Identification of a mutation in *LARS* as a novel cause of infantile hepatopathy

Jillian P Casey[a,b], Paul McGettigan[c], Niamh Lynam-Lennon[d], Michael McDermott[e], Regina Regan[a,b], Judith Conroy[a,b], Billy Bourke[a,b], Jacintha O' Sullivan[d], Ellen Crushell[f], SallyAnn Lynch[g,h], Sean Ennis[b,g]

[a]National Children's Research Centre, Our Lady's Children's Hospital, Crumlin, Dublin 12, Ireland (jillian.casey@ncrc.ie). [b]School of Medicine and Medical Science, University College Dublin, Belfield, Dublin 4, Ireland (regina.regan@ucd.ie, Judith.conroy@ucd.ie, billy.bourke@ucd.ie, sean.ennis@ucd.ie). [c]UCD School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Dublin 4, Ireland (paul.mcgettigan@ucd.ie). [d]Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Ireland (lynamlen@tcd.ie, OSULLIJ4@tcd.ie). [e]Pathology, Our Lady's Children's Hospital, Crumlin, Dublin 12, Ireland (Michael.McDermott@olchc.ie). [f]National Centre for Inherited Metabolic Disorders, Children's University Hospital, Temple Street, Dublin 1, Ireland (ellen.crushell@cuh.ie). [g]National Centre for Medical Genetics, Our Lady's Children's Hospital, Crumlin, Dublin12, Ireland (SallyAnn.Lynch@olchc.ie). [h]Genetics Department, Children's University Hospital, Temple Street, Dublin 1, Ireland.

Corresponding author: Sean Ennis (sean.ennis@ucd.ie)

Health Sciences Centre, University College Dublin, Belfield, Dublin 4, Ireland

Phone: +353 1 716 6668 Fax: +353 1 716 6651

Keywords: Infantile hepatopathy; Multisystem presentation; Irish Traveller; LARS; Aminoacyl-tRNA synthetase; Mitochondrial

Abbreviations

SNP, single nucleotide polymorphism; PDB, protein data-bank; siRNA, small-interfering ribonucleic acid; LARS, gene encoding the cytoplasmic leucyl-tRNA synthetase; MMP, mitochondrial membrane potential; ROS, reactive oxygen species; LeuRS, leucyl-tRNA synthetase enzyme; aaRS, aminoacyl-tRNA synthetase; MELAS, mitochondrial encephalomyopathy lactic acidosis and stroke-like symptoms; LARS2, gene encoding the mitochondrial leucyl-tRNA synthetase; BCAAs, branched chain amino acids.

Abstract

Infantile hepatopathies are life-threatening liver disorders that manifest in the first few months of life. We report on a consanguineous Irish Traveller family that includes six individuals presenting with acute liver failure in the first few months of life. Additional symptoms include anaemia, renal tubulopathy, developmental delay, seizures, failure to thrive and deterioration of liver function with minor illness. The multisystem manifestations suggested a possible mitochondrial basis to the disorder. However, known causes of childhood liver failure and mitochondrial disease were excluded in this family by biochemical, metabolic and genetic analyses. We aimed to identify the underlying risk gene using homozygosity mapping and whole exome sequencing. SNP homozygosity mapping identified a candidate locus at 5q31.3-q33.1. Whole exome sequencing identified 1 novel homozygous missense mutation within the 5q31.3-q33.1 candidate region that segregated with the hepatopathy. The candidate mutation is located in the *LARS* gene which encodes a cytoplasmic leucyl-tRNA synthetase enzyme responsible for exclusively attaching leucine to its cognate tRNA during protein translation. Knock-down of *LARS* in HEK293 cells did not impact on mitochondrial function even when the cells were put under physiological stress. The molecular studies confirm the findings of the patients' biochemical and genetic analyses which show that the hepatopathy is not a mitochondrial-based dysfunction problem, despite clinical appearances. This study highlights the clinical utility of homozygosity mapping and exome sequencing in diagnosing recessive liver disorders. It reports mutation of a cytoplasmic aminoacyl-tRNA synthetase enzyme as a possible novel cause of infantile hepatopathy and underscores the need to consider mutations in *LARS* in patients with liver disease and multisystem presentations.

1. Introduction

Infantile liver disease, or hepatopathy, affects approximately 1 in 2,500 live births [1] and can be a life-threatening condition. Liver failure in infancy is challenging to diagnose and manage, and has a high mortality rate [2]. First symptoms usually develop in early infancy and include poor feeding, failure to thrive, vomiting, jaundice, distended abdomen and hypoactivity. Clinically, infantile hepatopathies manifest with elevated liver transaminases, hypoglycemia, cholestasis, coagulopathy and hyperbilirubinemia. Biochemical and genetic investigations can help establish the underlying basis of the hepatopathy. However, the etiology of many liver diseases is not known and a poor understanding of the pathogenesis has hampered the development of improved diagnostic and therapeutic strategies [1]. Identification of the spectrum of disease-causing genes is crucial to further our understanding of the mechanisms that underlie these complex conditions.

Medical case reports of specific gene defects have been pivotal to our understanding of clinical phenotypes and individual genes that contribute to liver dysfunction [3]. The current study involves an extended Irish Traveller family that includes six individuals with a variable phenotype consisting of infantile liver failure, anaemia, failure to thrive, seizures, moderate developmental delay, renal tubulopathy and deterioration of liver function with minor illness (Fig. 1a and Supplementary Table 1). The defining phenotypic feature in this family is infantile liver failure, a rare condition with an estimated prevalence of <1/1,000,000 (Orphanet). The patients show signs of intermittent improvement and only develop acute symptoms when their bodies are under physiologic stress due to illness. The earliest age of presentation was at 2 months and the oldest affected member of this family is now aged 33 years.

Four liver biopsies were obtained from three of the patients. Initial biopsy of patient III:10 at 3 months of age showed a profound, pan-lobular and predominantly macrovesicular steatosis with lobular disarray, hepatocyte ballooning and individual hepatocyte necrosis (Fig. 1b). A subsequent biopsy at age 4 showed persistent if less pronounced steatosis, now largely confined to periportal region, with the development of bland bridging fibrosis. Biopsies from patient III:4 (Fig. 1c) and IV:1 at age 9 and 16 months respectively showed almost identical patterns of periportal steatosis and early portal fibrosis.

Known metabolic and genetic causes of liver failure, including galactosemia, tyrosinemia, fructose intolerance, peroxisomal and lysosomal disorders, fatty acid oxidation disorders, bile acid disorders, urea cycle disorders, glycogen storage disorders, neonatal hemachromatosis and congenital disorders of glycosylation were excluded through clinical, biochemical and genetic analyses (Supplementary Table 2). Due to the multisystem involvement in the patients, a possible mitochondrial basis for the disorder was investigated. Electron microscopy of muscle tissue showed normal mitochondrial appearance. Genetic analyses excluded major mitochondrial rearrangements, including Pearson's deletion, and sequence analysis of mitochondrial DNA did not identify any pathogenic mutations. There was no evidence of a mitochondrial respiratory chain defect in muscle or liver tissues. The patients were assessed for mitochondrial DNA content in muscle and liver tissue but no depletion was found. Two of the common mitochondrial disease genes in the Irish and Irish Traveller populations, *POLG1* and *SURF1*, were sequenced but no causative mutations were identified. We aimed to identify the underlying disease gene for an infantile hepatopathy using the combined approach of homozygosity mapping and whole exome sequencing.

2. Materials and Methods

2.1 Genotyping and homozygosity mapping

DNA samples from three patients (III:4, III:7 and III:10) and five unaffected relatives (II:5, II:6, II:7, II:8 and III:9) were available for molecular analysis. Ethical approval for the study was obtained from the ethics committee of Our Lady's Children's Hospital, Dublin 12, Ireland. Genomic DNA was extracted from peripheral lymphocytes and genotyped for 1 million single nucleotide polymorphisms (SNPs) on the Illumina platform (Illumina, California). Genome-wide SNP homozygosity mapping using 979,427 autosomal SNPs was performed with HomozygosityMapper [4].

2.2 Sequencing and test analysis of mutations

DNA from four individuals, 3 affected (III:4, III:7 and III:10) and 1 unaffected sibling (III:9), was selected for whole exome sequencing (GATC, Germany). The exome of each individual was enriched with the SureSelect 38 Mb Human All Exon Kit (Agilent Technologies, Santa Clara), and sequenced on an Illumina Genome Analyser II platform. The single sequence reads were aligned to the hg18 reference genome with the Burrows-Wheeler Alignment tool version 0.5.7 [5]. Reads of inadequate sequence quality and potential PCR duplicates were discarded. The quality scores for the aligned reads were recalibrated using GATK [6]. Regions containing clusters of SNPs (corresponding to local misalignment of reads around indels) were identified and the reads in these regions were realigned using GATK. Variants and indels were identified using SAMtools [7].

The c.245A>G and c.1118A>G variants were validated by Sanger sequencing in available family members (Supplementary Table 3). A control panel, comprising 186 control

chromosomes from the Irish (n=48 alleles) and Irish Traveller (n=138 alleles) populations, was screened for each variant to determine its frequency in the general population.

A parametric two-point LOD score was calculated for the *LARS* c.245A>G and c.1118A>G variants using Merlin (http://www.sph.umich.edu/csg/abecasis/Merlin/tour/linkage.html) assuming a recessive model with full penetrance and a disease allele frequency of 0.0001.

The predicted impact and cross-species conservation of LARS p.K82R and p.Y373C were predicted using PolyPhen2. PolyPhen2 models amino acid substitutions using both structure and sequence information, comparing a property of the wild-type (ancestral, normal) allele and the corresponding property of the mutant (derived, disease-causing) allele. The functional importance of an allele replacement is predicted from its individual features by a naive Bayes classifier.

The mutation-induced ΔΔG values were calculated for LARS p.Y373C using structure-based energy calculation algorithms available on the PEAT-SA web server (http://enzyme.ucd.ie/PEAT_SA). All calculations for LARS p.Y373C used the protein data-bank (PDB) structure 2WFD (chain B), which describes the crystal structure of the human cytosolic leucyl-tRNA synthetase editing domain. A PDB structure was unavailable for residue 82.

2.3 Cell culture

HEK293 cells were obtained from the European Collection of Cell Cultures. Cells were cultured as monolayers in EMEM medium (Lonza, Switzerland) supplemented with 10%

fetal calf serum (Lonza) (Complete medium). Cells were maintained at 37^{o}C in 95% humidified air containing 5% CO_2. For serum starvation treatments, cells were cultured in EMEM medium supplemented with 0.5% fetal calf serum for 16 h.

2.4 Small-interfering RNA transfection

Small-interfering RNA (siRNA) targeting LARS was purchased from Sigma (Sigma, St Louis)**.** Cells were transfected in 6-well plates with either siRNA or scrambled control (100 nM) (Sigma) using Lipofectamine 2000 transfection reagent (9 µL) (Invitrogen, New York) as per the manufacturer's instructions.

2.5 Mitochondrial functional studies

Mitochondrial function was assessed by measurement of mitochondrial mass, mitochondrial membrane potential (MMP) and reactive oxygen species (ROS) using the probes MitoTracker Green (300 nM) (Invitrogen, New York), Rhodamine 123 (5 µM) (Sigma. St Louis) and 2,7-dichlorofluorescein diacetate (15 µM) (Sigma, St Louis), respectively. Transfected cells were seeded in 96-well plates at a density of 1.5 x 10^3 cells/well. At 72 h post transfection, media was removed and cells were incubated at 37^{o}C for 30 min with the relevant probes in a buffer containing 130 mM NaCl, 5 mM KCl, 1 mM Na_2HPO_4,1 mM $CaCl_2$, 1 mM $MgCl_2$, and 25 mM Hepes (pH 7.4). Probes were then removed and fluorescence was measured using a Wallac Victor^2 1420 multilabel counter (Perkin Elmer, Massachusetts) with an excitation filter of 485 nm and an emission filter of 535 nm.

2.6 Western blotting

Cells were lysed in RIPA buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 2 mM EDTA, pH 8.0, 0.5% Triton X-100, protease inhibitor cocktail, 1 mM PMSF and 1 µM Na_3VO_4). Sample

protein concentrations were quantified using the BCA assay (Pierce, Illinois). Proteins (30 μg) were resolved on a 12% polyacrylamide gel and transferred to polyvinylidene membranes. Immunoblots were incubated with rabbit anti-human LARS antibody (Epitomics, California), 1:60 dilution, rabbit anti-human LARS2 antibody (Abcam, United Kingdom), 1:1,000 dilution or rabbit anti-human αβ-Tubulin (Cell Signaling Technology. Massachusetts), 1:1,000 dilution for 16 h at 4 ^{o}C. Immunoblots were incubated with HRP-labelled swine anti-rabbit IgG secondary antibodies (Dako, Stockport). Detection was performed using SuperSignal West Pico chemiluminescent substrate kit (Pierce, Illinois).

3. Results

3.1 Homozygosity mapping and whole exome sequencing

To identify the locus responsible for the infantile hepatopathy in this family, we undertook a genome-wide SNP homozygosity mapping analysis. We identified two candidate loci; a major peak on chromosome 5 (141,640,648-149,027,979) and a minor peak on chromosome 15 (40,862,494-42,243,939) containing 62 and 40 genes respectively (Fig. 2a). To identify the underlying causative mutation, DNA from four individuals, three affected and 1 unaffected sibling, was selected for exome sequencing. The exomes were sequenced to a mean coverage of 43-100 fold (Supplementary Table 4). On average 76,588 variants with a minimum SNP quality score of Q20 and 10x coverage were identified per individual. Assuming a recessive model, we selected novel and rare homozygous non-synonymous variants that segregated with the hepatopathy and that were located within the two candidate linkage regions (Table 1). Of the 126 homozygous segregating variants, 121 were reported to have a population frequency >1% (dbSNP130) and were considered unlikely to be disease-causing. Of the remaining 5 candidate variants, only 2 were located within the candidate linkage regions. The two candidate missense variants are located in exon 4 (c.245A>G;

p.K82R) and exon 11 (c.1118A>G; p.Y373C) of the leucyl-tRNA synthetase gene (*LARS*) at 5q32. The c.245A>G variant is reported as a rare variant in dbSNP (rs112954500) but none of the 4,550 samples are homozygous for the minor G allele. The second candidate variant, c.1118A>G, is novel.

Subsequently, an additional two children from this extended family (IV:1 and III:11) presented with similar symptoms; infantile liver dysfunction, failure to thrive, poor feeding, hypotonia and anaemia (Supplementary Table 1). Sanger sequence analysis of *LARS* exons 4 and 11 showed that the two patients are homozygous for both the c.245A>G and the c.1118A>G variants. Validation and segregation of LARS c.245A>G and c.1118A>G (LOD 2.6) was confirmed by Sanger sequencing of DNA from all available family members (Fig. 2b and c).

3.2 Frequency of LARS p.K82R and p.Y373C in the general population

To investigate the possibility that the *LARS* variants (c.245A>G and c.1118A>G) were previously unidentified polymorphisms in the Irish population, we sequenced both variants in 186 control chromosomes. The frequency of LARS c.245A>G (rs112954500) was 2.1% (1/48 alleles) and 0% (0/138 alleles) in the Irish and Irish Traveller control groups respectively. In contrast, LARS c.1118A>G was not present in any of the 186 control chromosomes and has not been reported as a common polymorphic variant in dbSNP or the 1000 Genomes project.

3.3 Functional classification and evolutionary conservation of LARS p.K82R and p.Y373C

The Polyphen-2 classification tool was used to predict the functional consequences of each of the two missense mutations (K82R and Y373C) in *LARS* and to examine cross-species

conservation. The K82R (lysine-to-arginine) variant is a conservative amino acid substitution and is predicted to be benign (Fig. 3a). Furthermore, the residue at position 82 of the LARS protein shows low levels of conservation (61%) and the arginine residue observed in the patients is present in 15 species (Supplementary Table 5). In contrast, Y373C (tyrosine-to-cysteine) is predicted to be potentially deleterious, is highly conserved across eukaryotic species (89%) and cysteine is only present in the blood fluke (Fig. 3a).

3.4 Predicted effect of LARS p.Y373C on protein structure and function

The K82R substitution is located within a beta-turn structure of a region of LARS annotated as the nucleotidyl transferase superfamily (cl00015). While the effect of the K82R amino acid change on LARS folding cannot be anticipated (the protein structure for this region is not available), lysine-to-arginine is considered to be a conservative substitution, predicted to preserve the structure and functional properties of the protein [8]. However, K-to-R creates a candidate citrullination site. Therefore, while lysine-to-arginine is quite conservative in terms of the retained charge properties, the modification of arginine (+1) to citrulline (neutral) can change protein structure due to the loss of a positive charge. Examination of the annotated LARS protein structure shows that the Y373C mutation is located within the editing domain of LARS, known as connective peptide 1 (CP1), which extends from residue 260 to 509 (Fig. 3b). Calculation of the mutation-induced energy changes predicted that the tyrosine-to-cytosine mutation at position 373 destabilises the wild-type protein structure by 5.9 kJ/mol (Supplementary Table 6).

3.5 Impact of LARS knock-down on mitochondrial function in HEK293 cells

Although the patients in the current study have normal mitochondrial morphology, normal mitochondrial DNA content and normal respiratory enzyme complexes, the multisystem

presentation directed diagnosis towards a possible mitochondrial hepatopathy. To investigate a functional role for *LARS* in modulating mitochondrial function, HEK293 cells were transfected with siRNA targeting LARS. LARS protein expression was measured at 24 h, 48 h and 72 h post transfection, with a maximum decrease in LARS protein expression (82%) demonstrated at 72 h post transfection, when compared to cells transfected with a scrambled control (Fig. 4a and b). Levels of the mitochondrial leucyl-tRNA synthetase enzyme (LARS2) were not affected by the LARS siRNA (Supplementary Figure 1). To assess the effect of *LARS* knockdown on mitochondrial function, mitochondrial mass, mitochondrial membrane potential (MMP) and reactive oxygen species (ROS) levels were measured in untransfected cells, cells transfected with a scrambled control and cells transfected with the siRNA, at 72 h post transfection. Mitochondrial mass, MMP and ROS levels were comparable in all treatment groups (Fig. 4c-e), suggesting that *LARS* may not play a functional role in modulating mitochondrial function. Similarly, there was no alteration in mitochondrial mass, MMP or ROS when transfected cells were subjected to stress conditions by serum starvation (Supplementary Figure 2), mimicking the physiologic stress demonstrated in patients at the onset of acute symptoms. This further suggests that *LARS* may not modulate mitochondrial function and supports the biochemical and genetic findings of normal mitochondria in these patients.

4. Discussion

We identified two missense variants (K82R and Y373C) in *LARS* that segregate with infantile hepatopathy in an extended Irish Traveller family. Defects in *LARS*, a nuclear-encoded cytoplasmic enzyme, represent a novel cause of infantile hepatopathy. *LARS* encodes a leucyl-tRNA synthetase (LeuRS) responsible for exclusively attaching leucine to its cognate tRNA. LeuRS has its own highly evolved proofreading and editing mechanisms to ensure

fidelity of protein translation and cell survival [9] and [10]. Incorporation of an incorrect amino acid into the nascent polypeptide could cause misfolding and production of defective or dominant interfering proteins [11]. Therefore, accuracy is a key aspect of aminoacyl-tRNA synthetase (aaRS) function.

We found that the affected members of the Irish Traveller family described in this study are homozygous for one novel missense mutation (Y373C) and one rare variant (K82R) in *LARS*. Of the two variants, Y373C is more likely to be disease-causing for a number of reasons. Firstly, LARS p.Y373C was not detected in 186 Irish control chromosomes and has not been reported as a common polymorphic variant in dbSNP or the 1000 Genomes project. This observation suggests that the Y373C mutation does not occur frequently in the general population and supports the likelihood that it is a rare disease-causing mutation. Secondly, of all the possible amino acid substitutions, the large hydrophobic residues (tryptophan, tyrosine and phenylalanine) are replaced least often and the low mutability of these residues reflects their importance in forming the hydrophobic core of protein folds. The folding free energy is an important characteristic of a protein's stability and is directly related to its function. Disease-causing point mutations often bring about a change in free energy which can affect protein stability. In the current study, LARS p.Y373C is predicted to destabilise the protein with a positive free energy change of 5.9 kJ/mol. The impact of destabilisation depends on which part of the protein secondary structure is altered. The Y373C substitution is located in the editing domain and may cause a reduction in fidelity which, in turn, can lead to the incorporation of erroneous amino acids into nascent polypeptides. Alterations in LARS secondary structure could impact on molecular interactions and formation of the multi-synthetase complex which can disturb normal cell regulatory networks as a result of non-canonical functions in angiogenesis, immune responses, transcriptional and translational

control, and signal transduction. Furthermore, tyrosine-to-cysteine substitutions have been shown to be pathogenic in a number of diseases [12], [13] and [14]. Thirdly, evolutionary comparison of the LARS sequence indicates that tyrosine-373 is highly conserved across eukaryotic species, supporting the likelihood that the residue is important for protein structure and/or function. Finally, the Y373C mutation is located within the editing domain of LARS, known as CP1. The LeuRS CP1 domain functions as a second fine sieve that excludes correctly charged amino acids, but hydrolyses mischarged tRNA [15] and [16]. Besides the aminoacylation reaction, editing is important for LeuRS to reduce mistakes during amino acid selection [17]. For example, introduction of a D399A mutation in the CP1 domain of LARS does not affect the amino acid activation of Leu but inactivates post-transfer editing leading to the accumulation of mischarged tRNA [18]. Proper functioning of this fidelity mechanism is critical to the cell and defects in editing can result in cell death and neurological disease in mammals [19] and [20].

LARS encodes a cytoplasmic amino-acyl tRNA synthetase enzyme (aaRS) called LeuRS. aaRSs are responsible for protein translation both in the cytoplasm and the mitochondria but are all encoded by nuclear genes. LeuRS is one of 9 known aaRS proteins that forms a macromolecular multi-synthetase complex that regulates transcription, translation, and various signalling pathways [21] and [22]. In addition, several aaRSs associate with auxiliary protein factors that are primarily used in cellular tasks other than translation, expanding the repertoire of functions aaRSs may pursue [23]. The aaRS enzymes have been implicated in a number of genetic disorders, including those with muscular, metabolic or mitochondrial involvement [24] and [25]. Of particular interest *LARS2*, the nuclear-encoded mitochondrial leucyl-tRNA synthetase enzyme, has been implicated in mitochondrial encephalomyopathy, lactic acidosis and stroke-like symptoms (MELAS). *LARS2* is synthesised on cytoplasmic

ribosomes and subsequently imported into the mitochondrion where it localises exclusively [26]. The cytoplasmic (*LARS*) and mitochondrial (*LARS2*) LeuRS are distinct enzymes encoded by different nuclear genes and there is no evidence of a possible inter-dependence between *LARS* and *LARS2*. In the current study, siRNA-medicated knock-down of LARS did not impact on levels of LARS2 protein.

We investigated a possible role for the cytoplasmic LARS enzyme in mitochondrial function. Knock-down of LARS in HEK293 cells and examination of mitochondrial mass, MMP and ROS levels showed that cells deficient in cytoplasmic LARS did not differ to wild-type cells. The symptoms of the patients are transient and are most evident when the children are ill or under physiological stress. Therefore, we induced stressful conditions in the siRNA transfected cells to determine if knock-down of LARS affected mitochondrial function when the cells are under stress, as was suggested by the patient phenotype. However, no difference was observed between wild-type and LARS knock-down cells with or without physiological stress. Our findings suggest that, under the conditions investigated, deficiency in LARS does not affect the normal function of the mitochondria and provides evidence that the hepatopathy observed in the patients is not mitochondrial-based but rather mimics a mitochondrial disease with its multisystem involvement. Exclusion of mitochondrial involvement is supported by the fact that the patients have normal mitochondrial morphology, normal respiratory chain complexes and do not have genetic abnormalities in the mitochondrial genome.

The most prominent phenotypic characteristic of the patients in the current study is liver failure before the first year of life, with recurrent episodes of liver dysfunction in subsequent years when the patients are challenged with minor illness. There are several ways by which

mutations in *LARS* could result in liver dysfunction. The liver is responsible for the synthesis of many proteins, some of which are vital for normal physiological function. Altered amino acid metabolism is a hallmark of liver disease which is characterised by low levels of circulating branched chain amino acids (BCAAs) and elevated levels of circulating aromatic amino acids and methionine [27]. LARS is responsible for attaching leucine, one of the three essential BCAAs, to its cognate tRNA. BCAAs help protect against diseases, including liver failure. Leucine is critical to human life and is particularly involved in stress, energy and muscle metabolism. It is possible that, similar to the A3243G mutation in $tRNA^{leu(UUR)}$ that causes MELAS, the *LARS* missense mutations identified in the current study may cause inefficient aminoacylation of tRNA-Leu which, in turn, may reduce the rate of protein synthesis in the liver [26].

We analysed the leucine content of all proteins in the human genome. We hypothesised that proteins with a high requirement for leucine were the most likely to be affected in patients with mutations in *LARS*, the enzyme responsible for making leucine. We found that immune-related proteins including the toll-like receptor family, interferons, various interleukins and other cytokines contain a very high percentage content of leucine. The observation that immune-related proteins have a high requirement for leucine correlates well with the phenotypic profile of the patients; the *LARS* mutation only appears penetrant/symptomatic when the patients are ill and require high levels of immunological proteins. In addition, we found that the phospholipid biosynthetic process is the most significantly enriched for proteins with a high percentage of leucine (corrected p value = 0.0017) (Supplementary Table 7). Liver biopsies of the patients show the presence of large fat deposits in the liver and defects in phospholipid biosynthesis have been linked to fatty liver disease.

The list of genetic diseases caused by mutations that affect mRNA translation is rapidly growing and the disease phenotypes show a surprising degree of heterogeneity [11]. Three of the seventeen cytoplasmic aaRS genes (*YARS*, *AARS* and *GARS*) have been implicated in disease to date [25]. The current study is the first to report mutations in *LARS* as a novel cause of infantile hepatopathy. Although infantile hepatopathies can be caused by mitochondrial disorders, we found that a mitochondrial-based hepatopathy was unlikely in these patients as knock-down of *LARS* had no impact on mitochondrial function. Exclusion of a mitochondrial basis is also supported by the biochemical and genetic analyses. Further investigation into the effects of translation defects in the cytoplasm on liver function is being explored in order to determine how mutations in *LARS* result in this complex phenotype. Given the high recurrence risk in endogamous Irish Traveller families, identification of the causative mutation in these patients will have considerable implications for their families both in terms of prognostic advice and genetic counselling.

Conflict of interest

All authors have nothing to disclose.

Acknowledgements

We sincerely thank the participating families for the use of genetic samples and clinical information. We would also like to thank Prof Eileen Treacy, Dr AA Monavari and Dr Joanne Hughes for obtaining DNA samples and providing clinical phenotypic details of the patients. This work was supported by the Children's Fund for Health, The Fundraising Office for Temple Street Children's University Hospital, Dublin, Ireland (PAC 09131). Jillian Casey

is supported by an EMBARK postgraduate award from the Irish Research Council for Science, Engineering and Technology (IRCSET) and a Medical Charities Research Group grant (MRCG/2011/17) from the National Children's Research Centre (Our Lady's Children's Hospital, Ireland) and the Health Research Board (Ireland). Paul McGettigan is funded by a Science Foundation Ireland strategic research grant (07/SRC/B1156).

References

1 R.J. Sokol. New North American research network focuses on biliary atresia and neonatal liver disease. J. Pediatr. Gastroenterol. Nutr., 36 (1) (2003), p. 1

2 P. McClean, S.M. Davison. Neonatal liver failure. Semin. Neonatol., 8 (5) (2003), pp. 393-401

3 C. Scharfe, H.H. Lu, J.K. Neuenburg, E.A. Allen, G.C. Li, T. Klopstock, T.M. Cowan, G.M. Enns, R.W. Davis. Mapping gene associations in human mitochondria using clinical disease phenotypes. PLoS Comput. Biol., 5 (4) (2009), e1000374

4 D. Seelow, M. Schuelke, F. Hildebrandt, P. Nürnberg. HomozygosityMapper--an interactive approach to homozygosity mapping. Nucleic Acids Res., 37 (Web Server issue) (2009), pp. 593-599

5 H. Li, R. Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25 (14) (2009), pp. 1754-1760

6 A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res., 20 (9) (2010), pp. 1297-1303

7 H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25 (16) (2009), pp. 2078-2079

8 T.D. Wu, D.L. Brutlag. Discovering empirically conserved amino acid substitution groups in databases of protein families. Proc. Int. Conf. Intell. Syst. Mol. Biol., 4 (1996), pp. 230-240

9 I. Apostol, J. Levine, J. Lippincott, J. Leach, E. Hess, C.B. Glascock, M.J. Weickert, R. Blackmore. Incorporation of norvaline at leucine positions in recombinant human hemoglobin expressed in Escherichia coli. J. Biol. Chem., 272 (46) (1997), pp. 28980-28988

10 S. Englisch, U. Englisch, F. von der Haar, F. Cramer. The proofreading of hydroxy analogues of leucine and isoleucine by leucyl-tRNA synthetases from E. coli and yeast. Nucleic Acids Res., 14 (19) (1986), pp. 7529-7539

11 G.C. Scheper, M.S. van der Knaap, C.G. Proud. Translation matters: protein synthesis defects in inherited disease. Nat. Rev. Genet., 8 (9) (2007), pp. 711-723

12 W. Zhou, C.R. Freed. Tyrosine-to-cysteine modification of human alpha-synuclein enhances protein aggregation and cellular toxicity. J. Biol. Chem., 279 (11) (2004), pp. 10128-10135.

13 J.Y. Ooi, Y. Yagi, X. Hu, Y.T. Ip. The Drosophila Toll-9 activates a constitutive antimicrobial defense. EMBO Rep., 3 (1) (2002), pp. 82-87

14 H. Sawai, A. Ida, Y. Nakata, K. Koyama. Novel missense mutation resulting in the substitution of tyrosine by cysteine at codon 597 of the type X collagen gene associated with Schmid metaphyseal chondrodysplasia. J. Hum. Genet., 43 (4) (1998), pp. 259-261

15 A.R. Fersht, M.M. Kaethner. Enzyme hyperspecificity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. Biochemistry, 15 (15) (1976), pp. 3342-3346

16 A.K. Betha, A.M. Williams, S.A. Martinis. Isolated CP1 domain of Escherichia coli leucyl-tRNA synthetase is dependent on flanking hinge motifs for amino acid editing activity. Biochemistry, 46 (21) (2007), pp. 6258-6267

17 O. Nureki, D.G. Vassylyev, M. Tateno, A. Shimada, T. Nakama, S. Fukai, M. Konno, T.L. Hendrickson, P. Schimmel, S. Yokoyama. Enzyme structure with two catalytic sites for double-sieve selection of substrate. Science, 280 (5363) (1998), pp. 578-582

18 X. Chen, J.J. Ma, M. Tan, P. Yao, Q.H. Hu, G. Eriani, E.D. Wang. Modular pathways for editing non-cognate amino acids by human cytoplasmic leucyl-tRNA synthetase. Nucleic Acids Res., 39 (1) (2011), pp. 235-247

19 V.A. Karkhanis, A.P. Mascarenhas, S.A. Martinis. Amino acid toxicities of Escherichia coli that are prevented by leucyl-tRNA synthetase amino acid editing. J. Bacteriol., 189 (23) (2007), pp. 8756-8758

20 J.W. Lee, K. Beebe, L.A. Nangle, J. Jang, C.M. Longo-Guess, S.A. Cook, M.T. Davisson, J.P. Sundberg, P. Schimmel, S.L. Ackerman. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature, 443 (7107) (2006), pp. 50-55

21 S.G. Park, K.L. Ewalt, S. Kim. Functional expansion of aminoacyl-tRNA synthetases and their interacting factors: new perspectives on housekeepers. Trends Biochem. Sci., 30 (10) (2005), pp. 569-574

22 S.W. Lee, B.H. Cho, S.G. Park, S. Kim. Aminoacyl-tRNA synthetase complexes: beyond translation. J. Cell Sci., 117 (2004), pp. 3725-3734

23 C.D. Hausmann, M. Ibba. Aminoacyl-tRNA synthetase complexes: molecular multitasking revealed. FEMS Microbiol. Rev., 32 (4) (2008), pp. 705-721

24 A. Montanari, C. De Luca, L. Frontali, S. Francisci. Aminoacyl-tRNA synthetases are multivalent suppressors of defects due to human equivalent mutations in yeast mt tRNA genes. Biochim. Biophys. Acta., 1803 (9) (2010), pp.1050-1057

25 A. Antonellis, E.D. Green. The role of aminoacyl-tRNA synthetases in genetic diseases. Annu. Rev. Genomics Hum. Genet., (9) (2008), pp. 87-107

26 R. Li, M.X. Guan. Human mitochondrial leucyl-tRNA synthetase corrects mitochondrial dysfunctions due to the tRNALeu(UUR) A3243G mutation, associated with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like symptoms and diabetes. Mol. Cell Biol., 30 (9) (2010), pp. 2147-2154

27 M. Charlton. Branched-chain amino acid enriched supplements as therapy for liver disease. J. Nutr., 136 (2006), pp. S295-S298

Figure titles and legends

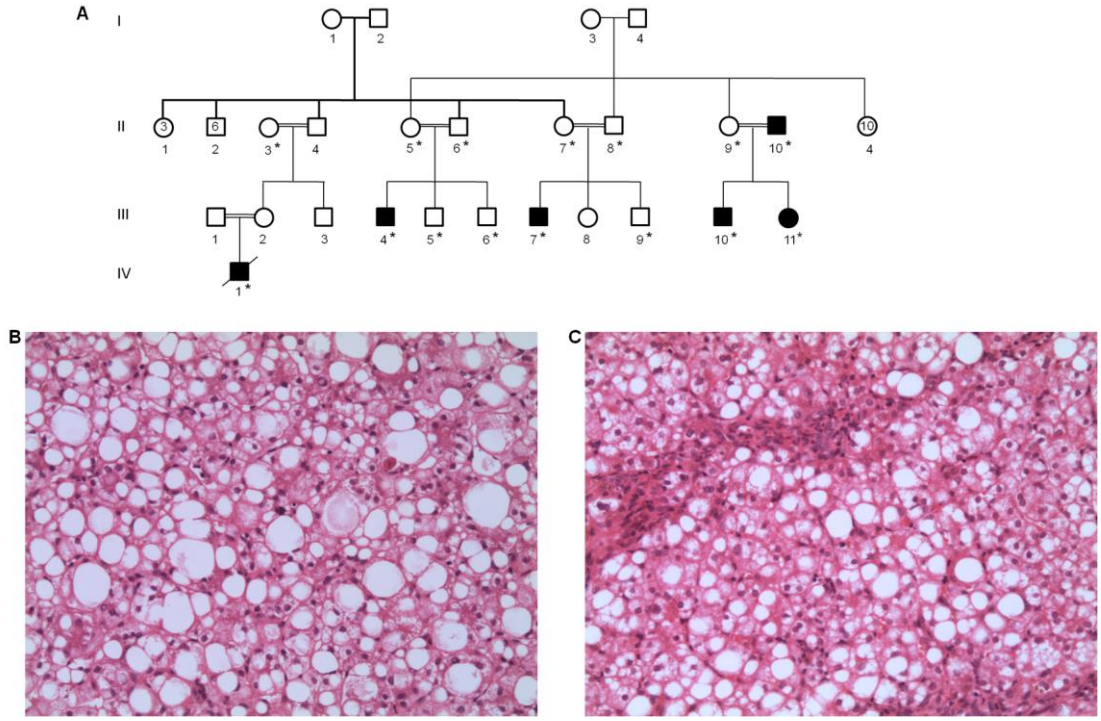


Fig. 1. An Irish Traveller family with an infantile hepatopathy (A) Six individuals presented with liver failure in the first few months of life. Individual IV:1 died age 4 years and 11 months from bilateral bronchopneumonia secondary to influenza H1N1 infection. DNA from three patients (III:4, III:7 and III:10) and five unaffected relatives (II:5, II:6, II:7, II:8 and III:9) was available for homozygosity mapping. DNA from additional patients (II:10, III:11 and IV:1) and unaffected relatives (II:3, III:5 and III:6) was later obtained for mutation testing. (B) Photomicrograph of patient III:10 liver showing marked steatosis and hepatocyte ballooning with a single necrotic hepatocyte. (C) Photomicrograph of patient III:4 liver showing periportal steatosis and mild portal fibrosis.

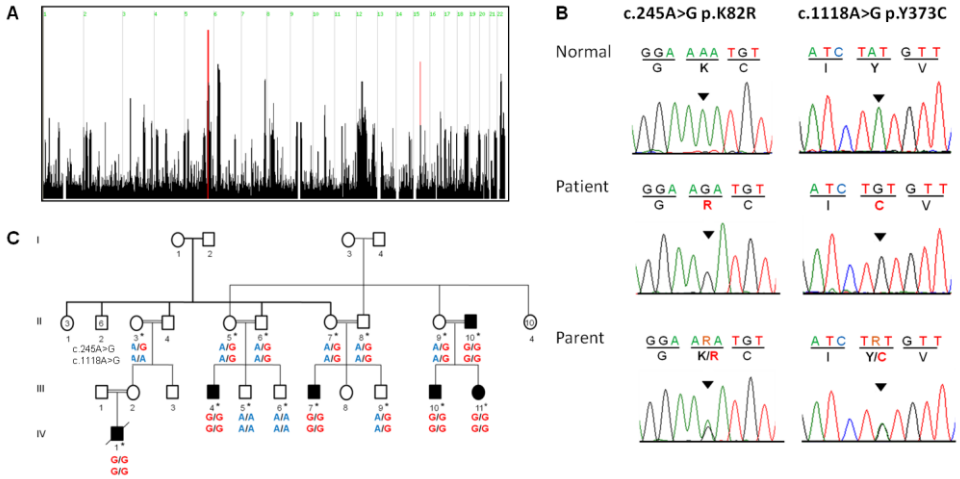


Fig. 2. Validation and segregation of LARS p.K82R and p.Y373C with an infantile hepatopathy (A) Homozygosity mapping identified one major (chromosome 5:141,640,648-149,027,979) and one minor peak (chromosome 15:40,862,494-42,243,939) containing 102 positional candidate genes. (B) The LARS c.245A>G (p.K82R) and c.1118A>G (p.Y373C) variants were validated by Sanger sequencing. The three rows represent Sanger traces corresponding to patients, parents and controls respectively. Sanger sequencing showed that all affected individuals are homozygous for K82R and Y373C. The unaffected parents are obligate carriers and unaffected siblings are heterozygous/homozygous for the normal allele at both loci. (C) DNA from available family members was tested for LARS c.245A>G and c.1118A>G. Sanger sequencing confirmed that both variants segregate with the hepatopathy.

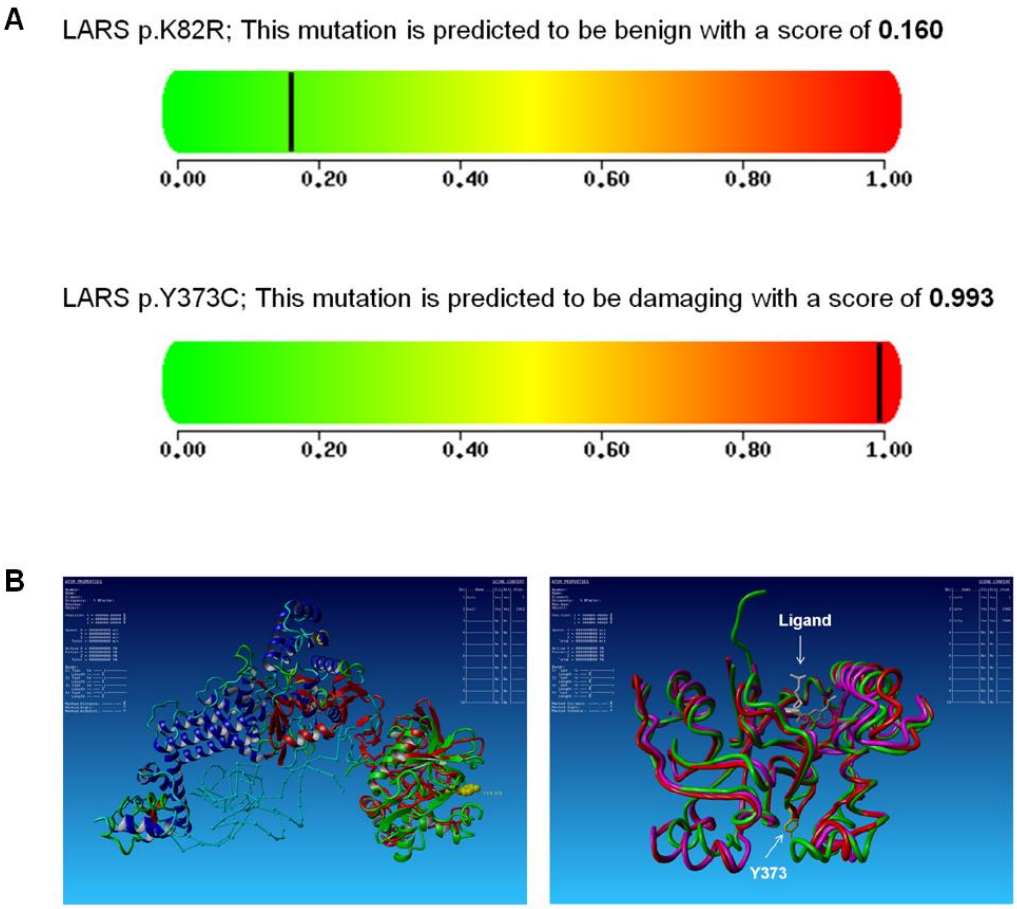


Fig. 3. Functional classification of LARS p.K82R and p.Y373 (A) PolyPhen2 predicted that the K82R variant is benign with a score of 0.160 (sensitivity: 0.91; specificity: 0.83). The Y373C mutation was predicted to be probably damaging with a score of 0.993 (sensitivity: 0.60; specificity: 0.96). (B) Residue 373, which is mutated in the patients, is shown as yellow spheres in the left-hand figure. The Y373C mutation (grey) is located within the editing domain of LARS, known as the connective peptide 1 (CP1) (right-hand figure). The position at which a ligand (grey) binds to the LARS CP1 domain is shown in the right-hand figure.

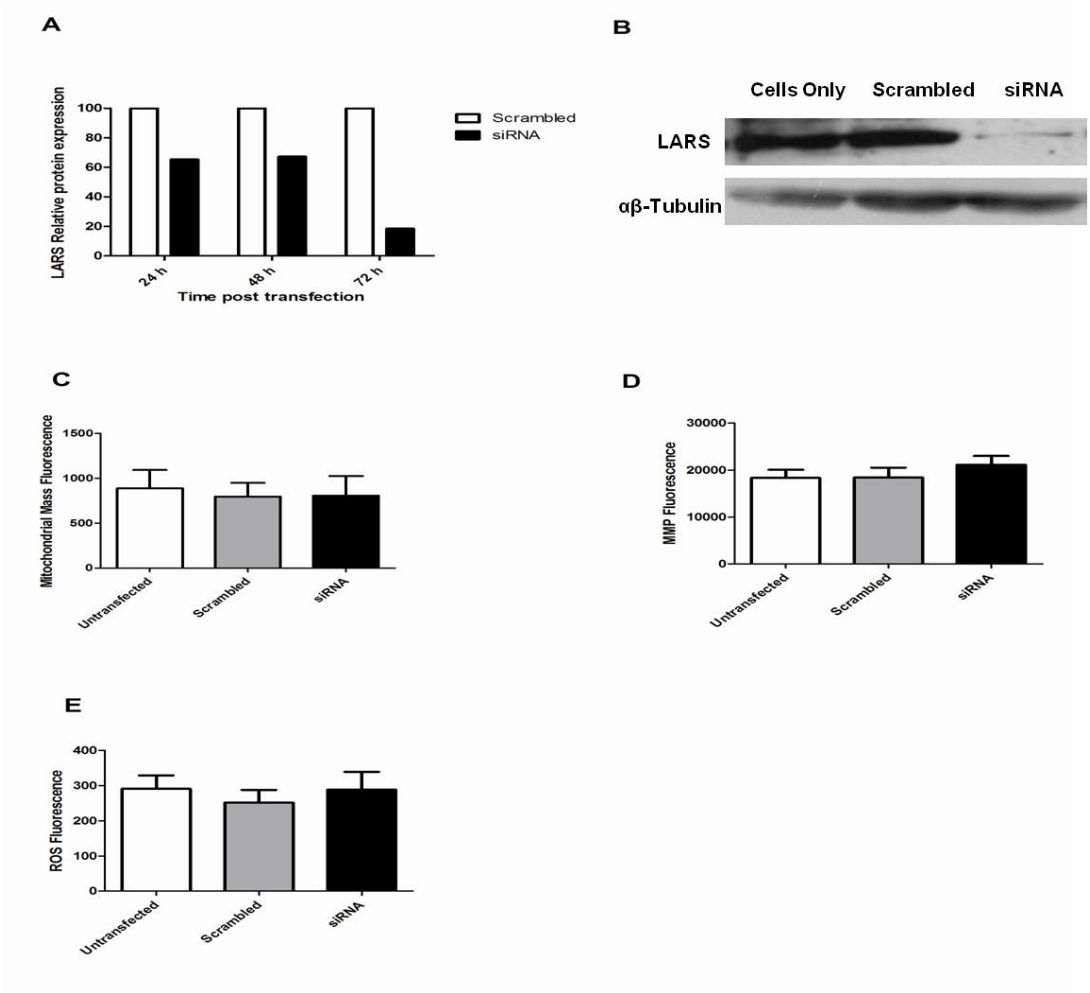
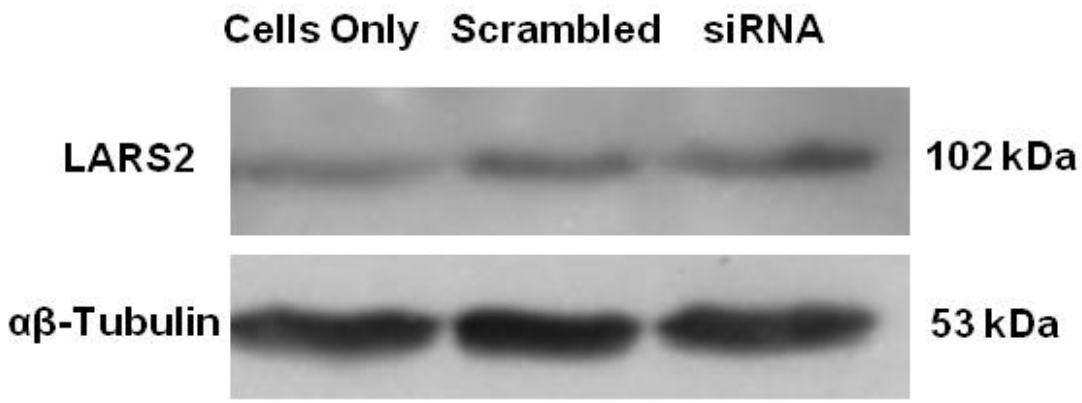
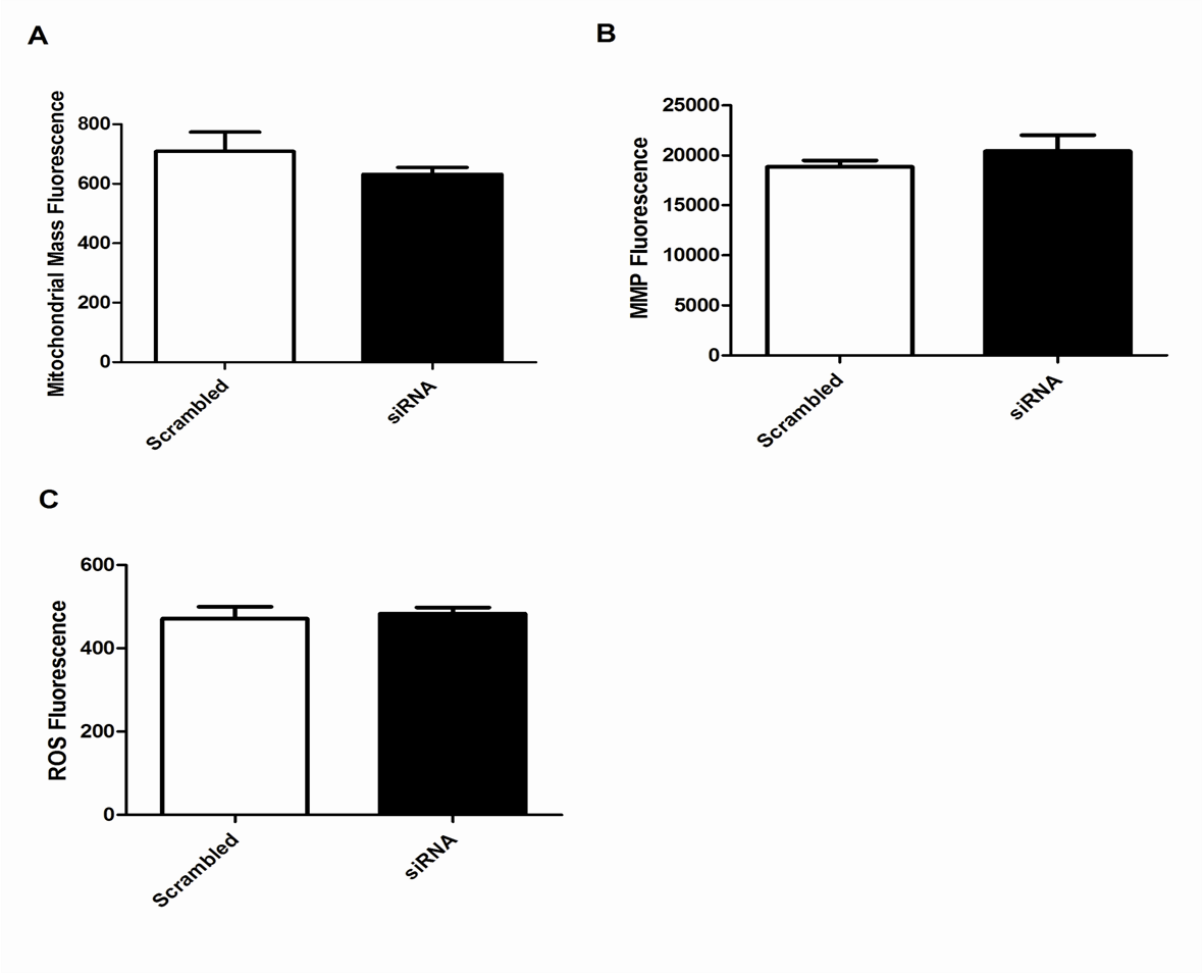


Fig. 4. Impact of LARS knock-down in HEK293 cells on mitochondrial function. (A) Knockdown of LARS protein in HEK293 cells was confirmed by western blot at 24 h, 48 h and 72 h post transfection. The greatest reduction in LARS protein expression was demonstrated at 72 h post transfection. (B) Representative immunoblot demonstrating knockdown of LARS protein at 72 h post transfection. (C) Mitochondrial mass, (D) MMP and (E) ROS levels were similar in untransfected cells, scrambled control transfected cells and siRNA transfected cells.

Supplementary Figures and Tables



Supplementary Figure 1. LARS2 expression is unaffected by knockdown of LARS. LARS2 protein expression was measured by western blot in untransfected HEK293 cells, cells transfected with a scrambled control and cells transfected with anti-LARS small interfering RNA (siRNA) at 72h post transfection. Similar levels of LARS2 expression was demonstrated in untransfected cells, cells transfected with a scrambled control and cells transfected with anti-LARS siRNA.



Supplementary Figure 2. Examination of mitochondrial function following knock-down of LARS in cells subjected to stress conditions.

Mitochondrial function was assessed in transfected cells subjected to serum starvation conditions. (A) mitochondrial mass, (B) mitochondrial membrane potential (MMP) and (C) reactive oxygen species **(**ROS) levels were similar in both scrambled control transfected cells and anti-LARS siRNA transfected cells.

Supplementary Table 1. Phenotypic characteristics of patients with an infantile hepatopathy of unknown genetic cause

Ind (Sex)	Age	Liv fail	FTT	HT	MA	Abn LFT	↑ LA	DD	Abn brain MRI	SRS	SF	Other
III:4 (M)	3 yrs 10 mos	+	+	+	+	+	+	+	+	-	NK	Unusually chubby cheeks, frontal bossing, abnormal thumbs, dysmorphic, renal tubulopathy, gross motor delay, and congenital adrenal hypoplasia
III:7 (M)	6 yrs 10 mos	+	+	+	NK	+	+	+	+	+	+	Renal dysfunction, decreased muscle bulk, coagulopathy, sensorineural hearing loss, microcephaly, head lag, continued hand flapping and excitable behaviour, and locomotor, social and speech problems
III:10 (M)	8 yrs 10 mos	+	+	+	+	+	+	+	-	+	+	Enlarged liver, cirrhosis of the liver, coagulopathy, gross motor function delay, congenital myopathy, mildly dysplastic bone marrow, long fingers and toes
III:11 (F)	4 mos	+	+	+	+	+	NK	NK	+	NK	NK	History of irritability, lethargy, paleness and poor feeding. Elevated tyrosine metabolites in her organic acids and elevated threonine in her amino acid profile
II:10 (M)	33 yrs 2 mos	+	+	+	+	+	+	+	-	+	NK	Suspected Reyes syndrome (1979), entered comatosed state as a result of severe measles but recovered, mild to moderate learning difficulties
IV:1 (M)	Died aged 4 yrs 11 mos	+	+	+	+	+	+	-	+	NK	+	Hyperactive, history of constipation, died from bilateral bronchopneumonia secondary to influenza H1N1 infection

Four members of an extended Irish Traveller family presented with acute liver failure before the first year of life. Abnormal liver function tests, elevated lactates, abnormal brain MRI's, failure to thrive, developmental delay and seizures are suggestive of a mitochondrial disorder but the underlying genetic cause is unknown. Features that are present or absent are represented by + and - respectively. Abbreviations are as follows: FTT; failure to thrive, HT; hypotonic, MA; microcytic anaemia, Abn LFT; abnormal liver function tests, ↑ LA; elevated serum lactates, Liv fail; acute liver failure, DD; developmental delay, Abn brain MRI; abnormal brain MRI, SRS; seizures, SF; steatosis and fibrosis of the liver, Other; other clinical features, NK; not known.

Supplementary Table 2. Exclusion of known causes of liver failure

Disorders investigated	Investigations performed
General investigations	▪ Serology for Hepatitis B, C, EBV, CMV ▪ Alpha-1-antitrypsin and Pi type ▪ Sweat test ▪ CGH microarray
Disorders that give rise to intoxication (intermediary metabolism)	▪ Ammonia, glucose, creatine kinase ▪ Plasma amino acids including homocysteine ▪ Urine organic acids (including orotic acid and succinylacetone) ▪ Serum ceruloplasmin, serum and urine copper ▪ Serum iron and ferritin ▪ Dried blood spot galactose ▪ Gal-1-PUT and epimerase activity ▪ *GALT*, *ACADS* and *ACADM* mutation analysis ▪ Acylcarnitine profile – blood spot and fibroblast ▪ Fibroblast fatty acid beta-oxidation (tritium release assay)
Disorders involving energy metabolism	▪ Pre and post prandial lactate, pyruvate, ketone body and glucose profiling ▪ Muscle biopsy: histology, histochemistry, electron microscopy, respiratory chain enzyme analysis, muscle mtDNA content, mtDNA analysis for deletions and rearrangements, mtDNA sequencing ▪ *POLG*, *SURF1*, *PEO1* and *TRMU* sequencing ▪ Liver biopsy: histology (see Fig. 1b), liver respiratory chain analysis, liver mtDNA content ▪ Red cell glycogen, leukocyte glycogen debrancher and phosphorylase activity ▪ Urine sugar and polyol analysis ▪ Tubular reabsorption of phosphate ▪ Ophthalmology and cardiology assessment ▪ MRI and MR spectroscopy of brain
Disorders involving complex molecules	▪ Isoelectric focusing of transferrin ▪ White cell lysosomal enzyme screen ▪ Urine mucopolysaccharides and oligosaccharides ▪ Bone marrow aspirate ▪ Plasma very long chain fatty acid analysis ▪ Plasma and urine bile acid analysis, plasma cholesterol

Known causes of liver failure were excluded in the patients through biochemical, metabolic and genetic analyses. All investigations listed had normal test results. The metabolic investigations are classified as per JM Saudubray 2012 [1]. Abbreviations are as follows:

EBV; Epstein-Barr virus, CMV; cytomegalovirus, CGH; comparative genomic hybridization, MRI; Magnetic Resonance Imaging.

Supplementary Table 3. Primer sequences for *LARS* exons 4 and 11-12

Exon	Forward 5'-3'	Reverse 5'-3'	Product bp
4	GTTGGTGAGTTTTGGGATGC	TGGATGCCCTTTATTTAAGCAC	415
11-12	CAACATAGCAAGGCTCTGTCTC	TTTAGCCTAGTCCCACAGCG	559

Primers were designed for *LARS* exons 4 and 11-12 to screen for the c.245A>G (p.K82R) and c.1118A>G (p.Y373C) variants respectively. The PCR reactions were performed using an annealing temperature of 62°C.

Supplementary Table 4. General statistics for exome sequence analysis

Patient	Total reads	Uniquely mapped reads (minus duplicates)	Called bases overlapping with target exons (CCDS)	CCDS % coverage	Mean fold coverage
III:4	59,553,308	55,407,130	1,309,383,561	98.43	47.81
III:7	49,863,568	49,190,866	1,516,011,850	98.43	55.47
III:10	66,990,204	65,091,965	1,209,383,087	98.84	43.97
III:9	110,488,100	108,957,034	2,737,713,032	98.82	99.57

DNA from 3 patients (III:4, III:7 and III:10) and 1 unaffected sibling (III:9) was selected for whole exome sequencing. An Agilent 38Mb CCDS exon array was used for enrichment and the captured DNA was sequenced on an Illumina GAII sequencer. The control (unaffected sibling III:9) was run in two lanes of the Illumina GAII which generated additional sequence data compared to the samples from the 3 patients.

Supplementary Table 5. Cross-species conservation analysis of the K82 and Y373 LARS residues

Query	Residue 82	Residue 373
sp|A6QLR2#1	K	Y
sp|UPI000179E9A2#1	K	Y
sp|Q8BMJ2#1	K	Y
sp|UPI0001B79C99#1	K	Y
sp|Q5PPJ6#1	K	Y
sp|Q6ZPT2#1	K	Y
sp|E1C2I9#1	K	Y
sp|UPI000194D10A#1	K	Y
sp|UPI0000F2B57B#1	K	Y
sp|Q5EB29#1	K	Y
sp|UPI000069F205#1	K	Y
sp|Q6AX83#1	K	Y
sp|UPI00016E27A7#1	K	Y
sp|C0H907#1	K	Y
sp|UPI00017B24E8#1	C	Y
sp|C3XXL0#1	K	Y
sp|UPI000180BA02#1	K	Y
sp|B7PFM8#1	K	Y
sp|D2A1C9#1	K	Y
sp|UPI0000D55F13#1	K	Y
sp|B3MP31#1	R	Y
sp|UPI0000519B35#1	K	Y
sp|B4LRU6#1	R	Y
sp|B4KK92#1	R	Y
sp|B4NXX9#1	R	Y
sp|Q29MC7#1	R	Y
sp|B4G9F6#1	R	Y
sp|B4I2Z8#1	R	Y
sp|B4JQG2#1	R	Y
sp|Q8MRF8#1	R	Y
sp|Q9VQR8#1	R	Y
sp|Q7Q495#1	K	Y
sp|B3N334#1	R	Y
sp|Q0IF77#1	K	Y
sp|B4MV75#1	K	Y
sp|B4Q9L5#1	R	Y
sp|Q09996#1	K	Y

sp|A8XXA4#1	K	Y
sp|A8PGZ5#1	K	Y
sp|E0VRG6#1	K	Y
sp|B3SBQ2#1	K	Y
sp|UPI000179292F#1	K	Y
sp|E0VD05#1	K	Y
sp|UPI00015B5E1F#1	K	Y
sp|C4QUA6#1	K	F
sp|Q54N83#1	K	Y
sp|A9UW56#1	K	Y
sp|D3B015#1	K	Y
sp|B0E752#1	E	Y
sp|C4M6P0#1	E	Y
sp|D0NG73#1	E	Y
sp|D7FTL8#1	K	Y
sp|D8LXV2#1	E	Y
sp|B9PT79#1	K	Y
sp|Q4RIR3#1	-	Y
sp|B6KL27#1	K	Y
sp|D2UXZ9#1	K	H
sp|C5KHS7#1	K	Y
sp|C5KBE7#1	K	Y
sp|B7FUC3#1	K	Y
sp|UPI00006CAA88#1	F	Y
sp|B8BYD7#1	K	Y
sp|B6AA20#1	K	F
sp|Q4CTR0#1	F	Y
sp|B0WIS0#1	K	Y
sp|Q386D9#1	Y	Y
sp|C6LVQ0#1	Y	F
sp|A8BY54#1	D	F
sp|A4H796#1	R	Y
sp|A7ANW3#1	K	Y
sp|Q4QG44#1	R	Y
sp|A4HVN8#1	R	Y
sp|A0BIY8#1	Q	Y
sp|Q4N1Y5#1	K	H
sp|Q4U991#1	K	Y
sp|A2EIB0#1	K	R
sp|B3L7I1#1	Q	Y
sp|Q4XWB3#1	D	Y
sp|A7S766#1	-	Y
sp|Q7RMJ9#1	D	C

sp|UPI0001923882#1	K	Y
sp|B9QEI1#1	-	Y
sp|UPI0001CF1A8A#1	K	-
sp|C6KT64#1	-	Y
sp|A5K262#1	-	Y
sp|Q9CQM4#1	K	-
sp|UPI0001CBA55E#2	K	-
sp|UPI00016E27C6#1	K	-
sp|D0A6K8#1	Y	-
sp|Q4CL52#1	L	-
sp|Q4Z2Y7#1	-	Y
sp|A8AB16#1	-	P
sp|Q5DC57#1	K	-
sp|E0S7E8#2	Y	-
sp|C6KT64#2	D	-
sp|Q8SRS8#2	Y	-
sp|B7XHP6#2	F	-
sp|Q5CG48#1	K	-
sp|C5KEA6#1	-	Y
sp|C4V8L1#2	F	-
sp|A7S765#1	K	-
sp|Q5BRJ7#1	-	F
sp|D6P6Z0#1	K	-
sp|A5K262#3	Q	-
sp|Q4Z2Y7#2	D	-

Multiple species alignment was performed using PolyPhen2 for the amino acids at positions 82 and 373 of the LARS protein. The residues at positions 82 and 373 are conserved in 58/96 and 78/88 species respectively.

Supplementary Table 6. Energy changes for LARS p.Y373C mutant protein

Mutation	Van der Walls	Electro-static	H-bond	Desolvation	Backbone entropy	Water bridges	Total ΔΔG
B:Y373C	9.52	0.10	2.42	-7.16	-0.33	1.42	5.97

The mutation-induced energy changes (ΔΔG) for LARS p.Y373C were calculated using the PEAT-SA webserver. All calculations for LARS p.Y373C used the protein data-bank (PDB) structure 2WFD (chain B). A positive value for ΔΔG indicates a less stable mutant compared to the wild-type protein structure.

Supplementary Table 7. Biological processes significantly enriched for proteins with high percentage leucine content

Biological Process	Observed	Expected	Raw p value	Adjusted p value
Phospholipid biosynthetic process (GO:0008654)	9	1.21	3.28e-06	0.0017
Glycerophospholipid biosynthetic process (GO:0046474)	6	0.79	0.0001	0.0058
Preassembly of GPI anchor in ER membrane (GO:0016254)	4	0.19	3.08e-05	0.0058
Phosphoinositide biosynthetic process (GO:0.0016254)	5	0.49	0.0001	0.0058
GPI anchor metabolic process (GO:0006505)	5	0.40	4.6e-05	0.0058
Phospholipid metabolic process (GO:0006644)	10	2.25	8.78e-05	0.0058
Lipid biosynthetic process (GO:0008610)	5	0.39	3.97e-05	0.0058
Organophosphate metabolism (GO:0019637)	10	2.38	0.0001	0.0058
Zinc ion transport (GO:0006829)	4	0.28	0.0002	0.0105
Protein amino acid lipidation (GO:0006497)	5	0.63	0.0004	0.0175
Glycerolipid biosynthetic process (GO:0045017)	6	0.95	0.0004	0.0175
Lipoprotein biosynthetic process (GO:0042158)	5	0.69	0.0006	0.0242

Glycoprotein metabolic process (GO:0009100)	9	2.43	0.0007	0.0262
Response to external stimulus (GO:0009605)	22	10.69	0.0010	0.0328
Response to wounding (GO:0009611)	16	6.962	0.0010	0.0328
Transition metal ion transport (GO:0000041)	5	0.80	0.0012	0.0370
Cellular lipid metabolic process (GO:0044255)	18	8.38	0.0018	0.0496
Phosphoinositide metabolic process (GO:0030384)	5	0.88	0.0018	0.0496

Proteins with ≥20% leucine content were analysed using WebGestalt [2] to identify biological processes that are significantly enriched for proteins with a high percentage leucine content. A hypergeometric analysis was applied with stringent Bonferonni correction to adjust for multiple testing. Biological processes with an adjusted *p* value < 0.05 are listed.

References

[1] J.M. Saudubray, Clinical approach to inborn errors of metabolism *In* Inborn Metabolic Diseases 5th Edition, Springer-Verlag Berlin Heidelberg 2012 Editors: Saudubray, van den Berghe, Walter.

[2] B. Zhang, S. Kirov, J. Snoddy, WebGestalt: an integrated system for exploring gene sets in various biological contexts Nucleic Acids Res., 33 (2005), pp. W741-748.