Inherited non-alcoholic fatty liver disease and dyslipidemia due to monoallelic *ABHD5* mutations

Leila Youssefian, Hassan Vahidnezhad, Amir Hossein Saeidian, Sara Pajouhanfar, Soheila Sotoudeh, Parvin Mansouri, Davoud Amirkashani, Sirous Zeinali, Michael A. Levine, Ketty Peris, Roberto Colombo, Jouni Uitto

Table of contents

Supplementary materials and methods

Patients Data

This study was approved by the Institutional Review Boards of the University Hospital, Rome, and the Pasteur Institute of Iran, Tehran, and all adult subjects and parents of children gave their written informed consent to participate in the study. A total number of 190 families of Iranian origin and one family of Italian ancestry were enrolled. Six Iranian families with Chanarin-Dorfman syndrome and also affected by the familial form of NAFLD and/or dyslipidemia, were diagnosed in various medical centers in Iran. These families were personally examined by two dermatologists (SS and PM) and by two medical geneticists (LY and HV). Medical records concerning the members of the family of Italian ancestry were collected by several general practitioners and gastroenterologists in the US and Italy where the subjects live.

Ultrasound

A diagnostic ultrasonography and/or transient elastography using Fibroscan 502 (5 MHZ; EchoSene, Paris, France) was performed.[1] Accepted criteria for diagnosis of the nonalcoholic fatty liver disease were used: elevated echogenicity of the liver compared with kidney parenchyma, reduced visualization of intra-hepatic blood vessels, and posterior beam attenuation.

Measurements

We measured height, weight, and fasting levels of plasma cholesterol and triglycerides. Serum alanine and aspartate amino-transferase and glutamyltransferase levels were assessed in the

clinical chemistry laboratory of the Tehran University of Medical Science or in US and Italian certified laboratories.

Whole Exome Sequencing

Genomic DNA was isolated from the peripheral blood of the proband (F1-III:3) and his affected father (F1-II:1) using the QIAamp DNA Blood Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA concentration was measured using the Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA) High Sensitivity Assay, and sample purity was checked by the Nanodrop (Thermo Fisher Scientific, Waltham, MA) by OD 260/280 ratio.

For each sample, 100 ng of DNA were used to prepare the library. DNA fragments obtained by sonication on a LE220 instrument (Covaris, Woburn, MA) were of 350 base pairs on average. The NimbleGen SeqCap EZ v3.0 (Roche NimbleGen Inc, Basel, Switzerland) was used for exome enrichment, and pre-capture multiplexing was applied following the manufacturer's protocol version 4.2. Libraries were assessed using Bioanalyzer DNA High Sensitivity chips (Agilent Technologies, Santa Clara, CA) and quantified by quantitative polymerase chain reaction using the Kapa Library Quantification Illumina/ABI Prism Kit protocol (KAPA Biosystems, Roche, Basel, Switzerland).

Whole exome sequencing (WES) was performed on the HiSeq 2500 (Illumina, La Jolla, CA) with the paired-end 100 protocol. Raw image files were analyzed and converted to base calls by real-time analysis using the default settings recommended by the manufacturer. Real time analysis output base call files (*.bcl) were converted to FASTQ files with consensus assessment of sequence and variation (CASAVA pipeline, version 1.8, Illumina). Processing of all data followed Genome Analysis Tool Kit instructions (GATK; Broad Institute, Cambridge,

MA). Specifically, Burrows-Wheeler Aligner software (BWA version 0.7.15) was used to map reads in paired-end mode to human reference genome hg19. The BAM files obtained were used for the subsequent steps in the analysis pipeline. Picard (version 2.5.0; SourceForge, Mountain View, CA) was used to sort mapped reads into coordinate order, ensure that all mate pair information was properly updated, and mark duplicate reads and low quality reads that were unpaired or unmapped, or that failed quality check. The GATK tool was also used for recalibration of the base quality score and for insertion/deletion (INDEL) realignment, with Picard used again for marking duplicates and obtain analysis-ready BAM files.

The average coverage of the targeted bases was 61.2×, with <0.01% of the target region not covered and 95.8% of the target region with ≥10× coverage. Exonic SNPs and indels identified were > 19,000. After filtering for genomic variations present in publicly-accessible databases, ExAC, gnomAD and GME with frequency >1%, we first focused on genes known to be involved in disorders and syndromes whose clinical features include fatty liver disease. Thereafter, we looked for nonsense, missense, frameshift and putative splicing mutations in any of the candidate gene. One monoallelic *ABHD5* variant was present in the proband, his father, and his sister. The presence of the same nonsense mutation was confirmed in the trio and identified in further members of their family by PCR amplification (primers were designed using the Primer3 program [http://biotools.umassmed.edu/bioapps/primer3_www.cgi]) and bidirectional Sanger sequencing.

NGS panel and data analysis

DNA was extracted from peripheral blood samples by a kit (QIAamp Blood Maxi Kit; Qiagen, Valencia, CA) or by salting-out method. DNA concentration was measured using a Qubit 2.0

fluorometer (Life Technologies, Carlsbad, CA). Target enrichment was performed using the TruSeq Custom Amplicon kit (Illumina, San Diego). DesignStudio (Illumina) was used for library design. All coding exons, at least 20 bp of the intron at each intron-exon boundary, and up to 50 bp of 5'- and 3′-untranslated regions were targeted. The designed library contained 38 genes (*ABCA12, ABHD5, AGPS, ALDH3A2, ALOX12B, ALOXE3, AP1S1, ARSE, CERS3, CLDN1, CYP4F22, EBP, ELOVL4, GJB2, GJB3, GJB4, GJB6, KRT1, KRT10, KRT2, KRT9, LIPN, LOR, NIPAL4, PEX7, PHYH, PNPLA1, PNPLA2, POMP, SLC27A4, SNAP29, SPINK5, ST14, STS, TGM1, TGM5, VPS33B, and ZMPSTE24*) divided into 351 targets covered by 558 amplicon probes, which were designed to cover 99% of targeted bases. A total of 93.2% of the reads were aligned to the human genome, with the mean coverage of the target region being 493×. In addition, only 0.4% of bases of the target region were not covered by any sequence read, indicating that 99.6% of all target region bases were sequenced at least once. Genomic DNA from 190 probands, each representative of a distinct extended family, and two Illumina controls were multiplexed using dual indexing with 12 primary indexes and eight secondary indexes. Dual-indexed samples were normalized to be equimolar and were pooled together following manufacturer's recommendations. The pool was sequenced on a single MiSeq flow-cell (Illumina). Reads were paired-end at 2 × 225 nucleotides with dual indexes, and a total of 8.49 giga-base pairs were generated. All samples passed quality control. Reads were aligned to the human genome version hg19[2] using the BaseSpace MiSeq Reporterworkflow version 2.4 TruSeq Amplicon App by Illumina. This software implements the banded Smith-Waterman algorithm in the targeted regions (specified in the Illumina manifest file). Picard Tools ReorderSam was used to properly format each alignment file (https://broadinstitute.github.io/picard/). Genome Analysis Tool Kit tools

RealignerTargertCreator, IndelRealigner, BaseRecalibrator, and PrintReads were used to pre-process the alignments, utilizing local alignments and base quality score distributions to optimize alignment accuracy. Variants were called with GATK HaplotypeCaller.

PCR and Sanger sequencing

PCR was performed using Taq polymerase (Qiagen) according to the manufacturer's instructions. Amplification of the *ABHD5* gene was performed with eight pairs of newly designed primers (sequences available upon request), spanning all seven exons and the flanking intronic sequences. The PCR products were bidirectionally sequenced using an automated sequencer (3730; Applied Biosystems, Foster City, CA). The mutation positions are reported in reference to NM_016006 (cDNA) and NP_001342115 (protein).

Lipid droplets imaging and quantification

Fresh EDTA-treated peripheral blood samples were collected after overnight fast, incubated at 37 °C for 12 h without and with 200 μM oleic acid (OA; Sigma-Aldrich, St Louis, MO, USA), and centrifuged at 3300 g for 10 min. OA (Sigma-Aldrich) stock solution was obtained by dissolving OA in a 2.1 mM solution of bovine serum albumin free of fatty acids (Sigma-Aldrich) previously prepared in 0.1 M Tris (pH 8). After OA was completely dissolved, the solution was heated to 37°C, then cooled, filtered (0.20 μm pore size) and stored at 4°C.

Buffy coats were carefully collected by gentle pipette suction, immediately smeared onto slide glasses, dried completely, and fixed with Biofix (Bio-Optica, Milan, Italy). Cells were sequentially washed with Dulbecco's phosphate-buffered saline (Thermo-Fisher, Carlsbad, CA, USA) and stained with Oil red O (ORO; Sigma-Aldrich) and 2-(4-amidinophenyl)-6-

indolecarbamidine (DAPI; Sigma-Aldrich) at room temperature as described by Tavian and Colombo.[3]

Stainings were evaluated on a Leica DM5500B microscope equipped with a 100x Fluortar oil-immersion objective. Fluorescence images were captured using a Leica DFC420 digital camera and a Leica Application Suite (LAS) software. Set-up (exposure time, saturation, gamma and gain, respectively) was as follows: ORO, 4 s, 1.2, 1.43 and 16; DAPI, 0.03–0.05 s, 1.1, 1.52 and 16. Lipid droplets (LDs) present in leucocytes were analyzed using ImageJ 1.51 software (originally developed by W.S. Rasband at NIH, Bethesda, MD, USA) that allows to isolate components having the same wavelength and to evaluate and quantify several parameters like area, numbers of selected subcellular units (LDs), and pixels per cell. The threshold value for LD area in order to discard fluorescent emissions due to impurity was set to 15 inche^2.

Statistical analysis

The statistical analysis of quantitative data on LDs identified in leukocytes by image analysis was performed with the SPSS v. 19 package (SPSS, Chicago, IL, USA) with a level of significance of 5%.

Fig. S1. **Filter-based annotation of all documented genetic variants of *ABHD5* in ExAC, gnomAD, and GME databases for estimation of NAFLD due to mono-allelic mutations.** In order to estimate the prevalence of the NAFLD-causing *ABHD5* deleterious alleles in the general population, we used ANNOVAR software tool to functionally annotate all 1,149 genetic variants of *ABHD5* reported in 188,951healthy individuals in these databases. Exonic and splicing variants were considered for further analysis, while intronic and synonymous variants were removed. Next, deleteriousness-scoring tools of ANNOVAR were applied to score the rest of 335 variants. Finally, 77 predicted-to-be pathogenic variants survived in the final list from these databases. These 77variants, present in 167 individuals, predicted a prevalence of *ABHD5*-associated NAFLD in the general population due to monoallelic *ABHD5* mutations as 1 out of 1,131 (weighted average of mean of the three databases) in apparently healthy individuals (non-CDS), a figure that falls to <0.001 if the genetic penetrance of mutations is assumed to be 95%.

Bioinformatics Filtering of Variants in the *ABHD5* Gene in Normal Population Databases			
	Database		
Summary of variants	**ExAC**	**gnomAD**	**GME**
Individuals	**60,706**	**125,748**	**2,497**
Total number of the variants in *ABHD5* locus	**357**	**775**	**17**
Exonic/splicing variants	**198**	**283**	**17**
Remove synonymous variants	**131**	**192**	**12**
Remove benign variants by prediction programs*	**31****	**43*****	**3******
Calculated population frequency in databases	**1:1,065**	**1:1,186**	**1:744**
*SIFT, PolyPhen-2,, MutationTaster, FATHMM			
**31 variants in 57 carriers out of 60,706 individuals			
***43 variants in 106 carriers out of 125,748 individuals			
****3 variants in 4 carriers out of 2497 individuals			

Table S1: List of Pathogenic/Likely Pathogenic Mutations in *ABHD5* in Normal Population Databases[a]													
No.	Database	Variant Types	Variants (*ABHD5*:NM_016006)	Allele Count	CADD[b] Score	SIFT[c]	Polyphen2[d]	LRT[e]	MutationTaster[f]	FATHMM[g]	PROVEAN[h]	MutationAssessor[i]
1	ExAC	exonic	Exon1:c.8delinsCGGA	5	.	.	.	.	.	.	.	.
2	ExAC	exonic	Exon1:c.G40T:p.G14X	1	33	.	.	N	A	.	.	.
3	ExAC	exonic	Exon3:c.A171G:p.I57M	1	22.4	D	P	D	D	D	N	M
4	ExAC	exonic	Exon3:c.C340T:p.R114X	3	38	.	.	D	A	.	.	.
5	ExAC	exonic	Exon3:c.372_374delAGA	1	.	.	.	.	.	.	.	.
6	ExAC	exonic	Exon3:p.A161Lfs*10	1	.	.	.		.	.		.
7	ExAC	splicing	Exon4:c.507-1G>A	2	26.2	D	.	.	D	D	D	.
8	ExAC	exonic	Exon4:p.V176Gfs*30	1	41	.	.	.	D	.	.	H
9	ExAC	exonic	Exon4:c.C550T:p.R184X	1	39	.	.	D	A	.	.	.
10	ExAC	exonic	Exon4:c.G551A:p.R184Q	2	24.7	D	B	D	D	D	D	M
11	ExAC	exonic	Exon4:c.A574T:p.R192X	1	41	.	.	D	A	.	.	.
12	ExAC	exonic	Exon4:c.G591A:p.W197X	1	40	.	.	D	A	.	.	.
13	ExAC	exonic	Exon4:c.C628G:p.P210A	12	27.7	D	D	D	D	D	D	H
14	ExAC	exonic	Exon4:c.C650T:p.A217V	1	24.1	D	P	D	D	D	D	L
15	ExAC	exonic	Exon4:p.F220Lfs*3	1	36.7	.	.	.	.	.	.	.
16	ExAC	exonic	Exon5:c.T683A:p.L228X	1	42	.	.	D	A	.	.	.
17	ExAC	exonic	Exon5:c.711delinsTTCAATGTTC	1	.	.	.	.	.	.	.	.
18	ExAC	exonic	Exon5:c.G727C:p.D243H	1	24.7	D	P	D	D	D	D	M
19	ExAC	exonic	Exon5:c.729delinsTA	4	.	.	.	.	.	.	.	.
20	ExAC	exonic	Exon5:c.A752C:p.H251P	1	26.5	D	D	D	D	D	D	M

21	ExAC	exonic	Exon5:c.G755A:p.C252Y	1	30	D	D	D	D	D	D	M
22	ExAC	exonic	Exon6:c.G778A:p.E260K	1	34	D	D	D	A	D	D	H
23	ExAC	exonic	Exon6:c.A820G:p.K274E	1	28.1	D	P	D	D	D	D	M
24	ExAC	exonic	Exon6:p.Q279Rfs*14	1	37.4	.	.	.	D	.	.	M
25	ExAC	exonic	Exon6:c.C835T:p.Q279X	1	41	.	.	N	A	.	.	.
26	ExAC	exonic	Exon6:c.C838G:p.R280G	2	33	D	D	D	D	D	D	M
27	ExAC	exonic	Exon6:c.T851G:p.M284R	1	25	D	P	D	D	D	D	L
28	ExAC	exonic	Exon6:c.886delinsGC	1	.	.	.	.	.	.	.	.
29	ExAC	exonic	Exon6:c.C934T:p.R312X	3	39	D	.	.	D	D	N	.
30	ExAC	exonic	Exon6:c.G935A:p.R312Q	2	33	D	D	D	D	D	D	M
31	ExAC	exonic	Exon6:c.1005_1006delAG	1	.	.	.	.	.	.	.	.
32	gnomAD	exonic	Exon1:c.2delinsTGGC	10	.	.	.	.	.	.	.	.
33	gnomAD	exonic	Exon1:c.8delinsCGGA	5	.	.	.	.	.	.	.	.
34	gnomAD	exonic	Exon1:c.18delinsGGAGGTGGACTCTGCCGACACCGGAGA	1	.	.	.	.	.	.	.	.
35	gnomAD	exonic	Exon2:c.T79A:p.W27R	2	25.6	D	D	D	D	D	D	M
36	gnomAD	exonic	Exon3:p.N71IfsX23	1	30.1	.	.	.	D	.	.	D
37	gnomAD	exonic	Exon3:c.C230A:p.P77Q	1	28.8	D	D	D	D	D	D	M
38	gnomAD	exonic	Exon3:c.T315A:p.Y105X	1	35	.	.	D	A	.	.	.
39	gnomAD	exonic	Exon3:c.G332A:p.G111D	1	28.7	D	D	D	D	D	D	H
40	gnomAD	exonic	Exon3:c.C340T:p.R114X	4	38	.	.	D	A	.	.	.
41	gnomAD	exonic	Exon3:c.372_374delAGA	1	.	.		.	.	.	.	.
42	gnomAD	exonic	Exon3:c.A505G:p.R169G	3	26.4	D	D	D	D	D	D	H
43	gnomAD	splicing	Exon4:c.507-1G>A	2	26.2	D	.	.	D	D	D	.
44	gnomAD	splicing	Exon4:c.507-1G>T	1	25.7	D	.	.	D	D	D	.
45	gnomAD	exonic	Exon4:p.V176GfsX30	1	.	.	.	.	.	.	.	.

46	gnomAD	exonic	Exon4:c.C550T:p.R184X	2	39	.	.	D	A	.	.	.
47	gnomAD	exonic	Exon4:c.G551A:p.R184Q	6	24.7	D	P	D	D	D	D	M
48	gnomAD	exonic	Exon4:c.C568T:p.Q190X	1	37	.	.	N	A	.	.	.
49	gnomAD	exonic	Exon4:c.A574T:p.R192X	1	41	.	.	D	A	.	.	.
50	gnomAD	exonic	Exon4:c.G590C:p.W197S	1	33	D	D	D	D	D	D	M
51	gnomAD	exonic	Exon4:c.C628G:p.P210A	21	27.7	D	D	D	D	D	D	H
52	gnomAD	exonic	Exon4:c.C628T:p.P210S	1	33	D	D	D	D	D	D	M
53	gnomAD	exonic	Exon4:c.C650T:p.A217V	1	24.1	D	P	D	D	D	D	L
54	gnomAD	exonic	Exon5:c.G678C:p.Q226H	1	23.4	D	D	D	D	D	D	M
55	gnomAD	exonic	Exon5:c.T683A:p.L228X	1	42	.	.	D	A	.	.	.
56	gnomAD	exonic	Exon5:c.711delinsTTCAATGTTC	1	.	.	.	.	.	.	.	.
57	gnomAD	exonic	Exon5:c.G727C:p.D243H	3	24.7	D	P	D	D	D	D	M
58	gnomAD	exonic	Exon5:c.729delinsTA	5	.	.	.	.	.	.	.	.
59	gnomAD	exonic	Exon5:c.C751A:p.H251N	1	33	D	D	D	D	D	D	M
60	gnomAD	exonic	Exon5:c.G755A:p.C252Y	4	30	D	D	D	D	D	D	M
61	gnomAD	exonic	Exon5:c.G760A:p.V254M	1	24.9	D	P	D	D	D	N	M
62	gnomAD	exonic	Exon6:c.G776A:p.G259D	1	32	D	D	D	D	D	D	H
63	gnomAD	exonic	Exon6:c.G778A:p.E260K	1	34	D	D	D	A	D	D	H
64	gnomAD	exonic	Exon6:c.A820G:p.K274E	1	28.1	D	D	D	D	D	D	M
65	gnomAD	exonic	Exon6:c.C826T:p.P276S	1	33	D	D	D	D	D	D	M
66	gnomAD	exonic	Exon6:c.C835T:p.Q279X	1	41	.	.	N	A	.	.	.
67	gnomAD	exonic	Exon6:c.C838G:p.R280G	4	33	D	D	D	D	D	D	M
68	gnomAD	exonic	Exon6:c.G839A:p.R280Q	1	35	D	D	D	D	D	D	M
69	gnomAD	exonic	Exon6:c.T851G:p.M284R	1	25	D	P	D	D	D	D	L
70	gnomAD	exonic	Exon6:c.886delinsGC	1	.	.	.	.	.	.	.	.

71	gnomAD	exonic	Exon6:c.C911T:p.S304F	1	33	D	P	D	D	D	D	M
72	gnomAD	exonic	Exon6:c.C934T:p.R312X	6	39	D	.	.	D	D	N	.
73	gnomAD	exonic	Exon6:c.G935A:p.R312Q	1	33	D	D	D	D	D	D	M
74	gnomAD	exonic	Exon6:c.1005_1006delAG	1	.	.	.	.	.	.	.	.
75	GME	exonic	Exon2:c.T79A:p.W27R	2	25.6	D	D	D	D	D	D	M
76	GME	exonic	Exon3:c.A171G:p.I57M	1	22.4	D	P	D	D	D	D	M
77	GME	exonic	Exon7:c.A964G:p.I322V	1	24	D	P	N	D	D	D	M

[a] http://exac.broadinstitute.org, https://gnomad.broadinstitute.org/, http://igm.ucsd.edu/gme/

[b] https://cadd.gs.washington.edu/snv: >20 Damaging

[c] http://sift.jcvi.org: SIFT: D: Deleterious (sift<=0.05); T: tolerated (sift>0.05)

[d] http://genetics.bwh.harvard.edu/pph2/. PolyPhen2: D: Probably damaging (>=0.957), P: possibly damaging (0.453<=pp2_hdiv<=0.956); B: benign (pp2_hdiv<=0.452)

[e] http://www.genetics.wustl.edu/jflab/lrt_query.html LRT: D: Deleterious; N: Neutral; U: Unknown

[f] http://www.mutationtaster.org MutationTaster: A" ("disease_causing_automatic"); "D" ("disease_causing"); "N" ("polymorphism"); "P" ("polymorphism_automatic")

[g] http://fathmm.biocompute.org.uk FATHMM: D: Deleterious; T: Tolerated

[h] http://provean.jcvi.org/index.php PROVEAN: Deleterious (<=0.05); T: tolerated (>0.05)

[i] http://mutationassessor.org/ MutationAssessor: H: high; M: medium; L: low; N: neutral. H/M means functional and L/N means non-functional

Supplementary references

1. Besharat S, Poustchi H, Mohamadkhani A, et al. Association of mutations in the basal core promoter and pre-core regions of the hepatitis B viral genome and longitudinal changes in HBV level in HBeAg negative individuals: Results from a cohort study in Northern Iran. Hepat Mon 2015;15:e23875.
2. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.
3. Tavian D, Colombo R. Improved cytochemical method for detecting Jordans' bodies in neutral lipid storage diseases. J Clin Pathol 2007;60:956-8.