EXPLORATORY ANALYSIS OF RARE AND NOVEL VARIANTS IN MEXICAN PATIENTS DIAGNOSED WITH SCHIZOPHRENIA AND DEMENTIA

José J. Martínez-Magaña1, Alma D. Genís-Mendoza1,2, Vanessa González-Covarrubias3, Janet Jiménez-Guench4, Aide G. Galindo-Chávez4, Andrés Roche-Bergua5, Carlos Castañeda-González4, Nuria Lanzagorta5, Xavier Soberón3*, Humberto Nicolini1,5*

1Genomics Laboratory of Psychiatric and Neurodegenerative Diseases, Instituto Nacional de Medicina Genómica, Mexico City; 2Psychiatric Care Services, Hospital Psiquiátrico Infantil “Juan N. Navarro”, SSA, Mexico City; 3Pharmacogenomics Laboratory, Instituto Nacional de Medicina Genómica, Mexico City; 4Psychiatric Care Services, Hospital Psiquiátrico “Fray Bernardino Alvarez,” SSA, Mexico City; 5Grupo de Estudios Médicos y Familiares, Mexico City, Mexico.

ABSTRACT

Background: Schizophrenia (SCZ) and dementia, often related, are two of the most common neuropsychiatric diseases; epidemiological studies have shown that SCZ patients present a 2-fold increased risk for dementia compared to non-schizophrenic individuals. We explored the presence of rare and novel damaging gene variants in patients diagnosed with late-onset dementia of Alzheimer's type (DAT) or SCZ. Methods: We included 7 DAT and 12 SCZ patients and performed high-depth targeted sequencing of 184 genes. Results: We found novel and rare damaging variants in 18 genes in these Mexican patients. Carriers of these variants showed extreme phenotypes, including, treatment-resistant SCZ or cognitive decline. Furthermore, we found a variation on ABCC1 as a possible link between psychosis and cognitive impairment. Discussion: As an exploratory analysis, we report some interesting variations that should be corroborated in larger sample size studies. (REV INVEST CLIN. 2019;71:246-54)

Key words: Next-generation sequencing. Dementia. Schizophrenia. Novel variants. Rare variants.

INTRODUCTION

Neuropsychiatric disorders affect approximately 30% of the population worldwide1-3. Schizophrenia (SCZ) and dementia, often related, are two of the most common neuropsychiatric diseases4, and epidemiological studies indicate that patients diagnosed with SCZ present a 2-fold increased risk for dementia compared to non-schizophrenic individuals5. Unfortunately, the etiology of these complex diseases remains to be fully elucidated. Genome-wide association studies (GWAS) have contributed to explain approximately 12% of phenotypic variation of these complicated disorders, including SCZ and dementia6,7, showing an

Corresponding author:
*Humberto Nicolini or Xavier Soberón
Instituto Nacional de Medicina Genómica
Mexico City, Mexico
E-mail: hnicolini@inmegen.gob.mx
xsoberon@inmegen.gob.mx

Received for publication: 27-11-2018
Approved for publication: 06-02-2019
DOI: 10.24875/RIC.19002923
apparent missing heritability\(^8\). One approach to find this missing heritability is to investigate rare highly-damaging (RHdv) and novel variants (Nv) which are not routinely considered in GWAS analyses. Several research groups have undertaken this quest using next-generation sequencing (NGS)\(^9\). One limitation could be that RHdv and Nv are potentially population-specific\(^{10,11}\). The collection of genetic variation in Mexican populations is still an ongoing and incipient endeavor, particularly for RHdv and Nv\(^{12}\). This study aimed to explore by NGS the presence of novel and damaging variants for 184 genes in 19 Mexican patients diagnosed with dementia or SCZ.

**METHODS**

**Study population**

Nineteen individuals from the Geriatric Clinic at the Psychiatric Hospital “Fray Bernardino Alvarez” and the Group of Medical and Family Studies Carracci in Mexico City, Mexico, were invited between 2011 and 2013 to participate. Of them, seven were diagnosed with late-onset dementia of probable Alzheimer’s type (DAT) and 12 with SCZ. All patients were invited to participate and signed informed consent. The study protocol complied with the Helsinki Declaration and was approved by the Ethics and Research Committee at the National Institute of Genomic Medicine (No. IMG/DI/136/2014).

DAT patients filled a demographic questionnaire and were evaluated by a geriatric psychiatrist at the Psychiatric Hospital “Fray Bernardino Alvarez.” Dementia was diagnosed based on the DSM-IVR criteria since our study group found memory impairment and at least one other cortical function affected\(^{13}\). All the patients had a family history of Alzheimer’s disease in at least one, first, or second degree relative, and fulfilled the criteria for probable Alzheimer’s diagnosis according to the National Institute of Neurological Disorders and Stroke and the Alzheimer’s Disease Related Disorders Association\(^{14}\). The patients were evaluated using the following scales: mini-mental state examination (MMSE), NEUROPSI, clock-drawing test, DIPAD, and the clinical dementia rating\(^{15-20}\).

Patients with paranoid SCZ were recruited from the Group of Medical Studies Carracci; all patients had a family history of at least one-, first-, or second-degree relative diagnosed with SCZ. Patients were evaluated with a diagnostic interview for genetic studies\(^{21}\), which is a structured interview, including the disorders contained in the Axis I of the DSM-IVR. In this respect, little changes have been made in the latest version of DSM for SCZ diagnosis\(^{22}\). Furthermore, when the medical record of the patient was available, we included a structured sequence of the response to the consumed medications. We established criteria for treatment-resistance, as previously published\(^{23}\). Positive and negative symptoms were evaluated with SAPS and SANS scales, and cognitive function was evaluated with the MMSE\(^{24}\). **APOE-E4** variant is the most extensively validated among the genetic markers associated with cognitive decline. To consider this variation, all the included individuals (i.e. 7 DAT and 12 SCZ) were negative for the E4 allele of the **APOE**; the **APOE** status was determined by real-time PCR, as previously described\(^{25}\).

**Targeted NGS**

Genomic DNA was extracted from peripheral leukocytes using the Gentra Puregene commercial kit (QIAGEN, USA). We designed synthetic probes for NGS, targeting genes associated with dementia, SCZ, and several pharmacogenetic targets. The selection of genes was based on a literature search for published works reporting an effect of common variations or rare variants in SCZ, dementia or drug response to different antipsychotics or antidementia drugs\(^ {6,7,26-36}\); a list of the captured genes is reported in Supplementary Table 1. Gene capture was performed using the Haloplex target enrichment system (Agilent Technologies, USA) with 1.51Mb with 40754 amplicons. Sequencing libraries were generated according to the manufacturer’s protocol (version D.5, May 2013). Briefly, all DNA samples (a total of 225 ng for each sample) were digested with 8-paired restriction enzymes; fragmentation pattern was analyzed in a 2100 Bioanalyzer (Agilent Technologies, USA). Sequencing libraries were generated according to the manufacturer’s protocol (version D.5, May 2013). Sequencing was performed using a NextSeq500 system (Illumina, USA), aiming for 200x depth coverage in paired-end reads.
Bioinformatic analyses

First, for quality control, we utilized trimmomatic to eliminate reads with a quality score Phred-QS <25 and length below 55 bp; indexes, adaptors, and 5 bp at both read ends were trimmed according to general practices. We then aligned reads to the human genome using BWA and SMALT with GRCh37/hg19 as reference. InDel realignment, base recalibration, and variant calling were done following the GATK best-practices recommendations. HaplotypeCaller was used for SNV detection, and copy number variants (CNV) were detected using the pipeline implemented by XHMM. A total of 1274 variants were called by both aligners, which were used for the following analyses. Variants were confirmed visually in the integrative genomic viewer IGV, and also, annotated using dbSNP version 147.

Analysis of rare and novel damaging variants selection

Variants were registered if detected in at least one SCZ or DAT patient, as heterozygous or homozygous. Variants were annotated utilizing different databases including: dbSNP, OMIC, ClinVar, GnomAD, rebuild, and 1000 Genomes, with Variant Effect Predictor, allowing the prediction of the functional impact, with queries to different algorithms and databases (SIFT, Polyphen-2, FATHMM, CADD, gene splicer, and splice region). As possible pathogenic variants, we selected loss-of-function (LoF) variants (frame shift, stop gained, splice-site acceptor, and splice-site donor) and missense variants if the three algorithms predicted the variants to be damaging (i.e., SIFT, FATHMM, and polyphen-2), and coding synonymous variants and non-coding variants were selected if the CADD score was higher than 25 (CADD). After filtering these variants, we included all the Nv, and for previously reported ones, we only included rare mutations (minor allele frequency <0.1%) using the Genome Aggregation Consortium (GnomAD) and the 1000 Genomes projects databases as reference for population allelic frequency. ClinVar, OMIM (Online Mendelian Inheritance in Man), and an own search in PubMed databases were used as reference for the clinical significance and disease-associated variants. Furthermore, a novel variation (Nv) was considered when it had not been reported. We used the Human Genome Variation Society (HGVS) nomenclature using:

<table>
<thead>
<tr>
<th>Variant</th>
<th>dbSNP</th>
<th>Gene</th>
<th>Reference MAF</th>
<th>Mendelian</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP_004792.1: p.Pro108Ala</td>
<td>rs199784029</td>
<td>NRXN1</td>
<td>0.0008</td>
<td>Pitt-Hopkins-like syndrome</td>
<td>Autism spectrum disorder and schizophrenia</td>
</tr>
<tr>
<td>NP_742054.1: p.Val33Met</td>
<td>rs765679790</td>
<td>KCNH2</td>
<td>0.0000008</td>
<td>Long QT Syndrome</td>
<td>Schizophrenia treatment response and lower intellectual coefficient in schizophrenia</td>
</tr>
<tr>
<td>NP_000515.2: p.Ala155Gly</td>
<td>rs145641566</td>
<td>HTR1A</td>
<td>0.0005</td>
<td>Periodic fever, menstrual cycle-dependent</td>
<td>Alcohol and nicotine dependence and Alzheimer’s disease with alcohol dependence comorbidity</td>
</tr>
<tr>
<td>NP_001748.1: p.Gly195Arg</td>
<td>rs146758729</td>
<td>CBR1</td>
<td>0.0015</td>
<td>NR</td>
<td>Drug toxicity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-coding variants</th>
<th>dbSNP</th>
<th>Gene</th>
<th>Reference MAF</th>
<th>Mendelian</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT_187607.1: g.1782677C&gt;T</td>
<td>rs28363996</td>
<td>ABCC1</td>
<td>0.0003</td>
<td>NR</td>
</tr>
</tbody>
</table>

References:

the web-tool mutalyzer\textsuperscript{54}, and included the “rs” dbSNP (version 147) identifier for the nonNv.

RESULTS

Summary of the total detected variants in the sample

Bioinformatic analyses detected 1274 variants on 184 genes, with an average depth of 96x (range: 55X-120X), and 91.2% coverage. Of these, 1148 were SNVs, 126 indels, and only one CNV on RELN gene. A total of 149 variants (11.7%) were located in coding regions and 1125 (88.3%) in non-coding regions. Frequency analyses showed that more than half of all variants (735 variants) were common (minor allele frequency >5%). In total, we also identified 86 Nv not previously reported. The genes with the highest number of Nv were \textit{PTGER3} (21 Nv), \textit{SLC6A3} (5 Nv), and \textit{ADD1} (5 Nv).

Rare and Nv in patients with DAT

In three of seven DAT patients (42.9%), we detected five damaging variants in five genes (\textit{NRXN1}, \textit{HTR1A}, \textit{KCNH2}, \textit{CBR1}, and \textit{ABCC1}) (Table 1). Novel or LoF variations were not observed. Four variants were missense: \textit{NRXN1} (p.Pro108Ser), \textit{HTR1A} (p.Ala155Gly), \textit{KCNH2} (p.Val33Met), and \textit{CBR1} (p.Gly195Arg), and one intronic \textit{ABCC1} (g.1782677C>T). LoF variation in three genes (\textit{NRXN1}, \textit{KCNH2}, and \textit{HTR1A}) has been reported to be causal of some syndromes with Mendelian inheritance type (Pitt-Hopkins-like syndrome-2, Long QT Syndrome 2, and menstrual cycle-dependent periodic fever), while \textit{CBR1} and \textit{ABCC1} have been reported in drug response. Furthermore, common variation in genes \textit{NRXN1} and \textit{KCNH2} has been previously associated with neuropsychiatric disorders (SCZ, autism spectrum disorder, and drug abuse and dependence), and only common variation on \textit{HTR1A} has been previously associated to DAT. One single DAT patient, DAT 1, carried three of the seven damaging variants, on \textit{NRXN1}, \textit{KCNH2}, and \textit{ABCC1}. This patient obtained the lowest scores in the MMSE = 7 (i.e., affecting almost all his cognitive areas). A summary of some sociodemographic and clinical characteristics of patients carrying the variants is shown in Supplementary Table 2.

Rare and Nv in patients with SCZ

In schizophrenic patients, we identified 13 variants on 13 genes: \textit{ANK2}, \textit{CYP3A4}, \textit{RELN}, \textit{HTR7}, \textit{DISC1}, \textit{TYMS}, \textit{CYP2B6}, \textit{MTHFR}, \textit{NRG1}, \textit{SLC6A5}, \textit{BDNF}, \textit{GRIN2B}, and \textit{ABCC1} (Table 2). Of these, four were LoF on \textit{ANK2}, \textit{CYP3A4}, \textit{RELN}, and \textit{HTR7}; three were missense on \textit{DISC1}, \textit{TYMS}, and \textit{CYP2B6}; and six were coding synonymous or non-coding on \textit{MTHFR}, \textit{NRG1}, \textit{SLC6A5}, \textit{BDNF}, \textit{GRIN2B}, and \textit{ABCC1}. We identified six Nv, which represented almost half of all variants detected for this patient group. In these patients, 10 of the 12 (83.33%) included individuals was a carrier of a damaging variant. Previously, LoF variants in \textit{ANK2}, \textit{RELN}, \textit{SLC6A5}, \textit{MTHFR}, and \textit{GRIN2B} have been reported to cause syndromes with Mendelian inheritance (Table 2). Interestingly, the patient carrier of the variants in \textit{DISC1} had the lowest cognitive function (mini-mental state = 15), and a patient carrier of the LoF in \textit{CYP3A4} had treatment-resistant SCZ. A summary of genetic variations and clinical and sociodemographic data of patients with SCZ are presented in Supplementary Table 3.

DISCUSSION

Here, we present a next-generation genome sequencing analysis to explore the existence of rare and novel damaging variants in patients with SCZ or DAT. Clearly, one of the main limitations of this study is the low number of patients included. However, as an exploratory study, we obtained interesting results that could prompt future studies with larger sample sizes. To the best of our knowledge, there are no reports using NGS to identify rare and novel gene variation for neuropsychiatric disorders in Mexican patients.

Our analyses showed that almost 10% of the targeted genes were carriers of one rare or novel damaging variant. For example, genes coding for drug-metabolizing enzymes (DME) (\textit{CBR1}, \textit{CYP3A4}, \textit{TYMS}, \textit{CYP2B6}, and \textit{MTHFR}), and genes involved in neurodevelopmental processes (\textit{ANK2}, \textit{RELN}, \textit{DISC1}, \textit{NRNX1}, \textit{NRG1}, and \textit{BDNF}) were the two main pathways observed in this study with relevant variation in these patients. Variants on genes \textit{ANK2}, \textit{RELN}, and \textit{NRNX1} have been associated with some syndromes with Mendelian inheritance affecting neurodevelopmental
mechanisms, which suggests that they may have a strong influence on the etiology of DAT or SCZ. The overall effect of these variants on the etiology of neuropsychiatric disorders is still under study, although some hypotheses have been proposed. For instance, a recent WES and WGS analysis of neuropsychiatric patients has proposed that an increase of damaging variants on these genes could decrease the age of Alzheimer’s onset, and that the age of onset of SCZ and autism-spectrum disorders could be influenced by the accumulation of de novo variants in genes involved in neurodevelopmental processes.

The effect of DME on brain processes has been understudied. Nevertheless, some, including CYP1A,

<table>
<thead>
<tr>
<th>Variant</th>
<th>dbSNP</th>
<th>Gene</th>
<th>Reference MAF</th>
<th>Mendelian</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LoF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP_001139.3:</td>
<td>rs750143580</td>
<td>ANK2</td>
<td>0.00006</td>
<td>Long QT syndrome and Cardiac Arrhythmia (Ankyrin-B-related)</td>
<td>Bipolar disorder with binge-eating</td>
</tr>
<tr>
<td>NC_000007.13:</td>
<td>NR</td>
<td>RELN</td>
<td>NR</td>
<td>Lissencephaly and Familial Temporal Lobe Epilepsy</td>
<td>Schizophrenia, autism spectrum disorder and Alzheimer’s disease</td>
</tr>
<tr>
<td>NP_059488.2.p.</td>
<td>rs67666821</td>
<td>CYP3A4</td>
<td>0.0002</td>
<td>NR</td>
<td>Treatment response in schizophrenia</td>
</tr>
<tr>
<td>NC_0000010.10:</td>
<td>NR</td>
<td>HTR7</td>
<td>NR</td>
<td></td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td><strong>Missense variants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP_001158010.1:</td>
<td>rs144959108</td>
<td>DISC1</td>
<td>0.0007</td>
<td>NR</td>
<td>Schizophrenia and Alzheimer’s disease</td>
</tr>
<tr>
<td>NP_001062.1:</td>
<td>NR</td>
<td>TYMS</td>
<td>NR</td>
<td></td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>NP_000758.1:</td>
<td>rs12721655</td>
<td>CYP2B6</td>
<td>0.0023</td>
<td>NR</td>
<td>Nicotine dependence</td>
</tr>
<tr>
<td><strong>Coding synonymous and non-coding variants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP_001305298.1:</td>
<td>rs77029901</td>
<td>SLC6A5</td>
<td>0.0003</td>
<td>Hyperekplexia</td>
<td>Schizophrenia</td>
</tr>
<tr>
<td>NP_005948.3:</td>
<td>rs2066466</td>
<td>MTHFR</td>
<td>0.0032</td>
<td>Homocystinuria due to MTHFR deficiency</td>
<td>Neural tube defects and schizophrenia</td>
</tr>
<tr>
<td>NC_000012.11:</td>
<td>NR</td>
<td>GRIN2B</td>
<td>NR</td>
<td>Autosomal dominant mental retardation and early infantile epileptic encephalopathy</td>
<td>Schizophrenia and autism spectrum disorder</td>
</tr>
<tr>
<td>NC_000002.10:</td>
<td>NR</td>
<td>NRG1</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC_0000008.10:</td>
<td>rs79141432</td>
<td>BDNF</td>
<td>0.0024</td>
<td></td>
<td>Schizophrenia</td>
</tr>
<tr>
<td>NC_0000012.11:</td>
<td>NR</td>
<td>ABCC1</td>
<td>NR</td>
<td></td>
<td>Drug resistance</td>
</tr>
</tbody>
</table>

*Reference MAF: Minor allele frequency reported in the GnomAD or in the 1000 Genomes Project. LoF variant reported to be disease-causing of Mendelian inheritance disorder. Common or rare variants reported to be associated to neuropsychiatric disorders. NR: No reported. LoF: Loss-of-function variants.*
CYP2B, CYP2C, and CYP3A, have been functionally linked to brain development\textsuperscript{112,113}. Our observations regarding DME include that among schizophrenic patients, two were carriers of the CYP3A4*20 (rs67666821) allele as homozygous, and this variant was present in a patient with treatment-resistant SCZ\textsuperscript{23}. CYP3A4*20 is an allele previously identified in the Brazilian population\textsuperscript{114}, and it has been found at high allele frequency in the Spanish population (minor allele frequency = 0.012)\textsuperscript{115}, but at low frequency in other European populations. This allele has been reported to affect the metabolism of clozapine, also associated with treatment-resistant SCZ\textsuperscript{116}.

In relation to carriage of damaging variants in neurodevelopmental genes that could affect SCZ and cognitive ability, two patients diagnosed with SCZ were carriers of the DISC1 missense rare variant (p.Arg418His) and clearly manifested a cognitive disability. DISC1 gene has been involved in the neurodevelopmental process and the development of normal cognitive function\textsuperscript{31}. The product of this gene is greatly involved in brain cortex development, including symmetry and orientation of neurons\textsuperscript{117-120}. Furthermore, a common variation in the DISC1 gene has been associated with Alzheimer’s disease, reinforcing the notion that this gene could have a strong effect on cognitive development.

An interesting finding was that ABCC1 (ATP-binding cassette, subfamily C, and member 1 gene) was the only gene where two patients in each group shared a variant. The patient diagnosed with DAT who was a carrier of the ABCC1 variant had a rapid cognitive decline, with severe manifestations of cognitive impairment. Likewise, the patient diagnosed with SCZ and was a carrier of a variant in this gene had a cognitive disability, mainly affecting memory function. ABCC1 has previously been implicated in the increased accumulation of amyloid-β, dependent on its expression in a mouse model of early Alzheimer’s disease\textsuperscript{121}. However, the effect of the observed novel and rare damaging variants in disease etiology would be under the scope of future studies. The development of NGS technologies has enabled the screening of many genetic variants, finding a large number that has not been previously reported. The substantial number of Nv found makes impractical to functionally validate each one; in this sense, computer methods have been developed to anticipate the effect of a variant at the molecular level. Here, we presented a sequencing data analysis utilizing different algorithms to prioritize the damaging effect of variants. We focused on those with a higher impact on disease etiology, based on distinct algorithms.

Our results may be limited by the small sample size; however, we explored genetic variation in 184 genes previously associated with neurodegenerative diseases and drug treatment. We located some rare and novel damaging variants on 18 genes formerly known to be involved in neuropsychiatric disorders in a Mexican population, and we discussed their potential role in these diseases. Future endeavors should focus on validating these observations.

ACKNOWLEDGMENTS

This work was supported by funding from the National Institute of Genomic Medicine to the Laboratory of Genomics of Psychiatric and Neurodegenerative Diseases, project IMG/DI/136/2014. The technical expertise of Alfredo Mendoza Vargas from the NGS unit is gratefully acknowledged.

SUPPLEMENTARY DATA

Supplementary data are available at Revista de Investigación Clínica online (www.clinicalandtranslational-investigation.com). These data are provided by the corresponding author and published online for the benefit of the reader. The contents of supplementary data are the sole responsibility of the authors.

REFERENCES

acid transporter genes SLC1A4, SLC1A5 and the glycine transporter genes SLC6A5, SLC6A9 with schizophrenia. BMC Psychiatry. 2008;8:58.


