Systematic Association Studies of Mitochondrial DNA Variations in Schizophrenia: Focus on the ND5 Gene

Mikhil N. Bamne, Michael E. Talkowski, Carlos T. Moraes, Stephen B. Manuck, Robert E. Ferrell, Kodavali V. Chowdari, and Vishwajit L. Nimgaonkar

Department of Psychiatry, University of Pittsburgh School of Medicine, Western Psychiatric Institute and Clinic, University of Pittsburgh, 3811 O’Hara Street, Pittsburgh, PA 15213; Department of Psychology, University of Miami School of Medicine, Miami, FL 33136; Department of Psychiatry, University of Pittsburgh School of Medicine, Western Psychiatric Institute and Clinic, Pittsburgh, PA 15213; Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15213

Introduction

Mitochondria are vital for cells especially that have constant high-energy demands, like muscle and brain cells. Several disorders can result from mitochondrial dysfunction. The prototypic mitochondrial diseases are Leber hereditary optic neuropathy (LHON), myoclonus epilepsy with ragged red fibers, and mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS). Mitochondrial DNA (mtDNA) mutations may also confer risk for common polygenic diseases, such as Parkinson disease, Alzheimer disease, and bipolar disorder.

Postmortem studies, as well as genetic association studies, have implicated mitochondrial dysfunction in schizophrenia (SZ). We conducted multistaged analysis to assess the involvement of mitochondrial DNA (mtDNA) variations in SZ. Initially, the entire mtDNA genome was sequenced in pools of DNA from SZ cases and controls (n = 180 in each group, set 1). Two polymorphisms localized to the NADH dehydrogenase subunit 5 (ND5) gene demonstrated suggestive case control allele frequency differences (mtDNA 13368 G/A, p = 0.019 and mtDNA 13708G/A, p = 0.043). Hence, the ND5 gene was sequenced in individual samples from the initial panel of cases and controls. Additional subjects from another independent set of cases and controls (set 2, cases, n = 244, controls n = 508) were also sequenced individually. No significant differences in allele frequencies for mtDNA 13368 G/A, and mtDNA 13708G/A were observed. However, we identified 216 other rare variants, 53 of which were reported earlier in association studies of other mitochondrial disorders. We compared the distribution of polymorphisms in both sets of cases and controls. No significant case-control differences were observed in the smaller, first set. In the second set, cases had more variants overall (p = 0.014), as well as synonymous variants (p = 0.02), but the difference for nonsynonymous variants was not significant (p = 0.19). Screening available first-degree relatives (n = 10) revealed 10 maternally inherited variations, suggesting that not all the variants are somatic mutations. Further investigations are warranted.

Key words: mitochondria/schizophrenia/NADH dehydrogenase subunit 5/pooled DNA sequencing/polymorphisms
Two sets of nonoverlapping cases and controls were studied (see table 1). Samples from set 1 were used for the initial analyses, including the pooled assays. Samples from set 2 were used for further detailed analysis. We amplified and sequenced mtDNA in each pool using 5 pairs of overlapping amplicons. All the primers were designed using the mtDNA sequence from human mitochondrial genome database and compared with genomic DNA sequences using NCBI BLAST to ensure specific amplification of regions of the mitochondrial genome (primer list provided in Supplementary Tables S1 and S2). Polymerase chain reactions (PCRs) (10 μl) included 5 pmol primers, 200 μM dNTP, 1.5 mM MgCl₂, 0.5 U AmpliTaq Polymerase (PE Biosystems, Foster City, CA), 1× buffer, and genomic DNA (60 ng). PCR cycles comprised initial denaturation at 94°C for 3 minutes, followed by 35 amplification cycles (94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 60 seconds), and final extension—72°C for 7 minutes, followed by storage at 4°C till further use. The amplified products were checked on 2.0% agarose gels using ethidium bromide staining.

Pooled DNA Sequencing and Individual Sequencing
Amplified PCR products were treated with ExoSAP-IT (USB, Cleveland, OH) at 37°C for 15 minutes followed by enzyme denaturation at 80°C for 15 minutes, and storage.
Treated PCR products were sequenced with appropriate primers and BigDye terminator V3.0 cycle sequencing reagents (Applied Biosystems, Foster City, CA) using the following conditions: 95°C for 30-second initial denaturation and 25 cycles of 95°C for 15 seconds, 50°C for 15 seconds, and 60°C for 4 minutes. Sequencing reactions were isopropanol precipitated and analyzed on an ABI PRISM 3700 DNA Analyzer (Applied Biosystems). Sequencing traces were analyzed using SEQUENCHER software (GeneCodes Corporation, Ann Arbor, MI). All sequencing traces were visually examined to ascertain base calls made by software. Each amplicon was sequenced in triplicate for both case and control pools. Sequencing traces were checked and confirmed by 2 investigators.

Table 2. Prevalence of ND5 Gene Variations Observed Among Cases and Controls

<table>
<thead>
<tr>
<th>Number of Participants With</th>
<th>Case 1 (n = 166)</th>
<th>Neonatal controls (n = 174)</th>
<th>Case 2 (n = 244)</th>
<th>Adult Controls (n = 508)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Variants</td>
<td>105 (63.2%)</td>
<td>109 (62.6%)</td>
<td>166 (68.0%)</td>
<td>316 (62.2%)</td>
</tr>
<tr>
<td>Synonymous Variants</td>
<td>92 (55.4%)</td>
<td>100 (57.5%)</td>
<td>148 (60.7%)</td>
<td>276 (54.3%)</td>
</tr>
<tr>
<td>Nonsynonymous Variants</td>
<td>54 (32.5%)</td>
<td>54 (31.0%)</td>
<td>73 (29.9%)</td>
<td>135 (26.6%)</td>
</tr>
<tr>
<td>Total Mutation Load</td>
<td>212 (0.70 x 10^{-3} per base)</td>
<td>189 (0.59 x 10^{-3} per base)</td>
<td>373 (0.94 x 10^{-3} per base)</td>
<td>603 (0.71 x 10^{-3} per base)</td>
</tr>
<tr>
<td>Expressed Mutation Load</td>
<td>59 (0.19 x 10^{-3} per base)</td>
<td>60 (0.19 x 10^{-3} per base)</td>
<td>107 (0.27 x 10^{-3} per base)</td>
<td>171 (0.20 x 10^{-3} per base)</td>
</tr>
</tbody>
</table>

Note: Mutation load: (total number of variations observed in each group)/(total number of bases analyzed for all samples in that group). Numbers of bases not covered in set 2 were considered while calculating mutation loads for set 2 samples. Individual sequencing of the ND5 gene in cases and controls from both the sets. The frequency of subjects showing variations did not vary significantly between cases and controls in both groups. Some subjects in each group had more than one variation and or both synonymous and nonsynonymous variations (set 1 cases, n = 64; neonatal controls, n = 57; set 2 cases, n = 105, and adult controls, n = 114). The total and expressed mutation loads are representing frequencies of total number of variations and nonsynonymous variations with in that group.

Some samples had more than one variant.
**Statistical Analysis**

We employed an R script console (version 1.9.1) to compare the differences in allele frequencies between cases and controls for pooled DNA sequencing as described by Chowdari et al.\textsuperscript{21} A nonparametric Mann-Whitney test was used to compare the distribution of mtDNA variations between cases and controls, and power calculations were carried out based on the samples available in set 2 (SAS, version 9.1).

**Results**

**Pooled DNA Sequencing Using Cases and Neonatal Cord Samples From Set 1**

We observed 44 common polymorphisms. The list of polymorphisms is available in Supplementary Table S3. All these polymorphisms have been reported in public databases.\textsuperscript{22,23,25,26}

Two variations in the \textit{ND5} gene showed significant case-control differences with regard to minor allele frequencies using the R script analysis (mt.13368G/A, cases: 29.5%, controls: 18.5%, \( Z \) score = 2.33, \( P = 0.019 \); mt.13708G/A (ND5: 458Ala \( \rightarrow \) Thr), cases: 13.7%, controls: 7.1%, \( Z \) score = 2.02, \( P = 0.043 \), see figure 1).

These suggestive findings prompted us to sequence the \textit{ND5} gene individually among cases and controls from set 1.

**Individual Sequencing of ND5 Gene in Cases and Controls**

**Set 1.** We confirmed the mt.13368G/A and mt.13708A/G (ND5: 458Ala \( \rightarrow \) Thr) variations in individual sequencing traces, initially observed by pooled DNA sequencing (mt.13368G/A, cases: 12/166 and controls: 8/166).
11/174; mt.13708A/G (ND5: 458Ala → Thr), cases: 14/166 and controls: 16/174). However, the case-control differences for these variants were not statistically significant. Though all samples comprising the DNA pools were sequenced individually, satisfactory sequencing traces could not be obtained from some samples (cases, $n = 14$; controls, $n = 6$). These traces were not used for case-control comparisons, and details for these individuals are not reported in table 1.

One or more mtDNA variant was present among 105 cases and 109 controls. The proportion of cases (63.2%) with variations was not significantly different compared with controls (62.6%, see table 2). A total of 92 cases (55.4%) had synonymous and 54 cases (32.5%) had nonsynonymous variations. In comparison, 100 controls (55.4%) had synonymous and 54 cases (32.5%) had nonsynonymous (31.0%) variations were observed.

We observed a total of 71 synonymous and 33 nonsynonymous variations in set 1 (listed in Supplementary Table S4). Some of these variations were restricted to either cases (synonymous: $n = 38$, nonsynonymous: $n = 14$) or to controls (synonymous: $n = 13$, nonsynonymous: $n = 6$). These proportions are not significantly different.

The total mutation load was marginally higher among the cases ($0.70 \times 10^{-3}$ per base), compared with the controls ($0.59 \times 10^{-3}$ per base, table 2). The expressed mutation load was also not significantly different between the case and control groups for ($0.19 \times 10^{-3}$ per base, for each group, table 2).

**ND5 Gene Sequence Analysis for Set 2 Participants**

Though the results from the pooled DNA analyses could not be confirmed by the individual sequencing, the presence of large numbers of other variants in the ND5 gene led us to investigate a second, larger, independent sample that included adult control individuals. We successfully sequenced approximately 90% of the ND5 gene sequence, individually in all cases and controls from set 2. However, 2 small regions of the ND5 gene (mt.12741 to mt.12941 and mt.13552 to mt.13634) could not be sequenced satisfactorily in any of the samples.

The cases and the screened adult controls showed nominal differences in the number of individuals with variants (cases = 68.0%, controls = 62.2%, $P = 0.014$; for synonymous variations, $Z$ score = 2.236, $P = 0.021$ and for nonsynonymous variations, $Z$ score = 1.323, $P = 0.191$).

In set 2, we observed 129 synonymous and 62 nonsynonymous mtDNA variations (see Supplementary Table S4). A total of 76 variations (56 synonymous and 20 nonsynonymous) overlapped with the variations observed in set 1. The cases had 63 synonymous and 42 nonsynonymous variations. We detected 97 synonymous and 45 nonsynonymous variations among the controls. There were 17 nonsynonymous variations restricted to the cases and 32 synonymous variations observed exclusively among the cases. There were 20 nonsynonymous and 66 synonymous variations that were present only among the controls.

The total mutation load was marginally higher among the cases ($0.94 \times 10^{-3}$ per base), compared with the controls ($0.71 \times 10^{-3}$ per base) (table 2). The expressed mutation load was not significantly different among cases ($0.27 \times 10^{-3}$ per base) and controls ($0.20 \times 10^{-3}$ per base) (table 2).

**Analysis of Combined Samples**

We identified a total of 219 variations among cases and controls from both sets (see Supplementary Table S4). From both sets 1 and 2, we identified 13 synonymous and 11 nonsynonymous novel variations in the ND5 gene not reported in the mitochondrial database. Of these 24 novel variations, 7 synonymous and 6 nonsynonymous variations were restricted to the cases.

Two nonsynonymous variations (mt.13129C → T, ND5: 265Pro → Ser and mt.14059A → G, ND5: 575Ile → Val) were restricted to controls and cases, respectively. Each of these variants was present in 5 or more subjects in each group. The proportions of controls showing variations were not different between set 1 and set 2 (set 1, overall: 62.6%, nonsynonymous: 31.0%, synonymous: 57.4% and set 2, overall: 62.1%, nonsynonymous: 26.5%, synonymous: 54.1%).

**Power Analysis**

In both sets of samples, the proportion of participants with variants was approximately 65%. If we consider only our set 2 samples (244 cases, 508 controls), we estimate that we had 73.8% power to detect a 10% difference in the proportion of individuals with variants between groups and 23.8% power to detect a 5% difference between groups (alpha threshold of 0.05 for both estimates).

**Polymorphisms Related to Haplogroups**

The Caucasian population can be classified into 9 main haplogroups, namely H, I, J, K, T, U, V, W, and X, based on polymorphic mtDNA variants. Two of these markers (mt13366 BamHI and mt13704 BstNI), associated with haplogroups T and I, respectively, were present in the ND5 gene. The frequencies of these markers were not different between cases and controls (mt.13366 BamHI, cases: 8.3% and controls: 6.9%; mt.13704 BstNI, cases: 8.5% and controls: 10.1%).

**Inheritance Patterns**

DNA was available from mothers and affected siblings of 10 cases who had mtDNA variants, enabling us to evaluate maternal inheritance for these variants. Maternal inheritance of mtDNA variations was observed in all these cases. The sequencing traces for one such family is shown...
in figure 2A. In this family, the mother and the affected sibling shared the mt.13924C → T (ND5: 530Pro → Ser). In another family, the proband had heteroplasmia for mt.14131A/C (ND5: 599Leu → Met), which was present in the mother (figure 2B). DNA was available from the affected siblings for 2 cases. We observed maternal inheritance for 2 nonsynonymous mtDNA variations in the first pair of affected siblings [mt.13924C → T (ND5: 530Pro → Ser) (figure 2A) and mt.14059A → G (ND5: 575Ile → Val)]. In the second pair, 2 nonsynonymous variations [mt.12940G → A (ND5: 202Ala → Thr) and mt.13879T → A (ND5: 515Ser → Thr)] were maternally inherited (data not shown).

**Discussion**

Our aim was to identify mtDNA variations that may be associated with SZ. Initially, we analyzed the entire mitochondrial genome among cases and controls, using the pooled DNA sequencing method. Pooled DNA sequence analysis enables rapid identification of common polymorphisms. We identified 44 common polymorphisms, all of which have been reported in public databases. The frequency of one ND5 polymorphism (mt.13704 BstNI: 9.1%) used for haplogroup definition is consistent between our samples and published data. Following pooled sequencing, we observed a significant case-control difference for 2 variants in the ND5 gene. We followed up this result by comprehensively sequencing the ND5 gene in all subjects. We confirmed the presence of homoplasmic variations at both positions showing case-control differences in the pooled analyses, but the significant allele frequency difference observed in the pooled DNA analysis was not reflected in the analyses of individual traces. These results are consistent with similar disparities noted in a recent whole-genome association study of bipolar disorder, in which results from pooling were also not confirmed by individual genotyping.

There are certain limitations to the pooling strategies. Peak heights generated in the pooled DNA sequencing traces may not indicate allele frequencies accurately due to the differential amplification of alleles during PCR. These workers have emphasized the necessity of applying correction factors generated by sequencing small number of heterozygotes for the estimation of allele frequency measurements. With genomic DNA, it is feasible to obtain precise correction factors using individuals with heterozygous genotypes. The precision arises from the fact that the heterozygous have equal number of chromosomes with each allele. Such precision is not possible when studying mtDNA because the number of mtDNA copies with the mutation is unknown among individuals with heteroplasy. It is also possible that the ratio of mtDNA to nuclear DNA in the blood cells might vary among individuals. These issues may explain the significant difference of allele frequencies observed in our pooled DNA sequencing compared with individual sequencing.

Intriguingly, we also observed an unexpectedly large number of ND5 variations among our subjects following individual sequence analyses. Our observations are consistent with a prior study that suggested the ND5 gene as a mitochondrial mutation hot spot. Most of the variants observed in our study have been reported elsewhere and thus are unlikely to be sequencing artifacts. The unusual diversity is consistent with the relatively high mutation rates known to occur in mtDNA. These rates have been cited in favor of mtDNA being a risk factor for SZ. However, the total number of variations as well as the frequency of nonsynonymous and synonymous variations was not significantly different between cases and controls in the first set. Nominally significant differences were noted in the second set, suggesting a need for evaluations using even larger samples. We analyzed unscreened neonatal controls in our first set of samples. Population stratification and the possibility that some of the controls might be diagnosed with SZ in adulthood could confound our results. Because the lifetime prevalence of SZ is approximately 1%, there is negligible loss of power when you consider these unscreened neonatal controls. Our power analyses suggested the samples used here had reasonable power to detect a modest or large effect size but relatively poor power if the expected effect sizes were small.

Earlier studies have reported on ND5 gene variations associated with prototypic mitochondrial disorders and other neurological/neuropsychiatric disorders. Parker and Parks observed differential expression of the ND5 gene variants among subjects with Parkinson disease. These ND5 variations were also previously associated with severe neurological damage in disorders like Leigh disease and MELAS. In a recent study, Zhadanov et al reported a de novo ND5 gene mutation associated with Leigh syndrome cases. They suggested a lethal effect of the ND5 gene mutation on the structure and function of the mitochondrial complex I. Vanniarajan et al identified a novel nonsynonymous variation in the ND5 gene (mt.13565C → A, ND5: 410Ser → Tyr) that has been implicated in late-onset MELAS. We also identified variations in the ND5 gene that have been reported in several disorders, including LHON, MELAS, idiopathic Parkinson disorder, and bipolar disorder. We are uncertain of the functional implications for these rare variations in our subjects.

Recently, Martorell et al reported a SZ case with specific nonsynonymous variations in the ND5 gene (mt.12403C → T, ND5: 23Leu → Phe and mt.12950A → C, ND5: 205Asn → Thr). Both variations were present in the same proband (cases: 1/6, controls: 0/95) in their samples. These 2 variations were also described among bipolar disorder patients. We observed mt.12403C → T (ND5: 23Leu → Phe) in 2 of our adult
controls but did not detect them in any of our cases. Interestingly, both adult controls with mt.12403C → T (ND5: 23Leu → Phe) also had mt.12950A → C, similar to Martorell et al. Only one of our cases had the mt.12950A → C (ND5: 205Asn → Thr) variation. Martorell et al also identified another variation (synonymous, mt.12705C → T) present in one of their probands but none of their controls. In our study, mt.12705C → T was identified among 36 cases and 79 controls. This dissimilarity could be due to the difference in sample numbers between these 2 studies. Samples from available family members enabled us to ascertain the inheritance of variants identified among 10 cases. DNA was available from the proband, the mother, and an affected sibling in 2 families. All the affected individuals showed maternal inheritance for the proband, the mother, and an affected sibling in 2 families. The affected individuals showed maternal inheritance for the observed ND5 variations, suggesting that the variants are not due to somatic mutations in these cases. Additional comprehensive screening of the entire mtDNA in individual cases and controls, along with their first-degree relatives, can help in ascertaining somatic versus germ-line origin for the other mtDNA variations.

In conclusion, our studies do not support associations between ND5 gene variations and SZ. However, we observed several variants among our patients that have also been reported in other mtDNA disorders. Thus, our result warrants further investigations of the ND5 gene association in larger sample. Also the functional impact of such variants and their relationship to SZ needs to be explored.

Supplementary Materials

Supplementary Tables S1–S4 are available online at http://schizophreniabulletin.oxfordjournals.org.

Funding

National Institute of Mental Health (MH56242, MH66263, and MH63480 to V.L.N.).

Acknowledgments

We thank our participants. We are also grateful to Dr D. J. Kupfer for helpful discussions and for sharing samples from Mental Health Intervention Research Center (MH30915).

References


