

Published in final edited form as:

*Ann Neurol.* 2011 March ; 69(3): 464–470. doi:10.1002/ana.22235.

## Exome Sequencing Allows for Rapid Gene Identification in a Charcot-Marie-Tooth Family

Gladys Montenegro, BSc<sup>1</sup>, Eric Powell, BSc<sup>1</sup>, Jia Huang, PhD<sup>1</sup>, Fiorella Speziani, MA<sup>1</sup>, Yvonne J.K. Edwards, PhD<sup>1</sup>, Gary Beecham, PhD<sup>1</sup>, William Hulme, BSc<sup>1</sup>, Carly Siskind, MSc<sup>2</sup>, Jeffery Vance, MD, PhD<sup>1</sup>, Michael Shy, MD<sup>2</sup>, and Stephan Züchner, MD<sup>1</sup>

<sup>1</sup> John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL

<sup>2</sup> Department of Neurology and Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, MI

### Abstract

**Objective**—Charcot-Marie-Tooth (CMT) disease comprises a large number of genetically distinct forms of inherited peripheral neuropathies. The relative uniform phenotypes in many patients with CMT make it difficult to decide which of the over 35 known CMT genes are affected in a given patient. Genetic testing decision trees are therefore broadly based on a small number of major subtypes (eg, CMT1, CMT2) and the observed mutation frequency for CMT genes. Since conventional genetic testing is expensive many rare genes are not being tested for at all.

**Methods**—Whole-exome sequencing has recently been introduced as a novel and alternative approach. This method is capable of resequencing a nearly complete set of coding exons in an individual. We performed whole-exome sequencing in an undiagnosed family with CMT.

**Results**—Within over 24,000 variants detected in 2 exomes of a CMT family, we identified a nonsynonymous *GJB1* (*Cx32*) mutation. This variant had been reported previously as pathogenic in X-linked CMT families. Sanger sequencing confirmed complete cosegregation in the family. Affected individuals had a marked early involvement of the upper distal extremities and displayed a mild reduction of nerve conduction velocities.

**Interpretation**—We have shown for the first time in a genetically highly heterogeneous dominant disease that exome sequencing is a valuable method for comprehensive medical diagnosis. Further improvements of exon capture design, next-generation sequencing accuracy, and a constant price decline will soon lead to the adoption of genomic approaches in gene testing of Mendelian disease.

In the past 20 years a large number of inherited neurological diseases have been deciphered and thousands of underlying genes have been identified. Once-singular clinical entities were genetically broken up into many subtypes. As much as this offers opportunities for basic research in understanding the underlying pathways and key proteins in a given disease, there are a number of complicating factors for the practicing neurologist. It is now increasingly difficult to truly know all the new genetically-defined forms of neurological disease. Furthermore, with current technology the genetic diagnosis is often cumbersome, expensive,

Address correspondence to Dr. Züchner, University of Miami Miller School of Medicine, Biomedical Research Building (BRB), Room 523, LC: M-860, 1501 NW 10th Avenue, Miami, FL 33136. szuchner@med.miami.edu.

#### Potential conflict of interest

C.S., E.P., G.B., G.M., J.H., J.V., M.S., C.Z., W.H., and Y.E. have received a research grant from the NIH.

Additional Supporting Information can be found in the online version of this article.

and in most cases does not evaluate all possible genes. In other words, the genetic knowledge is not being applied to the degree that would be warranted.

Charcot-Marie-Tooth disease (CMT) is a genetically heterogeneous disease that clinically falls into 2 major groups: CMT1, a primary demyelinating condition with reduced nerve conduction velocities (NCV); and CMT2, the axonal form of the disease. More than 35 genes have been identified thus far for autosomal dominant, recessive, and X-linked CMT forms.<sup>1–3</sup> However, for CMT2, only ~30% of the genetic causes are known and a considerable number of additional genes will be identified in the future.<sup>4</sup>

To improve genetic diagnostic procedures it is predicted that new developments in sequencing technology will soon play a crucial role. A recent study by Lupski and colleagues<sup>5</sup> demonstrated that whole human genome sequencing is capable of identifying mutations in a known gene for a recessive CMT. The authors estimated that such a study would still cost upward of \$50,000 per patient, however. An alternative approach has recently been successfully introduced, called exome sequencing.<sup>6</sup> This method allows for a targeted enrichment and resequencing of nearly all exons of protein-coding genes. Protein-coding exons account for only ~1% of the human genome, but 85% of Mendelian diseases are caused by mutations in this genomic space.<sup>7</sup> Thus, exome sequencing could provide a medical genetics diagnosis tool that is applicable today.<sup>8</sup> Indeed, recent reports identified genes for rare recessive syndromes using this approach.<sup>8,9</sup> To evaluate whole-exome sequencing for genetic diagnosis of more ubiquitous Mendelian diseases, CMT is an attractive model: (1) CMT is 1 of the most common inherited diseases in neurology with 1 in 2500 individuals affected<sup>10</sup>; (2) CMT is one of the genetically most heterogeneous diseases with >35 genes identified; and (3) CMT involves recessive as well as dominant disease alleles on autosomes and the X chromosome.

We have applied exome sequencing to a family with axonal to intermediate CMT that had been tested for the most common CMT2 gene *MFN2*. A self-reported family history that indicated male-to-male transmission argued against screening for X-linked CMT genes. The results of the exome sequencing screen in two affected family members of this family did allow for exclusion of known CMT genes and identified a previously reported mutation in the gap junction protein 1 (*GJB1*) also known as connexin 32 (*Cx32*).

## Patients and Methods

### Human Subjects

We studied a family of European descent ascertained in North America with CMT, consisting of 6 affected, 2 unclear/questionably affected, and 5 apparently unaffected individuals (Fig 1). Affected patients had been excluded for mutations in the gene *MFN2*. From all study participants, blood (~24ml) was collected in either ethylene diamine tetraacetic acid (EDTA) or acid citrate dextrose tubes and DNA was extracted in the biorepository of the Hussman Institute for Human Genomics. Informed consent was obtained from all individuals, and the Institutional Review Board (IRB) at the University of Miami Miller School of Medicine approved the study.

### Exome Sequencing

Enrichment of coding exons and flanking intronic regions totaling 38Mb was performed with the SureSelect human all exon kit (Agilent) following the manufacturer's standard protocol. The kit covers 1.22% of the human genome that corresponds to the consensus coding sequence (CCDS) collection of genes. The protocol allowed for integration of Illumina's Genome Analyzer II adapter sequences. Enriched DNA was subjected to standard

sample preparation for the Genome Analyzer II instrument. Two flow cell channels were sequenced per individual.

### Sanger Sequencing

Sanger sequencing was performed to exclude mutations in all coding exons and flanking intronic regions of *MFN2* in the affected patients of the studied family. We also applied Sanger sequencing to confirm the identified *GJB1* mutation and to demonstrate cosegregation with the phenotype in 13 family members. Exons and flanking intronic sequences were amplified on Applied Biosystems (ABI) Veriti 96-well Fast Thermal Cyclers using a touchdown protocol. Polymerase chain reaction (PCR) purification was completed with QuickStep<sup>TM</sup>2 SOPE resin (Edge BioSystems). Sequencing was performed using ABI BigDye Dye Terminator Cycle Sequencing Kit on an ABI 3730 sequencer. Sequence traces were analyzed using Sequencher (version 4.8; Gene Codes Corporation).

### Bioinformatics Analysis

Raw sequencing data were analyzed using version 1.6 of the Illumina analysis pipeline (Gerald). Sequence reads were aligned with the MAQ software. Variants were called with MAQ applying the standard parameters. All variants were submitted to Seattle Seq for further categorization into coding, noncoding, and novel single-nucleotide polymorphisms (SNPs). Resulting data were converged, filtered, and ranked by Genomic Evolutionary Rate Profiling (GERP) conservation scores.<sup>11</sup> All CMT genes were inspected for sufficient coverage and the presence of variations.

## Results

### Clinical Characteristics of the Studied CMT Family

The male patient 1007 of the family (see Fig 1) had normal developmental milestones. At the age of 18 years he developed ankle swelling and was found to have high arches. Gradually over several years he noted that he was prone to twisting his ankle while walking. In his 30s he developed difficulties walking due to the development of foot drop. He started using bilateral ankle-foot orthosis (AFO) and his hand function started deteriorating so that daily tasks such as opening door knobs or writing became difficult. He noted transient pains in hands and forearms and he was aware of a persistent numbness in both feet and hands. In his 40s he had several seizures. No other specific neurological symptoms were reported. The neurological exam revealed normal tone and strength of the proximal extremities. There was marked atrophy of the thenar eminence muscles and weakness of opposition in the hands. The lower extremities showed moderate to severe distal weakness and atrophy most prominently affecting the ankle dorsiflexor and intrinsic foot muscles. Sensory decrease was present for all modalities below elbows and knees. Achilles reflexes were absent; there was no Babinski sign. Electrophysiological studies showed reduced amplitude of the peroneal compound muscle action potential (CMAP) of 0.5mV (normal >2mV) and slightly slow conduction velocity of 35m/second (normal >41m/second). The ulnar CMAP amplitude was 9.5mV (normal >6mV) and CV was 49m/second (normal >49 m/second).

The patient's brother (individual 1000; see Fig 1) had high arches, and the brother's 2 daughters (individuals 0001 and 0100; see Fig. 1) both had flat feet. However, their mother also had flat feet, making clinical interpretation of this branch of the family difficult. One of them (individual 0001) was examined in more detail at 25 years of age. She was noted to have never done well in physical education in school and she often twisted her ankles. She reported some numbness at the tips of her fingers and showed minor signs of hand weakness (finger extension, thumb adduction) and mild distal atrophy in both upper and lower extremities. Reflexes at the ankle were diminished. Her electrophysiological studies showed

a slight reduction of NCV (ulnar 41m/second, normal >49m/second; peroneal 38m/second, and normal >41m/second). At this time her status was assessed as questionably affected with CMT.

The average age of clinical onset in the family was 29.5 years (range, 14–40 years). Two out of 4 females carrying the mutation undoubtedly developed the disease, albeit later than males (see Fig 1). The male affected individuals in this family had a more pronounced involvement of the hands compared to female mutation carriers. The pronounced symptoms in males are to be expected in an X-linked disease. Available electrophysiological studies indicated an axonal or intermediate type of CMT with mild reduction in nerve conduction and amplitudes.

### Exome Sequencing of a CMT Family

Exome sequencing was performed in two individuals of the family, 1007 and 0001 (see Fig 1). Overall, 92 and 93 million sequencing reads were produced for the two samples, comprising 6.9 and 7.0 billion bases, respectively. Approximately 82% of these were aligned to the human reference genome (hg18) and 67% of these fell onto targeted and enriched exons. The average sequence read depth was 57× and 69× in targeted exons. After filtering for a minimal sequence quality score (Phred) of 30 and a minimal sequence read depth of 5× we identified 64,568 and 61,153 single-nucleotide variants in the two individuals, respectively. Of these, 24,150 and 23,607 were coding variants. We further determined functionally significant variants as missense, nonsense, or splice-site affecting changes: 10,817 such variants were detected in case 1007 and 10,621 changes in individual 0001. When we excluded all variants reported in dbSNP version 131, excluding those variants that occurred in neuropathy associated genes, 507 and 697 variants were left in the 2 sequenced patients, respectively. This simple filtering strategy rapidly extracted the novel, functionally significant variants in the coding exons of two human genomes in an unbiased manner. The number of potentially causative variants was reduced by ~99%. Since the 2 sequenced individuals (1007 and 0001) were affected/likely affected they should segregate the same single causative variant. We intersected both datasets and identified only 86 unique changes in 82 genes segregating between both individuals (Supporting Table 1). The analysis of insertions and deletions showed that 146 high-quality insertion/deletions of up to 15bp in length segregated between both sequenced individuals, with 66 of these falling onto coding exons.

### Identification of a Causative CMT Mutation

The only remaining variant that fell on a CMT gene occurred in the gene *GJB1* (*Cx32*), c. 283G>A, V95M. The particular *GJB1* mutation had been reported in multiple families to cause CMT type 1X (CMT1X).<sup>12–17</sup> Sanger sequencing confirmed the presence of this mutation and complete cosegregation with the CMT phenotype (see Fig 1; Fig 2). A focused analysis of 39 neuropathy-related genes in all detected coding variants revealed nonsynonymous changes in 8 additional CMT genes (Table). These changes either did not segregate in the family and/or were known polymorphisms described in dbSNP and not associated with disease. However, it cannot be ruled out that these changes have a modifying effect on the phenotypic expression.

### Discussion

We performed whole-exome sequencing in one individual of a family with CMT and identified the causative mutation in the *GJB1* (*Cx32*) gene, c.283G>A, V95M. This particular variant had been reported, verifying its pathogenicity. The clinical characteristics are comparable to reported data as well: a moderate reduction of nerve conduction, a more

severe phenotype in males compared to females, and the description of involvement of upper distal extremities.<sup>12–17</sup> The atrophy of the thenar muscle was unusually pronounced in some of our patients. However, atrophy of the thenar muscles is a recognized early feature of CMT1X, and when accompanied by intermediately slowed nerve conductions, is in fact, suggestive of CMT1X in the absence of male-to-male transmission.<sup>16</sup>

What prevented us from screening *GJB1* (*Cx32*) with Sanger sequencing in this family? The wider pedigree (not shown) indicated the presence of male-to-male transmission, which would have excluded X-linked inheritance. The presence of such anecdotal evidence of inheritance patterns or symptoms can thus interfere with a rational selection of genetic tests. This raises a larger question; ie, whether it is even still economical to exclude genes on a 1-by-1 basis using Sanger sequencing, when a comprehensive, affordable, and relatively fast method such as exome sequencing is now available. We estimate that it will still be valuable to test the most common genes, including *PMP22*, *MFN2*, *MPZ*, and *GJB1* in CMT, in concordance with the phenotype and inheritance pattern (CMT1 vs CMT2). We have recently found that over 90% of 527 patients with a known genetic diagnosis had CMT caused by mutations in 1 of these 4 genes.<sup>18</sup> However, we have also identified over 200 patients seen in the CMT clinic at Wayne State University (Detroit, MI) for whom no genetic diagnosis has been made. Many of these neuropathies are likely to be caused by novel genes or genes for which it is difficult to obtain testing. Whole-exome sequencing would offer an attractive alternative approach to diagnose the causal gene in these patients. The novel approach of exome sequencing allowed us to finish the sequencing analysis in 8 weeks at a cost of ~\$5000 per individual. We estimate that by the end of 2010 the cost will decrease to below \$2000. This is potentially the most important argument when comparing this technique to whole-genome sequencing, which is still estimated in the range of \$50,000 for a comparable high sequencing-read depth. Read depth, or depth of coverage, relates to the number of individual short sequencing reads that span a given nucleotide. The number of measured nucleotides at a given location is the basis for determining the genotype (see Fig 2A). Thus, low depth of coverage (<8 sequencing reads) is more prone to error. A potential disadvantage of targeted exome sequencing is the poor uniformity of sequence depth, which requires larger sequence amounts. Uniformity in this context refers to the distribution of sequencing reads over a targeted area. There can be significant differences between areas of highest and lowest read depth, which usually requires the production of more sequencing reads to cover as many regions sufficiently as possible (Fig 3). In some genes the current hybridization probe design insufficiently covers alternative exons or areas with specific sequence characteristics (see Fig 3). For instance, the CMT-related genes *NEFL*, *HSN2*, and *SEPT9* were not investigated. However, recently updated exome kits of the second generation with a more comprehensive set of exons also cover these genes. In addition, longer sequencing reads of up to 125 bases on the Illumina platform should also improve the coverage issues of early studies. An general advantage of whole-genome or whole-exome approaches is that even a negative testing result will allow for storing of the complete library of coding variants identified in a patient for potential future in silico identification of mutations in newly-discovered genes.

We also have identified common, previously reported nonsynonymous polymorphisms in 8 other genes related to peripheral neuropathies (see Table). These were present in either or both sequenced individuals. Rather than causing CMT, such changes could contribute to modulating the phenotypic expression of age of onset, severity, and others. In future studies, exome sequencing will allow us to systematically record and correlate such common variants with subphenotypes and variations of clinical expressions in and across families.

In summary, we have successfully applied exome sequencing to diagnose a family with a highly heterogeneous and dominant Mendelian phenotype. The presented results support the



notion that genomic approaches in the form of targeted resequencing or whole-genome sequencing will replace traditional medical genetic testing in the foreseeable future.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

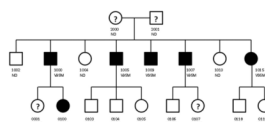
This research was supported by the National Institutes of Health (NIH) (NINDS R01NS052767 to S.Z., NINDS U54NS065712 to M.S. and S.Z.). This study is part of the Inherited Neuropathies Consortium of the Rare Disease Clinical Research Network (INC RDCRC).

We thank the patients for participating in this study. We also acknowledge the support of the Center for Computational Sciences at the University of Miami, namely Drs. Tsinoremas and Zysman, for hosting our data.

## References

1. Bernard R, De Sandre-Giovannoli A, Delague V, Levy N. Molecular genetics of autosomal-recessive axonal Charcot-Marie-Tooth neuropathies. *Neuromolecular Med.* 2006; 8:87–106. [PubMed: 16775369]
2. Kleopa KA, Scherer SS. Molecular genetics of X-linked Charcot-Marie-Tooth disease. *Neuromolecular Med.* 2006; 8:107–122. [PubMed: 16775370]
3. Zuchner S, Vance JM. Molecular genetics of autosomal-dominant axonal Charcot-Marie-Tooth disease. *Neuromolecular Med.* 2006; 8:63–74. [PubMed: 16775367]
4. Verhoeven K, Claeys KG, Zuchner S, et al. MFN2 mutation distribution and genotype/phenotype correlation in Charcot-Marie-Tooth type 2. *Brain.* 2006; 129:2093–2102. [PubMed: 16714318]
5. Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med.* 2010; 362:1181–1191. [PubMed: 20220177]
6. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature.* 2009; 461:272–276. [PubMed: 19684571]
7. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat Genet.* 2003; 33(suppl):228–237. [PubMed: 12610532]
8. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A.* 2009; 106:19096–101. [PubMed: 19861545]
9. Hoischen A, van Bon BW, Gilissen C, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet.* 2010; 42:483–485. [PubMed: 20436468]
10. Skre H. Genetic and clinical aspects of Charcot-Marie-Tooth's disease. *Clin Genet.* 1974; 6:98–118. [PubMed: 4430158]
11. Cooper GM, Stone EA, Asimenos G, et al. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 2005; 15:901–913. [PubMed: 15965027]
12. Bone LJ, Dahl N, Lensch MW, et al. New connexin32 mutations associated with X-linked Charcot-Marie-Tooth disease. *Neurology.* 1995; 45:1863–1866. [PubMed: 7477983]
13. Bone LJ, Deschenes SM, Balice-Gordon RJ, et al. Connexin32 and X-linked Charcot-Marie-Tooth disease. *Neurobiol Dis.* 1997; 4:221–230. [PubMed: 9361298]
14. 3rd workshop of the European CMT consortium: 54th ENMC International Workshop on genotype/phenotype correlations in Charcot-Marie-Tooth type 1 and hereditary neuropathy with liability to pressure palsies 28–30 November 1997, Naarden, The Netherlands. *Neuromuscul Disord.* 1998; 8:591–603. [PubMed: 10093067]
15. Ionasescu VV. X-linked Charcot-Marie-Tooth disease and connexin32. *Cell Biol Int.* 1998; 22:807–813. [PubMed: 10873293]
16. Hahn AF, Bolton CF, White CM, et al. Genotype/phenotype correlations in X-linked dominant Charcot-Marie-Tooth disease. *Ann N Y Acad Sci.* 1999; 883:366–382. [PubMed: 10586261]

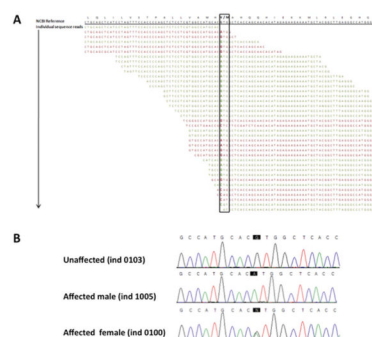
17. Park HK, Kim BJ, Sung DH, et al. Mutation analysis of the PMP22, MPZ, EGR2, LITAF, and GJB1 genes in Korean patients with Charcot-Marie-Tooth neuropathy type 1. *Clin Genet*. 2006; 70:253–256. [PubMed: 16922730]
18. Saporta ASD, Sottile SL, Miller LJ, et al. Charcot Marie Tooth (CMT) subtypes and genetic testing strategies. *Ann Neurol*. (in press).



**FIGURE 1.**

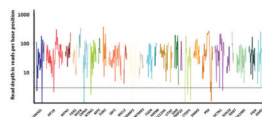
Pedigree of the studied family. *Filled black symbols* are phenotypically affected individuals. Each individual has an identifier and the mutational state is indicated. Question marks refer to unclear phenotypes. While (individual 2000) was not genetically tested and phenotypic information is unclear, it can be assumed that she was the carrier of the mutation, based on the inheritance pattern of CMT1X.





**FIGURE 2.**

Identification of a nonsynonymous mutation in *GJB1*. (A) Next-generation sequencing output at the position of the identified mutation in *GJB1* (black box). A coverage depth of 36 individual sequencing reads covered this particular site, with 22 of them being forward-aligned reads (green) and 14 being reverse-aligned reads (red). Out of 36 reads 12 (33%), showed the mutation in a female individual. (B) Sanger sequencing traces confirmed the change c.283G>A, V95M. As expected, male individuals showed a homozygous change and female participants carried a heterozygous change.



**FIGURE 3.**

Typical sequence coverage (log<sub>10</sub> scale) of peripheral neuropathy–related genes. Exome sequencing results in a range of sequence coverage depth. Some areas of targeted CMT genes (*x*-axis) were insufficiently covered in the current version of the whole-exome kits (<5 reads, *dotted line*). This will likely improve in future versions of exome-enrichment kits and with increasing sequence read length from ~75 bases to up to 125 bases on the Illumina Genome Analyzer platform.

**TABLE**

Sequence Changes Identified in Peripheral Neuropathy–Related Genes

Gene	Change	Rs Numbers in dbSNP	Segregation Between Individuals 1007 and 0001
<i>ARHGEF10</i>	–	–	–
<i>BSCL2</i>	–	–	–
<i>CTDP1</i>	THR340MET	2279103	Yes
<i>DCTN1</i>	–	–	–
<i>DNM2</i>	VAL818MET	Novel	No
<i>EGR2</i>	–	–	–
<i>FGD4</i>	–	–	–
<i>FIG4</i>	–	–	–
<i>GAN</i>	–	–	–
<i>GARS</i>	–	–	–
<i>GDAP1</i>	–	–	–
<i>GJB1</i>	VAL95MET	Novel	Yes
<i>HSPB1</i>	–	–	–
<i>HSPB8</i>	–	–	–
<i>IGHMBP2</i>	THR671ALA	622082	No
	LEU201SER	560096	Yes
<i>IKBKAP</i>	CYS1072SER	3204145	No
	PRO1158LEU	1538660	Yes
<i>KIAA1985</i>	–	–	–
<i>KIF1B</i>	–	–	–
<i>LITAF</i>	–	–	–
<i>LMNA</i>	–	–	–
<i>MFN2</i>	–	–	–
<i>MPZ</i>	–	–	–
<i>MTMR2</i>	GLU502GLN	61735578	No
<i>NDRG1</i>	–	–	–
<i>NGFB</i>	–	–	–
<i>NTRK1</i>	–	–	–
<i>PLEKHG5</i>	SER17ARG	17438786	Yes

Gene	Change	Rs Numbers in dbSNP	Segregation Between Individuals 1007 and 0001
<i>PMP22</i>	–	–	–
<i>PRX</i>	GLY1132ARG	268674	No
	ILE921MET	268673	No
<i>RAB7</i>	–	–	–
<i>SBF2</i>	–	–	–
<i>SETX</i>	ILE2587VAL	1056899	No
<i>SLC12A6</i>	–	–	–
<i>SPTLC1</i>	–	–	–
<i>YARS</i>	–	–	–
<i>AARS</i>	–	–	–
<i>TRPV4</i>	–	–	–
<i>PRPS1</i>	–	–	–
<i>SOX10</i>	–	–	–