Clinica Chimica Acta 479 (2018) 144–147

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Two de novo variations identified by massively parallel sequencing in 13 Chinese families with children diagnosed with autism spectrum disorder☆

Shi-Jun Li[a,*,1], Shan-shan Yu[b,**,1], Hong-yu Luo[b], Xin Li[b], Bin Rao[b], Yi Wang[c], Zhen-zhen Li[a], Gang Liu[d], Li-Ping Zou[e], Ji-Shui Zhang[f], Chen Feng[e], Jing Liu[g], Jian-Wei Liu[g], Nan Hu[h], Xiao-Qiao Chen[f], Sheng-Yuan Yu[i], Ke Li[i], Mian-Wang He[i], Xin-Guang Yu[j], Jun Wang[j], Sheng-Li Guo[j], Zhi-Ye Chen[d], Lei Zhang[k], Lin Ma[d,***]

[a] Department of Medical Instruments, Chinese PLA General Hospital, Beijing 100853, China
[b] BGI-shenzhen, Shenzhen 518083, China
[c] Department of Stomatology, Chinese PLA General Hospital, Beijing 100853, China
[d] Department of Radiology, Chinese PLA General Hospital, Beijing 100853, China
[e] Department of Pediatrics, Chinese PLA General Hospital, Beijing 100853, China
[f] Department of Neurology, Beijing Children's Hospital of Capital Medical University, Beijing 100045, China
[g] Institute of Geriatric Medicine, Chinese PLA General Hospital, Beijing 100853, China
[h] Department of Rehabilitation Medicine, Chinese PLA General Hospital, Beijing 100853, China
[i] Department of Neurology, Chinese PLA General Hospital, Beijing 100853, China
[j] Department of Neurosurgery, Chinese PLA General Hospital, Beijing 100853, China
[k] Department of Medical Information, Chinese PLA General Hospital, Beijing 100853, China

ARTICLE INFO

Keywords:
Autism spectrum disorder
Neurodevelopmental disorder
Chromosomal variants
De novo mutations
Massively parallel sequencing

ABSTRACT

Autism spectrum disorder (ASD) is a genetically heterogeneous neurodevelopmental disorder characterized by impairments in social interaction and communication, and by restricted and repetitive behaviors. The genetic architecture of ASD has been elucidated, including chromosomal rearrangements, de novo or inherited rare variants, and copy number variants. However, the genetic mechanism of Chinese families with ASD children is explored rarely. To identify genetic pathogenesis, we performed massively parallel sequencing on 13 Chinese ASD trio families, and found two de novo variations. The novel de novo splice alteration c.664 + 2T > G in the *DEAF1* gene and the novel de novo missense mutation c.95 C > T in the *AADAT* gene associated with ASD may be important clues for exploring the etiology of this disorder.

1. Introduction

Autism spectrum disorder (ASD) comprises a group of heterogeneous neurodevelopmental conditions affecting about 1% of the population, characterized by early-onset difficulties in social interaction and communication, and by restricted, repetitive behaviors and interests [1]. ASD often co-occurs with other neurological disorders, such as intellectual disability (ID), epilepsy, and schizophrenia [2]. The clinical heterogeneity of ASD is believed to reflect at least in part its heterogeneous genetic complexity [3]. To date, about 10%–20% of individuals with ASD have an identified genetic etiology, and copy number variations (CNVs) have been reported in 5%–10% of cases [4]. In addition, single-nucleotide variants (SNVs) and very small deletions and duplications (indels) have a substantial impact on risk [4]. Over the past three years, next-generation sequencing (NGS) of large ASD cohorts has revolutionized gene discovery. Currently, an estimated ~1000 genes are likely to be involved in ASD, and there are > 100 already identified recurrent genetic defects that can cause ASD [4,5]. These recent studies have pinpointed risk-conferring genes largely by focusing on de novo disrupting or deleterious variations because they can highly impact liability and are sufficiently rare that recurrent mutations in a gene provide strong support for causality [6–8]. Moreover, de novo

☆ This work was funded by National Natural Science Foundation of China (31100714) and Beijing Nova Program (2011112).
* Correspondence to: S.-J. Li, Department of Medical Instruments, Fuxing Road 28th, Beijing, China.
** Correspondence to: S.-s. Yu, BGI-shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, China.
*** Corresponding author.
E-mail addresses: shijunli07@yeah.net (S.-J. Li), yushan_mail@sina.cn (S.-s. Yu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cca.2018.01.025
Received 2 August 2017; Received in revised form 10 January 2018; Accepted 16 January 2018

protein-altering coding variants are more frequent in ASD individuals than in controls [6,8].

In this study, we performed massively parallel sequencing on 13 Chinese ASD trio families to explore genetic etiology. We applied comprehensive annotation to identify CNVs and SNVs or indels that may confer ASD risk. Among the variants, a novel de novo splice alteration in *DEAF1* and a novel de novo missense mutation in *AADAT* were associated with ASD.

2. Materials and methods

2.1. Participants

Children with ASD were recruited at the Department of Pediatrics of the Chinese PLA General Hospital for Autism Research. Exclusion criteria included metal implants, psychiatric or neurological disorders, structural brain abnormalities, or known genetic conditions. Informed consent and assent were obtained from subjects prior to participation in the study according to protocols approved by the Medical Ethics Committee of the Chinese PLA General Hospital. Thirteen families with preschool children with ASD, including 12 males and 1 female (mean age 4.56 ± 0.97), were recruited in the study as cases. Diagnosis was made according to Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) scores (8.00 ± 0.50) (Table 1). All participants had normal or corrected-to-normal vision, and no use of psychotropic medications was reported in any of the children.

2.2. Whole-genome low-coverage sequencing and analysis

Genomic DNA was extracted from whole blood from affected children and their parents. DNA (150 ng) was sheared to sizes between 200 and 250 base pairs (bp) and then repaired. The fragments were ligated to adapters after the addition of terminal A nucleotides. The adapter-ligated DNA underwent 10 cycles of PCR and was purified by an Agencourt AM Pure 450 mL Kit. The prepared DNA libraries were sequenced on Illumina HiSeq 2000 or 2500 analyzers to generate about 15 M50-bp single-end reads. Then, we used the population-scale CNV calling (PSCC) method previously described by Xuchao Li et al. to call CNVs [9]. CNVs were visualized along each chromosome with corresponding NCBI annotated gene information. To interpret the CNVs, the following databases were used: the Database of Genomic Variants (DGV) (http://dgv.tcag.ca/dgv/app/home), DECIPHER (https://decipher.sanger.ac.uk/syndromes#syndromes/overview), and the SFARI Gene database (https://gene.sfari.org/autdb/CNVHome.do).

2.3. Targeted sequencing and variant calling

We designed a custom panel (Roche NimbleGen, Inc., Madison, WI, USA) that captures 1331 genes known to be associated with ASD, and performed trio-base targeted sequencing. Approximately 3 μg of DNA was sheared into 200–250 bp fragments, which were then repaired, ligated to adaptors, and purified for subsequent PCR amplification. Amplified products were then captured using a solution hybrid selection method, in accordance with the manufacturer's instructions. After bound DNA had been isolated and re-amplified, the final prepared products were sequenced on Illumina HiSeq 2000 or 2500 analyzers with 90-bp paired-end reads. Image analysis and base calling were performed using the Illumina Pipeline. The human reference genome was obtained from the NCBI database, version hg19. BWA software (version 0.7.12-r1039) was used to align filtered reads to the hg19 reference genome, and both Picard (http://picard.sourceforge.net/) and GATK (version 2.8-1) software were used for marking PCR duplicates, sequence realignment, and quality score recalibration. Single-nucleotide polymorphisms (SNPs) and indels were called with the GATK HaplotypeCaller genotyper in single sample mode. The GATK Variant Filtration Genotyper was used for hard-filtering variants. The functional effects of variants were predicted by SIFT (http://sift.bii.a-star.edu.sg/index.html), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), LRT (http://www.genetics.wustl.edu/jflab/lrt_query.html), and Mutation Taster (http://mutationtaster.org) software.

Table 1
Demographic and neuropsychological characteristics of all subjects.

	Min.	Max.	Median	QL
DSM	7.0	8.0	8.0	1.0
CARS	23.0	40.0	30.0	8.0
ABC	41.0	70.0	58.0	11.5

3. Results

3.1. CNV analysis

Whole-genome low-coverage sequencing was performed to identify CNVs in ASD families. We successfully detected 20 CNVs that were ≥100 kb in all probands (Table S1). We assessed the impact of rare CNVs in cases based on the following conditions: CNV inheritance status, estimated CNV size, key genetic content of the region, the association of the CNV with reported cases, and whether CNVs overlapped in the Database of Genomic Variants (DGV) [10]. Of the CNVs detected in these 13 cases, 5 were de novo (Table 2). However, no CNVs explained the clinical phenotype.

3.2. Targeted sequencing and de novo variants

We designed a custom panel to capture the exons of 1331 genes associated with ASD and performed sequencing on 13 ASD families, obtaining an average yield mapped to a target region of 809 Mb for each sample, which represented an average depth of 137×. More than 92% of targeted exon regions had at least 30× coverage, which is sufficient for variant calling.

To identify disease-causing mutations, we applied a three-step procedure. First, we filtered out the variants located in introns, except for potential splice sites. Next, we screened rare variants for which the allele frequency was < 5% in the dbSNP, HapMap, 1000 Genomes, and local (100 Chinese healthy) databases. Third, the remaining variants were screened according to three different models of inheritance that may explain the proband phenotypes: autosomal recessive inheritance including homozygous variants or compound heterozygous variants, X chromosome-linked inheritance, and a de novo model.

We identified 563 variants in all 13 ASD families, including 486 autosomal recessive variants, 75 X-linked variants, and 2 de novo variants (Table S2). We then focused our analyses on variants that lead to protein-coding changes, that is, stop (nonsense), missense variants, exonic insertions/deletions (especially frameshift indels), and variants at potential splice sites, to best identify functional variation. Among these variations, two de novo mutations were the most plausible candidates to explain the phenotype exhibited by the probands. Neither mutation was observed in the dbSNP, HapMap, 1000 Genomes, or local databases, and they were confirmed by Sanger sequencing. The genetic variants of the probands and both parents are shown in Fig. 1.

The intronic T > G variant in *DEAF1* (c.664 + 2T > G, NM_021008.2) found in family F60 is de novo and was predicted to disrupt splicing. The child of family F60 suffered from difficulty in expressing his needs using typical words or motions. *DEAF1* is associated with autosomal dominant mental retardation (OMIM615828), and to date, four heterozygous de novo mutations have been identified in four patients with ID, impaired speech development, and behavioral problems [11–13]. Two homozygous *DEAF1* mutations have also been reported to be highly pathogenic [14,15]. The DEAF1 protein and the location of its mutations are shown in Fig. 2A.