Short Report

Whole exome sequencing identifies a novel mutation in the transglutaminase 6 gene for spinocerebellar ataxia in a Chinese family


Autosomal dominant spinocerebellar ataxias (SCA) constitute a heterogeneous group of inherited disorders. The transglutaminase 6 (TGM6) gene was recently suggested as a SCA causative gene in Chinese SCA families. In this study, two affected members of a three-generation Chinese family with SCA characterized by progressive cerebellar ataxia and lower limb pyramidal signs were subjected to whole exome sequencing. Through bioinformatics analysis of the sequence variants in these two individuals, we identified a novel mutation in the TGM6 gene (c.1528G>C) which showed perfect co-segregation with disease phenotype in all nine members of this family. This finding confirms that mutations in TGM6 gene represent an important cause of SCA in Chinese. This study also shows that whole exome sequencing of a small number of affected individuals, leveraged on bioinformatics analysis, can be an efficient strategy for identifying causative mutations in rare Mendelian disorders.

Conflict of interest

The authors declare no conflict of interest.

Autosomal dominant spinocerebellar ataxias (SCA) are a highly heterogeneous group of neurodegenerative disorders characterized by progressive cerebellar ataxias associated with a variable combination of pyramidal and extrapyramidal signs, neuropathy, ophthalmoplegia, cognitive impairment and epilepsy. To date, over 30 different SCAs have been described and causative mutations identified in over 20 of them (1). Most of the causative mutations involve trinucleotide repeat expansion, which can either be translated or located in the introns and untranslated 5' or 3' regions. Insertion/deletion and missense mutations have also been described.

Exome sequencing is superior to traditional positional cloning strategies (2) in identifying mutations in diseases which are found in only a small number of
cases/families, have locus heterogeneity, or substantially reduce reproductive fitness (3, 4). Many bioinformatics and statistical genetic tools, for example (5, 6), are being developed to process the large amount of data generated by exome sequencing for disease gene mapping. Recently, also using the exome sequencing strategy, Wang et al. (5, 7) identified transglutaminase 6 (TGM6) as a novel causative gene of SCA in two Chinese families (7). Additional replication studies in independent samples are necessary to clarify the causal role of this novel gene in the development of SCA.

In this study, we examined the features of a three-generation Chinese family with SCA and used exome sequencing to identify the underlying mutation.

Subjects and methods

Subjects

A three-generation Chinese family with a SCA was recruited in Hong Kong. The proband was a 53-year-old woman with slowly progressive unsteady gait. Blood was collected from nine family members (four symptomatic and five asymptomatic). DNA from two affected family members was subjected to exome sequencing. This study was approved by the Hong Kong Hospital Authority/Hong Kong West Cluster Institutional Review Board Ethics Committee. All subjects gave written informed consent to participate in the study.

Genomic DNA extraction and exome sequencing

Genomic DNA was extracted and purified from peripheral blood leukocytes using the QIAamp blood kit according to the manufacturer’s protocols (QIAGEN, Hilden, Germany). Exome capture was conducted by a NimbleGen 2.1M HD array to collect protein coding regions of human genome DNA (Roche NimbleGen Inc., Madison, WI). The exon-enriched DNA were sequenced by the Illumina Genome Analyzer II platform, following the manufacturer’s instructions (Illumina, San Diego, CA) at deCODE (http://www.decode.com/).

Reads mapping and variants calling

The paired-end 76-bp short reads of exome sequencing were mapped onto the University of California, Santa Cruz (UCSC) human reference genome, version hg18, by the Burrows-Wheeler Aligner (BWA) (8). Duplicated reads were marked by Picard (http://picard.sourceforge.net/). The Genome Analysis Toolkit (GATK) (9) was used to recalibrate the alignments and to call single-nucleotide variants (SNVs) (by UnifiedGenotyper) and short insertion–deletion variants (Indels) (by IndelGenotyperV2). All genotype calls with a read coverage ≤×4, a Phred-scaled base calling quality of ≤50, and a Phred-scaled genotype calling quality of ≤20 were excluded.

Variant filtration, prioritization and validation analysis

We annotated and prioritized the SNVs and Indels using a software known as a Knowledge-based mining platform for Genomic and Genetic studies using Sequence data (KGGSeq) (http://statgenpro.psychiatry.hku.hk/kgsseq) (10). The settings of the prioritization are described in Appendix S1 of the Supporting information.

We validated the prioritized candidate variants by testing for co-segregation of the genotype with the phenotype in the remaining seven family members, using conventional Sanger sequencing method. The linkage analysis program Merlin was used to calculate the single-point logarithm of odds (LOD) score to evaluate the statistical significance of the co-segregation (11).

Computer modeling

AGGRESCAN (http://bioinf.uab.es/aggrescan) was used to predict folding properties of normal and D510H mutant TGM6 protein. The TGM6 protein tertiary structure was downloaded from SWISS-MODEL Repository (http://swissmodel.expasy.org/repository) and the effect of D510H mutation was evaluated through DeepView/SWISS-PDBVIEWER (version 4.0.4, http://spdbv.vital-it.ch/). PopMuSic-2.0 (http://babylone.ulb.ac.be/PopPV2) was used to predict thermodynamic protein stability changes based on the downloaded protein tertiary structure.

Results

Clinical examination of the SCA family

The proband was a 53-year-old woman (II-6) (see pedigree tree in Fig. 1). She reported unsteadiness of gait with frequent falls since early teenage with slow progression. Physical examination at age 53 showed cerebellar ataxia with intention tremor and dysmetria and pyramidal signs with generalized hyperreflexia and bilateral Babinski’s sign. She required two crutches for walking. There was no associated sensory impairment, ophthalmoplegia or impaired cognition. Her father had walking difficulties for a few years before his death from diabetes-related complications at age 59. She had two affected sisters, both with cerebellar ataxia and pyramidal signs of spasticity, hyperreflexia and Babinski’s sign. Her son (III-8) had a history of delayed speech development. When he was assessed by pediatricians at age 12 for evaluation of unsteadiness and easy falls, he was found to have mild ataxia. At age 20, there were more prominent pyramidal signs in the lower limbs with increased tone and brisk tendon reflexes. Table 1 shows the clinical examination results of the patients in this pedigree.

Exome sequencing and identification of candidate genes

Two affected individuals (II-1 and III-8) from this family were sequenced by the exome sequencing platforms. Over 30 million read pairs (76bp × 2)
Mutation of TGM6 in a Chinese SCA family

Fig. 1. Pedigree chart of the Chinese spinocerebellar ataxias family. Filled and unfilled symbols indicate affected and unaffected individuals, respectively. Squares and circles represent males and females, respectively. The genotypes at the suggested variant of TGM6 gene are listed below each individual where the mutation is in bold and italic font.

Table 1. Clinical features of symptomatic family members

<table>
<thead>
<tr>
<th></th>
<th>II-1</th>
<th>II-6</th>
<th>II-8</th>
<th>III-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Age of onset</td>
<td>40s</td>
<td>Teenage</td>
<td>20s</td>
<td>Childhood</td>
</tr>
<tr>
<td>Presenting symptoms</td>
<td>Unsteady gait</td>
<td>Unsteady gait, frequent falls</td>
<td>Unsteady gait</td>
<td>Delayed speech, easy falls</td>
</tr>
<tr>
<td>Spasticity</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ataxia</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Deep tendon reflexes</td>
<td>Brisk</td>
<td>Brisk</td>
<td>Brisk</td>
<td>Brisk</td>
</tr>
<tr>
<td>Associated features</td>
<td>Nil</td>
<td>Nil</td>
<td>Nil</td>
<td>Mild mental retardation</td>
</tr>
<tr>
<td>Ambulation</td>
<td>Independent</td>
<td>With crutches</td>
<td>Wheelchair</td>
<td>Independent</td>
</tr>
</tbody>
</table>

were mapped onto UCSC human reference genome (version hg18). The GATK pipeline finally called 21,364 unique SNVs and 610 unique short Indels in the two subjects which passed the quality assessment and aligned onto the exonic regions. The mean coverages at the consensus coding sequence (CCDS) regions for SNVs and short Indels were ~28.5-fold and ~22.4-fold; 90% SNVs and 80% Indels have at least fourfold coverage in both exome sequencing samples.

Table 2 lists the summary results of the three-level filtration and prioritization procedure in KGGSeq. Only ~24.6% SNVs (out of the 21,364) and ~40% Indels (out of the 610) were consistent with the assumption of rare autosomal dominant disease mode (i.e. being heterozygous in both patients) and were kept for further analysis. We then focused on the 2676 non-synonymous [missense, stopgain (a mutation that changes an amino acid codon to a stop codon), stoploss (a mutation that changes a stop codon to an amino acid codon)] and splicing SNVs and 247 frameshift-, non-frameshift-, stoploss-, stopgain- and splicing Indels. After the exclusion of variants with non-rare alleles (Minor Allele Frequency, MAF > 0.005), 77 SNVs and 8 Indels were left. Fifty-nine non-synonymous SNVs which were not predicted to be disease causing were further excluded. In the knowledge level prioritization, four non-synonymous SNVs and two Indels variants located in genes whose protein products have protein–protein interaction, or share the same known biological pathways with at least one of the 20 known causal genes of SCA, or co-mentioned with the term ‘spinocerebellar ataxias’ in the titles or abstracts of published papers of NCBI PubMed database were highlighted.

Table 2. Summary results of the three-level filtration and prioritization procedure in KGGSeq

<table>
<thead>
<tr>
<th>Steps</th>
<th>Number of SNVs (genes)</th>
<th>Number of Indels (genes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>21,364</td>
<td>610</td>
</tr>
<tr>
<td>Inheritance patterna</td>
<td>5,150</td>
<td>247</td>
</tr>
<tr>
<td>Non-synonymousb</td>
<td>2,676</td>
<td>247</td>
</tr>
<tr>
<td>Rare in dbSNP + 1000 Genome and an in-house datasetc</td>
<td>77</td>
<td>8</td>
</tr>
<tr>
<td>Predicted to be disease causal</td>
<td>18 (18)</td>
<td>–</td>
</tr>
<tr>
<td>Knowledge-relatedd</td>
<td>4 (4)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>PPI</td>
<td>1 (1)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Pathway</td>
<td>2 (2)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>PubMed</td>
<td>2 (2)</td>
<td>0</td>
</tr>
</tbody>
</table>

KGGSeq, Knowledge-based mining platform for Genomic and Genetic studies using Sequence data; SNV, single-nucleotide variant.

aDominant mode only considered with variants in heterozygous genotypes and with shared alleles between the two patients.

bNon-synonymous includes missense, stopgain, stoploss and splicing SNVs and insertions/deletions causing frameshift, non-frameshift, stoploss, stopgain and splicing differences.

cThe rare variants referred to variants with MAF ≤ 0.005 in the datasets.

dKnowledge-related variants/genes refer to those variants’ genes having PPI(s) or sharing pathway(s) with provided candidate gene(s), and those variants fell into region(s) or gene(s) which co-occurred in the titles or abstracts of papers in the PubMed database.
Li et al.

Mutation screening at the prioritized exome sequencing variants

In the prioritized list by KGGSeq, two missense variants were of particular interest. One SNV is within the gene ADRBK1 [c. 302G>A(p.R101H)] whose product has protein-protein interaction (PPI) with an SCA casual gene PRKCG and shares the same pathway with another SCA causal gene ITPR1. The other SNV is within the gene TGM6 [c. 1528G>C(p.D510H)] which was recently reported as a causal gene for two Chinese SCA families by exome sequencing (7). We first carried out follow-up validation of the two missense variants by the conventional Sanger DNA sequencing in nine of the family members who had agreed to participate in this study. The missense mutation p.D510H of TGM6 was found to co-segregate with the disease phenotype in the family with all four symptomatic individuals being heterozygous at this variant. This mutation was not present in any of the asymptomatic family members tested. The LOD score for linkage under rare dominant mode was 1.71 corresponding to the p-value 0.005 (chi-squared statistic, $\chi^2 = 8.19$). Conversely, the SNV of ADRBK1 did not co-segregate with disease phenotype and was found to be present in II-1, II-6, II-9, III-3, III-4, III-7 and III-8; it was absent in II-2 and II-8.

Computer modeling at a novel mutation of TGM6

Using AGGREGSCAN, the wild-type (WT) residue 510D was located at the boundary of a ‘hotspot area’ (denoting an aggregation-prone segment) and the mutant H residue at this site extended the hotspot area to 509th site, suggesting that this mutation may perturb normal protein folding. According to the predicted structure (Fig. S1) by Swiss-model, D 510 in the WT TGM6 protein made a hydrogen bond with SER 568 while the mutant H510 was predicted to make two hydrogen bonds with THR 566, which may change the local surface structure of the Transglutaminase_C domain. Using PopMUTSc2-2.0, this mutation would result in high-folding free energy ($\Delta \Delta G = 0.08$ kcal/mol) that had a destabilizing effect on TGM6 protein structure.

Discussion

The mutation we identified in the TGM6 gene (NM_198994c. 1528G>C; NP_945345.2p.D510H) is highly conserved across multiple species with a normalized conservation score of 0.946 by PHYLOP (12). It is within the Pfam protein domain PF00927, Transglutaminase_C. The mutant residue H has a side chain which is 10% positive and 90% neutral compared with D which is negative (13) and is much larger in size. Protein structure analysis suggested that this mutation may perturb normal protein folding, change the local surface structure, and destabilize TGM6 protein structure. Both SIFT (14) and POLYPHEN2 (15) predicted this mutation to be highly deleterious. It is in close proximity to a previously reported mutation c. 1550T>G transition (p.L517W) in the same gene in another Chinese SCA family. The mutation co-segregated perfectly with the phenotype in our SCA family. Although the linkage LOD score did not achieve the conventional significance level 3 due to relatively small pedigree size, the corresponding p-value is small (0.005), which suggests that only approximately 0.5% of the genome is expected to show perfect co-segregation with the disease in nine family members.

There is considerable heterogeneity in the phenotype within this family. While all the symptomatic subjects had cerebellar ataxias together with pyramidal signs in the lower limbs, the age of onset, the rate of progression and severity of disability varied considerably. One affected member also showed associated mental retardation which was absent in the other symptomatic family members. This clinical heterogeneity even within the same family with the same mutation highlights the variability in symptoms and severity that is characteristic of SCA. Their clinical features of predominantly cerebellar ataxia and prominent lower limb pyramidal signs are nevertheless similar to the two previously described SCA families with mutations in TGM6 by Wang et al. (5, 7), which is supportive of a common or similar genetic basis in the three Chinese pedigrees.

The protein product of the TGM6 gene has not been well studied. TGM6 is expressed in the human central nervous system and TGM6 deposits have been shown to be present in the cerebellum of patients with gluten ataxia (16). Injection of antibodies against TGM2/3/6 caused ataxia in mice (17). These findings suggested a role of TGM6 in the development of ataxia but more biological functional studies are required to define the pathogenic mechanisms of TMG6.

This study can also serve as proof-of-principle example of our KGGSeq tool, showing the effectiveness of bioinformatics analysis in exome sequencing studies of monogenic disorders. With the use of a comprehensive bioinformatics prioritization framework implemented in KGGSeq for the downstream analysis, we quickly pinpointed the underlying causal mutation. We hope that this case will encourage more bioinformatics analysis for sequencing-based genetic mapping.

Supporting Information

The following Supporting information is available for this article: Fig. S1. The predicted tertiary structures of wild-type (a) and mutant (b) TGM6.

Appendix S1. Variant filtration and prioritization analysis by KGGSeq.

Additional Supporting information may be found in the online version of this article.
Acknowledgements

This study was funded by the Faculty of Medicine Development Fund, the University Research Committee for the Strategic Research Theme on Genomics and the Small Project Funding HKU 201007176166. M. L. has salary support from the Genome Research Centre and the Centre for Reproduction, Development and Growth at University of Hong Kong.

References