TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing

Jun Ling Wang,1,2,* Xu Yang,3,∗ Kun Xia,2,* Zheng Mao Hu,2 Ling Weng,1 Xin Jin,3,4 Hong Jiang,1,5 Peng Zhang,3 Lu Shen,1,5 Ji Feng Guo,1,5 Nan Li,1 Ying Rui Li,3 Li Fang Lei,1 Jie Zhou,1 Juan Du,1 Ya Fang Zhou,1 Qian Pan,2 Jian Wang,3 Jun Wang,3,6 Rui Qiang Li3,6 and Bei Sha Tang1,2,5

1 Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan province 410008, China
2 National Key Lab of Medical Genetics of China, Changsha, Hunan province 410008, China
3 BGI-Shenzhen, Shenzhen, Guangdong province 518083, China
4 Innovative Program for Undergraduate Students, School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510641, China
5 Neurodegenerative Disorders Research Center, Central South University, Changsha, Hunan province 410008, China
6 Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark

*These authors contributed equally to this work.

Correspondence to: Dr Bei Sha Tang,
Department of Neurology,
Xiangya Hospital,
Changsha, Hunan Province 410008, China
E-mail: bstang7398@yahoo.com.cn

Correspondence may also be addressed to: Rui Qiang Li,
BGI-Shenzhen, Beishan Industrial Zone,
Yantian Distince,
Shenzhen, 518083,
China.
E-mail: lirq@genomics.org.cn

Autosomal-dominant spinocerebellar ataxias constitute a large, heterogeneous group of progressive neurodegenerative diseases with multiple types. To date, classical genetic studies have revealed 31 distinct genetic forms of spinocerebellar ataxias and identified 19 causative genes. Traditional positional cloning strategies, however, have limitations for finding causative genes of rare Mendelian disorders. Here, we used a combined strategy of exome sequencing and linkage analysis to identify a novel spinocerebellar ataxia causative gene, TGM6. We sequenced the whole exome of four patients in a Chinese four-generation spinocerebellar ataxia family and identified a missense mutation, c.1550T–G transition (L517W), in exon 10 of TGM6. This change is at a highly conserved position, is predicted to have a functional impact, and completely cosegregated with the phenotype. The exome results were validated using linkage analysis. The mutation we identified using exome sequencing was located in the same region (20p13–12.2) as that identified by linkage analysis, which cross-validated TGM6 as the causative spinocerebellar ataxia gene in this family. We also showed that the causative gene could be mapped by a combined method of linkage analysis and sequencing of one sample from the family. We further confirmed our finding by identifying another missense mutation c.980A–G transition (D327G) in exon seven of TGM6 in an additional spinocerebellar ataxia family, which...
also cosegregated with the phenotype. Both mutations were absent in 500 normal unaffected individuals of matched geographical ancestry. The finding of TGM6 as a novel causative gene of spinocerebellar ataxia illustrates whole-exome sequencing of affected individuals from one family as an effective and cost efficient method for mapping genes of rare Mendelian disorders and the use of linkage analysis and exome sequencing for further improving efficiency.

**Keywords:** whole-exome sequencing; spinocerebellar ataxia; linkage analysis; TGM6

**Abbreviations:** NS/SS/Indel = non-synonymous/splice acceptor and donor site/insertions or deletions; SCA = autosomal-dominant spinocerebellar ataxias; SNP = single-nucleotide polymorphism; TGM6 = transglutaminase 6

## Introduction

Autosomal-dominant spinocerebellar ataxias (SCA) are a complex and heterogeneous group of neurodegenerative diseases characterized by progressive cerebellar ataxia of gait and limbs variably associated with ophthalmoplegia, pigmentary retinopathy, pyramidal and extrapyramidal signs, dementia and peripheral neuropathy (Schols et al., 2004; Matilla-Duenas et al., 2010). Epidemiological data indicate a prevalence of 5–7 cases per 100 000 people in some populations, and as many as 18.5/100 000 in Japan when dominant, recessive, sporadic SCAs and familial spinal paraplegia are included (Tsuji et al., 2008; Matilla-Duenas et al., 2010). So far, 31 different genetic loci have been reported to be associated with SCA subtypes, and 19 causative genes have been identified, including trinucleotide or polynucleotide repeat expansions, such as SCA1–3, 6–8, 12, 17, 31 and dentatorubral-pallidolysiasis atrophy; and non-repeat mutations, such as SCA5, 11, 13–16, 27, 28 and 16q-linked autosomal dominant cerebellar ataxia (Sato et al., 2009; Di Bella et al., 2010; Matilla-Duenas et al., 2010). There are no genes or mutations identified for the remaining subtypes.

Previously, the primary means for disease gene identification has been through traditional positional cloning strategies. The scope of this method however, is limited, especially in situations where there are small family sizes, locus heterogeneity, substantially reduced reproductive fitness and an abundance of candidate genes present in the mapped region. However, a recently developed combinational strategy using whole-exome capture and high-throughput sequencing technology provides a powerful and affordable means to identify causative genes even in these difficult cases. Recent studies have shown the sensitivity and accuracy of the exome sequencing approach (Choi et al., 2009; Ng et al., 2009, 2010), in the identification of the causal genes of several rare monogenic diseases such as Miller syndrome (Ng et al., 2010) and Bartter syndrome (Choi et al., 2009).

Here, we report the application of whole-exome capture to identify a novel causative gene for SCA, TGM6 (transglutaminase 6). We identified this gene using only four affected individuals in one four-generation Chinese SCA family. Several lines of evidence supported the causal role of this gene including: (i) validation through linkage analysis; (ii) the conservative nature of the mutation site; (iii) the cosegregation of the mutation with the phenotype and (iv) its absence in 500 normal unaffected individuals of matched geographical ancestry. We also identified a different missense mutation present in TGM6 in a second (two-generation) Chinese SCA family. The identification of a new gene underlying this severe disorder provides another tool for understanding the genetic basis of this disease. Additionally, the ability to identify a causal gene using a single family relatively quickly indicates that exome sequencing will be a powerful tool for discovering not only the remaining SCA genes, but a broad spectrum of rare Mendelian disorders, of which neurological disorders are a major component.

## Materials and methods

### Subjects

Genomic DNA was prepared from venous blood by standard procedures, from a variety of affected and unaffected individuals. The first set of analyses was of a four-generation Chinese family (Family CS; pedigree Fig. 1A) with SCA, characterized by late onset, slow progressive gait, limb ataxia and, in some cases, spasmodic torticollis. All the available affected individuals (nine patients, including five males and four females) were subjected to a thorough neurological examination by two or more experienced neurologists. A total of 24 subjects (II:10, II:12, II:14, II:16, II:17, III:1–4, III:6–10, III:14–20, III:22, IV:1 and IV:2) in Family CS (both affected and their healthy spouse) were included in the linkage study. An additional six subjects (III:11, III:12, III:13, III:21, IV:3 and IV:4) were included with the former 24 samples, for the fine mapping analysis. Four individuals (III:6, III:7, III:17 and IV:1) were used for exome sequencing.

A cohort of 84 additional unrelated autosomal-dominant SCA pedigrees was chosen for further analysis. A diagnosis of SCA was determined following the Harding diagnostic criteria (Harding, 1983) and other known SCAs causative gene mutations had been excluded previously from this affected set. Analyses also included 500 unaffected individuals of matched geographical ancestry as a healthy control. The study was approved by the Expert Committee of Xiangya Hospital of the Central South University in China (equivalent to an Institutional Review Board). Written informed consent was obtained from all subjects.

### Genome-wide genotyping for linkage analysis

To localize the disease-causing gene, we carried out whole-genome genotyping using the Infinium HumanLinkage-12 Genotyping BeadChip (Illumina, San Diego, CA, USA). The BeadChip included 6090 single-nucleotide polymorphism (SNP) markers with an average gap of 441 kb and 0.58 cM across the genome. The genotype assignments were determined with Bead Studio genotyping module software (Illumina). Thirteen additional markers were analysed to narrow the
Figure 1 Partial pedigree of Families CS and LY, brain magnetic resonance imaging of proband of Family CS, and mapped region on chromosome 20. (A) Haplotype analysis is shown for chromosome 20 using 13 markers. The black bar indicates the haplotype assumed to carry the disease allele. The bars with black and white haplotypes indicate the existence of recombination event. (B) Pedigree of Family LY. All sampled subjects in A and B are identified by their Roman numerals below the symbol. Arabic numerals denote each individual in a generation. Open symbols = unaffected; filled symbols = affected; symbols with a diagonal line = deceased subjects; symbols with a question mark = clinically unknown; squares = male; circles = female; arrow = the proband; minus = the wild-type allele; + in panel A (Family CS) = the heterozygous missense mutation, c.1550T–G transition (L517W), in exon 10 of TGM6; + in panel B (Family LY) = the heterozygous missense mutation c.980A–G transition (D327G), in exon seven of TGM6. (C) Brain magnetic resonance imaging of proband III.7 in Family CS. Left: axial T1-weighted image showing atrophy of the cerebellar vermis. Right: mid-line sagittal T1-weighted image showing cerebellar atrophy, particularly evident in the superior vermis, with enlargement of the fourth ventricle. (D) Mapped region of Family CS on chromosome 20. Genetic map of the chromosome 20 markers used in this study. Loci that appear on the same line map to the same genetic location. The order of these markers was obtained from the chromosome 20 P-arm metric physical map. The minimal in-linkage interval is 8.4 Mb of genomic DNA, 18.45 cM in size between D20S199 and D20S917 on chromosome 20p13-12.2.
positive candidate region on chromosome 20 by the method as described (Tang et al., 2004). Relative positions were derived from the human genome draft sequence and the Marshfield map (www.ncbi.nlm.nih.gov). Pairwise logarithm of the odds scores were calculated using the MLINK programs of the Linkage 5.2 package. The allele frequencies of markers were assumed to be equal, as well as the recombination fractions in males and females. The disease was considered to be autosomal dominant with a frequency of 0.0001 and 95% penetrance (Lathrop and Lalouel, 1984).

**Exon capture**

We obtained 15 μg genomic DNA samples from each of the four affected members of Family CS and sheared the genomic DNA by sonication. The sheared genomic DNA of each individual was then hybridized with a NimbleGen 2.1M-probe sequence capture array to enrich the exonic DNA in each library. The array covered 18,654 (92%) of the total 20,091 genes in the Consensus Coding Sequence Region database 2008.

**Next-generation sequencing, read mapping and single nucleotide polymorphism detection**

Exon-enriched DNA was sequenced by the Illumina Genome Analyser II platform following the manufacturer’s instructions (Illumina). Raw image files were processed by the Illumina pipeline (version 1.3.4) for base calling and generating the reads set. The sequencing reads were aligned to the NCBI human reference genome (NCBI36.3) using SOAPaligner (Li et al., 2008). We collected reads that were aligned to the designed target regions for SNP identification and subsequent analysis. The consensus sequence and quality of each allele was calculated by SOAPsnp. The low-quality variations were filtered out using the following criteria: (i) quality score >20 (Q20); (ii) average copy number at the allele site <2; (iii) distance of two adjacent SNPs ≥5 bp; and (iv) sequencing depth ≥4 and ≤500.

**Detection of insertions and deletions**

Insertions and deletions (indels) in the exome regions were identified through de novo assembly of the sequencing reads. We assembled the reads using SOAPdenovo (Li et al., 2010) with the 31-mer option enabled. We then aligned the assembled consensus sequences to the reference genome by LASTZ, and passed the alignment result to axtBest (Schwartz et al., 2003) to separate orthologous from paralogous alignment. Finally, we identified the breakpoints in the alignment and annotated the genotypes of insertions and deletions.

**Public data used**

Data from dbSNP129 and 1000 Genome Project (20100208 release) were downloaded from the NCBI FTP site (http://www.ncbi.nlm.nih.gov/ftp/). The exome variation set of eight HapMap samples were obtained from ‘Supplementary data 2’ of 12 exomes in Ng et al. (2009).

---

**Results**

**The clinical feature of two spinocerebellar ataxia families**

The proband (III:7) of Family CS was a 60-year-old female of Chinese descent and presented with walking difficulty, spasmodic torticollis and speech disturbance. The symptom of poor coordination began at ~47 years of age. One year later, she experienced intermittent, irregular brief muscle jerks in the neck and upper extremities. She also developed progressive gait ataxia, mild dysarthria, bucking while eating or drinking, in-coordination of her hands, and increasing muscle weakness in the neck two years from onset. The disease progressed slowly and clinical severity correlated with the duration of the disease. Wheelchair use was required 10 years after onset.

On admission at the age of 60, the general physical examination was unremarkable. Neurological examination showed severe cerebellar gait ambulate with assistance, intention tremor in both arms and mild cerebellar dysarthria. There was prominently intermittent, irregular neck muscle contraction, leading to rotatory movement of the head and upper extremities. Eye movements were intact in all directions; no saccadic movements and nystagmus were observed. She had difficulty standing with feet together, but was able to ambulate with assistance. She had truncal ataxia, finger-to-nose dysmetria and trouble accomplishing the heel-to-knee test. Tendon hyperreflexia and Babinski’s sign were found. MRI of the brain at the age of 57 years showed moderate and mild atrophy in the cerebellum and brainstem, respectively (Fig. 1C). The International Ataxia Cooperative Rating Scale (ICARS) (Trouillas et al., 1997) and Scale for the Assessment and Rating of Ataxia (SARA) (Weyer et al., 2007), used to quantify the severity of cerebellar symptoms, were 66 and 22, respectively (Table 1).

The clinical features of the nine affected Family CS members included in this study are summarized in Table 1. The age of onset ranged from 40–48 years (mean 43.7 ± 2.9 years) and disease duration varied from 5–31 years (mean 15.9 ± 8.8 years). There was no obvious anticipation of age of onset between generations. Walking difficulty and cerebellar dysarthria were early features in all individuals, while ataxia of the upper extremities was a relatively late symptom. After onset, all patients displayed a slow rate of progression. The International Ataxia Cooperative Rating Scale and Scale for the Assessment and Rating of Ataxia scales clearly showed that clinical severity grossly correlated with the duration of the disease; after disease duration of ~10 years, most patients needed unilateral support or wheelchairs outdoors, while patients with a disease duration of >20 years, showed position sense defect and Babinski’s signs. Four of the nine patients had hand tremors; eight had brisk reflexes and five had extensor plantar responses. Three other patients (II:16, III:3 and III:6) also had spasmatic torticollis as did the index (III:7), but their symptoms were relatively mild. Cognitive deterioration, epilepsy, signs of saccadic eye movements, nystagmus, visual problems, ophthalmoplegia and peripheral nerve involvements were absent in all patients.

The proband (II:2) of Family LY (Fig. 1B) is a 55-year-old female who presented at 40 years with a progressive cerebellar syndrome.
Table 1 Clinical features in nine affected members of Family CS

<table>
<thead>
<tr>
<th></th>
<th>Family CS</th>
<th>Family LY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>II/16</td>
<td>II/17</td>
</tr>
<tr>
<td>Sex</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Age at exam (years)</td>
<td>70</td>
<td>68</td>
</tr>
<tr>
<td>Age at onset (years)</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>Disease duration (years)</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>Walking ability</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td>Upper/lower limb ataxia</td>
<td>++/+++</td>
<td>++/+++</td>
</tr>
<tr>
<td>Nystagmus Horizontal/Vertical</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Saccade slowing</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Spasticity</td>
<td>++/+++</td>
<td>++/+++</td>
</tr>
<tr>
<td>Brachial palsy</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Tremor</td>
<td>Hand</td>
<td>Right hand</td>
</tr>
<tr>
<td>Spasmodic tics</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Position sense</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>Tendon reflexes</td>
<td>N</td>
<td>↑</td>
</tr>
<tr>
<td>Ankle tone</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Babinski’s sign</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SARA</td>
<td>56</td>
<td>59</td>
</tr>
<tr>
<td>Brain MRI</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

Clinical signs are graded as follows: – = absent; ± = subtle; + = mild; ++ = moderate; +++ = severe; ↑ = increased; N = normal; D = defect; ND = not determined; CA = cerebellar atrophy; w = wheelchair; u = unilateral support; I = independent.

characterized by lower limb ataxia and severe dysarthria. At the age of 55, the patient was wheelchair bound. Upon neurological examination at age 55, she showed severe cerebellar gait, intention tremor in arms, severe cerebellar dysarthria and slow eye movements. No nystagmus, mental retardation or sensory injuries were detected. MRI showed marked cerebellar atrophy (data not shown). Her younger sister manifested with a progressive cerebellar syndrome at 41 years of age. Our neurological examination showed dysarthria, slow eye movements and mild movement in-coordination when examined at 45 years of age. MRI showed mild atrophy of the cerebellum (data not shown). Their father had experienced poor coordination and speech disturbance beginning at the age of 40 years; he died at the age of 55 due to an accident. No other information was available about the parents.

**Linkage mapping of the syndrome to chromosome 20p13-12.2**

Parallel inspection of SNP data from Family CS identified a single shared region on chromosome 20p13-12.2 flanked by SNP markers rs12624577 (proximally) and rs674630 (distally) with a maximum two-point logarithm of the odds score of 3.85 (p = 0.00) at rs976192 (Supplementary Table 1). A maximum two-point logarithm of the odds score of 5.36 (p = 0.00) was obtained at D20S437 using thirteen additional markers for fine mapping (Supplementary Table 2). The highest probability haplotype in chromosome 20p13-12.2. Another important recombination event was present between D20S115 and D20S917 in III/3, mapping the disease gene centromeric border to marker D20S917. Affected descendents IV:1 and IV:2 of this individual inherited this recombinant chromosome. According to these two key recombinations, a particular single haplotype, comprising identical alleles spanned by microsatellite markers D20S199 and D20S917 was identified to cosegregate with the disease phenotype in all examined affected family members with SCA and in two clinical unknown members (III:21 and IV:2). This locus was found to span an 18.45 cM region, ~8.4 Mb of genomic DNA, and included 91 reference genes (Fig. 1D).

**Exome sequencing identifies a candidate gene**

We sequenced the exomes of four affected individuals in Family CS (SampleIII:6, III:7, III:17 and IV:1, Table 2). We generated an average of 4.5 billion bases of sequence per affected individual as paired-end, 76 bp reads; 4.1 billion bases (91.14%) passed the quality assessment and aligned to the human reference sequence; and 49.82% of the total bases mapped to the targeted bases with a mean coverage of 65.1-fold. At this depth of coverage, ~97% of the targeted bases were sufficiently covered to pass our thresholds for variant calling. Using de novo assembly of exon sequences (Li et al., 2010), we also detected 125 coding insertions and 226 deletions. After identification of variants, we focused only on non-synonymous (NS) variants, splice acceptor and donor site mutations (SS), and short, frame-shift coding insertions or deletions (indel) that were more likely to be pathogenic mutations.
found that the results were also sufficient to reveal the four affected samples (Sample III:6, III:7, III:17 and IV:1) and pared all possible permutations of two affected individuals from the SNP release of the 1000 Genome Project as filters, we combined the predicted results (Table 3, column 2–5). These linkage variants against dbSNP129, the HapMap 8 and HapMap samples (HapMap 8), and the SNP release of the 1000 Genome Project (20100208 release) (Table 3; row 2, 3 and 4), and removed the shared SNPs. After filtering, the candidate gene pool was reduced ~20-fold compared to the whole NS/SS/Indel variant set. Considering that most pathogenic variants either affect highly conserved sequences and/or are predicted to be deleterious, we used PolyPhen to assess the non-synonymous variants for a likely functional impact (Table 3, row 5) (Sunnyav et al., 2001). This further reduced the candidate gene pool ~30-fold compared to the whole NS/SS/Indel variants. Because SCA in this family is an autosomal dominant inherited disease, all affected individuals should share the same causal variant. Thus comparison of exome data from four samples (III:6, III:7, III:17 and IV:1 in Family CS) to identify the shared variant were sufficient to identify TGM6 as the sole candidate gene, which had a variant c.1550T–G transition (L517W) in exon 10.

By taking advantage of the linkage results to find the shared variant in the locus region, using dbSNP129, the HapMap 8 and the SNP release of the 1000 Genome Project as filters, we compared all possible permutations of two affected individuals from the four affected samples (Sample III:6, III:7, III:17 and IV:1) and found that the results were also sufficient to reveal TGM6 as the only shared candidate gene (Table 3, column 9). In addition, the number of genes for each sample could be narrowed down to between three and six candidates in the mapped locus genomic region; TGM6 was also identified as a candidate gene when combining the predicted results (Table 3, column 2–5). These linkage analysis results were consistent with the exome data from four samples with the identification of TGM6 as the candidate gene.

**Mutation of the TGM6 gene**

To assess the likelihood that the variant c.1550T–G transition (L517W) of TGM6 had a functional impact on the protein, we first used PolyPhen to predict the biophysical consequences of this change. We found that the variant occurred in a highly conserved position (Fig. 2A–C and G) and was likely to be functionally damaging. We then used Sanger sequencing to screen all clinically affected subjects in Family CS, the clinically unknown subjects, and 25 presumptive unaffected family members. All nine patients were heterozygous for this mutation and no clinically unaffected family members carried this variant, except for two clinically unknown (presymptomatic) individuals (III:21 and IV:2) who were also heterozygous. These two individuals were also confirmed to be carriers using single haplotype analysis (Fig. 1A). Thus, the mutation in TGM6 completely cosegregated with the SCA phenotype within this family. Additionally, this mutation was not detected in 500 unaffected control individuals of matched geographical ancestry who were examined using Sanger sequencing, providing further evidence that it is not a polymorphism.

To further confirm that the mutation in TGM6 was responsible for the SCA subtype that maps to chromosome 20p13-12.2, we sequenced all the exons of the TGM6 gene in probands of a cohort of 84 additional unrelated autosomal dominant SCA pedigrees where the known SCA-causative gene mutations had been excluded. Here, in one of the families, a two-generation Chinese SCA family (Family LY; Fig. 1B), we again found a mutation in TGM6: a missense variant c.980A–G transition (D327G) in exon 7 in affected persons II:2 and II:3; i.e. different from that in Family CS (Fig. 2D–G). We carried out additional confirming analyses, as those done for Family CS, and found that this novel variant was at a highly conserved position, was predicted to be damaging, completely cosegregated with the SCA phenotype in Family LY, and was absent among the 500 geographical matched controls.

**Discussion**

We have identified a novel SCA causative gene, TGM6, which had different missense mutations in two different autosomal dominant SCA families [c.1550T–G transition (L517W) in exon 10 and c.980A–G transition (D327G) in exon 7]. We provide several lines of evidence to support the contention that the TGM6 gene is the pathogenic causative gene: (i) the missense variant identified in Family CS via exome sequencing was also located in the associated region we identified using linkage analysis, providing two separate experimental methods for localizing the gene of interest; (ii) a different missense mutation was identified in another SCA family (Family LY); (iii) these variants completely cosegregated with the disease phenotype in Family CS, including in two pre-symptomatic carriers (III:21 and IV:2) where the mutations were confirmed by single haplotype analysis, and in Family LY; (iv) both TGM6 mutations were present at a conserved site among different vertebrates; (v) both mutations were predicted to be functionally damaging by PolyPhen; (vi) neither mutation was seen in 500 unaffected healthy

---

**Table 2 Summary of original exome sequencing data**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Bases (Gbp)</th>
<th>Map Bases (Gbp)</th>
<th>Map Bases Rate (%)</th>
<th>Exon Map Bases (Gbp)</th>
<th>Exon Map Bases Rate (%)</th>
<th>Exon Length (Mbp)</th>
<th>Covered Length (Mbp)</th>
<th>Coverage (%)</th>
<th>Mean Depth Mode Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample III:6</td>
<td>4.23</td>
<td>3.92</td>
<td>92.75</td>
<td>2.33</td>
<td>55.16</td>
<td>34.11</td>
<td>33.98</td>
<td>99.62</td>
<td>68.39</td>
</tr>
<tr>
<td>Sample III:7</td>
<td>4.16</td>
<td>3.79</td>
<td>91.10</td>
<td>1.93</td>
<td>46.42</td>
<td>34.11</td>
<td>34.01</td>
<td>99.71</td>
<td>56.67</td>
</tr>
<tr>
<td>Sample III:17</td>
<td>4.25</td>
<td>3.93</td>
<td>92.51</td>
<td>2.20</td>
<td>51.77</td>
<td>34.11</td>
<td>33.91</td>
<td>99.43</td>
<td>64.46</td>
</tr>
<tr>
<td>Sample IV:1</td>
<td>5.26</td>
<td>4.64</td>
<td>88.21</td>
<td>2.42</td>
<td>45.94</td>
<td>34.11</td>
<td>33.99</td>
<td>99.65</td>
<td>70.89</td>
</tr>
<tr>
<td>Mean</td>
<td>4.48</td>
<td>4.07</td>
<td>91.14</td>
<td>2.22</td>
<td>49.82</td>
<td>34.11</td>
<td>33.97</td>
<td>99.60</td>
<td>65.10</td>
</tr>
</tbody>
</table>

---

Downloaded from https://academic.oup.com/brain/article-abstract/133/12/3510/311482 by guest on 17 July 2018
Table 3 Direct identification of the causal gene for Family CS by exome sequencing

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NS/SS/Indel</td>
<td>5796/34</td>
<td>5649/40</td>
<td>5780/40</td>
<td>5842/37</td>
<td>3964/26</td>
<td>3099/26</td>
<td>2434/20</td>
<td>3736-3964/26-30</td>
</tr>
<tr>
<td>Not in dbSNP 129</td>
<td>869/6</td>
<td>734/9</td>
<td>931/8</td>
<td>891/8</td>
<td>288/3</td>
<td>134/3</td>
<td>68/2</td>
<td>207-288/3-5</td>
</tr>
<tr>
<td>Not in dbSNP 129, nor in eight HapMap exomes</td>
<td>616/6</td>
<td>520/6</td>
<td>674/7</td>
<td>661/7</td>
<td>153/5</td>
<td>43/3</td>
<td>15/2</td>
<td>87-155/3-4</td>
</tr>
<tr>
<td>Not in dbSNP 129, eight HapMap exomes, nor in dbSNP thousands genome</td>
<td>309/4</td>
<td>262/3</td>
<td>341/6</td>
<td>384/5</td>
<td>75/1</td>
<td>5/1</td>
<td>1/1</td>
<td>48-101/1</td>
</tr>
<tr>
<td>Predicted to be damaging</td>
<td>211/1</td>
<td>203/1</td>
<td>214/1</td>
<td>212/1</td>
<td>48/1</td>
<td>3/1</td>
<td>1/1</td>
<td>36-52/1</td>
</tr>
</tbody>
</table>

Rows show the effect of excluding consideration variants found in dbSNP129, the eight HapMap exomes, the dbSNP 1000 genomes or all. Columns show the effect of requiring that NS/SS/Indel variants be observed in each affected individual (Columns 2–5) or two to four affected individuals (Columns 6–8). Column 9 provide ranges of observations that occur when all possible permutations of two affected individuals (sample III:6, III:7, III:17, IV:1) are used. (Whole/Locus): indicate NS/SS/Indel variants be observed in whole exome or in Locus region by linkage analysis.
genetically heterogeneous disorders. Therefore different kindreds may present with similar disease phenotypes, but are the result of different mutated genes. Next-generation sequencing technologies using exome capture have been successfully used for causative gene identification in several rare monogenic diseases. Current studies, however, have used different patients from several unrelated families to identify the causative gene (Choi et al., 2009; Ng et al., 2009, 2010). However, it is difficult to detect the causative gene from Sanger sequencing.
genetic mutations for the clinically and genetically heterogeneous disorders such as SCA. Here, we identified the SCA causative gene using four samples from affected individuals, all from one family.

We further showed that by using a strategy that combined linkage mapping, exome sequencing and functional predicted results, we can quickly identify the causal mutation using only one sample from a single family. To date, fewer than 2400 genes have been shown to be responsible for Mendelian phenotypes. Genetic alterations responsible for over 1600 other mapped Mendelian disorders have not been identified (Amberger et al., 2009). Considering the huge number of rare monogenic diseases that remain uncharacterized, but have linkage data, our combinational approach seems to be a promising, highly accurate and affordable method for identifying a large number of these genes using a very small number of samples.

In conclusion, we found a novel causative gene, TGM6, of SCA. Further functional studies of these specific gene variants are needed to gain new insights into the pathogenic mechanism(s) involved in SCA. We also demonstrate that whole-exome sequencing of just a few samples from one family can be used to identify the causative gene and further, that identification of disease-causing genes can be achieved using the combined strategy of exome sequencing and linkage analysis with even fewer samples required. This approach together with continuing improvement of sequencing technology will make medical genomics of rare Mendelian disorders a reality.

Acknowledgements

We are indebted to all the patients and family members for their generous participation in this work.

Funding

This work was supported by Projects in the Major State Basic Research Development Program of China (973 Program) (Grant number 2006cb500700 and 2011CB510000) and the National Natural Science Foundation of China.

Supplementary material

Supplementary material is available at Brain online.

References


