VARIACIÓN EXÓMICA DE POLIMORFISMOS DE NUCLEÓTIDO ÚNICO EN GENES ASOCIADOS A TRASTORNOS DEL ESPECTRO AUTISTA (TEA) EN EL SUR OCCIDENTE DE COLOMBIA

EXOMIC SINGLE NUCLEOTIDE POLYMORPHISMS VARIATION IN GENES ASSOCIATED TO AUTISM SPECTRUM DISORDER (ASD) IN SOUTHWEST COLOMBIA

Camila Azcárate Gómez[1], Alejandra Rodríguez[2], Adalberto Sánchez[3], Julio Cesar Montoya[4], José María Satizabal[5], Felipe García Vallejo[6], Laboratorio de Biología Molecular y Patogénesis. Departamento de Ciencias Fisiológicas. Facultad de Salud. Universidad de Valle. Cali. Colombia

1. Estudiante de sexto semestre del programa de Medicina. Facultad de Salud. Universidad del Valle.
2. Candidata a Doctora en Ciencias Biomédicas. Escuela de Ciencias Básicas. Facultad de Salud. Universidad del Valle.
3. Profesor Asociado. Departamento de Ciencias Fisiológicas. Facultad de Salud. Universidad de Valle.
4. Profesor Titular. Departamento de Ciencias Fisiológicas. Facultad de Salud. Universidad de Valle.
5. Profesor Titular. Departamento de Ciencias Fisiológicas. Facultad de Salud. Universidad de Valle.
6. Profesor Titular y Director Científico del grupo LABIOMOL. Departamento de Ciencias Fisiológicas. Facultad de Salud. Universidad de Valle.

Recibido: Agosto 1 de 2017
Aceptado: Octubre 2 de 2017
*Correspondencia del autor: Laboratory of Molecular Biology and Pathogenesis. Department of Physiological Sciences. School of Basic Sciences. Faculty of Health. Universidad del Valle. Cali. Colombia., E-mail: labiomol@gmail.com

RESUMEN

Introducción: Los trastornos del espectro autista (TEA), son desordenes genéticos del neurodesarrollo clínicamente complejos. Recientemente con el avance en las tecnologías de secuenciación de nueva generación (NGS), el enfoque de la patología ha cambiado hacia investigar el papel que tienen las nuevas sustituciones de nucleótido único (SNS) en el progreso de los TEA; además de investigar qué papel tienes las variantes raras en el desarrollo de la condición autista. **Objetivo:** De secuencias de exomas completos de individuos del Suroccidente de Colombia caracterizar los SNS en genes asociados a TEA. **Metodología:** Se secuenciaron los exomas completos de muestras de mucosa bucal de 190 individuos del Suroccidente colombiano; del alineamiento de los exomas empleando la plataforma de Illumina Inc. se registraron y asignaron SNS en diez y ocho genes previamente caracterizados como marcadores para TEA. **Resultados**: Para las muestras del Suroccidente colombiano incluidas en este estudio, se obtuvo un número variable de SNS por gen. Los genes con un mayor número de variantes SNS fueron: RELN con 1.628 sustituciones (945/1.628; C→T); CNTNAP2 con 1.529 (841/1.529; C→T), y SHANK3 con 843 (578/843; C→T). **Discusión:** La secuenciación de exomas completos permitió identificar mutaciones nuevas y variantes poblacionales raras que podrían estar asociadas a ETS. Este trabajo es el primero en reportar la variación SNP en genes TEA en el Suroccidente colombiano. Es de gran importancia pues impulsará nuevos estudios a partir de muestras de otras regiones colombianas para construir una base nacional de la variación SNP de genes TEA en Colombia.

Palabras claves: SNP, Desorden del Espectro Autista, Sur Occidente Colombiano, Exomas, Secuenciación de Siguiente Generación

ABSTRACT

Introduction: Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder with genetic and clinical heterogeneity. In the more recent years and with the advent of next- generation sequencing (NGS) technologies, the focus on the pathology, shifted towards investigating the role of inherited and de-novo single nucleotide substitutions (SNS) in the onset of ASD and the interplay of de novo and inherited rare variants in the development of ASD. **Objective**: To search and characterize Single Nucleotide Substitutions (SNS) of ASD associated genes in exome samples from South-West Colombia. **Methodology**: We collected oral mucose samples of 190 individuals from the South West Colombia. The full exome sequencing of them allows to characterize SNS in eighteen genes which were previously associated as markers for ASD. Data analysis was performed using the Illumina platform alignments and gene assignment which were supplemented with dynamic tables in Excel. **Results**: Our results showed a variable number of new SNS in the samples of southwest Colombia include in this study. The genes with most single nucleotide variation were: RELN with 1,628 SNS (945/1,628 C→T); CNTNAP2 with 1,529 SNS, (841/1,529 C→T), and SHANK3 with 843 SNS, (578/843 C→T). **Discussion:** Whole exome sequencing allows to identify new and rare mutations that would cause ASD at a genomic scale. This is the first study of SNPs variation in ASD genes in the Southwest Colombian region. Its importance is to highlight that further studies from samples from other Colombian regions to generate a more enriched national database of ASD SNP variability in Colombia.

Keywords: SNP, Autism Spectrum Disorder, Colombian South-West, Next Generation Sequencig.

INTRODUCTION

Autism spectrum disorder (ASD) is a neurological and developmental disorder that begins early in childhood and lasts throughout a person's life. It affects how a person acts and interacts with others, communicates, and learns. It includes what used to be known as Asperger syndrome and pervasive developmental disorders. It is manifested in a series of symptoms based on the Wing triad that includes: communication, flexibility and imagination and social interaction (1). To date, knowledge about ASDs suggests a multifactorial structure, which includes a genetic condition added to an environmental factor in the early stages of central nervous system (CNS) development (gestational stage and the first 2 years of postnatal life). However, in 90% of the cases, the causal factor is unknown and in the other 10% of the cases chromosomal and non-chromosomal events of genetic character have been identified, joint to others of environmental character which affect the development of the CNS. (2-5)

In 2005, the Autism Genome Project (AGP) was launched as a large-scale collaboration led by the National Alliance for Autism Research and the National Institutes of Health (6). The aim of this project was to explore through the human genome in the search for genes that confer susceptibility to autism. Phase I of the project included of large-scale scans of the genome using SNP and micro-satellite microarray technology. Subsequently with this information, linkage analysis was performed on approximately 1,500 pedigrees in order to make a fine mapping and sequencing of the critical regions that have been identified associated with the pathology. Finally, they identified the exact SNP variants located in the genes that confer a high predisposition to ASD. (6,7). Variants can be found in the human genome, not only in the exome regions, but also in regions of DNA that do not encode for proteins.

The exome of the human genome is made up of about 180,000 exons constituting approximately 1.5% of the total genome. Mutations occurring in exomic sequences have a much greater probability of being expressed in comparison with any other component of human genome. It is estimated that the exome contains 85% of the mutations which are capable of generating a disease or disorder. In fact, exome sequencing has been an effective strategy for determining the genetic basis of dozens of hereditary diseases and for identifying de novo mutations involved in developmental disorders such as schizophrenia, intellectual disability, epilepsy and ASD (8, 9).

By exome sequencing and analysis of rare variants in 3,871 cases of autism and 9,937 ancestral controls, 107 autosomal genes strongly associated with ASDs were identified (10, 11). In this context, the present study was performed with the main objective to search and characterize Single Nucleotide Substitutions (SNS) of 18 ASD associated genes in 190 exomes of individuals coming from South-West Colombia. To execute the study, we used Next Generation sequencing (NGS) - Whole exome sequencing and bioinformatics analysis. Our results showed a variable number of SNS in the samples of southwest Colombia included in this study; some of them have not been previously reported in any other studies worldwide.

Methodology

DNA extraction

To obtain the DNA samples, each of the 190 patients was subjected to perform with a cotton a swab a buccal mucosa sample. The samples were resuspended in phosphate buffer and the DNeasy kit™ of QIAGEN Co. (Hilden, Alemania) was used. Each extraction was later quantified and verified in quality to undergo the sequencing process.

Sequencing-Protocol

DNA aliquots for each of the samples were lysed with the TruSeq Exome Library prep ™ kit and subsequently the libraries obtained were normalized for sequencing using the TruSeq Rapid Exome ™ kit. The above-mentioned kits are supplied by the Illumina Company of San Diego, California, USA. The standardized fragments with the sequencing adapters were loaded into a HiSeq2500 computer.

Exome analysis

The absolute and relative frequency of each SNP was quantified in the 18 genes associated with ASD. We also collected information about the position in the chromosome, Clinvar RS and type of mutation (transition or transversion) as well as the effect of each. All the data was organized in a dynamic table in Excel®.

RESULTS

Our results showed a variable number of SNS in the samples of southwest Colombia include in this study. The table 1 shows the count of different single nucleotide substitution (SNS) across the genes included in the present study.

Table 1. Statistical results of news SNS in 18 genes associated with ASD in the exomic sequences of Southwest Colombian subjects from a sample of 190 exomes sequenced.

Gene	Total SNS	New SNS	Percentage
TBR1	28	27	96.42
CDH8	111	106	95.50
NRNX1	123	116	94.30
EN2	43	40	93.02
SHANK3	262	242	92.40
FOXP1	57	52	91.22
DYRK1A	32	29	90.62
NLGN3	126	121	90.03
RELN2	127	112	88.18
BCDKC	42	37	88.09
PTHD1-A	40	34	85.0
SCN2A	40	29	72.5
GRIN2BA	43	31	72.10
ADNP	28	19	67.86
GABR3	27	18	66.67
PTEN	30	19	63.33
CNTNAP2	67	37	55.22
ARID1B	239	133	51.46

The genes with most single nucleotide variation were: RELN with 1,628 SNS (945/1,628 C>T); CNTNAP2 with 1,529 SNS, (841/1,529 C>T), and SHANK3 with 843 SNS, (578/843 C>T). The most frequent SNP (relative frequency of 0.91), corresponded to that with ID rs7638391 at the positon 71015021 in the FOXP1A gene by the clinvar information it is not directly associated with the pathology.

DISCUSSION

Whole exome sequencing allows to identify new and rare mutations that would cause ASD at a genomic scale. As Fatemi (12) stated, the causes for autism might be genetic and/or environmental. However, the preponderance of evidence supports genetic causes for evolution of this disabling disorder, with high linkage between autism and several markers on chromosome (7).

The present study is the first to show SNPs variation in 18 ASD genes in the Southwest Colombian region. From 18 genes, those with more single nucleotide substitutions (SNS) were RELN, CNTNAP2 and SHANK3. According with the NCBI database the gene Reelin RELN (Gene ID: 5649), encodes a large secreted extracellular matrix protein thought to control cell-cell interactions critical for cell positioning and neuronal migration during brain development. This protein may be involved in schizophrenia, autism, bipolar disorder, major depression and in migration defects associated with temporal lobe epilepsy. The study made by Lammert and Howell (13) explained that since 2001 when the International Molecular Genetic Study of Autism Consortium (IMGSAC) described a region on chromosome 7q as the peak region of linkage and first autism susceptibility locus (AUTS1), the gene RELN emerged as a number one candidate for autism as its location is at chromosome 7q22. Serajee *et al*. (14) analyzed 34 single nucleotide polymorphisms (SNPs) in the RELN gene from Caucasian families, finding significant differences in the transmissions of the alleles of exon 22 and intron 59 SNP to autistic subjects showing a role for this gene in the susceptibility to autism. In our sample two indel events at 103629803 (87/190) T >TGCCGCC and 103206001 (85/190) G>GAA, located at splice-region-variant:intron-variant: elongation splice-egion were reported as a possible translocations process occurred in 45.7% of our samples. Mutational process at splice region produce an abnormal splicing of the primary transcript and the consequence is the absence of a functional RELN mRNA.

On the other hand, the contactin associated protein like 2 CNTNAP2 (Gene ID: 26047), encodes a member of the neurexin family which functions in the vertebrate nervous system as cell adhesion molecules and receptors. This gene encompasses almost 1.5% of chromosome 7 and is one of the largest genes in the human genome. It is directly bound and regulated by forkhead box protein P2, a transcription factor related to speech and language development. In the review made by Peñagarikano and Geschwind (15), it is established that there is strong evidence of the relation between CNTNAP2 and ASD such as its major role in language development in ASD patients (16). Sampath *et al.* (17) studied 2148 common SNPs using transmission disequilibrium test (TDT) across the entire ~3.3 Mb CNTNAP2 locus in 186 multiplex and 323 simplex families with ASD without finding an association of the variants discovered with ASD. However they did found differential expression of this gene in autistic brains and controls. The indel146805228 (1177190) presviously reportad as ID rs35167289, T>TG at intrón-variant: feature-elongation, and 148106465 (93/190) ID rs72031591, rs142426153, C>CTCTT intron_variant:feature_elongation were frequent in the exome from Southwest Colombia. Mutational process at splice region produce an abnormal splicing of the primary transcript and the consequence is the absence of a functional CNTNAP2 mRNA.

The gene multiple Ankyrin repeat domains 3 SHANK3 (Gene ID: 85358), provides instructions for making a protein that is found in many of the body's tissues but is most abundant in the brain. The SHANK3 protein plays a role in the functioning of synapses and dendritic spine maturation and many mutations in this gene have been pointed out as cause of ASD (18). Disruptions of the Shank3 gene in mouse models, have resulted in synaptic defects and autistic-like behaviors including anxiety, social interaction deficits, and repetitive behavior. In fact, the research results by Mei *et al*. (19) generated a novel Shank3 conditional knock-in mouse model, and show that re-expression of the Shank3 gene in adult mice, led to improvements in synaptic protein composition, spine density and neural function in the striatum.

With complete exome sequencing it is possible to identify rare mutations linked to diseases in an efficient way, and considering that the exome is the coding region of the genome, is precisely where the majority of mutations are found, this technique has been of great support for genomic studies addressing the etiology of certain

diseases.

The present study collected and sequence by the first time 190 exomes from the Colombian South-West population and found SNPs in 18 genes associated with ASD, contributing to the knowledge of the genomic structure of this population and making a way to further studies taking samples from other Colombian regions in order to construct by the first time a more enriched national database of ASD SNP variability in Colombia.

ACKNOWLEDGMENT

Authors tanks to Universidad del Valle for the support to the experimental work. To Genomics Co. and Georgia-Tech, USA for the sequencing and part of bioinformatics analyses. To all donors of the DNA samples who make possible to perform the present study. The experiments involved human samples were make following the ethics regulations at the Universidad del Valle, according with National legislation for the experiments involving human samples, and WHO.

REFERENCIAS

1. Wing L, Gould J. Severe Impairments of Social Interaction and Associated Abnormalities in Children: Epidemiology and Classification. Journal of Autism and Developmental Disorders. 1979; 9:11-29.
2. Newschaffer CJ, Fallin D, Lee NL. Heritable and Nonheritable Risk Factors For Autism Spectrum Disorders. Epidemiologic Reviews 2002;24:137-53.
3. Postorino V, Kerns CM, Vivanti G, Bradshaw J, Siracusano M, Mazzone L. Anxiety Disorders and Obsessive-Compulsive Disorder in Individuals with Autism Spectrum Disorder. Curr Psychiatry Rep. 2017 Oct 30;19(12):92
4. De Rubeis S, Buxbaum JD. Recent advances in the genetics of autism spectrum disorder. Curr Neurol Neurosci Rep. 2015;15(6):365.
5. Sykes NH, Lamb JA. Autism: the quest for the genes. Expert Rev Mol Med. 2007;9(24):1-15.
6. Hu-Lince D, Craig DW, Huentelman MJ, Stephan DA. The Autism Genome Project: goals and strategies. Am J Pharmacogenomics. 2005;5(4):233-46
7. Weiss LA. Autism genetics: emerging data from genome-wide copy-number and single nucleotide polymorphism scans. Expert Rev Mol Diagn. 2009;9(8):795-803.
8. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nature Genet 2011;43:860–863.
9. XU B, Roos JL, Dexheimer P, Boone B, Plummer B, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nature Genet.2011; 43, 864–868.
10. De Rubeis S, He AP, Goldberg CS, Poultney K, Samocha AE, et al.Synaptic, transcriptional, and chromatin genes disrupted in autism. Nature. 2014; 515 (7526): 209-215
11. Yuen RK, Merico D, Bookman M, L Howe J, Thiruvahindrapuram B, Patel RV, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci. 2017;20(4):602-61112.
12. Fatemi SH. The role of Reelin in pathology of autism. Molecular Psychiatry. 2002; 7: 919-920.
13. Lammert DB, Howell BW. *RELN* Mutations in Autism Spectrum Disorder. *Frontiers in Cellular Neuroscience*. 2016;10:84.
14. Serajee FJ, Zhong H, Mahbubul AHM. Association of Reelin gene polymorphisms with autism. Genomics. 2006; 87(1): 75-83
15. Peñagarikano O, Geschwind DH. What does CNTNAP2 reveal about Autism Spectrum Disorder? Trends in molecular medicine. 2012;18(3):156-163
16. Alarcón M, et al. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am. J. Hum. Genet. 2008;82(1):150–159
17. Sampath S, Bhat S, Gupta S, et al. Defining the Contribution of CNTNAP2 to Autism Susceptibility. Devaney J, ed. PLoS ONE. 2013;8(10):e77906.
18. Nemirovsky SI, Córdoba M, Zaiat JJ, Completa SP, Vega PA, et al. Whole Genome Sequencing Reveals a De Novo SHANK3 Mutation in Familial Autism Spectrum Disorder. Hu VW, ed. PLoS ONE. 2015;10(2):e0116358
19. Mei Y, Monteiro P, Zhou Y, Kim JA, Gao X, et al. Adult Restoration of Shank3 Expression Rescues Selective Autistic-Like Phenotypes. Nature. 2016;530(7591):481-484