doi:10.1111/ncn3.72

ORIGINAL ARTICLE

Exome sequencing shows a novel *de novo* mutation in *ATL1*

Kishin Koh,[1] Hiroyuki Ishiura,[2] Michiaki Miwa,[1] Koichiro Doi,[3] Jun Yoshimura,[3] Jun Mitsui,[2] Jun Goto,[2] Shinichi Morishita,[3] Shoji Tsuji[2,4] and Yoshihisa Takiyama[1]

[1]Department of Neurology, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, [2]Department of Neurology, Graduate School of Medicine, The University of Tokyo, [3]Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, and [4]Medical Genome Center, The University of Tokyo Hospital, Tokyo, Japan

Key words
ATL1, *de novo* mutation, exome sequencing, spastic paraplegia, spastic paraplegia type 3A.

Accepted for publication 13 January 2014.

Correspondence
Shoji Tsuji, Department of Neurology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. E-mail: tsuji@m.u-tokyo.ac.jp

Abstract

Background and Aim: We experienced treating a male patient born to healthy parents, who presented with spastic gait from childhood, and his son and daughter similarly presented with spastic gait, raising the possibility of a *de novo* mutation in the patient. With the hypothesis of a *de novo* mutation in the patient, we carried out exome sequencing of the patient and his parents to identify the gene with a disease-causative mutation in the patient.

Methods: The genomic DNA obtained from the patient and his parents were subjected to exome sequencing. We applied various filters to identify candidate *de novo* mutations.

Results: We identified three *de novo* mutations, namely, *PLB1* c.3601T>A p.C1201S, *PHF2* c.2678C>T p.S893L and *ATL1* c.1259A>C p.Q420P. Two of these mutations (*ATL1* and *PHF2*) cosegregated with the disease phenotype in this family. Because the clinical presentations of the affected individuals in this family are typical for spastic paraplegia type 3A, the novel mutation (p.Q420P) in *ATL1* is likely the cause of early-onset spastic paraplegia in this family.

Conclusion: In the present, study we confirm the efficacy and usefulness of exome sequencing of parent–child trios for establishing the molecular diagnosis of patients with neurological diseases, in whom *de novo* mutations are suspected.

Introduction

The search for *de novo* mutations has been intensively carried out to identify disease-causative mutations focusing on patients with severe phenotypes and early ages at onset, and who were born to healthy parents. This strategy has been successfully applied to the discovery of causative genes for many diseases including those for intellectual disability,[1] autism[2] and schizophrenia.[3,4] For disorders hypothesized to result from *de novo* mutations, a potentially powerful approach is exome sequencing of parent–child trios, which should facilitate the identification of *de novo* mutations present only in the patients.

We have recently experienced treating a male patient presenting with spastic gait from childhood, and his son and daughter similarly presented with spastic gait, suggesting an autosomal dominant mode of inheritance. His parents, however, were unaffected. With the hypothesis of a *de novo* mutation in the patient, we carried out exome sequencing of the patient and his parents to identify the gene with a disease-causative mutation in the patient.

Methods

Patients. The pedigree chart of this family is shown in Figure 1. The patient (II-2) presented with spastic gait from the age of 4 years, and his son (III-1) and daughter (III-2) also presented with spastic gait from the age of 2 years. The patient was examined by neurologists, and underwent physiological, radiological and biochemical examinations among others. The patient showed only lower limb spasticity and exaggerated tendon reflexes with extensor plantar responses without sensory or cranial nerve dysfunctions, loss of vibratory sense, neurogenic bladder or cognitive decline. Brain magnetic resonance imaging (MRI) of the patient did not show any abnormalities including thickness of the corpus callosum. MRI findings of the spinal cord were unremarkable. His children (III-1 and III-2) showed mild spastic gait and exaggerated tendon reflexes with extensor plantar responses. On the basis of these observations, we made the diagnosis of autosomal dominant hereditary spastic paraplegia, although the parents (I-1 and I-2) did not show any neurological abnormalities on examinations by neurologists.

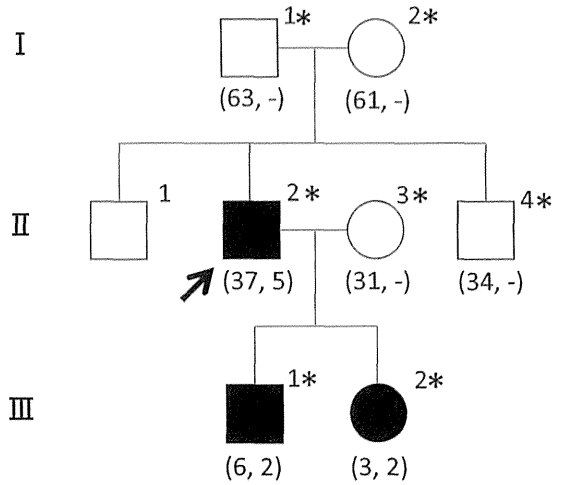

Figure 1 Pedigree chart. Squares and circles indicate male and female, respectively. Filled symbols indicate affected individuals, whereas open symbols indicate unaffected individuals. The arrow indicates the patient; (*x*, *y*) indicates age at examination and onset age. *Individuals from whom genomic DNA were collected.

The brother of the proband (II-4) did not show any neurological abnormalities either.

Methods. Written informed consent was obtained from all the participants. The children's informed assent was obtained with the parents' consent. The genomic DNA obtained from the patient and his parents were subjected to exome sequencing. Exome capture was carried out using an Agilent SureSelect Human ALL Exon 50 Mb Kit (Agilent, Santa Clara, CA, USA), followed by massively parallel sequencing using Illumina Hiseq2000 (100 bp paired end; Illumina, San Diego, CA, USA). We aligned the exome data by Burrows-Wheeler Aligner,[5] and extracted single-nucleotide variations (SNV) using SAMtools.[6] When these SNV are used to identify variations that are present in the patient, but not in the parents, too many candidate *de novo* mutations are listed, presumably attributable to errors originating at various steps. To more efficiently identify candidate *de novo* mutations, we applied the following filters, which were previously applied for the identification of *de novo* mutations:[7]

1 Use variations that are homozygous references in both parents and are supported by 10 or more high-quality reads at the mutated sites for each member of the trio.
2 Use a reliable homozygous reference in each parent such that the likelihood of heterozygosis, ${}_nC_i\ (1/2)^i(1/2)^{n-i}$, is lower than the likelihood of homozygosis, ${}_nC_i\ (999/1000)^i(1/1000)^{n-i}$, where the average error rate is assumed to be one in 1000, n represents the number of total reads, and i is the number of reads consistent with the reference.
3 Use reliable *de novo* mutations of the proband such that the number of alternative allele reads is at least 30% among the total reads, which is the condition proposed by Kong *et al.*[8]

Candidate *de novo* mutations were validated by Sanger sequencing.

Single nuceotide polymorphism (SNP) typing was performed carried out using Genome-Wide Human SNP array 6.0 (Affymetrix, Santa Clara, CA, USA), followed by linkage analysis and haplotype reconstruction using SNP-HiT-Link,[9] and Allegro version 2.[10] The present study was approved by the institutional review boards of the University of Tokyo and the University of Yamanashi.

Results

We identified 10 502, 10 541, and 10 667 variations in the father, mother and patient, respectively (Table 1). The candidate variants were selected as described in the Methods section. First, by selecting the variations that are supported by read depths of 10 times or more high-quality reads, we obtained 9798, 10 020, and 10 171 variations in the father, mother, and patient, respectively (Table 1). Second, by filtering for a reliable homozygous reference in each parent, we were able to narrow the candidate variations to 12. By the third filter in which the number of alternative allele reads in the patient is at least 30% of the total number of reads, we found five variations as the candidate *de novo* mutations (Table 2).

By validating the five candidate variations by Sanger sequencing, we identified three *de novo* mutations, namely, *PLB1* c.3601T>A p.C1201S, *PHF2* c.2678C>T p.893L and *ATL1* c.1259A>C p.Q420P (Fig. 2). *PLB1* c.3601T>A p.C1201S and *ATL1* c.1259A>C p.Q420P are novel mutations present in neither dbSNP137 in the Human Gene Mutation Database nor the in-house variation database including data from Japanese control subjects (n = 373). *PHF2* c.2678C>T p.893L is registered in dbSNP137 as rs149241220, although its frequency is one in 4300 in the European American population (NHLBI Exome Sequencing Project). It is registered in neither the Human Gene Mutation Database nor the in-house variation database. The *ATL1* mutation (c.1259A>C p.Q420P) and *PHF2* mutation

Table 1 Summary of whole-exome sequencing

	Total read (base)	Coverage	No. variations	No. regions with read depths ≧ 10 (%)	No. variations in the regions with coverages ≧ 10×
Father	140 606 140	235.5×	10 502	92.6	9798
Mother	116 931 610	196.2×	10 541	94.3	10 020
Patient	85 656 114	143.0×	10 667	95.9	10 171

Table 2 Filtering of variations

Filter	No. variations
All variations	10 667
Variations (coverage ≧ 10×)	10 171
Heterozygous in the patient and homozygous in the parents	75
Reliable homozygous reference in the parents	12
Reliable heterozygous variations in the patient	5
De novo mutations (validated by Sanger sequencing)	3

(c.2678C>T p.S893L) were present in the patient, and his affected his son and daughter, but was not present in an unaffected sibling of the patient, thus confirming the cosegregation of these two *de novo* mutations with the disease. The *PLB1* mutation (c.3601T>A p.C1201S) did not cosegregate with the disease phenotype. Haplotype analysis showed that all three *de novo* mutations occurred in the paternal chromosomes (Fig. 2).[8]

We have identified *de novo* mutations in two genes (*ATL1* and *PHF2*) that cosegregated with the disease phenotype in this family. *ATL1* has been established as the causative gene of spastic paraplegia type 3A (SPG3A). Because the clinical presentations of the patients in this family are typical for SPG3A, the novel mutation (p.Q420P) is likely the cause of the early-onset spastic paraplegia in this family. *PHF2* encodes plant homeodomain finger protein 2, which belongs to a diverse group of transcriptional regulators affecting eukaryotic gene expression by affecting the chromatin structure.[11] So far, no human diseases have been found to be associated with mutations in *PHF2*, and the *de novo* mutation (p.S893L) is registered in dbSNP build 137 as rs149241220. In addition, both PolyPhen2 and SIFT predicted the *ATL1* mutation as probably damaging, and the *PHF2* mutation as possibly damaging and tolerated. Taken together, the present findings strongly support the interpretation that the *de novo* novel mutation in *ATL1* caused the disease in this family, and that the affected individuals in this family carry SPG3A.

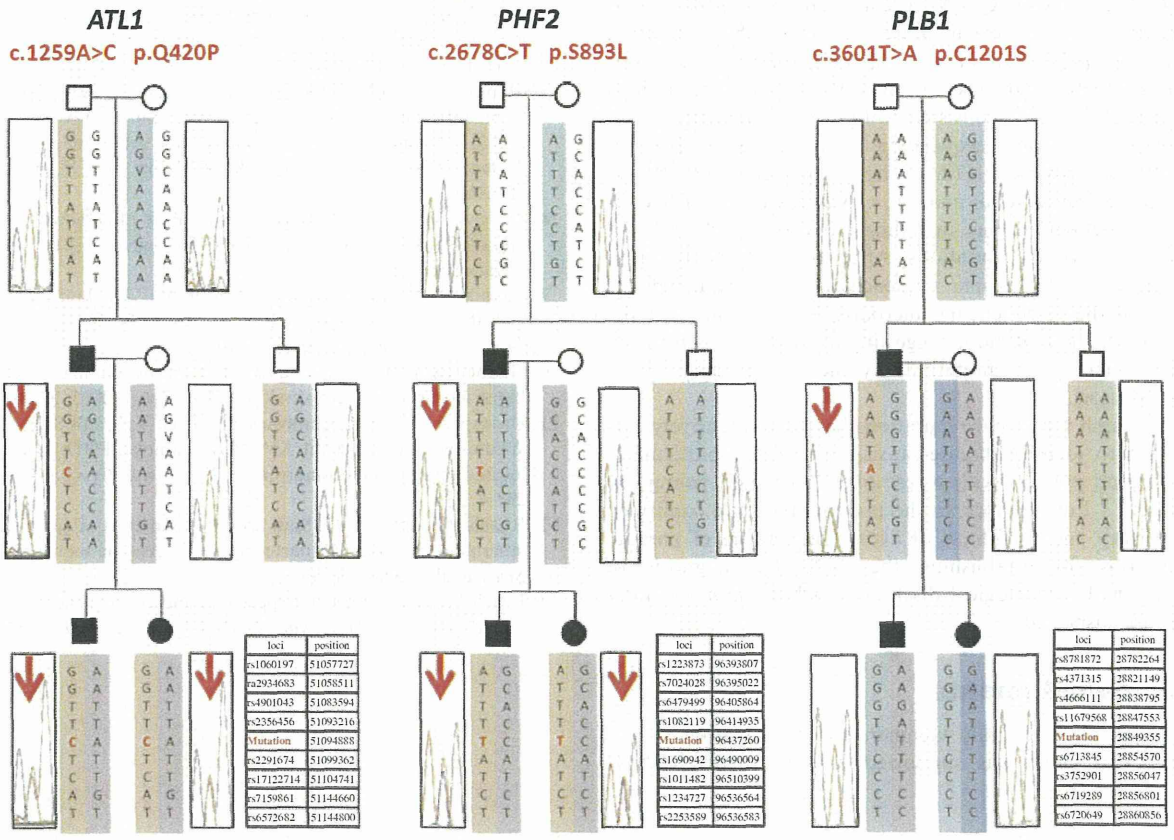

Figure 2 *De novo* mutations and haplotype analysis. Haplotypes were manually reconstructed on the basis of single nuceotide polymorphism (SNP) typing using Genome-Wide Human SNP array 6.0 (Affymetrix), and the electropherograms of the *de novo* mutations are shown along with the pedigree chart. Green, red, black, and blue curves indicate A, T, G and C, respectively. Red arrows show the positions of the mutations. The mutation and the SNP used for haplotype reconstruction are shown in the boxes along with the positions on the chromosomes. Appropriate informative SNP flanking the mutations are selected for *ATL1*, *PHF2* and *PLB1*. *ATL1* and *PHF2* mutations cosegregated with the disease phenotype, whereas the *PLB1* mutation did not.

Discussion

The present study showed the capability of massively parallel sequencing to pinpoint the *de novo* mutations through whole-exome sequencing. It has been shown that we encounter a substantial number of errors in the called variations derived from short reads obtained by massively parallel sequencing. In this study, we showed that we can efficiently narrow down to a reasonable number of candidate *de novo* variations by applying various filters including those of sufficient read depths and likelihood (Table 2).

De novo mutation rates have been estimated to be between 7.6×10^{-9} and 2.2×10^{-8}/base/generation.[12,13] Hence, the average number of *de novo* mutations in exonic regions is estimated to be one or two. Indeed, a recent study has shown that the average number of *de novo* mutations in exonic regions is 0.9.[7] In the present study, we identified three *de novo* mutations that are in good agreement with previous estimations, one of which turned out to be the cause of the disease and was found in *ATL1*.

Of note, the three *de novo* mutations occurred in the paternal chromosomes. A recent study has shown that paternal mutations occur more frequently than maternal mutations, and the number of paternal mutations increases with age.[8] Thus, our observation of three *de novo* mutations in paternal lineage is consistent with previous reports.

Before the exome sequencing analysis, we carried out mutational analysis of *ATL1* using resequencing microarrays designed for mutational analysis of various forms of hereditary spastic paraplegias including SPG3A, which did not show any causative mutations. After the p.Q420P mutation was identified by exome sequencing, we re-evaluated the results of the resequencing microarrays, and confirmed the presence of the mutation, suggesting difficulties in sensitively calling heterozygous mutations by the resequencing microarray analysis.

The search for *de novo* mutations has been intensively carried out to identify disease-causing mutations, particularly focusing on patients with severe phenotypes and young age at onset.[1–4] Furthermore, in the present study, we confirmed the efficacy and usefulness of exome sequencing of parent–child trios for establishing the molecular diagnosis of patients with neurological diseases, in whom *de novo* mutations are suspected.

Acknowledgments

We thank all the family members for participating in this study. This work was supported in part by KAKENHI (Grants-in-Aid for Scientific Research on Innovative Areas; No.22129001 and 22129002), the Global COE Program from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and a Grant-in-Aid (H23-Jitsu-yoka (Nanbyo)-Ippan-004) from the Ministry of Health, Welfare and Labor, Japan.

References

1 Hamdan FF, Saitsu H, Nishiyama K *et al.* Identification of a novel in-frame *de novo* mutation in SPTAN1 in intellectual disability and pontocerebellar atrophy. *Eur. J. Hum. Genet.* 2012; **20**: 796–800.
2 O'Roak BJ, Deriziotis P, Lee C *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nat. Genet.* 2011; **43**: 585–9.
3 Girard SL, Gauthier J, Noreau A *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat. Genet.* 2011; **43**: 860–3.
4 Xu B, Roos JL, Dexheimer P *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat. Genet.* 2011; **43**: 864–8.
5 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; **38**: 1767–71.
6 Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9.
7 Isojima T, Mitsui J, Doi K *et al.* A recurrent de novo FAM111A mutation causes Kenny-Caffey syndrome type 2. *J. Bone Miner. Res.* 2013. doi: 10.1002/bbmr.2091. [Epub ahead of print].
8 Kong A, Frigge ML, Masson G *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* 2012; **488**: 471–5.
9 Fukuda Y, Nakahara Y, Date H *et al.* SNP-HiTLink: a high-throughput linkage analysis system employing dense SNP data. *BMC Bioinformatics* 2009; **10**: 121.
10 Gudbjartsson DF, Thorvaldsson T, Kong A, Gunnarsson G, Ingolfsdottir A. Allegro version 2. *Nat. Genet.* 2005; **37**: 1015–6.
11 Hasenpusch-Theil K, Chadwick BP, Theil T, Heath SK, Wilkinson DG, Frischauf AM. PHF2, a novel PHD finger gene located on human chromosome 9q22. *Mamm. Genome* 1999; **10**: 294–8.
12 Roach JC, Glusman G, Smit AF *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010; **328**: 636–9.
13 Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl Acad. Sci. USA* 2010; **107**: 961–8.