Squalene Synthase Deficiency: Clinical, Biochemical, and Molecular Characterization of a Defect in Cholesterol Biosynthesis

David Coman,[1,2,4,5,20,*] Lisenka E.L.M. Vissers,[6,20] Lisa G. Riley,[7,8] Michael P. Kwint,[6] Roxanna Hauck,[8] Janet Koster,[9] Sinje Geuer,[6] Sarah Hopkins,[10] Barbra Hallinan,[11] Larry Sweetman,[12] Udo F.H. Engelke,[13] T. Andrew Burrow,[14] John Cardinal,[2] James McGill,[1,3,4] Anita Inwood,[1] Christine Gurnsey,[3] Hans R. Waterham,[9] John Christodoulou,[7,8,15,16,17] Ron A. Wevers,[18,21] and James Pitt[17,19,21,*]

Mendelian disorders of cholesterol biosynthesis typically result in multi-system clinical phenotypes, underlining the importance of cholesterol in embryogenesis and development. *FDFT1* encodes for an evolutionarily conserved enzyme, squalene synthase (SS, farnesyl-pyrophosphate farnesyl-transferase 1), which catalyzes the first committed step in cholesterol biosynthesis. We report three individuals with profound developmental delay, brain abnormalities, 2-3 syndactyly of the toes, and facial dysmorphisms, resembling Smith-Lemli-Opitz syndrome, the most common cholesterol biogenesis defect. The metabolite profile in plasma and urine suggested that their defect was at the level of squalene synthase. Whole-exome sequencing was used to identify recessive disease-causing variants in *FDFT1*. Functional characterization of one variant demonstrated a partial splicing defect and altered promoter and/or enhancer activity, reflecting essential mechanisms for regulating cholesterol biosynthesis/uptake in steady state.

The isoprenoid biosynthesis pathway (Figure 1), essential for numerous biological processes, can be divided into a pre-squalene (also named the mevalonate pathway) and a post-squalene pathway, the latter specifically involved in the synthesis of sterol isoprenoids, including cholesterol. Cholesterol serves as an important structural component of the cell membrane and myelin but is also the precursor of oxysterols, steroid hormones, and bile acids.[1] The two-step enzymatic conversion of farnesyl-pyrophosphate (FPP) to squalene by squalene synthase is the first committed step in cholesterol biosynthesis.[2,3]

We identified three individuals, a sibship and an unrelated child, whose urine metabolic profiles were suggestive of a cholesterol biosynthesis defect. (For all individuals, informed consent for genetic testing was obtained. For individuals 1 and 2, additional written consent was obtained for photo publication. Ethics approval was provided by Royal Children's Hospital and Health Service Ethics Committee and Sydney Children's Hospitals Network Human Research Ethics Committee.) Salient clinical features include facial dysmorphisms (Figure 2), dry skin with photosensitivity, generalized tonic-clonic seizures, structural brain malformations, cortical visual impairment, profound global developmental delay, and genital malformations in the two males (Table 1, see Supplemental Note). The gas chromatography-mass spectrometry and nuclear magnetic resonance spectroscopy profiles yielded a consistent and complex pattern of abnormal metabolites including the accumulation of methylsuccinic acid, mevalonate lactone, mesaconic acid, and 3-methyladipic acid. Additionally, we observed saturated and unsaturated branched-chain dicarboxylic acids and glucuronides derived from farnesol (Figure 3). A similar metabolite profile has previously been observed in the urine of animal models and humans treated with pharmacological inhibitors of squalene synthase, as well as in animals loaded with farnesol.[4–6] To confirm

[1]Department of Metabolic Medicine, Lady Cilento Children's Hospital, Brisbane, QLD 4101, Australia; [2]Department of Paediatrics, Wesley Hospital, Brisbane, QLD 4066, Australia; [3]Department of Chemical Pathology, Pathology Queensland, Royal Brisbane & Women's Hospital, Brisbane, QLD 4006, Australia; [4]School of Medicine, University of Queensland, Brisbane, QLD 4067, Australia; [5]School of Medicine, Griffith University, Gold Coast, QLD 4222, Australia; [6]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, 6525 GA Nijmegen, the Netherlands; [7]Genetic Metabolic Disorders Research Group, Sydney Children's Hospital Network, Sydney, NSW 2145, Australia; [8]Discipline of Child & Adolescent Health, Sydney Medical School, University of Sydney, Sydney, NSW 2006, Australia; [9]Laboratory Genetic Metabolic Diseases, Academic Medical Center, University of Amsterdam, 105 AZ Amsterdam, the Netherlands; [10]Division of Neurology, Children's Hospital, Philadelphia, PA 19104, USA; [11]Division of Neurology, Cincinnati Children's Medical Centre, Cincinnati, OH 45229, USA; [12]Institute of Metabolic Disease, Baylor Scott & White Research Institute, Baylor University Medical Center, Dallas, TX 75204, USA; [13]Translational Metabolic Laboratory - 830 TML, Department of Laboratory Medicine, Radboud University Medical Centre, Geert Grooteplein 10, 6525 GA Nijmegen, the Netherlands; [14]University of Arkansas for Medical Sciences College of Medicine, Department of Pediatrics Little Rock, Arkansas, AR 72205, USA; [15]Neurodevelopmental Genomics Research Group, Murdoch Children's Research Institute, Melbourne VIC 3052, Australia; [16]Genetic Metabolic Disorders Research Group, Sydney Children's Hospital Network, Sydney NSW 2145, Australia; [17]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Melbourne, VIC 3052, Australia; [18]Translational Metabolic Laboratory - 830 TML, Department of Laboratory Medicine, Radboud University Medical Centre, Geert Grooteplein 10, 6525 GA Nijmegen, the Netherlands; [19]Department of Paediatrics, University of Melbourne, Melbourne, VIC 3052, Australia
[20]These authors contributed equally to this work
[21]These authors contributed equally to this work
*Correspondence: david.coman@health.qld.gov.au (D.C.), james.pitt@vcgs.org.au (J.P.)
https://doi.org/10.1016/j.ajhg.2018.05.004.

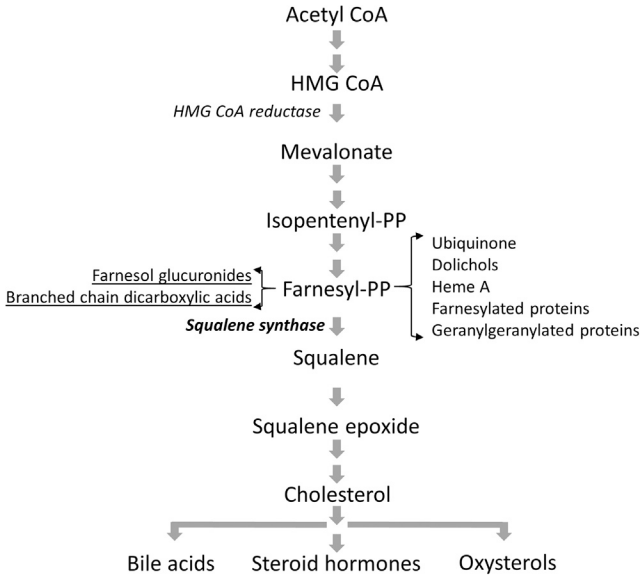


Figure 1. Schematic Representation of the Isoprenoid/Cholesterol Biosynthesis Pathway
Metabolites that accumulate as a result of squalene synthase deficiency are underlined (pathway re-drawn based on Tansey and Shechter[13]).

that the abnormal organic acids observed in the affected individuals derive from farnesol, we analyzed urine from a control subject after an oral farnesol load, which resulted in a similar metabolite profile (Figure S1). Plasma cholesterol in the three affected individuals was decreased (2.5–2.8 mmol/L; reference 3.0–5.5), and both HDL and LDL cholesterol levels were decreased or low normal (Table 1). Plasma total farnesol levels (the sum of free farnesol and farnesyl-pyrophosphate) in affected individuals were, however, significantly increased (1.5–3.9 μmol/L; reference $<$ 0.12; Table 1) while plasma squalene levels were reduced or normal (0.17–0.93 μmol/L; reference 0.36–1.04). Thus, the body fluid metabolite profile of the affected individuals was strongly suggestive of a cholesterol biosynthesis defect at the level of squalene synthase, converting farnesyl-pyrophosphate into squalene.

The affected individuals and their unaffected parents were subjected to whole-exome sequencing (WES). Informed consent for genetic testing was obtained for all individuals under protocols approved by the Royal Children's Hospital and Health Service Ethics Committee and Sydney Children's Hospitals Network Human Research Ethics Committee. Assuming a recessive mode of inheritance, prioritization of variants revealed compound heterozygous variants affecting *FDFT1*, encoding the SS enzyme, in the sibship (individuals 1 and 2): a maternally inherited 120 kb deletion (chr8(GRCh37):g.11667760–11787743), resulting in loss of exons 6–10 of *FDFT1* and the entire coding sequence of the neighboring *CTSB* gene (encoding cathepsin B [MIM: 116810], haploinsufficiency for which is asymptomatic in the mother), and a paternally inherited variant GenBank: NM_001287742.1(FDFT1); c.880−24_880−23delinsAG, predicted to create a novel splice acceptor site. The latter prediction was functionally tested using a minigene splice assay, which indeed showed the retention of 22 bp of intron 8 sequence (Figure S2). *FDFT1*cDNA analysis using RNA isolates generated from fibroblasts of the sibship confirmed the partial splicing defect, showing both normally spliced *FDFT1* cDNA and mis-spliced *FDFT1* cDNA (Figure S2). Subsequent western blot analysis demonstrated a marked reduction in squalene synthase protein in cultured lymphoblasts and fibroblasts in individuals 1 and 2 (Figures 4B and S2). An even greater reduction was observed when fibroblasts were cultured in lipid-depleted media. While the FDFT1 level is further reduced relative to the control in the LD-FBS (compared to FBS), we hypothesize that the absolute level may not be decreased, but *FDFT1* expression in the control is upregulated in LD-FBS. We hypothesize that the residual squalene synthase protein remaining is likely to arise from the correctly spliced paternal allele, which would also explain our observation at cDNA level. In individual 3, no

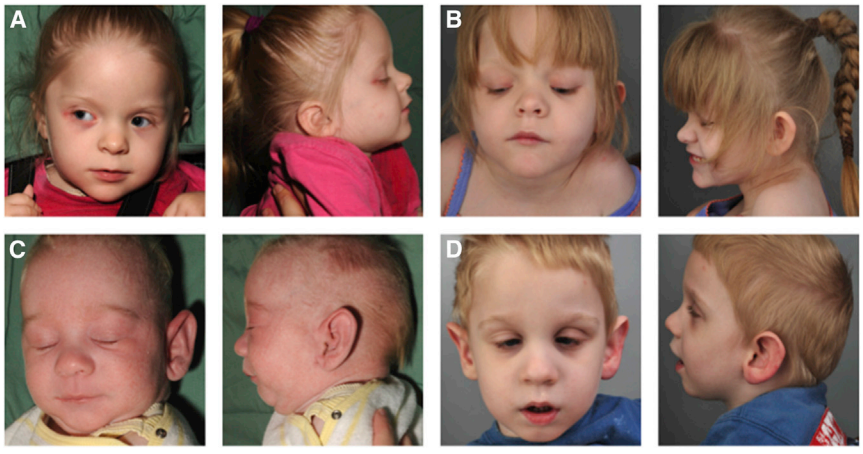


Figure 2. Facial Characteristics
Individual 1 at 3 years of age (A) and 6 years of age (B) and individual 2 at 5 weeks of age (C) and 3 years of age (D). Dysmorphic features include depressed nasal bridge, low-set posteriorly rotated ears, square nasal tip, epicanthic folds, mild micrognathia, and retrognathia.

Table 1. Summary of Key Clinical, Biochemical, and Molecular Features

	Individual 1	Individual 2	Individual 3
Age at presentation	day 5 with seizures	day 5 with seizures	first week of life with seizures
Current age	10 years old	7 years old	9 years old
Gender	Female	Male	Male
Gestation	37/40	39/40	39/40
Birth weight	2,600 g (40^{th} percentile)	2,750 g (10^{th} percentile)	3,033 g (25^{th} percentile)
Birth length	47 cm (50^{th} percentile)	49 cm (30^{th} percentile)	50 cm (50^{th} percentile)
Birth head circumference	32.5 cm (50^{th} percentile)	34 cm (50^{th} percentile)	33 cm (10^{th} percentile)
Ethnicity	European	European	European
Parental relationship	non-consanguineous	non-consanguineous	non-consanguineous
MRI brain abnormalities	hypoplastic corpus callosum, white matter loss	hypoplastic corpus callosum, white matter loss	diffuse polymicrogyria, central white matter and cortical volume loss
Seizures	yes, neonatal onset	yes, neonatal onset	yes, neonatal onset
Global developmental delay	profound across all developmental modalities	profound across all developmental modalities	profound across all developmental modalities
Irritability	yes	yes	yes
Optic nerve hypoplasia	yes	yes	no
Cataracts	no	no	no
Visual impairment	cortical VI	cortical VI	cortical VI
Dysmorphism	depressed nasal bridge low set posteriorly rotated ears square nasal tip epicanthic folds mild micrognathia mild retrognathia 2-3 toe syndactyly	depressed nasal bridge large ears square nasal tip epicanthic folds mild micrognathia mild retrognathia 2-3 toe syndactyly dorsal foot fat pads	coarse facial features bitemporal narrowing prominent ears triangular facies
Genitals	normal	bilateral cryptorchidism	hypospadias
Cardiac malformations	no	no	bicuspid aortic valve
Gastrointestinal	IUGR with FTT PEG feeds constipation	IUGR with FTT PEG feeds constipation	IUGR with FTT PEG feeds constipation
Skeletal	thin gracile bones fixed flexion deformity knees	thin gracile bones fixed flexion deformity knees	fixed flexion deformities in elbows
Skin	dry skin photosensitivity lack of hair pigment on LM/EM	dry skin photosensitivity lack of hair pigment on LM/EM	dry skin photosensitivity
Sleep cycle	poor sleep initiation	poor sleep initiation	poor sleep initiation and maintenance
Cholesterol	2.5 (3–5.5 mmol/L)	2.8 (3–5.5 mmol/L)	2.7 (3–5.5 mmol/L)
Triglyceride	1.1 (0.55–2.0 mmol/L)	1.6 (0.55–2.0 mmol/L)	0.8 (0.55–2.0 mmol/L)
HDL-C	0.7 (0.9–2.2 mmol/L)	1.2 (0.9–2.2 mmol/L)	1.06 (0.9–2.2 mmol/L)
LDL-C	1.3 (2.0–3.4 mmol/L)	0.8 (2.0–3.4 mmol/L)	1.3 (2.0–3.4 mmol/L)
VLDL-C	0.5 (0.1–20.65 mmol/L)	0.7 (0.1–20.65 mmol/L)	not performed
Plasma total farnesol	1.5, 1.6 (<0.12 μmol/L)	1.7, 3.9 (<0.12 μmol/L)	not performed

(*Continued on next page*)

Table 1. ***Continued***

	Individual 1	Individual 2	Individual 3
Plasma squalene	0.81, 0.93 (0.36–1.04 μmol/L)	0.17, 0.37 (0.36–1.04 μmol/L)	not performed
Molecular	compound heterozygous mutations: chr8(GRCh37): g. 11667760-11787743, maternal; chr8(GRCh37): g.11689003_ 11689004delinsAG; NM_001287742.1 (FDFT1): c.880-24_880-23delinsAG, paternal	compound heterozygous mutations: chr8(GRCh37): g. 11667760-11787743, maternal; chr8(GRCh37): g.11689003_11689004delinsAG; NM_001287742.1 (FDFT1): c.880-24_880-23delinsAG, paternal	homozygous mutation: chr8(GRCh37): g.11660095_ 11660110del; NM_001287742.1: c.-75+131_-75+146del

Abbreviations: IUGR, intra-uterine growth retardation; PEG, percutaneous endoscopic gastrostomy; FTT, failure to thrive; LM/EM, light microscopy and electro-electron microscopy.

pathogenic bi-allelic variants affecting the coding sequence of *FDFT1* could be identified. Interestingly, however, some of the *FDFT1* deeper intronic sequence was interpretable from WES data, in allowing the identification of a rare homozygous intronic 16 bp deletion (Chr8(GRCh37):g.11660095_11660110del). The variant was validated using Sanger sequencing in individual 3 and confirmed to be heterozygous in both non-consanguineous parents (Figure S3). Notably, the deletion is absent from our in-house database containing WES data of >15,000 individuals, nor has it been reported in GnomAD (access date October 25, 2017),[7] suggesting that this is a private mutation (Figure S3). The residual variation intolerance score (RVIS) of *FDFT1* shows that it is among the 6.5% of human genes most intolerant to functional variation.[8] We next determined the functional consequence of the 16 bp intronic deletion in individual 3.

FDFT1 exists in 11 different isoforms, encoding 5 different protein variants (Figure S4). Using fetal and adult organ-specific cDNA libraries, we determined the normal pattern for 10/11 isoforms and detected ubiquitous *FDFT1* presence in all major organ tissues tested (Figure S5). Next, we tested cDNA generated from a fibroblast cell line of individual 3, as well as from a control individual, for the presence of these *FDFT1* isoforms. Whereas all ten isoforms tested were present in the control fibroblast line, only seven of ten were present in individual 3 (Figure S5), and three isoforms remained undetected (GenBank: NM_001287742.1, NM001287743.1, and NM00128774.4; Figure 4C). The addition of cycloheximide (CHX) failed to result in isoform detection (Figure 4C), suggesting that the absence of these isoforms is a consequence of abnormal regulation rather than erroneous splicing degraded by nonsense-mediated RNA decay. Under normal conditions, the absent *FDFT1* isoforms are observed in fetal and adult skeletal muscle, as well as in adult brain, spleen, testis, lung, and kidney (Figure S5). In line with the presence of the seven other *FDFT1* isoforms, western blot analysis on fibroblasts from individual 3 identified squalene synthase at protein levels comparable to a normal control individual (Figure S6).

We next explored the possibility that the 16 bp intronic deletion in *FDFT1* generates pathogenicity as a consequence of a regulatory disturbance. Annotation of the chromatic state for *FDFT1* indicates that the region containing the 16 bp deletion is predicted to have promoter and/or enhancer effects (Figure S7). To assess the effect of the deletion on the putative promoter activity, two constructs were generated and tested in a luciferase assay: one construct containing the 1,024-bp wild-type promoter sequence and one containing the same fragment but with the 16 bp deletion. Analysis of the wild-type sequence showed that indeed the sequence has promoter capacity (Figure 4D, Table S1). Moreover, the normalized luciferase readout of the construct containing the 16-bp deletion shows a significantly reduced promoter activity sequence (two-tailed t test, $p = 2.90 \times 10^{-6}$), suggesting that the aberrant expression of *FDFT1* isoforms in individual 3 may indeed be the consequence of diminished promoter activity due to the homozygous 16 bp deletion (Figure 4D, Table S1).

Farnesol and its products exhibit a wide variety of biological activities, including cell growth inhibition, induction of apoptosis, and regulation of bile acid secretion.[9] The primitive streak gives rise to the endoderm and mesoderm in embryonic stem cells (ES). Mouse ES primitive streak development is inhibited by statin medication, which is driven by protein farnesylation.[10] Farnesol-induced apoptosis is associated with ER stress and activation of the unfolded protein response, inhibition of the phosphatidylcholine system, activation of MAP-kinases, and activation of the apoptosome intrinsic mitochondrial-dependent caspases.[9] Evidence is emerging that dysregulation of the mevalonate pathway may be involved in the progression of neurodegeneration in disorders such as Alzheimer disease (MIM: 104300).[11]

Cholesterol biosynthesis defects are noted for significant birth malformations.[1] These are also evidenced in our individuals with *FDFT1* disruptive variants (Table 1). Pathogenic mechanisms could involve direct toxicity of accumulated metabolites, abnormal processes that involve farnesyl-pyrophosphate (such as protein farnesylation,

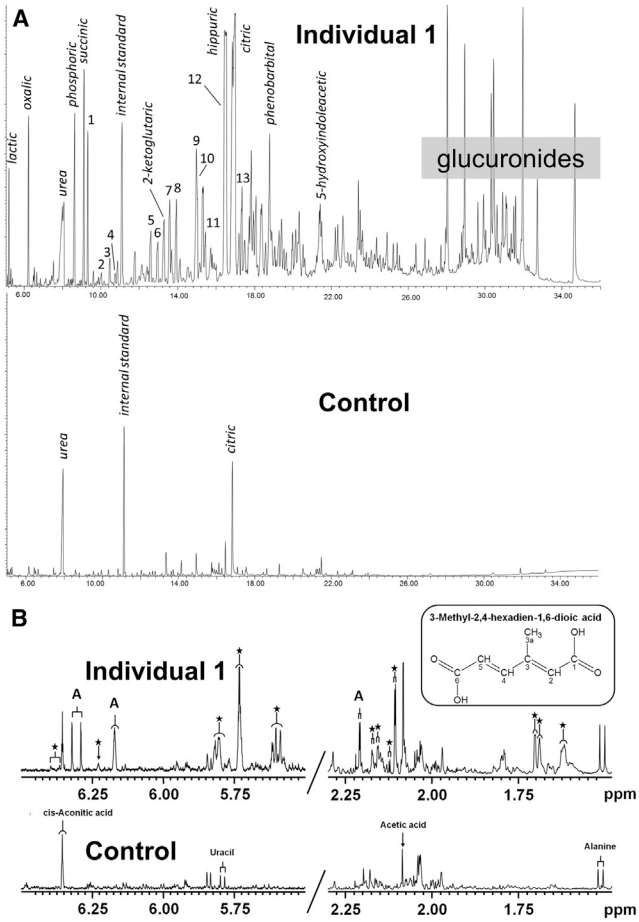


Figure 3. Urine Metabolite Profiles

(A) Urine organic acid GC-MS profiles of individual 1 (top) and a control subject (bottom). Normal urine components are indicated in italics. Numbered peaks indicate abnormal metabolites: (1) methylsuccinic, (2) mevalonic lactone, (3) mesaconic acid, (4) 2-methylgluatic acid, (5) 3-methyladipic acid, (6) 3-methylhex-3-enedioic acid, (7) 3-methylhex-2-endioic acid, (8) 2,6-dimethylheptanedioic acid, (9) unknown, (10) 3-methylhex-2,4-dienedioic, (11) 2,6-dimethylhept-2-enedioic acid, (12) 3,7-dimethyloctanedioic, and (13) 3,7-dimethyl-2,6-dienedioic. Control values are given in Figure S1.

(B) One-dimensional 500 MHz ^{1}H-NMR spectra of urine of individual 1 and age-matched control subject measured at pH 2.5 (regions 5.60–6.50 ppm and 1.50–2.25 ppm). The insert shows the structure of 3-methylhex-2,4-dienedioic acid. Proton assignments of 3-methylhex-3,4-dienedioic acid (A) and other farnesol-derived dicarboxylic acids (asterisk).

Figure 1), or reduced cholesterol moiety attachments to the hedgehog patterning gene family.[1] Also, tissue-specific analyses of the various *FDFT1* isoforms show that at least one isoform (GenBank: NM_001287743.1) is detected in brain, which is absent in individual 3 (Figure S5), and could potentially explain the severe neurodevelopmental disorder observed. Importantly, *Fdft1*-null mice demonstrate embryonic lethality at day 12.5 in conjunction with growth restriction and neurodevelopmental disorders.[12] The fact that the *FDFT1* variants in individuals 1, 2, and 3 are compatible with life may be explained by the fact that all individuals have some form of residual FDFT1 activity, either resulting from the diminished levels of correctly spliced *FDFT1* in individuals 1 and 2, or alternatively, by functional compensation for disrupted regulation in individual 3.

In conclusion, we describe squalene synthase deficiency due to pathogenic variants in *FDFT1* leading to altered splicing and transcriptional deregulation of the *FDFT1* isoforms. The clinical phenotype resembles other known cholesterol biosynthesis defects. The accumulation of farnesyl pyrophosphate in this disorder initiates a complex metabolic cascade involving glucuronidation, hydroxylation, and oxidation to shorter chain molecules. Our cohort exhibits a urine metabolite profile which is an important diagnostic indicator and provides insight into the important role of *FDFT1* in embryogenesis and morphogenesis.

Supplemental Data

Supplemental Data include seven figures, one table, Supplemental Case Reports, and Supplemental Methods and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.05.004.

Acknowledgments

The authors would, first and foremost, like to thank the families, without whose participation this work would not have been possible. This work was supported by the Kevin Milo Trust and internal research funds at the Children's Hospital at Westmead. The authors would like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at http://gnomad.broadinstitute.org/about. The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program.

Declaration of Interests

The authors declare no competing financial interests.

Received: February 24, 2018
Accepted: May 11, 2018
Published: June 14, 2018

Web Resources

dbSNP, https://www.ncbi.nlm.nih.gov/projects/SNP/
DECIPHER, https://decipher.sanger.ac.uk/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
gnomAD Browser, http://gnomad.broadinstitute.org/
MutationTaster, http://www.mutationtaster.org/
NIST Standard Reference Databases: Analytical Chemistry, http://www.nist.gov/srd/analy.cfm
OMIM, http://www.omim.org/
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
Primer3, http://bioinfo.ut.ee/primer3
SIFT, http://sift.bii.a-star.edu.sg/
UCSC Genome Browser, https://genome.ucsc.edu

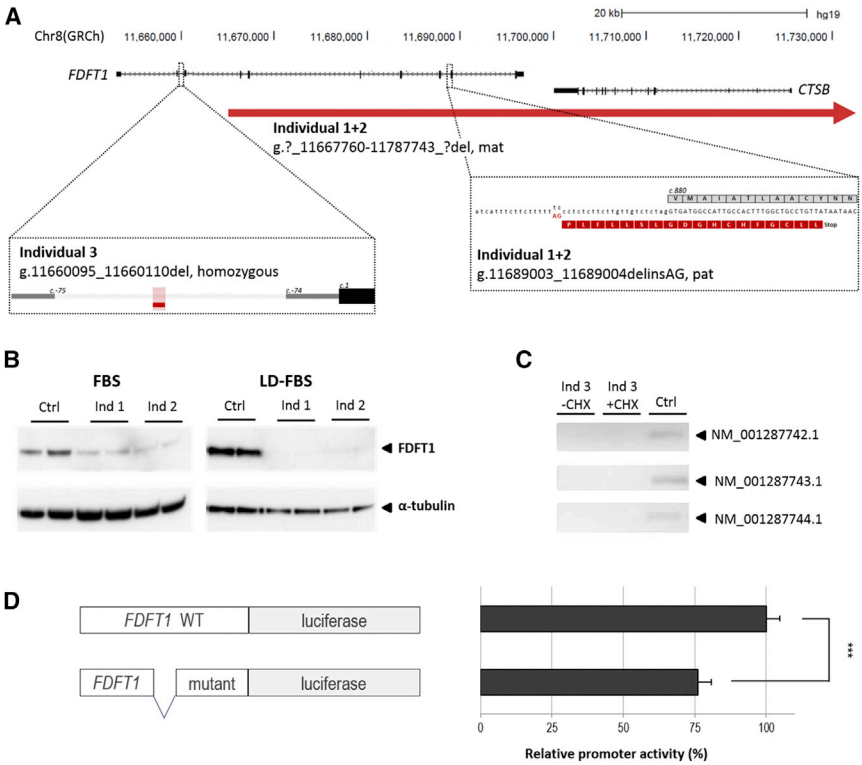


Figure 4. Schematic Representation of *FDFT1*, the Variants Identified in Individuals 1, 2, and 3, and Functional Follow-up

(A) The maternal deletion identified in the sibship extends ~50 kb more proximally, but no other genes are included. Genomic positions are based on hg19. Red horizontal lines indicate deleted sequence. For the paternal mutation in individuals 1 and 2, the wild-type protein amino acid sequence is provided in gray, whereas the predicted sequence of the splice mutation is shown in red.

(B) Western blot analysis of individuals 1 (Ind 1) and 2 (Ind 2), using two different culturing conditions (fetal bovine serum [FBS] and lipid depleted [LD-] FBS), using an antibody to detect FDFT1 in fibroblast of individuals 1 and 2. Alpha-tubulin was used as loading control. FDFT1 reduction is more prominent when cells were cultured under lipid depletion.

(C) Isoform-specific PCR using cDNA generated from EBV-transformed PBMCs of individual 3 (Ind 3) was used for the detection of FDFT1 isoforms. Three FDFT1 isoforms could not be detected in individual 3 whereas these were present in EBV-transformed PBMCs of a control individual. The addition of cycloheximide (CHX; - absent, + present) did not facilitate expression detection. Profile of all isoforms is presented in Figure S5.

(D) The genomic segment containing the homozygous 16 bp deletion in individual 3 was tested for promoter activity using a luciferase assay, which showed significantly reduced activity when compared to the control construct containing the wild-type sequence (two-tailed t test, $p = 2.9 \times 10^{-6}$). ***$p < 0.001$. Error bars represent one SD.

References

1. Waterham, H.R. (2006). Defects of cholesterol biosynthesis. FEBS Lett. *580*, 5442–5449.
2. Schechter, I., Conrad, D.G., Hart, I., Berger, R.C., McKenzie, T.L., Bleskan, J., and Patterson, D. (1994). Localization of the squalene synthase gene (FDFT1) to human chromosome 8p22-p23.1. Genomics *20*, 116–118.
3. Do, R., Kiss, R.S., Gaudet, D., and Engert, J.C. (2009). Squalene synthase: a critical enzyme in the cholesterol biosynthesis pathway. Clin. Genet. *75*, 19–29.
4. Jemal, M., and Ouyang, Z. (1998). Gas chromatography-mass spectrometric method for quantitative determination in human urine of dicarboxylic (dioic) acids produced in the body as a consequence of cholesterol biosynthesis inhibition. J. Chromatogr. B Biomed. Sci. Appl. *709*, 233–241.
5. Vaidya, S., Bostedor, R., Kurtz, M.M., Bergstrom, J.D., and Bansal, V.S. (1998). Massive production of farnesol-derived dicarboxylic acids in mice treated with the squalene synthase inhibitor zaragozic acid A. Arch. Biochem. Biophys. *355*, 84–92.
6. Bostedor, R.G., Karkas, J.D., Arison, B.H., Bansal, V.S., Vaidya, S., Germershausen, J.I., Kurtz, M.M., and Bergstrom, J.D. (1997). Farnesol-derived dicarboxylic acids in the urine of animals treated with zaragozic acid A or with farnesol. J. Biol. Chem. *272*, 9197–9203.
7. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
8. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. *9*, e1003709.
9. Joo, J.H., and Jetten, A.M. (2010). Molecular mechanisms involved in farnesol-induced apoptosis. Cancer Lett. *287*, 123–135.
10. Okamoto-Uchida, Y., Yu, R., Miyamura, N., Arima, N., Ishigami-Yuasa, M., Kagechika, H., Yoshida, S., Hosoya, T., Nawa, M., Kasama, T., et al. (2016). The mevalonate pathway regulates primitive streak formation via protein farnesylation. Sci. Rep. *6*, 37697.
11. Hottman, D.A., and Li, L. (2014). Protein prenylation and synaptic plasticity: implications for Alzheimer's disease. Mol. Neurobiol. *50*, 177–185.
12. Tozawa, R., Ishibashi, S., Osuga, J., Yagyu, H., Oka, T., Chen, Z., Ohashi, K., Perrey, S., Shionoiri, F., Yahagi, N., et al. (1999). Embryonic lethality and defective neural tube closure in mice lacking squalene synthase. J. Biol. Chem. *274*, 30843–30848.
13. Tansey, T.R., and Shechter, I. (2000). Structure and regulation of mammalian squalene synthase. Biochim. Biophys. Acta *1529*, 49–62.

The American Journal of Human Genetics, Volume *103*

Supplemental Data

Squalene Synthase Deficiency: Clinical, Biochemical, and Molecular Characterization of a Defect in Cholesterol Biosynthesis

David Coman, Lisenka E.L.M. Vissers, Lisa G. Riley, Michael P. Kwint, Roxanna Hauck, Janet Koster, Sinje Geuer, Sarah Hopkins, Barbra Hallinan, Larry Sweetman, Udo F.H. Engelke, T. Andrew Burrow, John Cardinal, James McGill, Anita Inwood, Christine Gurnsey, Hans R. Waterham, John Christodoulou, Ron A. Wevers, and James Pitt

Supplemental Notes

Supplemental Case Reports

This research was approved by the Sydney Children's Hospitals Network Human Research Ethics Committee (HREC ID 10/CHW/114). Written informed consent was obtained from the parents of the affected individuals investigated in this study.

Individual 1 and 2:

These siblings are the product of non-consanguineous parents of European background. The salient features common to both include; intrauterine growth restriction (IUGR), facial dysmorphism, generalised seizures, hypoplastic corpus callosum, reduced age appropriate myelination, bilateral optic nerve hypoplasia with visual impairment, profound global developmental delay, irritability, post-natal failure to thrive, poor sleep initiation and maintenance, dry skin with photosensitivity, and gracile bones.

Urine organic acid profiles in both siblings consistently demonstrated an abnormal pattern of elevated mevalonate lactone, methylsuccinate and multiple branched-chain dicarboxylic acids which bore similarity to the patterns of animals treated with squalene synthase inhibitor drugs (5 samples in subject 1 and 4 samples in subject 4). Fasting cholesterol studies are summarized in Table 1, and consistently demonstrated low normal total cholesterol (TC) levels and reduced low-density lipoprotein cholesterol (LDL-C).

Individual 1

This female infant was delivered via a spontaneous vaginal delivery (SVD) at 37 weeks gestation after an uneventful pregnancy, with Apgar scores of 9 and 9 at one and five minutes respectively, a birth weight 2600g, head circumference 32.5cm, and length 47cm.

She developed generalised tonic-colonic seizures at five days of age, which were well controlled with phenobarbitone. Seizures have not been evident beyond the infantile period. An MRI of the brain at 9 months of age demonstrated a hypoplastic corpus callosum, and reduced white matter. Optic nerve hypoplasia was detected on fundoscopy and was associated with poor visual acuity. Anterior and posterior pituitary endocrine function were intact. Dysmorphic features included depressed nasal bridge, low set and posteriorly rotated ears, square nasal tip, epicanthic folds, mild micrognathia and mild retrognathia, and 2-3 toe syndactyly. Genitalia were normally formed.

Currently she is 10 years of age, and demonstrates profound global developmental delay in all modalities, in keeping with the milestone acquisition of a 6-month old child. She can roll and sit independently. She can manipulate objects to her mouth but is unable to feed herself. She is non-verbal, but does smile in response to verbal social interactions. Neurological examination demonstrates central hypotonia, hyperreflexia, and fixed flexion joint contractures at the knees. Irritability and autistic-like behaviors, such as habitual eye poking, have been evident since the first year of life. Hypersalivation has required medical interventions. Postnatal failure to thrive, compounded by an unsafe swallow, required the insertion of a PEG at 4 years of age for enteral feeding. Sleep initiation and maintenance are significant issues which have responded to melatonin supplementation. Dry skin and photosensitivity is evident with clinically significant UV sun damage after only 10 minutes in direct sunlight. A skeletal survey demonstrated thin gracile bones with reduced bone mineralization, but no skeletal axial fractures have been observed.

Individual 2

This male infant was delivered via a SVD at 39 weeks gestation after an uneventful pregnancy, with Apgar scores of 9 and 9 at one and five minutes respectively, a birth weight 2750g, head circumference 34cm, and length 49cm. His clinical course has mirrored that of his elder sister. He showed the same brain MRI changes, dysmorphic features, and bilateral undescended testes that required surgical intervention. He is currently 7 years of age, and demonstrates profound global developmental delay in all modalities, with an overall skill acquisition of a 9-month old child. He can sit independently, and pull to stand and will ambulate short distances on his knees. He can babble, smiles in response to verbal social interactions, but does not display meaningful social language.

Individual 3

This male infant was delivered at 39 weeks gestation after an uneventful pregnancy, with Apgar scores of 9 and 9 at one and five minutes respectively, a birth weight 3033grams, head circumference 33cm, and length 50cm. His clinical presentation in infancy was significant for severe static encephalopathy, intractable epilepsy, spasticity, opisthotonos, global developmental delay with cortical blindness, hypospadias, bicuspid aortic valve, neonatal hepatitis, dry skin, facial dysmorphism, failure to thrive and sleep dysfunction. Dysmorphism included coarse facial features, bitemporal narrowing, prominent ears, and triangular facies. MRI of the brain at 3 months of age demonstrated diffuse polymicrogyria involving the frontal, parietal, and temporal lobes with some central white matter and cortical volume loss. EEG demonstrated multifocal epileptiform discharges, but seizure control was high doses of levetiracetam, with only occasional breakthrough seizures during illnesses. He is

now 9 years of age and demonstrates severe global developmental delay, he attempts to sit and vocalizes without formation of clear words.

Supplemental Figures

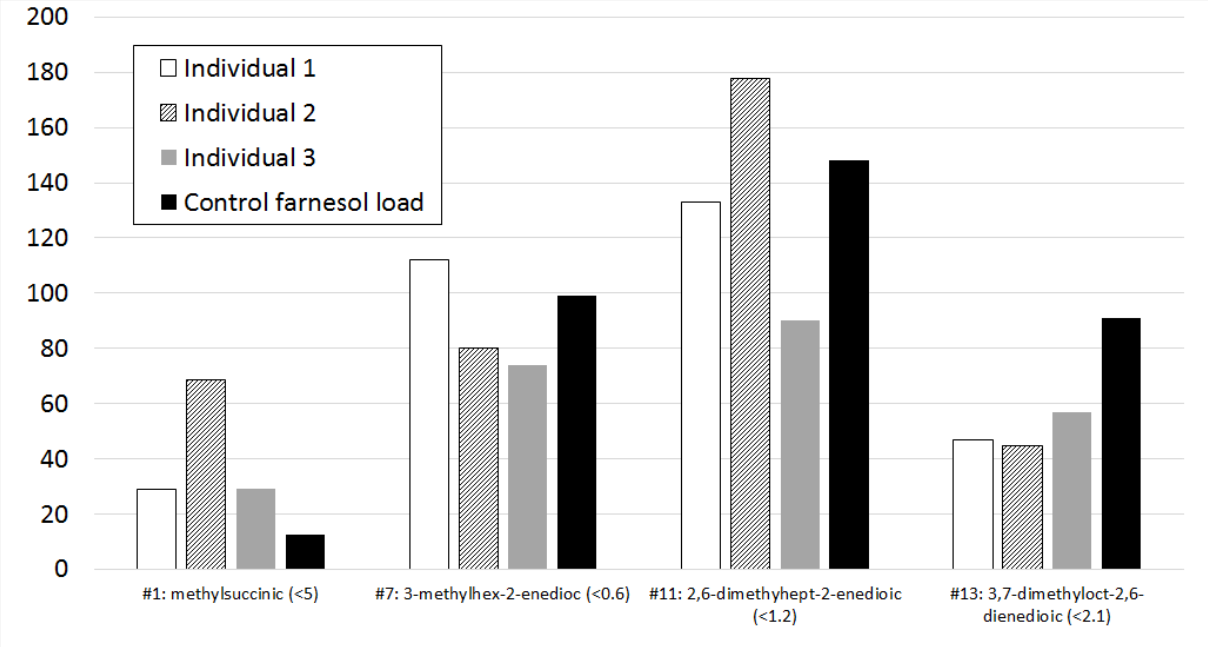
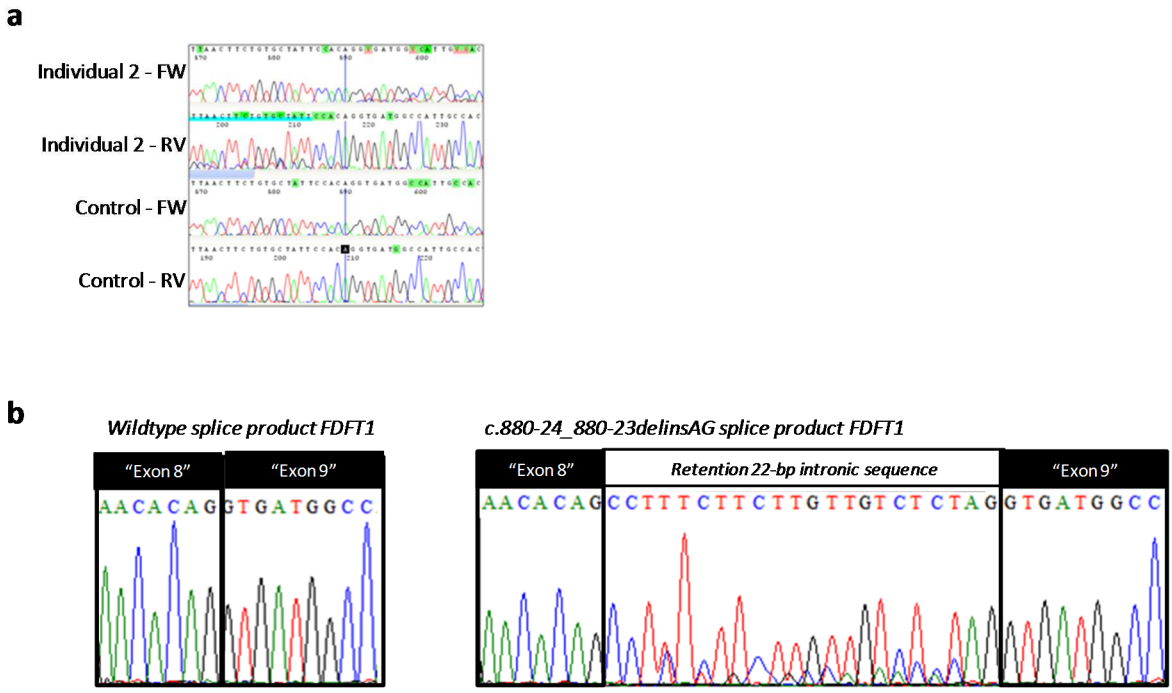


Figure S1: Urine metabolite concentrations in individuals and in a control after oral farnesol loading (urine collected 4.25 hours post load). Concentrations of the major dicarboxylic farnesol metabolites are in µmol/mmol creatinine, numbers refer to peak labels in Figure 1a and values in brackets are levels in unloaded controls.



<u>Figure S2:</u> Transcriptional consequences of paternal FDFT1 splice site mutation in Individuals 1 and 2. (**a**) PCR results using *FDFT1* cDNA generated from Individual 2, with primers in exon 8 and exon 9. Top two lanes show the forward and reverse electropherograms obtained using standard Sanger sequencing. Given the deletion on the maternal allele, cDNA product is only obtained from the paternal allele. The c.880-24_880-23delinsAG variant is predicted to lead to aberrant splicing, characterized by the retention of 22 bases of intron 8 sequence. This aberrantly spliced product will result in a premature termination codon in exon 9, and is therefore expected to be degraded by nonsense mediated mRNA decay (NMD). With the wt splice acceptor still intact, it is expected that normally spliced *FDFT1* can also be detected. cDNA analysis of individual 2 indeed shows the normal splice product (from the wt acceptor splice site) as well as a lower, 2nd trace reflecting the residual aberrant splice product from incomplete NMD. (**b**) To provide further functional proof that the c.880-24_c.880-23delinsAG variant as observed in Individuals 1 and 2, results in aberrant splicing, a mini-gene splice assay was set up, using sequence of exon 8, intron 8 and exon 9. In a wildtype scenario (left), exon 8 and exon 9 are spliced together. Targeted mutagenesis to introduce the variant shows that indeed aberrant splicing occurs, as is shown by the retention of 22 bases of intron 8 sequence.

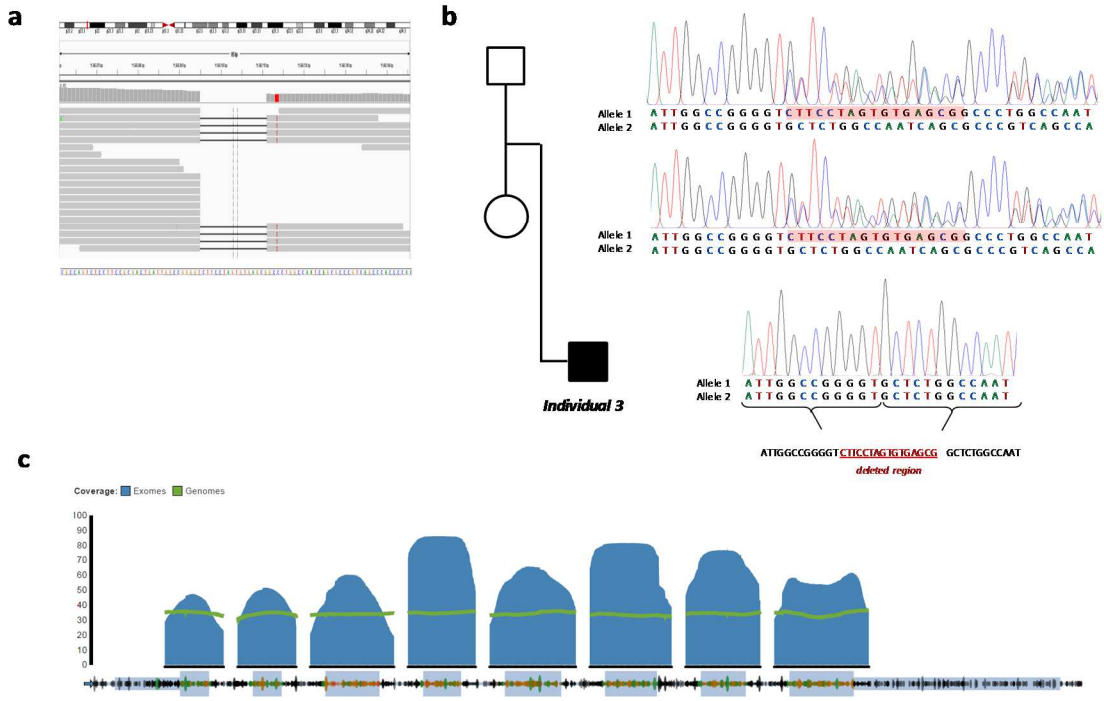


Figure S3: Screenshot of IGV displaying the BAM file from individual 3, demonstrating the coverage for this region, as well as the homozygous deletion identified in 8 reads. (**b**) Sanger validation of 16 bp deletion in Individual 3 and his parents. The deleted sequence is highlighted in red. (**c**) Screenshot from gnomAD, displaying *FDFT1* coverage information from both genomes (green) and exomes (blue), indicating overall sufficient coverage for *FDFT1* and the detection of the 16 bp deletion, if it were present in data captured in gnomAD.

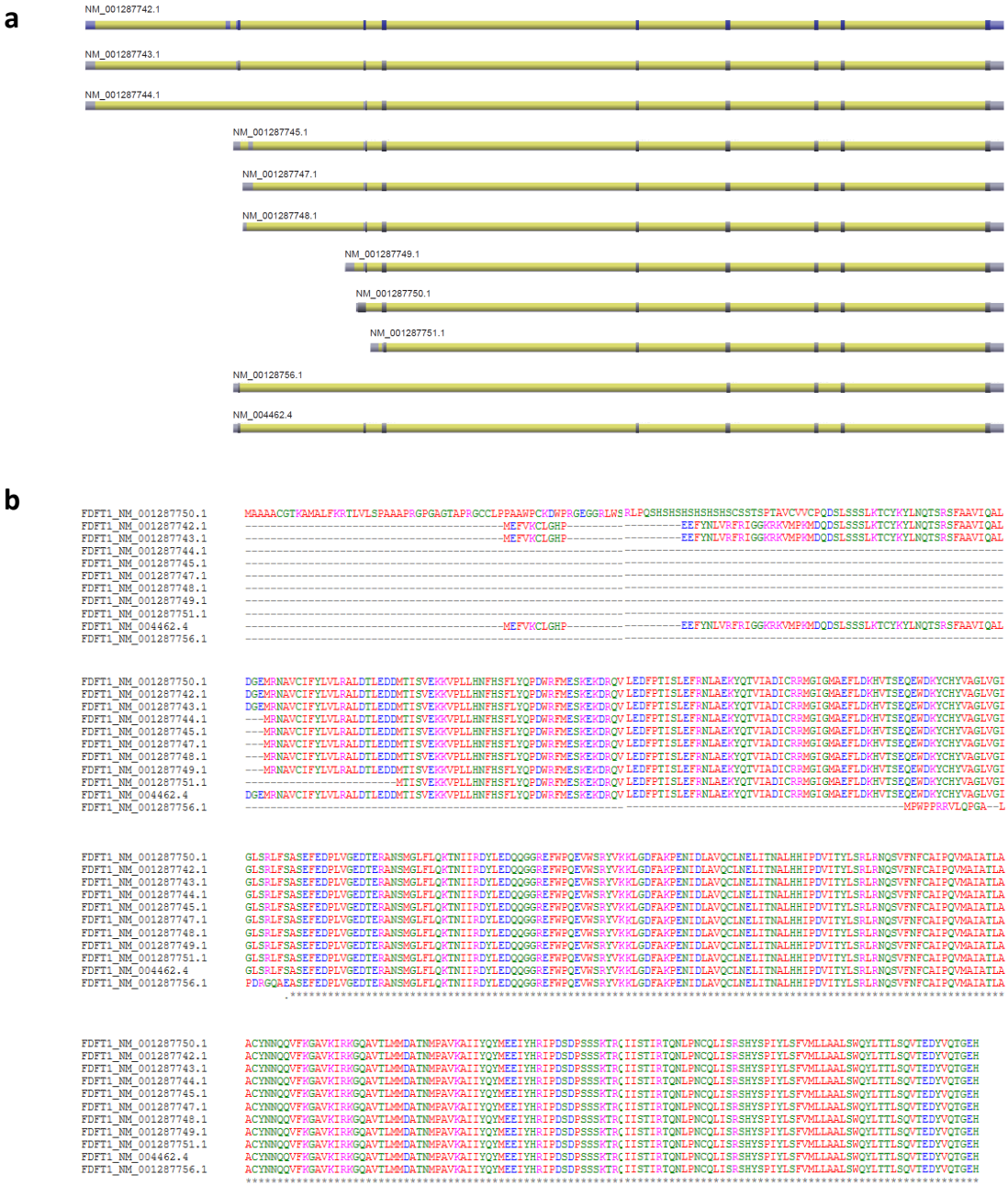
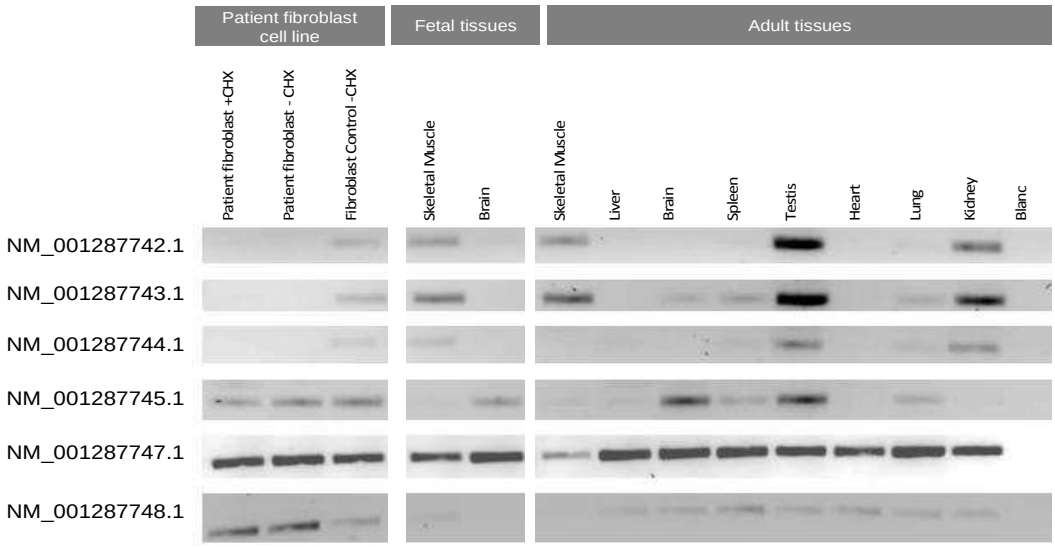


Figure S4: Schematic representation of FDFT1 and protein isoforms. (a) Overview of FDFT1 transcript isoforms based on Ref Seq. In total, 11 different transcripts for FDFT1 are reported, with the main differences being at the 5' end. Coding exons are shaded in dark grey, whereas non-coding exonic sequence is shaded in lighter grey. Intronic sequence is shown in yellow. (**b**) The 11 transcript FDFT1 isoforms result in proteins of five lengths, with the main difference being the N-terminus of the protein:

1) 476 amino acid (aa) protein is encoded by NM_01287750.1;
2) 417 aa isoform by NM_01287742.1, NM_01287743.1 and NM_04462.4;
3) 353 aa isoform by NM_01287744.1, NM_01287745.1, NM_01287747.1, NM_01287748.1 and NM_01287749.1;
4) 332 aa isoform by NM_01287750.1
5) 250 aa isoform by NM_01287756.1

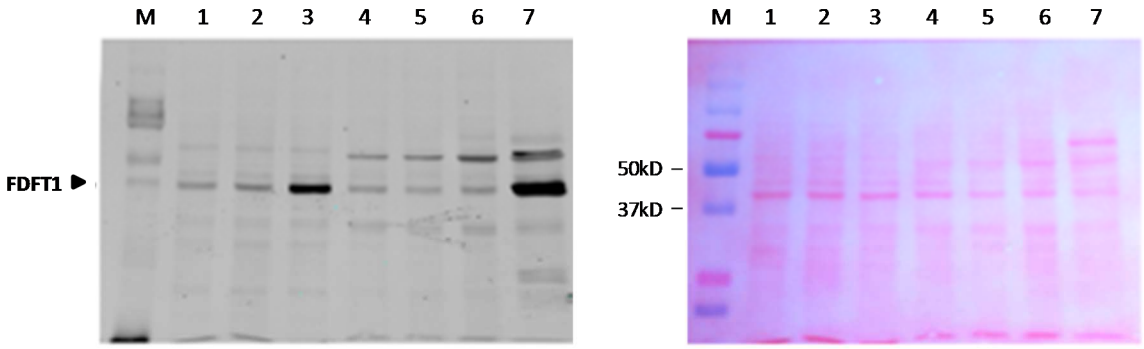
a



b

	Patient fibroblast			Fetal tissues		Adult tissues								
Isoform	**Patient fibroblast + CHX**	**Patient fibroblast - CHX**	**Control fibroblast – CHX**	**Skeletal muscle**	**Brain**	**Skeletal muscle**	**Liver**	**Brain**	**Spleen**	**Testis**	**Heart**	**Lung**	**Kidney**	**Blanc**
NM_001287742.1	-	-	+	+	-	+	-	-	-	+	-	-	+	-
NM_001287743.1	-	-	+	+	-	+	-	+	+	+	-	+	+	-
NM_001287744.1	-	-	+	+	-	-	-	-	-	+	-	+	+	-
NM_001287745.1	+	+	+	-	+	-	-	+	+	+	-	+	-	-
NM_001287747.1	+	+	+	+	+	+	+	+	+	+	+	+	+	-
NM_001287748.1	+	+	+	-	-	-	+	+	+	+	+	+	+	-
NM_001287749.1	+	+	+	+	+	-	+	+	+	+	+	+	+	-
NM_001287750.1	+	+	+	+	+	+	+	+	+	+	+	+	+	-
NM_001287751.1	+	+	+	+	+	+	+	+	+	+	+	+	+	-
NM_001287756.1/ NM_004462.4	+	+	+	+	+	+	+	+	+	+	+	+	+	-

<u>Figure S5:</u> FDFT1 isoforms, as well as detection of these isoforms in fibroblasts from Individual 3.(**a**) FDFT1 isoforms in fibroblasts of Individual 3 grown with cycloheximide (+CHX) and without cycloheximide (-CHX) to differentiate between absence of transcripts due to NMD or due to regulatory effects. Fibroblasts from a control individual were used to determine normal expression patterns of these isoforms in fibroblasts. Also, various organ specific cDNA libraries were tested for the normal expression patterns of these isoforms. (**b**) Simplified overview of the results presented in (a) using the following codes: +: detected; -: absent; Red: no isoform detection expected based on related control experiment; Yellow: tissues where the respective isoform is normally identified and likely absent in the patient.



Legend:

M Marker

1. Individual 2, lymphoblasts
2. Individual 1, lymphoblasts
3. Control a, lymphoblasts
4. Individual 3, fibroblasts
5. Control b, fibroblasts
6. Control c, fibroblasts
7. Control b + FDFT1 overexpression

Figure S6: Western blot analysis Individual 3. Western blot for FDFT1 protein levels detection in Individuals 1, 2 and 3 (left). Whereas a significant FDFT1 reduction can be shown in Individuals 1 and 2 (lanes 1 and 2, compared to lane 3), no significant reduction can be identified in Individual 3 (lane 4, compared to lanes 5 and 6). Right: Ponceau S staining for loading control.

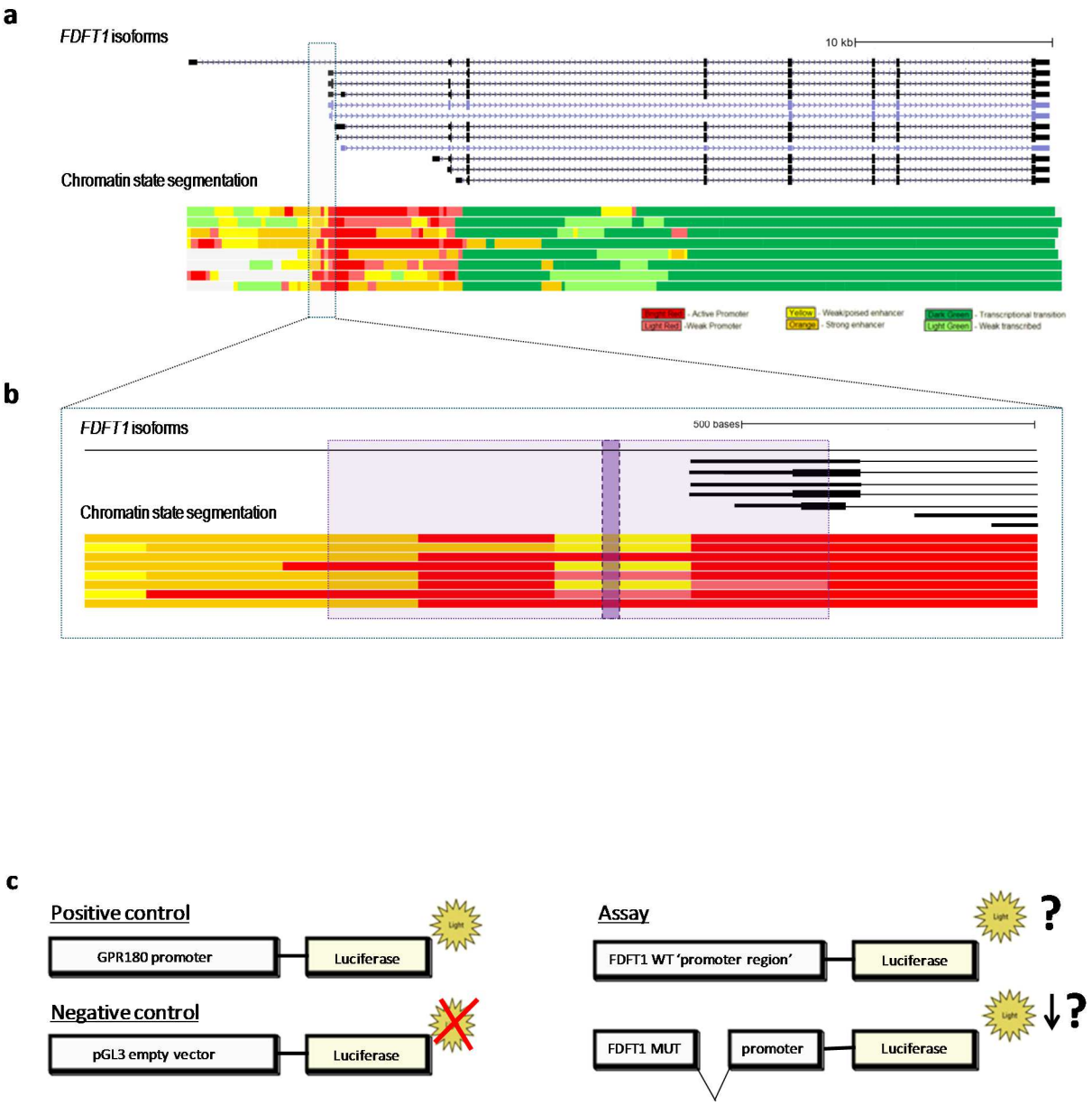


Figure S7: Schematic representation chromatin state FDFT1 isoforms and experimental set-up to assess its promoter activity.(**a**) Schematic representation of FDFT1 isoforms with annotation of chromatic state segmentation indicating that the region containing the deletion in patient 3 is predicted to have promoter (in shades of red) and/or enhancer (in shades of orange) activity based on mapping and analysis of human cell types[5,6]. (**b**) Zoom-in of the region containing the 16 bp deletion in Individual 3. The 1,024 bp-region assessed for promoter activity is highlighted in purple, with the mutation located indicated in darker purple. (**c**) Experimental set-up to test promoter activity of FDFT1. Results thereof are presented in Figure 4, and Table S1.

Supplemental Tables

Table S1: Raw and processed data of luciferase assay for assessing promoter activity of FDFT1 intronic sequence.

FDFT1 wildtype sequence								
	measurement	*Firefly Luciferase (raw data)*	*Renilla Luciferase (raw data)*	Normalization for transfection efficiency	standard deviation	Relative strength of promoter compared to empty vector	Average relative promoter strength	standard deviation
Biological replicate 1	1.1	5,253,639	18,853,700	0.28	0.01	26.11	**25.98** (100%)	**1.22** (4.68%)
	1.2	4,650,605	16,362,319	0.28		26.63		
	1.3	3,470,537	12,675,186	0.27		25.66		
Biological replicate 2	2.1	5,366,001	19,665,446	0.27	0.02	25.57		
	2.2	4,485,676	15,132,068	0.30		27.78		
	2.3	3,364,588	13,070,420	0.26		24.12		

FDFT1 mutant								
	measurement	*Firefly Luciferase (raw data)*	*Renilla Luciferase (raw data)*	Normalization for transfection efficiency	standard deviation	Relative strength of promoter compared to empty vector	Average relative promoter strength	standard deviation
Biological replicate 1	1.1	3,284,846	15,446,939	0.21	0.01	19.93	**19.78** (76%)	**0.90** (4.56%)
	1.2	4,340,171	20,590,486	0.21		19.75		
	1.3	5,530,909	24,804,902	0.22		20.89		
Biological replicate 2	2.1	3,254,900	16,763,116	0.19	0.01	18.20		
	2.2	4,440,535	21,228,426	0.21		19.60		
	2.3	5,490,335	25,354,536	0.22		20.29		

GPR180 promoter (positive control)								
	measurement	*Firefly Luciferase (raw data)*	*Renilla Luciferase (raw data)*	Normalization for transfection efficiency	standard deviation	Relative strength of promoter compared to empty vector	Average relative promoter strength	standard deviation
Biological replicate 1	1.1	320,052	9,949,642	0.03	1.66E-04	3.01	3.01	0.11
	1.2	252,404	7,858,441	0.03		3.01		
	1.3	296,761	9,151,764	0.03		3.04		
Biological replicate 2	2.1	342,483	10,659,048	0.03	1.84E-03	3.01		
	2.2	204,668	6,801,323	0.03		2.82		
	2.3	286,700	8,488,724	0.03		3.16		

pGL3 empty vector (negative control)								
	measurement	*Firefly Luciferase (raw data)*	*Renilla Luciferase (raw data)*	Normalization for transfection efficiency	standard deviation	Relative strength of promoter compared to empty vector	Average relative promoter strength	standard deviation
Biological replicate 1	1.1	328,714	30,048,130	0.01	1.13E-03	1.03	1.00	0.09
	1.2	391,524	32,767,584	0.01		1.12		
	1.3	266,326	27,482,966	0.01		0.91		
Biological replicate 2	2.1	315,404	30,272,910	0.01	1.08E-03	0.98		
	2.2	363,185	31,327,826	0.01		1.09		
	2.3	265,387	28,117,710	0.01		0.88		

None Transfected			
	measurement	*Firefly Luciferase (raw data)*	*Renilla Luciferase (raw data)*
Biological replicate 1	1.1	186	172
	1.2	135	287
	1.3	196	n.d.
Biological replicate 2	2.1	-	-
	2.2	-	-
	2.3	-	-

Supplemental Methods

Urine organic characterization GC-MS and NMR

Urine organic acids. Urine organic acid screening was performed by standard methodology involving solvent extraction, conversion to trimethylsilyl derivatives and analysis by GC-MS[1].

Plasma total farnesol and squalene measurements. Two hundred µL plasma was mixed with 200 µL of 100 mM Tris buffer pH 9.5 containing 100 mM sodium chloride and 50 mM magnesium chloride. Ten µL of alkaline phosphatase (Sigma P4439, 0.2 units/µL) was added and the mixture incubated for 15 hours at 37^{o}C. Absolute ethanol (800 µL) and 10 M sodium hydroxide (200 µL) was added and the mixture heated at 80^{o}C for 1 hour. After cooling, 1 mL of n-hexane containing paraffin mix (Supelco 4-7106 2% in n-octane, 50 µL in 50 mL of n-hexane) was added followed by vortexing and centrifuging. The upper n-hexane layer was transferred and dried under a nitrogen stream at room temperature. Pyridine (50 µL) and *N*-*tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide containing 1% *tert*-butyldimethylchloro-silane (50 µL) was added followed by incubation at 80^{o}C for 30 mins followed by addition of 150 µL of iso-octane. Samples were analysed on an Agilent 5973 GC-MS fitted with a 30 m HP-5MS column, 0.25 mm ID, 0.25 µm film. Splitless 1 µL injection were performed and the oven temperature was held at 150^{o}C for 1 min, then ramped to 320^{o}C at 10^{o}C/min and held for 2 minutes. Ions 279 (farnesol), 69 (squalene) and 71 *m/z* (C22 internal standard) were measured by selected ion monitoring. Calibrators were normal plasma spiked with free farnesol and squalene. The efficiency of alkaline phosphatase conversion of farnesyl- pyrophosphate to farnesol was checked by measuring a plasma sample spiked with farnesyl-pyrophosphate (77 and 117% recovery as farnesol).

Urine NMR. Urine of the three individuals with squalene synthase deficiency was measured with one-dimensional- and 2-dimensional COSY proton NMR spectroscopy at 500 MHz at pH 2.50 via previously described methods[2,3]

Genetic analysis

DNA isolation. Genomic DNA was extracted from whole blood using QIAamp DNA Blood Mini Kit (Qiagen, Cat# 69506), following manufacturer's specifications. Genomic DNA (gDNA) was subjected to agarose gel and OD ratio tests to confirm the purity and concentration.

Genome build. For mapping and annotation purposes referred to in this study, genome Reference Consortium Human Build 37, GRCh37, also known as hg19, was used.

Genomic microarray analysis. Genome-wide array analysis was performed by Mater Pathology Queensland using the SNP-based Cytoscan 750K array (Affymetix), with an effective resolution of 200 kb for copy number variation and 5 Mb resolution for contiguous stretches of homozygosity. Software analysis was performed using Chromosome Analysis Suite (ChAS) (Affymetrix).

Whole exome sequencing. WES was using DNA isolated from whole blood. For individuals 1 and 2, WES was performed by Otogenetics Corporation (Norcross, GA USA).) In essence, exome enrichment was based on Agilent SureSelect AV5 Exon Coverage (Agilent Technologies, Wilmington, DE USA, catalog# 5190-6208) following manufacturer's instructions, after which PE sequencing was performed with a read length of 100-106 nucleotidesat, covering 97.7% of targeted regions 100-fold and 99.7% 10-fold.

Analysis pipeline included mapping (hg19) with BWA-MEM and SAMtools, realignment and base calling with GATK Lite (2.3), variants annotation with snpEff and snpSift. The annotated

files were further analysed using Genegrid (Genomatix). Only deleterious variants with an allele frequency <0.01 according to the 1000 genomes phase 3 data set and the ExAC data set and shared by both affected siblings were considered candidates. For Individual 3 and his parents a Trio whole exome sequencing was performed commercially by Macrogen (Korea) using the Agilent SureSelect V5 capture kit, and run on an Illumina HiSeq4000 apparatus. The average read length was 101 bp and the mean depth of on-target reads was 92X. Subsequently, variants were filtered for a minor allele frequency <0.1% in ExAC and our in-house database containing 7,788 exomes, and for a strong effect on protein level (e.g. nonsense, missense, splice site and indels). Notably, after analyzing WES data of Individuals 1 and 2, pointing towards the involvement of *FDFT1*, all variants in this gene identified were searched for, including those in intronic sequences of *FDFT1* captured by WES. The latter filtering identified the private 16-bp homozygous intronic deletion. Due to its proximity to the exon, the intronic variant sequence was captured by the exome capture at a coverage over 10 reads.

Functional validation

Splice site variant analysis. Exon 9 of *FDFT1*, with flanking exon/intron boundaries, was PCR amplified from control gDNA using primers 5´-AGCAATTGCCCATTCAACAGA-3´ and 5´-ACCTGGTTAAACAGTGACATTACTA-3´. PCR product was cloned into pCR2.1-TOPO for screening and then subcloned into the pEF-1α promoter of p.EZe1.1, to create a wild-type FDFT1 clone (pEZe1.1/WT FDFT1 Ex9). The *FDFT1* c.880-24_880-23delinsAG variant was introduced using site-directed mutagenesis (QuikChange II kit, Agilent) according to the manufacturer's instructions to produce pEZe1.1/MUT FDFT1 Ex9. Clones were confirmed by Sanger sequencing. pEZe1.1/WT FDFT1 Ex9, pEZe1.1/MUT FDFT1 Ex9 and pEZe1.1 were

individually transfected into HEK293 cells using Lipofectamine 3000 (Life Technologies) according to the manufacturer's instructions. Cells were harvested 24h post-transfection and RNA extracted using an RNeasy mini kit (Qiagen). cDNA synthesis was performed on 2 µg RNA using SuperScript III reverse transcriptase (Life Technologies). Primers targeted to pEZe1.1 were used to amplify the cDNA region containing the spliced FDFT1 exon 9. Products were analysed by 1.5% (w/v) agarose gel electrophoresis and Sanger sequencing (AGRF).

To assess the effect of the c.880-24_880-23delinsAG mutation at cDNA level in individual 1 and 2, fibroblasts were grown and harvested for mRNA extraction according to standard procedures. A similar procedure was followed using control individuals. Primers designed for the full length of *FDFT1* cDNA (1441 bp) failed to produce a product, after which primers were designed to amplify part of the cDNA (993 bp). The relevant PCR products were isolated from gel and purified, after which they were sent for Sanger sequencing.

Western Blot. Immunoblotting and densitometry were performed as previously described[4], with the following modifications: membranes were probed with a 1:1000 dilution of anti-FDFT1 (Proteintech Group 13128-1-AP) or a 1:5000 dilution of anti-α-tubulin (Sigma T6199), overnight at 4°C. Control (Ctrl) and affected individuals (Ind1, Ind2Ind1, Ind2) fibroblasts were grown in media supplemented with FBS or lipid-depleted FBS (LD-FBS).

(Tissue specific) FDFT1 isoform analysis. Fibroblast cell lines of individual 3 were grown in the absence (-) and presence (+) of cycloheximide (CHX) to inhibit nonsense mediated mRNA decay. Using isoform specific PCR primers, cDNA generated from the fibroblast cell lines of individual 3 and from a control individual was tested for the presence of 10 of 11 *FDFT1* isoforms. Unique primers to distinguish NM_001287756.1 from NM_004462.4 could not be

made. In addition, organ specific cDNA libraries from fetal and adult tissues were tested for the normal patterns of these isoforms. Obtained PCR products were sequenced to verify the isoform specificity.

Luciferase assay. Gateway-tailed PCR primers (FW: GGGGACAAGTTTGTACAAAAAA GCAGGCTTCCCAAAGTGTTGCGATTACA; RV: GGGGACCACTTTGTACAAGAAAGCTGGGTGATC TTGGGCATCACCTTCC) were used to amplify a 1024 bp genomic PCR product using DNA from a control and individual 3 to assess the promoter activity under normal condition as well as when containing the 16 bp deletion respectively. The promoter sequence was subsequently introduced into entry-vector pDONR201 following the manufacturer's instructions, and transformed into DH5alpha chemically competent cells. Colonies were grown overnight at 37°C after which routine miniprep was performed for plasmid DNA isolation. The DNA insert was sequenced to exclude genetic variation other than the 16-bp deletion. Both the WT 1024 bp sequence as well as the one containing the 16 bp deletion observed in Individual 3, were cloned into pGL3 to perform the luciferase assay. pGL3 containing GPR180 and its relevant transcription were used as technical positive control, whereas an empty pGL3 basic vector was used as negative control. All constructs were co-transfected with *Renilla* to determine transfection efficiency. Luciferase activity was measured from two biological replicates (100 ng per assay) using luminometry, in triplicate. Relative promoter activity was determined by correcting for transfection efficiency, and normalization against the activity of the empty pGL3 vector.

Supplemental References

1. Pitt JJ, Peters H, Boneh A, Yaplito-Lee J, Wieser S, Hinderhofer K, Johnson D, Zschocke J. (2015). Mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase deficiency: urinary

organic acid profiles and expanded spectrum of mutations. J Inherit Metab Dis. 38:459-66.

2. Bostedor RG, Karkas JD, Arison BH, Bansal VS, Vaidya S, Germershausen JI, Kurtz MM, Bergstrom JD. (1997). Farnesol-derived dicarboxylic acids in the urine of animals treated with zaragozic acid A or with farnesol. J Biol Chem 272:9197-203.
3. Wevers, R.A. Engelke UF, Moolenaar SH, Bräutigam C, de Jong JG, Duran R, de Abreu RA, van Gennip AH. (1999). 1H-NMR spectroscopy of body fluids: inborn errors of purine and pyrimidine metabolism. Clin Chem 45:539-48.
4. Riley LG, Menezes MJ, Rudinger-Thirion J, Duff R, de Lonlay P, Rotig A, Tchan MC, Davis M, Cooper ST, Christodoulou J. (2013). Phenotypic variability and identification of novel YARS2 mutations in YARS2 mitochondrial myopathy, lactic acidosis and sideroblastic anaemia. Orphanet J Rare Dis. 8:193.
5. Ernst, J. & Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol 28:817-25.
6. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al., (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473(7345):43-9.