuropeanAmerican (EA) allele frequency based ESP data (6500 Exome Project).

[a] Transmitting parents presenting a broader autism phenotype.

father discarded a paternal origin. Apart from the *ARID1B* c.736G > A variant, the ASD-35 patient presented an intermediate *FMR1* allele (53CGGs) (Fig. 1c).

Clinical re-evaluation of the three patients (ASD-2, ASD-35 and ASD-57) harboring the *ARID1B* c.736G > A variant revealed that apart from the ASD their clinical features were similar with a coarse face with thick eyebrows and prominent eyelashes, a large mouth and alae nasi which were thicker than usual. One also presented agenesia of the left teste. None of these three children had intellectual disability (ID), having average IQ of 104.82 (SD ± 2.34); however presented delayed speech development. On the other hand, EEG tests were normal with no history of seizures.

Among the 22 rare SNVs detected, only the *SCN1A* C.1811G > A variant has been previously reported and is associated with intractable childhood epilepsy [17,18]. This variant was detected in the ASD-25 patient and his unaffected mother. This patient is currently 10 years old and to date has no history of epileptic seizures.

4. Discussion

Understanding of the molecular drivers of ASD has evolved thanks to the emergence of high-throughput techniques such as whole genome microarrays and NGS. Many studies have demonstrated the important contribution of rare CNVs to the genetic puzzle of ASD, estimating that 5–10% of sporadic ASD cases carry a *de novo* CNV [19,20]. Genes involved in CNVs are considered to have dose-dependent rather than severe loss-of-function effects on the behavior of affected individuals. Furthermore, since pathogenic CNVs implicating ASD-associated genes often show variable expressivity, a complex disease/rare variant model for ASD has been proposed, in which a proportion of etiologic risk is conferred by very rare variants and *de novo* mutations [reviewed in 12]. However, it has recently been described that common SNPs make an important genetic contribution to the overall variance observed in ASD and other psychiatric disorders [6,21].

Exome and whole-genome sequencing studies have estimated that *de novo* SNV events probably explain ~6% of the overall variance in autism risk [5,8,22] and that an additional 5% may be explained by rare inherited recessive or X-linked loss-of-function SNVs [23,24]. Although these studies have provided a catalogue of most of the variations in the genome, the profound locus heterogeneity of ASD makes it challenging to distinguish variants that confer risk from the background noise of non significant SNVs. However, despite this heterogeneity, targeted large-scale resequencing studies have confirmed the significance of specific loci [8,25]. Furthermore, it has been shown that many of these genes are biologically related and converge in interconnected functional pathways [5,8,26]. On the basis of these observations, we aimed to design and validate an ASD candidate gene panel following a pilot study in which we performed a comprehensive molecular analysis of a cohort of 50 Spanish ASD patients. Taking into account the great phenotypic heterogeneity of ASD, including mild to severe levels of cognitive impairment often accompanied by seizures and other medical problems, we carefully evaluated and clinically selected our cohort, with the majority of cases presenting high cognitive functioning ASD without dysmorphologies.

Since a relatively high percentage of FRAXA and FRAXE patients present ASD traits [27], we first discarded *FMR1* and *FMR2* repeat expansions. An intermediate *FMR1* allele of 53CGGs was detected in one patient (ASD-35). Although several studies have reported an excess of intermediate *FMR1* alleles in patients with cognitive and/or behavioral phenotypes [28,29], the frequency herein observed (2%) was similar to that reported in the general population [30].

Secondly we performed a whole-genome CNV analysis and identified only one *de novo* duplication (~253 KB) at Xp22.33/Yp11.32-p11.31 in one patient (ASD-61) in whom the severity of autistic symptoms and cognitive dysfunction was higher than the mean of the rest of the cohort. Moreover, as reported previously, chromosome 15 was the most common chromosome involved in the ASD group. The BP1-BP2 genomic interval within band 15q11.2 was found to be duplicated in two cases (4%), with the duplication showing paternally inheritance. Similar to most of the cases reported in the literature [31]. The effect of the BP1-BP2 duplication on ASD is controversial and according to the current American College of Medical Genetics guidelines for interpretation of postnatal CNV [32], this CNV remains classified as a variant of unknown significance.

Finally, we identified CNVs involving the two brain-expressed genes *SORCS3* and *DLGAP2*. Only one case of a duplication affecting *SORCS3* and rare genetic variants of *DLGAP2* has previously been implicated in the pathogenesis of ASD [33,34]. The CNV affecting the *SORCS3* gene was detected in patient ASD-02, who also carried a rare SNV in the *ARID1B* gene (Fig. 1a, Table 2). Familial study showed that while the affected patient inherited both the CNV and the SNV from an affected father, two other siblings only inherited the CNV. These siblings do not meet ASD criteria, despite being clinically affected and presenting learning disabilities and social anxiety disorder. Although further studies are necessary in order to attribute a causal role to these rare variants, the co-occurrence of CNVs and rare SNVs altering neuronal genes may play a role in the genetic susceptibility to ASD [10,35]. To date, the number of loci needed to regulate synaptic homeostasis and how these variants interact with each other to modulate the risk for ASD remain unclear [9].

Massively parallel sequencing identified rare SNVs predicted to be deleterious in 13.6% (6/44) of the patients. Although the targeted gene panel designed included strong ASD candidate genes such as *CHD8*, *GRIN2B*, *DYRK1A*, no rare variants were detected among them. Furthermore, contrary to what has previously been described in family-based exome studies of ASD and ID [12][reviewed in 12], most of the variants identified were inherited and not disruptive with respect to the biological function of the protein. Although our patients were diagnosed with autism on the basis of the current gold-standard criteria, their cognitive functioning was higher than that reported in other cohorts. Therefore, it can be argued that the SNVs detected in this study may have milder defects than the *de novo* risk variants reported in family-based exome studies of ASD.

Among the 44 genes included in the panel (Supplementary Table 1), different SNVs we observed in three: *EHMT1*, *LAMC3* and *ARID1B*. In particular, the *ARID1B* c.736G > A variant was identified in three unrelated cases. Further familial studies showed that this variant was inherited in two cases (ASD-2 and ASD-57, Fig. 1a and b). This variant has not been described in over 6500 exomes in the ESP6500. However, it has been recorded with different MAFs in other databases in a European population (dbSNP, rs#375160616: MAF 0.002 [A], and Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org/) rs#375160616: MAF 0.01 [A]). In the 1000 Genomes Project Phase 3, this variant was recorded in a Spanish population with a frequency of 0.005. Even though only one *in silico* tool predicted this variant to be deleterious (Table 2), its frequency was higher in our ASD cohort (3.4%) compared to that reported in the 1000 Genomes Project Phase 3 in a Spanish population (0.5%). Over the last few years *ARID1B* has been recognized as a major ID gene. All the pathogenic mutations reported to date have been truncating heterozygous mutations associated with ID and a diagnosis of Coffin-Siris syndrome [reviewed in 36]. Clinical re-evaluation of the three patients carrying the *ARID1B* c.736G > A variant revealed additional features such as, delayed speech development and dysmorphic traits including a coarse face,

thick eyebrows and prominent eyelashes. Although the *ARID1B* phenotype varies greatly, these features are frequently observed in other patients with *ARID1B* mutations [36]. Further research is needed to investigate the implication of the *ARID1B* c.736G > A variant in the Spanish ASD population.

Among all the rare variants detected, only the *SCN1A* c.1811G > A (ASD-25) has previously been described and associated with intractable childhood epilepsy [17,18]. However, none of the individuals (ASD-25 and his unaffected mother) carrying the *SCN1A* c.1811G > A variant had a history of seizures. Moreover, since this variant has been described in the ExAC (rs#121918769: MAF 0.001[A]), it could be considered insufficient to cause seizures.

This study has 3 main limitations: first, the limited number of patients enrolled in the study; second, the limited number of genes selected for the study, and finally, the inability to distinguish between rare variants unrelated to disease and those that may confer risk, highlighting the difficulties involved in the clinical interpretation of novel variants in ASD. Variant databases of healthy controls are still necessary in order to unravel the clinical implications of novel variants since the frequency and distribution of these variants have not been well characterized in matched unaffected controls. Furthermore, close follow-up of individuals and families and studies in additional patients are required to interpret the clinical significance of these variants identified by NGS techniques.

5. Conclusions

Collectively, our results underscore the difficulty of the molecular diagnosis of ASD and confirm its genetic heterogeneity. Although comprehensive molecular diagnosis of ASD is now within reach, our results demonstrate that the application of NGS with targeted gene panels in children with high cognitive functioning ASD has a low diagnostic yield due to the challenge of distinguishing between rare variants unrelated to disease and those that are contributory. Nevertheless, information gained from this and other genetic screenings is important for better interpreting novel variants of ASD.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We thank all the patients and families for their cooperation. This work was supported by Fundación Alicia Koplowitz (AKOPLOWITZ11_006), AGAUR 2014-SGR603. The CIBERER is an initiative of the ISCIII. Genotyping services were provided by the "Centro Nacional de Genotipado – PRB2-Instituto de Salud Carlos III (CeGen-ISCIII)"

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.mrfmmm.2015.12.006.

References

[1] M. Elsabbagh, et al., Global prevalence of autism and other pervasive developmental disorders, Autism Res. 5 (3) (2012) 160–179.
[2] E. Courchesne, et al., Mapping early brain development in autism, Neuron 56 (2) (2007) 399–413.
[3] A. Ronald, R.A. Hoekstra, Autism spectrum disorders and autistic traits: a decade of new twin studies, Am. J. Med. Genet. B Neuropsychiatr. Genet. 156 (B(3)) (2011) 255–274.
[4] J. Sebat, et al., Strong association of de novo copy number mutations with autism, Science 316 (5823) (2007) 445–449.
[5] S.J. Sanders, et al., De novo mutations revealed by whole-exome sequencing are strongly associated with autism, Nature 485 (7397) (2012) 237–241.
[6] Cross-Disorder Group of the Psychiatric Genomics Consortium, et al., Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs, Nat. Genet. 45 (9) (2013) 984–994.
[7] C.P. Schaaf, et al., Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders, Hum. Mol. Genet. 20 (17) (2011) 3366–3375.
[8] B.J. O'Roak, et al., Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders, Science 338 (6114) (2012) 1619–1622.
[9] C.S. Leblond, et al., Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders, PLoS Genet. 8 (2) (2012) e1002521.
[10] S. Girirajan, E.E. Eichler, Phenotypic variability and genetic susceptibility to genomic disorders, Hum. Mol. Genet. 19 (R2) (2010) R176–187.
[11] S.M. Engel, J.L. Daniels, On the complex relationship between genes and environment in the etiology of autism, Epidemiology 22 (4) (2011) 486–488.
[12] N. Krumm, et al., A de novo convergence of autism genetics and molecular neuroscience, Trends Neurosci. 37 (2) (2014) 95–105.
[13] C. Lord, et al., Diagnosing autism: analyses of data from the autism diagnostic interview, J. Autism Dev. Disord. 27 (5) (1997) 501–517.
[14] S. Ehlers, et al., a screening questionnaire for asperger syndrome and other high-functioning autism spectrum disorders in school age children, J. Autism. Dev. Disord. 29 (2) (1999) 129–141.
[15] J. Kaufman, et al., Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data, J. Am. Acad. Child Adolesc. Psychiatry 36 (7) (1997) 980–988.
[16] D.T. Miller, et al., Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies, Am. J. Hum. Genet. 86 (5) (2010) 749–764.
[17] A. Escayg, et al., A novel SCN1A mutation associated with generalized epilepsy with febrile seizures plus–and prevalence of variants in patients with epilepsy, Am. J. Hum. Genet. 68 (4) (2001) 866–873.
[18] N. Fendri-Kriaa, et al., Genetic screening of two Tunisian families with generalized epilepsy with febrile seizures plus, Eur J Neurol 16 (6) (2009) 697–704.
[19] D. Levy, et al., Rare de novo and transmitted copy-number variation in autistic spectrum disorders, Neuron 70 (5) (2011) 886–897.
[20] S.J. Sanders, et al., Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism, Neuron 70 (5) (2011) 863–885.
[21] T. Gaugler, et al., Most genetic risk for autism resides with common variation, Nat. Genet. 46 (8) (2014) 881–885.
[22] B.M. Neale, et al., Patterns and rates of exonic de novo mutations in autism spectrum disorders, Nature 485 (7397) (2012) 242–245.
[23] T.W. Yu, et al., Using whole-exome sequencing to identify inherited causes of autism, Neuron 77 (2) (2013) 259–273.
[24] E.T. Lim, et al., Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders, Neuron 77 (2) (2013) 235–242.
[25] J. de Ligt, et al., Diagnostic exome sequencing in persons with severe intellectual disability, N. Engl. J. Med. 367 (20) (2012) 1921–1929.
[26] D. Pinto, et al., Convergence of genes and cellular pathways dysregulated in autism spectrum disorders, Am. J. Hum. Genet. 94 (5) (2014) 677–694.
[27] M.K. Belmonte, T. Bourgeron, X. Fragile, Syndrome and autism at the intersection of genetic and neural networks, Nat. Neurosci. 9 (10) (2006) 1221–1225.
[28] L.A. Haddad, et al., Fully mutated and gray-zone FRAXA alleles in Brazilian mentally retarded boys, Am. J. Med. Genet. 84 (3) (1999) 198–201.
[29] S.A. Youings, et al., FRAXA and FRAXE: the results of a five year survey, J. Med. Genet. 37 (6) (2000) 415–421.
[30] I. Madrigal, et al., Intermediate FMR1 alleles and cognitive and/or behavioural phenotypes, Eur. J. Hum. Genet. 19 (8) (2011) 921–923.
[31] P. Chaste, et al., Modest impact on risk for autism spectrum disorder of rare copy number variants at 15q11.2, specifically breakpoints 1 to 2, Autism Res. 7 (3) (2014) 355–362.
[32] H.M. Kearney, et al., American college of medical genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants, Genet. Med. 13 (7) (2011) 680–685.
[33] A.C. Lionel, et al., Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD, Sci. Transl. Med. 3 (95) (2011) 95ra75.
[34] W.H. Chien, et al., Deep exon resequencing of DLGAP2 as a candidate gene of autism spectrum disorders, Mol. Autism 4 (1) (2013) 26.
[35] D. Pinto, et al., Functional impact of global rare copy number variation in autism spectrum disorders, Nature 466 (2010) 368–372.
[36] G.W. Santen, et al., The ARID1B phenotype: what we have learned so far, Am. J. Med. Genet. C Semin. Med. Genet. 166 (3) (2014) 276–289.