Determination of candidate genes involved in schizophrenia using the whole-exome sequencing

Senormanci O1, Karakas Celik S2, Valipour E2, Dogan V3, Senormanci G1

University of Health Sciences Bursa Yuksek Ihtisas Training and Research Hospital, Yildirim, Bursa, Turkey. senorman_7@hotmail.com

ABSTRACT

OBJECTIVES: We used the whole-exome sequencing to evaluate several genes suspected of being involved in the pathogenesis of schizophrenia.

METHODS: The study sample was composed of two families. In the first family, two siblings had schizophrenia, and the parents were healthy. In the second family, two siblings had schizophrenia, while the other sibling and the parents did not.

RESULTS: Indels were detected in some genes, including SPON1, GRIA3, SMAD5, PCLO, KMT2C, SRD5A2, SEMA3B, NCOR2, GPHB5, FAM174B, CLTCL1, and TMEM216. The insertion of three nucleotides (TGA) was detected in the sequence of the PCLO gene. The mutation resulted in the insertion of aspartic acid (Asp, D) in the amino acid sequence of the PCLO protein. Indels detected in SPON1, GRIA3, SMAD5, KMT2C, SRD5A2, SEMA3B, GPHB5, CLTCL1, and TMEM216 had a damaging effect, while the indel in PCLO had a non-damaging effect on protein function.

CONCLUSION: To the best of our knowledge, no previous studies have examined the relationship between the pathogenesis of schizophrenia and the gene mutations identified in this study (Tab. 1, Ref. 42). Text in PDF www.elis.sk.

KEY WORDS: schizophrenia, genetic, whole-exome sequencing, family.

Introduction

Schizophrenia is a psychiatric disorder that causes chronic and significant problems worldwide, affecting 0.5 % to 1 % of the adult population (1). It is believed that genetic factors play an important role in the etiology of schizophrenia, and the heritability of schizophrenia is estimated to be approximately 80 % (2). Genome-wide association studies (GWAS) are technologies that can identify small effect genes by examining single nucleotide polymorphism (SNP) and copy number variants by studying more than 1,000 patient’s and control samples. These studies are very useful in cases of multiple genetic disorders such as schizophrenia (3). Although the common risk loci identified with GWAS cannot account for a large portion of the genetic risk (4), a multi-stage schizophrenia GWAS has been previously reported (5). Even if GWAS offers an important perspective for diseases, it has limitations, such as not being effective in the study of variations within communities, where the frequencies are at the lower ranges as well as being a time-consuming and costly method. Whole-exome sequencing (WES) methods are a new generation of sequencing methods used to study the mutations and diversities of rare diseases. The WES method is especially useful in developing clinical diagnoses and personalized disease risk profiles. Currently, the method is widely used due to its lower cost and the amount of time saved in obtaining results (6–8). To this point, very few studies have been conducted on schizophrenia using the WES method. However, one such WES-based study found that minor de novo mutations between N-methyl-D-aspartate receptor (NMDAR) complexes and glutamatergic post-synaptic proteins containing cytoskeleton-related proteins that regulate the plasticity of glutamatergic synapses play a role in the etiology of schizophrenia (9). In a case-controlled study conducted by Purcell et al., enriched gene sets were found to be voltage-gated calcium channels and signaling complexes created by activity-regulated cytoskeleton-associated scaffold proteins of post-synaptic density (PSD) (10). Somatic variations were investigated using the WES method for monozygotic twin schizophrenic patients and their parents in China. Although a somatic single nucleotide variation and two somatic additions/deletions were detected in the first stage of the study, no pathogenic somatic variants were detected in the Sanger sequencing (11). In Japan, an unc-13 homologous B (Caenorhabditis elegans) (UNC13B) gene was identified using the WES.
method on two schizophrenic patients and one healthy volunteer in a family with schizophrenia patients (12). Therefore, in order to enlarge previous studies, this research was aimed to define gene variants involved in schizophrenia by comparing individuals with schizophrenia and healthy individuals in the same family using a precise method over a short time period.

Methods

Participant characteristics and evaluation

This study was conducted on schizophrenia patients, who were regularly monitored psychiatrically in the Department of Psychiatry of the Medicine Faculty at Bulent Ecevit University, and on their first-degree relatives, who did not have any psychiatric disease. The patients’ diagnoses were assessed by the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I). The SCID-I is a structured clinical interview developed to determine the diagnosis of psychiatric disorders in the first axis according to DSM-IV (13, 14). All the patients had the paranoid schizophrenia subgroup criteria according to the DSM-IV-TR criteria (15). Ethics committee approval for the study was obtained from the Bulent Ecevit University Clinical Research Ethics Committee, protocol number of 2014-111-17/06. All the participants were informed about the study prior to enrolling in it.

Family 1 consisted of healthy parents and two siblings with schizophrenia. The first was a 36-year-old single female, who was a high school graduate and not working. She had been sick for an average of seven years; during that time, she was admitted to the psychiatric hospital six times. Her latest drug treatment consisted of clozapine 600 mg/day, haloperidol 10 mg/day, and biperiden 2 mg/day (Family 1 on the charts, Sibling 1). The second sibling was a 37-year-old single male, a primary school graduate, who was not working. He has been suffering from schizophrenia for 20 years and had been hospitalized many times in psychiatric hospitals. His latest drug treatment consisted of clozapine 700 mg/day (Family 1, Sibling 2 on tabulations). Family 2 consisted of a healthy mother, a healthy child, and two schizophrenic siblings. One of the siblings was a 50-year-old male patient, a primary school graduate, who was not working. He had a 30-year history of illness with four admissions to the Bulent Ecevit University clinical hospital (Family 2, Sibling 2 on the charts). The last medications he used were paliperidone 12 mg/day and quetiapine 200 mg/day. The second sibling was a single 40-year-old female, a primary school graduate, who was not working. She had been suffering from schizophrenia for 20 years and had had three admissions to the psychiatric hospital. The last medications she used were paliperidone 12 mg/day and quetiapine 200 mg/day (Family 2, Sibling 3 on the tabulations). The healthy control group did not have any DSM-IV diagnoses and no evidence of previous psychiatric referral.

Whole-exome sequencing

DNA samples were sheared into random fragments with Covaris technology, and the size of the library fragments were mainly distributed from 200 bp to 300 bp. Then, adapters were ligated to both ends of the fragments. The DNA was amplified by ligation-mediated polymerase chain reaction (LM-PCR), purified, and hybridized to exome sequencing array. Non-captured fragments were washed out. The captured LM-PCR products were subjected to the Agilent 2100 Bioanalyzer and quantitative PCR to estimate the magnitude of enrichment. Each captured DNA sample was loaded onto Illumine Hiseq platforms. In this study, high throughput sequencing was performed for each library to ensure that each sample met the desired average sequencing coverage. Illumine base-calling software was used for base-calling with default parameters.

Bioinformatics analysis

Bioinformatics analysis was carried out on sequenced samples, which were unprocessed data generated by the Illumine machine. All the clean data of each sample were mapped to the human reference genome (GRCh37/HG19). Burrows-Wheeler Aligner (BWA) software was used to align computation and to map the reads (16, 17). To ensure an accurate variant calling, we followed the recommended best practices for variant analysis using the Genome Analysis Toolkit (GATK, https://www.broadinstitute.org/gatk/guide/best-practices).

A 3-stage data filtering was performed to reduce the contamination on the sequenced fragments: (1) removing reads, which had an adapter sequence, (2) filtering out low quality calls (defined as having a Phred quality score equal to or less than 5) that were more than 50 %, and (3) removing error rates (“N” base) that were more than 10 %. Statistical analysis of the data and bioinformatics analysis were performed using this high-quality data (“clean data”). After high-confidence indels were identified, the SnpEff tool (http://snpeff.sourceforge.net/SnpEff_manual.html) and the SIFT program (http://sift.bii.a-star.edu.sg/) were used as predictor programs. The SnpEff program was used to determine whether the indels had protein-coding changes that had affected amino acids. The SIFT program was used to predict whether an amino acid substitution had affected protein function based on sequence homology and the physical properties of amino acids.

Results

In this project, 50.39 Mb target regions were captured, on which we performed variant calling. Total clean reads per sample were aligned to the human reference genome (GRCh37/HG19) using BWA (16, 17). On an average, 99.58 % were mapped successfully. The strict data quality control (QC) was performed in the whole analysis pipeline for the clean data, the mapping data, the variant calling, etc.

In this research, 6,129 indels were detected, of which 928 were new indels. Of all indels, 380 were frameshifffs, six were stoploss, six were startloss, and 56 were insertion. Of the all identified indels, 223 were only in patients and not in healthy family members.

In the present study, indels were detected in some genes, including SPON1, GRIA3, SMAD5, PCLO, KMT2C, SRD5A2, Sema3B, NCOR2, GPHB5, FAM174B, CLTCL1, and TMEM216. The insertion of TGA was detected in the sequence of the PCLO gene. The mutation resulted in insertion of Asp, D in the amino acid sequence of the PCLO protein. Indels detected in
Tab. 1. The results obtained by analysis of the genes using SIFT program.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Coordinates</th>
<th>Gene ID</th>
<th>Nucleotide change</th>
<th>Substitution Type</th>
<th>Region</th>
<th>Effect</th>
<th>Score</th>
<th>Substitution</th>
<th>Gene ID</th>
<th>Coordinates</th>
<th>Nucleotide change</th>
<th>Substitution Type</th>
<th>Region</th>
<th>Effect</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLTCL1</td>
<td>22,192</td>
<td>ENSG00000070371</td>
<td>GTG-GGGG</td>
<td>Frameshift</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.85</td>
<td>ccaag-C-gcaaa</td>
<td>FAM174B</td>
<td>15,926</td>
<td>554-555</td>
<td>AA-Deletion</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.943</td>
</tr>
<tr>
<td>SRD5A2</td>
<td>24,192</td>
<td>ENSG00000026634</td>
<td>TGG-GGGG</td>
<td>Frameshift</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.85</td>
<td>ccaag-C-gcaaa</td>
<td>SRD5A2</td>
<td>24,192</td>
<td>554-555</td>
<td>AA-Deletion</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.943</td>
</tr>
<tr>
<td>NCOR2</td>
<td>15,192</td>
<td>ENSG00000012171</td>
<td>GTG-GGGG</td>
<td>Frameshift</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.85</td>
<td>ccaag-C-gcaaa</td>
<td>NCOR2</td>
<td>15,192</td>
<td>554-555</td>
<td>AA-Deletion</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.943</td>
</tr>
<tr>
<td>GPHB5</td>
<td>15,192</td>
<td>ENSG00000012171</td>
<td>GTG-GGGG</td>
<td>Frameshift</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.85</td>
<td>ccaag-C-gcaaa</td>
<td>GPHB5</td>
<td>15,192</td>
<td>554-555</td>
<td>AA-Deletion</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.943</td>
</tr>
<tr>
<td>FAM174B</td>
<td>15,192</td>
<td>ENSG00000012171</td>
<td>GTG-GGGG</td>
<td>Frameshift</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.85</td>
<td>ccaag-C-gcaaa</td>
<td>FAM174B</td>
<td>15,192</td>
<td>554-555</td>
<td>AA-Deletion</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.943</td>
</tr>
<tr>
<td>TMEM216</td>
<td>15,192</td>
<td>ENSG00000012171</td>
<td>GTG-GGGG</td>
<td>Frameshift</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.85</td>
<td>ccaag-C-gcaaa</td>
<td>TMEM216</td>
<td>15,192</td>
<td>554-555</td>
<td>AA-Deletion</td>
<td>CDS</td>
<td>Damaging</td>
<td>0.943</td>
</tr>
</tbody>
</table>

Discussion

Exome sequencing is a method, by which only coding regions of the genome are sequenced to detect disease-gene relationships. There are many technologies involved in determining the causes of genetic-associated diseases. Each technique has its own technical, financial, and material restrictions. Although WES is an expensive method compared to the other technologies, it provides a large amount of information with new data analysis programs. Because of its advantages and capabilities, we selected this method for use in this study. Some of the genes determined in this work have been known to be expressed in the brain, but there were not any published reports about their role in the pathogenesis of schizophrenia.

Common gene mutations of both sick and healthy individuals for each family were identified; then, a common pool of these gene mutations was made for both families. The number of genes carrying the insertion/deletion (ins/del) mutation that were obtained from the individual patients of both families was 200. By analyzing these genes, we were able to determine some genes that are involved in schizophrenia.

In the study performed by Jahanshad et al (2013), SPON1 variants were detected in rs2618516 on chromosome 11 (11p15.2). The variants have been shown to cause structural changes in the brains of the elderly and to be responsible for various levels of dementia. Further analysis suggested that this gene and its surrounding gene networks probably had a significant association with autism, developmental disturbances, and mental retardation. The spondin-1 protein, also called F-spondin, is encoded by the SPON1 gene. Spondin protein that is secreted by ventral plaque cells has been found to aid axonal targeting of nerve cells during embryonic development and is a protein containing six thrombospondin domains (807 amino acids), a reelin domain, and a spondin domain. Due to its important role in the axon nerve junction in embryonic development, it is thought that mutations in the gene (especially frameshift mutations that lead to complete protein dysfunction) may be related to schizophrenia (21). In our study, the addition of C in position 602 of the SPON1 gene was detected in the common gene pool of the patients. Therefore, SPON1 can be evaluated as a candidate gene for schizophrenia.

Previously, Northern blot analysis revealed the GRIA3 transcripts in fetal and adult brains. The CHRNA3 gene has been acknowledged to have associations with the clinical features of schizophrenia patients. Similarly, the results of this research showed that the CHG33 gene of schizophrenic patients had a "CTG" deletion between 67–69 nucleotides. It was shown that the GRIA3 gene in rats causes pharmacological and kinetic changes to L-glutamate or α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) (24). Although the GRIA3 gene is considered to be a candidate gene for broad bipolar disorder and X-linked mental retardation, it has been reported to be associated with schizophrenia (25). The genes in the glutamate neurotransmitter system could be considered the potential candidate genes for schizophrenia. The effect of the deficiency of the glutamate signaling pathway in schizophrenia is supported by the fact that there are many mutations at the post-synaptic proteins involved in the NMDR and AMPA receptor (AMPAR) signaling pathways (9, 10). In a systematic study of the AMPAR, GRIA3 was also found to be associated with schizophrenia in women (26). GRIA3 promoter methylation has been shown to be significantly increased in schizophrenia patients compared to the non-disease control group (27). In our study, an ins/del mutation was detected in the GRIA3 gene of schizophrenic patients. This mutation leads to frameshift. Unlike the promoter methylation, frameshift mutation leads to the production of nonfunctional
proteins. Therefore, GRIA3 can be evaluated as a candidate gene for schizophrenia.

Previously, a study was conducted on 300 Bulgarian schizophrenic patients and 600 healthy relatives of the patients to identify susceptibility genes for schizophrenia on chromosome 5q31–32, which is known to be a risk locus for schizophrenia. As the result of the analyses, the SMAD5 gene was found to be one of the candidate genes related to the 5q31–32 linkage region (28). Similarly, in this research, the insertion of C at position 1315–1316, which leads to SMAD5 frameshift, was detected in the common gene pool of patients.

Mutations in H3K4me regulators were observed in neurodevelopmental disorders, such as schizophrenia. KMT2C (lysine methyltransferase 2C) is one of the H3K4me methyltransferases involved in the development of schizophrenia. It is thought that corticotropin-releasing hormone receptors may be associated with schizophrenia (32). In this study, insertion and deletion were detected in the DNAPK gene in the common gene pool of patients. Therefore, this result is in agreement with previous reports.

Dysfunction of the hypothalamus-pituitary-adrenal axis is involved in the development of schizophrenia. It is thought that cortisol metabolizing enzymes such as schizophrenia 5 alpha reductase have an increased systemic activity (33). In the case-control study of schizophrenic patients and a healthy control group, the study did not find a significant difference in cortisol, free cortisone and its metabolites, and SNP in the common gene pool of patients.

In a meta-analysis of schizophrenia, in which synaptic plasticity is believed to play an important role, dysbindin-1 has been identified as the primary indicator of susceptibility to schizophrenia (31). The DNAPK complex binds and phosphorylates dysbindin-1. Functional associations between dysbindin-1 and DNAPK may be associated with schizophrenia (32). In this study, insertion and deletion were detected in the DNAPK gene in the common gene pool of patients. Therefore, this result is in agreement with previous reports.

In this study, in addition to these genes, the following mutations were also detected: addition of C between the 81–82 nucleotides in the SEMA3B gene, addition of nine nucleotides (AGCAGCGGC) between the 5517v5518 nucleotides in the NCOR2 gene, insertion of C between the 156–157 nucleotides in the GPHB5 gene, deletion of the GCTCCA sequence between the 206–211 nucleotides in the FAM174B gene, and insertion of G between the 3602–3603 nucleotides in the CLTCL1 gene.

SEMA3B is thought to play an important role in the guidance of growth cones during neuronal development (36). Northern blot analysis revealed an increase in NCOR2 protein expression in the brain tissue of mice (37). It has been reported that the lactoferrin level of serum is high in schizophrenic patients and is not affected by antipsychotic treatment (38). PCR analysis revealed that the amount of GPHB5 protein increases in human brain tissues (39). In situ hybridization of human embryonic tissues revealed an increase in the expression of TMEM216 in the central nervous system (40). In the patient with behavioral problems, epilepsy, and autism spectrum disorder, chromosomal deletion was detected at a region including FAM174B and two other genes (41). It has been found that the expression of the CLTCL1 gene increases in the developing human brain after conception and is effective in nerve growth (42). Although the role of these gene mutations in the pathogenesis of schizophrenia is unknown, both proteins are detected in patients, but are not found in healthy individuals. Further research is needed on the role of these genes and the proteins synthesized by them in the pathogenesis of schizophrenia.

References


Received May 7, 2018. Accepted June 4, 2018.