Original Article

Exome Sequencing and Linkage Analysis Identified Novel Candidate Genes in Recessive Intellectual Disability Associated with Ataxia

Roshanak Jazayeri MD[1], Hao Hu PhD[2], Zohreh Fattahi PhD student[1], Luciana Musante PhD[2], Seyedeh Sedigheh Abedini PhD student[1], Masoumeh Hosseini PhD student[1], Thomas F. Wienker MD PhD[2], Hans Hilger Ropers MD PhD[2], Hossein Najmabadi PhD[1], Kimia Kahrizi MD•[1]

Abstract

Background: Intellectual disability (ID) is a neuro-developmental disorder which causes considerable socio-economic problems. Some ID individuals are also affected by ataxia, and the condition includes different mutations affecting several genes.

Methods: We used whole exome sequencing (WES) in combination with homozygosity mapping (HM) to identify the genetic defects in five consanguineous families among our cohort study, with two affected children with ID and ataxia as major clinical symptoms.

Results: We identified three novel candidate genes, *RIPPLY1*, *MRPL10*, *SNX14*, and a new mutation in known gene *SURF1*. All are autosomal genes, except *RIPPLY1*, which is located on the X chromosome. Two are housekeeping genes, implicated in transcription and translation regulation and intracellular trafficking, and two encode mitochondrial proteins. The pathogenesis of these variants was evaluated by mutation classification, bioinformatic methods, review of medical and biological relevance, co-segregation studies in the particular family, and a normal population study.

Conclusions: Linkage analysis and exome sequencing of a small number of affected family members is a powerful new technique which can be used to decrease the number of candidate genes in heterogenic disorders such as ID, and may even identify the responsible gene(s).

Keywords: Ataxia, exome sequencing, gene, intellectual disability, recessive

Cite this article as: Jazayeri R, Hu H, Fattahi Z, Musante L, Abedini SS, Hosseini M, Wienker TF, Ropers HH, Najmabadi H, Kahrizi K. Exome Sequencing and Linkage Analysis Identified Novel Candidate Genes in Recessive Intellectual Disability Associated with Ataxia. *Arch Iran Med*. 2015; 18(10): 670 – 682.

Introduction

Intellectual disability (ID) affects around 1%–3% of the general population and is diagnosed based on the simultaneous presence of these essential conditions: an intelligence quotient (IQ) below 70 and significant limitations in two or more adaptive skill areas. Etiologically, genetic causes are now believed to be the most important, and have been linked to more than 50% of patients with ID.[1–3] Based on clinical manifestations, ID is subdivided into non-syndromic ID, in which intellectual deficits appear without other abnormalities, and syndromic ID, in which intellectual deficits are associated with other medical and behavioral signs and symptoms such as malformations, dysmorphic features, or neurological abnormalities such as ataxia. In 2004, the Genetics Research Center (GRC) in Tehran started an ongoing "ARID (autosomal recessive intellectual disability) genetic causes" cohort study in consanguineous families because of the prevalence of consanguinity (about 40%) and youth (more than half of the population are 30 years of age or younger) in Iran.[4,5] In 2011, using target enrichment and next generation sequencing, we identified 50 novel candidate ARID genes.[5] More recently, the application of exome sequencing for recognizing a diseased gene in the coding areas of the human genome has enabled the recognition of causative variations in many known and unknown ID genes.[6,7] ID associated with ataxia, reported in some known and recently identified syndromes, is also phenotypically and genetically heterogeneous. From our large cohort study with ARID and cerebellar ataxia, we report five families here who have been investigated using linkage analysis and whole-exome sequencing (WES).

Authors' affiliations: [1]Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. [2]Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany.
•**Corresponding author and reprints:** Kimia Kahrizi MD, Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. Telefax: +98-21-22180138; E-mail: kkahrizi@uswr.ac.ir, kahrizi@yahoo.com
Accepted for publication: 12 August 2015

Materials and Methods

Ethics statement

The study was approved by the Ethics Committee of the University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. Written informed consent was obtained from the parents of all patients.

Family ascertainment and clinical diagnosis

We identified five consanguineous Iranian families with at least two individuals with ID and ataxia in a large ARID cohort collected by the Genetics Research Center (GRC), University of Social and Welfare Rehabilitation Sciences, in Tehran (Figure 1). The Wechsler Intelligence Scales for Children (WISC) and the Wechsler Adult Intelligence Scales (WAIS) were used to assess the IQ of children and parents. Behavioral and mental status were evaluated by psychiatrists. All participants underwent clinical evaluation and physical neurological examination. Whenever possible, a brain MRI scan was also performed on one affected individual from each family to confirm cerebellar involvement as a

cause of ataxia (Figure 2). Chromosomal abnormalities (by karyotyping), Fragile X syndrome (by molecular methods or Southern blotting), metabolic diseases (by tandem mass spectrometry) and autosomal recessive primary microcephaly (MCPH) (by linkage to known loci for microcephaly) had previously been excluded in these families. Peripheral blood samples were obtained and genomic DNA was extracted according to standard procedures.

SNP genotyping and homozygosity mapping

We performed genome-wide single nucleotide polymorphism (SNP) genotyping on all available affected and unaffected individuals (at least two affected individuals, two parents and one normal sibling) from the five families under investigation (seven individuals from family M-244 and five individuals from the others), using an Axiom® Genome-Wide CEU 1 Array Plate (Affymetrix, Santa Clara, CA, USA). We used Alohomora software version 0.32 to prepare SNP data for linkage analysis, drawing haplotypes with the Merlin program. We performed multipoint linkage analysis compatible with an autosomal recessive inheritance model and a disease allele frequency of 0.001, using Merlin 13 software, which is based on Lander-Green algorithm.[8] Thus, for each family, we searched for common homozygous regions shared among the affected individuals and absent within the unaffected family members. Haplotypes were constructed using HaploPainter software.

Whole-exome sequencing (WES) and data analysis

We performed WES on the genomic DNA from five affected individuals of the five families (only one affected individual from each family). For whole exome sequencing, genomic DNA was extracted from peripheral blood, and 3 µg (in a volume of 100 µL) was sheared into fragments of about 250 bp. The fragmented DNA was end-repaired and processed with an Illumina multiplex adapter ligation kit. About 1 µg of processed DNA was used as input for the Agilent SureSelect Whole Exome Enrichment kit version 4. Whole-exome sequencing was carried out using a Hiseq2500 sequencer in the mode of 101 bp paired-end sequencing. More than 95% of the targeted coding exons were covered by at least 10 non-redundant sequencing reads. Variant detection, annotation, filtering, and prioritization were conducted with the MERAP package.[9] Variations, which were located in linkage intervals identified by homozygosity mapping, were considered as candidate causal variants. Mutations fitting the model of recessive inheritance, with a probable damaging effect on protein function (splice site, nonsense, missense, and frameshift) were filtered against multiple databases: dbSNP138, the 1000 Genomes Project, the ESP6500 project, and the exomes from 200 Danish individuals. The final candidate mutations were confirmed by Sanger sequencing in the affected individuals using the ABI PRISM Big Dye Terminator Cycle Sequencing Kit and ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Alleles were discriminated using CodonCode Aligner software. In addition, siblings and parents were also sequenced, whenever available, to confirm appropriate co-segregation of the allele within the family. Furthermore, 250 unaffected individuals with matched geographical ancestry were recruited as controls for this study.

Results

Family recruitment and clinical features

A total of five consanguineous families, composed of 10 affected and 17 unaffected individuals (parents and normal siblings), were recruited for this study (Figure 1). Affected members in these five families showed ID and progressive truncal ataxia, some had generalized seizure. The average age of patients at the time of recruitment was 27 years (range 15–41 years). All patients were male. Comprehensive examination of the affected individuals and their medical history identified some other clinical features in certain families (Table 1). Details of clinical examination and brain MRI results are summarized in Table 1.

Linkage analysis

We used genotyping results from at least five DNA samples from each family (Figure 1) for the purpose of homozygosity mapping (Figure 3) and multipoint parametric linkage analysis (Figure 4). In family M244, a single prominent (ca 62 Mbp) peak with a LOD (logarithm of odds) score of 2.63, and including a total of 986 genes was identified on chromosome 1. In family M9000121, we found a 25 Mbp interval with LOD score 2.22 on chromosome 6, containing 187 genes. In three other families, several homozygous regions and multiple peaks with maximum LOD scores of about 2 (1.92) were identified. All regions of homozygosity identified in the five families under investigation are included in Table 2.

Exome sequencing identified candidate genes

Five affected individuals, one from each family (M244_4781, M9000065_36340, M9000097_36980, M9000111_37260, and M9000121_37460), were included in a WES study. The data analysis based on a recessive inheritance model, identified 10 likely protein damaging mutations (one in family M9000097, three in family M9000111, and two in each of the other families M244, M9000065 and M9000121). Prediction of the pathogenicity of variants using different programs is given in Table 4.

After variant filtering, *in silico* prediction of variant pathogenicity, confirmation by Sanger sequencing and co-segregation studies, only four variants (in *RIPPLY1, SNX14, MRPL10* and *SURF1* genes) were confirmed. In family M244, a deletion of a nucleotide (C) near the end of the 3rd exon in *RIPPLY1* gene on chromosome X, results in a frameshift change from amino acid position 93 and causes a premature stop codon that affects the protein product, which is a developmental transcription regulator. In family M9000121, a point mutation G>T in *SNX14* gene on chromosome 6, causes a premature stop codon (TAG) at position 153 in *SNX14*, which is an important neuronal protein. In family 9000097, a point mutation G>A in *MRPL10* gene on chromosome 17, changes arginine to tryptophan at position 176 in the mitochondrial ribosomal protein MRPL10. In family M9000065, we identified a missense variant, due to a point mutation C>A in exon 8 of *SURF1* gene on chromosome 9, which encodes a protein involved in the biogenesis of the cytochrome C oxidase complex. No variants co-segregated with the disease in family M9000111, therefore remaining unresolved. All changes identified by WES in this study are included in Table 3.

Discussion

Using linkage analysis, we studied five consanguineous Iranian families with ID and cerebellar ataxia. Non-parametric linkages analysis are used when disease model is unknown. Here, consistent with recessive inheritance model of our families, we used classic parametric approach whereby the probability that a gene

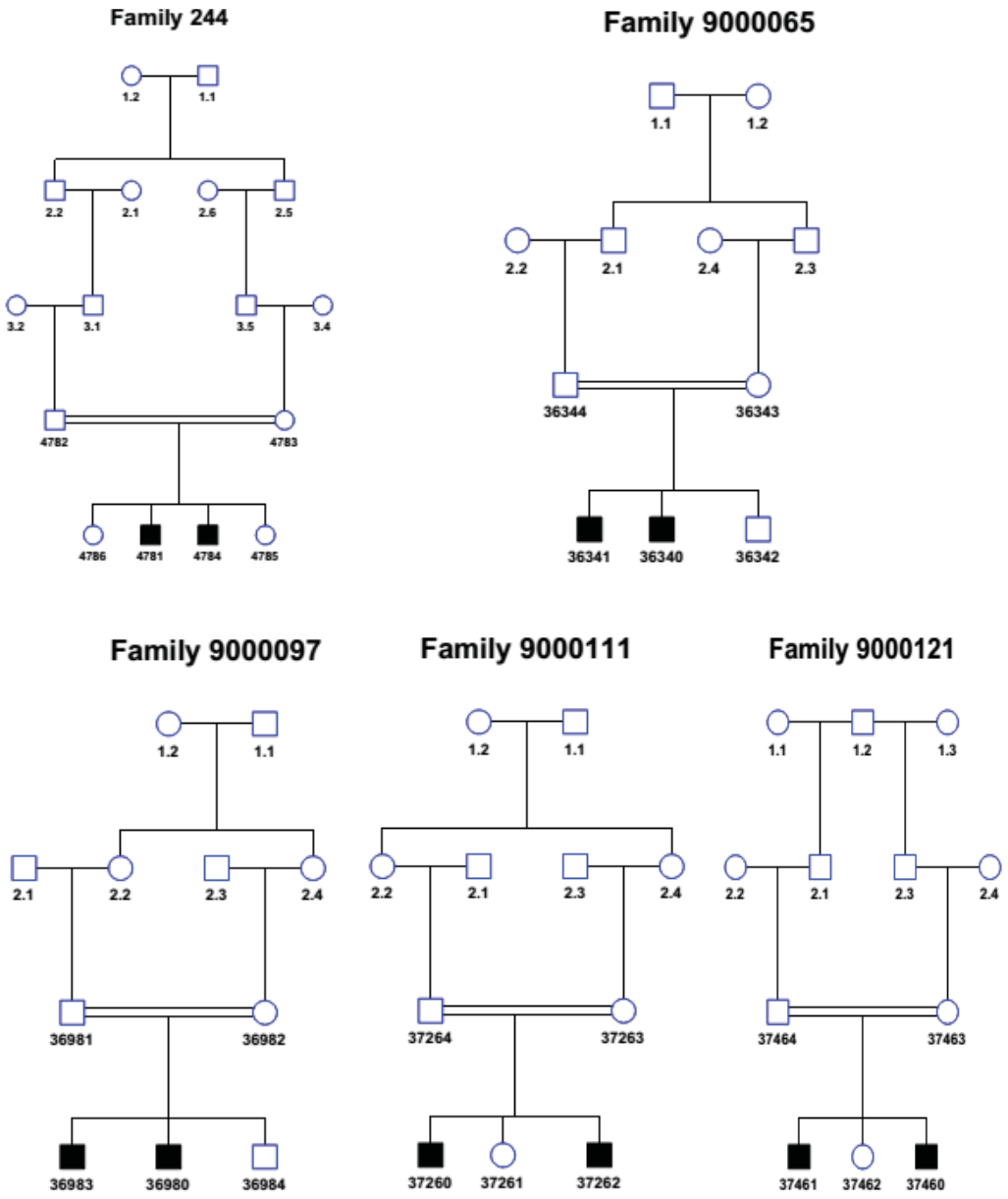


Figure 1. Pedigrees of five families with ID and associated ataxia.

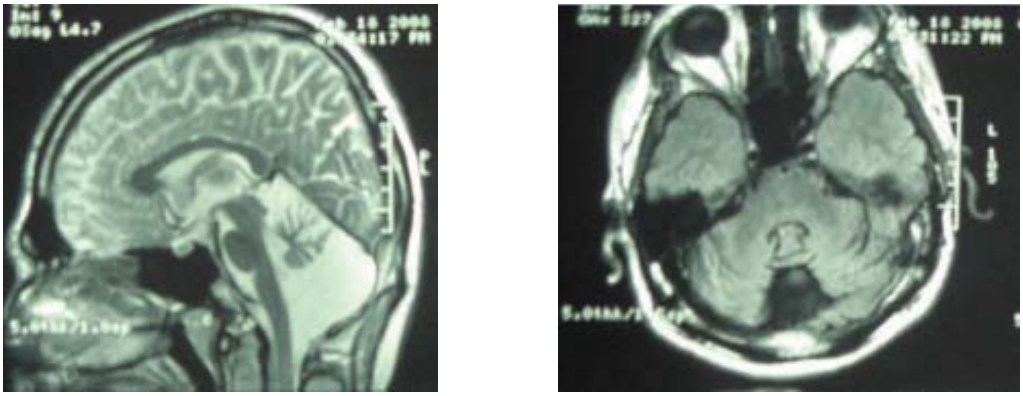


Figure 2. Brain MRI scans showing cerebellar hypoplasia in family M244, I_4781.

Table 1. Clinical data summary.

Family		Family M244		Family M9000065		Family M9000097		Family M9000111		Family M9000121	
R coefficient		1/32		1/8		1/8		1/8		1/64	
ID		I-4781	I-4784	I-36340	I-36341	I-36980	I-36983	I-37260	I-37262	I-37460	I-37461
Gender		M	M	M	M	M	M	M	M	M	M
Birth		Term	Term	N/A	N/A	N/A	N/A	Term	Preterm	Term	Term
	OFC (cm)	34 (-0.2 SDS)	35 (+0.1 SDS)	N/A	N/A	N/A	N/A	35 (+0.1 SDS)	34 (-0.2 SDS)	35 (+0.1 SDS)	35 (+0.1 SDS)
Age (yrs) at examination		20	18	15	19	31	42	28	21	27	38
Age (yrs) at present		24	22	18	22	34	45	31	24	30	41
IQ		40	36	24	26	42	40	39	42	24	26
HC (cm)		56 (0 SDS)	58 (+2SDS)	53.5 (–1 SDS)	55 (0 SDS)	56 (+1 SDS)	56 (+1 SDS)	53 (–2 SDS)	53 (–2 SDS)	58 (+2 SDS)	54 (–1 SDS)
Length (cm)		176 (0 SDS)	171 (-1 SDS)	148 (-3 SDS)	168 (+1 SDS)	164 (-2 SDS)	170 (-1 SDS)	133 (-6 SDS)	150 (-4 SDS)	145 (-5 SDS)	138 (-6 SDS)
Initial cerebellar signs		GA/D	GA/D	GA/D	GA/D	GA/D	GA/D	GA/D	GA	GA/D	GA/D
Quadrupedal gait		+	+	-	-	-	-	-	-	+	+
Seizures/onset		+/16y	+/15y	-	-	+/4y	+/1y	-	-	-	+/1y
Seizures therapy response		+	+	-	-	+	+	-	-	-	+
Muscular hypotonia		-	-	-	-	-	-	+	-	-	-
Spasticity		+	-	-	-	-	-	-	-	+	-
Cerebellar atrophy (MRI)		+	N/A	N/A	N/A	N/A	N/A	+	N/A	N/A	N/A
Facial dysmorphism		Bulbous nose, thin upper lip, square mandible	Bulbous nose, thin upper lip, square mandible	_	_	_	_	Down-slanting palpebral fissures, strabismus, short philtrum	Strabismus, short philtrum	Long face, prognathism, big nose, strabismus	Long face, prognathism
Behavior (social development)		Low activity, quiet temperament (parent report)	Low activity, quiet temperament (parent report)	Aggression	Normal	Aggression	Aggression	Normal	Aggression	Aggression (DASH-II)	Quiet temperament (parent report)
Musculoskeletal	Upper	-	-	-	-	-	Left hand paralysis	-	-	-	-
	Lower	-	-	-	-	-	-	-	-	-	-
Eyes	Keratoconus	+	-	-	-	-	-	-	-	-	-
	Cataracts	-	-	+	+	-	-	+	+	-	-
	Nystagmus	-	+	-	-	-	-	-	-	-	-
Ears	SNHL	+/ 12 y	+/ 10y	-	-	-	-	-	-	-	-
Skin	Hyperpigmentation	-	-	-	-	-	-	+	-	-	-
Other clinical signs		-	-	-	-	-		-	-	-	Scoliosis

D, dysarthria; GA, gait ataxia; HC, head circumference; M, male; N/A, not assessed; OFC, occipitofrontal circumference; R, coefficient of relationship; SNHL, sensory neural hearing loss

Table 2. Details of intervals detected by homozygosity mapping.

Family	Chromosome	LOD score	Start	End	No. of genes	Size of interval (Mbp)
M244	Chr. 1	2.63	rs680714	rs12742614	986	61.8
M9000097	Chr. 5	1.92	rs2173759	rs4869452	120	19.62
	Chr. 4	1.92	rs11737133	rs9992616	30	2.62
	Chr. 17	1.92	rs11652705	rs12453046	208	7.07
	Chr. 19	1.92	rs10415034	rs278183	258	14.47
M9000121	Chr. 6	2.22	rs12204438	rs6935364	187	24.97
M9000065	Chr. 1	1.92	rs6657679	rs9662349	222	28.06
	Chr. 2	1.92	rs300780	rs17494521	42	8.31
	Chr. 3	1.92	rs2900640	rs2290165	25	2.38
	Chr. 7	1.92	rs314616	rs6954232	82	5.29
	Chr. 9	1.92	rs2236547	rs7868484	12	1.92
	Chr. 10	1.81	rs11199849	rs11245450	44	3.66
	Chr. 14	1.92	rs1350270	rs6575862	131	3.6
	Chr. 22	1.92	rs5752639	rs737976	26	1.71
M9000111	Chr. 20	1.92	rs34301726	rs2663004	95	6.14
	Chr. 1	1.92	rs12564866	rs11807941	57	4.05
	Chr. 2	1.92	rs2357486	rs10206954	10	0.55
	Chr. 12	1.92	rs2137564	rs10843894	36	4.82
	Chr. 13	1.92	rs9318814	rs9575186	1	0.97
	Chr. 18	1.79	rs8097943	rs1689030	25	2.35
	Chr. 22	1.81	rs131408	rs5996674	30	0.71

Table 3. Candidate variants identified by WES in this study.

Sample	Candidate genes	Genomic positions (Hg19)	Gene	Nucleotide change	Protein change	Status
M244_4781	2	Chr.X:13794385	*GPM6B*	c.C809T	p.A270V	X-linked recessive (not co-segregating)
		Chr.X:106144724	*RIPPLY1*	c.276del	p.K93Sfs*71	X-linked recessive (co-segregating)
M9000065_36340	2	Chr.9:136218979	*SURF1*	c.G770T	p.G257V	Co-segregating
		Chr.9:134735990	*MED27*	c.G871A	p.G291S	Not co-segregating
M9000097_36980	1	Chr.17:45904039	*MRPL10*	c.C526T	p.R176W	co-segregating
M9000111_37260	3	Chr.X:99934379	*SYTL4*	c.T1589C	p.V530A	Not co-segregating
		Chr.22:50278793	*ZBED4*	c.A1483G	p.S495G	Not co-segregating
		Chr.10:119799727	*RAB11FIP2*	c.C703T	p.H235Y	Not co-segregating
M9000121_37460	2	Chr.16:89178551	*ACSF3*	c.G874A	p.A292T	Not co-segregating
		Chr.6:86277254	*SNX14*	c.C459A	p.Y153X	Co-segregating

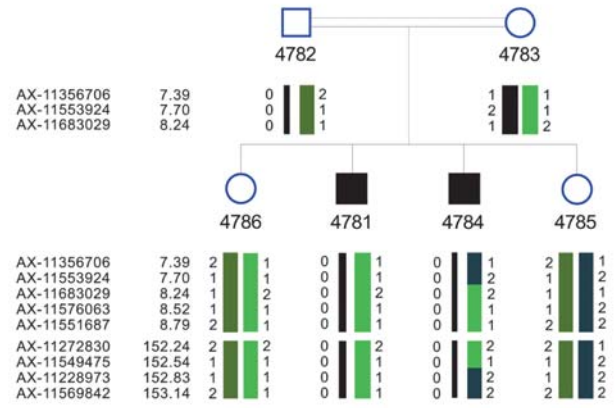
important for a disease is linked to a genetic marker is studied through the LOD score, which assesses how probably a given pedigree, where the disease and the marker are co-segregating, is due to the existence of linkage (with a given linkage value) or to chance. In most of our families, linkage analysis resulted in several large homozygous intervals. Even in family M244, in which we identified a single interval (62 Mbp) on chromosome 1, with a LOD score of 2.63, a total of 986 genes were present. Although it is lower than significant LOD score of 3, positive LOD scores favor the presence of linkage, and LOD score of 2.63 means that the likelihood of linkage occurring at this distance is more than 100 times greater that no linkage. In highly consanguineous families, individuals tend to share many large homozygous intervals that contain dozens or hundreds of genes. On the other hand, we obtained a non-significant LOD score because of the small sizes of our families (only two affected individuals). Thus, linkage analysis, by itself, may not be a suitable method for detecting monogenic defects in highly consanguineous families. Because of the

Table 4. Prediction of pathogenicity by different programs.

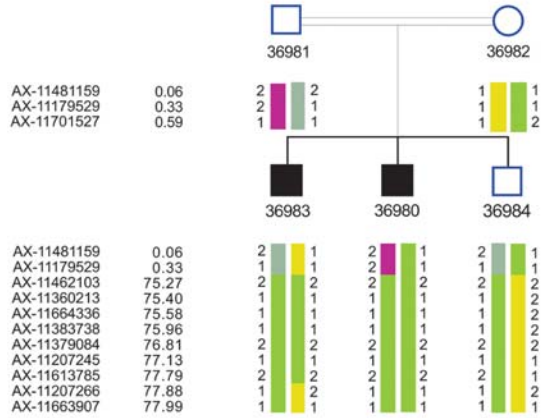
Sample	Variants	Grantham	Phylop	GERP	SIFT	PolyPhen2	MutationTaster
M244_4781	X:13794385G>A	64	5,502	5.76	tolerated(0.54)	probably damaging(0.997)	disease causing(0.9997)
	X:106144724del	NA	-1,006	-5.53	NA	NA	NA
M9000065_36340	9:136218979C>A	109	5,015	5.39	tolerated(0.49)	probably damaging(1.0)	disease causing(0.9998)
	9:134735990C>T	56	5,399	5.3	tolerated(0.45)	probably damaging(1.0)	disease causing(0.9997)
M9000097_36980	17:45904039G>A	101	1,313	5.62	damaging(0.00)	probably damaging(1.0)	disease causing(0.9931)
M9000111_37260	X:99934379A>G	64	4,232	4.87	NA	probably damaging(0.999)	disease causing(0.9982)
	22:50278793A>G	56	4,540	5.23	damaging(0.00)	probably damaging(1.0)	disease causing(0.9927)
	10:119799727G>A	83	5,592	5.5	damaging(0.03)	probably damaging(0.991)	NA
M9000121_37460	16:89178551G>A	58	4,953	4.49	damaging(0.01)	possibly damaging(0.941)	NA
	6:86277254G>T	NA	2,283	4.89	NA	NA	NA

Note: GERP and Phylop just are metrics indicating how the changed nucleotide is conserved in different species. Grantham is also a metrics showing how much the changed amino acid is different from its replacement. SIFT and PolyPhen are directly predicting programs which predict the pathogenesis of the variants.

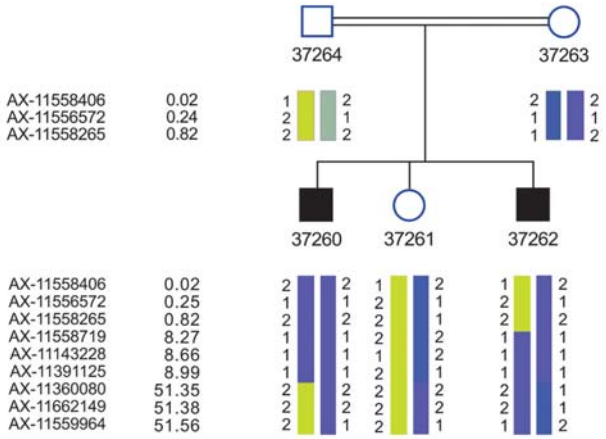
Family M244



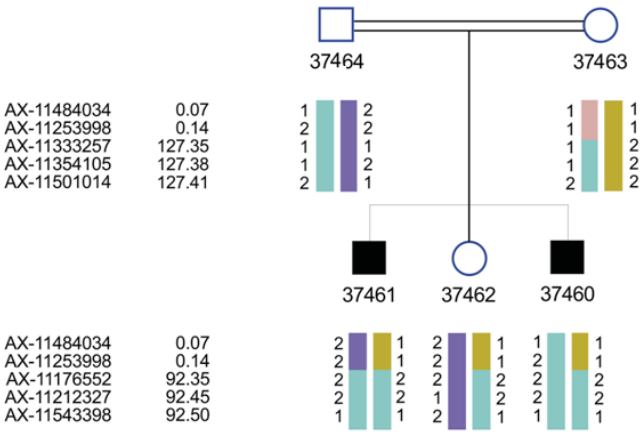
Family M9000097



Family M9000111



Family M9000121



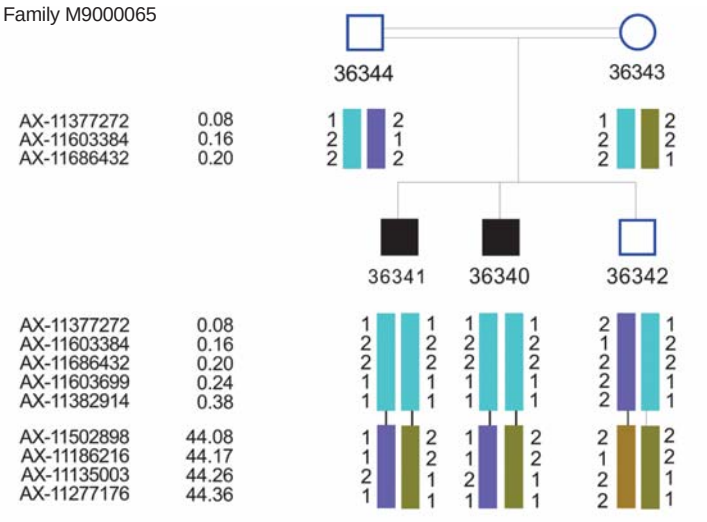
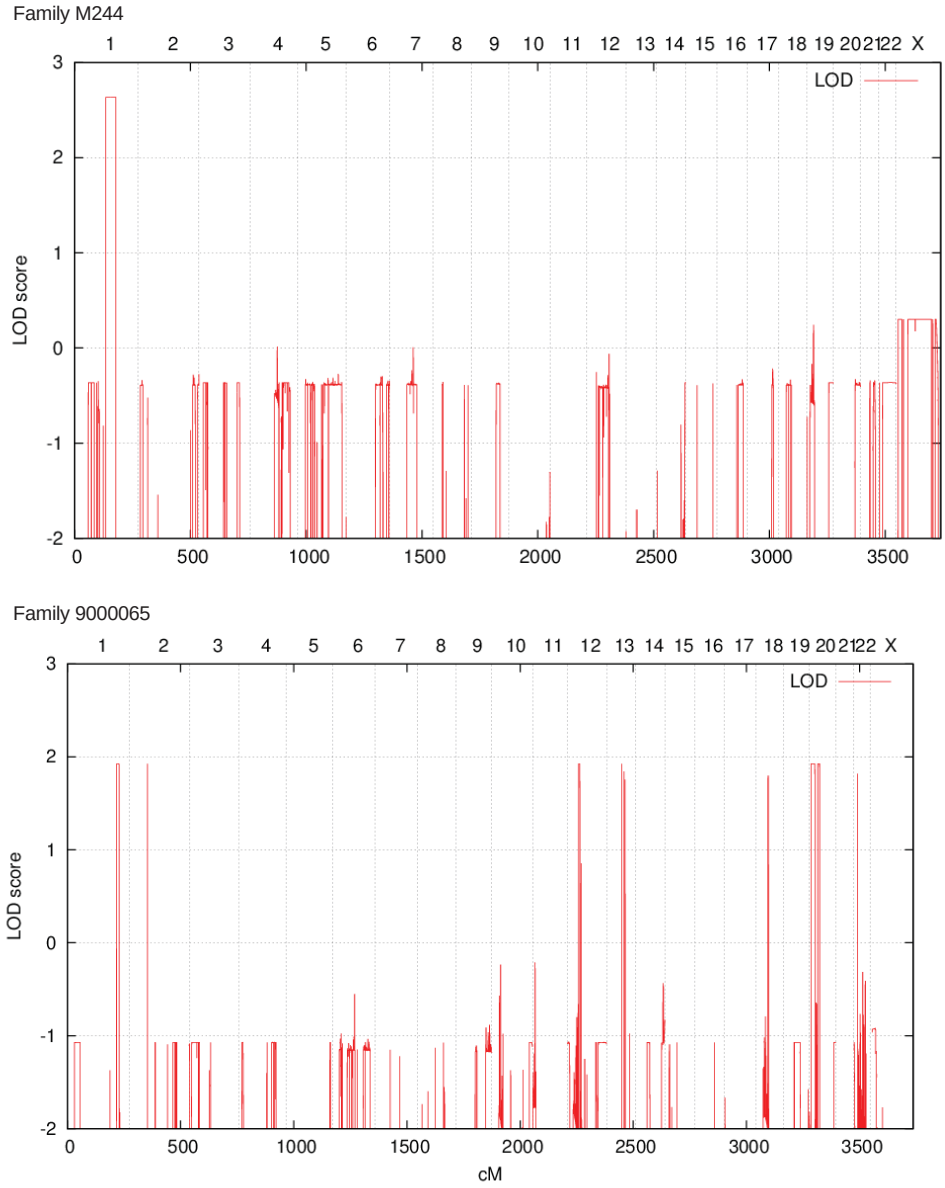


Figure 3. Outputs of homozygosity mapping for all five families across the whole genome.



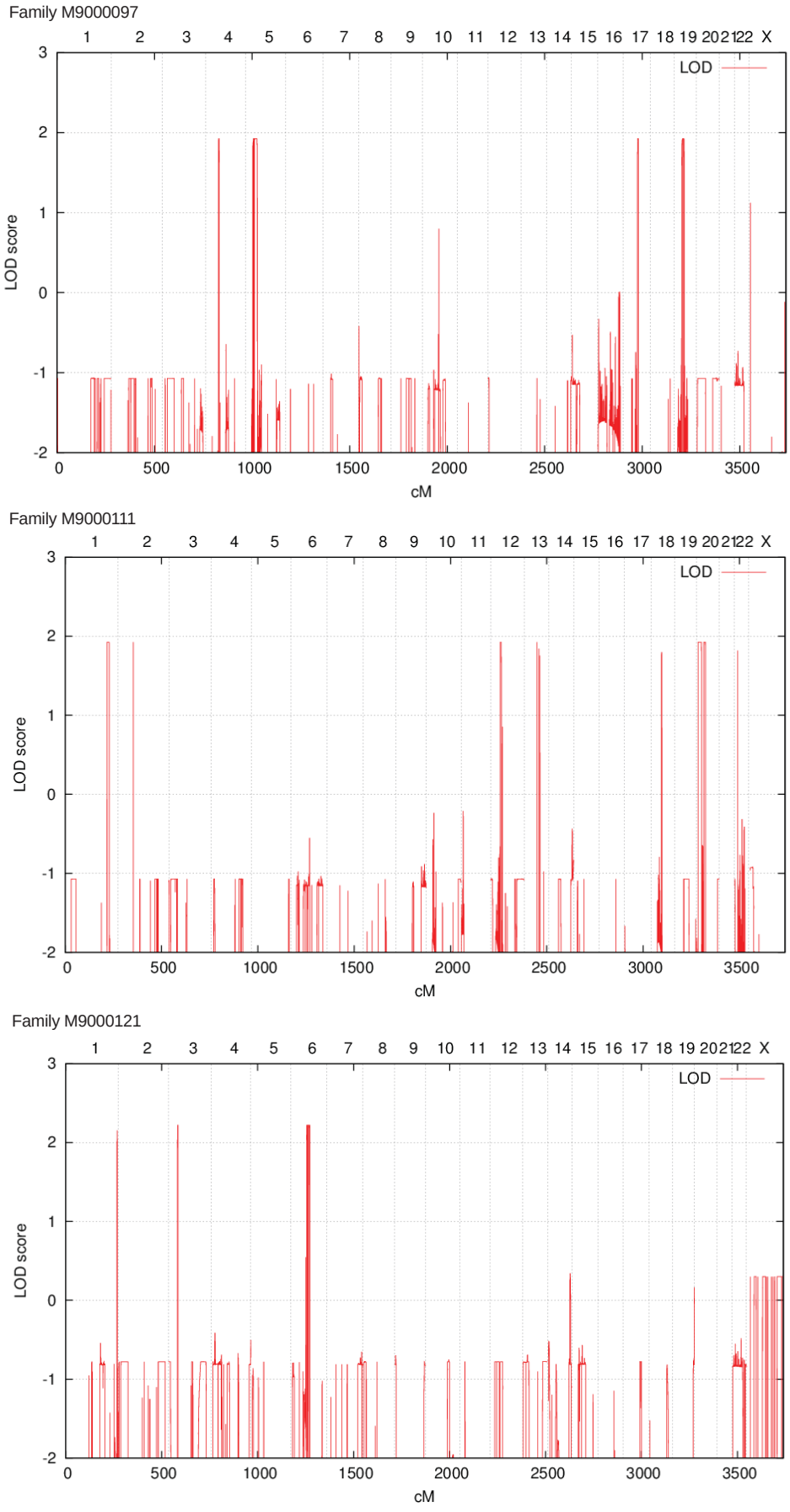


Figure 4. Merlin plots for all five families.

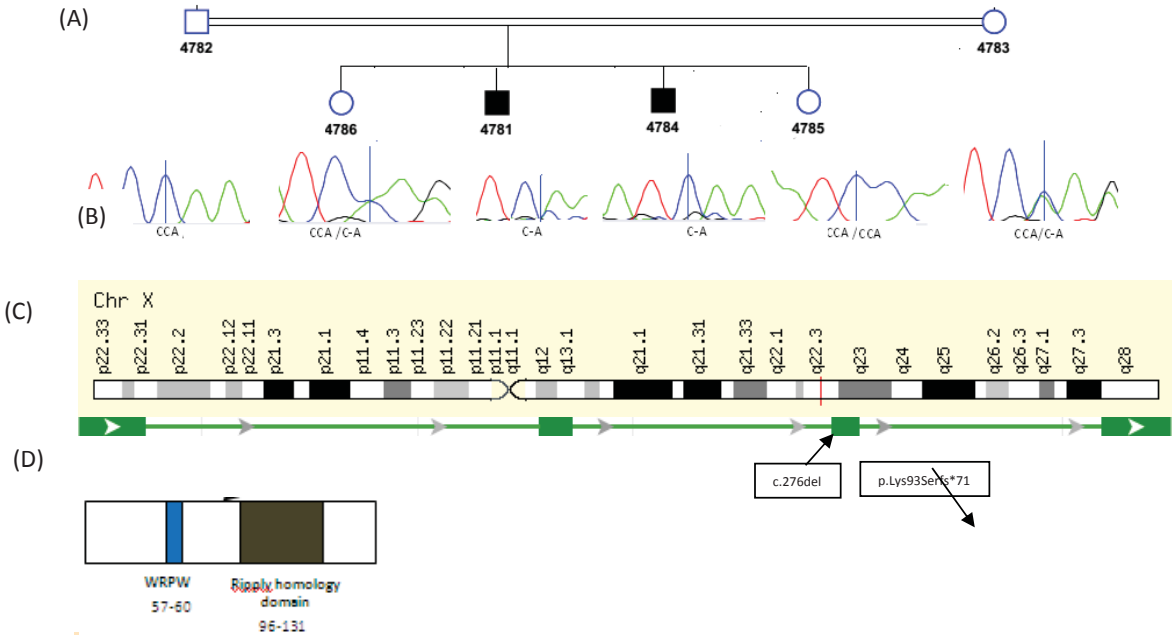


Figure 5. Identification of *RIPPLY1* mutation in two affected individuals in Family M244. **(A)** Pedigree of the family, the individuals with numbers under them are genotyped samples. **(B)** Sequence traces for two affected individuals, two normal sisters and their parents, show deletion in genomic DNA (two affected sons, hemizygous mutant; father, wild type; mother and elder sister, heterozygote; younger sister, normal homozygote). **(C)** Schematic diagram of the *RIPPLY1* genomic locus. **(D)** Schematic diagram of conserved domains (with numbers of their amino acids) of *RIPPLY1* indicating the protein approximate location of the frameshift mutation "p.Lys93Serfs*71", detected in this family. [http://www.uniprot.org/uniprot/Q0D2K3]

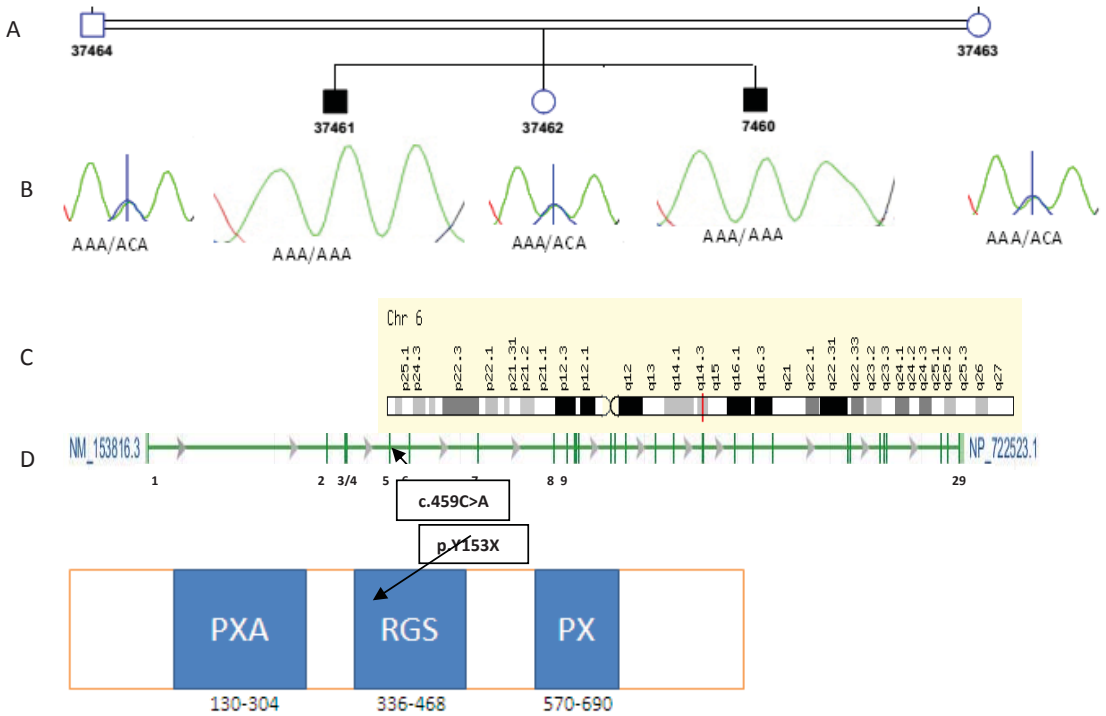


Figure 6. Identification of *SNX14* mutation in two affected individuals in Family 9000121. **(A)** Pedigree of the family, the individuals with numbers under them are genotyped samples. **(B)** Sequence traces for two affected individuals, a normal sister and their parents, show point mutations in genomic DNA (two affected individuals, homozygous mutant; the parents and normal sister, heterozygous). **(C)** Schematic diagram of the *SNX14* genomic locus. **(D)** Position of the nonsense variant on *SNX14* transcript (NM_153816) (top) and the *SNX14* protein (bottom) indicating the approximate location of the mutation detected in this family. The protein consists of three major domains: PXA domain, RGS domain and conserved PX phosphoinositide binding domain, indicating exact amino acid numbers of each domain, and approximate location of mutation in the gene and its protein. [http://www.uniprot.org/uniprot/Q9Y5W7].

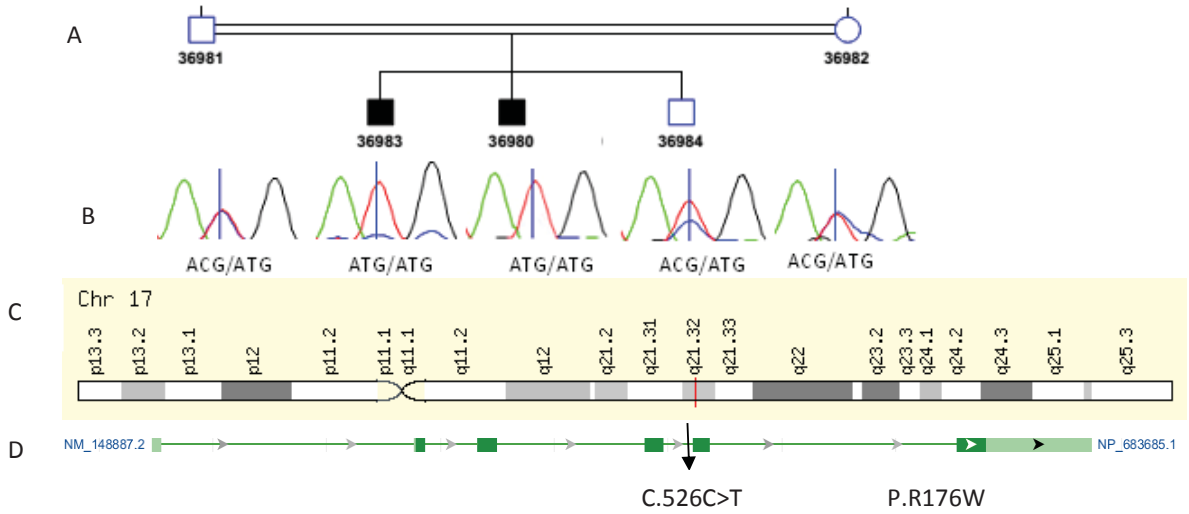


Figure 7. Identification of *MRPL10* mutation in two affected individuals in Family 9000097. **(A)** Pedigree of the family, the individuals with numbers under them are genotyped samples. **(B)** Sequence traces for two affected individuals, a normal brother and their parents, show point mutations in genomic DNA (two affected individuals, homozygous mutant; the parents and normal brother, heterozygous). **(C)** Schematic diagram of the *RIPPLY1* genomic locus. **(D)** Position of the missense variant (black arrow) on *MRPL10* transcript (NM_148887).

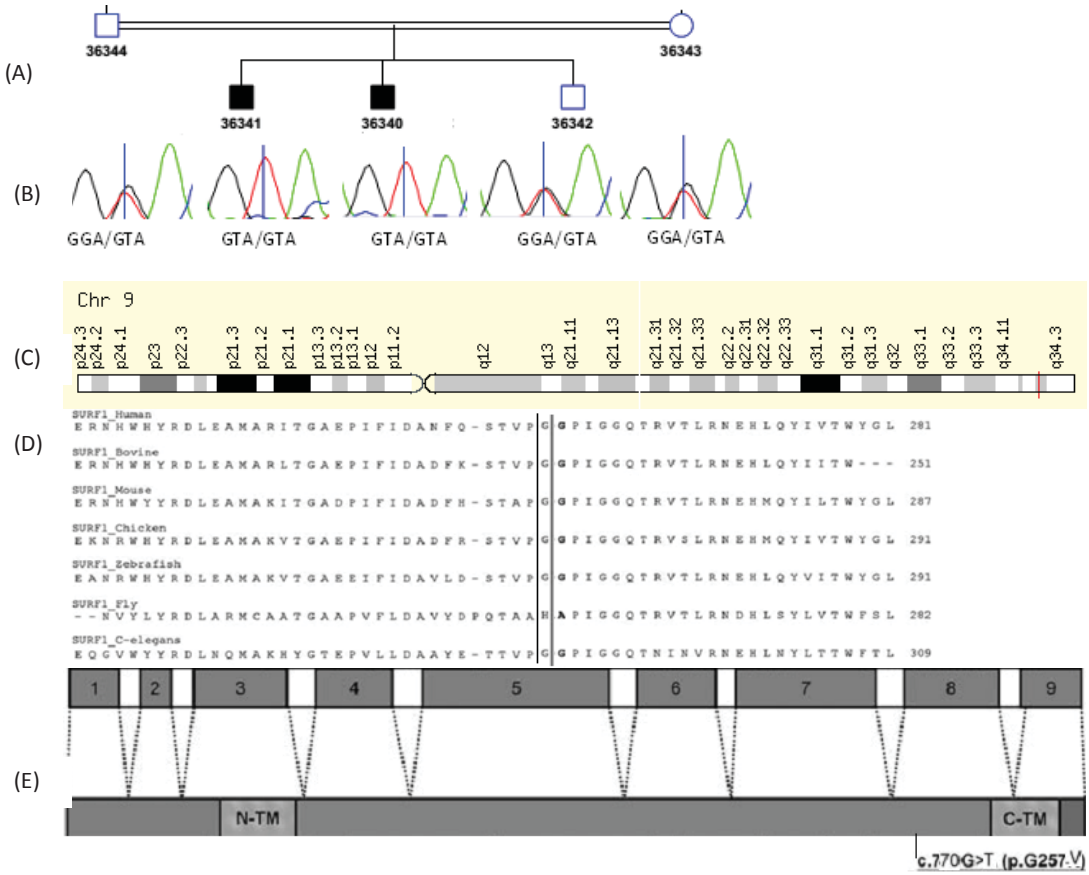


Figure 8. Identification of *SURF1* mutation in two affected individuals, Family 9000065. **(A)** Pedigree of the family, the individuals with numbers under them are genotyped samples. **(B)** Sequence traces for two affected individuals, a normal brother and their parents, show point mutations in genomic DNA (two affected individuals, homozygous mutant; the parents and normal brother, heterozygous). **(C)** Schematic diagram of the *SURF1* genomic locus. **(D)** The *SURF1* protein alignment in *Caenorhabditis elegans*, fruitfly, zebrafish, chicken, mouse, bovine, and human. The mutation c.770G>T (p.G257V) alters glycine at position 257, which is evolutionarily conserved from zebrafish to human. The proteins were aligned using ClustalW2 at the European Bioinformatics Institute, European Molecular Biology Laboratory website (http://www.ebi.ac.uk/Tools/clustalw2/index.html). **(E)** The alignment of the mutation of *SURF1*, described in our report, on the DNA and the encoded protein. N-TM indicates the N-terminal transmembrane domain of SURF1 protein extending from protein position 61 to 79, C-TM indicates C-terminal transmembrane domain extending from protein position 274 to 290. The coding sequence includes: exon 1 (1 to 55), exon 2 (56 to 107), exon 3 (108 to 240), exon 4 (241 to 323), exon 5 (324 to 515), exon 6 (516 to 588), exon 7 (589 to 751), exon 8 (752 to 835), and exon 9 (836 to 903). This figure is provided from paper of Lee, et al., 2008 (ref no. 27).

considerable hindrance of these large gene lists and the cost and time-consuming traditional Sanger sequencing of these numerous genes, WES was applied to one patient from each family. WES analyzes the total exons of all genes, and so is a suitable method to evaluate the total exons of the whole genome.

Identification and confirmation of potentially pathogenic variants located in linkage intervals corroborates the validity of our results. In four families, only one variant identified in WES co-segregated with the disease, and is therefore considered a probable causal variant in that particular family. The four variants identified by WES include three novel candidate genes *RIPPLY1*, *SNX14*, and *MRPL10*, and a new mutation in a known gene *SURF1*; all located in homozygous regions. Two (*RIPPLY1* and *SNX14*) are housekeeping genes involved in basic cellular functions: transcription and translation regulation and intracellular trafficking, and two (*MRPL10* and *SURF1*) encode mitochondrial proteins. In family M9000111, where none of the three variants identified in WES co-segregated with the disease in the family, the causative defect might be located in regions other than exons, such as introns or extragenic regulatory sequences, and so whole-genome sequencing (WGS), if possible, would be considered as the next step. A nonsense mutation in *SNX14* and frameshift in *RIPPLY1* genes can be confidently predicted to destroy the protein and are very likely to be pathogenic. On the other hand, a pathogenic effect of the missense mutations identified in *MRPL10* and *SURF1* was predicted using different prediction programs. Likewise, at the beginning of our study, we had considered autosomal recessive to be the most likely mode of inheritance for all of our families, because of consanguineous marriage and more than one affected sibling; however, an X-linked pattern could not have been excluded, especially because of the male sex of our affected individuals. In one of our families (M244), after filtration, WES revealed only one variation on the X chromosome, which co-segregated with the disease phenotype in the family, confirming an X-linked inheritance of ID associated with truncal ataxia in a consanguineous family. However, this is not unexpected, as shown in a study by Pouya and colleagues aiming to estimate the frequency of Fragile X syndrome (FXS) in intellectually disabled patients from Iran. They showed that 3.4% of FXS cases were present in consanguineous families, therefore indicating that, even in families with consanguineous parents, FXS has to be ruled out before assuming that familial mental retardation (MR) is the result of autosomal recessive gene defects.[10] Here, we separately discuss each of these four variants which we identified:

RIPPLY1

The *RIPPLY1* gene encodes Ripply Transcriptional Repressor 1 (RIPPLY1) protein with 151 amino acids. It is a transcription regulator, and belongs to the RIPPLY family which is thought to play a role in somitogenesis, development, and transcriptional repression.[11] *RIPPLY1* c.276del mutation, which is the only variation identified in the M-244 family, will lead to "p.Lys93Serfs*71" mutation. This frameshift mutation is predicted to alter the amino acid sequences beginning from Lysine at position 93, introducing a later stop codon (12 amino acids later). Thus, this variant loses the Ripply homology domain (96–131 amino acids), a conserved domain of 36 amino acids found in other Ripply proteins (Figure 5), which is required for transcriptional repression.[11] Further supporting proof for the pathogenicity of this variant is provided by Kawamura, *et al*.[12] Searching for sequences similar to that of zebrafish Ripply1, they identified DSCR6 Down Syndrome Critical Region Protein 6 homolog, which they called RIPPLY3. Shibuya, *et al.* mapped the *DSCR6* gene to chromosome 21q22.2.[13] Furthermore, Girirajan, *et al*. have recently shown that copy-number variants (CNVs) within this region were significantly enriched in cases with developmental delay, intellectual disability, and other neurodevelopmental disorders (including autism spectrum disorder [ASDS]), compared to controls; however, the pathogenicity needs to be confirmed by other reports and functional studies.[14]

SNX14

SNX14 gene encodes a member of the sorting nexin family. Members of this family contain a phox (PX) domain, which is a phosphoinositide binding domain, regulate adhesion to organelle membranes of the secretory and endocytic system, and play important roles in membrane trafficking, membrane remodeling, organelle motility, cell signaling, and protein sorting.[15] The encoded protein also contains a regulator of the G protein signaling (RGS) domain, acts as GTPase activating protein for G alpha subunits of heterotrimeric G proteins.[16,17] The variant c.459C>A is located near (the 3rd nucleotide from) the end of the 5th exon of the *SNX14* gene, which has 29 exons (NM_153816; isoform a), and changes tyrosine amino acid (TAC) at position 153 to a premature stop codon (TAA) (Figure 6). However, genes containing premature termination codons seldom cause production of the truncated protein as might be predicted, because these proteins will usually be degraded because of the nonsense-mediated decay (NMD) mechanism, which prevents from production of harmful truncated proteins. Thus, the usual result of a nonsense mutation is to prevent any expression of the gene. However, NMD does not apply to premature stop codons that are in the last exon of a gene or less than about 50 nucleotides upstream of the last splice junction. Considering the location of our variant, the mutated protein is predicted to be degraded, although residual amount of protein might still be produced. In addition to the absence of the protein, truncated proteins can also be potentially pathogenic. In this case, SNX14 truncated protein ends at amino acid position 153 (full protein has 946 amino acids), and therefore lacks PX and also RGS domains (Figure 6), which are in turn involved in intracellular trafficking and act as GTPase activating proteins for G alpha subunits of heterotrimeric G proteins. Further supporting evidence for the pathogenicity of this variant comes from studies that showed the SNX family members as disease-related genes in patients with microcephaly, and Down syndrome.[18,19] Furthermore, dysregulation of *SNX14* has been suggested to occur in bipolar disorder and in 6q14 microdeletion syndrome as one of four possible candidate genes for intellectual disability.[20,21] More evidence for the pathogenicity of SNX14 comes from a recent study by Huang, *et al.* on knockdown of *Snx14* in mouse, which dramatically reduced neuronal excitability and synaptic transmission. They demonstrated an important neuronal role in intrinsic neuronal excitability and synaptic transmission, and confirmed that SNX14 is predominantly expressed in the brain.[22]

MRPL10

MRPL10 gene encodes a 39S mitochondrial ribosomal protein L10 with 261 amino acids which is involved in protein synthesis within the mitochondrion. The deduced 261-amino acid human MRPL10 protein contains a 28-amino acid N-terminal mitochondrial import signal. Mitochondria have their own translation

system for production of 13 proteins essential for oxidative phosphorylation. MRPL10 is one of more than 70 protein components of mitochondrial ribosomes that are encoded by the nuclear genome.[23] Defective neuronal differentiation in the presence of extra human chromosome 21, an *in vitro* model of TT2F mouse embryonic stem cells, identified MRPL10 as one of the primarily post-transcriptional and translational modifiers in the early neuronal stem cell stage.[24] MRPL10 is the major acetylated protein in the mitochondrial ribosome and is also a substrate of the NAD(+)-dependent deacetylase, SIRT3, which regulates mitochondrial protein synthesis by deacetylation of the mitochondrial ribosome.[25] Sirtuins have a role in regulation of transcription and apoptosis leading to a substantial interest in inhibitors of these enzymes as possible antineoplastic agents. We found a missense variant in *MRPL10* (c.526C>T; p.R176W), which is within the homozygous region. It changes electrically charged (positive) arginine to nonpolar (hydrophobic) tryptophan in position 176 in the 4th exon (it has 5 exons) (Figure 7).

SURF1

SURF1 gene encodes a protein localized to the inner mitochondrial membrane which is thought to be involved in the biogenesis of the cytochrome C oxidase complex.[26] SURF1 contains a central part and two transmembrane domains, one is in the N-terminus and the other in the C-terminus, and are believed to be essential for the function of the protein (Figure 8). Defects in this gene are a cause of Leigh syndrome (LS), a severe early-onset progressive neurodegenerative disorder with subacute neurodegenerative encephalopathy. Typical LS is a progressive neurological disease with signs and symptoms of brainstem or basal ganglia disease and raised lactate levels in blood or cerebrospinal fluid, associated with MRI features, postmortem neuropathological changes or neuropathology in a similarly affected sibling. However, the clinical presentation can be highly variable, and some patients with LS can have an atypical course without lactic acidosis or the unique brain lesions. It is genetically heterogeneous, but mutations in the complex IV assembly genes, particularly *SURF1*, are an important cause.[27] SURF1-associated LS is panethnic, with mutations in *SURF1* gene being reported in Caucasian, Asian, and Hispanic patients. Although mutations in *SURF1* have been mainly associated with typical LS, a significant number of patients can exhibit atypical LS and there is no definite genotype-phenotype correlation; however, frameshift mutations resulting in protein truncation closer to the C-terminus may carry a better prognosis. Lee and colleagues reviewed the previously reported *SURF1* mutations and observed a clustering of mutations in exon 8 of *SURF1*, suggesting a vital function for this region.[28] Our patients were 15 and 19 years old; positive findings were severe ID, ataxia, short stature, absence of language, gait ataxia and cataract, and they had no dysmorphism, hypertrichosis, microcephaly, failure to thrive (FTT), swallowing difficulty, seizures, and no evidence of elevated blood sugar. They were not available for further assessment including performing brain MRI, or investigation of lactate or pyruvate, however, with only this information, they could be considered to be atypical Leigh syndrome which is mild. Based on a previous study reported by Najmabadi, *et al*. (2011), another missense mutation in this gene caused autosomal recessive intellectual disability without any other manifestations of Leigh syndrome.[5] The novel missense variant, c.770G>T (p.G257V), changes glycine residue that is evolutionarily conserved (Figure 8) from zebrafish to humans.[28] The computer-based algorithms SIFT and PolyPhen predict p.G257V to be deleterious. The two patients were homozygous for variant p.G257V, they had two normal siblings, one sister and one brother, and their parents were heterozygous, thus the variant co-segregated with the disease phenotype (Figure 8). Recently, a missense mutation in exactly this position, which alters amino acid glycine (p.G257R), has been reported to be a probable pathogenic change (compound heterozygous) with another mutation in the other allele).[28] Therefore, the collective evidence indicates that the c.770G>T (p.G257V) variant is predicted to be a deleterious mutation. Our findings support the view that missense mutations in the *SURF1* gene may correlate with a milder course, with longer survival.

In conclusion, the study of linkage analysis and whole exome sequencing in one affected individual from each of five Iranian families with ID identified three families with novel mutations in the candidate genes *RIPPLY1, MRPL10* and *SNX14*. The proteins encoded by these novel candidate genes have important roles in the regulation of transcription, translation, intracellular trafficking, and mitochondrial metabolism. In a fourth family, a novel mutation was found in the known gene *SURF1* co-segregating with the disease in the family. The fifth family was not found to have any variant co-segregated with the phenotype, and so further studies are needed to identify the underlying genetic cause in this family. The phenotype studied in one of our families was in agreement with the reported phenotype of known mutations in the same gene (*SURF1*), thus in this case functional studies are not necessary to conclude that this is a pathogenic mutation. In the other cases, it may still be that not all of these alterations will prove to be causative, even with comprehensive checking of our data and strict filtering versus all known neutral and pathogenic sequence variants. Functional studies will be required to exclude the possibility of uncommon polymorphisms not relevant to ID and ataxia, especially for missense mutations.

This study illustrates that exome sequencing of a small number of affected family members is a powerful new advanced technique for significantly decreasing the number of candidate genes in heterogenic disorders such as ID, and may even specifically identify the gene(s) responsible. However, it should be noted that each homozygous nonsynonymous variation is not necessarily causative and should be interpreted with caution, more specifically for a missense change. These findings, after being confirmed, will also provide the framework for potential carrier detection and prenatal diagnosis (PND) of this form of ID, associated with cerebellar ataxia.

Conflict of interest

We attest that we have herein disclosed any and all financial or other relationships which could be considered to be a conflict of interest and that all sources of financial support for this study have been disclosed and are indicated in the Acknowledgements.

Acknowledgments

We gratefully thank all family members and affected individuals who participated in this study. This article is extract of PhD student project. This research was supported by the Max-Planck Society, EU FP 7 project GENCODYS, grant no 241995; and Iranian National Science Foundation, INSF, grant no 93/ Page /37653.

References

1. Firth H, Carter N. Database aids for the evaluation of mental retardation. In: Knight SJL, ed. *Genetics of Mental Retardation*. Oxford: Karger Publishers; 2010: 67-76.
2. Ropers HH. Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet*. 2010; 11: 161-87.
3. Curry CJ, Stevenson RE, Aughton D, Byrne J, Carey JC, Cassidy S, et al. Evaluation of mental retardation: recommendations of a consensus conference. *Am J Med Genet*. 1997; 72(4): 468-77.
4. Najmabadi H, Motazacker MM, Garshasbi M, Kahrizi K, Tzschach A, Chen W, et al. Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Hum Genet*. 2007; 121(1): 43-8.
5. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*. 2011; 478(7367): 57-63.
6. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010; 42(9): 790-3.
7. Topper S, Ober C, Das S. Exome sequencing and the genetics of intellectual disability. *Clin Genet*. 2011; 80(2): 117-26.
8. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. 2001; 30: 97–101
9. Hu H, Wienker TF, Musante L. Integrated sequence analysis pipeline provides one-stop solution for identifying disease-causing mutations. *Hum Mutat*. 2014; 35(12): 1427-35.
10. Pouya AR, Abedini SS, Mansoorian N, Behjati F, Nikzat N, Mohseni M, et al. Fragile X syndrome screening of families with consanguineous and non-consanguineous parents in the Iranian population. *Eur J Med Genet*. 2009; 52(4): 170-3.
11. Hitachi K, Danno H, Tazumi S, Aihara Y, Uchiyama H, Okabayashi K, et al. The Xenopus Bowline/Ripply family proteins negatively regulate the transcriptional activity of T-box transcription factors. *Int J Dev Biol*. 2009; 53: 631-9.
12. Kawamura A, Koshida S, Hijikata H, Ohbayashi A, Kondoh H, Takada S. Groucho-associated transcriptional repressor ripply1 is required for proper transition from the presomitic mesoderm to somites. *Dev Cell*. 2005; 9(6): 735-44.
13. Shibuya K, Kudoh J, Minoshima S, Kawasaki K, Asakawa S, Shimizu N. Isolation of two novel genes, DSCR5 and DSCR6, from Down syndrome critical region on human chromosome 21q22. 2. *Biochem Biophys Res Commun*. 2000; 271(3): 693-8.
14. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *New Engl J Med*. 2012; 367(14): 1321-31.
15. 15. Tea SD, Sale RD, Collins BM. Insights into the PX (phox-homology) domain and SNX (sorting nexin) protein families: structures, functions and roles in disease. *Biochem J*. 2012; 441: 39-59
16. Carroll P, Renoncourt Y, Gayet O, De Bovis B, Alonso S. Sorting nexin-14, a gene expressed in motoneurons trapped by an in vitro preselection method. *Dev Dynam*. 2001; 221(4): 431-42.
17. Worby CA, Dixon JE. Sorting out the cellular functions of sorting nexins. *Nature Rev Mol Cell Biol*. 2002; 3(12): 919-31.
18. Vervoort VS, Viljoen D, Smart R, Suthers G, DuPont BR, Abbott A, et al. Sorting nexin 3 (SNX3) is disrupted in a patient with a translocation t(6;13)(q21;q12) and microcephaly, microphthalmia, ectrodactyly, prognathism (MMEP) phenotype. *J Med Genet*. 2002; 39(12): 863-9.
19. Wang X, Zhao Y, Zhang X, Badie H, Zhou Y, Mu Y, et al. Loss of sorting nexin 27 contributes to excitatory synaptic dysfunction by modulating glutamate receptor recycling in Down's syndrome. *Nat Med*. 2013; 19(4): 473-80.
20. Seelan RS, Khalyfa A, Lakshmanan J, Casanova MF, Parthasarathy RN. Deciphering the lithium transcriptome: microarray profiling of lithium-modulated gene expression in human neuronal cells. *Neuroscience*. 2008; 151(4): 1184-97.
21. Becker K, Di Donato N, Holder-Espinasse M, Andrieux J, Cuisset JM, Vallée L, et al. *De novo* microdeletions of chromosome 6q14. 1-q14. 3 and 6q12. 1-q14. 1 in two patients with intellectual disability-further delineation of the 6q14 microdeletion syndrome and review of the literature. *Eur J Med Genet*. 2012; 55(8): 490-7.
22. Huang HS, Yoon BJ, Brooks S, Bakal R, Berrios J, Larsen RS, et al. Snx14 regulates neuronal excitability, promotes synaptic transmission, and is imprinted in the brain of mice. *PLoS One*. 2014; 9(5): e98383.
23. Kenmochi N, Suzuki T, Uechi T, Magoori M, Kuniba M, Higa S, et al. The human mitochondrial ribosomal protein genes: mapping of 54 genes to the chromosomes and implications for human disorders. *Genomics*. 2001; 77(1): 65-70.
24. Wang CC, Kadota M, Nishigaki R, Kazuki Y, Shirayoshi Y, Rogers MS, et al. Molecular hierarchy in neurons differentiated from mouse ES cells containing a single human chromosome 21. *Biochem Biophys Res Commun*. 2004; 314(2): 335-50.
25. Yang Y, Cimen H, Han M-J, Shi T, Deng J-H, Koc H, et al. NAD+-dependent deacetylase SIRT3 regulates mitochondrial protein synthesis by deacetylation of the ribosomal protein MRPL10. *J Biol Chem*. 2010; 285(10): 7417-29.
26. Zhu Z, Yao J, Johns T, Fu K, De Bie I, Macmillan C, et al. SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. *Nature Genet*. 1998; 20(4): 337-43.
27. Yao J, Shoubridge EA. Expression and functional analysis of SURF1 in Leigh syndrome patients with cytochrome C oxidase deficiency. *Hum Mol Genet*. 1999; 8(13): 2541-9.
28. Lee IC, El-Hattab AW, Wang J, Li FY, Weng SW, Craigen WJ, et al. SURF1□associated leigh syndrome: a case series and novel mutations. *Hum Mutat*. 2012; 33(8): 1192-200.