Supplementary Material

1. **Supplementary Methods**
2. **Supplementary Figure**
3. **Supplementary Table**
4. **Supplementary References**

1. Supplementary Methods

Exome sequencing and variants analysis

Genomic DNA was sent for whole exome sequencing at the Broad Institute Genomic Services. Sequencing reads were aligned to reference genome hg19 using Burrows Wheeler Aligner (Li & Durbin, 2009). Exome coverage was 92.9% with a mean target coverage of 82 reads. Aligned reads were sorted and duplicates marked using Picard Tools (Broad Institute). The Genome Analysis Toolkit was used to call variants, recalibrate base quality scores, then recall variants based on the recalibration scores using the best practices protocol for variant analysis (Van der Auwera et al., 2013). We used Annovar to annotate variants, loaded the variants into an SQL database, and used custom SQL queries to identify rare, homozygous and compound heterozygous nonsynonymous or truncating variants (Wang, Li, & Hakonarson, 2010). Variant frequency of less than 1% was filtered using data from the Genome Aggregation Database (Lek et al., 2016), the Greater Middle East Variome Project (Scott et al., 2016) and Iranome (Akbari et al., 2017). Protein pathogenicity of variants was predicted using CADD (Kircher et al., 2014), SIFT (Ng & Henikoff, 2003), and Polyphen-2 (Adzhubei et al., 2010). Further annotation on the clinical significance of variants was gathered from the databases UCSC Genome Browser (Kent et al., 2002), Uniprot (Poux et al., 2017), Online Mendelian Inheritance of Man (McKusick-Nathans Institute of Genetic Medicine), and The Human Gene Mutation Database (Stenson et al., 2017). The methodology of exome sequencing and variant analysis for family PKMR97 (Thr198Met) has been reported in detail previously (Nguyen et al., 2014).

Generation of human point mutations in human Mfsd2a

The five human mutations of Mfsd2a, Pro493Leu (P493L), Thr198Met (T198M), Pro164Thr (P164T), and compound heterozygote Arg326His (R326H) and Val250Phe (V250F) were individually generated through the amplification of human Mfsd2a using gene-specific and site-

specific mutagenic primers and ligated into pcDNA3.1 after digestion with restriction enzymes EcoRV and XbaI.

3D structural modeling of the T198M, P164T, P493L, R326H and V250F mutants

Starting from the published 3D model of MFSD2A WT in the outward occluded state, single point mutations T198M and P493L were generated independently by sidechain prediction using SCWRL (Quek et al., 2016). This initial model of T198M or P493L was subjected to local structural optimization by loop modeling implemented in MODELLER (Sali et al., 1993), resulting in 2500 models that were evaluated by the DOPE (discrete optimized protein energy) score to select the best ranked model.[3] For P164T, R326H, and V250F, point mutations were generated from the same starting model[1] using the mutagenesis function followed by local sphere regularization with secondary structure restraints in COOT (Emsley et al., 2010). Molecular graphics were created in PyMOL (The PyMOL Molecular Graphics System, 2002).

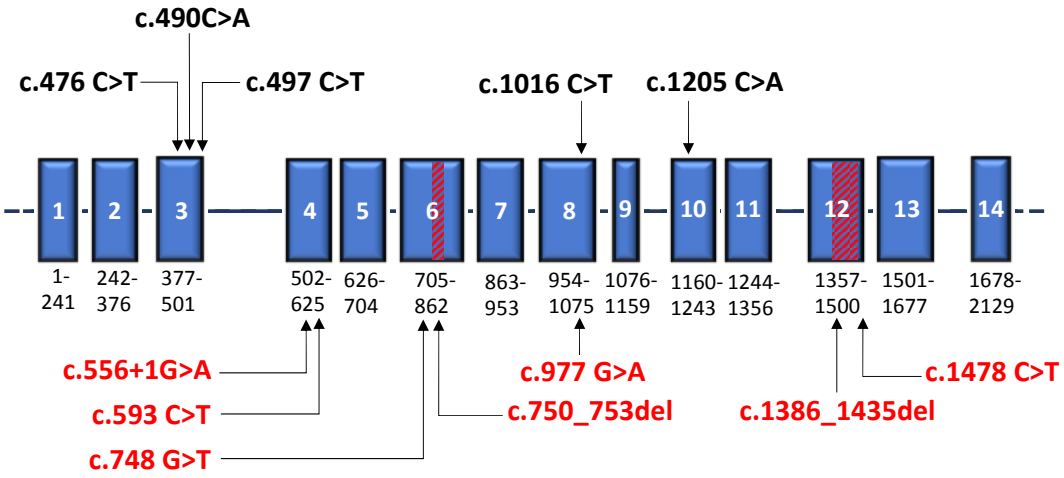
Western Blot and Immunofluorescence analysis of mutant transiently transfected in HEK293 cells

Cellular expression of the human mutants was compared with the wild-type (WT) Mfsd2a, and the non-functional sodium binding mutant Asp97Ala (D97A) expression constructs by immunoblotting using a rabbit polyclonal antibody against Mfsd2a on transiently transfected HEK293 cells (Chan et al., 2018). Using the same antibody against MFSD2A, the cellular localizations of the mutants transiently transfected into HEK293 were also visualized together with its WT Mfsd2a as a control using confocal immunofluorescence microscopy (Zeiss). Cell transfected with an empty pcDNA3.1 was used as a negative control. Details of these methods were previously described (Nguyen et al., 2014; Quek et al., 2016).

Mfsd2a Transport assays

In vitro transport of the Mfsd2a ligand, ^{14}C-Lysophosphatidylcholine-Docosahexaenoic acid (LPC-DHA) (ARC Radiochemicals), spiked into unlabeled 10 mM LPC-DHA (Vanteres Pte Ltd) was tested in HEK293 cells transiently transfected with wild-type (WT) Mfsd2a and mutants for 24 hours (Nguyen et al., 2014; Quek et al., 2016). Uptake activity of 14C-LPC-DHA for all constructs were measured after 30 minutes incubation with 50 mM LPC-DHA diluted in serum-free DMEM (Gibco). The cells were washed two times in serum-free DMEM (Gibco) containing 0.5% fatty-acid free bovine serum albumin and harvested with RIPA buffer into 4 ml of scintillation fluid (Ecolite, MP-biopharmaceuticals). Disintegrations Per Minute (DPM) of the incorporated LPC-DHA in each well of transfected HEK293 cells were counted using a scintillation counter (Tricarb, Perkin Elmer). All transport assays were carried out in triplicates using a 12-well plate.

2. Supplementary Figure



Supplementary Figure 1. Schematic drawing of *MFSD2A* with previously reported variants (in black) and the variants identified in this study (in red). Intragenic deletions are indicated by diagonal lines within the affected exon.

3. Supplementary Table

Table S1. Other potential causative variants in the reported *MFSD2A* families.

Families	Gene	Variant (hg19)	Status	gnomAD	GME	Iranome	ClinVar (ID)	SIFT	Mutation Taster	GERP score	CADD score	ACMG class
A	*MACF1*	chr1:39765977 C>A	hom	0	0	0	-	0.238	0.977	5.81	16.5	III (BP4, PM2, PP3)
	SZT2	chr1:43908592 C>T	hom	0.00003	0	0	-	0.002	1	5.67	34	III (BP1, PM2, PP3)
	CACHD1	chr1:65016278 G>A	hom	0.00140	0.00302	0.00375	-	0.178	1	6.02	27.6	II (BS1)
	TTN	chr2:179395282 G>C	hom	0	0	0	-	1	1	5.23	13.1	II (BP1, BP4, PM2)
	PARD3B	chr2:206057991 C>T	hom	0.00003	0	0	-	0.102	0.999	5.63	22.5	III (BP4)
	ABCA12	chr2:215802262 T>C	hom	0	0	0	-	0.11	0.801	5.67	21.3	II (BP1, BP4, PM2)
	TBL1XR1	chr3:176752064 T>C	hom	0	0	0.00063	-	1	1	5.65	17.9	III (PM2, PP2)
	ALG3	chr3:183960623 G>A	hom	0.00007	0	0.00063	-	0.007	0.999	5.09	25	III (PM2, PP2, PP3)
	ATP13A5	chr3:193039554 C>T	hom	0.00020	0	0	-	0.592	1	5.82	5.8	II (BS1, BP4)
	LRRC15	chr3:194081159 T>C	hom	0.00020	0	0	-	0.029	1	5.02	17.5	II (BS1, BP4)
	RGS12	chr4:3344267 T>C	hom	0.00460	0	0.00625	-	0	0	1.49	1.8	II (BS1, BP4)
	ADAMTS8	chr11:130289012 C>T	hom	0.00040	0.00151	0.00438	-	0.041	1	5.62	13.8	II (BS1, BP4)
	MPP2	chr17:41960701 G>C	hom	0.00001	0	0	-	0.487	1	4.15	17.9	III (BP4, PM2)
	FAM187A	chr17:42982324 C>T	hom	0.00390	0	0.03062	-	0.465	1	5.54	11.3	II (BS1, BP4)
	STH	chr17:44077019 C>G	hom	0.00002	0	0.00063	-	0	1	2.03	34	III (BP4, PM1, PM2)
	ZDHHC8P1	chr22:23742049 G>A	hom	0	0	0.05882	-	0	0	1.82	4.7	I (BA1, BP4)
	CRYBB2P1	chr22:25853368 T>C	hom	0	0	0.1181	-	0	0	2.22	12.3	I (BA1, BP4)
B†	-	-	-	-	-	-	-	-	-	-	-	-
C	*SCP2*	chr1:53393072 T>G	hom	0.00005	0	0	-	0.778	0.885	3.14	-	III (PM2, BP4)
	TMCC2	chr1:205241169 C>T	hom	0.00006	0.00251	0.00437	-	0.492	0.999	5.18	-	II (BS1, BP4)
	NAGK	chr2:71297921 G>C	hom	-	0	0	-	0.26	0.995	4.95	20.7	III (PM2, BP4)
	NAGK	chr2:71295842 G>T	hom	-	0	0	-	0.002	1	5.11	36	III (PM2, PP3)
	MAP6	chr11:75378664 C>T	hom	-	0	0	-	0.438	0.999	4.5	6.9	III (PM2, BP4)
	POSTN	chr13:38166262 C>T	hom	0.00074	0	0.00063	-	0.331	0.999	5.18	22.2	II (BS1, BP4)
D†	-	-	-	-	-	-	-	-	-	-	-	-
E	*CDKL5*	chrX:18668586 C>T	het	0.00019	0	0	RCV000475262	0	0.999	-7.59	0	I (BS1, BS2, BP4, BP6)
	TUBB3	chr16:90002195 G>A	het	0.00035	0	0	RCV000903349	0.001	1	4.66	16	II (PP2, BS1, BP4, BP6)
F	*BRWD1*	chr21:40608526 T>C	hom	0	0	0	-	0.154	0.899	5.44	15.4	III (PM2, BP4)
G	*EXOSC8*	chr13:37583420 G>C	hom	0.00385	0.00554	0.00875	RCV000418794	0	1	5.85	14	II (PP3, PP5, BS1, BP1)
	ALDH5A1	chr6:24495252 T>C	hom	0.000096	0	0	-	0.212	0.999	1.27	4	II (PM2, BP1, BP4)

ACMG American College of Medical Genetics and Genomics, *BA* Benign stand alone, *BS* Benign Strong, *BP* Benign supporting, *CADD* Combined Annotation Dependent Depletion, *GERP* Genomic Evolutionary Rate Profiling, *GME* Greater Middle East Variome Project, *PM* Pathogenic Moderate, *PP* Pathogenic supporting, *SIFT* Sorting Intolerant From Tolerant, *VEP* Variant Effect Predictor, *VUS* variant of unknown significance. † In these two families, no other possible causative variant could be identified.

3. Supplementary References

Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit 7.20.

Chan JP, Wong BH, Chin CF, et al The lysolipid transporter Mfsd2a regulates lipogenesis in the developing brain. PLoS Biol. 2018;16:e2006443.

Emsley P, Lohkamp B, Scott WG, et al. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66:486-501.

Fattahi Z, Beheshtian M, Mohseni M, et al. Iranome: A catalog of genomic variations in the Iranian population. Hum Mutat. 2019;40:1968-1984.

Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res. 2002;12:996-1006.

Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310-315.

Krivov GG, Shapovalov MV, Dunbrack RL Jr. Improved prediction of protein side-chain conformations with SCWRL4. Proteins. 2009;77:778-795.

Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285-291.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-1760.

Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812-3814.

Nguyen LN, Ma D, Shui G, et al. Mfsd2a is a transporter for the essential omega-3 fatty acid docosahexaenoic acid. Nature. 2014;509:503-506.

Poux S, Arighi CN, Magrane M, et al. On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. Bioinformatics. 2017;33:3454-3460.

Quek DQ, Nguyen LN, Fan H, Silver DL. Structural Insights into the Transport Mechanism of the Human Sodium-dependent Lysophosphatidylcholine Transporter MFSD2A. J Biol Chem. 2016;291:9383-9394.

Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234:779-815.

Scott EM, Halees A, Itan Y, et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. Nat Genet. 2016;48:1071-1076.

Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006;15:2507-2524.

Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet. 2017;136:665-677.

The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. 2002.

Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1-11.10.33.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.