SUPPLEMENTARY MATERIAL FOR:

P5CS expression study in a new family with *ALDH18A1*-associated hereditary spastic paraplegia SPG9

Pamela Magini[1,*], Clara Marco-Marín[2, 3, *], Juan M Escamilla-Honrubia[2, 3], Diego Martinelli[4], Carlo Dionisi Vici[4], Francesca Faravelli[5], Francesca Forzano[6], Marco Seri[1, 7, a], Vicente Rubio[2, 3, a], Emanuele Panza[1, 7, a].

[1]Medical Genetics Unit, S. Orsola-Malpighi Hospital, Bologna, Italy

[2] Instituto de Biomedicina de Valencia of the CSIC, Valencia, Spain

[3] Group 739, Centro para Investigación Biomédica en Red sobre Enfermedades Raras CIBERER-ISCIII, Valencia, Spain;

[4] Division of Metabolism, Bambino Gesù Children's Research Hospital, Rome, Italy;

[5] Clinical Genetics, NE Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK;

[6] Medical Genetics Unit, University of Genova, Italy.

[7] Department of Medical and Surgical Sciences, University of Bologna, Italy.

*Magini and Marco-Marín contributed equally to this work

[a] Corresponding Authors:

Marco Seri, MD and Emanuele Panza, PhD
S. Orsola-Malpighi University Hospital
Department of Medical and Surgical Sciences
University of Bologna, Italy
E-mails: marco.seri@unibo.it emanuele.panza@unibo.it

Vicente Rubio, MD, PhD
Instituto de Biomedicina de Valencia (IBV-CSIC) and CIBERER-ISCIII
Jaime Roig 11
Valencia-46010
Email: rubio@ibv.csic.es
Phone: +34 963391772

This supplementary material is composed of:

- Supplementary Methods
- Supplementary Results
- Supplementary References
- Supplementary Table 1
- Supplementary Figure 1
- Supplementary Figure 2

SUPPLEMENTARY METHODS

Human P5CS production

The coding sequence for the mature long form of human P5CS (lacking the mitochondrial targeting sequence, residues 1-56[1]) was PCR-amplified (see Supplementaty Table 1 for the oligonucleotides used) from commercial pCR4-TOPO-P5CS-transformed *E. coli* cells (product MHS6278-211687672 from Open Biosystems/Dharmacon/Horizon Discovery Ltd., Cambridge UK). The product was ligation-free cloned into pOPINM (from Clontech, Madrid, Spain) (Supplementary Fig. 1), and the resulting vector was subjected to site directed mutagenesis (Quickchange System, Stratagene, La Jolla, CA) using appropriate oligonucleotides (Supplementary Table 1) to remove codons 238-239, to render P5CS sensitive to ornithine inhibition[1]. To introduce the appropriate mutation this vector was used for a second round of site-directed mutagenesis as above with appropriate oligonucleotide pairs (Supplementary Table 1). Sf9 insect cells were co-transfected with equal amounts of the wild type or mutant vector (Supplementary Fig. 1) and of the linearized bacmid DH10:KO1629, to generate the engineered baculovirus[2,3]. Sanger sequencing corroborated the correctness and fidelity of the constructions. Protein production within the cells is schematised in Fig. 1B of the main text, essentially as done previously for CPS1[4].

For wild-type or mutant P5CS purification (4ºC), a 200-ml insect cell culture was centrifuged 15 min (2500 × g), the pellet was suspended in 15 ml of buffer A (50 mM HEPES pH 7.5, 0.2 M NaCl, 1 mM EDTA, 1 mM dithiothreitol) supplemented with 0.25% Nonidet P-40, 0.1 mM phenyl methyl sulphonyl fluoride (PMSF), and 10 μM each pepstatin and leupeptin. This suspension was homogeneized (electrically operated Potter-Elvehjem glass/teflon homogeneizer followed by brief sonication), centrifuged (30 min, 15000 × g), the supernatant was syringe-filtered (0.45 μm pore)

and applied to a 1-ml MBPTrap HP column (GE Healthcare, Barcelona, Spain) equilibrated with buffer A. After a 15-ml wash the outlet of this column was connected to the inlet of a buffer A-equilibrated 1-ml Histrap HP column (GE Healthcare), and 3 ml of buffer A containing 10 mM maltose were applied to the columns. Then the tandem was disassembled and the HisTrap HP column was washed with 15 ml of buffer A followed by application of 3 ml buffer A containing 125 mM imidazole to elute doubly-tagged P5CS. The effluent from this last 3-ml application was supplemented with 10 µg/ml PreScission protease (GE Healthcare), followed by overnight incubation at 4ºC, before application to the regenerated MBPTrap HP column to retain MBP and MBP-tagged P5CS. Untagged P5CS was collected in the flow-through and a 1-ml column wash of buffer A. It was concentrated (centrifugal ultrafiltration, Amicon Ultra 50K; Merck-Millipore, Burlington, USA) to ~2 mg protein/ml, 25% glycerol was added, and the mixture was aliquoted and stored at -80ºC. SDS-PAGE analysis of the protein was done in 10% polyacrylamide gels[5], with Coomassie staining. P5CS purity was estimated densitometrically from these gels (MultiGauge program V2.1 from Fujifilm).

Enzyme activity assays.

All assays were carried out at least in duplicate, and they were repeated on separate days with different enzyme preparations. Results are given as units/mg of pure P5CS. One unit is one µmol NADPH (P5CS and G5PR reaction) or NADH (G5K reaction) oxidized per min. The assay solution used for the P5CS global reaction contained 50 mM HEPES pH 7.5, 10 mM ATP, 40 mM $MgCl_2$, 0.8 M Na L-glutamate, 0.2 mM NADPH, 0.1 M KCl, 5 mM K-phosphoenolpyruvate, 0.04 mg/ml pyruvate kinase (from rabbit muscle) and 1 mg/ml bovine serum albumin. The same solution was used for the G5K reaction except for the replacement of the NADPH by NADH and for the addition of 0.03 mg/ml lactate dehydrogenase (from rabbit muscle).

The assay mixture for the G5PR partial reaction was the same as for the global P5CS reaction except for the lowering of the L-glutamate concentration to 0.2 M and for the addition of 15 µg/ml *E. coli* G5K (EcG5K)[6]. At 0.2 M glutamate, production of G5P by the G5K partial reaction of P5CS is virtually nil (data not shown) whereas EcG5K is essentially saturated by this substrate[7]. In this way, G5P is produced in situ by the EcG5P added. Since G5P is unstable[8], a steady state level of this compound should be attained when produced by EcG5K. This level should be directly proportional to the

concentration of G5K, according to the expression [G5P] = k × $t_{1/2}$ × [EcG5K], where k is a constant and $t_{1/2}$ is the half-life of G5P (not known but believed to be quite low[8]), a value that should also be constant at a given temperature and pH. Thus, the Michaelis-Menten equation for the dependency of the velocity on the concentration of G5P, v = V_{max} × [G5P] / (K_m + [G5P]), can be rewritten to relate the velocity to the concentration of EcG5K, v = V_{max} × [EcG5K] / (K'_m + [EcG5K]); where K'_m= K_m / (k × $t_{1/2}$). Therefore, the G5PR activity of P5CS should be a hyperbolic function of the concentration of exogenously added EcG5K. Thus, when determining substate kinetics for the G5PR reaction either the EcG5K or the NADPH concentrations were varied as indicated.

Thermofluor assays

They were carried out in sealed microwell plates in a 7500 real-time PCR instrument from Applied Biosystems (Foster City, CA), with excitation at 488 nm, measuring the increase in SYPRO Orange fluorescence (emission monitored at 610 nm) with increasing temperature (1º C /min) of 20 µl of a solution of 0.2 mg/ml P5CS (wild-type or mutant) in 50 mM HEPES pH 7.5, 0.2 M NaCl, 1 mM dithiothreitol, 25 % glycerol and 1:1000 dilution of the commercial preparation of SYPRO Orange (Invitrogen, Carlsbad, CA).

DNA Sequencing studies.

DNA for sequencing studies was obtained from peripheral blood by standard methods. For WES, Illumina TruSeq technology (library preparation) and the HiSeq2000 sequencing instrument were used. Reads were checked with FastQC (http://www.bioinformatics.babraham.ac.uk/publications.html), aligned to the h19 reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.38) with BWA[9], utilizing GATK[10] for realignment and base quality score recalibration, and utilizing PicardTools (http://picartools.sourceforge.net) for duplicate removal. SAMtools[9] and GATK were used for collecting alignment statistics, and GATK was employed to calculate coverage statistics over the targeted regions and for variant calling and filtering by quality, prior to annotation against Ensembl (http://www.ensembl.org) of the variants passing quality filters. Mean redundancy of sequence coverage was 74× while 89% of the bases were covered >20×. To identify significant variants of potential pathological relevance we filtered out the changes found in 5'- and 3' untraslated regions, those changes corresponding to synonymous base

substitutions, and the variants already reported in public databases with >1% frequency or in homozygous (autosomes) or hemyzygous (sex chromosomes) individuals. The databases used were GnomAD (http://gnomad.broadinstitute.org), ExAC (http://exac.broadinstitute.org), EVS (http://evs.gs.washington.edu/EVS) and 1000 Genomes Browser (http://browser.1000genomes.org/index.html). We also excluded from further analyses variants observed in polymorphic genes (such as mucine and taste receptor genes) and those found in high frequency in our in-house exome database.

Standard automated Sanger sequencing was used to confirm *ALDH18A1* variants in patients and family members, utilizing the BigDye Terminator v1.1 Cycle Sequencing Kit and an ABI 3100 Sequencer (ThermoFisher Scientific).

Genomic positions are those of GRCh37/hg19 genome build. Reference sequences for the gene, the cDNA and the protein for *ALDH18A1* correspond, respectively, to ENSG00000059573, ENST00000371224 and ENSP00000360268 (http://www.ensembl.org/).

Other methods.

Protein concentrations were determined according to Bradford[11] using a commercial reagent (Bio-Rad, Madrid, Spain) and bovine serum albumin as standard. Kinetic data were fitted to hyperbolae using GraphPad Prism software (San Diego, CA). In silico tools used for assessing base and amino acid conservation and for predicting mutation effects included PhyloP[12], Clustal omega[13], Human Splicing Finder server[14] (http://www.umd.be/HSF3), Polyphen-2[15], MutPred2[16] and Mutation Taster 2[17]. Pymol (DeLano Scientific; http://www.pymol.org) was used for visual structural analysis, for depicting the structure of the G5PR component of human P5CS, and for structural superimpositions involving ALDH[18] and aminoadipate dehydrogenase[19].

SUPPLEMENTARY RESULTS

Clinical presentations

The two patients are the sons of healthy non-consanguineous Italian parents and they have a healthy sister (Fig. 2A of main text). There was no evidence of psychomotor retardation or congenital motor alterations in their extended family (grandparents, uncles and cousins and their offspring). In patient 1, now of 40 years of age, developmental delay became noticeable in the first year of life. Motor impairment was

progressive throughout his entire life, with stiffness, coordination deficits, abnormal gait and development of pes cavus. Since age 35 years the patient requires a walker or a wheelchair to move. At the age of 10 years detailed neuro-psychiatric evaluation documented intellectual disability (IQ 58), spastic paraparesis with muscular hypotrophy, athetoid movements of upper limbs and bilateral pes cavus. An electroencephalogram (EEG) detected widespread abnormal cerebropathic and irritative features. Psychiatric evaluation suggested chronic psychosis grafted on his organic mental condition. Symptomatic treatment with valproic acid and haloperidol led to the disappearance of the neuroirritative and psychiatric manifestations. Mitral valve prolapse was documented by cardiological examination. The patient had complained and continues to complain of mild swallowing difficulties. At 22 years of age physical examination showed hyposomia (weight, height and cranial circumference < 5th percentile), long face, prominent nasal root, prognathism, long brachial segment, long fingers and bilateral pes cavus, that was more marked in the left foot. At 32 years of age a new neurological examination showed exaggerated tendon reflexes in lower limbs, polykinetic patellar reflexes and inexhaustible ankle jerk reflexes, as well as Hoffman and Babinski signs. Spastic gait with hypertonia of all four limbs (more pronounced in the lower limbs), and distal muscular hypotrophy was recorded. Somatosensory and visual evoked potentials revealed significant anomalies of central conduction in all limbs, indicating an involvement of the sensory pathways. Brain NMR only identified an increase in the prominence of the cortical sulci. Karyotyping, X-fragile molecular test, FISH with subtelomeric probes and Sanger sequencing of the *SPG11* and *SPG15* genes exons and exon/intron boundaries yielded normal results. Whole exome sequencing (WES) was performed at the age of 37 years (see next section) and identified the *ALDH18A1* mutations, that were confirmed by Sanger sequencing (supplementary Fig. 2).

Patient 2 was the 2-year younger brother of patient 1 (Fig. 2A of the main text). He exhibited at 3 years of age mild gait disturbances and learning difficulties, with a slowly progressive course. At the age of 5 years an EEG detected irritative features leading to treatment with valproic acid and thioridazine. Five years later neurologic examination revealed intellectual disability (IQ 46), athetoid movements, mild left *pes cavus* and a clinical picture similar to that presented by his brother. EEG detected again irritative, mainly frontal anomalies, tending to spread over the other cortical fields in hyperpnea condition. At 18 years, when under stressful situations, the patient presented

events of tonic-clonic seizures that could be controlled pharmacologically. At 20 years of age physical examination evidenced microcephaly and relatively short stature (weight and cranial circumference < 5th centile, height 5-10th centile), long brachial segments, long fingers and toes, scoliosis, long face, prominent prognathism and strabismus. At 35 years his gait rapidly worsened and equilibrium deficits arose. A neurologic picture similar to that of sibling 1 was observed at 37 years. The Wechsler Adult Intelligence Scale (WAIS) test results were consistent with mild to moderate intellectual disability (Verbal IQ 63, Performance IQ 57, Total IQ 56). Karyotyping, X-fragile test, array comparative genetic hybridization (aCGH), and Sanger sequencing of the exons and exon/intron boundaries of genes *PLP1*, *SPAST* and *SPG7*, as well as serum and urine amino acid levels did not reveal pathologic findings. The diagnosis was based on automated Sanger DNA sequencing after mutations in the *ALDH18A1* gene were detected in both alleles of his brother (see next section).

Identification of compound heterozygosity for two *ALDH18A1* missense mutations predicted to be damaging.

The lack of manifestations in the parents and sister and their presence in the two male siblings (Fig. 2A of the main text), as well as the lack of family history of disabilities in the extended family (not shown) suggested recessive inheritance, although the exclusive affectation of both male sibs could also be consistent with X-linked inheritance. Failure to identify genetic anomalies in focused genetic studies in both patients led us to perform WES sequencing in the older sib (age at the moment of study, 37 years). We analyzed the findings focusing on low population frequency (<1%) variants present biallelically in autosomal genes of the patients or hemizygotically on chromosome X, with particular focus on variants not present in homozygous (autosomal genes) or hemizygous (X-chromosome) state in public databases. We found 21 possible compound heterozygous, 1 homozygous and 2 X-linked variants. Of the 13 genes affected by these changes, the only gene that had been previously associated with spastic paraparesis was *ALDH18A1*.

The mutations found in this gene leading to the protein changes p.(Arg371Gln) and p.(Ser497Asn) are described in the main text. Arg371 is invariant in all the vertebrate and plant sequences sampled, it is highly conserved (with some conservative replacements to lysine) in invertebrates, and is largely conserved in fungi, and in

bacteria and archaea, with most replacements being conservative (to histidine or lysine) (Fig. 2B of main text). Of the 95 sequences examined, only in one microbial sequence a glutamine was observed at this position. These comparisons indicate a very strong bias at this position against the glutamine found in the mutated form, and in favor of arginine or of another positively charged residue. Similarly, the alignment of 95 sequences (Fig. 2B and Table 1 of main text) reveals a strong bias for serine at the position corresponding to Ser497 of human P5CS, with invariance of the serine or conservative replacement by threonine, and with complete exclusion of the mutant Asn in the 95 sequences examined. This very high conservation is expected if Arg371 and Ser497 have important roles either for the structure of the function of P5CS, while the virtual and complete exclusion of glutamine and asparagine at one and the other positions strongly point to the causation of ill-effects by the two observed mutations. In agreement with their disease-causing role, both mutations cosegregated with the disease in the family tree as expected for recessive inheritance, with the parents being heterozygotes for one or the other mutation and the wild-type allele, and both affected siblings being compound heterozygotes for these two mutant alleles, while the healthy sister was homozygous for both normal alleles (Fig. 2A of the main text).

SUPPLEMENTARY REFERENCES

1. Hu CA, Lin WW, Valle D. Cloning, characterization, and expression of cDNAs encoding human delta 1-pyrroline-5-carboxylate dehydrogenase. J Biol Chem 1996; 271: 9795-9800.
2. Zhao Y, Chapman DA, Jones IM. Improving baculovirus recombination. Nucleic Acids Res 2003; 31: E6-6.
3. Berrow NS, Alderton D, Sainsbury S et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res 2007; 35: e45.
4. Diez-Fernandez C, Martínez AI, Pekkala S et al. Molecular characterization of carbamoyl-phosphate synthetase (CPS1) deficiency using human recombinant CPS1 as a key tool. Hum Mutat 2013; 34:1149-1159.
5. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970; 227:680-685.

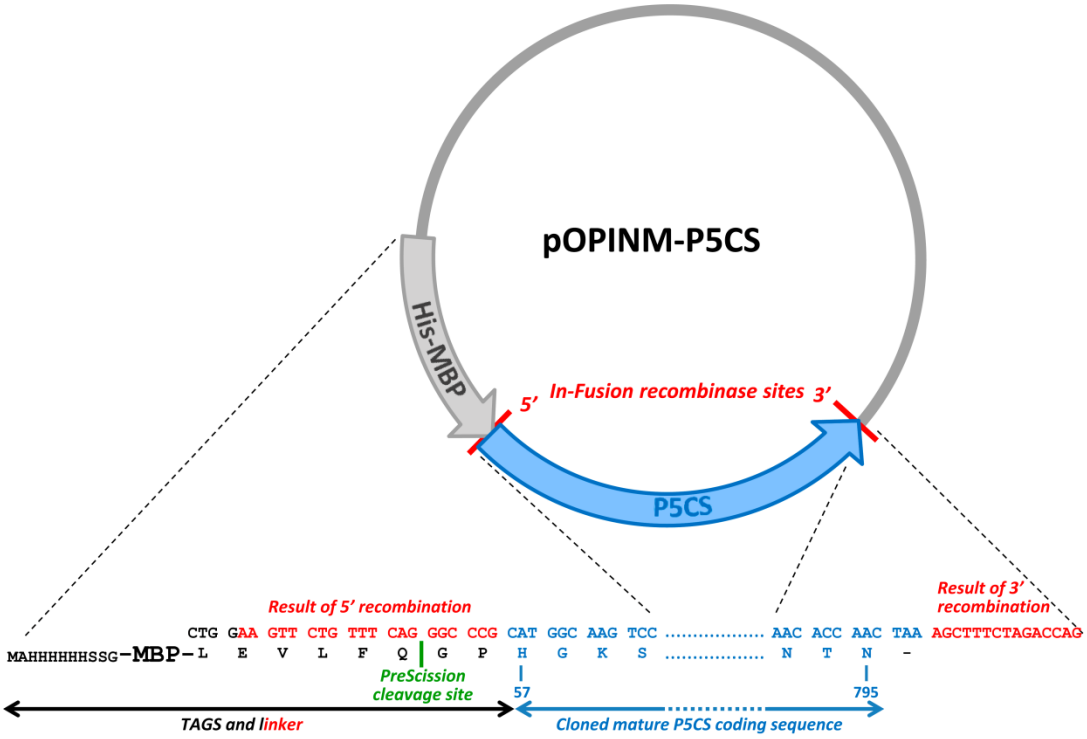
6. Pérez-Arellano I, Gil-Ortiz F, Cervera J, Rubio V. Glutamate-5-kinase from *Escherichia coli*: gene cloning, overexpression, purification and crystallization of the recombinant enzyme and preliminary X-ray studies. Acta Crystallogr D Biol Crystallogr. 2004;60:2091-2094
7. Pérez-Arellano I, Rubio V, Cervera J. Mapping active site residues in glutamate-5-kinase. The substrate glutamate and the feed-back inhibitor proline bind at overlapping sites. FEBS Lett. 2006; 580:6247-6253.
8. Seddon AP, Zhao KY, Meister A. Activation of glutamate by gamma-glutamate kinase: formation of gamma-cis-cycloglutamyl phosphate, an analog of gamma-glutamyl phosphate. J Biol Chem. 1989;264:11326-11335.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-1760.
10. DePristo MA, Banks E, Poplin R et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011; 43:491-498.
11. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976; 72:248-254.
12. Pollard KS, Hubisz MJ, Siepel A. Detection of non-neutral substitution rates on mammalian phylogenies. Genome Res 2010;20:110-121.
13. Sievers F, Wilm A, Dineen D et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011;7:539.
14. Desmet FO, Hamroun D, Lalande M et al.. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acid Research 2009;37:e67.
15. Adzhubei IA, Schmidt S, Peshkin L et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-249.
16. Pejaver V, Urresti J, Lugo-Martinez J et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv 2017;134981.
17. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014;11:361-362.
18. Liu ZJ, Sun YJ, Rose J et al. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nat Struct Biol 1997;4:317-326.

19. Luo M, Tanner JJ. Structural basis of substrate recognition by aldehyde dehydrogenase 7A1. Biochemistry 2015;54:5513-5522.

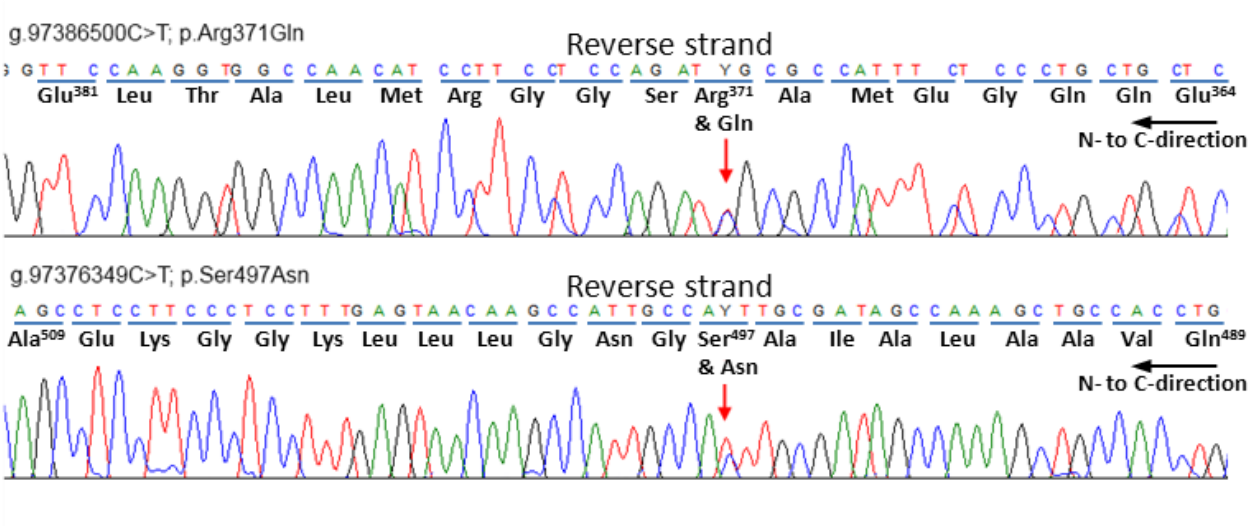
Supplementary Table 1. Synthetic oligonucleotides used in this work

Use	Name	Sense	Sequence (5'-3')
PCR-cloning of mature P5CS	P5CS-F	Forward	*AAGTTCTGTTTCAGGGCCCG*CATGGCAAGTCCTTCGCCCAC[a]
	P5CS-R	Reverse	*ATGGTCTAGAAAGCTTTA*GTTGGTGTTTCTGAGGAATAGGG[a]
Mutagenesis: deletion residues 238-239	Short-F	Forward	CCCAACAGTGACCTGCAGGG_GGTTATTAGTGTTAAAG[b]
	Short-R	Reverse	CTTTAACACTAATAACC_CCCTGCAGGTCACTGTTGGG[b]
Mutagenesis: p.Arg371Gln	R371Q-F	Forward	GGAGAAATGGCGC**A**ATCTGGAGGAAG[c]
	R371Q-R	Reverse	TCCTTCCTCCAGAT**T**GCGCCATTTCTC[c]
Mutagenesis: p.Ser497Asn	S497N-F	Forward	TGGCTATCGCAA**A**TGGCAATGGCTTG[c]
	S497N-R	Reverse	ACAAGCCATTGCCA**T**TTGCGATAGCC[c]

[a] P5CS coding sequence in normal type, cloning extensions in italics.
[b] The underlining marks the site of the two-codon deletion.
[c] Bold type marks base substitutions to introduce the desired mutations.



Supplementary Figure 1. Scheme of pOPINM-P5CS. Zooms, nucleotide or/and amino acid (in single letter notation) sequences corresponding to selected parts of the construct. MBP, maltose binding protein. The In-Fusion recombinase and the PreScission protease are commercial products from Clontech and GE Healthcare, respectively. The final protein product after PreScission cleavage should correspond to amino acids 57-795 of the short form of P5CS (residues 238 and 239 missing) preceded by two amino acids (Gly-Pro) from the linker.



Supplementary Figure 2. Validation of the two novel *ALDH18A1* mutations through Sanger sequencing in one of the two patients. Electropherograms for reverse strand are shown for both mutations. The two substitutions are indicated with red arrows.Y means the coexistence in heterozygosity of C and T. Genomic (GRCh37) and protein coordinates are reported. The horizontal arrows signal the direction of the protein sequence from N- to C-terminus.