| | | | | | |
|---|---|---|---|---|---|
| Stefanou N, Papanikolaou V, Furukawa Y, Nakamura Y, Tsezou A. | Leptin as a critical regulator of hepatocellular carcinoma development through modulation of human telomerase reverse transcriptase. | BMC Cancer | 10(1) | 442 | 2010 |
| Yamamoto K, Ishida T, Nakano K, Yamagishi M, Yamochi T, Tanaka Y, Furukawa Y, Nakamura Y, Watanabe T. | SMYD3 interacts with HTLV-1 Tax and regulates sub-cellular localization of Tax. | Cancer Science | 102 | 260-266 | 2011 |

# Ⅴ．　研究成果の刊行物・別刷

Neurogenetics
DOI 10.1007/s10048-010-0271-4

ORIGINAL ARTICLE

# Posterior column ataxia with retinitis pigmentosa in a Japanese family with a novel mutation in FLVCR1

**Hiroyuki Ishiura · Yoko Fukuda · Jun Mitsui · Yasuo Nakahara · Budrul Ahsan · Yuji Takahashi · Yaeko Ichikawa · Jun Goto · Tetsuo Sakai · Shoji Tsuji**

Received: 2 December 2010 / Accepted: 3 December 2010


**Abstract** Posterior column ataxia with retinitis pigmentosa (PCARP) is an autosomal recessive neurodegenerative disorder characterized by retinitis pigmentosa and sensory ataxia. Previous studies of PCARP in two families showed a linkage to 1q31–q32. However, detailed investigations on the clinical presentations as well as molecular genetics of PCARP have been limited. Here, we describe a Japanese consanguineous family with PCARP. Two affected siblings suffered from childhood-onset retinitis pigmentosa and slowly progressive sensory ataxia. They also showed mild mental retardation, which has not been described in patients with PCARP. Parametric linkage analysis using high-density single nucleotide polymorphism arrays supported a linkage to the same locus. Target capture and high-throughput sequencing technologies revealed a novel homozygous c.1477G>C (G493R) mutation in *FLVCR1*, which cosegregated with the disease. A recent study has identified three independent mutations in *FLVCR1* in the original and other families. Our results further confirmed that PCARP is caused by mutations in *FLVCR1*.

**Keywords** Posterior column ataxia with retinitis pigmentosa · Linkage analysis · Target capture · Massively parallel sequencing · FLVCR1

H. Ishiura · Y. Fukuda · J. Mitsui · Y. Nakahara · B. Ahsan · Y. Takahashi · Y. Ichikawa · J. Goto · S. 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

T. Sakai
Department of Neurology, Himeno Hospital,
2316 Niishiro, Hirokawa Town, Yame County,
Fukuoka 834-0115, Japan

Published online: 26 January 2011

## Introduction

Posterior column ataxia with retinitis pigmentosa (PCARP, MIM 609033) is an autosomal recessive, childhood onset neurodegenerative disorder characterized by sensory ataxia and retinitis pigmentosa. Previous studies [1, 2] on American and Spanish families revealed a linkage to chromosome 1q31–q32 defined by D1S2692 (206.10M in NCBI36/hg18 assembly, http://genome.ucsc.edu/) and D1S2141 (213.26M). Because only two families have been reported with proven linkage to 1q31–q32, detailed investigations on the clinical presentations as well as the molecular genetics of PCARP have been limited. We have recently identified a Japanese family with PCARP with supportive linkage to 1q31–q32. Employing target capture and high-throughput sequencing technologies, we herein identified a novel mutation in *FLVCR1*.

## Patients and methods

### Patients

The pedigree chart of the Japanese family with PCARP is shown in Fig. 1. Two affected siblings and an unaffected sibling were born to consanguineous parents. Written informed consent was obtained from all the participants. All the participants were clinically evaluated by a neurologist (T.S.). The study was approved by the ethical committee of The University of Tokyo.

### Linkage analysis

Genomic DNAs were extracted from peripheral blood leukocytes according to standard protocols. Five of the

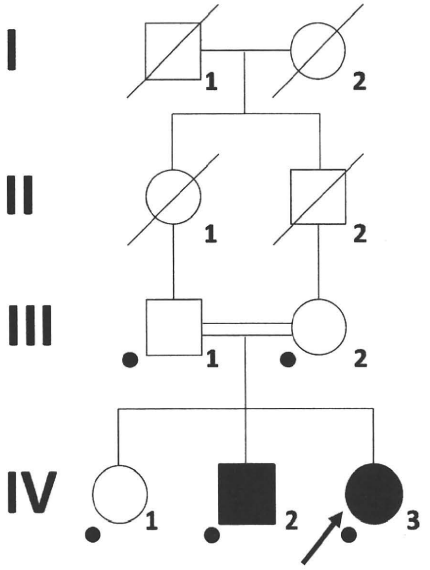


**Fig. 1** Pedigree chart. Pedigree chart of a Japanese family with PCARP. *Squares and circles* indicate males and females, respectively. Affected persons are designated with filled symbols. A *diagonal line* through a symbol represents a deceased person. A person *with the arrow* is the index patient. Persons with available genomic DNAs are indicated by *dots*

family members were genotyped using Affymetrix 50K Xba and 50K Hind arrays (Affymetrix, Santa Clara, CA) following the manufacturer's instructions. Using pipeline software SNP-HiTLink [3], single nucleotide polymorphisms (SNPs) with a $p$ value of >0.05 in the Hardy–Weinberg test, a call rate of >0.95, a confidence score of genotyping <0.1, a minor allele frequency in the controls >0, and intermarker distances of 80 to 120 kb were selected for the linkage analysis. Parametric multipoint linkage analysis (autosomal recessive model with complete penetrance) was performed with Allegro version 2 [4]. Haplotypes were reconstructed using Allegro.

### Target capture

Using NimbleGen's custom human sequence capture 2.1M array (Roche NimbleGen, Madison, WI), we designed probes corresponding to the target regions (chromosome 1: 200,106,833–213,208,193 and chromosome 20: 15,311,130–32,500,997) avoiding repetitive sequences in the regions. Twenty micrograms of genomic DNA of an affected person (IV-2) was captured according to the manufacturer's instructions [5], followed by quantification of average fold enrichment of the captured sample.

### Massively parallel sequencing

Since the target capture procedure was optimized for 454 Sequencer (454 Life Sciences, Branford, CT), the enrichment sample was nebulized for 16 min for further fragmentation to obtain appropriate lengths of DNA fragments suitable for sequencing using Genome Analyzer IIx (GAIIx, Illumina, San Diego, CA). We then carried out single-end library preparation for GAIIx. Massively parallel sequencing was accomplished using two lanes of GAIIx (100-bp-long single-end read).

### Short read alignment and variant calling

After removing the tag sequences designed for 454 sequencing system, short reads were aligned to the reference genome (NCBI36/hg18 assembly) with bwa [6] using default parameters. After removing multiple aligned reads (mapping quality of 0), single nucleotide variants (SNVs) and short insertion/deletion variants (indels) were called with SAMtools [7]. Quality threshold for SNVs and indels were set to 20 and 50, respectively.

### Annotation and confirmation of variant calls

After annotation with RefSeq (http://www.ncbi.nlm.nih.gov/projects/RefSeq/) and dbSNP130/dbSNP131 (http://www.ncbi.nlm.nih.gov/projects/SNP/), all the novel nonsynonymous variant calls were subjected to direct nucleotide sequence analysis for confirmation. Confirmed amino acid changes were then subjected to PolyPhen (http://genetics.bwh.harvard.edu/pph/) for prediction of functional effects.

### Direct nucleotide sequence analysis for confirmation of mutation in *FLVCR1*

Polymerase chain reaction was performed using a primer pair of FLVCR1-F 5′-GCAATTCGCCTACCTCAACT-3′ and FLVCR1-R 5′-ACACAAGTCCTTTTGCCAGG-3′ and LATaq (TaKaRa, Ohtsu, Shiga, Japan). Direct nucleotide sequence analysis was performed using ExoSAP-IT (USB, Cleveland, OH), a BigDye Terminator v3.1 kit, and XTerminator employing an ABI PRISM3100 sequencer (Life Technologies Corporation, Carlsbad, CA).

## Results

### Clinical manifestations of the family

The index patient (IV-3 in Fig. 1) was a 31-year-old female, who was noted to be night-blind at the age of five by her mother. Thereafter, she developed gait disturbance. She consulted with an ophthalmologist at the age of 31. Ophthalmologic examinations revealed retinitis pigmentosa of the bone corpuscle type with optic atrophy. On neurological examination, she was found to be mildly

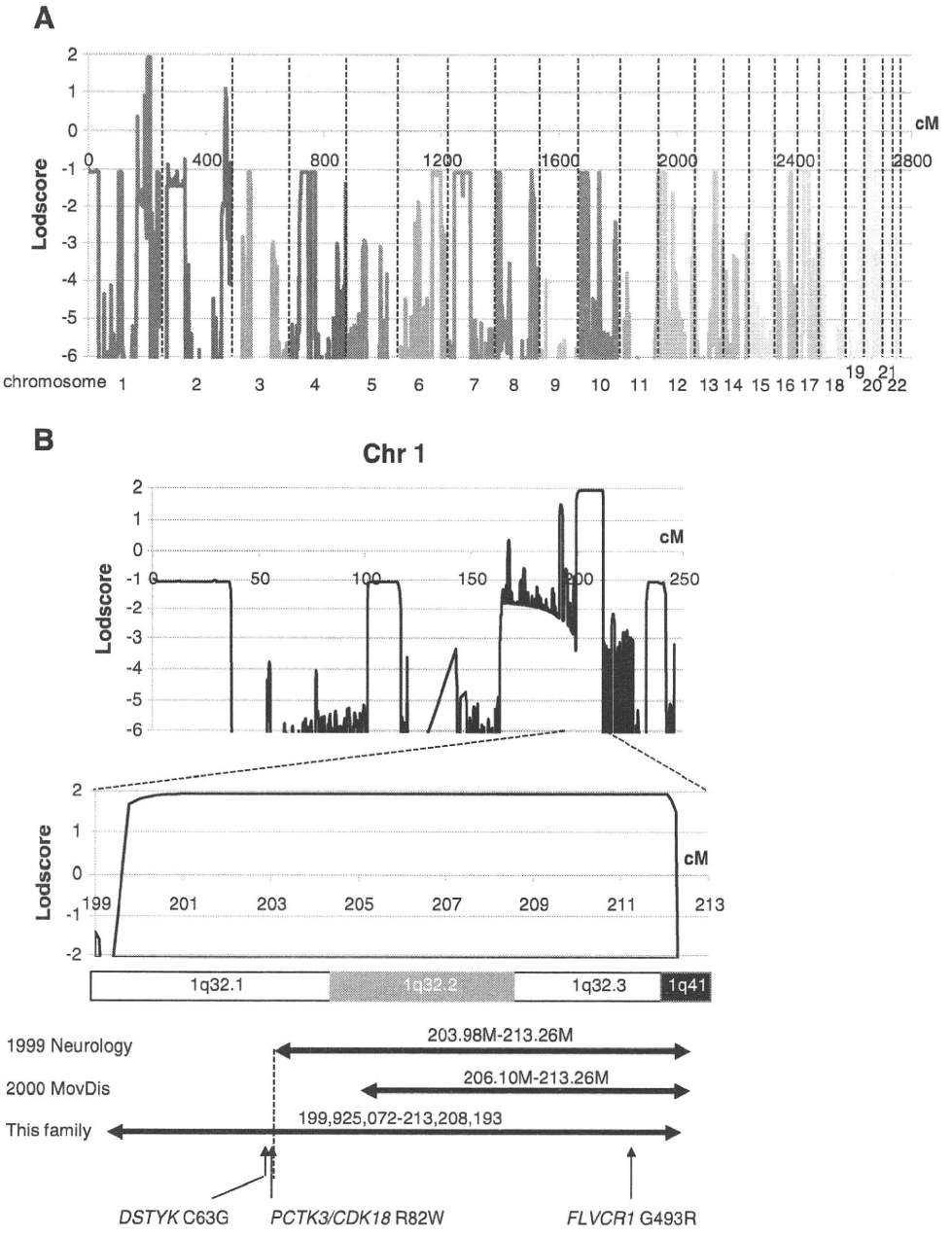


**Fig. 2** Multipoint linkage analysis and candidate regions. **a** Parametric multipoint linkage analysis (autosomal recessive model) of the family revealed linked regions on chromosomes 1 and 20. Multipoint LOD scores spanning all the chromosomes are shown. The *horizontal axis* is the cumulative genetic distance (centimorgan) starting at the short arm of chromosome 1. The vertical axis represents LOD scores. Regions on chromosomes 1 and 20 give the highest multipoint LOD scores of 1.93. **b** Parametric multipoint linkage analysis of chromosome 1. Regions with a multipoint LOD score of 1.93 are enlarged below. The *horizontal axis* is the genetic distance (centimorgan) starting at the short arm of chromosome 1. The *vertical axis* shows multipoint LOD scores. *Below the graphs*, the candidate regions demonstrated by this study as well as by previous studies [1, 2] are shown along with the diagram of chromosome 1q32.1–q41. Novel nonsynonymous variants detected in this study are also shown. *FLVCR1* G493R is the only variant that is located inside the minimum candidate region

retarded. Muscle tone was decreased in the limbs with normal strength. Coordination was preserved in the arms and legs, but with moderately ataxic gait and truncal titubation. Romberg's sign was positive. Deep tendon reflexes were decreased in the arms and absent in the legs with flexor plantar responses. Superficial sensations were intact, whereas vibratory and position senses were lost in the toes. Normal values were found in the following tests: complete blood count, blood vitamin E level, and plasma phytanic acid level. Her peripheral blood smears showed no acanthocytes. Axial T2-weighted images of the cervical spinal cord on magnetic resonance imaging demonstrated a hyperintense signal in the posterior half of the cord. Her brother (IV-2 in Fig. 1) was examined early in his thirties and was found to have mental retardation, retinitis pigmentosa, and posterior column ataxia. The other family members were neurologically normal.

**Table 1** Variants in target regions of chromosomes 1 and 20

| | No. of variants | No. of variants in exon/SS | No. of novel variants in exon/SS | No. of novel nonsynonymous variants in exon/SS |
|---|---|---|---|---|
| chr1 | 13,616 | 60 | 5 | 4 |
| chr20 | 10,545 | 30 | 1 | 1 |

*SS* splice site (splice donor and acceptor sites including two adjacent nucleotides in introns)

**Table 2** Novel nonsynonymous variants detected in target regions

*Chr* chromosome, *homo* homozygous

| Chr | Physical position | Variant | Gene | Amino acid change | Polyphen |
|---|---|---|---|---|---|
| 1 | 203447100 | A>C (homo) | *DSTYK* | C63G | Probably damaging |
| 1 | 203759347 | C>T (homo) | *PCTK3/CDK18* | R82W | Possibly damaging |
| 1 | 211129174 | G>C (homo) | *FLVCR1* | G493R | Possibly damaging |

## Linkage analysis

Multipoint parametric linkage analysis revealed the highest LOD scores of 1.93 spanning regions on chromosome 1 (defined by rs950114 and rs10494988) and chromosome 20 (defined by rs2876404 and rs6082269, Fig. 2a). The region on the chromosome 1 overlapped the previously defined locus of PCARP (Fig. 2b).

## Massively parallel sequencing analysis

Average fold enrichment for QC loci of the captured library was 129. From two lanes of GAIIx, we obtained 37,165,950 reads. Of these, 15,865,704 reads (42.7%) had tag sequences for 454 in the first 20 bases. In these reads, tag sequences were eliminated and we used them as 80 bp sequences. Aligned uniquely to the reference genome were 32,332,900 reads (87.0%), and 29,693,695 reads (79.9%) were aligned to the target region. The average coverage of target regions was 89.6X.

In the 30.3 Mb of target region on chromosomes 1 and 20, 24161 variants were called. Of these, 90 were located in coding regions and splice sites in the target regions, six of which were not registered in dbSNP131 (http://www.ncbi.nlm.nih.gov/projects/SNP/), and five of which were concluded to be novel nonsynonymous SNV (Tables 1 and 2). Two of the five novel variant calls were heterozygous, and direct nucleotide sequence analysis revealed that they were false positives.

Considering previous linkage studies [1, 2], two of the three novel nonsynonymous SNVs were located outside the overlapping candidate region (Fig. 2b and Table 1). Thus, the only novel nonsynonymous variant within the minimum candidate region was a homozygous c.1477G>C (G493R) of *FLVCR1* (Fig. 3a). The mutation was further confirmed by direct nucleotide sequence analysis (Fig. 3b). The two

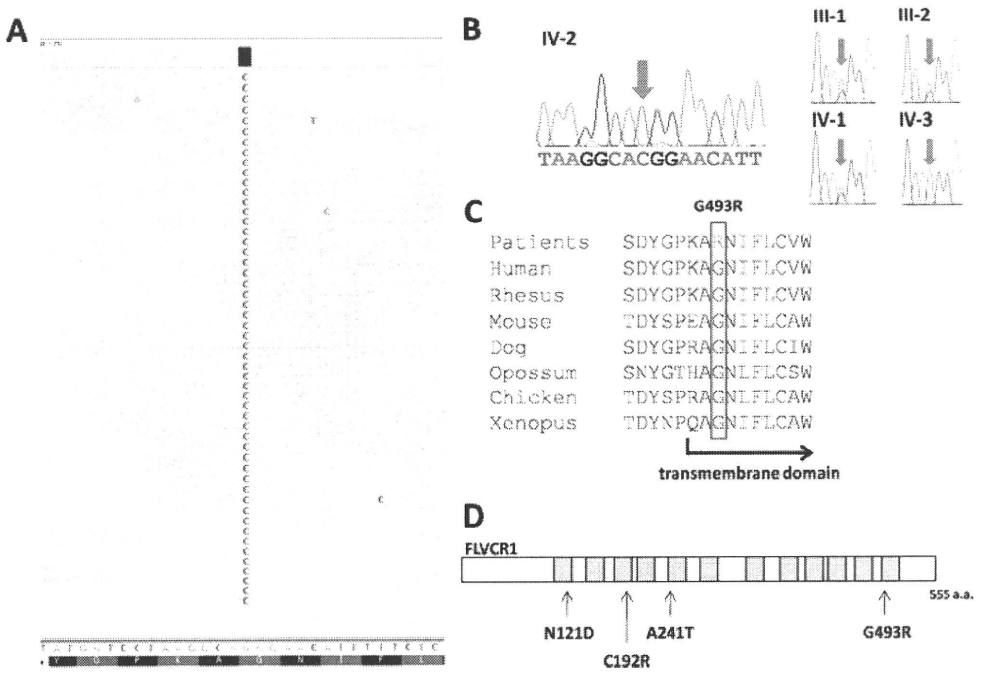


**Fig. 3** Identification of causative mutation in *FLVCR1*. **a** Aligned short reads showing homozygous *FLVCR1* c.1477G>C mutation. Aligned reads are viewed using Integrative Genomic Viewer (http://www.broadinstitute.org/igv/). Each short read is represented as a horizontal bar. Only mismatched bases are explicitly shown. All the 52 reads aligned in the position show the C allele, suggesting a homozygous mutation. **b** Direct nucleotide sequence analysis confirms the mutation, which cosegregates with the disease. **c** Partial FLVCR1 amino acid sequence alignment reveals that G493 is evolutionally conserved among species. A putative transmembrane domain is also indicated by an arrow. **d** Schematic representation of FLVCR1 protein. Mutations detected to date are shown. Putative transmembrane domains are *shaded*

affected individuals carried the homozygous mutation, whereas the parents and the unaffected sibling carried the heterozygous mutation. Because R493 is evolutionally well conserved (Fig. 3c) and the amino acid change was not observed in 192 control chromosomes, we concluded it as a pathogenic mutation of PCARP.

## Discussion

We described two cases of a Japanese family with PCARP. Linkage analysis supported the linkage to the previously defined locus, and we identified a novel mutation in *FLVCR1* employing targeted capture and massively parallel sequencing as the cause of PCARP. Very recently, Rajadhyaksha et al. have conducted massively parallel sequencing and found independent mutations in *FLVCR1* (N121D, A241T, and C192R) in three families [8]. Our report further confirmed that PCARP is caused by mutations in *FLVCR1*.

FLVCR1 is a 555 amino acid protein that has 12 transmembrane domains. Intriguingly, three previously reported mutations are located in the first, third, and fifth putative transmembrane domains of FLVCR1. The mutation which we found is also located in the 12th transmembrane domain (Fig. 3d). Moreover, all the mutations in FLVCR1 found in PCARP are substitution of a hydrophilic amino acid for a hydrophobic amino acid (A214T) or substitutions of charged amino acids for uncharged amino acids (N121D, C192R, and G493R). These finding suggest a possibility that disruption of transmembrane domains of FLVCR1 is involved in the pathogenesis of PCARP.

Although childhood-onset retinitis pigmentosa and sensory ataxia found in the affected siblings were characteristics of PCARP, they also had mild mental retardation. Because no cognitive deficits have been reported in the original PCARP families [2], careful interpretation would be necessary. One possibility is that the clinical presentations can be more heterogeneous depending on mutations and G493R mutation in *FLVCR1* is associated with mental retardation. Another possibility is that other gene(s) are responsible for mental retardation. Because there are at least two other novel homozygous amino acid changes in the candidate regions as determined on the basis of the linkage analysis of this family under an autosomal recessive model (Table 2), some of these substitutions may contribute to mental retardation.

Previous studies suggested that FLVCR1 is a heme transporter, and *FLVCR1* null mice present a phenotype with a lack of erythropoiesis and craniofacial and limb deformities resembling Diamond–Blackfan anemia [9]. Because neither changes in the shape of erythrocytes nor anemia was observed in the index patient, the discrepancy between human disease and mouse model should be further investigated in the future.

In conclusion, we identified a novel mutation in *FLVCR1* in a Japanese PCARP family. The study showed that target capture and massively parallel sequencing technologies enable us to identify causative genes even in a small family and they are expected to further unveil molecular pathogeneses of neurodegenerative disorders.

**Acknowledgments** We thank the family members for their participation. We also thank Dr. S. Ogawa (Cancer Genome Project, The University of Tokyo) for his kind help to sequencing. This work was supported in part by KAKENHI (Grant-in-Aid for Scientific Research) on Scientific Research on Innovative Areas (Exploring Molecular Basis for Brain Diseases Based on Personal Genomics), Priority Areas (Applied Genomics), Integrated Database Project, and Scientific Research (A) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. H. I. is supported by a Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists.

## References

1. Higgins JJ, Morton DH, Loveless JM (1999) Posterior column ataxia with retinitis pigmentosa (AXPC1) maps to chromosome 1q31–q32. Neurology 52:146–150
2. Higgins JJ, Kluetzman K, Berciano J, Combarros O, Loveless JM (2000) Posterior column ataxia and retinitis pigmentosa: a distinct clinical and genetic disorder. Mov Disord 15:575–578
3. Fukuda Y, Nakahara Y, Date H, Takahashi Y, Goto J, Miyashita A, Kuwano R, Adachi H, Nakamura E, Tsuji S (2009) SNP-HiTLink: a high-throughput linkage analysis system employing dense SNP data. BMC Bioinformatics 10:121
4. Gudbjartsson DF, Thorvaldsson T, Kong A, Gunnarsson G, Ingolfsdottir A (2005) Allegro version 2. Nat Genet 37:1015–1016
5. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR (2007) Genome-wide in situ exon capture for selective resequencing. Nat Genet 39:1522–1527
6. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760
7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (1000) Genomes Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079
8. Rajadhyaksha AM, Elemento O, Puffenberger EG, Schierberl KC, Xiang JZ, Putorti ML, Berciano J, Poulin C, Brais B, Michaelides M, Weleber RG, Higgins JJ (2010) Mutations in FLVCR1 cause posterior column ataxia and retinitis pigmentosa. Am J Hum Genet 87:643–654
9. Keel SB, Doty RT, Yang Z, Quigley JG, Chen J, Knoblaugh S, Kingsley PD, De Domenico I, Vaughn MB, Kaplan J, Palis J, Abkowitz JL (2008) A heme export protein is required for red blood cell differentiation and iron homeostasis. Science 319:825–828

Neurogenetics
DOI 10.1007/s10048-010-0253-6

ORIGINAL ARTICLE

# Identification of novel SNPs of *ABCD1*, *ABCD2*, *ABCD3*, and *ABCD4* genes in patients with X-linked adrenoleukodystrophy (ALD) based on comprehensive resequencing and association studies with ALD phenotypes

**Takashi Matsukawa · Muriel Asheuer · Yuji Takahashi · Jun Goto · Yasuyuki Suzuki · Nobuyuki Shimozawa · Hiroki Takano · Osamu Onodera · Masatoyo Nishizawa · Patrick Aubourg · Shoji Tsuji**

Received: 15 May 2010 / Accepted: 9 July 2010


**Abstract** Adrenoleukodystrophy (ALD) is an X-linked disorder affecting primarily the white matter of the central nervous system occasionally accompanied by adrenal insufficiency. Despite the discovery of the causative gene, *ABCD1*, no clear genotype–phenotype correlations have been established. Association studies based on single nucleotide polymorphisms (SNPs) identified by comprehensive resequencing of genes related to *ABCD1* may reveal genes modifying ALD phenotypes. We analyzed 40 Japanese patients with ALD. *ABCD1* and *ABCD2* were analyzed using a newly developed microarray-based resequencing system. *ABCD3* and *ABCD4* were analyzed by direct nucleotide sequence analysis. Replication studies were conducted on an independent French ALD cohort with extreme phenotypes. All the mutations of *ABCD1* were identified, and there was no correlation between the genotypes and phenotypes of ALD. SNPs identified by the comprehensive resequencing of *ABCD2*, *ABCD3*, and *ABCD4* were used for association studies. There were no significant associations between these SNPs and ALD phenotypes, except for the five SNPs of *ABCD4*, which are in complete disequilibrium in the Japanese population. These five SNPs were significantly less frequently represented in patients with adrenomyeloneuropathy (AMN) than in controls in the Japanese population ($p$= 0.0468), whereas there were no significant differences in patients with childhood cerebral ALD (CCALD). The replication study employing these five SNPs on an independent French ALD cohort, however, showed no significant associations with CCALD or pure AMN. This study showed that *ABCD2*, *ABCD3*, and *ABCD4* are less likely the disease-modifying genes, necessitating further studies to identify genes modifying ALD phenotypes.

**Keywords** Adrenoleukodystrophy · *ABCD1* · *ABCD2* · *ABCD3* · *ABCD4* · DNA microarray

**Electronic supplementary material** The online version of this article (doi:10.1007/s10048-010-0253-6) contains supplementary material, which is available to authorized users.

T. Matsukawa · Y. Takahashi · J. Goto · S. Tsuji (✉)
Department of Neurology, Graduate School of Medicine,
The University of Tokyo,
7-3-1 Hongo, Bunkyo,
Tokyo 113-8655, Japan
e-mail: tsuji@m.u-tokyo.ac.jp

M. Asheuer · P. Aubourg
UMR Inserm 745, University Paris-Descartes,
Hôpital Saint-Vincent de Paul,
Paris, France

Y. Suzuki · N. Shimozawa
Gifu University,
Gifu, Japan

H. Takano · O. Onodera · M. Nishizawa
Niigata University,
Niigata, Japan

Published online: 27 July 2010

## Introduction

Adrenoleukodystrophy (ALD) is a demyelinating disease caused by mutations of *ABCD1* [1]. This disease affects primarily the white matter of the central nervous system occasionally accompanied by adrenal insufficiency [2–4]. Diagnosis of ALD is usually made by the increased contents of very long chain saturated fatty acids (VLCFAs; >C22:0) in plasma as well as by mutational analysis of *ABCD1* [5–7].

Since 15% of obligate female carriers have normal VLCFA levels [7], mutational analysis is essential for the diagnosis of the carriers. Since the first report of allogenic HSCT for childhood ALD, there has been an increasing number of reports showing efficacies of HSCT for the childhood cerebral form of ALD, if HSCT is performed at early stages of the disease [8–10]. Thus, availability of rapid molecular diagnosis for patients with ALD and carriers is mandatory in the clinical practice for ALD.

ALD is characterized by a broad spectrum of clinical presentations including childhood cerebral form, adrenomyeloneuropathy (AMN), AMN complicated by cerebral demyelination, adulthood cerebral form, and Addison disease. From clinical experience, patients with different clinical phenotypes can be observed even in a single pedigree. In support of this, no clear genotype–phenotype correlations have been observed [11–16], raising the possibility that other genetic or environmental factors are involved in the pleiomorphic clinical presentations of ALD.

*ABCD1* gene encodes a half-ATP-binding cassette (ABC) transporter, adrenoleukodystrophy protein (ALDP), which is localized to the peroxisomal membrane. *ABCD2*, *ABCD3*, and *ABCD4* genes are the closest homologues of the *ABCD1* gene [17, 18]. It has been shown that the majority of mouse liver ALDP and the 70-kDa peroxisomal membrane protein (PMP70) that is encoded by *ABCD3* are homomeric proteins [19]. Furthermore, it has been shown that ALDP can form homodimers or a heterodimer with the adrenoleukodystrophy-related protein (ALDR) that is encoded by *ABCD2* or the PMP70 that is encoded by *ABCD3* [16, 20–22], raising the possibility that these ABCD1-related genes function as disease-modifying genes for ALD.

To provide a rapid mutational analysis for ALD, we developed a microarray-based high-throughput resequencing system of *ABCD1* (TKYPD01) [23]. Furthermore, to explore the possibility that these *ABCD1*-related genes function as disease-modifying genes, we established a comprehensive resequencing system for *ABCD1*-related genes, *ABCD2*, *ABCD3*, and *ABCD4*. On the basis of the comprehensive resequencing of *ABCD1*, *ABCD2*, *ABCD3*, and *ABCD4* genes, we identified 11 novel single nucleotide polymorphism (SNPs). Using these novel SNPs as well as previously described SNPs of these genes, we conducted detailed association studies of these SNPs with the clinical phenotypes of ALD.

## Materials and methods

### Participants

Forty Japanese ALD patients, consisting of 14 patients with childhood cerebral ALD (CCALD), 8 patients with adulthood cerebral ALD (AdultCer), 2 patients with AMN with later development of cerebral ALD (AMN-Cer), 13 patients with AMN, 1 asymptomatic patient, and 2 patients with unknown form, were enrolled in this study. Among the patients, mutations were previously identified in 16 ALD patients by direct nucleotide sequence analysis, while no mutational analyses were conducted for 24 patients.

For replication studies of the results of association studies on Japanese ALD patients showing potential associations of SNPs in *ABCD1*-related genes with ALD phenotypes, an independent French ALD cohort with well-defined extreme phenotypes consisting of 118 patients with CCALD and 71 patients with pure AMN (AMN with age >45 years as well as with normal brain magnetic resonance imaging) was studied. In addition, 51 ALD patients with AMN-Cer were also analyzed in the French ALD cohort.

### Procedures

Primers specific for *ABCD1*, *ABCD2*, *ABCD3*, and *ABCD4* were designed using BLAST search and Smith–Waterman method to avoid amplification of the related homologous genes (Fig. 1; ESM Tables 1, 2, 3, and 4). In particular, since there were many segments homologous to exons 8, 9, and 10 of *ABCD1*, a specific primer pair was designed (Fig. 1). Fifty nanograms of genomic DNA were subjected to polymerase chain reaction (PCR) amplification in a total volume of 50 μL. The PCR conditions were as follows: 94°C for 1 min, followed by five cycles consisting of 94°C for 30 s, 62°C for 30 s, and 68°C for 2 min; five cycles consisting of 94°C for 30 s, 60°C for 30 s, and 68°C for 2 min; and 25 cycles consisting of 94°C for 30 s, 58°C for 30 s, and 68°C for 2 min, followed by a final extension at 68°C for 7 min, using the LA Taq with GC Buffer PCR system (Takara Bio, Otsu, Shiga, Japan).

Resequencing DNA microarrays were used in the analyses of the sequences of *ABCD1* (TKYPD01) and *ABCD2* (TKYAD01). TKYPD01 and TKYAD01 were designed using the platform of GeneChip CustomSeqTM Resequencing Microarray (Affymetrix, Santa Clara, CA, USA). Since there are substantial homologies between *ABCD1* and *ABCD2*, these genes were placed in independent microarrays (TKYPD01 and TKYAD01). Each PCR product of *ABCD1* and *ABCD2* was quantified using PicoGreen (Molecular Probes, Eugene, OR, USA) and equimolarly pooled. Pooled PCR products of *ABCD1* and *ABCD2* were fragmented using DNAse I, labeled with biotin, hybridized to DNA microarrays, and subjected to scan and analyses of nucleotide sequences of *ABCD1* (TKYPD01) and *ABCD2* (TKYAD01) according to the manufacturer's instructions (Affymetrix, Santa Clara, CA, USA). The base calls that were undetermined using the GDAS analysis software (Affymetrix, Santa Clara, CA, USA) were further analyzed

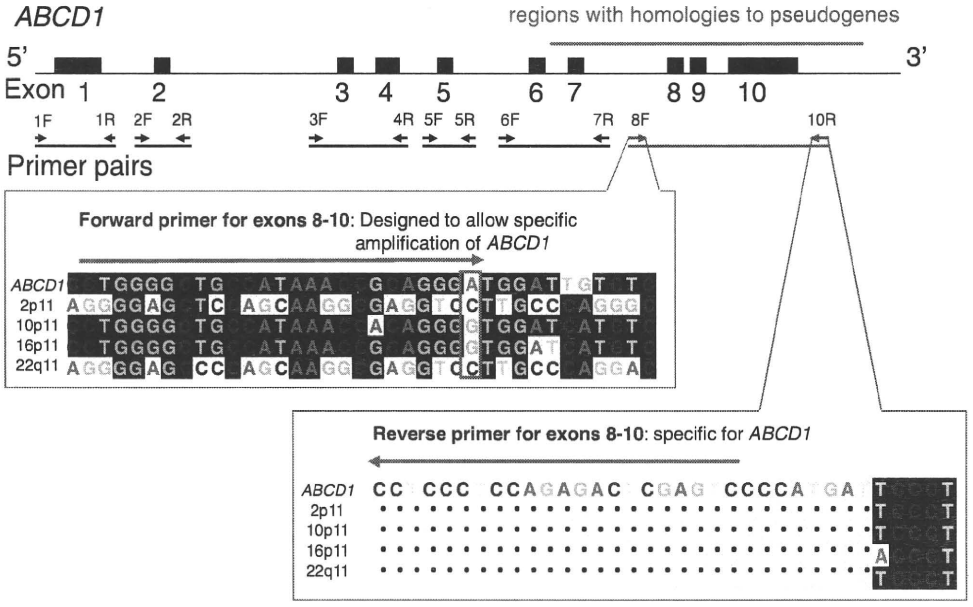


**Fig. 1** Primer design for *ABCD1*. All the exons of *ABCD1* were amplified using six primer pairs. There were pseudogenes at 2p11, 10p11, 16p11, and 22q11, which were similar in sequence to exons 7–10 of *ABCD1* gene (92–96%). The forward primer for exons 8–10 was designed to avoid amplification of the related homologous genes. We could design a specific reverse primer for exons 8–10

by manual inspection. Identified mutations and SNPs were confirmed by the direct nucleotide sequence analysis of the PCR products. All the PCR products of *ABCD3* and *ABCD4* were analyzed by the direct nucleotide sequence analysis. Identified SNPs of *ABCD2*, *ABCD3*, and *ABCD4* were examined as to whether they were novel SNPs or known SNPs using the J SNP (http://snp.ims.u-tokyo.ac.jp/index_ja.html) and DB SNP (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Snp&cmd=Limits).

### Statistical analyses

We compared the allele frequencies of detected SNPs between the subgroups of ALD or between the individual subgroup and the controls by Fisher's exact test using the JMP 7 software (SAS Institute, Cary, NC, USA). Deviation of the SNP genotypes from the Hardy–Weinberg equilibrium was evaluated using the PEDSTATS program [24]. Linkage disequilibria among the neighboring SNPs were evaluated using Haploview version 4.1 [25].

## Results

### Resequencing DNA microarray-based mutational analysis of *ABCD1* gene in Japanese ALD patients

All the mutations of *ABCD1* were clearly identified using the resequencing DNA microarray system including 26 missense, 2 nonsense, and 12 insertion/deletion mutations

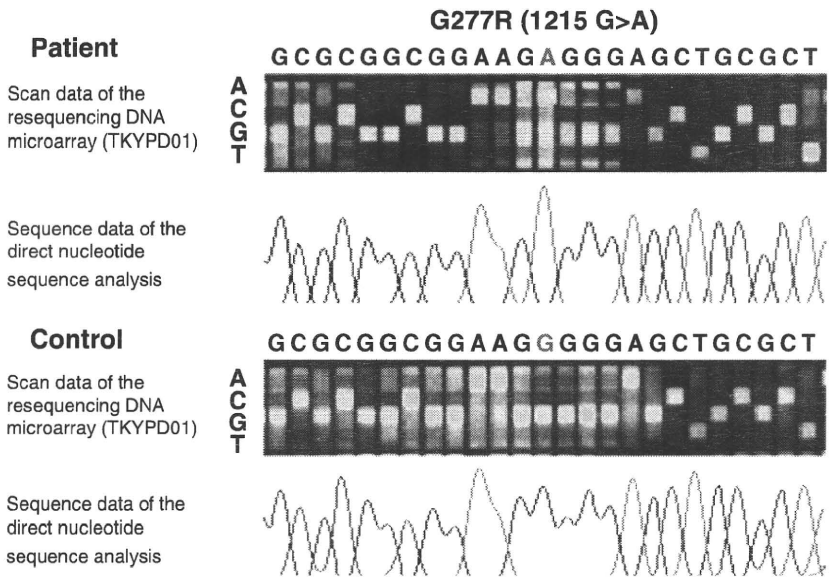


**Fig. 2** Scan data of the resequencing DNA microarray and sequence data of the direct nucleotide sequence analysis (*upper panel*: patient, *lower panel*: control). Each *column* shows a base position, and each *row* shows a base call DNA in the scan data of the resequencing DNA microarray. Here, a mutation (G277R) was detected, and the signal intensities around the mutation were reduced because of the mismatch of the mutation site. The sequence data of the direct nucleotide sequence analysis confirmed the scan data of the resequencing DNA microarray

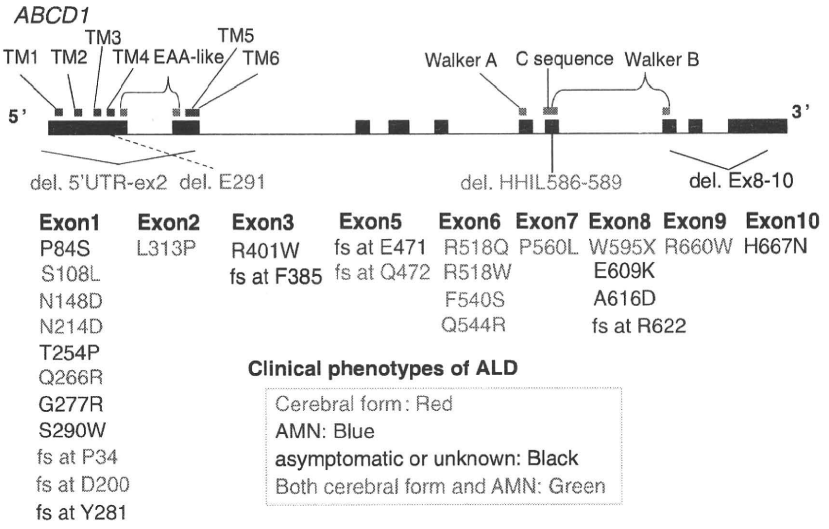


**Fig. 3** Identified mutations of *ABCD1*. Mutations of *ABCD1* gene were widely scattered in the entire region of *ABCD1* gene. All types of *ABCD1* mutations were distributed among all the phenotypes of adrenoleukodystrophy. *TM* transmembrane domain, *EAA-like* EAA-like protein motif, *Walker A* Walker A motif, *C sequence* nucleotide binding fold conserved sequence, *Walker B* Walker B motif, *fs* frameshift

of *ABCD1* (Figs. 2 and 3; Tables 1 and 2). Mutations of *ABCD1* gene were widely scattered in the entire region of *ABCD1* gene (Fig. 3; Tables 1 and 2). All types of *ABCD1* mutation were distributed among all the phenotypes of ALD (Fig. 3; Tables 1 and 2). Among the 40 mutations, 11 mutations were novel (Tables 1 and 2). Among the deletion/frameshift mutations that are expected to result in complete loss of ALDP functions, the mutations were distributed among all the phenotypes of ALD (Tables 1 and 2), supporting the previous observations of no genotype–phenotype correlations.

## Identification of SNPs of *ABCD2*, *ABCD3*, and *ABCD4* genes by comprehensive resequencing

Comprehensive resequencing of *ABCD2*, *ABCD3*, and *ABCD4* genes of the 40 Japanese patients with ALD revealed two novel SNPs, nine SNPs (six known and three novel SNPs), and 13 SNPs (seven known and six novel SNPs), respectively (Fig. 4; Tables 3, 4, 5, and 6; ESM Table 5). Hardy–Weinberg equilibrium was fulfilled for each SNP. The five known SNPs (rs17782508, rs2301345, rs4148077, rs4148078, and rs3742801) were in complete

**Table 1** Identified *ABCD1* mutations: mutations of *ABCD1* that result in devastating effects (frame shifts or nonsense mutations) on adrenoleukodystrophy protein (ALDP)

| Patient number | Phenotype | Mutation of *ABCD1* | Effect of mutation of *ABCD1* |
|---|---|---|---|
| 1 | CCALD | 488C>AT | Frameshift at P34 |
| 2 | CCALD | 2171G>A | W595X |
| 3 | CCALD | 5′UTR-Ex2 1.4-kb deletion[a] | Disruption of gene structure |
| 4 | AdultCer | Del. 986C[a] | Frameshift at D200 |
| 5 | AdultCer | Del. 1801–1802AG[a] | Frameshift at Q472 |
| 6 | AMN-Cer | Del. 2251 GGTG ins. TGTTCT[a] | Frameshift at R622 |
| 7 | AMN | Ins. 1237T[a] | Frameshift at Y281 |
| 8 | AMN | Del. 1801–1802AG[a] | Frameshift at Q472 |
| 9 | AMN | 2171G>A | W595X |
| 10 | AMN | Del. 2251 GGTG ins. TGTTCT[a] | Frameshift at R622 |
| 11 | Unknown | Del. 1541C[a] | Frameshift at F385 |
| 12 | Unknown | Ex8-10 0.3-kb deletion[a] | Disruption of gene structure |

Amino acid residue numbers in ALDP are based on Mosser et al. [1]. The domains and motifs in the ALDP are based on Mosser et al. [1]

*CCALD* childhood cerebral ALD, *AdultCer* adult with cerebral ALD, *AMN-Cer* AMN with cerebral ALD, *AMN* adrenomyeloneuropathy, *TM* transmembrane domain, *Loop 1* loop 1 motif, *EAA-like* EAA-like protein motif, *Walker A* Walker A motif, *Cons* nucleotide binding fold conserved sequence, *Walker B* Walker B motif

[a] Novel mutation

**Table 2** Identified *ABCD1* mutations: mutations of *ABCD1* that result in amino acid substitutions or in-frame deletions

| Patient number | Phenotype | Mutation of *ABCD1* | Effect of mutation of *ABCD1* | Position of mutation |
|---|---|---|---|---|
| 13 | CCALD | 709C>T | S108L | Loop1 |
| 14 | CCALD | 709C>T | S108L | Loop1 |
| 15 | CCALD | 829A>G | N148S | TM2 |
| 16 | CCALD | 1026A>G | N214D | TM3 |
| 17 | CCALD | 1182G>A | G266R | Between TM4 and EAA-like |
| 18 | CCALD | 1324T>C[a] | L313P | Between EAA-like and TM5 |
| 19 | CCALD | 1938C>T | R518W | Walker A |
| 20 | CCALD | 1939G>A | R518Q | Walker A |
| 21 | CCALD | 2017A>G | Q544R | Between Walker A and Cons |
| 22 | CCALD | 2017A>G | Q544R | Between Walker A and Cons |
| 23 | CCALD | 2065C>T | P560L | Between Walker A and Cons |
| 24 | CCALD | 2065C>T | P560L | Between Walker A and Cons |
| 25 | CCALD | Del. 2145–2156 | Del. HILQ587-590 | Between Walker A and Cons |
| 26 | AdultCer | Del. 1257–1259 | Del.E291 | EAA-like |
| 27 | AdultCer | 2005T>C | F540S | Between Walker A and Cons |
| 28 | AdultCer | 2358C>T | R660W | C-terminal to Walker B |
| 29 | AdultCer | 2385C>A | H667N | C-terminal to Walker B |
| 30 | AMN-Cer | 1146A>C | T254P | TM4 |
| 31 | AMN | 636C>T | P84S | TM1 |
| 32 | AMN | 709C>T | S108L | Loop1 |
| 33 | AMN | 1182G>A | G266R | Between TM4 and EAA-like |
| 34 | AMN | 1197G>A | E271K | Between TM4 and EAA-like |
| 35 | AMN | 1215G>A[a] | G277R | Between TM4 and EAA-like |
| 36 | AMN | 1255C>G | S290W | EAA-like |
| 37 | AMN | 1581C>T | R401W | Between TM6 and Walker A |
| 38 | AMN | 2233C>A | A616D | Cons |
| 39 | AMN | 2385C>A | H667N | C-terminal to Walker B |
| 40 | Asymptomatic | 2211G>A | E609K | Cons |

Amino acid residue numbers in ALDP are based on Mosser et al. [1]. The domains and motifs in the ALDP are based on Mosser et al. [1]

*CCALD* childhood cerebral ALD, *AdultCer* adult with cerebral ALD, *AMN-Cer* AMN with cerebral ALD, *AMN* adrenomyeloneuropathy, *TM* transmembrane domain, *Loop 1* loop 1 motif, *EAA-like* EAA-like protein motif, *Walker A* Walker A motif, *Cons* nucleotide binding fold conserved sequence, *Walker B* Walker B motif

[a] Novel mutation

linkage disequilibrium in the Japanese patients with ALD as well as in the controls, as determined using Haploview version 4.1 (Fig. 4).

## Association studies of SNPs of *ABCD2*, *ABCD3*, and *ABCD4* with the clinical phenotypes of ALD

Using the 11 novel SNPs and 13 previously described SNPs in *ABCD2*, *ABCD3*, and *ABCD4*, we conducted association studies of these SNPs with the clinical phenotypes of ALD (Tables 5 and 6).

For *ABCD2*, we analyzed two novel SNPs (novel SNP1 and novel SNP2). There were no significant differences in the allele frequencies between patients with cerebral form and those with AMN, or between the patients with individual phenotypes of ALD and the controls. For *ABCD3*, we analyzed three novel SNPs (novel SNP3, novel SNP4, and novel SNP5) and six previously described SNPs (rs4148058, rs2147794, rs16946, rs681187, rs662813, and rs337592). However, we did not also detect any significant associations.

For *ABCD4*, we analyzed six novel SNPs (novel SNP6, novel SNP7, novel SNP8, novel SNP9, novel SNP10, and novel SNP11) and seven previously described SNPs (rs17782508, rs17182959, rs17158118, rs2301345, rs4148077, rs4148078, and rs3742801). Interestingly, the five previously described SNPs (rs17782508, rs2301345, rs4148077, rs4148078, and rs3742801) that are in complete linkage disequilibrium were significantly less frequently represented in the patients with Japanese AMN than in the controls in the Japanese population ($p$=0.0468), whereas

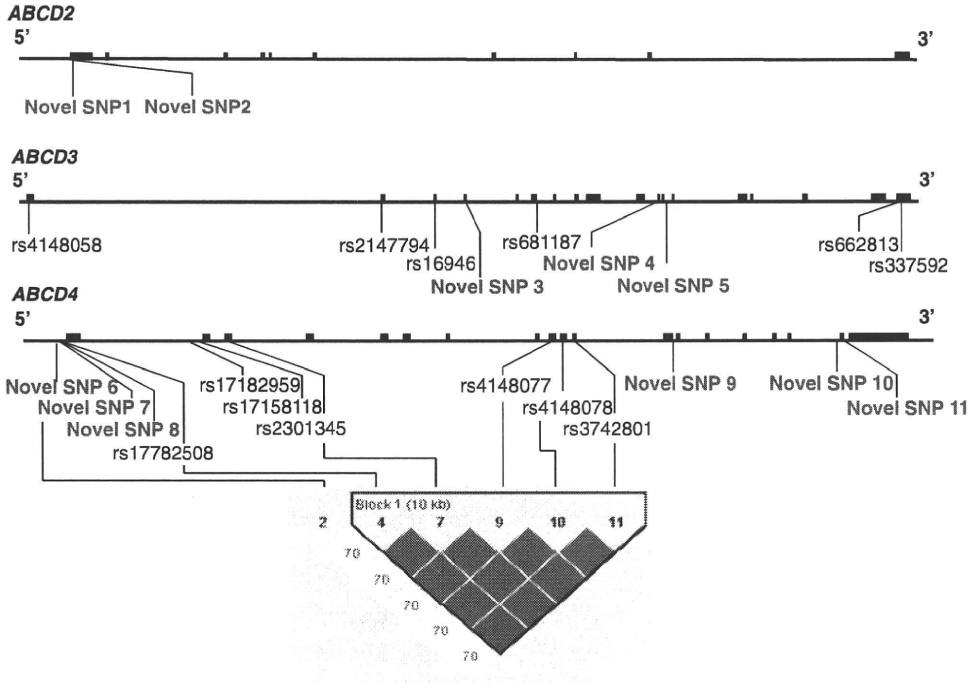


**Fig. 4** Identified single nucleotide polymorphisms (SNPs) of *ABCD2*, *ABCD3*, and *ABCD4* (*upper panel*). Comprehensive resequencing of *ABCD2*, *ABCD3*, and *ABCD4* genes of the 40 patients with adrenoleukodystrophy (ALD) revealed two novel SNPs, nine SNPs (six known and three novel SNPs), and 13 SNPs (seven known and six novel SNPs), respectively. *Red characters* indicate the novel SNPs, *blue characters* indicate the SNPs identified in the coding region, and *black characters* indicate the SNPs identified in the noncoding region. Linkage disequilibrium (LD) map of SNPs of *ABCD4* in Japanese patients with ALD and the controls using the Haploview version 4.1 (*lower panel*). The five known SNPs (rs17782508, rs2301345, rs4148077, rs4148078, and rs3742801) were in complete disequilibrium in Japanese patients with ALD and the controls (LOD=43.97, $r^2$=1.0, D′=1.0). Novel SNP7 and the five known SNPs (rs17782508, rs2301345, rs4148077, rs4148078, and rs3742801) were not in strong disequilibrium in Japanese patients with ALD and the controls (LOD=1.15, $r^2$=0.037, D′=0.706), although novel SNP7 and the five known SNPs (rs17782508, rs2301345, rs4148077, rs4148078, and rs3742801) were strong disequilibrium only in Japanese patients with ALD (LOD=2.02, $r^2$= 0.221, D′=1.0). The *number in the box* indicates the data of D′. The *color of the box* is determined from the LOD score and D′. The block was determined using a confidence interval algorithm [33]

**Table 3** Summary of identified single nucleotide polymorphism (SNPs) of *ABCD2*, *ABCD3*, and *ABCD4* in 40 adrenoleukodystrophy patients: novel SNPs

| Gene | Name | Fragment | Position (UCSC hg18) | Base call | Category | Amino acid change |
|---|---|---|---|---|---|---|
| *ABCD2* | Novel SNP1 | Exon1 | 38299954 | A/T | 5′ untranslated region | |
| | Novel SNP2 | Exon1 | 38299659 | G/C | Coding nonsynonymous | A9G |
| *ABCD3* | Novel SNP3 | Exon4 | 94706096 | A/G | Coding nonsynonymous | M94V |
| | Novel SNP4 | Exon14 | 94727816 | T/G | Intron | |
| | Novel SNP5[a] | Exon15 | 94728352 | G/C | Intron | |
| *ABCD4* | Novel SNP6 | 5′UTR | 73840784 | T/C | Upstream at the transcription start site | |
| | Novel SNP7 | 5′UTR | 73839945 | T/C | Upstream at the transcription start site | |
| | Novel SNP8 | 5′UTR | 73839604 | G/A | Upstream at the transcription start site | |
| | Novel SNP9[b] | Exon12 | 73826720 | A/G | Intron | |
| | Novel SNP10 | Exon18 | 73823320 | A/G | Intron | |
| | Novel SNP11[a] | Exon18 | 73823116 | T/C | Intron | |

A total of 24 SNPs of *ABCD2*, *ABCD3*, and *ABCD4* were identified in 40 ALD patients. Among them, 11 SNPs (45.8%) were novel SNPs. The positions of these novel SNPs were based on the UCSC genome browser hg18

[a] These SNPs were identified only in the cerebral form (childhood cerebral ALD and adult with cerebral ALD)

[b] These SNPs were identified only in the AMN form

**Table 4** Summary of identified single nucleotide polymorphism (SNPs) of *ABCD2*, *ABCD3*, and *ABCD4* in 40 adrenoleukodystrophy patients: known SNPs

A total of 24 SNPs of *ABCD2*, *ABCD3*, and *ABCD4* were identified in 40 ALD patients. Among them, 11 SNPs (45.8%) were novel SNPs. The positions of these novel SNPs were based on the UCSC genome browser hg18

[a] These SNPs were in complete disequilibrium in the Japanese population

| Gene | Fragment | SNP ID | Category | Amino acid change |
|---|---|---|---|---|
| *ABCD3* | Exon1 | rs4148058 | 5′ untranslated region | |
| | Exon2 | rs2147794 | Intron | |
| | Exon3 | rs16946 | Coding synonymous | |
| | Exon7 | rs681187 | Intron | |
| | Exon23 | rs662813 | 3′ untranslated region | |
| | Exon23 | rs337592 | 3′ untranslated region | |
| *ABCD4* | 5′UTR | rs17782508[a] | Upstream at the transcription start site | |
| | Intron1 | rs17182959 | Intron | |
| | Intron1 | rs17158118 | Intron | |
| | Exon3 | rs2301345[a] | Coding synonymous | L62L |
| | Exon9 | rs4148077[a] | Coding nonsynonymous | A304T |
| | Exon10 | rs4148078[a] | Coding synonymous | L320L |
| | Exon11 | rs3742801[a] | Coding nonsynonymous | E368K |

there were no significant differences in the Japanese patients with the cerebral form compared with the controls (Tables 5 and 6).

Given the significant association of the five SNPs (rs17782508, rs2301345, rs4148077, rs4148078, and rs3742801) with the phenotypes of AMN, we then conducted a replication study on an independent French ALD cohort with extreme phenotypes (117 CCALD cases and 71 pure AMN cases). However, we did not find any significant association of these five SNPs with AMN or CCALD. Interestingly, the combination of two intronic SNPs (A (rs17182959) and G (rs7158118)) was significantly more frequently represented in the 51 patients with AMN-Cer than in those with CCALD in the French ALD cohort ($p$=0.0049). The combination of these two intronic SNPs (A (rs17182959) and G (rs7158118)), however, was not present in any of the Japanese patients with ALD, although one combination of two intronic SNPs (A (rs17182959) and G (rs7158118)) was present in the Japanese controls (Tables 5 and 6 ; ESM Table 5).

## Discussion

Our microarray-based high-throughput mutational analysis system was accurate to detect all the mutations, which were confirmed by direct nucleotide sequence analysis. This system should be highly useful for the mutational analysis of *ABCD1* for the diagnosis of patients with ALD, and the diagnosis of the carriers with ALD.

**Table 5** Association studies of detected single nucleotide polymorphism (SNPs) with the clinical phenotypes of Japanese adrenoleukodystrophy patients: novel SNPs

| Gene | SNP name | Allele frequency (number) | | | $p$ value[a] | | |
|---|---|---|---|---|---|---|---|
| | | Cerebral form (a total of 44 chromosomes) | AMN (a total of 26 chromosomes) | Control (number of detected SNPs/total number) | Cerebral form vs AMN | Cerebral form vs control | AMN vs control |
| *ABCD2* | Novel SNP1 | 3 | 0 | 5/164 | 0.2894 | 0.3700 | 1.0000 |
| | Novel SNP2 | 1 | 0 | 3/164 | 1.0000 | 1.0000 | 1.0000 |
| *ABCD3* | Novel SNP3 | 1 | 0 | 0/164 | 1.0000 | 0.2115 | 1.0000 |
| | Novel SNP4 | 0 | 0 | 3/134 | 1.0000 | 1.0000 | 1.0000 |
| | Novel SNP5 | 1 | 0 | 0/160 | 1.0000 | 0.2157 | 1.0000 |
| *ABCD4* | Novel SNP6 | 17 | 9 | 67/164 | 0.8019 | 0.8635 | 0.6680 |
| | Novel SNP7 | 2 | 1 | 2/164 | 1.0000 | 0.1974 | 0.3585 |
| | Novel SNP8 | 1 | 0 | 5/164 | 1.0000 | 1.0000 | 1.0000 |
| | Novel SNP9 | 0 | 1 | 0/164 | 0.3714 | 1.0000 | 0.1368 |
| | Novel SNP10 | 4 | 5 | 14/160 | 0.2766 | 1.0000 | 1.0000 |
| | Novel SNP11 | 1 | 0 | 5/160 | 1.0000 | 1.0000 | 1.0000 |

[a] Results of two-sided Fisher's exact test

**Table 6** Association studies of detected single nucleotide polymorphism (SNPs) with the clinical phenotypes of Japanese adrenoleukodystrophy patients: known SNPs

| Gene | SNP ID | Allele frequency (number) | | | *p* value[b] | | |
|---|---|---|---|---|---|---|---|
| | | Cerebral form (a total of 44 chromosomes) | AMN (a total of 26 chromosomes) | Control (number of detected SNPs/total number) | Cerebral form vs AMN | Cerebral form vs control | AMN vs control |
| *ABCD3* | Rs4148058 | 11 | 5 | 22/164 | 0.7697 | 0.1010 | 0.3820 |
| | Rs2147794 | 6 | 2 | 34/152 | 0.7009 | 0.2880 | 0.0740 |
| | Rs16946 | 6 | 3 | 35/164 | 1.0000 | 0.2933 | 0.3019 |
| | Rs681187 | 17 | 9 | 75/158 | 1.0000 | 0.3109 | 0.2890 |
| | Rs662813 | 18 | 10 | 42/152 | 1.0000 | 0.0984 | 0.1435 |
| | Rs337592 | 2 | 4 | 19/152 | 0.1855 | 0.1714 | 0.7512 |
| *ABCD4* | Rs17782508[a] | 9 | 2 | 42/164 | 0.1921 | 0.5575 | 0.0468 |
| | Rs17182959 | 10 | 7 | 40/128 | 0.5599 | 0.3386 | 0.8163 |
| | Rs17158118 | 10 | 7 | 33/162 | 0.5599 | 0.8344 | 0.4454 |
| | Rs2301345[a] | 9 | 2 | 42/164 | 0.1921 | 0.5575 | 0.0468 |
| | Rs4148077[a] | 9 | 2 | 42/164 | 0.1921 | 0.5575 | 0.0468 |
| | Rs4148078[a] | 9 | 2 | 42/164 | 0.1921 | 0.5575 | 0.0468 |
| | Rs3742801[a] | 9 | 2 | 42/164 | 0.1921 | 0.5575 | 0.0468 |

[a] Five SNPs (rs17782508, rs2301345, rs4148077, rs4848078, and rs3742801) were in complete disequilibrium in the Japanese population

[b] Results of two-sided Fisher's exact test

Although reverse transcription (RT)-PCR has been preferentially used for the analysis of *ABCD1* gene [11, 12, 26] to overcome the difficulty of specifically amplifying the *ABCD1* gene owing to the existence of the related highly homologous genes [13, 27], primers allowing the specific amplification of *ABCD1* enable the PCR analysis of genomic DNA, which is much easier than RT-PCR analysis. Similar approaches have also been used for SSCP-based [13, 28] and DNA-based diagnostic testing methods [29].

In the Japanese ALD patients, mutations of *ABCD1* gene were widely scattered in the entire region of *ABCD1* gene. All types of *ABCD1* mutation were distributed among all the phenotypes of ALD, including childhood cerebral form, AMN, and adulthood cerebral form, suggesting that there is no association of a particular phenotype of ALD with individual mutations as previously observed in other ALD populations [1, 11–13, 15] (http://www.x-ald.nl/). Even among the frameshift mutations that are clearly expected to cause a complete loss of ALDP functions, such *ABCD1* mutations were distributed among all the phenotypes of ALD.

On the basis of the comprehensive resequencing of *ABCD2*, *ABCD3*, and *ABCD4* genes, we searched for SNPs of these related genes to explore the possibility of these genes as candidate disease-modifying genes for ALD, although it was shown that ALD phenotypes are independent of *ABCD2* genotype in two independent association studies of *ABCD2* polymorphisms and ALD phenotypes [30]. Although our study did not reveal SNPs significantly associated with the clinical phenotypes irrespective of the ethnic background, the SNPs in *ABCD4* with suggestive association in the Japanese patients (rs17782508, rs2301345, rs4148077, rs4148078, and rs3742801) and French patients (rs17182959 and rs7158118) may still deserve further investigation including association studies on other independent cohorts and studies on the biological effects related to these SNPs. Among the SNPs with suggestive association in the Japanese patients (rs17782508, rs2301345, rs4148077, rs4148078, and rs3742801), rs4148077 (A304T) substitutes a hydrophilic amino acid for a hydrophobic amino acid, and rs3742801 (E368K) substitutes a basic amino acid for an acidic amino acid. It would be interesting to investigate if these amino acid substitutions may have relevance to the function of *ABCD4*.

In this study, we identified as many as 11 novel SNPs in *ABCD2*, *ABCD3*, and *ABCD4* genes in addition to the 13 previously described SNPs. These findings indicate that there are still numerous novel SNPs with the number comparable to that of previously described SNPs; furthermore, this study places great emphasis on the role of comprehensive resequencing in the discovery of novel SNPs in relevant genes. The novel SNPs as well as previously described ones in *ABCD2*, *ABCD3*, and *ABCD4* genes should be useful for further association studies on ALD and other peroxisome diseases and on the biological implications associated with these SNPs.

The diverse phenotypic variations of ALD still remain enigmatic. Recent studies suggest the role of perosixomes of oligodendrocytes in axonal loss and neuroinflammation [31] and microglial apoptosis as an early pathogenic change in CCALD [32]. With the advancement in our understanding of the pathophysiology of ALD, we hope that we can further probe into the disease-modifying factors on the basis of the molecular pathogenesis of ALD. Genome-wide association studies may well serve as an alternative approach for the identification of disease-modifying genetic factors.

**Acknowledgements** This study was partially supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas, Applied Genomics, the Twenty-first Century COE Program, Center for Integrated Brain Medical Science, and Scientific Research (A) from the Ministry of Education, Culture, Sports, Science and Technology of Japan; a Grant-in-Aid for "the Research Committee for Ataxic Diseases" of the Research on Measures for Intractable Diseases from the Ministry of Health, Labour and Welfare, Japan; and a grant from the Takeda Foundation. MA and PA were supported by grants from INSERM and European Leukodystrophy Foundation (2008-001C4A).

**Disclosures** The experiments comply with the current laws of the country in which they were performed. The authors report no conflicts of interest.



## References

1. Mosser J, Douar AM, Sarde CO, Kioschis P, Feil R, Moser HW, Poustka AM, Mandel JL, Aubourg P (1993) Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. Nature 361:726–730
2. Schaumburg HH, Powers JM, Raine CS, Suzuki K, Richardson EP Jr (1975) Adrenoleukodystrophy. A clinical and pathological study of 17 cases. Arch Neurol 32:577–591
3. Moser HW (1997) Adrenoleukodystrophy: phenotype, genetics, pathogenesis and therapy. Brain 120:1485–1508
4. van Geel BM, Assies J, Wanders RJA, Barth PG (1997) X-linked adrenokeukodystrophy: clinical presentation, diagnosis, and therapy. J Neurol Neurosurg Psychiatry 63:4–14
5. Igarashi M, Schaumburg HH, Powers J, Kishimoto Y, Kolodny EH, Suzuki K (1976) Fatty acid abnormality in adrenoleukodystrophy. J Neurochem 26:851–860
6. Moser HW, Moser AB, Kawamura N, Murphy J, Suzuki K, Schaumburg H, Kishimoto Y (1980) Adrenoleukodystrophy: elevated C26 fatty acid in cultured skin fibroblasts. Ann Neurol 7:542–549
7. Moser AB, Kreiter N, Bezman L, Lu S, Raymond GV, Naidu S, Moser HW (1999) Plasma very long chain fatty acids in 3, 000 peroxisome disease patients and 29, 000 controls. Ann Neurol 45:100–110
8. Aubourg P, Blanche S, Jambaqué I, Rocchiccioli F, Kalifa G, Naud-Saudreau C, Rolland MO, Debre M, Chaussain JL, Griscelli C, Fischer A, Bougnères PF (1990) Reversal of early neurologic and neuroradiologic manifestations of X-linked adrenoleukodystrophy by bone marrow transplantation. N Engl J Med 322:1860–1866
9. Malm G, Ringden O, Anvret M, von Dobeln U, Hagenfeldt L, Isberg B, Knuutila S, Nennesmo I, Winiarski J, Marcus C (1997) Treatment of adrenoleukodystrophy with bone marrow transplantation. Acta Paediatr 86:484–492
10. Shapiro E, Krivit W, Lockman L, Jambaque I, Peters C, Cowan M, Harris R, Blanche S, Bordigoni P, Loes D, Ziegler R, Crittenden M, Ris D, Berg B, Cox C, Moser H, Fischer A, Aubourg P (2000) Long-term effect of bone-marrow transplantation for childhood-onset cerebral X-linked adrenoleukodystrophy. Lancet 356:713–718
11. Krasemann EW, Meier V, Korenke GC, Hunneman DH, Hanefeld F (1996) Identification of mutations in the ALD-gene of 20 families with adrenoleukodystrophy/adrenomyeloneuropathy. Hum Genet 97:194–197
12. Takano H, Koike R, Onodera O, Sasaki R, Tsuji S (1999) Mutational analysis and genotype-phenotype correlation of 29 unrelated Japanese patients with X-linked adrenoleukodystrophy. Arch Neurol 56:295–300
13. Kok F, Neumann S, Sarde CO, Zheng S, Wu KH, Wei HM, Bergin J, Watkins PA, Gould S, Sack G, Moser HW, Mandel J, Smith KD (1995) Mutational analysis of patients with X-linked adrenoleukodystrophy. Hum Mutat 6:104–115
14. Ligtenberg MJ, Kemp S, Sarde CO, van Geel BM, Kleijer WJ, Barth PG, Mandel JL, van Oost BA, Bolhuis PA (1995) Spectrum of mutations in the gene encoding the adrenoleukodystrophy protein. Am J Hum Genet 56:44–50
15. Berger J, Molzer B, Fae I, Bernheimer H (1994) X-linked adrenoleukodystrophy (ALD): a novel mutation of the ALD gene in 6 members of a family presenting with 5 different phenotypes. Biochem Biophys Res Commun 205:1638–1643
16. Smith KD, Kemp S, Braiterman LT, Lu JF, Wei HM, Geraghty M, Stetten G, Bergin JS, Pevsner J, Watkins PA (1999) X-linked adrenoleukodystrophy: genes, mutations, and phenotypes. Neurochem Res 24:521–535
17. Broccardo C, Troffer-Charlier N, Savary S, Mandel JL, Chimini G (1998) Exon organisation of the mouse gene encoding the adrenoleukodystrophy related protein (ALDRP). Eur J Hum Genet 6:638–641
18. Lombard-Platet G, Savary S, Sarde CO, Mandel JL, Chimini G (1996) A close relative of the adrenoleukodystrophy (ALD) gene codes for a peroxisomal protein with a specific expression pattern. Proc Natl Acad Sci USA 93:1265–1269
19. Guimarães CP, Domingues P, Aubourg P, Fouquet F, Pujol A, Jimenez-Sanchez G, Sá-Miranda C, Azevedo JE (2004) Mouse liver PMP70 and ALDP: homomeric interactions prevail in vivo. Biochim Biophys Acta 1689:235–243
20. Hillebrand M, Verrier SE, Ohlenbusch A, Schäffer A, Söling HD, Wouters FS, Gärtner J (2007) Live cell FRET microscopy: homo- and heterodimerization of two human peroxisomal ABC transporters, the adrenoleukodystrophy protein (ALDP, ABCD1) and PMP70 (ABCD3). J Biol Chem 282:26997–27005
21. Liu LX, Janvier K, Berteaux-Lecellier V, Cartier N, Benarous R, Aubourg P (1999) Homo- and heterodimerization of peroxisomal ATP-binding cassette half-transporters. J Biol Chem 274:32738–32743
22. Tanaka AR, Tanabe K, Morita M, Durisu M, Kasiwayama Y, Matsuo M, Kioka N, Amachi T, Imanaka I, Ueda K (2002) ATP binding/hydrolysis by and phosphorylation of peroxisomal ATP-binding cassette proteins PMP70 (*ABCD3*) and adrenoleukodystrophy protein (*ABCD1*). J Biol Chem 277:40142–40147
23. Takahashi Y, Seki N, Ishiura H, Mitsui J, Matsukawa T, Kishino A, Onodera O, Aoki M, Shimozawa N, Murayama S, Itoyama Y, Suzuki Y, Sobue G, Nishizawa M, Goto J, Tsuji S (2008) Development of

high-throughput microarray-based resequencing system for neurological disorders and its application to molecular genetics of amyotrophic lateral sclerosis. Arch Neurol 65:1326–1332
24. Wigginton JE, Abecasis GR (2005) PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. Bioinformatics 21:3445–3447
25. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265
26. Lachtermacher MB, Seuanez HN, Moser AB, Moser HW, Smith KD (2000) Determination of 30 X-linked adrenoleukodystrophy mutations, including 15 not previously described. Hum Mutat 15:348–353
27. Sarde CO, Mosser J, Kioschis P, Kretz C, Vicaire S, Aubourg P, Poustka A, Mandel JL (1994) Genomic organization of the adrenoleukodystrophy gene. Genomics 22:13–20
28. Feigenbaum V, Lombard-Platet G, Guidoux S, Sarde C, Mandel JL, Aubourg P (1996) Mutational and protein analysis of patients and heterozygous women with X-linked adrenoleukodystrophy. Am J Hum Genet 58:1135–1144
29. Boehm CD, Cutting GR, Lachtermacher MB, Moser HW, Chong SS (1999) Accurate DNA-based diagnostic and carrier testing for X-linked adrenoleukodystrophy. Mol Genet Metab 66:128–136
30. Maier EM, Mayerhofer PU, Asheuer M, Köhler W, Rothe M, Muntau AC, Roscher AA, Holzinger A, Aubourg P, Berger J (2008) X-linked adrenoleukodystrophy phenotype is independent of ABCD2 genotype. Biochem Biophys Res Commun 377:176–180
31. Kassmann CM, Lappe-Siefke C, Baes M, Brügger B, Mildner A, Werner HB, Natt O, Michaelis T, Prinz M, Frahm J, Nave KA (2007) Axonal loss and neuroinflammation caused by peroxisome-deficient oligodendrocytes. Nat Genet 39:969–976
32. Eichler FS, Ren JQ, Cossoy M, Rietsch AM, Naqpal S, Moser AB, Frosch MP, Ransohoff RM (2008) Is microglial apoptosis an early pathogenic change in cerebral X-linked adrenoleukodystrophy? Ann Neurol 63:729–742
33. Grabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adenemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229

*Human Molecular Genetics, 2010, Vol. 19, Review Issue 1* **R65–R70**
*doi:10.1093/hmg/ddq162*
*Advance Access published on April 22, 2010*

# Genetics of neurodegenerative diseases: insights from high-throughput resequencing

**Shoji Tsuji***

Department of Neurology, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655, Japan

Received April 14, 2010; Revised and Accepted April 19, 2010

**During the past three decades, we have witnessed remarkable advances in our understanding of the molecular etiologies of hereditary neurodegenerative diseases, which have been accomplished by 'positional cloning' strategies. The discoveries of the causative genes for hereditary neurodegenerative diseases accelerated not only the studies on the pathophysiologic mechanisms of diseases, but also the studies for the development of disease-modifying therapies. Genome-wide association studies (GWAS) based on the 'common disease–common variants hypothesis' are currently undertaken to elucidate disease-relevant alleles. Although GWAS have successfully revealed numerous susceptibility genes for neurodegenerative diseases, odds ratios associated with risk alleles are generally low and account for only a small proportion of estimated heritability. Recent studies have revealed that the effect sizes of the disease-relevant alleles that are identified based on comprehensive resequencing of large data sets of Parkinson disease are substantially larger than those identified by GWAS. These findings strongly argue for the role of the 'common disease–multiple rare variants hypothesis' in sporadic neurodegenerative diseases. Given the rapidly improving technologies of next-generation sequencing next-generation sequencing (NGS), we expect that NGS will eventually enable us to identify all the variants in an individual's personal genome, in particular, clinically relevant alleles. Beyond this, whole genome resequencing is expected to bring a paradigm shift in clinical practice, where clinical practice including diagnosis and decision-making for appropriate therapeutic procedures is based on the 'personal genome'. The personal genome era is expected to be realized in the near future, and society needs to prepare for this new era.**

## INTRODUCTION

Neurodegenerative diseases are usually characterized by onset in late adulthood, a slowly progressive clinical course and neuronal loss with regional specificity in the central nervous system. In Alzheimer disease, Parkinson disease (PD), spinocerebellar ataxias and amyotrophic lateral sclerosis, neurodegeneration preferentially involves the cerebral cortex, extrapyramidal system, cerebellum and spinal cord, respectively. Although the majority of neurodegenerative diseases are sporadic, Mendelian inheritance patterns have been well documented. Intriguingly, the clinical presentations and neuropathological findings of hereditary forms of these neurodegenerative diseases are often indistinguishable from the sporadic diseases, raising the possibility that common pathophysiologic mechanisms underlie both hereditary and sporadic neurodegenerative diseases.

During the past three decades, there have been remarkable advances in our understanding of the etiologies of hereditary neurodegenerative diseases, which have been accomplished by 'positional cloning' efforts (1–4). The identification of the causative genes for hereditary neurodegenerative diseases has accelerated studies on the pathophysiologic mechanisms of diseases and the development of disease-modifying therapies based on these discoveries has now become a reality.

### Molecular bases of neurodegenerative disease with Mendelian traits

Establishment of positional cloning strategies (1–4) including high throughput linkage analysis employing microarrays (5–6) has further accelerated the search for the causative genes for diseases with Mendelian traits. Furthermore, the availability of the human genome sequence (7) has tremendously acceler-

*To whom correspondence should be addressed. Email: tsuji@m.u-tokyo.ac.up

ated the discovery of causative genes. Despite this progress, however, the mutations causing a substantial number of hereditary diseases remain to be identified. In familial amyotrophic lateral sclerosis (FALS), in particular, its causative genes have been identified in only 25–30% of FALS cases, suggesting that the majority of FALS genes remain to be identified (8–10). When the pedigree size is limited, it is difficult to narrow the candidate region by linkage analysis; hence, tremendous effort is still required to identify the causative genes based on positional cloning strategies. In diseases such as FALS in which the clinical severity is substantial, the number of living affected individuals is often limited. Thus, for many disorders, we need high throughput comprehensive resequencing capability to identify the causative mutations located in broad candidate regions of 10–100 Mb.

## MOLECULAR BASIS OF SPORADIC NEURODEGENERATIVE DISEASES

For sporadic neurodegenerative diseases, which comprise the majority of the cases, we are still far from understanding their molecular etiologies despite the clues obtained on the basis of neuropathological findings. For example, although we know that accumulation of senile plaques in which the β amyloid protein is the major component underlies both sporadic and familial Alzheimer disease, we have little knowledge on the molecular etiologies of sporadic Alzheimer disease. We occasionally observe affected siblings or relatives with neurodegenerative diseases, which raises the possibility of involvement of genetic factors in these diseases. To identify susceptibility genes that account for the heritability seen for complex traits, genome-wide association studies (GWAS) employing common single nucleotide polymorphisms (SNPs) have been conducted. The theoretical framework for GWAS is the 'common disease–common variant hypothesis', in which common diseases are attributable in part to allelic variants present in more than 1–5% of the population (11–13).

Although GWAS have successfully revealed numerous susceptibility genes for common diseases such as diabetes as well as neurodegenerative diseases, the odds ratios associated with these risk alleles are generally low and account for only a small proportion of estimated heritability (14–16). It is assumed that risk alleles with large effect size may be rare in frequency and hard to detect by GWAS employing common SNPs. Emerging new technologies of next-generation sequencers will eventually enable the identification of all the variants including 'rare variants' in single subjects. In this review, future directions for identifying disease-relevant genetic variations on the basis of comprehensive resequencing of the human genome employing the next-generation sequencers are discussed.

## ROLE OF RARE VARIANTS IN NEURODEGENERATIVE DISEASES

The general finding that the odds ratios associated with risk alleles identified for disease susceptibility by GWAS are low indicates that GWAS based on the 'common disease–common variants hypothesis' are not effective in identifying genetic risks with large effect sizes. High λs, estimating recurrent risks for siblings of affected individuals have been demonstrated in many diseases with complex traits, but the genetic risk factors identified by GWAS do not account for the high λs. Current experience with GWAS strongly suggests that rarer variants that are hard to detect by GWAS may account for the 'missing' heritability. Such rare variants may have large effect sizes as genetic risk factors for diseases. Thus we need a paradigm shift from the 'common disease–common variants hypothesis' to a 'common disease–multiple rare variants hypothesis' to identify disease-relevant alleles with large effect sizes.

The prominent role of rare variants in neurodegenerative disease is best highlighted by the recent discovery of the glucocerebrosidase gene (GBA) as a robust genetic risk factor for PD (17–18). PD, which is characterized by tremor, rigidity, bradykinesia, and postural instability, is the second most common neurodegenerative disease after Alzheimer disease, with onset typically in late adulthood. The prevalence of PD has been estimated to be 0.3% in the general population and 1% in people over 60 years of age. Although α-synuclein (SNCA), leucine-rich repeat kinase 2 (LRRK2), UCHL-1, Parkin (PARK2), PTEN-induced putative kinase 1 (PINK1) and DJ-1 have been identified as causative genes for familial PD, PD patients with pathogenic mutations in these genes are rare, and most of the PD cases are sporadic, the etiologies of which are poorly understood. A population-based study coupled with genealogy information demonstrated that the estimated risk ratio for PD for siblings of patients with PD was significantly elevated ($\lambda s = 6.3$), indicating that genetic factors substantially contribute to the development of sporadic PD (18). Recent clinical observations (19) suggested the association of sporadic PD with heterozygous mutations in the glucocerebrosidase gene (GBA) encoding the enzyme that is deficient in patients with Gaucher disease, an autosomal recessive lysosomal storage disease. Furthermore co-morbidity of PD and Gaucher disease had previously been described (20). We conducted an extensive resequencing analysis of GBA in PD patients and controls, and found that GBA variants that are pathogenic for Gaucher disease confer a robust susceptibility to sporadic PD, and, even account for familial clustering of PD (18) (Fig. 1). The combined carrier frequency of the 'pathogenic variants' was as high as 9.4% in PD patients and significantly more frequent than in controls (0.37%) with a markedly high odds ratio of 28.0 (95% CI, 7.3 to 238.3) for PD patients compared with controls.

The molecular effects of the 'pathogenic variants' in PD remain to be elucidated. Gain of toxic functions of the mutant glucocerebrosidase proteins independent of enzyme activity might be involved in the pathogenesis. Intriguingly, however, all the variants associated with PD are 'pathogenic variants' for Gaucher disease, raising the possibility that a decreased glucocerebrosidase activity plays a role in the pathogenesis of PD. Identification of a splice junction mutation, 'IVS6+1g>a', which is predicted to lead to a loss of function due to a premature stop codon, in this study may further support this notion (18).

Many GWAS have recently been conducted to identify susceptibility genes for PD. Satake *et al.* (21) have recently published the results of their GWAS on Japanese PD cases and

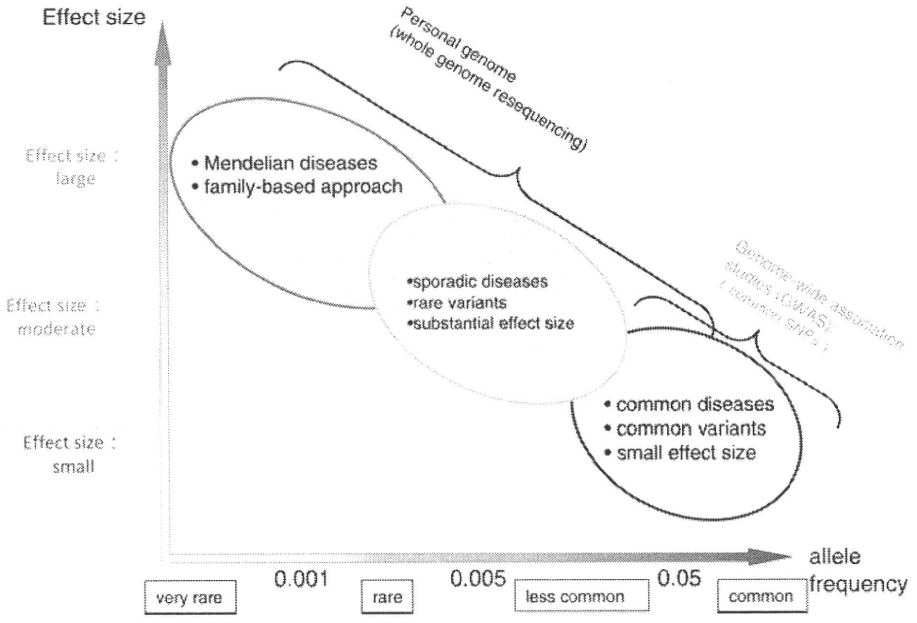


**Figure 1.** Research paradigm to identify disease-related variations based on comparison of effect sizes of variants and allele frequencies of the variants in population. Adapted by permission from Macmillan Publishers Ltd: *Nature*, **461**: 747–53 (2009) (16).

**Table 1.** Comparison of allele frequencies and odds ratios of disease-relevant variations

| Variants | Parkinson disease (%) | Controls (%) | Odds ratio (95% confidence interval) |
|---|---|---|---|
| GBA[a] | 9.4 | 0.4 | 28.0 (7.3–238.3) |
| SNCA (rs11931074)[b] | 32 | 42 | 1.50 (1.34–1.68) |
| LRRK2 (rs1994090)[b] | 11 | 8 | 1.43 (1.20–1.70) |
| BST1 (rs11931532)[b] | 45 | 40 | 1.22 (1.09–1.35) |
| PARK16 (rs947211)[b] | 43 | 48 | 1.23 (1.11–1.37) |

[a]Mitsui *et al.* (18).
[b]Satake *et al.* (21).

controls. They found four genetic risk factors including LRRK2 and SNCA. As shown in the Table 1, the odds ratios are relatively low (1.24–1.37) despite the significant *P*-values. More importantly, the GBA locus on chromosome 1 was not detected in their GWAS, presumably because rare variants such as those in GBA are hard to detect by GWAS using common SNPs (tag SNPs). These results provide the following lessons: (i) risk factors with substantially high odds ratio are present in common diseases such as PD; (ii) rare variants are present at low frequencies; (iii) multiple rare GBA variants were detected only through comprehensive resequencing of the gene; and (iv) GWAS on the same population did not identify the locus for the susceptibility gene harboring multiple rare variants. These data demonstrate the power of resequencing strategies for the identification of rare variants in neurodegenerative disease.

As Manolio *et al.* (16) recently described that GWAS have identified hundreds of genetic variants associated with complex human diseases and traits, and have provided valuable insights into their genetic architecture. However, because most variants identified to date confer relatively small increments in risk and explain only a small proportion of familial clustering, a remaining challenge will be to define the genetic basis of the 'missing' heritability.

## APPLICATION OF HIGH THROUGHPUT RESEQUENCING FOR THE IDENTIFICATION OF RARE GENETIC VARIANTS WITH LARGE EFFECT SIZES

As discussed earlier, GWAS are inefficient in identifying rare variants associated with disease susceptibility and instead whole genome resequencing will be required. For PD, clinical observations suggested an association of PD and Gaucher disease (19–20). Without such clinical observations, however, comprehensive whole genome or exome resequencing will be required to identify rare variants relevant to disease. To accomplish this goal, high throughput resequencing efforts employing next-generation sequencers will be the most promising approach.

### Developing next-generation sequencing technologies

The automated Sanger method is considered a 'first-generation' technology, and newer methods are referred to as next-generation sequencing (NGS) (22). As shown in Figure 2, the throughput of NGS is dramatically increasing. As of 2010, the throughput is 100–200 Gb/run. Since the cost for whole genome resequencing for a read depth sufficient to identify variants with a high accuracy is still expensive, it is not easy to resequence the whole genome of a large number of individuals. Thus, we need to develop strategies to efficiently identify disease-relevant variants employing technologies with high accuracy and reasonable cost.