Defining the Role of Cysteinyl-tRNA Synthetase (CARS1) in Human Recessive Disease

by

Molly E. Kuo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2021

Doctoral Committee:

Professor Anthony Antonellis, Chair
Professor Donna Martin
Professor Miriam Meisler
Associate Professor Jordan Shavit

Molly Kuo

mollykuo@umich.edu

ORCID iD: 0000-0002-1190-4115

© Molly Kuo 2021

Dedication

To the patients and families

who generously shared their genetic information,

medical records, samples, and stories with me.

Acknowledgements

This work is the result of the support and guidance from an entire community of individuals who have helped me throughout these years.

The first person I want to thank is Tony Antonellis, who has been a fantastic thesis advisor. Tony is a mentor who both helped me when I needed it and also pushed me to become more independent. He made a specific effort to ensure that I developed all of the skills I needed; he helped me prepare for countless presentations, guided me through the publication and grant application processes, and encouraged me to form and develop collaborations. Additionally, with my clinical interests, he readily helped facilitate connections with clinicians and shadowing experiences for me. Tony has treated me with an incredible amount of respect and understanding. I cannot thank him enough for his genuine interest in my training and overall well-being.

I am so fortunate to have worked with some amazing individuals while in the Antonellis lab. Though Laurie Griffin and I never actually overlapped in the lab, I owe her a big thank you for helping me decide to join the lab and continuing to serve as a fantastic mentor to me ever since. To Stephanie Oprescu, thank you for teaching me and making my rotation an amazing experience, and to Chetna Gopinath, thanks for being my first bay-mate and welcoming me to the lab. To Liz Fogarty, thank you for being a great role model and always being willing to help. Rebecca Meyer-Schuman, thank you for sticking with me on the ARS side of the lab throughout the years. I am so appreciative of Natasha Golovchenko and Jenn Pierluissi, who were great lab

managers and made things so much easier in the lab. Megan Forrest, Christina Del Greco, Sheila Marte, and Allison Cale were great additions to the lab, and I thank them for great conversations and valuable input. I have particularly enjoyed being a part of Christina's and Sheila's early graduate school careers and look forward to their many successes. I am so appreciative of Maclaine Parish, an extraordinary undergraduate student, for her contributions to my projects and for helping me become a better mentor. I am very thankful for a wonderful group of colleagues and friends.

I received a tremendous amount of guidance and support from the University of Michigan biomedical community. I want to thank my committee members, Donna Martin, Miriam Meisler, and Jordan Shavit, and past committee member, Bill Dauer, who provided valuable feedback and guidance. I so appreciate the time and effort they have put towards helping me and my projects progress. Thank you to the members of the Meisler and Kitzman labs for helpful input during joint lab meetings. I sincerely thank Miriam Meisler for her help with all of the mouse work and for her overall mentorship. Thank you, also, to Guy Lenk, Young Park, and Sydney Musser for their advice and help with mouse work. I will never forget the hours Guy spent helping me with mouse dissections. I also want to say thank you to Isaac Jia and Bala Bharathi Burgula in the Kitzman lab for readily giving me advice and helping with sample preparation and analyses for sequencing experiments. I am very appreciative of the input from faculty and students in the Department of Human Genetics and the Cellular and Molecular Biology Program, who have listened to countless talks and poster presentations. I have enjoyed being a part of the research community at Michigan and am amazed at how many people have taken an interest in helping me.

I am very grateful for the many opportunities I had to participate in collaborative research; I learned immensely from our collaborators and their valuable contributions to this work. In particular, I want to thank Bill Gahl, May Malicdan, and Wendy Introne at the NIH Undiagnosed Diseases Program for including me in their project, which then launched my entire thesis research. Thank you also to Vikram Shakkottai who included me on a project to investigate variants identified in his patients and also invited me into his clinic to meet his patients. That was truly one of the most meaningful experiences of my graduate work. I will always be grateful to the patients and families who have generously contributed to this work.

I received incredible administrative and financial support from multiple programs and departments. I am very grateful to Ron Koenig, Kathy Collins, Katy Keegan, Carey Lumeng, Goutham Narla, Justine Hein, Gretchen Aland, Liz Bowman, Ellen Elkin, Hilkka Ketola, and Laurie Koivupalo with the MSTP for all of their help and support throughout these years. I am especially thankful to Ron and Kathy for their mentorship and guidance, and for Katy, who not only provided guidance and support regularly but also invited me to shadow in the Pediatric Genetics clinic. I would like to thank Bob Fuller, Kathy Collins, Ken Cadigan, Ben Allen, Manoj Puthenveedu, Patricia Ocelnik, Lauren Perl, and Jessica Kijek for all of their help and support during my time as a Cellular and Molecular Biology graduate student. Thank you to Molly Martin, Dhammika Dewasurendra, Sue Kellogg, Karen Sturtz, Kally Hang, Shaina Vera, and the rest of the Department of Human Genetics administrative staff for their support while I have been in Tony’s lab.

In addition to the Michigan community, I want to thank the countless people who helped me get to where I am and who continue to support me. To Beth Sullivan and Hunt Willard, who were fantastic mentors to me during my undergraduate training, thank you for preparing me for and encouraging me to pursue a research career. To my friends from home, high school and college, the MSTP, medical school, and graduate school, thank you for helping me keep perspective and enjoy life. To my mom, dad, and brother, thank you for your encouragement and for always being excited to celebrate my accomplishments. I am forever grateful to my parents for their love and support throughout my entire life. To Matt, my husband and very best friend, thank you for your love and unwavering support. I cannot imagine doing any of this without you.

Table of Contents

List of Tables

List of Figures

Abstract

Proteins are essential to nearly all aspects of biology, and the synthesis of proteins from genetic information (i.e., translation) is a vital cellular process. The first step of translation is performed by aminoacyl-tRNA synthetases (ARSs), a group of essential enzymes that ligate tRNA molecules to cognate amino acids. Mutations in all 37 human ARS-encoding loci have been associated with two distinct clinical presentations: (***i***) dominant axonal peripheral neuropathy; and (***ii***) severe, early-onset, multi-system recessive disease. However, our understanding of the allelic, locus, and clinical heterogeneity of ARS-related phenotypes is incomplete, and the molecular, cellular, and organismal consequences of ARS variants are poorly defined. In this dissertation, we: (**1**) expand the allelic and phenotypic heterogeneity of ARS-mediated dominant and recessive disease; (**2**) provide the first report of pathogenic cysteinyl-tRNA synthetase 1 (*CARS1*) variants in a multi-system recessive human disease; and (**3**) investigate the downstream consequences of *CARS1* variants on translation.

In Chapter 2, we evaluate the evidence for pathogenicity of 13 previously unreported ARS variants. For each variant, we assess segregation with disease, frequency in the general population, conservation of the affected amino-acid residues, and effects on gene and protein function. We expand the allelic heterogeneity of alanyl-tRNA synthetase 1 (*AARS1)*- and glycyl-tRNA synthetase 1 (*GARS1*)-associated dominant disease. Additionally, we expand the phenotypic spectrum of alanyl-tRNA synthetase 2 (*AARS2*)- and histidyl-tRNA synthetase 1 (*HARS1*)-associated recessive disease to include ataxia. Functional studies reveal that the

majority of ARS variants result in reduced function and suggest that impaired tRNA charging may contribute to disease pathogenesis in both dominant and recessive ARS-associated diseases.

In Chapter 3, we present clinical, genetic, and functional data that implicate cysteinyl-tRNA synthetase 1 *(CARS1*) variants in a complex, multi-system, recessive disease that includes microcephaly, developmental delay, and brittle hair and nails. Functional studies reveal that each variant results in complete or partial loss-of-function effects, suggesting that $tRNA^{Cys}$ charging is impaired, which may lead to defects in translation. In Chapter 4, we utilize cell, yeast, and mouse models to investigate effects of variants on global translation and specifically on the translation of cysteine-rich proteins. Assessments of patient fibroblast cells reveal no significant differences in global translation between patient and control cells; however, preliminary data from dual luciferase assays in yeast expressing disease-associated *CARS1* variants suggest that translation of cysteine-rich sequences may be impaired. Furthermore, to evaluate the effects of *CARS1* variants in a tractable mammalian model with disease-relevant tissues, a mouse model homozygous for a disease-associated *CARS1* variant is generated. Homozygous mice recapitulate the growth restriction phenotype observed in patients but display no other phenotypes. Future investigation of a mouse that is compound heterozygous for the disease-associated *CARS1* variant and a null allele will test if this more severe genotype causes effects on translation in patient-relevant tissues.

Overall, this thesis research expands the clinical, locus, and allelic heterogeneity of ARS-associated disease, provides insight into the pathogenic mechanisms and downstream consequences of ARS variants, and may lead to potential avenues for therapeutic development.

Chapter 1

Introduction to Aminoacyl-tRNA Synthetases and Disease

1.1 Translation is an essential cellular process

Nearly all aspects of biology involve and rely on proteins[1]. Proteins are polymers of amino acids that carry out myriad functions in cells such as catalyzing chemical reactions and providing structure[1]. The process of synthesizing proteins from genetic information is called translation; DNA is first transcribed into RNA, which is then translated into protein by the ribosome, a large molecular machine composed of RNA and protein[1]. Transfer RNAs (tRNAs) are also critical for deciphering messenger RNA (mRNA) sequences; each tRNA recognizes a codon (a sequence of three nucleotides) through pairing with its anticodon and delivers a specific amino acid corresponding to that codon to the translational machinery[1]. The enzymes that ligate tRNAs with the appropriate amino acids are called aminoacyl-tRNA synthetases (ARSs)[1,2], and these enzymes are the focus of this dissertation. A subset of data in this chapter was previously published in *Trends in Genetics* (Volume 36, Issue 2, pages 105-117, February 1, 2020)[3]. Permission was requested for reproduction of the article through the Copyright Clearance Center; as an author of the article, I retain the right to include the article in a dissertation, and permission is not required. The author of this thesis wrote and designed all of the figures and tables in this chapter with assistance from Matthew Pun and Jonathon Kuo for the computational analyses of the human proteome (Table 1.2). The figures were enhanced using Elsevier Illustration Services.

1.1.1 Stages of translation

There are four stages of translation: initiation, elongation, termination, and ribosome recycling[1]. First, the 7-methyl guanosine cap at the 5' end of the mRNA is bound by initiation factor eIF4E, and the poly(A) tail at the 3' end of the mRNA is bound by poly(A) binding protein[1]. These factors assemble with other initiation factors to generate a closed loop complex on the mRNA[1]. Additionally, initiation factor eIF2, GTP, and an initiator methionine-bound tRNA form the ternary complex, which binds the small ribosomal subunit to assemble the pre-initiation complex[1]. This complex associates with the mRNA closed loop complex and, along with other factors, scans the mRNA for an AUG codon and, in the case of eukaryotic cells, an associated Kozak sequence, which signals the start of an open reading frame, or protein coding region[1]. The initiator methionine-bound tRNA recognizes the AUG codon and is loaded into the P site of the small subunit of the ribosome[1]. The large ribosomal subunit then joins, and the process of elongation adds amino acids to a growing polypeptide chain as the ribosome translates the mRNA[1]. There are three basic steps of elongation: (*i*) the codon located in the A site of the ribosome indicates which aminoacyl-tRNA is loaded into the ribosome next; (*ii*) the amino acid attached to the tRNA is added to the polypeptide chain via a peptide bond; and (*iii*) an elongation factor moves the tRNA-mRNA complex through the ribosome, opening the A site for the next aminoacyl-tRNA[1]. The elongation cycle repeats until the ribosome reaches a stop codon, which is recognized by release factors that signal for the polypeptide chain to be released[1]. Finally, the large and small ribosomal subunits dissociate and are then available to bind to initiation factors toward initiating another round of translation[1].

1.1.2 Translational regulation: control and fidelity

There are multiple mechanisms that regulate translation to appropriately respond to external conditions and to ensure fidelity. Translational control is used to modulate gene expression and acts at a global level and at an mRNA-specific level[4]. Cells globally regulate translation largely through the modification of translation initiation factors[4]. One important mechanism is the phosphorylation of serine 51 on the alpha subunit of translation initiation factor eIF2 (eIF2α)[1]. Different kinases phosphorylate eIF2α in response to different stressors[5]; for example, protein kinase double-stranded RNA-dependent (PKR) responds to viral invasion, PKR-like ER kinase (PERK) responds to unfolded proteins in the endoplasmic reticulum, heme-regulated inhibitor (HRI) responds to heme deficiency, and general control non-derepressible-2 (GCN2) responds to amino acid starvation, which is of particular relevance to this thesis. During times of nutrient deprivation and decreased availability of amino acids, there is an increase of uncharged tRNA, which GCN2 binds through a domain that is related to histidyl-tRNA synthetase[1]. Upon binding to uncharged tRNA, the kinase domain of GCN2 is activated and phosphorylates eIF2α[1]. The phosphorylation of eIF2α then renders eIF2 unavailable to form the ternary complex that is needed for translation initiation, and translation is reduced globally[1]. Additionally, translation initiation factor eIF4E, which binds the mRNA 5' cap, is a target of global translational control[1]. Both the transcription and activation of eIF4E through phosphorylation are responsive to external stimuli[1]. There are also binding proteins that sequester eIF4E (4E-BPs), which decreases overall translation; the activity of 4E-BPs can be modified by phosphorylation that occurs in response to different stimuli[1]. Translation is also regulated in an mRNA-specific manner[4]. MicroRNAs (miRNAs) are small non-coding RNAs that have been found to hybridize to mRNA sequences in the 3' untranslated region (3'-UTR) and repress translation[4]. Furthermore, RNA-binding proteins

can modulate the translation of specific mRNAs by blocking or interfering with translation initiation factors[4].

Accurate translation is important for maintaining the genetic code and normal cellular function[6]. This fidelity is partly achieved by appropriate generation of aminoacyl-tRNAs by ARSs (described in more detail below). Additional points of regulation include ribosome selection of the aminoacyl-tRNA and ribosome translocation along the mRNA[6]. Translation elongation factor 1A (eEF1A) contributes to fidelity by mediating the appropriate codon and anticodon pairing[7]. During elongation, eEF1A binds and delivers aminoacyl-tRNAs to the ribosome[7]. When there is correct codon-anticodon matching, eEF1A undergoes a conformational change and hydrolyzes GTP, which allows release of eEF1A and continued elongation[7]. In addition, eEF1A may also prevent the use of misacylated tRNAs through weaker binding to misacylated tRNAs compared to appropriately aminoacylated tRNAs[6,8].

1.1.3 Translation and disease

Dysfunction of the different components of translation can lead to a variety of human disease phenotypes. Mutations in ribosomal proteins cause ribosomopathies, which are characterized by hypoproliferative phenotypes including craniofacial developmental abnormalities and anemia[6]; variants in 20 ribosomal protein genes have been implicated in the blood disorder Diamond-Blackfan anemia (MIM 105650)[9,10]. Additionally, variants in the genes encoding the subunits of the translation initiation factor EIF2B have been implicated in leukoencephalopathy with vanishing white matter (MIM 603896)[11]. Mutations in multiple elongation factor genes have been implicated in neurological diseases; *de novo* missense mutations in *EEF1A2* are associated

with epilepsy, intellectual disability, ataxia, and autistic behavior; and a missense mutation in *EEF2* was identified in patients with autosomal dominant spinocerebellar ataxia[6]. Furthermore, another group of proteins that are essential for translation and that have been linked to disease is the aminoacyl-tRNA synthetases (ARSs). In this chapter, we will review ARS function, their role in human disease, their associated tissue-predominant phenotypes, and the outstanding questions that this thesis work aims to address.

1.2 Aminoacyl-tRNA synthetases and their role in translation

ARSs are ubiquitously-expressed, essential enzymes that are responsible for charging tRNA with cognate amino acids, the first step in translating genetic information into proteins[2]. ARSs charge tRNA through a two-step reaction that involves: (*i*) activating the amino acid with ATP to form an aminoacyl adenylate intermediate; and (*ii*) transferring the aminoacyl group to the terminal ribose of the conserved CCA sequence at the 3' end of the tRNA and releasing the charged tRNA (Figure 1.1)[2]. Each ARS enzyme contains peptide sequences that are important for activating the amino acid (*e.g.*, a catalytic domain) as well as sequences important for recognizing the appropriate tRNA molecules (*e.g.*, an anticodon binding domain)[12]. There are two structural classes of ARSs, class I and class II[13]. These differ based on where they bind tRNA (class I enzymes bind the minor groove of the tRNA acceptor stem, and class II enzymes bind the major groove) and where they aminoacylate ribose (class I enzymes aminoacylate at the 2'-OH, and class II enzymes aminoacylate at the 3'-OH)[13]. ARS activity is specific to amino acid and tRNA pairings, and there are multiple mechanisms that contribute to this specificity including recognition of amino acids by the ARS active site and molecular signals in tRNAs that indicate specific aminoacylations[8,14,15]. Many ARSs have proofreading or editing activities that

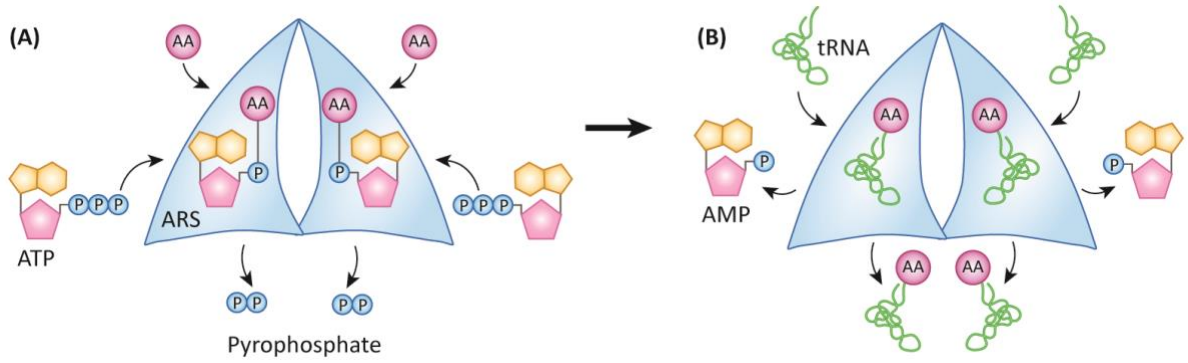


Figure 1.1. The two-step aminoacylation reaction.

(A) First, the aminoacyl-tRNA synthetase (ARS) binds the amino acid (AA) and an ATP molecule to form the aminoacyl adenylate intermediate; this process results in the release of a pyrophosphate molecule. (B) Second, a cognate tRNA molecule binds the ARS and the amino acid is transferred to the tRNA; an AMP molecule and the charged tRNA are then released. The image represents a dimeric enzyme where each subunit is capable of charging tRNA. This figure was illustrated by the author and enhanced by Elsevier Illustration Services.

prevent misincorporation of non-cognate amino acids during translation by distinguishing between cognate and non-cognate amino acids based on size or hydrophobicity and hydrolyzing misactivated amino acids[8]. The active site has been proposed to exclude non-cognate amino acids larger than the cognate amino acid due to steric hindrance[8]. If a non-cognate amino acid is misactivated, it may be hydrolyzed before and/or after the amino acid is transferred to the tRNA[8]. Some ARSs have distinct editing domains while others perform editing activities at the active site[8].

In total, there are 37 ARS genes encoded in the human nuclear genome: 18 encode cytoplasmic enzymes, 17 encode mitochondrial enzymes, and two ARS genes encode enzymes that function in both compartments[2]. Glycyl-tRNA synthetase 1 (*GARS1* [MIM 600287]) uses alternative translation initiation sites to encode both cytoplasmic and mitochondrial enzymes[16], and lysyl-tRNA synthetase 1 (*KARS1* [MIM 601421]) leads to both cytoplasmic and mitochondrial enzymes through alternative splicing of an internal exon that encodes for a mitochondrial targeting sequence[17]. There is no mitochondrial glutaminyl-tRNA synthetase identified in humans; to account for this, glutamate is incorrectly ligated to $tRNA^{Gln}$ in the mitochondria, and the transamidation pathway then converts glutamate to glutamine[18]. ARSs are abbreviated using the single-letter code for the amino acid followed by 'ARS1' for cytoplasmic enzymes or 'ARS2' for mitochondrial enzymes[2]. For example, cytoplasmic cysteinyl-tRNA synthetase (MIM 123859) is abbreviated as 'CARS1', and mitochondrial cysteinyl-tRNA synthetase (MIM 612800) is abbreviated as 'CARS2'. In vertebrates, glutamyl-prolyl-tRNA synthetase 1 (EPRS1; MIM 138295) is a bifunctional enzyme that ligates both glutamate and proline to cognate tRNA[13].

1.3 Aminoacyl-tRNA synthetases and non-canonical functions

In addition to tRNA charging, secondary functions have been described for some ARS enzymes[19,20]. This expansion of ARS functions has been linked to the addition of appended domains[21] and to protein isoforms arising from alternative splicing[22,23]. ARSs have been reported to participate in a variety of processes including mTOR signaling, inflammation, angiogenesis, apoptosis, and tumorigenesis[19,20]. For example, leucyl-tRNA synthetase 1, LARS1, has been reported to function as a leucine sensor that activates mTORC1 when leucine is present; during leucine starvation, mTORC1 is inactive, and mRNA translation initiation is inhibited[19]. In addition, fragments of tryptophanyl-tRNA synthetase 1, WARS1, are produced through alternative splicing or proteolysis and prevent angiogenesis by inhibiting vascular endothelial-cadherin[19]. Further work is needed to fully understand the spectrum of non-translational ARS functions.

1.4 Aminoacyl-tRNA synthetases and human inherited disease

Over the past 15 years, variants in ARS loci have been associated with a variety of human inherited disease phenotypes (Table 1.1). In 2003, variants in *GARS1* were first reported in patients with dominant peripheral neuropathies (also referred to as Charcot-Marie-Tooth [CMT] disease)[24]. Since then, variants in four additional ARS loci have been implicated in dominant peripheral neuropathies[25-28]. The mechanism by which ARS variants affect the peripheral nervous system has remained elusive. Additionally, variants in 36 of the 37 ARS loci have been implicated in early-onset, multi-system recessive disease phenotypes[3,29-31]. In these cases, most patients present with overlapping clinical phenotypes; however, some tissues appear to be more

Table 1.1. Phenotypes associated with ARS variants

Gene	Location of Protein Function	Mode of Inheritance	Disease Phenotype(s)	References
AARS1	Cytoplasm	Autosomal Recessive	Early-onset epileptic encephalopathy with myelination defect	32
			Microcephaly with hypomyelination, epileptic encephalopathy, and spasticity	33
			Recurrent acute liver failure	34
		Autosomal Dominant	Charcot-Marie-Tooth Disease Type 2N	26,35-38
			Distal hereditary motor neuropathy	39
AARS2	Mitochondria	Autosomal Recessive	Leukoencephalopathy with ovarian failure	40-46
			Cardiomyopathy	47,48
			Multiple respiratory chain complex defects	49
			Optic atrophy, peripheral retinal bone spicule pigmentation, demyelinating polyneuropathy	50
			Primary pulmonary hypoplasia	51
			Cerebellar ataxia	52
			Tremor, downbeat nystagmus, primary amenorrhea	48
CARS1	Cytoplasm	Autosomal Recessive	Microcephaly, developmental delay, liver dysfunction, brittle hair and nails	53
CARS2	Mitochondria	Autosomal Recessive	Epileptic encephalopathy	54
			Progressive myoclonic epilepsy	55
DARS1	Cytoplasm	Autosomal Recessive	Hypomyelination with brain stem and spinal cord involvement and leg spasticity	56,57
DARS2	Mitochondria	Autosomal Recessive	Leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation	58-65
EARS2	Mitochondria	Autosomal Recessive	Leukoencephalopathy with thalamus and brainstem involvement and high lactate	66-70
			Neonatal lactic acidosis, recurrent hypoglycemia, agenesis of corpus callosum	71
			Multiple respiratory chain complex defects	49
EPRS1	Cytoplasm	Autosomal Recessive	Hypomyelinating leukodystrophy	72
FARSA	Cytoplasm	Autosomal Recessive	Interstitial lung disease with brain calcifications	31,73
FARSB	Cytoplasm	Autosomal Recessive	Interstitial lung disease, cerebral aneurysms, brain calcifications, cirrhosis	73-76
FARS2	Mitochondria	Autosomal Recessive	Hereditary spastic paraplegia	77
			Alpers Syndrome	78
			Early onset epilepsy	79-83
			Global delay, dysarthria, and tremor	84
GARS1	Mitochondria and Cytoplasm	Autosomal Recessive	Systemic mitochondrial disease	85
			Cardiomyopathy	49
			Growth retardation, microcephaly, thinning of corpus callosum, decreased white matter, sensorineural hearing loss, dysmorphic features	86
			Charcot-Marie-Tooth Disease Type 2D	24,87-89

		Autosomal Dominant	Distal hereditary motor neuropathy	24,87,90-94
			Infantile-onset spinal muscular atrophy-like phenotype	89,90,95,96
HARS1	Cytoplasm	Autosomal Recessive	Usher Syndrome	97
			Multisystem ataxic syndrome	98
		Autosomal Dominant	Charcot-Marie-Tooth Disease Type 2W	27,99
			Distal hereditary motor neuropathy	99
HARS2	Mitochondria	Autosomal Recessive	Perrault Syndrome	100
IARS1	Cytoplasm	Autosomal Recessive	Prenatal growth retardation, neonatal cholestasis, muscular hypotonia, intellectual disability, infantile hepatopathy	101,102
IARS2	Mitochondria	Autosomal Recessive	Cataracts, growth hormone deficiency, sensory neuropathy, sensorineural hearing loss, skeletal dysplasia syndrome; Leigh Syndrome	103-105
KARS1	Mitochondria and Cytoplasm	Autosomal Recessive	Nonsyndromic hearing loss (DFNB89)	106
			Recessive intermediate Charcot Marie Tooth Disease Type B, dysmorphic features, developmental delay, self-abusive behavior, vestibular Schwannoma	107
			Visual impairment, microcephaly, seizures	108
			Impaired cognitive ability, seizure, hypotonia, ataxia	109
LARS1	Cytoplasm	Autosomal Recessive	Infantile hepatopathy	110,111
LARS2	Mitochondria	Autosomal Recessive	Perrault syndrome	112,113
			Hydrops, lactic acidosis, sideroblastic anemia, multisystem failure	114,115
MARS1	Cytoplasm	Autosomal Recessive	Interstitial lung disease and liver disease	116-119
			Pulmonary alveolar proteinosis	120
		Autosomal Dominant	Charcot-Marie-Tooth Disease Type 2U	118,121-123
MARS2	Cytoplasm	Autosomal Recessive	Developmental delay, sensorineural hearing loss	124
			Autosomal recessive spastic ataxia with leukoencephalopathy	125
NARS1	Cytoplasm	Autosomal Recessive	Neurodevelopmental disorder with microcephaly, impaired language, and gait abnormalities	30,126
		Autosomal Dominant	Neurodevelopmental disorder with microcephaly, impaired language, epilepsy, and gait abnormalities	30
NARS2	Mitochondria	Autosomal Recessive	Alpers Syndrome	127
			Developmental delay, intellectual disability, epilepsy, myopathy	128-130
			Nonsyndromic deafness	131
			Leigh Syndrome	131
PARS2	Mitochondria	Autosomal Recessive	Alpers Syndrome	127
			Infantile-onset developmental delay and epilepsy	128,132
QARS1	Cytoplasm	Autosomal Recessive	Progressive microcephaly, cerebral-cerebellar atrophy, hypomyelination, intractable seizures, developmental delay	133,134
			Severe growth deficiency, microcephaly, intellectual disability, characteristic facial features	135

RARS1	Cytoplasm	Autosomal Recessive	Hypomyelination	[136,137]
RARS2	Mitochondria	Autosomal Recessive	Pontocerebellar hypoplasia	[138-144]
			Early onset epileptic encephalopathy	[145]
			Lactic acidosis with or without neurological symptoms (microcephaly, seizures, developmental delay)	[146]
			Intellectual disability	[147]
			Myoclonic epilepsy, mental retardation, spasticity, extrapyramidal features	[148]
SARS1	Cytoplasm	Autosomal Recessive	Intellectual disability, ataxia, microcephaly, speech impairment, aggressive behavior	[149,150]
SARS2	Mitochondria	Autosomal Recessive	Hyperuricemia, pulmonary hypertension, renal failure, and alkalosis	[151]
			Spastic paresis	[152]
TARS1	Cytoplasm	Autosomal Recessive	Trichothiodystrophy	[153]
TARS2	Mitochondria	Autosomal Recessive	Axial hypotonia and severe psychomotor delay	[154]
VARS1	Cytoplasm	Autosomal Recessive	Microcephaly, seizures, cortical atrophy	[155-159]
VARS2	Mitochondria	Autosomal Recessive	Mitochondrial encephalopathy	[154,160,161]
			Encephalocardiomyopathy	[162]
			Multiple respiratory chain complex defects	[49]
WARS1	Cytoplasm	Autosomal Dominant	Distal hereditary motor neuropathy	[28,163]
WARS2	Mitochondria	Autosomal Recessive	Intellectual disability, ataxia, microcephaly, speech impairment, aggressive behavior	[149]
			Infantile onset leukoencephalopathy	[164,165]
			Infantile onset Parkinsonism	[166]
YARS1	Cytoplasm	Autosomal Recessive	Developmental delay, failure to thrive, hypertriglyceridemia, liver dysfunction, lung cysts, abnormal subcortical white matter	[167]
		Autosomal Dominant	Dominant Intermediate Charcot-Marie-Tooth Disease Type C	[25,168]
YARS2	Mitochondria	Autosomal Recessive	Myopathy, lactic acidosis, sideroblastic anemia	[169-174]
			Multiple respiratory chain complex defects	[49]

sensitive to the reduced function of a particular ARS, and the mechanisms behind this tissue specificity are not defined. Here, we review our current understanding of ARS-related inherited diseases and the associated tissue-predominant features. Below, I outline our current understanding of ARS-related genetic disease.

1.4.1 Dominant peripheral neuropathies: tRNA charging deficits and axon function

Variants in five ARS genes (*GARS1*, *YARS1* [MIM 603623], *AARS1* [MIM 601065], *HARS1* [MIM 142810], and *WARS1* [MIM 191050]) have been convincingly implicated in dominant peripheral neuropathies[29]. Inherited or *de novo* variants in four ARS genes (*GARS1*, *YARS1*, *AARS1*, and *HARS1*) cause an inherited axonal peripheral neuropathy that impairs both motor and sensory function in the distal extremities (Table 1.1)[24-27]. Variants in *AARS1*, *GARS1*, *HARS1*, and *WARS1* have also been implicated in a dominant peripheral neuropathy that is characterized by distal limb muscle atrophy without sensory involvement (Table 1.1)[28,39,92,94,99,163]. Additionally, variants only in *GARS1* have been identified in patients with a severe, dominantly-inherited, infantile-onset spinal muscular atrophy-like phenotype[89,90,95,96]. While variants in *MARS1* (MIM 156560) have been reported in patients with dominant CMT[118,121,122,175,176], a genetic argument for pathogenicity has not been established; some variants identified in patients have also been identified in relatives who are non-symptomatic[118,121,175]. Also, *de novo* variants in *NARS1* (MIM 108410) were reported in patients with a dominantly-inherited neurodevelopmental phenotype[30], but further investigation is needed to predict the pathogenicity of these variants, and whether any additional variants in noncoding sequences are present (the mutations were identified through exome sequencing).

Neuropathy-associated ARS variants are uniformly missense or in-frame deletion variants[25]. The majority of these variants show loss-of-function effects in functional assays including *in vitro* aminoacylation assays and yeast complementation assays[29,177], which are both established methods to study the functional consequences of ARS variants[177]. *In vitro* aminoacylation assays involve incubating recombinant purified ARS enzyme, tRNA, ATP, and radiolabeled amino acid[177]. The reaction is then spotted on filter paper, and the tRNA is precipitated[177]. The amount of tRNA charged is assessed by measuring radioactivity levels and calculating steady-state kinetics[177]. Yeast complementation assays test enzyme function *in vivo* by testing the ability of yeast cells to survive in the presence of the mutant ARS allele and the absence of the wild-type ARS gene[177]. Yeast cell growth is then used as a proxy for enzyme function[177].

Though data indicate loss-of-function effects for dominant-neuropathy-associated alleles, haploinsufficiency is unlikely to be the disease mechanism based on: (*i*) studies showing that heterozygous null *Gars1* mice do not have a neuropathic phenotype while mice heterozygous for pathogenic missense variants do[178,179]; and (*ii*) the frequency of null alleles in the heterozygous state in gnomAD for each of the five ARS genes implicated in dominant neuropathy[180]. Indeed, the probability of loss-of-function intolerance (pLI) has been calculated for these five genes based on the number of loss-of-function variants identified in gnomAD (pLI is 0 for *YARS1*, *AARS1*, and *HARS1*; 0.3 for *WARS1*; and 0.31 for *GARS1*), which predicts that these genes are tolerant to the loss of one allele (the closer the pLI is to one, the more intolerant the gene is to the loss of one allele)[180], further indicating that haploinsufficiency is unlikely.

Two non-mutually-exclusive hypotheses have been proposed to explain the mechanism of ARS variants in dominant neuropathy (Figure 1.2). First, since all five implicated enzymes function as homodimers in the cytoplasm, and since the variants show complete or partial loss-of-function effects, a dominant-negative mechanism may explain the dominant phenotype (Figure 1.2A)[29]. Non-functional mutant subunits may dimerize with wild-type subunits in a heterozygous patient cell and drastically reduce tRNA charging. An impairment in ARS function—and therefore translation—may be particularly detrimental to long peripheral nerve axons, potentially explaining the tissue-specificity[29]. Observations of peripheral neuropathy in patients with ARS-related recessive disease is consistent with this notion[53]. Further work investigating if dimerization is necessary for pathogenicity and if variants in monomeric enzymes can cause dominant neuropathy will help assess the role of dominant-negative effects in disease pathogenesis.

A second hypothesis states that gain-of-function effects contribute to disease pathogenesis (Figure 1.2B)[29,181]. Variants in some ARSs induce conformational openings that permit aberrant interactions with other proteins[181]. Mutant GARS1 was reported to aberrantly interact with neuropilin 1 (NRP1 [MIM 602069])[182], tropomyosin receptor kinase (TRK) receptors[183], and histone deacetylase 6 (HDAC6 [MIM 300272])[184,185]. Mutant AARS1 was also recently reported to aberrantly interact with NRP1[186]. Similarly, mutant YARS1 was reported to aberrantly interact with tripartite motif-containing protein 28 (TRIM28 [MIM 602742])[187]. These interactions have been proposed to explain the sensitivity of the peripheral nervous system to *GARS1* and *YARS1* variants via neuron-specific pathways[181]; however, it should be noted that

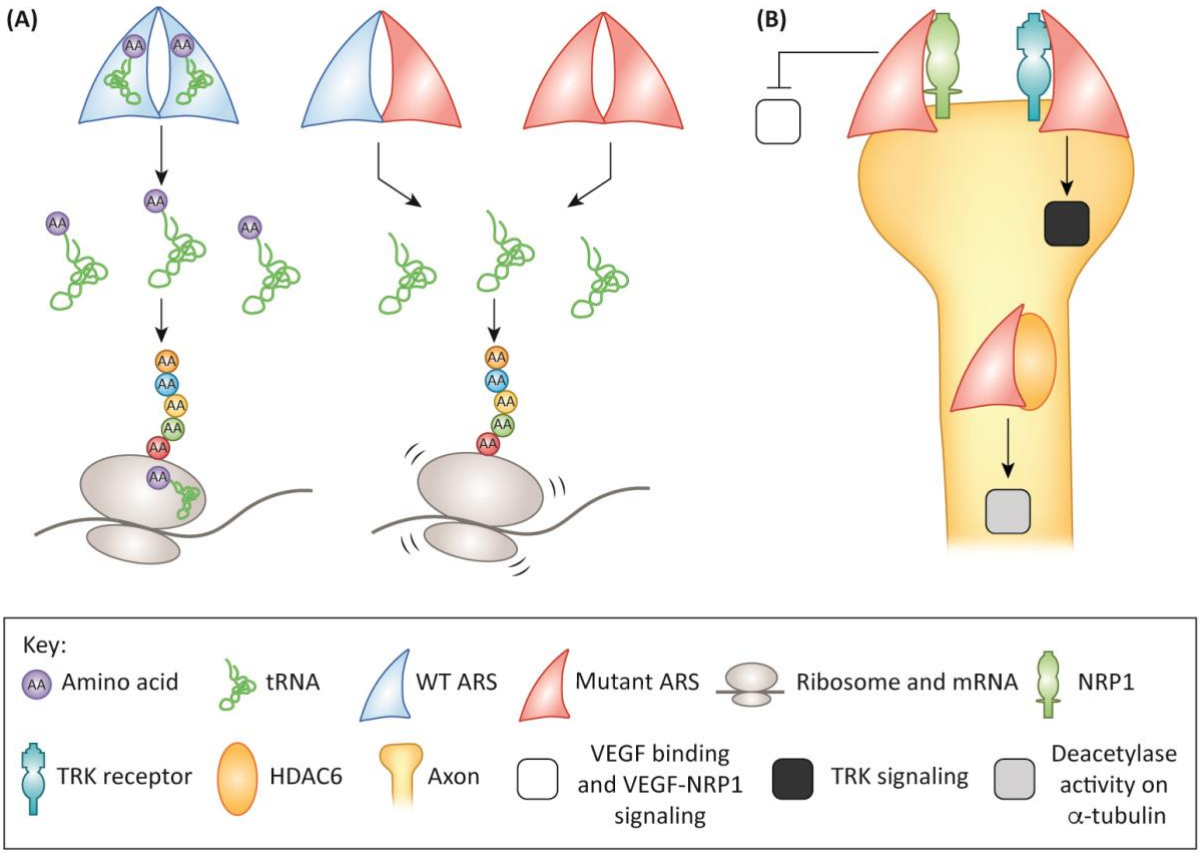


Figure 1.2. Potential mechanisms of ARS-associated dominant peripheral neuropathy.

(A) Non-functional mutant ARS subunits may dimerize with wild-type subunits in a heterozygous patient cell and drastically reduce tRNA charging—and subsequently translation—via a dominant-negative effect. (B) Mutant ARS enzymes may aberrantly bind NRP1, TRK, or HDAC6 and perturb neuronal signaling pathways. A key is provided to define each component; see text for details. This figure was illustrated by the author and enhanced by Elsevier Illustration Services.

interactions between GARS1 and NRP1 do not explain early-onset, severe, GARS1-associated neuropathy[188].

Regardless of the molecular mechanism, it is clear that the pathogenic alleles are dominantly toxic. In support of this notion, allele-specific RNAi rescues aspects of dominant neuropathy in a mouse model of *GARS1* variants[188], which is consistent with both dominant-negative and gain-of-function mechanisms.

1.4.2 Variants in tRNA synthetase genes cause a spectrum of recessive disease phenotypes

To date, genetic variants in 36 of the 37 ARS loci have been implicated in human recessive disease phenotypes (Table 1.1)[29,30,153]. These 36 loci include all 17 genes encoding an ARS enzyme that functions only in the mitochondria[29]. Unsurprisingly, disease-associated variants affecting mitochondrial ARS enzymes typically cause clinical manifestations in highly metabolic tissues. These include, but are not limited to, the central nervous system (CNS) (*CARS2*[54], *FARS2* [MIM 611592][81], *NARS2* [MIM 612803][127,128], *PARS2* [MIM 612036][127,128], *RARS2* [MIM 611524][189], *TARS2* [MIM 612805][154], *DARS2* [MIM 610956][190], *EARS2* [MIM 612799][68-70,191], *MARS2* [MIM 609728][124,125], and *WARS2* [MIM 604733][164]), muscle (*YARS2* [MIM 610957][192]), ovaries (*HARS*2 [MIM 600783][100], *LARS2* [MIM 604544][112], and *AARS2* [MIM 612035][43]) and the skeletal system (*IARS2* [MIM 612801][104])[29,193]. Diseases associated with mitochondrial ARSs have been classified based on whether variants cause clinical manifestations in the CNS and/or other tissues, highlighting the spectrum of tissues involved[193]. It is currently unclear why variants in mitochondrial ARSs cause phenotypes affecting different tissues depending on which enzyme is affected. Two models for explaining this phenomenon have been

presented[193]. First, since mitochondrial translation is important for oxidative phosphorylation, the tissue-predominant phenotypes may arise due to temporal and spatial differences in energy requirements[193]. Second, mitochondrial ARSs may have secondary functions that have yet to be discovered and that may be affected by the pathogenic variants[193]. In addition, amino-acid concentrations and mitochondrial tRNA expression may be variable in different tissues[194], which would convey differential susceptibility to impaired mitochondrial ARS function. Further work investigating tissue-specific energy requirements, secondary functions of mitochondrial ARS enzymes, and the basic biology of tRNA expression and amino-acid availability in mitochondria is needed to address these questions.

Variants in 19 of the 20 ARS loci that encode cytoplasmic enzymes have been implicated in recessive diseases that often affect a wide array of tissues (Table 1.1)[29-31,53,74,153]. To date, *WARS1* is the one remaining ARS locus that has yet to be associated with recessive disease[29]. This may be reflective of the low disease prevalence and a lack of genetic information from patients; with time and increased clinical genetic sequencing, patients will likely be identified with bi-allelic *WARS1* variants. Pathogenic alleles in the 19 ARS genes that have been implicated in recessive disease largely cause overlapping recessive clinical presentations, which often include neurodevelopmental phenotypes such as developmental delay and microcephaly[29]. Interestingly, some tissues appear to be more sensitive to defects in a particular ARS gene. For example, *CARS1* and *TARS1* variants cause brittle hair[53,153], *KARS1* variants cause sensorineural hearing loss[106], *MARS1* variants cause lung disease[120], and *LARS1* variants cause infantile liver dysfunction[110,111].

Extensive genetic and functional data implicate a partial loss-of-function molecular pathology in ARS-associated recessive disease[29]. Patients are either compound heterozygous for one missense mutation and one null allele (a variant that completely inhibits the function of the gene product), compound heterozygous for two missense mutations, or homozygous for a single missense mutation[3,29,30,153,195]. Furthermore, functional studies indicate that recessive disease-associated ARS mutations cause reduced protein levels in western blot assays, decreased mutant enzyme activity via *in vitro* aminoacylation assays, and/or diminished ability of the mutated gene to support cellular growth in yeast complementation assays[29,177]. Given that all ARSs share the common function of tRNA charging, a logical notion would be that disease-associated ARS variants cause defects in translation, which then contribute to disease pathogenesis (Figure 1.3A). However, while a preponderance of evidence points to reduced ARS function as the culprit in ARS-related recessive phenotypes, the role of secondary functions cannot be ruled out (Figure 1.3B).

While loss-of-function ARS alleles have been convincingly implicated in a range of recessive phenotypes, a critical question has not been addressed: Why do some tissues appear to be more sensitive to impairments of particular ARSs? We speculate that differing requirements across tissues to translate proteins enriched for specific amino acids may contribute to the tissue-predominant phenotypes. That is, some tissues may have high demands for proteins rich in a certain amino acid and therefore may be more sensitive to defects in charging tRNA with that amino acid. In addition, it is known that different tissues have differing availabilities of specific tRNAs[194,196] and amino acids[197,198], which could also influence whether a tissue is more or less sensitive to defects in specific ARSs. Here, we explore potential mechanisms to explain tissue-

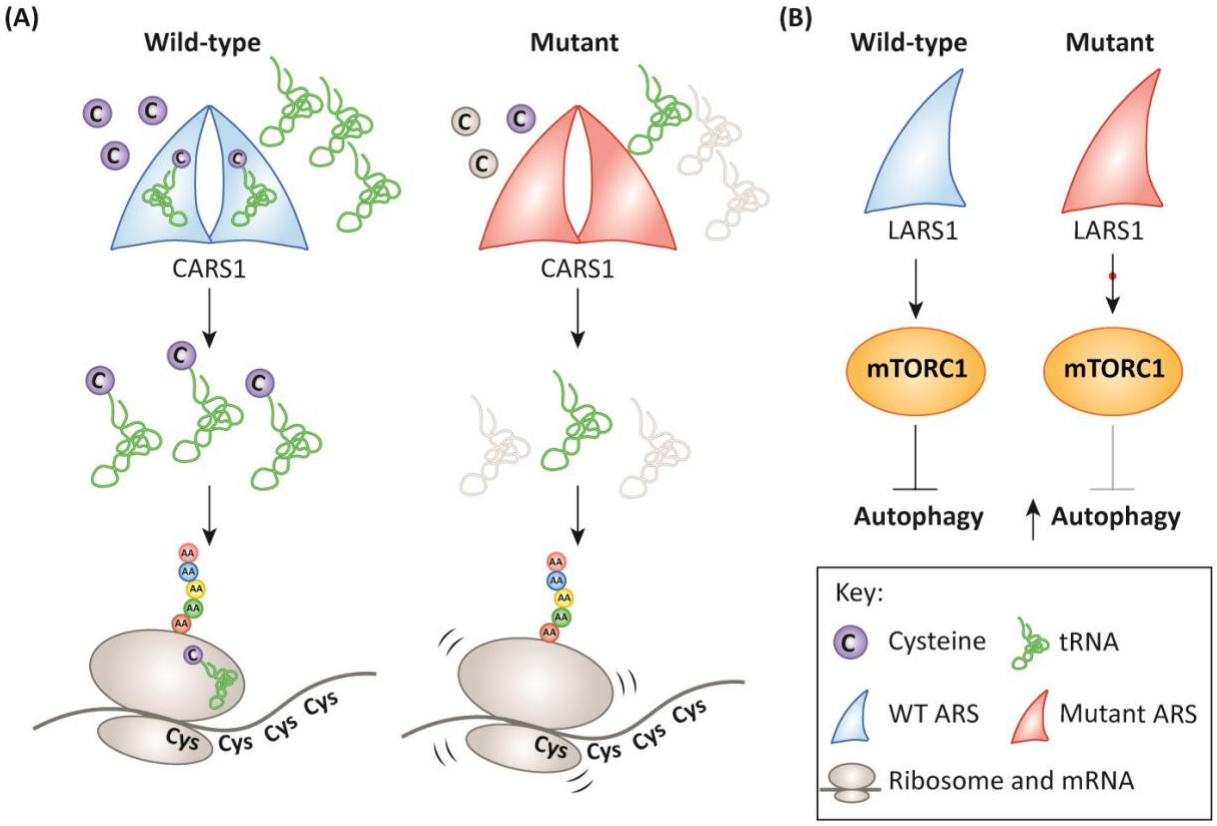


Figure 1.3. Potential mechanisms of ARS-associated recessive disease.

(A) Wild-type CARS1 (left) charges $tRNA^{Cys}$ with cysteine, and ribosomes are then able to translate transcripts with many cysteine codons. In contrast, mutant CARS1 (right) has reduced charging activity, which could result in uncharged $tRNA^{Cys}$ and the stalling of ribosomes along transcripts with many cysteine codons (indicated by small, curved lines around ribosome). The impaired charging activity may be exacerbated by reduced levels of $tRNA^{Cys}$ (grayed out tRNA) and/or cysteine (grayed out circles) in specific tissues. (B) Wild-type LARS1 (left) is an activator of mTORC1, which inhibits autophagy. Activation of mTORC1 may be impaired by mutant LARS1 (right), which could lead to abnormally increased autophagy. A key is provided to define each component; see text for details. This figure was illustrated by the author and enhanced by Elsevier Illustration Services.

predominant features of ARS-related disease using examples of ARS loci that cause recessive phenotypes (Figure 1.4).

1.4.3 Brittle hair and nails: cysteine, threonine, and keratin biology

Hair is a keratinized tissue that contains two major structural components: (***i***) hair keratins that form keratin intermediate filaments; and (***ii***) keratin-associated proteins that are located in the matrix around—and form disulfide bonds with—the keratin intermediate filaments[199]. The nail plate is another keratinized structure that also has a high content of keratin proteins and keratin-associated proteins[200,201]. Pathogenic variants in genes encoding hair and nail keratins result in diseases affecting the structure of these tissues including monilethrix (MIM 158000) and ectodermal dysplasia of hair and nail type[202,203]. Variants in two ARS loci, cysteinyl-tRNA synthetase 1 (*CARS1*) and threonyl-tRNA synthetase 1 (*TARS1*; MIM 187790), cause brittle hair and nails, suggesting that keratinocytes may be particularly sensitive to impaired tRNA charging with cysteine and threonine.

CARS1 encodes the enzyme that charges $tRNA^{Cys}$ with cysteine in the cytoplasm[2] and bi-allelic variants in *CARS1* cause a multi-system, recessive disease (see Chapter 3)[53]. *CARS1* variants were identified in four patients from three families, and each of the variants showed loss-of-function effects on CARS1 protein expression, tRNA charging, or cell viability in a yeast model[53]. The patients presented with features that overlap with other ARS-related recessive phenotypes including developmental delay, microcephaly, and liver dysfunction[53]. Interestingly, all patients also presented with brittle hair and nails[53], features that had not been reported in ARS-associated diseases until recently when patients with *TARS1* variants were reported with brittle hair[153]. In the latter study, two patients were diagnosed with trichothiodystrophy (TTD), a

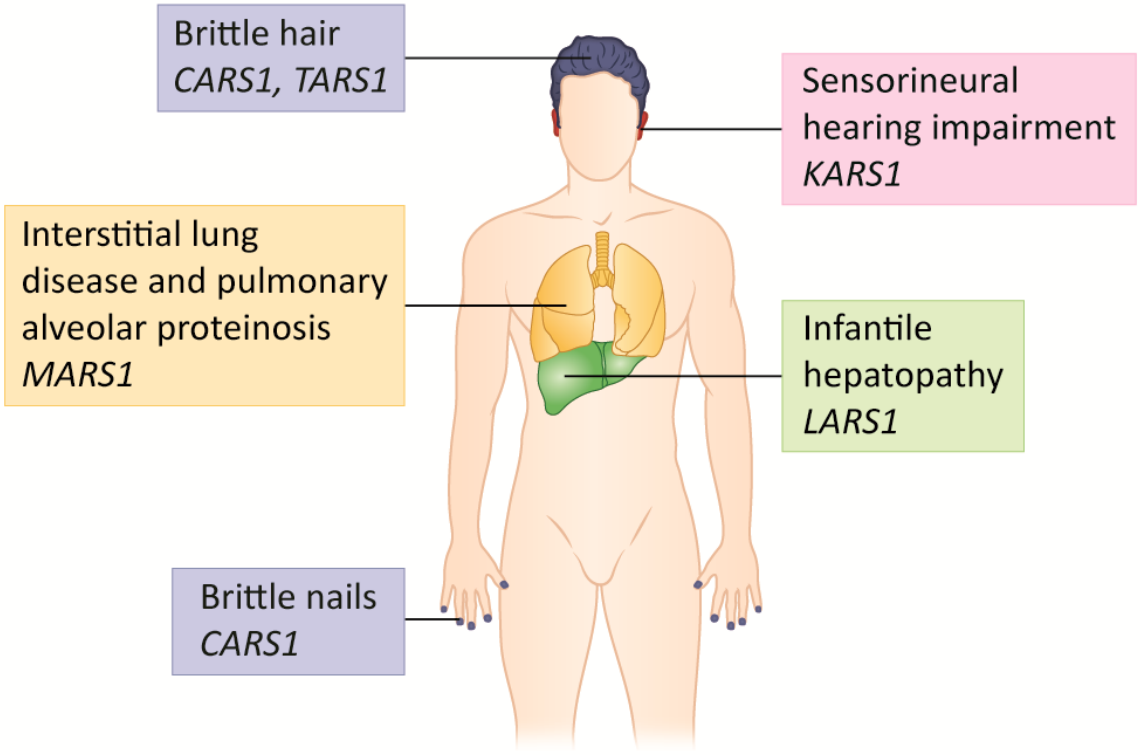


Figure 1.4. Variants in ARS genes have been associated with tissue-specific or tissue-predominant recessive phenotypes.

Variants in *CARS1* and *TARS1* have been implicated in brittle hair (blue), and *CARS1* variants have also been implicated in brittle nails (blue). Variants in *KARS1* have been associated with sensorineural hearing impairment (red). Variants in *MARS1* have been implicated in interstitial lung disease and pulmonary alveolar proteinosis (yellow). Variants in *LARS1* have been implicated in infantile hepatopathy (green). This figure was illustrated by the author and enhanced by Elsevier Illustration Services.

rare, recessive, neuroectodermal disease with characteristic brittle hair[153]. Each patient carried bi-allelic *TARS1* variants that cause loss-of-function effects on TARS1 protein levels and tRNA charging[153]. One of the patients additionally had ichthyosis, follicular keratosis, delayed physical development, recurrent respiratory tract infections, and acromandibular dysplasia[153]. The second patient was born encased in a tight membrane and had ichthyosis[153]. A nail phenotype was not reported in these patients[153].

How *CARS1* variants lead to brittle hair and nails, and how variants in *TARS1* lead to brittle hair is not understood. For *CARS1* variants, this phenotype may reflect the high cysteine requirements of hair and nails[53], since hair and nail keratins and keratin-associated proteins contain many cysteine residues that participate in disulfide bonding. In fact, we performed a computational analysis of the human proteome to identify proteins with cysteine content above the proteome average (3%) and found that 25 of the top 30 proteins with the highest percentages of cysteine (35-41%) are keratin-associated proteins (Table 1.2). Additionally, gene ontology analysis of genes encoding proteins with higher-than-average cysteine content showed an enrichment for gene products associated with 'keratinization' (GO:0031424; p-value = 4.35e-28) (Table 1.3). Since disease-associated *CARS1* variants cause a loss-of-function effect, $tRNA^{Cys}$ charging is likely impaired, which may lead to defects in cysteine-rich translation, including keratins and keratin-associated proteins; such an effect may explain the brittle hair and nail phenotype.

For *TARS1*, it has been suggested that pathogenic variants cause brittle hair due to reduced translation[153]; however, this does not explain why this phenotype is only observed in patients with *TARS1* and *CARS1* variants. Though keratin-associated proteins are not the most

Table 1.2. Top 30 human proteins with the highest percentages of each indicated amino acid.

Cysteine		*Threonine*		*Lysine*		*Methionine*		*Leucine*	
Protein	**% Cys**	**Protein**	**% Thr**	**Protein**	**% Lys**	**Protein**	**% Met**	**Protein**	**% Leu**
SCYGR5	41%	IGHD1-1	60%	TMA7	31%	SNRPC	14%	PPY2P	38%
SCYGR1	40%	MUC2	35%	H1F0	29%	ROMO1	13%	MYMX	30%
SCYGR3	39%	MUC22	34%	HIST1H1B	29%	MT-RNR1	13%	SFTA2	26%
SCYGR7	39%	MUC3A	29%	HIST1H1E	28%	RPL41	12%	MFSD3	25%
SCYGR6	38%	TRBJ2-3	25%	RPL41	28%	SPTSSA	11%	OCLM	25%
SCYGR4	37%	TRBD1	25%	HIST1H1C	28%	RTL9	10%	MT-ND3	24%
KRTAP4-6	37%	MUC17	24%	HIST1H1D	28%	WDR83OS	10%	TTTY13	24%
SCYGR8	37%	MUC12	22%	H1-1	27%	SYCE3	10%	SCGB3A2	24%
KRTAP4-12	37%	MUC3B	21%	SREK1IP1	26%	MT-ND4L	10%	MT-ND4L	23%
KRTAP4-4	37%	MUC5B	21%	CYLC2	26%	ERVK-16	10%	TMEM216	23%
SCYGR2	37%	MUC5AC	20%	FAM133B	25%	VENTXP1	9%	GLYCAM1	23%
KRTAP4-5	36%	PRG4	20%	FAM133A	25%	SLC31A1	9%	TSPO2	23%
KRTAP4-3	36%	TRBJ2-7	20%	HMGN2	24%	TIMM10	9%	GP1BB	23%
KRTAP4-7	36%	TRBJ1-2	20%	HMGN4	23%	MT-ATP8	9%	TMEM203	23%
KRTAP17-1	36%	TRBJ1-1	20%	LLPH	23%	DNM1P34	9%	SMIM40	23%
KRTAP5-7	36%	MUC19	20%	RPL38	23%	HNRNPM	9%	SLC39A5	23%
KRTAP9-1	36%	MUC21	19%	CYLC1	23%	TMEM262	9%	TMEM82	23%
KRTAP5-9	36%	TRGJ1	19%	FAU	22%	SCGB2A1	8%	BPIFA1	23%
KRTAP4-11	35%	TRDJ1	19%	TRDN	22%	MTRNR2L5	8%	IL11	23%
KRTAP5-8	35%	HCG22	19%	RPS25	22%	S100A10	8%	SLC35A4	23%
KRTAP5-11	35%	HAVCR1	18%	RPL39P5	22%	MED18	8%	NRM	23%
KRTAP4-9	35%	KLF18	18%	FAM32A	21%	SS18	8%	SFT2D3	22%
KRTAP4-8	35%	MUCL1	18%	C19orf53	21%	BEX3	8%	IGFALS	22%
KRTAP5-6	35%	MUC16	18%	RPL35	21%	SMAP2	8%	TMEM35B	22%
KRTAP4-2	35%	TRBJ1-6	18%	HMGN1	21%	S100A7L2	8%	CYB561D2	22%
MT1B	34%	PLAC4	17%	RPL36A	21%	SNRPG	8%	BIK	22%
KRTAP4-1	34%	MUC6	17%	RPL13AP3	21%	SOX2	8%	IL27	22%
KRTAP5-3	33%	TMEM123	17%	TMSB4X	20%	ERVK-21	8%	LRRC32	22%
MT1F	33%	MTRNR2L9	17%	SERF2	20%	MLF2	8%	TMEM140	22%
MT1E	33%	MTRNR2L6	17%	RPL29	20%	CHMP3	8%	MYMX	30%

The human proteome was obtained from UniProt (Proteome ID UP000005640; accessed November 3, 2019). Proteins from the Swiss-Prot UniProt database are shown and uncharacterized proteins were removed. Biostrings in R was used to determine the number of each amino acid in each protein. For each protein, percent for a given amino acid was calculated by dividing the number of that specific amino acid by the total number of amino acids. Here we show the top 30 proteins for each of the listed amino acids.

Table 1.3. Gene ontology results for genes encoding proteins with higher-than-average percentages of each indicated amino acid

Cysteine

Biological Process	Fold Enrichment	P-Value
lymphocyte chemotaxis (GO:0048247)	3.96	1.04E-04
keratinization (GO:0031424)	3.80	4.35E-28
negative regulation of signaling receptor activity (GO:2000272)	3.76	1.49E-05
chemokine-mediated signaling pathway (GO:0070098)	3.53	1.03E-06
monocyte chemotaxis (GO:0002548)	3.50	3.13E-02
keratinocyte differentiation (GO:0030216)	3.40	4.91E-26
antimicrobial humoral immune response mediated by antimicrobial peptide (GO:0061844)	3.33	4.55E-08
disruption of cells of other organism (GO:0044364)	3.33	6.71E-06
killing of cells of other organism (GO:0031640)	3.33	6.71E-06
cellular response to chemokine (GO:1990869)	3.20	1.07E-05
response to chemokine (GO:1990868)	3.20	1.07E-05
lymphocyte migration (GO:0072676)	3.19	9.97E-04
epidermal cell differentiation (GO:0009913)	3.08	5.87E-24
negative regulation of BMP signaling pathway (GO:0030514)	3.08	1.07E-02
neutrophil chemotaxis (GO:0030593)	3.04	2.73E-04
granulocyte chemotaxis (GO:0071621)	3.00	1.20E-04
defense response to Gram-negative bacterium (GO:0050829)	2.94	2.71E-04
antimicrobial humoral response (GO:0019730)	2.94	4.82E-08
neutrophil migration (GO:1990266)	2.91	4.57E-04
granulocyte migration (GO:0097530)	2.85	1.55E-04
skin development (GO:0043588)	2.84	1.01E-23
defense response to Gram-positive bacterium (GO:0050830)	2.80	3.57E-04
epidermis development (GO:0008544)	2.78	9.23E-25
cell killing (GO:0001906)	2.76	5.28E-05
leukocyte chemotaxis (GO:0030595)	2.76	7.02E-07
myeloid leukocyte migration (GO:0097529)	2.69	3.58E-05
regulation of BMP signaling pathway (GO:0030510)	2.62	3.89E-03
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	2.59	4.12E-21
positive regulation of leukocyte chemotaxis (GO:0002690)	2.58	2.27E-02
regulation of leukocyte chemotaxis (GO:0002688)	2.51	3.29E-03

Threonine

Biological Process	Fold Enrichment	P-Value
homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156)	2.15	1.88E-06
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	2.14	8.25E-21
sensory perception of smell (GO:0007608)	2.10	7.96E-21
detection of chemical stimulus involved in sensory perception (GO:0050907)	2.07	1.02E-20
sensory perception of chemical stimulus (GO:0007606)	2.01	4.10E-21
detection of chemical stimulus (GO:0009593)	2.01	3.36E-20
cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742)	1.99	3.04E-08
detection of stimulus involved in sensory perception (GO:0050906)	1.96	3.10E-19
detection of stimulus (GO:0051606)	1.79	1.33E-17
cell-cell adhesion (GO:0098609)	1.75	1.83E-10
lymphocyte mediated immunity (GO:0002449)	1.63	1.93E-02
adaptive immune response (GO:0002250)	1.59	8.18E-08
sensory perception (GO:0007600)	1.58	5.44E-14
G protein-coupled receptor signaling pathway (GO:0007186)	1.57	2.40E-19
cell adhesion (GO:0007155)	1.57	2.21E-12
biological adhesion (GO:0022610)	1.56	5.48E-12
nervous system process (GO:0050877)	1.49	2.63E-15
leukocyte mediated immunity (GO:0002443)	1.40	9.98E-04
system process (GO:0003008)	1.39	2.12E-14
regulation of immune response (GO:0050776)	1.31	1.83E-03

immune response (GO:0006955)	1.31	1.56E-07
positive regulation of immune system process (GO:0002684)	1.29	1.00E-02
locomotion (GO:0040011)	1.27	5.81E-03
transmembrane transport (GO:0055085)	1.27	1.06E-02
regulation of immune system process (GO:0002682)	1.26	7.89E-04
immune system process (GO:0002376)	1.24	1.86E-07
cell surface receptor signaling pathway (GO:0007166)	1.21	2.92E-04
signal transduction (GO:0007165)	1.21	4.62E-14
signaling (GO:0023052)	1.21	2.15E-14
response to chemical (GO:0042221)	1.20	2.95E-10

Lysine

Biological Process	Fold Enrichment	P-Value
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	2.86	3.83E-08
cotranslational protein targeting to membrane (GO:0006613)	2.78	5.60E-08
protein targeting to ER (GO:0045047)	2.75	1.61E-08
establishment of protein localization to endoplasmic reticulum (GO:0072599)	2.71	1.85E-08
translational initiation (GO:0006413)	2.52	7.19E-09
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	2.49	9.53E-07
protein localization to endoplasmic reticulum (GO:0070972)	2.46	1.29E-07
mitotic cytokinesis (GO:0000281)	2.39	3.36E-02
viral transcription (GO:0019083)	2.27	3.19E-04
viral gene expression (GO:0019080)	2.17	3.32E-04
cellular protein complex disassembly (GO:0043624)	2.09	1.17E-03
nucleosome organization (GO:0034728)	2.09	1.33E-04
chromatin assembly or disassembly (GO:0006333)	2.07	3.25E-04
nucleosome assembly (GO:0006334)	2.06	1.52E-02
chromatin assembly (GO:0031497)	2.06	3.98E-03
rRNA processing (GO:0006364)	2.06	1.49E-06
translation (GO:0006412)	2.05	1.98E-14
DNA packaging (GO:0006323)	2.05	6.70E-05
translational elongation (GO:0006414)	2.04	9.89E-03
rRNA metabolic process (GO:0016072)	2.02	1.64E-06
ribosome biogenesis (GO:0042254)	2.00	3.39E-09
protein folding (GO:0006457)	1.99	6.69E-06
peptide biosynthetic process (GO:0043043)	1.98	2.03E-13
protein targeting to membrane (GO:0006612)	1.97	5.69E-04
DNA conformation change (GO:0071103)	1.96	4.25E-08
protein-containing complex disassembly (GO:0032984)	1.94	1.57E-05
nuclear-transcribed mRNA catabolic process (GO:0000956)	1.93	2.87E-04
M phase (GO:0000279)	1.93	3.18E-03
mitotic prometaphase (GO:0000236)	1.93	3.18E-03
mitotic M phase (GO:0000087)	1.93	3.18E-03

Methionine

Biological Process	Fold Enrichment	P-Value
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	2.66	1.26E-41
sensory perception of smell (GO:0007608)	2.6	7.07E-42
detection of chemical stimulus involved in sensory perception (GO:0050907)	2.56	9.08E-42
detection of chemical stimulus (GO:0009593)	2.45	1.08E-39
sensory perception of chemical stimulus (GO:0007606)	2.45	2.42E-41
detection of stimulus involved in sensory perception (GO:0050906)	2.4	1.23E-39
detection of stimulus (GO:0051606)	2.1	2.40E-34
proton transmembrane transport (GO:1902600)	2	5.68E-03
organic acid transmembrane transport (GO:1903825)	1.95	5.66E-03
carboxylic acid transmembrane transport (GO:1905039)	1.94	7.39E-03
amino acid transport (GO:0006865)	1.92	2.96E-02
sodium ion transport (GO:0006814)	1.91	1.11E-03

sensory perception (GO:0007600)	1.87	1.04E-30
organic acid transport (GO:0015849)	1.81	9.50E-06
G protein-coupled receptor signaling pathway (GO:0007186)	1.79	3.27E-37
anion transmembrane transport (GO:0098656)	1.74	5.73E-09
carboxylic acid transport (GO:0046942)	1.69	5.77E-03
inorganic ion transmembrane transport (GO:0098660)	1.68	3.97E-12
inorganic cation transmembrane transport (GO:0098662)	1.66	3.94E-09
nervous system process (GO:0050877)	1.65	8.77E-27
cation transmembrane transport (GO:0098655)	1.63	8.90E-10
ion transmembrane transport (GO:0034220)	1.63	4.65E-18
transmembrane transport (GO:0055085)	1.61	1.41E-21
organic anion transport (GO:0015711)	1.59	4.61E-03
cation transport (GO:0006812)	1.56	3.36E-10
system process (GO:0003008)	1.53	7.55E-28
monocarboxylic acid metabolic process (GO:0032787)	1.49	1.03E-03
oxoacid metabolic process (GO:0043436)	1.48	1.83E-07
organic acid metabolic process (GO:0006082)	1.47	1.68E-07
carboxylic acid metabolic process (GO:0019752)	1.46	2.00E-06

Leucine

Biological Process	Fold Enrichment	P-Value
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	2.44	7.55E-35
detection of chemical stimulus involved in sensory perception (GO:0050907)	2.39	6.60E-37
sensory perception of smell (GO:0007608)	2.37	1.96E-34
detection of chemical stimulus (GO:0009593)	2.32	2.60E-36
sensory perception of chemical stimulus (GO:0007606)	2.28	1.09E-35
detection of stimulus involved in sensory perception (GO:0050906)	2.24	3.84E-34
detection of stimulus (GO:0051606)	2.00	5.66E-30
G protein-coupled receptor signaling pathway (GO:0007186)	1.82	4.93E-43
membrane lipid metabolic process (GO:0006643)	1.73	1.05E-02
anion transmembrane transport (GO:0098656)	1.70	2.37E-04
sensory perception (GO:0007600)	1.70	2.58E-22
organic anion transport (GO:0015711)	1.63	6.98E-07
organic acid transport (GO:0015849)	1.61	6.27E-03
carboxylic acid transport (GO:0046942)	1.61	6.27E-03
anion transport (GO:0006820)	1.58	7.86E-08
transmembrane transport (GO:0055085)	1.52	1.36E-17
ion transmembrane transport (GO:0034220)	1.49	1.88E-09
lipid biosynthetic process (GO:0008610)	1.47	1.79E-04
nervous system process (GO:0050877)	1.46	1.44E-14
ion transport (GO:0006811)	1.45	2.46E-13
cation transport (GO:0006812)	1.43	3.11E-06
cation transmembrane transport (GO:0098655)	1.41	1.50E-03
metal ion transport (GO:0030001)	1.40	3.01E-03
inorganic ion homeostasis (GO:0098771)	1.39	9.18E-04
cellular ion homeostasis (GO:0006873)	1.39	3.94E-03
inorganic ion transmembrane transport (GO:0098660)	1.39	4.21E-03
cation homeostasis (GO:0055080)	1.39	1.79E-03
ion homeostasis (GO:0050801)	1.38	3.25E-04
inorganic cation transmembrane transport (GO:0098662)	1.38	3.46E-02
cellular cation homeostasis (GO:0030003)	1.38	8.84E-03

(*Legend on next page*)

Proteins with higher than average cysteine (>3% cysteine; 3,578 proteins), threonine (>5% threonine; 7,496 proteins), lysine (>6% lysine; 6,724 proteins), methionine (>2% methionine), and leucine (>10% leucine; 8,438 proteins) were analyzed using geneontology.org[204,205] to identify biological processes that are significantly enriched for proteins with a high percentage of the indicated amino acid. PANTHER Overrepresentation Test (Released 20190711 or Released 20210224 for methionine)[206] using the GO Ontology database (Released 2019-10-08 or Released 2021-02-01 for methionine) was used with Fisher's Exact test and Bonferroni correction for multiple testing. The top 30 biological processes with a p-value < 0.05 are listed and ranked by fold enrichment.

threonine-rich proteins in the proteome, many cysteine-rich, keratin-associated proteins also have threonine content higher than the proteome average (>5%). For example, keratin-associated protein 9-2 (KRTAP9-2), which is 33% cysteine, is also 16% threonine. Gene ontology analysis of genes encoding proteins with higher than average threonine content showed an enrichment for gene products associated with adhesion including 'cell-cell adhesion' (GO:0098609; p-value = 1.83e-10) (Table 1.3), which is an important process for keratinocytes[207]. Further work is needed to determine if cysteine- and threonine-rich proteins are affected by *CARS1* and *TARS1* variants.

1.4.4 Sensorineural hearing loss: lysine and the inner ear

The inner ear has a complex cellular organization and, therefore, variants in a large panel of genes can cause hearing loss[106,208]. Lysyl-tRNA synthetase 1 (*KARS1*) encodes the enzyme that charges $tRNA^{Lys}$ with lysine in both the cytoplasm and the mitochondria[17]. Interestingly, bi-allelic *KARS1* variants have been associated with recessive, non-syndromic hearing impairment[106], suggesting that inner ear cells may be particularly sensitive to impairments in tRNA charging with lysine. In addition to non-syndromic hearing impairment, *KARS1* variants have been implicated in recessive, syndromic hearing loss. In these latter patients, hearing loss is comorbid with other symptoms commonly associated with ARS variants including developmental delay, neurocognitive decline, seizures, and leukoencephalopathy[109,209,210].

It is not currently understood how the inner ear is particularly sensitive to mutations in *KARS1*. It has been suggested that aminoacylation impairments in the cochlea might affect many cellular processes and lead to hearing impairment[106]. A general sensitivity to impaired translation may explain why variants in other ARS loci (*LARS2*[112], *HARS2*[100], *HARS1*[97], *AARS1*[35], and *RARS2*[140])

have also been associated with broader syndromes that include hearing impairment, but it does not explain the non-syndromic hearing loss associated with *KARS1* variants. We wondered if inner ear cells have high lysine requirements that may explain this tissue-predominant phenotype. We analyzed the human proteome to identify proteins with percentages of lysine higher than the proteome average (6%) and then considered if those proteins had roles in the inner ear (Table 1.2). One protein of particular interest with high lysine content is growth arrest-specific 8 (GAS8 [MIM 605178]), which contains 12% lysine. GAS8 is a component of the dynein regulatory complex that is important for ciliary motility and for ear development[211], and loss-of-function variants in *GAS8* were implicated in autosomal recessive ciliary dyskinesia (MIM 616726)[212]. Similarly, the transmembrane inner ear expressed gene (*TMIE* [MIM 607237]) has been associated with autosomal recessive non-syndromic hearing loss[213] and encodes a protein with 15% lysine. Gene ontology analysis of genes encoding proteins with higher-than-average lysine content showed the greatest enrichment for gene products associated with protein targeting including 'SRP-dependent cotranslational protein targeting to membrane' (GO:0006614; p-value = 3.83e-8) and 'protein targeting to ER' (GO:0045047; p-value = 1.61e-8) (Table 1.3); however, it is unclear how perturbations in these processes would specifically affect the process of hearing. Future work is needed to investigate the roles of lysine-rich proteins in the inner ear and if the translation of lysine-rich proteins is specifically affected by *KARS1* variants.

1.4.5 Interstitial lung disease and pulmonary alveolar proteinosis: methionine and lung function

In the lung, alveoli allow gas exchange[214,215], and in disorders that affect the alveoli, patients often have difficulty breathing. Methionyl-tRNA synthetase (*MARS1*) encodes the enzyme that

charges $tRNA^{Met}$ with methionine in the cytoplasm[2], and interestingly, variants in *MARS1* have been associated with lung-predominant disease phenotypes. Bi-allelic variants in *MARS1* cause interstitial lung and liver disease (ILLD; MIM 615486), a recessive phenotype of respiratory insufficiency and liver disease[116-118,120,216]. In interstitial lung diseases, damage and scarring occurs in the interstitium, the tissue around the alveoli[217,218]. Van Meel *et al.* first reported a patient with a multi-system disorder including interstitial lung disease, liver failure, and failure to thrive; aminoacylation assays showed that each variant was associated with reduced charging activity[116]. Hadchouel *et al.* then identified *MARS1* variants in patients with a severe childhood form of pulmonary alveolar proteinosis (PAP) prevalent on Réunion Island[120]. In PAP, lipoproteinaceous material builds up in the alveoli and blocks oxygen exchange[219,220]. These patients often develop lung fibrosis and present with liver involvement, failure to thrive, and systemic inflammation[120]. These variants showed loss-of-function effects in aminoacylation assays[221]. This association between *MARS1* variants and pulmonary disease suggests that alveolar cells and other related lung cells may be particularly susceptible to impairments in charging tRNA with methionine.

Since ATG is the initiator codon, it might be expected that variants in *MARS1* would cause the broadest effects; why the lung is so severely affected by *MARS1* variants is unclear. We wondered whether the lung has high requirements for methionine, which may explain this sensitivity. We sorted the human proteome based on percentage methionine, and no obvious trends were immediately identified in our analysis (Table 1.2). Surfactant pulmonary-associated surfactant protein C (SFTPC [MIM 178620]), which has been implicated in the interstitial lung disease surfactant metabolism dysfunction type 2 (MIM 610913)[218,222], has 5% methionine,

which is higher than the average 2%. Gene ontology analysis of genes encoding proteins with higher-than-average methionine content showed enrichment for gene products associated with transport including 'proton transmembrane transport' (GO:1902600; p-value = 5.68e-3) and 'organic acid transmembrane transport' (GO:1903825; p-value = 5.66e-3), and it is unclear how disruptions to these processes may specifically affect pulmonary processes. More work is needed to determine whether methionine-rich proteins are important in the lung and whether their expression is affected by *MARS1* variants.

1.4.6 Infantile hepatopathy: leucine and liver function

Infantile liver disease is a life-threatening condition that develops in early infancy and can be difficult to diagnose[110]. Patients present with failure to thrive, jaundice, and distended abdomen, and may have elevated liver transaminases and hyperbilirubinemia[110]. Leucyl-tRNA synthetase 1 (*LARS1*) encodes the enzyme that charges $tRNA^{Leu}$ with leucine in the cytoplasm[2]. Bi-allelic variants in *LARS1* cause infantile liver failure syndrome type 1 (ILFS1; MIM 615438), a multi-system disorder involving infantile liver failure, developmental delay, anemia, seizures, and recurrent liver dysfunction triggered by febrile illness[110,111,223,224]. Notably, bi-allelic variants in many ARS genes result in phenotypes that affect the liver[29,74], which may indicate that this organ is generally sensitive to impaired translation. However, *LARS1* variants seem to predominantly affect the liver and cause earlier onset and a more severe liver phenotype; perhaps this liver-predominant effect may be due to impaired translation of proteins with high leucine content. It has been proposed that proteins with high levels of leucine were most likely to be affected in patients with *LARS1* variants[110]. Previous analysis of the human proteome determined that immune-related proteins contain very high percentages of leucine, which correlates with the

patient phenotype of hepatic dysfunction triggered with illness[110]. Additionally, leucine-rich proteins are involved in the phospholipid biosynthetic process, and in support of this observation, large fat deposits were seen in affected livers[110]. Our gene ontology analysis of genes encoding proteins with higher-than-average leucine content showed an enrichment for similar terms including 'lipid biosynthetic process' (GO:0008610; p-value = 1.79e-4) (Table 1.3). Furthermore, when analyzing the human proteome for leucine-rich proteins, we found that neuroblastoma-amplified sequence (*NBAS*; MIM608025), which is associated with infantile liver failure syndrome 2 (MIM 616483)[225], encodes a protein with 13% leucine; this is a higher percentage than the proteome average of 10%.

As mentioned above, loss of a secondary function could explain the tissue-predominant effects of ARS variants, and there are data suggestive of this for *LARS1*-associated disease. It has been reported that LARS1 is a leucine sensor and an activator of mTORC1 (an autophagy inhibitor), and that LARS1 deficiency results in reduced activation of mTORC1 and increased autophagy[226]. Reduced hepatic mTORC1 activity has been shown to cause liver cell damage; therefore, variants in *LARS1* may contribute to disease through reduced mTORC1 activity and abnormal autophagy[111]. Aberrant mTORC1 signaling may also contribute to the seizure phenotype observed in these patients since mTOR signaling has been associated with neurological disorders including epilepsy[227]. Future research is needed to determine if leucine-rich protein expression is specifically affected by *LARS1* variants, and if the disease-associated alleles affect mTORC1 activity.

1.5 Summary

While variants in all 37 human ARS loci have been implicated in genetic disease, our understanding of the clinical and allelic heterogeneity of ARS-related disease is incomplete. Many questions on the mechanisms of disease pathogenesis remain unanswered, including: What are the downstream consequences of impaired ARS function on translation? Does impaired ARS function result in a global defect in translation and/or a more dramatic effect on the translation of a specific subset of proteins? Why do mutations in these ubiquitously-expressed enzymes result in tissue-specific or tissue-predominant phenotypes?

Five ARS genes have been linked with dominant peripheral neuropathies, which may reflect sensitivity of the peripheral nervous system to impairments in translation and/or to aberrant protein interactions that affect signaling pathways. Variants in the majority of ARSs have been linked with recessive phenotypes, and certain ARSs are associated with tissue-predominant phenotypes. The explanation for the observed tissue specificity is not understood and may reflect: (*i*) differing requirements across tissues to translate proteins enriched for specific amino acids; (*ii*) differing availability of tRNA and amino acids across tissues; and/or (*iii*) impairment of secondary functions of ARS enzymes that are important for specific tissues.

In this thesis, I will present studies implicating newly identified ARS variants in human disease, expanding the clinical and allelic spectrum of dominant and recessive ARS-associated disease (Chapter 2). Additionally, I will describe clinical, genetic, and functional data that for the first time implicated cysteinyl-tRNA synthetase 1 (*CARS1*) in disease (Chapter 3), increasing the number of ARSs associated with disease and linking impaired CARS1 function with a tissue-

predominant brittle hair and nail phenotype. Finally, I will describe studies investigating the downstream consequences of *CARS1* variants on global translation and the translation of cysteine-rich proteins (Chapter 4). This work aims to provide insights into the biological function of ARS enzymes; expand the clinical, locus, and allelic spectrum of ARS variants; advance our understanding of the mechanisms of ARS-associated diseases; and contribute towards understanding how mutations in ubiquitously-expressed genes can result in tissue-predominant phenotypes.

Chapter 2

Expanding the Allelic and Phenotypic Heterogeneity of ARS-Mediated Dominant and Recessive Disease

2.1 Introduction

Aminoacyl-tRNA synthetases (ARSs) are essential enzymes responsible for charging tRNA molecules with cognate amino acids. To date, variants in five genes encoding an ARS have been implicated in dominant peripheral neuropathies[29]. The majority of these variants demonstrate complete or partial loss-of-function effects in functional assays; however, our understanding of how variants in ubiquitously expressed ARS loci cause this tissue-specific neuropathy phenotype is incomplete. Haploinsufficiency is unlikely to be the disease mechanism due to the frequency of null alleles in the heterozygous state in presumably healthy individuals in the general population[180,228], and heterozygous parents of patients with ARS-associated recessive disease are unaffected[29]. Additionally, a mouse heterozygous for one null glycyl-tRNA synthetase 1 (*Gars1*) allele did not display a phenotype[178]. As a result of these observations, two viable hypotheses have been proposed: a dominant-negative mechanism and toxic gain-of-function effects[29]. First, since all five implicated enzymes function as homodimers in the cytoplasm, and the variants show loss-of-function effects, a dominant-negative mechanism may explain the dominant phenotype[29]. Non-functional mutant subunits may dimerize with wild-type subunits in a heterozygous patient cell and drastically reduce tRNA charging. Second, mutations in ARSs may cause gain-of-function effects. Variants in three ARS loci have been reported to induce conformational openings that permit aberrant interactions with other proteins[181]: (1) mutant

alanyl-tRNA synthetase 1 (AARS1) interacts with neuropilin 1 (NRP1)[186]; (2) mutant glycyl-tRNA synthetase 1 (GARS1) interacts with NRP1[182], tropomysin receptor kinase (TRK) receptors[183], and histone deacetylase 6 (HDAC6)[184,185]; and (3) mutant tyrosyl-tRNA synthetase 1 (YARS1) interacts with tripartite motif-containing protein 28 (TRIM28)[187]. These interactions may explain the sensitivity of the peripheral nervous system to these ARS variants through neuron-specific pathways[181]. Investigating newly identified ARS variants in patient populations will improve our understanding of the clinical and allelic spectrum of ARS-associated dominant disease and lead to insights into how these mutations cause peripheral neuropathies.

Consistent with the essential function and ubiquitous expression of ARSs, mutations in 36 of the 37 ARS-encoding loci have been implicated in severe, early-onset recessive phenotypes[29]. There is an abundance of evidence supporting partial loss-of-function as the molecular mechanism of ARS-mediated recessive disease[29]. Patients are homozygous for a single missense mutation, compound heterozygous for two missense mutations, or compound heterozygous for one missense mutation and one null allele[3,29,30,153,195]. Additionally, recessive disease-associated ARS mutations cause decreased mutant enzyme activity in aminoacylation assays and/or support reduced cellular growth in yeast complementation assays[29,177]. Importantly, the evidence to date suggests a reduction of ARS function but not a complete loss of function; a complete loss of ARS function is predicted to be lethal given the essential nature of these genes. ARS variants have been implicated in a range of recessive phenotypes[3,29], and the degree to which the two ARS alleles impair ARS function may contribute to the clinical heterogeneity of ARS-associated recessive phenotypes. Interestingly, variants in particular ARS genes have been associated with

tissue-specific or tissue-predominant phenotypes (see Chapter 1), and how these mutations lead to these cell-specific phenotypes is not understood[3].

In this chapter we investigate: (1) *AARS1* and *GARS1* variants identified in patients with peripheral neuropathies; and (2) alanyl-tRNA synthetase 2 (*AARS2*), histidyl-tRNA synthetase 1 (*HARS1*), *AARS1*, and tyrosyl-tRNA synthetase 1 (*YARS1*) variants identified in patients with recessive phenotypes (Table 2.1). We evaluate the evidence for pathogenicity of these newly identified ARS variants by assessing: (1) segregation of the variants with disease, (2) absence or presence of the variants in the general population, (3) conservation of the affected residues, and (4) functional characterization of the variants. We consider how the variants may affect the functional domains of the proteins; the majority of ARS enzymes contain: (1) a catalytic (or aminoacylation) domain that is important for activating the amino acid; and (2) an anticodon binding domain that is important for recognizing tRNA[12]. Additionally, AARS1 and AARS2 each have an editing domain that prevents misincorporation of non-cognate amino acids and a C-terminal domain that acts as a bridge between the editing and aminoacylation domains[229]. Mitochondrial enzymes, like AARS2, contain a mitochondrial targeting sequence (MTS) that signals translocation to the mitochondria[230]. *GARS1* uses alternative translation initiation start sites to encode the enzymes that charge $tRNA^{Gly}$ with glycine in both the cytoplasm and mitochondria[16]; as a result, the cytoplasmic and mitochondrial forms share the same polypeptide sequence except for an MTS in the mitochondrial GARS1. Some ARSs, including GARS1 and HARS1, contain a WHEP domain that may contribute to tRNA or protein binding, but the function of the WHEP domain has not been fully defined[231]. In GARS1, the WHEP domain is not essential for enzyme activity, and in yeast complementation assays, expression of GARS1

Table 2.1. Summary of newly identified patients with ARS variants

Patient Identifier	Mode of Inheritance	Genotype	Disease Phenotype
AARS1 Patients 1 and 2 (AARS1 Family 1)	Dominant	+/S627L *AARS1*	Charcot-Marie-Tooth Disease Type 2N
AARS1 Patients 3-8 (AARS1 Family 2)	Dominant	+/E337K *AARS1*	Charcot-Marie-Tooth Disease Type 2N
AARS1 Patients 9-19 (AARS1 Family 3)	Dominant	+/R326W *AARS1*	Charcot-Marie-Tooth Disease Type 2N
AARS1 Patient 20	Recessive	C115R/R750W *AARS1*	Microcephaly, failure to thrive, seizures, and hypomyelination,
AARS2 Patient 1	Recessive	F131del/I328M *AARS2*	Dysarthria, balance impairment, and impaired hand dexterity
AARS2 Patient 2	Recessive	F131del/I328M *AARS2*	Vision problems, gait changes, dysarthria, and cognitive impairment
GARS1 Patients 1 and 2	Dominant	+/I334N *GARS1*	Infantile Spinal Muscular Atrophy
HARS1 Patient 1	Recessive	D206Y/V244Cfs*6 *HARS1*	Intellectual disability, ataxia, developmental delay, dysarthria, and dysmetria
HARS1 Patients 2 and 3	Recessive	L305dup/I465L *HARS1*	Ataxia, microcephaly, absent knee and ankle jerks, and distal muscle weakness
YARS1 Patients 1 and 2	Recessive	I59T/I59T *YARS1*	Cholestatic jaundice, developmental delay, and facial dimorphism

with the WHEP and MTS domains deleted robustly rescues growth of yeast lacking *GRS1*, the yeast ortholog of *GARS1*[86,232]. Assessing whether newly identified variants may affect these protein domains provides insight into the functional consequences of the variants.

To explore whether ARS variants affect function, we employ aminoacylation assays and yeast complementation assays, which are established methods for evaluating the effects of ARS variants[177]. *In vitro* aminoacylation assays involve: (1) incubating recombinant purified ARS enzyme, tRNA, ATP, and radiolabeled amino acid; (2) spotting the reaction on filter paper (3) precipitating the tRNA; and (4) measuring radioactivity levels and calculating steady-state kinetics to determine aminoacylation activity[177]. While these assays directly assess the tRNA charging capacity of mutant enzymes, they test one allele at a time, and the results may not reflect the function of enzymes in a living cell[177]. To test enzyme function *in vivo*, yeast complementation assays are used to test the ability of yeast cells to survive in the presence of the mutant ARS allele and the absence of the wild-type ARS gene[177]. Yeast cell growth is then used as a proxy for enzyme function[177]. There are limitations to this assay including that yeast are small, single-celled organisms, and again, one allele is tested at a time. Recently, an aminoacylation assay has been developed that uses lysates from patient cells as the source of ARS enzymes[72]. This importantly allows assessment of tRNA charging in the context of the patient genotype, as opposed to one allele at a time; however, results from this assay may not reflect ARS activity *in vivo* and from different cell types. Data from multiple functional studies can strengthen arguments regarding pathogenicity of variants identified in patients.

Uncovering the allelic and clinical heterogeneity of ARS-associated disease will expand our understanding of ARS variants in disease and increase our ability to uncover the mechanisms underlying ARS-associated tissue-predominant phenotypes. Importantly this work may also improve patient diagnosis by informing screening practices for ARS variants and improving appropriate filtering and interpretations of clinical sequencing data. A subset of data in this chapter was published in *Human Molecular Genetics* (Volume 27, Issue 23, pages 4036-4050, December 1, 2018, License Numbers 5085540453751 and 5085540583958)[38], *American Journal of Medical Genetics Part A* (Volume 182, Issue 5, pages 1167-1176, March 17, 2020, License Number 5085540697830)[95], *Cerebellum* (Volume 19, Issue 1, pages 154-160, February 2020, License Number 5085540802672)[52], and *Human Mutation* (Volume 41, Issue 7, pages 1232-1237, July 2020, License Number 5085540933311)[98]. Permission was requested for reproduction of data through the Copyright Clearance Center (see license numbers above). The author performed all the studies in this chapter with the following exceptions: clinical collaborators performed the clinical evaluations, phenotype classification, and mutation sequencing; the UM Advanced Genomics Core performed the Sanger sequencing reactions described; Ya-Ming Hou at Thomas Jefferson University performed *in vitro* aminoacylation assays on AARS1 protein variants (Figure 2.4); Daniele Galatolo and Filippo Santorelli at IRCCS Stella Maris Foundation performed *HARS1* quantitative real-time PCR and western blot analysis (Figure 2.12); Patrick Mullen and Christopher Francklyn at University of Vermont performed the HARS1 fibroblast lysate aminoacylation assays (Figure 2.14); the E337K *AARS1*, S627L *AARS1*, and K337E *ALA1* constructs were generated and tested by Stephanie Oprescu (Antonellis Laboratory, University of Michigan; Figure 2.3); the D206Y *HARS1* construct was generated and tested by Rebecca Meyer-Schuman (Antonellis Laboratory, University of Michigan; Figure 2.13); and the

V244Cfs*6 *HARS1* construct was generated and tested by Natasha Golovchenko (Antonellis Laboratory, University of Michigan; Figure 2.13).

2.2 Methods

2.2.1 Patient sample collection and identification of ARS variants

The patients with *AARS1* variants (referred to as AARS1 Patients 1-19; Table 2.1) and dominant axonal Charcot-Marie-Tooth (CMT) disease were examined by experienced neurologists, and all samples were obtained with patients' consent. The variants were identified using a CMT-specific custom-made capture array that was developed for 45 CMT-associated genes in addition to the exons of 24 genes of other neurologic disorders (NimbleGen, Roche). Sequencing libraries were prepared according to the manufacturer's protocol (Newbler v.2.3, NimblGen, Roche) and sequenced on a Titanium FLX pyrosequencer (Roche). All high-quality differences were confirmed using Sanger sequencing analysis.

Patients with the I334N *GARS1* variant (referred to as GARS1 Patients 1 and 2; Table 2.1) were identified at Phoenix Children's Hospital and Baylor Genetics. The I334N *GARS1* variant was identified through exome sequencing for GARS1 Patient 1 and through a hereditary neuropathy gene panel testing for GARS1 Patient 2.

Patients with *AARS2* variants (referred to as AARS2 Patients 1 and 2; Table 2.1) were identified by Dr. Vikram Shakkottai at the University of Michigan Ataxia Clinic. The University of Michigan Institutional Review Board determined that the description of these cases did not require IRB approval; we obtained authorization from the patients to publish the description of

their cases. The F131del and I328M *AARS2* variants in the proband (AARS2 Patient 1) were determined through exome sequencing (University of Chicago), and commercial genetic testing for *AARS2* showed the same variants in the proband's symptomatic brother (AARS2 Patient 2). Neurological examinations and sequencing were performed for both parents.

Patients with *HARS1* variants (referred to as HARS1 Patients 1-3; Table 2.1) were identified in Italy, and the pediatric ethics committee of the Tuscany Region, Italy approved the study. Parents provided written informed consent for cellular and DNA studies. The *HARS1* variants for HARS1 Patient 1 were determined through exome sequencing (TruSightOne, Illumina). For HARS1 Patients 2 and 3, the variants were identified through a targeted multigene resequencing panel analysis for genes related to ataxias (Nimblegen, Roche). Genomic DNA was isolated from patient fibroblasts using the Wizard Genomic DNA Purification Kit (Promega), and Sanger sequencing (UM Advanced Genomics Core) of PCR-amplified products (primer sequences in Appendix) confirmed all of the variants.

The patient with bi-allelic *AARS1* variants (referred to as AARS1 Patient 20; Table 2.1) was identified by Dr. Renée Perrier at the University of Calgary. Testing was performed as part of a large comprehensive intellectual disability panel (Prevention Genetics).

Patients with *YARS1* variants (referred to as YARS1 Patients 1 and 2; Table 2.1) were identified by Ronen Spiegel at the Emek Medical Center in Israel. The variants were determined through exome sequencing.

2.2.2 Population frequency and conservation

The prevalence of each variant in the general population was determined by searching gnomAD[180,228]. Conservation of each variant was examined by aligning ARS protein orthologs from multiple species with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The accession numbers for the AARS1 protein sequences used were: human (*Homo sapiens*, NP_001596.2), mouse (*Mus musculus*, NP_666329.2), zebrafish (*Danio rerio*, NP_001037775.1), fly (*Drosophila melanogaster*, AAF05593.1), worm (*Caenorhabditis elegans*, O01541.1), yeast (*Saccharomyces cerevisiae*, EDN63655.1), and bacteria (*Escherichia coli*, BAA16559.1). The accession numbers for the GARS1 protein sequences used were: human (*Homo sapiens*, NP_002038.2), mouse (*Mus musculus*, NP_851009.2), zebrafish (*Danio rerio*, NP_001373382.1), fly (*Drosophila melanogaster*, NP_001261867.1), worm (*Caenorhabditis elegans*, NP_498093.1), yeast (*Saccharomyces cerevisiae*, NP_009679.2), and bacteria (*Escherichia coli*, OMI63373.1). The accession numbers for the AARS2 protein sequences used were: human (*Homo sapiens*, NP_065796.2), chimpanzee (*Pan troglodytes*, PNI77549.1), horse (*Equus caballus*, XP_023480797.1), cow (*Box taurus*, NP_001178140.1), rat (*Rattus norvegicus*, NP_001100361.1), mouse (*Mus musculus*, NP_001344929.1), chicken (*Gallus gallus*, NP_001026227.1), zebrafish (*Danio rerio*, XP_021329187.1), and yeast (*Saccharomyces cerevisiae*, NP_014980.3). The accession numbers for the HARS1 protein sequences used were: human (*Homo sapiens*, P12081.2), mouse (*Mus musculus*, Q16035.2), zebrafish (*Danio rerio*, NP_001289185.1), fly (*Drosophila melanogaster*, AAN09471.2), worm (*Caenorhabditis elegans*, P34183.3), yeast (*Saccharomyces cerevisiae*, P07263.2), and bacteria (*Escherichia coli*, OSM92786.1). The accession numbers for the YARS1 protein sequences used were: human (*Homo sapiens*, NP_003671.1), mouse (*Mus musculus*, NP_598912.4), zebrafish (*Danio rerio*,

NP_958473.1), fly (*Drosophila melanogaster*, NP_648895.1), worm (*Caenorhabditis elegans*, NP_740947.2), yeast (*Saccharomyces cerevisiae*, NP_011701.3), and bacteria (*Escherichia coli*, BAA15398.2).

2.2.3 Quantitative real-time PCR

Daniele Galatolo and Filippo Santorelli (IRCCS Stella Maris Foundation) performed quantitative real-time PCR of *HARS1* mRNA from patient fibroblasts (HARS1 Patients 1-3). Patient RNA was obtained from cultured skin fibroblasts using High Pure RNA Isolation Kit (Roche). Total RNA (1 μg) was retrotranscribed using Transcriptor First Strand cDNA Synthesis Kit (Roche) following the manufacturer's instructions. qPCR was performed using qPCRBIO SyGreen Mix Hi-ROX (PCR Biosystems) and 7500 Fast Real-Time PCR Systems (Thermo Fisher Scientific). The ribosomal protein 18S was used as an internal control. Gene expression levels were normalized to the internal control, and the relative quantification was carried out using the ΔCt method. The results were expressed relative to mRNA expression.

2.2.4 Western blot analysis

Daniele Galatolo and Filippo Santorelli (IRCCS Stella Maris Foundation) performed western blot analysis of HARS1 protein levels from patient fibroblasts (HARS1 Patients 1-3). Total protein (30-50 μg) was loaded on Novex 8-16% Tris-Glycine Mini Gels, WedgeWell format (Thermo Fisher Scientific) in reducing conditions. Protein transfers to membrane were performed using the Trans-Blot Turbo RTA Mini PVDF Transfer Kit (Bio-Rad Laboratories) and the Trans-Blot Turbo Transfer System (Bio-Rad Laboratories). Membranes were blocked with TBS/0.1% Tween 20 (TTBS) containing 5% non-fat dry milk and then incubated overnight

at 4°C with primary antibodies in TTBS with 2.5% non-fat dry milk. Rabbit monoclonal anti-HARS antibody (Abcam; ab140640) was used diluted 1:500, and mouse monoclonal anti-GAPDH antibody (Abcam; ab8245) was used diluted 1:8000. Membranes were incubated with peroxidase-conjugated anti-mouse and anti-rabbit secondary antibodies (Jackson ImmunoResearch Laboratories Inc.) for 1 hour at room temperature in the same buffer used for the primary antibodies (2.5% non-fat dry milk in TTBS). Bands were imaged using ChemiDoc Imaging System (Bio-Rad Laboratories Inc). ImageJ (https://imagej.nih.gov/ij/) was used for densitometry analysis.

2.2.5 Yeast complementation assays

Yeast complementation assays were performed as previously described[27,35,37,86,168,233-235]. The *AARS1* variants were modeled in the human *AARS1* open-reading frame. Each *AARS2* variant was modeled in yeast *ALA1* and human *AARS1* (Table 2.2). The *GARS1* variant was modeled in the human *GARS1* open-reading frame lacking the mitochondrial targeting sequence (MTS) and WHEP domain (ΔMTSΔWHEP); *GARS1* with both the MTS and WHEP domain deleted allow robust yeast cell growth in complementation assays[86,232]. The *HARS1* variants were modeled in the human *HARS1* open-reading frame. The *YARS1* variants were modeled in the human *YARS1* open-reading frame. Mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit, a sequence-validated wild-type ARS pDONR221 construct, and mutation-specific primers (primer sequences in Appendix). Reactions were transformed into *E. coli*, and plasmid DNA was purified from individual colonies and Sanger sequenced (UM Advanced Genomics Core) to confirm the presence of the mutation and the absence of PCR-induced errors. Sequence-validated constructs for wild-type and each studied mutation were cloned into the

LEU2-bearing yeast expression constructs, pYY1[232] or pRS315 (for *ALA1*), using Gateway cloning technology (Invitrogen). The pYY1 construct is a high-copy-number yeast vector that contains a constitutive *ADH* promoter[232], and the pRS315 vector is a yeast centromere vector[236]. The resulting reactions were transformed into *E. coli*, purified, and digested with *Bsr*GI (New England BioLabs) to confirm proper recombination and the presence of an appropriately sized insert.

To assess the ability of wild-type and mutant ARS alleles to support cellular growth, previously validated haploid yeast strains with the endogenous *ALA1*, *GRS1*, *HTS1*, or *TYS1* locus deleted[27,35,37,86,233-235], and viability maintained via a pRS316 vector bearing wild-type *ALA1*, *GRS1*, *HTS1*, or *TYS1*, respectively, were transformed with the wild-type ARS, the mutant ARS, or an empty construct. Transformed yeast cells were selected for the presence of pYY1 or pRS315 by growth on solid media lacking leucine (pYY1 and pRS315 both harbor the *LEU2* gene) and uracil (pRS316 harbors the *URA3* gene). Colonies were grown to saturation in 2 ml liquid medium lacking leucine and uracil at 30°C and 275 rpm for 48 hours. A 1 ml aliquot from each 2 ml culture was spun down at 10,000 rpm and re-suspended in 50 ul UltraPure RNase/DNase-free water. Undiluted cultures and dilutions of 1:10 and 1:100 were spotted on 0.1% 5-FOA complete solid medium (Teknova) to select for cells that spontaneously lost the *URA3*-bearing maintenance vector[237]. *AARS2* variants modeled in the yeast gene, *ALA1*, were also spotted onto yeast plates containing glycerol (yeast extract-peptone-glycerol [YPG]: 10 g/L yeast extract, 20 ml/L glycerol, 10 g/L peptone, 20 g/L agar). Yeast viability was assessed by visual inspection after 5 days of incubation at 30°C (or 37°C where indicated). At least two colonies per transformation were evaluated.

2.2.6 Yeast growth curves

To generate growth curves for the *AARS1* variants, yeast colonies from strains transformed with E337K *AARS*, R750W in *cis* with E337K *AARS*, G913D in *cis* with E337K *AARS1*, or empty pYY1 were inoculated in liquid medium lacking leucine and uracil and incubated 2 days at 30°C 275 rpm until the cultured cells reached saturated levels. For the *YARS1* variants, yeast colonies from strains transformed with wild-type *YARS1*, I59Y *YARS1*, or empty pYY1 were used. Saturated cultures (10 μl) were spiked into 4 ml liquid medium lacking leucine and uracil, and the OD_{600} was measured in triplicate at 0, 6, 23, 30, 47, 53, 72, 97, and 120 hours for the *AARS1* variants and at 0, 24, 30, 48, 54, 72, 97, and 120 hours for the *YARS1* variants using a GloMax-Multi Detection System (Promega). The average of the measurements at each time point was graphed using GraphPad Prism, with error bars representing standard deviation. For the *AARS1* variants, colonies from two independent transformations using independently generated constructs were tested. Four independently generated I59T *YARS1* constructs were tested.

2.2.7 Aminoacylation assays

For AARS1 aminoacylation assays, wild-type and E337K *AARS1* were cloned into the pET-DEST42 gateway vector (Life Technologies) for expression of AARS1 with a C-terminal fusion to 6xHis in *E. coli*. Protein purification was performed with the Ni^{2+}-affinity resin according to the manufacturer's protocol (Novagen). Following the T7 transcription, purification, and annealing of human $tRNA^{Ala}$, steady-state aminoacylation assays were performed at 37°C as previously described[35,238]. The incorporation of ^{3}H-Ala into ^{3}H-Ala-$tRNA^{Ala}$ was monitored as acid precipitable counts on filter pads. Each reaction was initiated by addition of an AARS1

enzyme (1.67 nM) to a mixture of increasing concentrations of $tRNA^{Ala}$ (0.2-0.18 μM). The data were fit to the Michaelis-Menten equation, and an average of at least three replicates was calculated.

For HARS1 aminoacylation assays, whole cell extracts were obtained from skin fibroblasts from patients or controls. Aminoacylation activity was tested by incubating the whole cell extracts with (^{14}C) histidine, $tRNA^{His}$, and ATP using established methods[133]. HARS1 aminoacylation activity was determined by the slope of counts per minute over time and normalized to control fibroblasts. The mean ± SEM of five independent experiments was calculated. One-way ANOVA tests were used for statistical analysis.

2.3 Results

2.3.1 AARS1 *variants identified in patients with Charcot-Marie-Tooth disease type 2N*

Alanyl-tRNA synthetase 1 (*AARS1* [MIM 601065]) encodes the enzyme responsible for charging tRNA with alanine in the cytoplasm[2]. Variants in *AARS1* have been previously implicated in dominant peripheral neuropathies including Charcot-Marie Tooth disease type 2N[26,35] and distal hereditary motor neuropathy[39], as well as recessive encephalopathies[32,33] and liver disease[34]. We identified three families with CMT type 2 (referred to as AARS1 Families 1-3; Table 2.1; Figure 2.1). In AARS1 Family 1, a mother and daughter (referred to as AARS1 Patients 1 and 2) both presented in their 30s with distal weakness, and the mother also had wasting of the legs and arms and distal sensory disturbances. Electromyography (EMG) demonstrated a severe sensorimotor axonal neuropathy in both individuals. In AARS1 Family 2, six individuals (referred to as AARS1 Patients 3-8) presented in their first to third decade of life with atrophy and weakness of

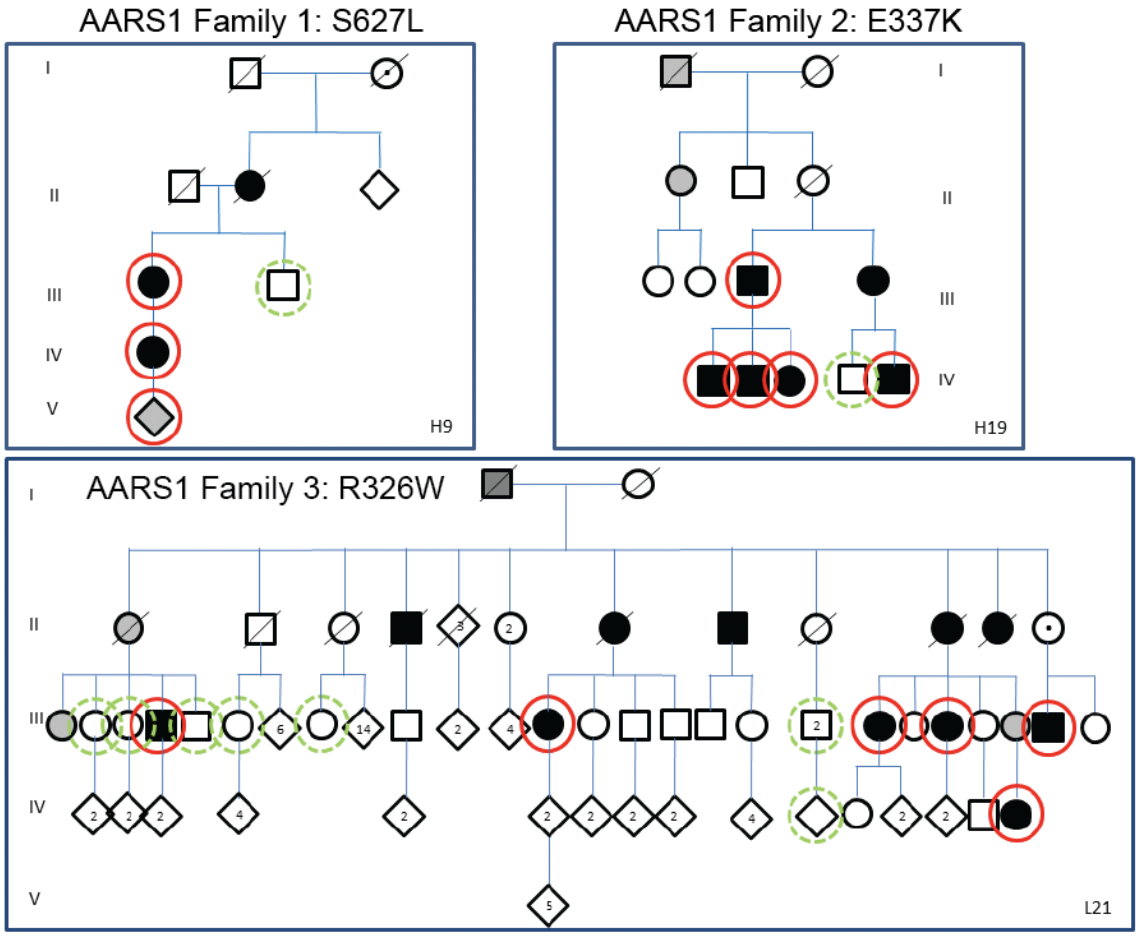


Figure 2.1. Segregation of *AARS1* variants in families with Charcot-Marie-Tooth disease type 2.

The variant identified in each family is listed above the corresponding pedigree. Circles indicate female individuals, and squares represent male individuals. Tested individuals are marked by circles; solid line indicates the individual tested positive for the respective *AARS1* mutation, and dashed line indicates the individual tested negative for the *AARS1* mutation. Filled symbols indicate affected patients. Grey symbols indicate individuals who were probably affected, were said to be affected, or had no detailed neurologic examination. Open symbols indicate individuals who were not affected. The family individuals are numbered consecutively from left to right.

the distal legs with some individuals also presenting with distal arm weakness. There were variable minor distal sensory disturbances. EMG demonstrated a sensorimotor polyneuropathy with axonal features in four of the six individuals. In one individual, demyelinating features were also observed. In Family 3, eleven affected individuals (referred to as AARS1 Patients 9-19) presented in their first to third decade of life with variable weakness of distal legs and no or minimal sensory disturbances. EMG was consistent with a severe sensorimotor axonal neuropathy.

A custom-made sequence capture array containing all exons of known CMT-associated genes followed by next-generation sequencing revealed an *AARS1* variant in each family: S627L (NM_001605:c.1880C>T; p.Ser627Leu) in AARS1 Family 1, E337K (c.1006G>A, p.Glu337Lys) in AARS1 Family 2, and R326W (c.976C>T; p.Arg326Trp) in AARS1 Family 3. The variants segregated with disease in a manner consistent with a dominant inheritance pattern (Figure 2.1). None of the variants is present in gnomAD[180,228]. The S627 residue is located in the editing domain (Figure 2.2A; see Chapter 1 for details about editing domains) and is conserved among all species analyzed (Figure 2.2B). The E337 residue is located in the tRNA recognition domain (Figure 2.2A) and is conserved among human, mouse, zebrafish, fly, and worm (Figure 2.2B). The R326 residue is located in the tRNA recognition domain (Figure 2.2A) and is conserved among all species analyzed (Figure 2.2B).

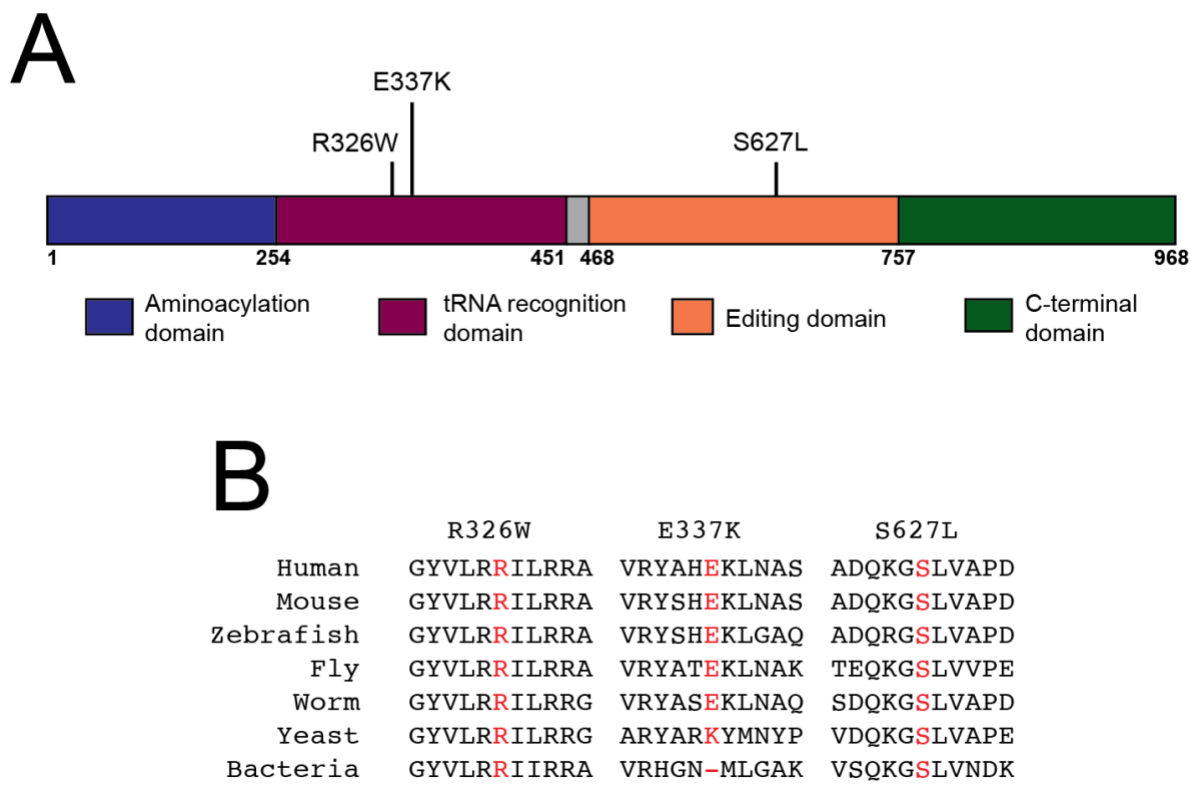


Figure 2.2. Localization and conservation of *AARS1* variants.

(A) AARS1 functional domains are indicated in blue (aminoacylation domain), purple (tRNA recognition domain), orange (editing domain), and green (C-terminal domain). The positions of the variants are shown across the top, and numbers along the bottom indicate amino-acid positions for the AARS1 protein. (B) Conservation of the affected amino-acid residues. The position of each variant is shown along with flanking AARS1 amino-acid residues from multiple, evolutionarily diverse species. The position of the affected residue is shown in red for each species.

2.3.2 R326W and S627L AARS1 *support reduced growth while E337K* AARS1 *supports increased growth in yeast complementation assays*

Yeast complementation assays have revealed loss-of-function characteristics of CMT-associated *AARS1* variants[35,37]. To test for functional consequences of R326W, E337K, and S627L *AARS1*, we compared the ability of wild-type and mutant human *AARS1* to complement loss of endogenous yeast *ALA1* (the yeast ortholog of *AARS1*). Complementation studies were performed in a previously validated haploid yeast strain deleted for endogenous *ALA1* and maintained with *ALA1* on a *URA3*-bearing vector[35,37]. Wild-type or mutant *AARS1* expression constructs or empty constructs (with no *AARS1* insert) were transformed into the above yeast strain and grown on 5-FOA, which selects for spontaneous loss of the *URA3*-bearing maintenance vector (Figure 2.3A, 2.3B)[237]. The empty plasmid did not support yeast growth at 30°C or 37°C, which indicates that *ALA1* is an essential gene. The wild-type *AARS1* expression construct supported some yeast growth at 30°C (Figure 2.3A) and more robust yeast growth at 37°C (Figure 2.3B), which indicates that human *AARS1* can complement the loss of the endogenous *ALA1* locus, consistent with previous findings[239]. The R326W *AARS1* variant was unable to support yeast growth at 30°C (Figure 2.3A), suggesting that R326W represents a loss-of-function allele. The S627L *AARS1* variant demonstrated reduced yeast viability compared to wild-type *AARS1* at 30°C (Figure 2.3A) and an even more dramatic reduction in viability at 37°C (Figure 2.3B), consistent with a hypomorphic allele. The E337K *AARS1* variant, however, showed increased yeast cell growth compared to wild-type *AARS1* at both 30°C and 37°C (Figure 2.3A, 2.3B). Interestingly, lysine is the wild-type amino acid at this residue in yeast. To determine the effect of changing lysine to glutamic acid in yeast *ALA1*, a complementation assay was performed with a K337E *ALA1* expression construct (Figure 2.3C, 2.3D). The K337E

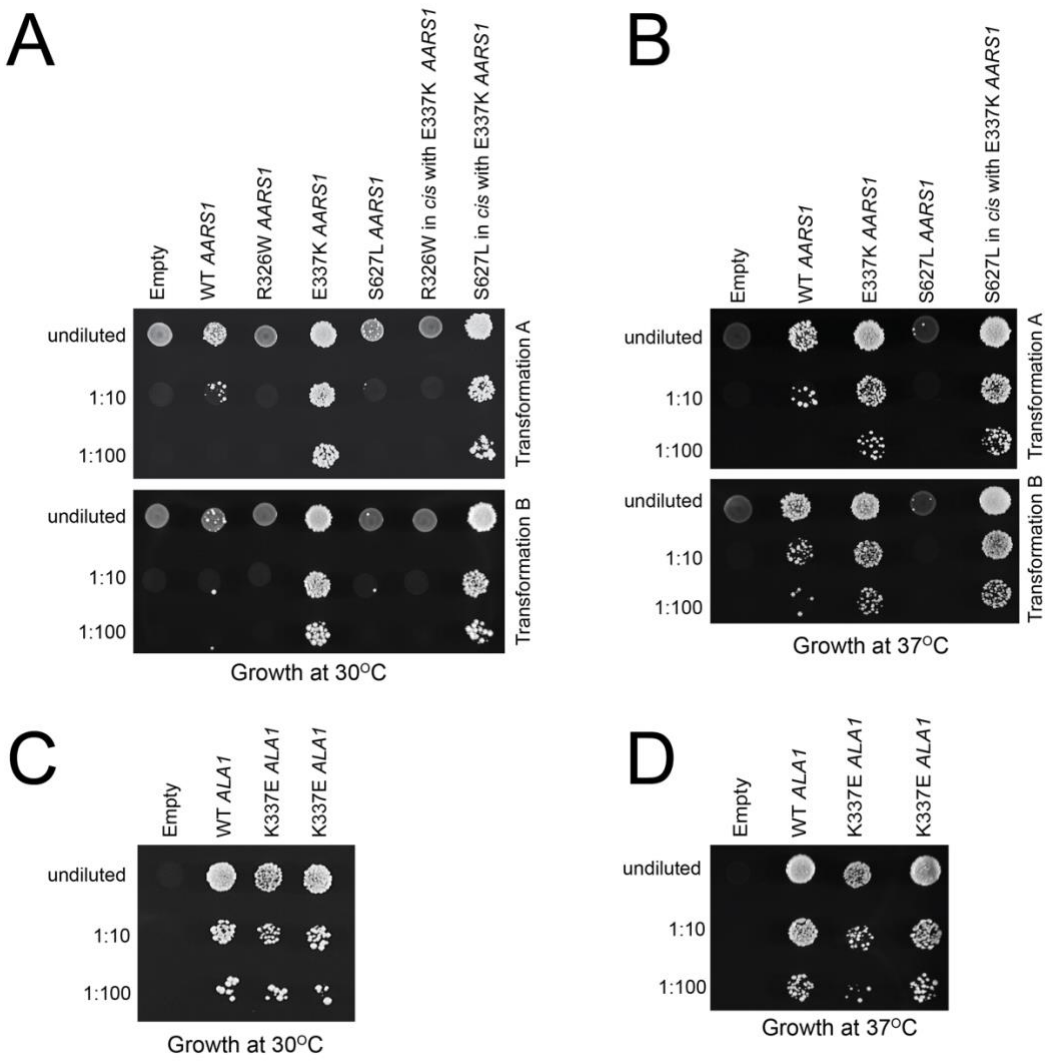


Figure 2.3. The R326W and S627L *AARS1* variants display loss-of-function effects in yeast complementation assays.

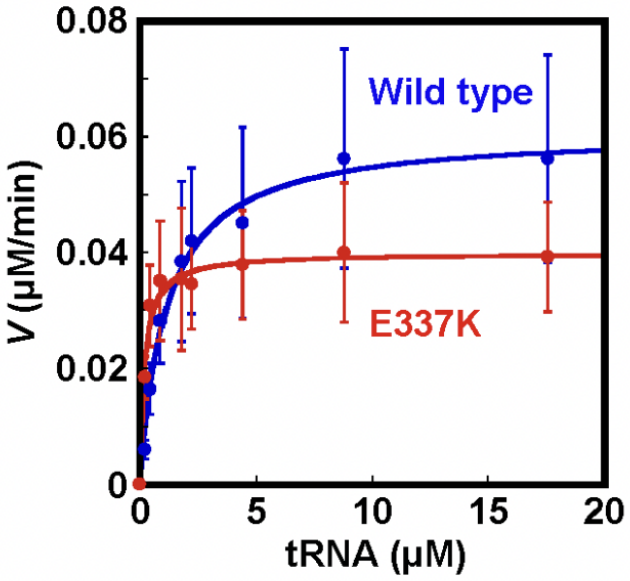
(A) Yeast lacking endogenous *ALA1* (the yeast ortholog of *AARS*) were transformed with vectors containing wild-type *AARS1*or mutant *AARS*, or a vector with no *AARS1*insert ('empty'). Two constructs for wild-type or mutant *AARS1*were transformed ('Transformation A' and 'Transformation B'). Resulting cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C. (B) Cultures resulting from the same transformations as in panel A were plated undiluted or diluted on media containing 5-FOA and grown at 37°C. (C) Vectors containing wild-type *ALA1* or K337E *ALA1*, or a vector with no *ALA1* insert were transformed into the same haploid yeast strain as in panel A. Resulting cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C. (D) Cultures resulting from the same transformations as in panel C were plated undiluted or diluted on media containing 5-FOA and grown at 37°C.

variant supported yeast cell growth slightly less than wild-type *ALA1* at 30°C and 37°C, suggesting that enzymes with a lysine at this residue function better in yeast. Therefore, the increased viability of yeast expressing E337K *AARS1* may be the result of improving the function of the human enzyme in yeast; further studies are needed to test the enzyme activity of E337K *AARS1* (see Chapter 2.3.3).

Since wild-type *AARS1* supported moderate, but not robust, yeast cell growth, the identification of a hypermorphic variant (E337K *AARS1*) in yeast provides a more robust assay to test the loss-of-function effects of *AARS1* variants. Therefore, R326W and S627L were tested in *cis* with E337K *AARS1*. R326W in *cis* with E337K was unable to support any yeast cell growth (Figure 2.3A), consistent with loss-of-function effects. In contrast, S627L in *cis* with E337K supported yeast growth similar to E337K only (Figure 2.3A, 2.3B), indicating that the reduced growth caused by S627L *AARS1* has the potential to be rescued. In sum, our yeast complementation assay data are consistent with R326W and S627L *AARS1* having complete or partial loss-of-function effects.

2.3.3 E337K AARS1 *demonstrates increased tRNA charging in* in vitro *aminoacylation assays*

To further explore the hypermorphic effect of E337K *AARS1* observed in yeast, we performed aminoacylation assays, which evaluate the kinetic properties of ARS variants[177]. We tested the ability of recombinant human AARS1 proteins (wild-type and E337K) to charge $tRNA^{Ala}$ with ^{3}H-alanine. In steady-state assays under multiple turnover conditions, the mutant enzyme exhibited a 5-fold decrease in K_m for tRNA relative to the wild-type enzyme, but a small decrease in the catalytic turnover number k_{cat} (Figure 2.4). These effects on kinetic parameters



	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} $μM^{-1}$)
Wild type	0.61 ± 0.02	1.13 ± 0.1	0.54 ± 0.2
E337K	0.40 ± 0.01	0.19 ± 0.03	2.09 ± 0.3

Figure 2.4. E337K AARS1 demonstrates increased tRNA charging.

Steady-state assays of aminoacylation of $tRNA^{Ala}$ with ^{3}H-alanine were performed at 37°C with each enzyme at 1.67 nM, and $tRNA^{Ala}$ at varying concentrations. Error bars show standard deviations derived from averaging at least three independent sets of experiments. These experiments were performed by Ya-Ming Hou at Thomas Jefferson University.

result in a nearly 4-fold increase in the catalytic efficiency of aminoacylation (k_{cat}/K_m; Figure 2.4). This indicates that E337K *AARS1* has increased enzyme activity consistent with a hypermorphic allele.

2.3.4 A recurrent de novo GARS1 *variant identified in patients with infantile SMA-like phenotype*

Glycyl-tRNA synthetase 1 (*GARS1* [MIM 600287]) encodes the enzyme that charges tRNA with glycine in both the mitochondria and the cytoplasm, and *GARS1* was the first ARS gene to be linked with disease[2,24]. Variants in *GARS1* have been implicated in dominant peripheral neuropathy phenotypes including Charcot-Marie Tooth disease Type 2D (CMT2D) and distal hereditary motor neuropathy (dHMN) type V[24,29], as well as recessive multi-system phenotypes[86]. Additionally, patients with a dominant infantile spinal muscular atrophy (iSMA)-like phenotype and *GARS1* variants have been reported[89,90]. We identified two unrelated patients (referred to as GARS1 Patients 1 and 2; Table 2.1) who presented with very early onset neuropathy and respiratory distress, which has a similar presentation to iSMA. GARS1 Patient 1 presented at 9 weeks of age with respiratory distress, decreased movements, and poor feeding. GARS1 Patient 2 presented at 6 weeks of age with repeated events of respiratory failure requiring multiple intubations, and at two months of age he required a tracheostomy and has remained on mechanical ventilation. An EMG/NCV was performed at 3 years of age and revealed a severe, active, diffuse, axonal, predominantly motor peripheral neuropathy.

Exome sequencing of GARS1 Patient 1 and hereditary neuropathy gene panel testing of GARS1 Patient 2 revealed that both patients were heterozygous for I334N (NM_002047.2:c.1001T>A;

p.Ile334Asn). In both patients the variant was a *de novo* event. Additionally, the variant is absent from gnomAD[180,228]. The I334 residue is located in the catalytic domain (Figure 2.5A) and is conserved among all species analyzed except bacteria (Figure 2.5B). A different missense substitution at this codon, I334F (also reported as I280F), has been identified in patients with features of later onset distal hereditary motor neuropathy[89,240], and functional studies showed that I334F reduced aminoacylation activity *in vitro*[235]. These data suggest that I334N may affect protein function.

2.3.5 The I334N GARS1 *variant does not support yeast cell growth in complementation assays*

Yeast complementation assays have been performed to reveal loss-of-function characteristics of CMT-associated *GARS1* variants[234,235]. In recent publications, full length *GARS1* allowed minimal cell growth in yeast, but *GARS1* with both the mitochondrial targeting sequence (MTS) and WHEP domain deleted (ΔMTSΔWHEP) allowed robust cell growth in yeast[86,232]. We therefore compared the ability of wild-type and I334N human *GARS1* ΔMTSΔWHEP to complement loss of endogenous yeast *GRS1* (the yeast ortholog of *GARS1*). Complementation studies were performed in a previously validated haploid yeast strain deleted for endogenous *GRS1* and maintained with *GRS1* on a *URA3*-bearing vector[241]. Wild-type or I334N *GARS1* ΔMTSΔWHEP expression constructs or empty constructs (with no *GARS1* insert) were transformed into the yeast strain and grown on 5-FOA, which selects for spontaneous loss of the *URA3*-bearing maintenance vector (Figure 2.6)[237]. The empty plasmid did not support yeast growth, consistent with *GRS1* being an essential gene (Figure 2.6). The wild-type *GARS1* ΔMTSΔWHEP expression construct supported robust yeast growth (Figure 2.6), which indicates that human *GARS1* can complement the loss of the endogenous *GRS1* locus, consistent with

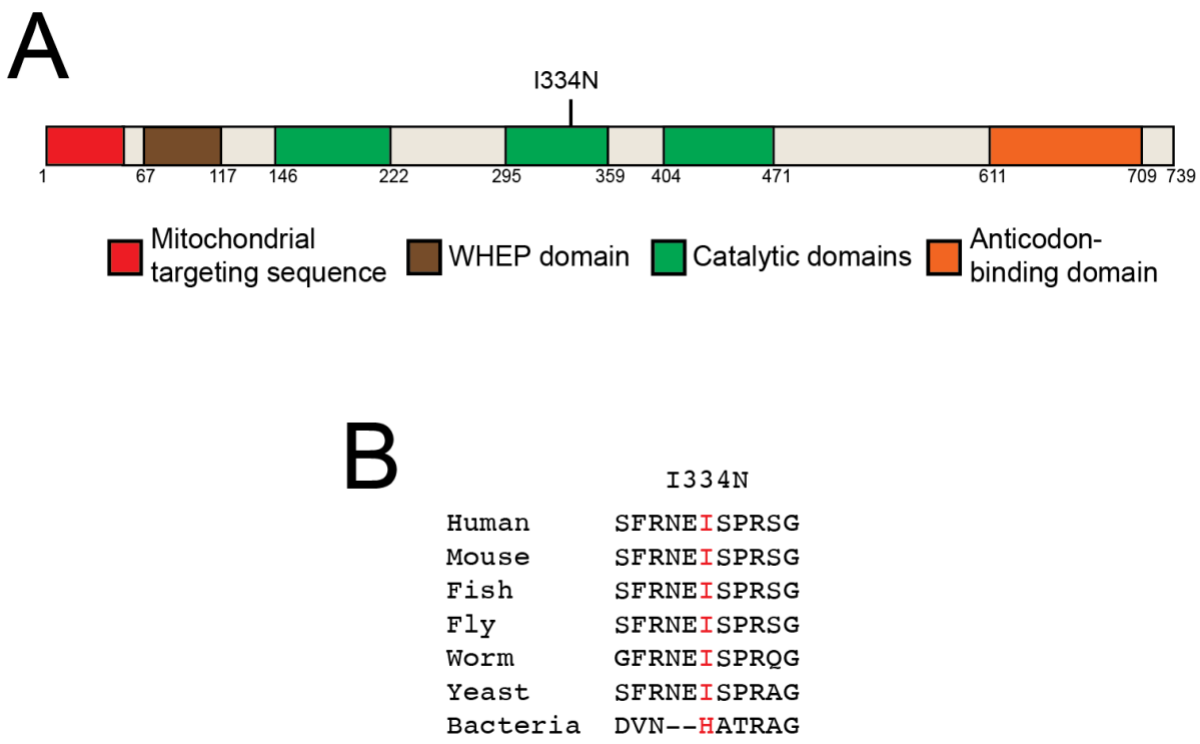


Figure 2.5. Localization and conservation of *GARS1* variants.

(A) GARS1 functional domains are indicated in red (mitochondrial targeting sequence), brown (WHEP domain), green (catalytic domains), and orange (anticodon-binding domain). The position of the variant is shown across the top, and numbers along the bottom indicate amino-acid positions for the GARS1 protein. (B) Conservation of the affected amino-acid residue. The position of the I334N variant is shown along with flanking GARS1 amino-acid residues from multiple, evolutionarily diverse species. The position of the affected residue is shown in red for each species.

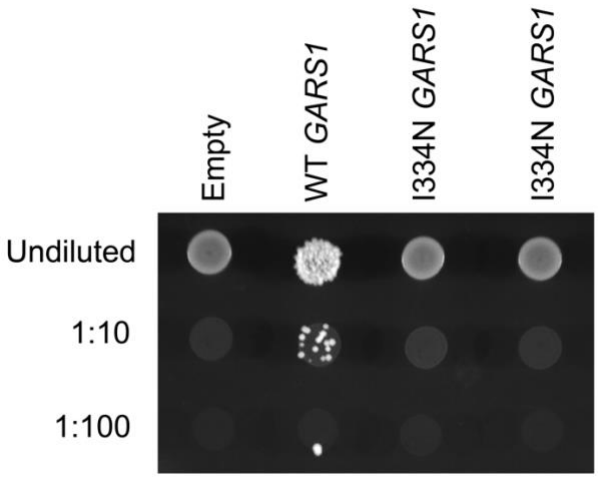


Figure 2.6. The I334N *GARS1* variant displays loss-of-function effects in yeast complementation assays.

Yeast lacking endogenous *GRS1* (the yeast ortholog of *GARS1*) were transformed with vectors containing wild-type *GARS1* or mutant *GARS1*, or a vector with no *GARS1* insert ('Empty'); the vector used in each experiment is indicated across the top. Resulting cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C.

previous findings[86,232]. The I334N *GARS1* variant was unable to support yeast growth (Figure 2.6), suggesting that I334N represents a loss-of-function allele.

2.3.6 Bi-allelic AARS2 *variants identified in patients with ataxia without leukoencephalopathy*

Mutations in mitochondrial alanyl-tRNA synthetase (*AARS2* [MIM 612035]) have previously been associated with infantile cardiomyopathy[242] or with later-onset leukoencephalopathy with ataxia, cerebellar atrophy, and ovarian failure[243]. Additionally, infantile primary pulmonary hypoplasia (without cardiomyopathy)[51], cardiomyopathy at age 15[48], and leukoencephalopathy with optic atrophy and retinopathy[50] were associated with *AARS2* variants. We identified two siblings (referred to as AARS2 Patients 1 and 2; Table 2.1) with ataxia but no leukoencephalopathy, a phenotype not previously associated with *AARS2* variants. A female patient (AARS2 Patient 1) presented at age 26 with dysarthria, balance impairment, and impaired hand dexterity. Her brother (AARS2 Patient 2) presented with vision problems and gait changes in childhood and later developed dysarthria and cognitive impairment. Brain MRI studies revealed cerebellar atrophy and no abnormal white matter signal in either patient.

Exome sequencing revealed AARS2 Patient 1 was heterozygous for F131del (NM_020745.4:c.390-392del; p.Phe131del) and I328M (c.984C>G; p.Ile328Met) *AARS2*, and commercial genetic testing for *AARS2* showed the same variants in AARS2 Patient 2. Both parents had a normal neurological examination, and DNA sequencing revealed that the two variants were on separate alleles (Figure 2.7A). The F131del *AARS2* variant was paternally inherited and present in gnomAD[180,228] at a low frequency (1 in 251,492 alleles). The maternally inherited variant (I328M) was not present in gnomAD. Both F131del and I328M are located in

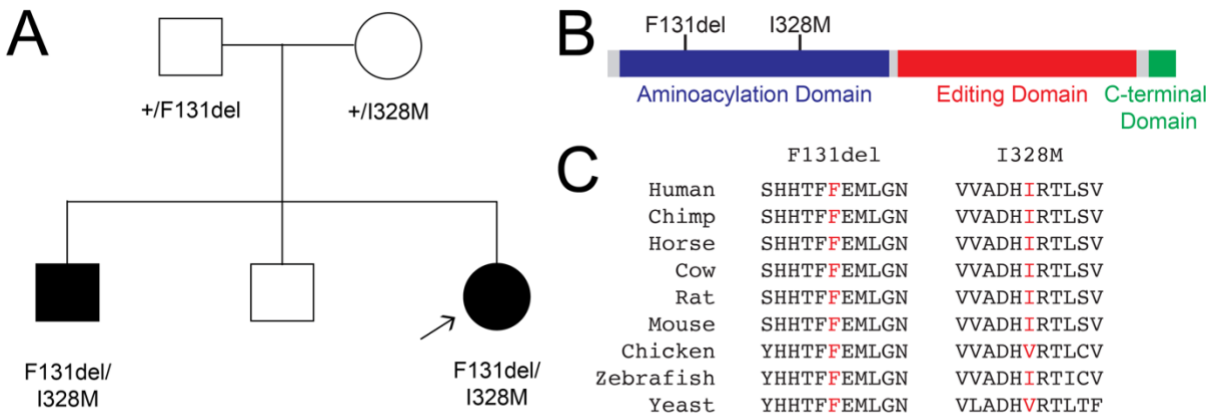


Figure 2.7. Segregation, localization, and conservation of *AARS2* variants.

(A) The simplex pedigree for the affected family. Squares represent males and circles represent females. Genotypes are indicated under each symbol, and filled shapes indicate affected individuals. (B) AARS2 functional domains are indicated in blue (aminoacylation domain), red (editing domain), and green (C-terminal domain). The positions of the variants are shown across the top, and numbers along the bottom indicate amino-acid positions for the AARS2 protein. (C) Conservation of the affected amino-acid residues. The position of each variant is shown along with flanking AARS2 amino-acid residues from multiple, evolutionarily diverse species. The position of the affected residue is shown in red for each species.

the aminoacylation domain of AARS2 (Figure 2.7B). The F131 residue is conserved among all analyzed species, and the I328 residue is conserved among all species analyzed except chicken and yeast (Figure 2.7C). These data suggest that the variants may affect protein function.

2.3.7 F131del and I328M AARS2 *variants exhibit loss-of-function effects in yeast complementation assays*

To assess the functional consequences of F131del and I328M *AARS2*, we modeled each variant in yeast *ALA1*, which uses alternative translation initiation sites to encode yeast enzymes that charge tRNA in both the cytoplasm and mitochondria[244]. We then compared the ability of mutant and wild-type *ALA1* to complement the loss of *ALA1* in yeast complementation assays. Briefly, we used a previously validated haploid yeast strain deleted for the endogenous *ALA1* locus, with viability maintained due to the presence of *ALA1* on a *URA3*-bearing vector[35,37,38]. Wild-type, F102del (corresponding to human F131del *AARS2*; Table 2.2) or V306M (corresponding to human I328M *AARS2*; Table 2.2) *ALA1* or a plasmid with no *ALA1* insert ('Empty' in Figure 2.8A) was transformed into yeast, and growth was evaluated on 5-FOA medium, which selects for cells that have spontaneously lost the maintenance vector[237]. The plasmid with no *ALA1* insert was unable to support yeast growth (Figure 2.8A), consistent with *ALA1* being an essential gene. The F102del *ALA1* variant was unable to support yeast growth (Figure 2.8A), indicating that F102del *ALA1* represents a loss-of-function allele. The V306M *ALA1* variant showed similar yeast cell growth compared to wild-type *ALA1* (Figure 2.8A) indicating that this is not a loss-of-function allele in this assay.

Table 2.2. AARS2 variants modeled in yeast ALA1 and human AARS1

Human *AARS2* Nucleotide Change[a]	Human AARS2 Amino Acid Change[b]	Yeast ALA1 Amino Acid Change[c]	Human AARS1 Amino Acid Change[d]
c.390-392del	F131del	F102del	F98del
c.984C>G	I328M	V306M	A302M

[a] Human *AARS2* nucleotide positions correspond to GenBank Accession number NM_020745.4
[b] Human AARS2 amino acid positions correspond to GenBank Accession number NP_065796.2
[c] Yeast ALA1 amino acid positions correspond to GenBank Accession number EDN63655.1
[d] Human AARS1 amino acid positions correspond to GenBank Accession number NP_001596.2

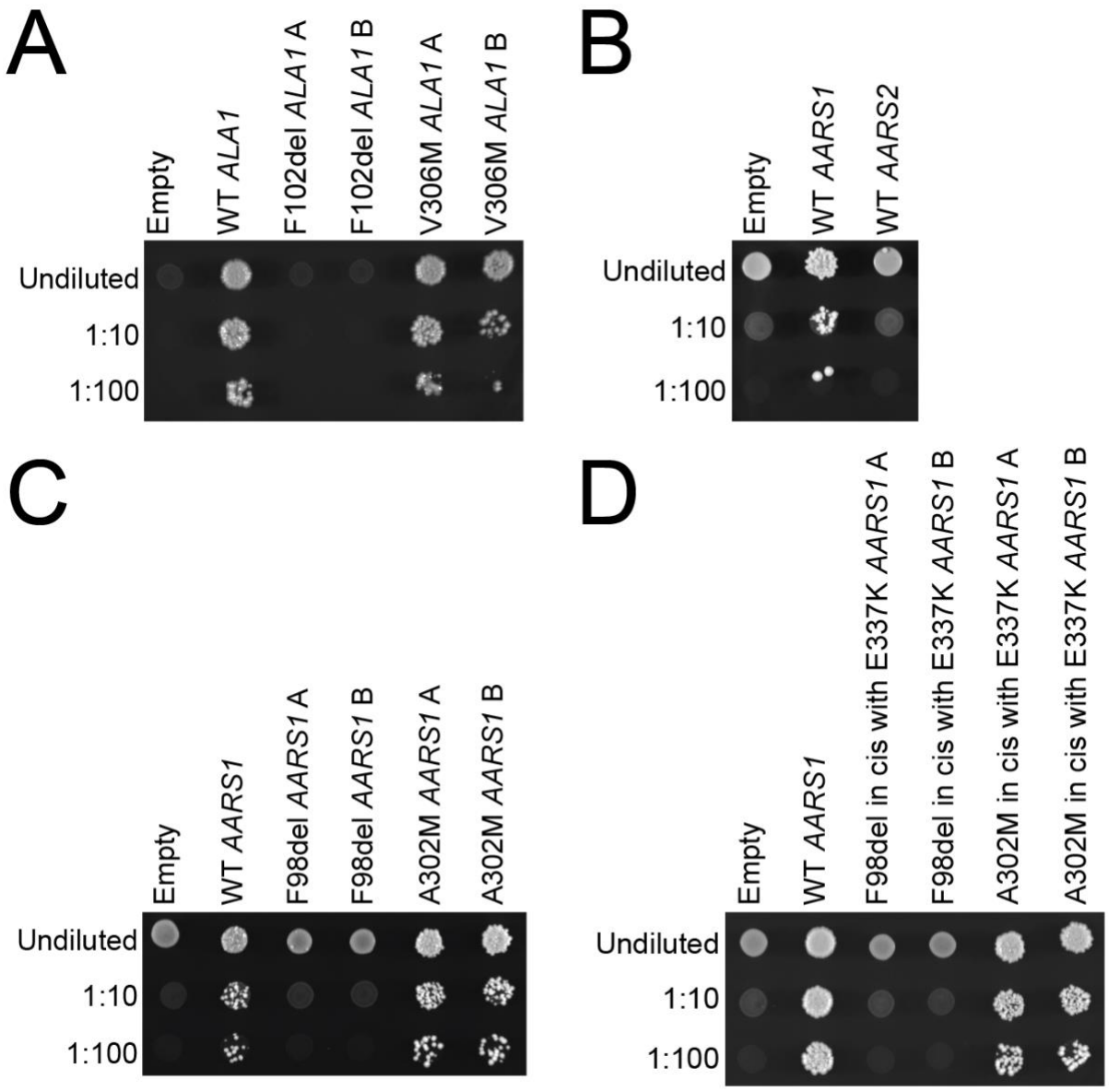


Figure 2.8. AARS2 variants display loss-of-function effects in yeast complementation assays.

Yeast lacking endogenous *ALA1* were transformed with a vector with no *ALA1* or *AARS1* insert ('Empty'), or with vectors containing wild-type *ALA1* or mutant *ALA1* (A), wild-type *AARS1*or *AARS2* (B), wild-type *AARS1*or mutant *AARS1*(C), or E337K *AARS1*or mutations in *cis* with E337K *AARS1*(D). The vector used in each experiment is indicated across the top. Cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C. Two independently generated mutant constructs ('A' and 'B') were tested.

In the experiments described so far, yeast growth was evaluated on medium containing glucose, a fermentable carbon source that yeast metabolize by glycolysis. Glycerol, however, is a non-fermentable carbon source that requires mitochondrial respiration[245]. Since the patients have variants in *AARS2*, which encodes a mitochondrial enzyme, we also evaluated yeast growth on medium containing glycerol to test mitochondrial function. Similar results as on glucose medium were observed for F102del and V306M *ALA1*; F102del did not support yeast growth, and V306M supported similar yeast growth compared to wild-type *ALA1* (Figure 2.9), indicating that F102del *ALA1* is a loss-of-function allele and V306M *ALA1* is not a loss-of-function allele in this assay.

Multiple human ARS genes can complement loss of the endogenous yeast gene, and the effects of disease-associated variants on yeast growth can be evaluated in the context of the human gene. To determine whether we could assess variants in the human *AARS2* gene in yeast, we next tested whether wild-type *AARS2* rescues loss of yeast *ALA1*. Unfortunately, wild-type *AARS2* does not complement the loss of *ALA1* (Figure 2.8B). However, wild-type human *AARS1*, which encodes the cytoplasmic enzyme, has been shown to complement the loss of *ALA1*[38], so we tested the ability of each *AARS2* variant modeled in *AARS1* (Table 2.2) to complement the loss of *ALA1*. Wild-type, F98del (corresponding to F131del *AARS2*), or A302M (corresponding to I328M *AARS2*) *AARS1* or a plasmid with no *AARS1* insert were transformed into yeast, and growth was evaluated on 5-FOA medium as above. The F98del *AARS1* variant was unable to support yeast growth, consistent with a loss-of-function allele (Figure 2.8C). The A302M *AARS1* variant showed similar yeast cell growth compared to wild-type *AARS1* (Figure 2.8C). As described above (see Chapter 2.3.2), wild-type *AARS1* supports moderate yeast cell growth;

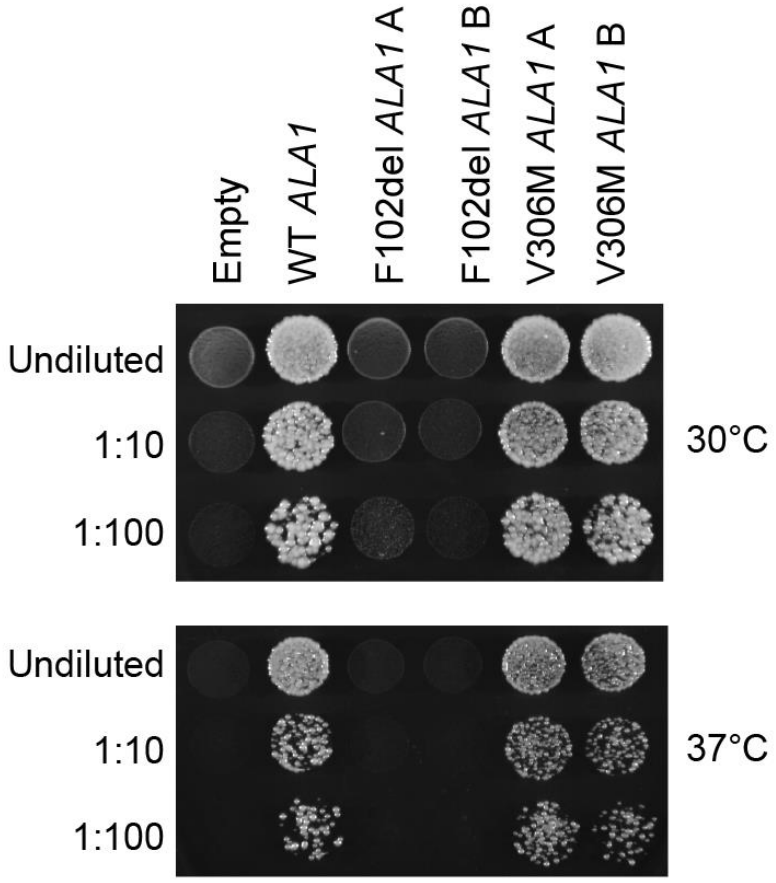


Figure 2.9. *AARS2* variants modeled in *ALA1* on glycerol.

Yeast lacking endogenous *ALA1* were transformed with a vector with no *ALA1* or *AARS1*insert ('Empty'), or with vectors containing wild-type *ALA1* or mutant *ALA1*. The vector used in each experiment is indicated across the top. Cultures were plated undiluted or diluted (1:10 or 1:100) on media containing glycerol and 5-FOA and grown at 30°C or 37°C. Two independently generated mutant constructs ('A' and 'B') were tested.

however, E337K *AARS1* supports increased cell growth compared to wild-type AARS1[38]. Therefore, to improve the sensitivity of our yeast complementation experiments by increasing our ability to detect subtle differences in growth, each variant was also tested in *cis* with E337K *AARS1*. The F98del *AARS1* variant in *cis* with E337K was unable to support yeast cell growth (Figure 2.8D). In contrast, the A302M in *cis* with E337K *AARS1* variant supported slightly reduced yeast growth compared to E337K only (Figure 2.8D) suggesting that this is a hypomorphic allele. In summary, our yeast complementation assays indicate that F131del *AARS2* dramatically impairs gene function, consistent with pathogenicity in the recessive phenotype described here. While our data suggest that I328M *AARS2* is a hypomorphic allele, additional studies will be required to verify this finding.

2.3.8 HARS1 *variants identified in patients with a multi-system ataxic syndrome*

Histidyl-tRNA synthetase 1 (*HARS1* [MIM 142810]) variants have previously been associated with autosomal recessive Usher syndrome type III[97] and with autosomal dominant peripheral neuropathy[27,99]. We identified three individuals (referred to as HARS1 Patients 1-3; Table 2.1) from two families who presented with a multi-system ataxic syndrome. HARS1 Patient 1 presented at 3 years of age with mild intellectual disability, gait ataxia with leg spasticity, developmental delay, dysarthria, and dysmetria. HARS1 Patients 2 and 3 are siblings from a separate family who presented with neurodevelopmental disorders including ataxic gait, microcephaly, absent knee and ankle jerks, and distal muscle weakness. HARS1 Patient 3 also had conductive hearing deficits. Nerve conduction studies in HARS1 Patient 2 revealed slowed motor velocity and increased latency.

Exome sequencing revealed that HARS1 Patient 1 is compound heterozygous for V244Cfs*6 (NM_002109.6:c.730delG; p.Val244Cysfs*6) and D206Y (c.616G>T; p.Asp206Tyr). A targeted multigene resequencing panel analysis for genes related to inherited ataxias revealed HARS1 Patients 2 and 3 are both compound heterozygous for I465L (c.1393A>C; p.Ile465Leu) and L305dup (c.910_912dupTTG; p.Leu305dup). Sanger sequencing of genomic DNA from patient fibroblasts confirmed the genotypes of each individual (Figure 2.10). HARS1 Patient 1 inherited V244Cfs*6 from the father and D206Y from the mother. HARS1 Patients 2 and 3 inherited L305dup from their mother (Figure 2.10), and we inferred that I465L was inherited from the father since we could not genotype paternal DNA. The V244Cfs*6, D206Y, and L305dup variants are absent in gnomAD[228], and the I465L variant is present in gnomAD at low frequency (2/251,424; no homozygotes). The D206 residue is located in the catalytic domain (Figure 2.11A) and is conserved among human, zebrafish, fly, and yeast (Figure 2.11B). The L305 residue is located in the catalytic domain (Figure 2.11A) and is conserved among all species analyzed (Figure 2.11B). The I465 residue is located in the anticodon-binding domain (Figure 2.11A) and is conserved among human, mouse, zebrafish, and yeast (Figure 2.11B). The V244Cfs*6 variant introduces a stop codon in the catalytic domain (Figure 2.11A) and is predicted to ablate 260 amino acids including the entire anticodon-binding domain. These data suggest that the variants may affect protein function.

2.3.9 Patient fibroblasts have reduced HARS1 *mRNA and protein levels*

To determine the effect of each genotype on gene and protein expression, RNA and protein samples were obtained from primary fibroblasts from each patient, and *HARS1* mRNA and protein levels were assessed. Real-time quantitative PCR (qPCR) indicated a significant

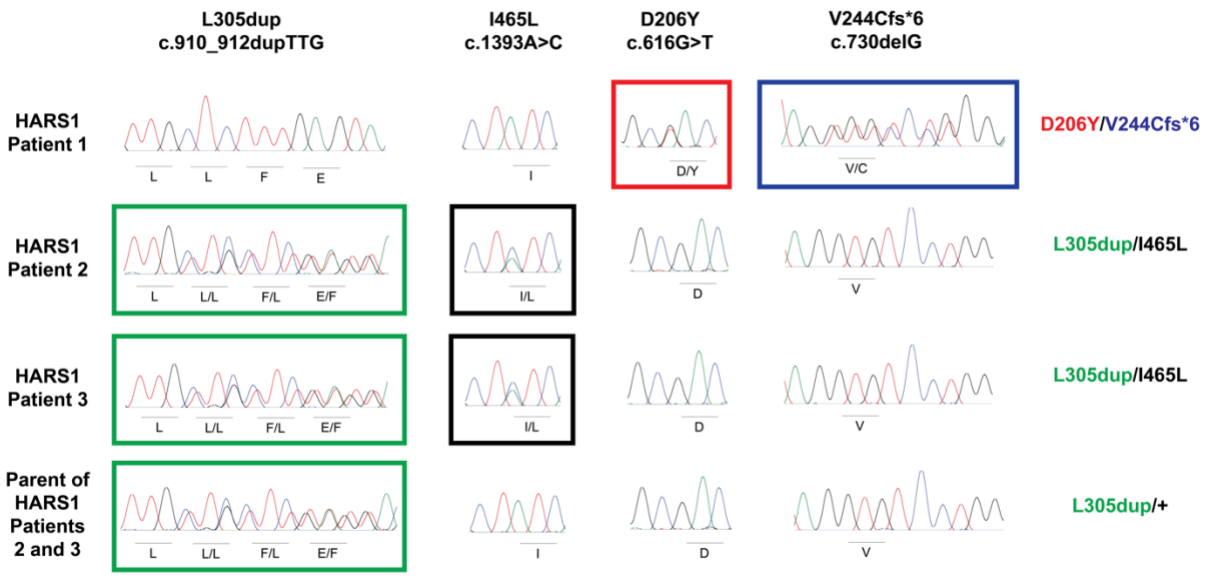


Figure 2.10. Sanger sequencing chromatograms from patients with *HARS1* variants.

Representative sequence chromatograms are shown for the individuals indicated on the left. The protein and cDNA annotations of the identified variants are above, and the effect on amino-acid sequence is indicated below. Regions surrounding each mutation were amplified and sequenced in all four individuals. Boxes indicate a variant was identified. The genotype of each individual is on the right.

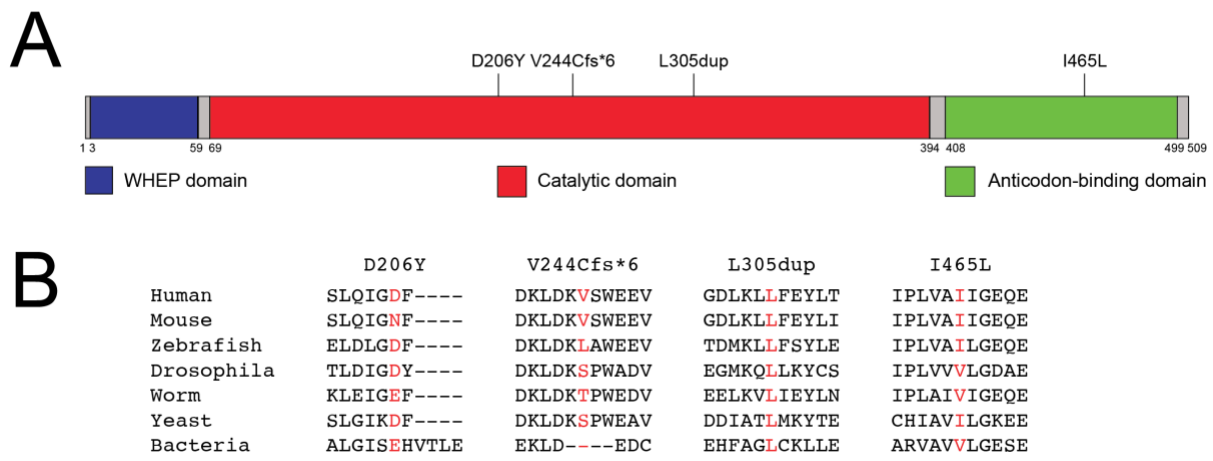


Figure 2.11. Localization and conservation of *HARS1* variants.

(A) HARS1 functional domains are indicated in blue (WHEP domain), red (catalytic domain), orange (editing domain), and green (anticodon-binding domain). The positions of the variants are shown across the top, and numbers along the bottom indicate amino-acid positions for the HARS1 protein. (B) Conservation of the affected amino-acid residues. The position of each variant is shown along with flanking HARS1 amino-acid residues from multiple, evolutionarily diverse species. The position of the affected residue is shown in red for each species.

reduction in *HARS1* mRNA in all three patients compared to controls (Figure 2.12A), suggesting *HARS1* mRNA may be unstable or subjected to nonsense-mediated decay in patient fibroblasts. Consistent with the reduced *HARS1* mRNA levels, western blot analyses also indicated a significant reduction of HARS1 protein levels in all three patient fibroblasts compared to controls (Figure 2.12B). These results indicate that the patient fibroblasts have reduced, but not ablated, HARS1 expression.

2.3.10 D206Y and I465L HARS1 *reduce growth in yeast complementation assays*

To test for functional consequences of D206Y, V244Cfs*6, L305dup, and I465L *HARS1 in vivo*, yeast complementation assays were employed using established methods[27,233]. Complementation studies were performed using a previously validated haploid yeast strain with the endogenous *HARS1* ortholog (*HTS1*) deleted and viability maintained via a maintenance vector harboring *HTS1* and *URA3*[27,233]. Each variant was modeled in human *HARS1* and then transformed into the yeast strain described above (Figure 2.13). Growth was evaluated on 5-FOA, which is toxic to strains expressing URA3 and selects for spontaneous loss of the maintenance vector[246]. These analyses revealed that: (1) wild-type human *HARS1* supports yeast cell growth, while a vector with no *HARS1* insert does not (Figure 2.13); (2) D206Y and I465L *HARS1* allow growth consistent with wild-type *HARS1* in this assay (Figure 2.13); and (3) V244Cfs*6 and L305dup do not support any yeast growth (Figure 2.13), consistent with these alleles having a loss-of-function effect.

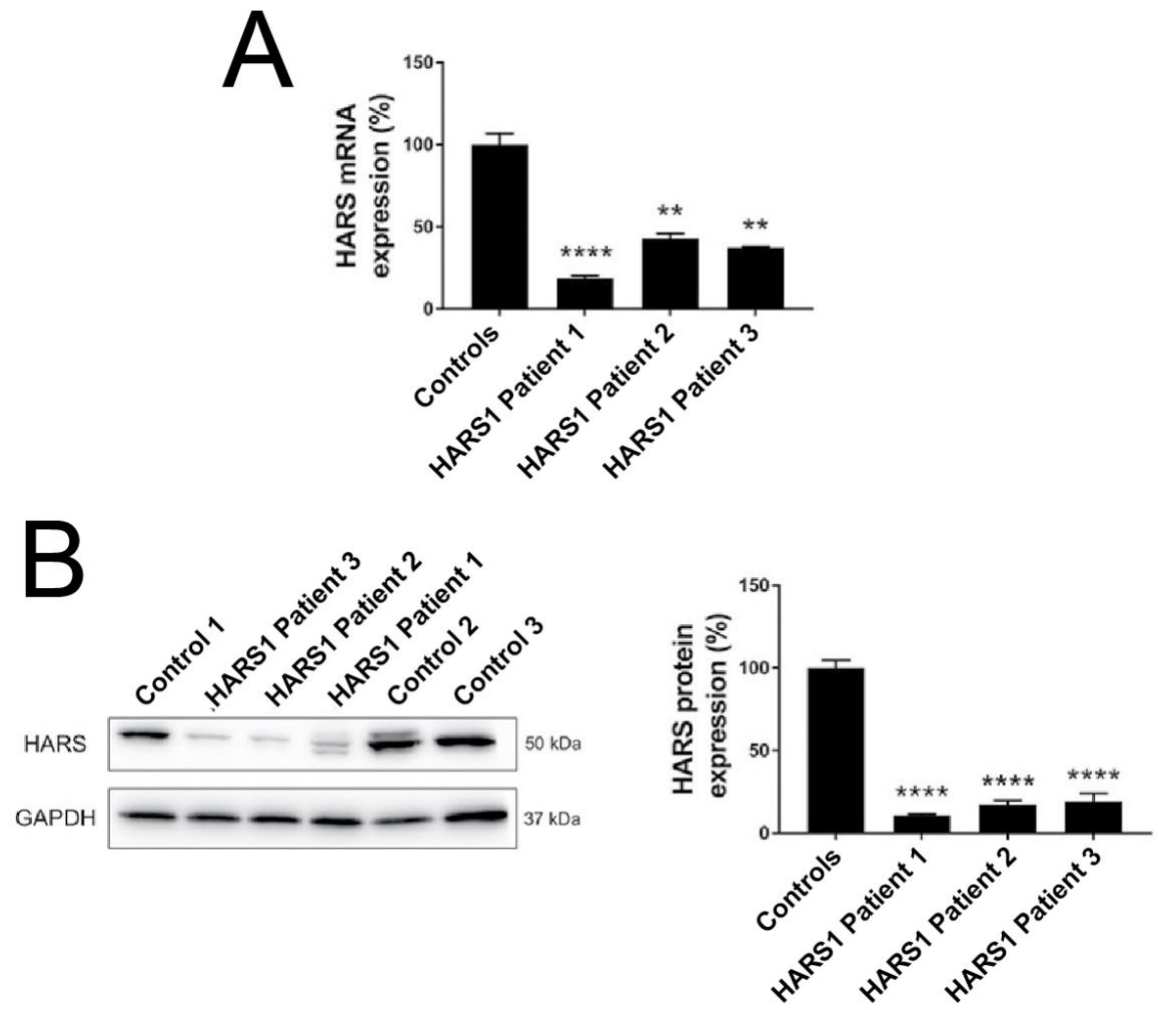


Figure 2.12. *HARS1* mRNA and protein levels are reduced in patient fibroblasts.

(A) Real-time quantitative PCR showed a significant reduction of *HARS1* mRNA level in all patients. Data are shown as mean ± SEM and were normalized to control fibroblasts. Controls n = 22; n = 3 in each patient. One-way ANOVA was used for statistical analysis. **p < .01, ****p < .0001. (B) Representative western blot (left) used for HARS1 protein detection. Quantification (right) showed a severe reduction of protein expression in all patients. Data are shown as mean ± SEM of three independent experiments (n = 9) and were normalized to control fibroblasts. One-way ANOVA was used for statistical analysis. ****p < .0001. These experiments were performed by Daniele Galatolo and Filippo Santorelli at IRCCS Stella Maris Foundation.

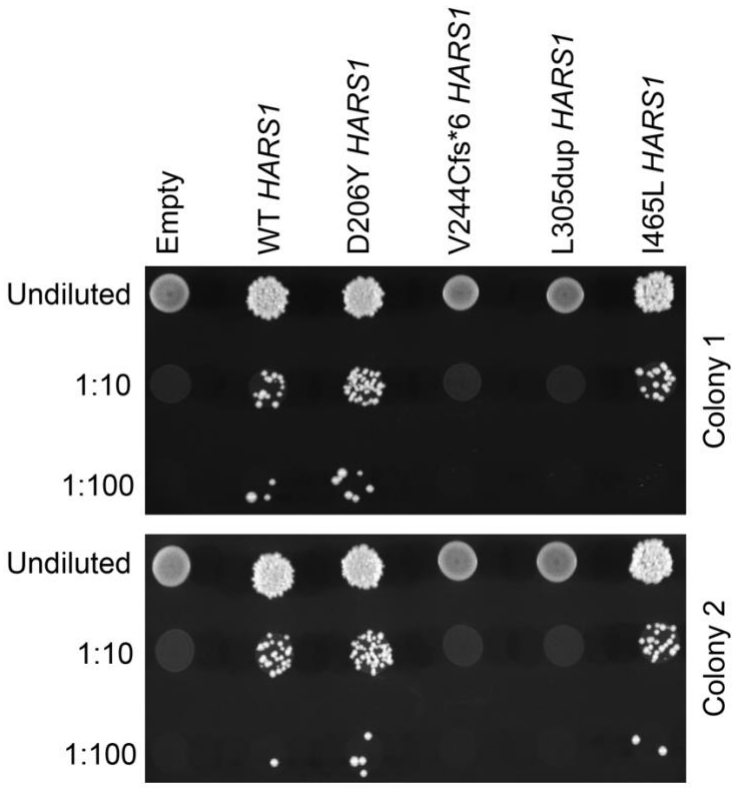


Figure 2.13. *HARS1* variants display loss-of-function effects in yeast complementation assays.

Yeast lacking endogenous *HTS1* were transformed with a vector with no *HARS1* insert ('Empty'), or with vectors containing wild-type or variant *HARS1*. The vector used in each experiment is indicated across the top. Cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C. Two independently transformed colonies were tested (upper and lower panels).

2.3.11 HARS1 enzymes from patient fibroblasts exhibit decreased aminoacylation activity

To investigate the effects of the *HARS1* variants in the context of the affected patient genotypes, we performed aminoacylation reactions using whole-cell extracts from patient and control fibroblasts as the sources of ARS enzymes (see Chapter 2.1 for further discussion of this assay). We tested aminoacylation activity by incubating the whole-cell extracts with (^{14}C) histidine, $tRNA^{His}$, and ATP. tRNA charging activity was significantly decreased using cell extracts from all of the patients compared to controls (Figure 2.14). These data are consistent with reduced *HARS1* function in each patient.

2.3.12 AARS1 *variants identified in a patient with microcephaly, hypomyelination, seizures, and failure to thrive*

In addition to dominant peripheral neuropathy as described above, variants in alanyl-tRNA synthetase 1 (*AARS1*) have been associated with recessive disease phenotypes including epileptic encephalopathy with myelination defects[32,33] and liver failure[34]. We identified a three-year-old girl (referred to as AARS1 Patient 20; Table 2.1) with postnatal onset microcephaly, failure to thrive, seizures, hypomyelination, thin corpus callosum, and developmental delay. Using a comprehensive intellectual disability panel, we determined that the patient was compound heterozygous for two variants in *AARS1* (NM_001605.2): C115R (c.343T>C; p.Cys115Arg) and R750W (c.2248C>T; p.Arg750Trp). The C115R variant is located in the aminoacylation domain (Figure 2.15A) and is also present in gnomAD[228] at low frequency (3/251,274; no homozygotes). The R750W variant is located in the editing domain (Figure 2.15A) and is present in gnomAD at low frequency (3/251,392; no homozygotes). The C115 residue is conserved among human,

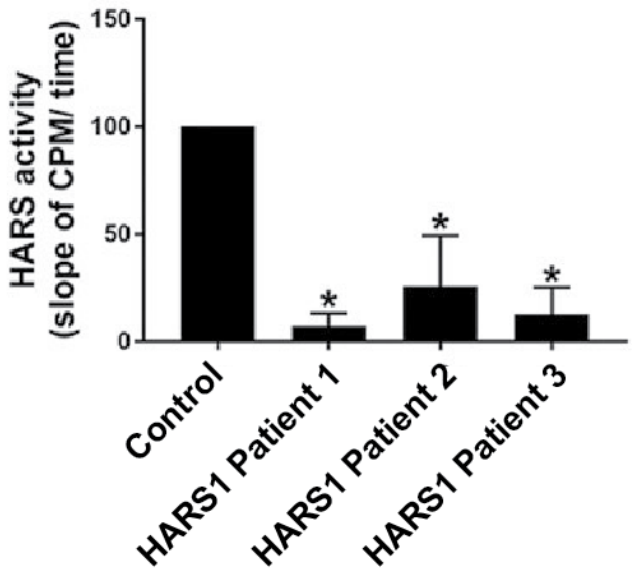


Figure 2.14. HARS1 enzymes from patient fibroblasts exhibit decreased aminoacylation activity.

HARS1 aminoacylation activity was determined by incubating whole cell extracts with (^{14}C) histidine, $tRNA^{His}$, and ATP and depicted as the slope of counts per minute over time. Data shown as mean ± SEM of five independent experiments and were normalized to control fibroblasts. One-way ANOVA was used for statistical analysis. *$p < .02$. These experiments were performed by Patrick Mullen and Christopher Francklyn at University of Vermont.

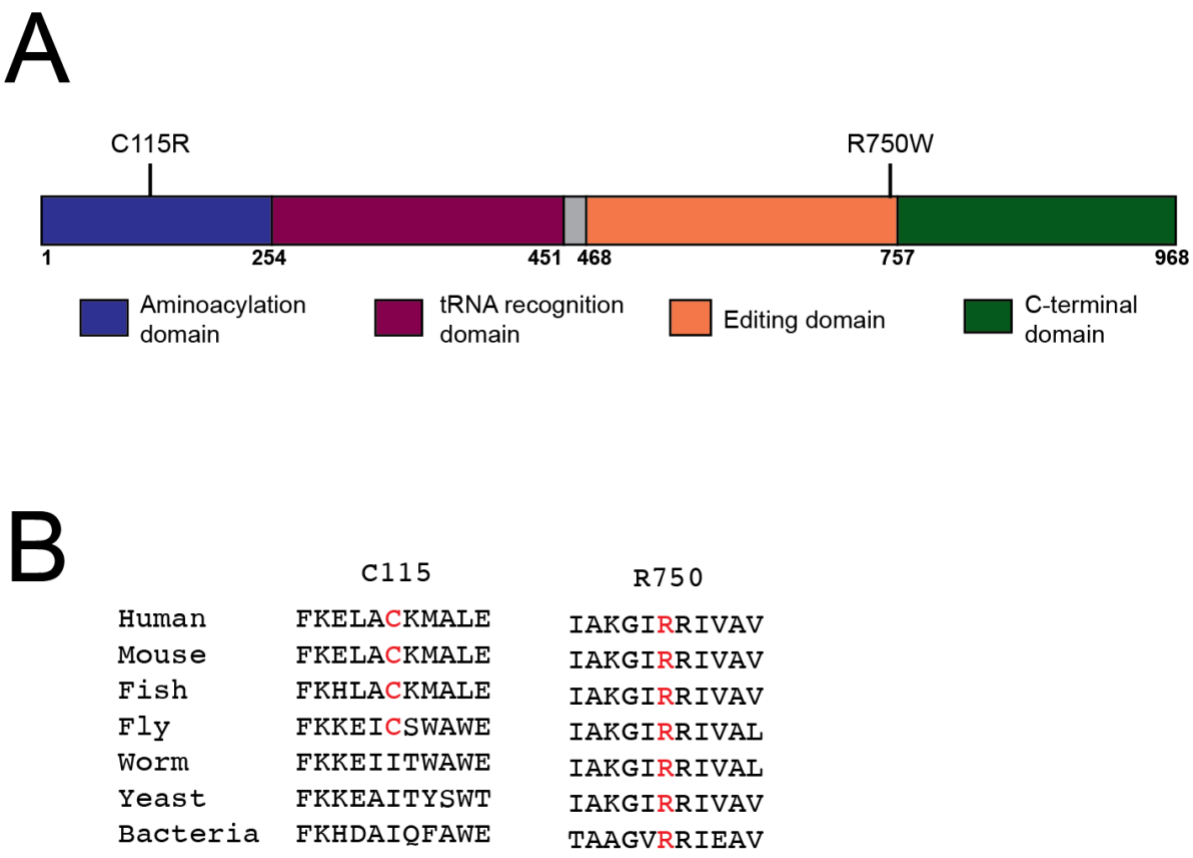


Figure 2.15. Localization and conservation of *AARS1* variants.

(A) AARS1 functional domains are indicated in blue (aminoacylation domain), purple (tRNA recognition domain), orange (editing domain), and green (C-terminal domain). The positions of the variants are shown across the top, and numbers along the bottom indicate amino-acid positions for the AARS1 protein. (B) Conservation of the affected amino-acid residues. The position of each variant is shown along with flanking AARS1 amino-acid residues from multiple, evolutionarily diverse species. The position of the affected residue is shown in red for each species.

mouse, zebrafish, and fly, and the R750 residue is conserved among all analyzed species (Figure 2.15B). These data suggest that the variants may affect AARS1 function.

2.3.13 C115R and R750W AARS1 *display loss-of-function effects in yeast complementation assays*

Yeast complementation assays have been performed to reveal loss-of-function characteristics of disease-associated *AARS1* variants[32,35,37,38]. To test for functional consequences of C115R and R750W *AARS1*, we compared the ability of each mutation to complement the loss of endogenous yeast *ALA1* compared to wild-type *AARS1*. Briefly, we used a previously validated haploid yeast strain deleted for endogenous *ALA1* and with viability maintained with *ALA1* on a *URA3*-bearing vector[35,37,38]. Wild-type or mutant *AARS1* expression constructs or empty constructs (with no *AARS1* insert) were transformed into the yeast strain, and growth was evaluated on 5-FOA medium, which selects for spontaneous loss of the *URA3*-bearing maintenance vector[237]. The empty plasmid did not support yeast growth, consistent with *ALA1* being an essential gene (Figure 2.16A). Wild-type *AARS1* expression constructs supported yeast growth, which indicates that human *AARS1* can complement loss of the endogenous *ALA1* locus (Figure 2.16A), consistent with previous findings[239]. The C115R *AARS1* variant was unable to support yeast growth (Figure 2.16A), suggesting that C115R represents a loss-of-function allele. The R750W *AARS1* variant showed similar yeast cell growth compared to wild-type *AARS1* (Figure 2.16A). We also tested Y690Lfs*3 and G913D *AARS1*; these variants were previously reported in siblings with an autosomal recessive syndrome including progressive microcephaly, hypomyelination, and epileptic encephalopathy[33]. In the yeast complementation assay, the Y690Lfs*3 and G913D *AARS1* variants were unable to support yeast growth (Figure 2.16A),

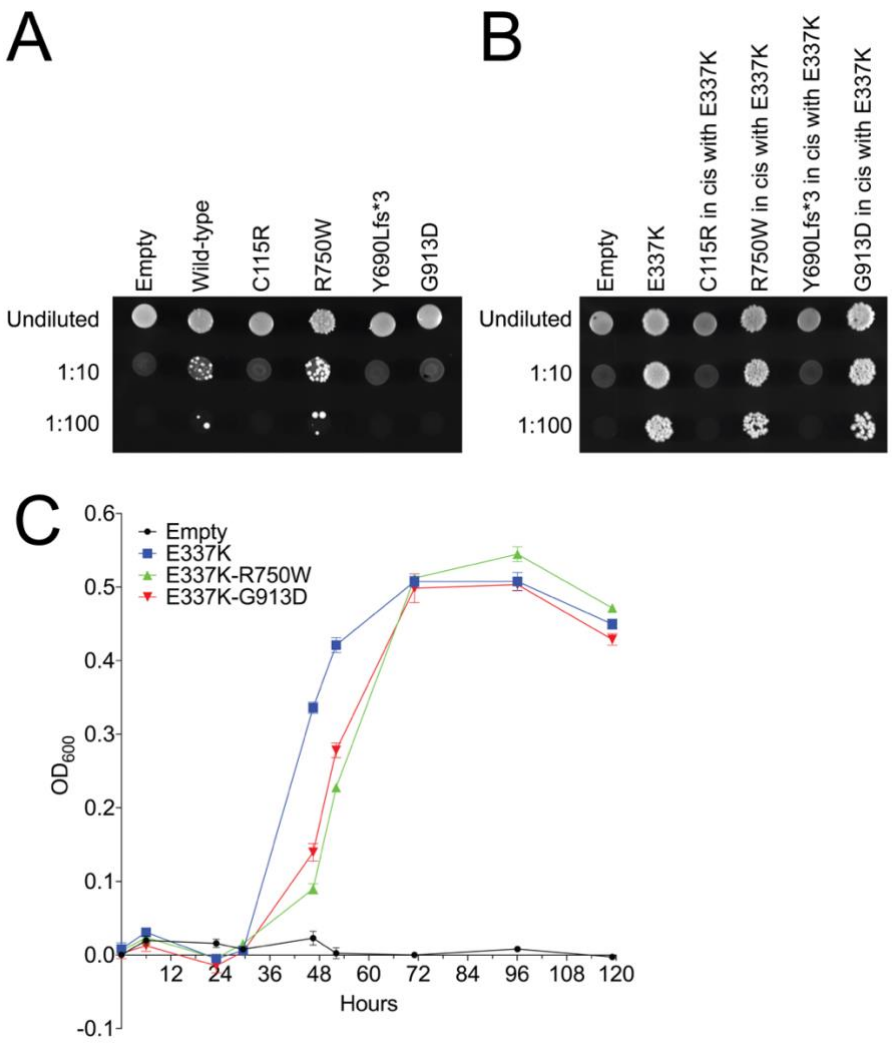


Figure 2.16. The C115R and R750W *AARS1* variants display loss-of-function effects in yeast complementation assays.

(A) Yeast lacking endogenous *ALA1* (the yeast ortholog of *AARS1*) were transformed with vectors containing wild-type *AARS1* or mutant *AARS1*, or a vector with no *AARS1* insert ('Empty'); the vector used in each experiment is indicated across the top. Resulting cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C. (B) Yeast lacking *ALA1* were transformed with vectors containing E337K *AARS1* or mutations in *cis* with E337K *AARS1*; the vector used in each experiment is indicated across the top. Resulting cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C. (C) Yeast transformed with E337K *AARS1*, R750W in *cis* with E337K *AARS1*, G913D in *cis* with E337K *AARS1*, and empty pYY1 were grown at 30°C in liquid media containing 5-FOA. Growth was assessed via OD_{600} absorbance readings measured at the indicated time points. Each time point represents the average of three replicate readings from each culture, and error bars represent standard deviation.

which is consistent with previously reported findings that the mutations decrease aminoacylation activity[33].

As described above (see Chapter 2.3.2), wild-type *AARS1* supports moderate yeast cell growth, and E337K *AARS1* supports increased cell growth compared to wild-type AARS1[38]. Therefore, to increase the ability to detect subtle differences in growth, each variant was also tested in *cis* with E337K *AARS1*. C115R in *cis* with E337K and Y690Lfs*3 in *cis* with E337K *AARS1* were still unable to support any yeast cell growth (Figure 2.16B) consistent with these alleles being complete loss-of-function alleles in this assay. R750W in *cis* with E337K and G913D in *cis* with E337K *AARS1* supported some yeast cell growth, but this was visibly reduced compared to E337K only (Figure 2.16B). To gain a more sensitive assessment of yeast expressing R750W in *cis* with E337K and G913D in *cis* with E337K *AARS1*, we tested these alleles in yeast growth curve assays in liquid 5-FOA selective media. Yeast with R750W in *cis* with E337K and G913D in *cis* with E337K *AARS1* demonstrated slower growth than E337K *AARS1* (Figure 2.16C). Mid-log phase growth of yeast expressing R750W in *cis* with E337K and G913D in *cis* with E337K *AARS1* is delayed (averaged across two experiments) 10 h and 9.5 h, respectively, compared to yeast expressing E337K *AARS1*. In sum, our humanized yeast complementation assays show that C115R and R750W *AARS1* impair gene function, consistent with pathogenicity in the recessive phenotype described.

2.3.14 Identification of YARS1 *variants in patients with a multi-system phenotype including cholestatic jaundice, sensorineural hearing loss, and developmental delay*

Variants in cytoplasmic tyrosyl-tRNA synthetase (*YARS1* [MIM 603623]) were previously associated with dominant intermediate Charcot-Marie-Tooth disease type C[25,122,168] and recessive multi-system phenotypes including developmental delay, liver dysfunction, lung dysfunction, and/or sensorineural hearing loss[167,247,248]. We identified two sisters (referred to as YARS1 Patients 1 and 2; Table 2.1) who presented with neonatal onset cholestatic jaundice, mild-moderate global developmental delay, and facial dimorphism. One sister also presented with sensorineural hearing loss that required hearing aids. Exome sequencing revealed that both patients were homozygous for I59T *YARS1* (NM_003680.3:c.176T>C; p.Ile59Thr), a variant that was not previously reported. Both of the unaffected parents were determined to be heterozygous for I59T *YARS1*. I59T is present in gnomAD[180,228] at a low frequency (7/282,858), and no homozygous individuals were present in this database. The I59 residue maps to the catalytic domain[167] (Figure 2.17A) and is conserved in all analyzed species except yeast and bacteria (Figure 2.17B). These data suggest that I59T may have functional consequences.

2.3.15 I59T YARS1 *supports yeast viability*

To determine the functional consequences of I59T *YARS1 in vivo*, we performed yeast complementation assays using established methods[168]. We used a previously validated haploid yeast strain deleted for endogenous *TYS1* and maintained with *TYS1* on a *URA3*-bearing vector[168]. Wild-type or I59T *YARS1* expression constructs or empty constructs (with no *YARS1* insert) were transformed into the yeast strain, and growth was evaluated on 5-FOA medium, which selects for spontaneous loss of the *URA3*-bearing maintenance vector[237]. The empty

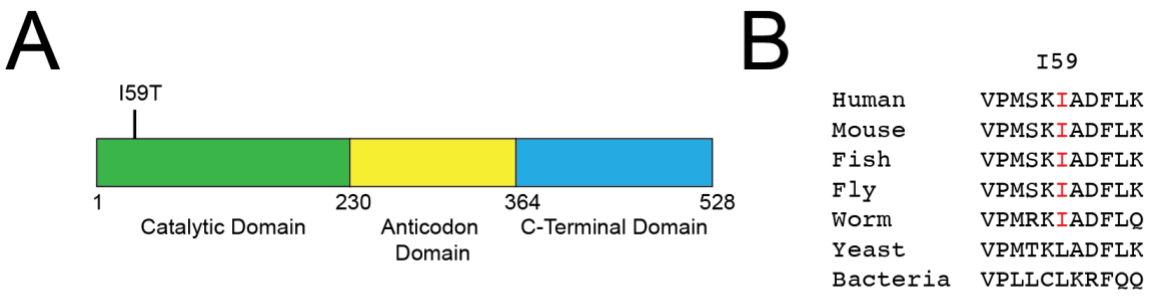


Figure 2.17. Localization and conservation of the I59T *YARS1* variant.

(A) YARS1 functional domains are indicated in green (catalytic domain), yellow (anticodon domain), and blue (C-terminal domain). The position of the variant is shown across the top, and numbers along the bottom indicate amino-acid positions for the YARS1 protein. (B) Conservation of the affected amino-acid residue. The position of the I59T variant is shown along with flanking YARS1 amino-acid residues from multiple, evolutionarily diverse species. The position of the affected residue is shown in red for each species.

plasmid did not support yeast growth, consistent with *TYS1* being an essential gene (Figure 2.18A). Wild-type *YARS1* supported yeast growth, which indicates that human *YARS1* can complement the loss of endogenous *TYS1* (Figure 2.18A). I59T *YARS1* also supported yeast growth, similar to wild-type *YARS1* (Figure 2.18A). To gain a more sensitive assessment of yeast growth, we performed yeast growth curves in liquid 5-FOA medium (Figure 2.18B). Yeast with wild-type *YARS1* and I59T *YARS1* expression constructs grew similarly, both reaching mid-log phase growth at about 48 h (Figure 2.18B). All together, these experiments indicate that when tested in yeast, I59T *YARS1* does not reduce yeast cell growth and is not a loss-of-function allele in this assay.

2.4 Discussion

In summary, we present clinical, genetic, and functional data that implicate newly identified ARS variants in human disease. We studied *AARS1* variants in dominant Charcot-Marie-Tooth (CMT) disease and *GARS1* variants in dominant infantile spinal muscular atrophy (iSMA), expanding the allelic heterogeneity of *AARS1*- and *GARS1*-associated dominant disease. Additionally, we described *AARS2*, *HARS1*, *AARS1*, and *YARS1* variants identified in patients with recessive disease, increasing the allelic and phenotypic heterogeneity of ARS-associated recessive disease. This work expands our understanding of the phenotypes associated with ARS variants, increases our knowledge of the ARS variants that cause disease, and provides insight into the underlying disease mechanisms.

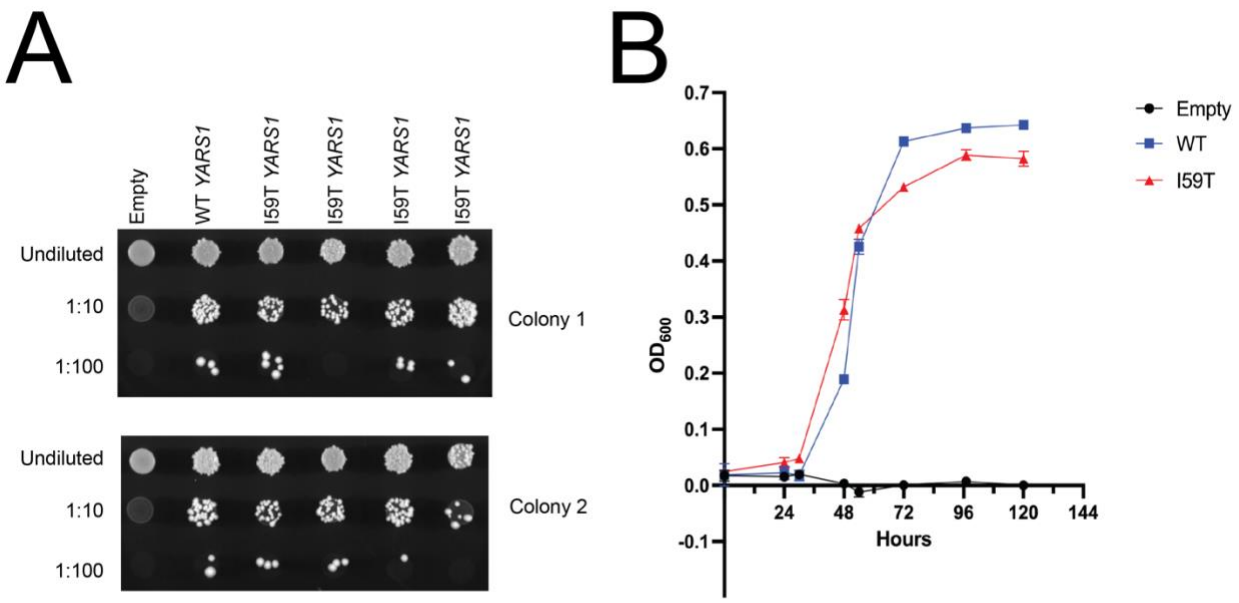


Figure 2.18. I59T *YARS1* supports yeast cellular growth.

(A) Yeast lacking endogenous *TYS1* (the yeast ortholog of *YARS1*) were transformed with vectors containing wild-type *YARS1* or K59T *YARS1*, or a vector with no *YARS1* insert ('Empty'); the vector used in each experiment is indicated across the top. Resulting cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C. Two colonies were analyzed. (B) Yeast transformed with wild-type *YARS1*, I59T *YARS1*, or empty pYY1 were grown at 30°C in liquid media containing 5-FOA. Growth was assessed via OD_{600} absorbance readings measured at the indicated time points. Each time point represents the average of three replicate readings from each culture, and error bars represent standard deviation.

2.4.1 ARS variants and dominant disease

We identified three families with CMT2 that carry variants in *AARS1*, and the genetic evidence supports pathogenicity of the identified *AARS1* variants. The variants segregate with disease, consistent with dominant inheritance; affect conserved residues; and are absent from gnomAD. In yeast complementation assays, R326W and S627L *AARS1* displayed loss-of-function or partial loss-of-function effects (Figure 2.3). Of note, R326 is only three amino-acid residues apart from the previously reported R329H mutation, which also demonstrates loss-of-function effects in yeast complementation assays and *in vitro* aminoacylation assays[35]. Interestingly, S627L is located in the editing domain of *AARS1*, and while variants in the *AARS1* editing domain have previously been associated with recessive disease[32,33], this is the first time that variants in the editing domain have been associated with dominant disease. In contrast to R326W, S627L, and the majority of previously published CMT-associated ARS variants, E337K *AARS1* improved cell growth in yeast complementation assays (Figure 2.3). Since lysine is the wild-type amino acid at this residue in yeast, this increased growth may reflect the mutation human enzyme functioning better specifically in yeast. However, E337K *AARS1* also resulted in increased tRNA charging compared to wild-type *AARS1* in *in vitro* aminoacylation assays (Figure 2.4). This suggests that E337K *AARS1* represents a hypermorphic mutation, and the dominant neurotoxicity may be associated with inappropriately increased enzyme kinetics. However, how increased tRNA charging activity may contribute to the disease pathogenesis in this patient is currently unclear. Further studies are needed to determine how an enzyme with increased aminoacylation could be toxic, including an assessment on how a rapidly charging enzyme alters the levels of available tRNA and amino acid in a cell, and whether that results in any downstream consequences on translation or other cellular processes, such as metabolic

pathways involving glycine. Importantly, the functional studies performed here tested this allele in isolation and not in the presence of the wild-type allele. Since the patients are heterozygous for E337K *AARS1*, future work is needed to investigate the effect of this mutation on the mutant-wild-type dimer (see Chapter 5).

In addition to CMT, variants in *GARS1* have also been associated with dominant infantile spinal muscular atrophy (iSMA). We identified two patients with severe, dominant, early-onset spinal muscular atrophy, who both had a *de novo* I334N *GARS1* variant. Clinical and genetic evidence is consistent with pathogenicity of I334N *GARS1*. The patients presented with similar clinical presentations to each other. The variant affects a conserved residue in the catalytic domain and is not present in gnomAD. Assessment of the previously identified, hereditary-neuropathy-associated I334F *GARS1* revealed reduced activity in *in vitro* aminoacylation assays[235], suggesting that the I334 residue is important for the function of the enzyme. We determined that I334N *GARS1* does not support yeast cell growth in complementation assays (Figure 2.6), demonstrating loss-of-function effects. In sum, our clinical, genetic, and functional data suggest that I334N *GARS1* causes iSMA. These findings expand the allelic heterogeneity of *GARS1*-associated disease and support that severe early-onset SMA can be caused by variants in this gene. How variants in *GARS1* can result in a spectrum of dominant phenotypes ranging from later-onset upper limb predominant peripheral neuropathy to early-onset SMA-like phenotypes is currently unclear. Work investigating whether differing effects on tRNA charging and/or genetic modifiers may be playing a role in the phenotypic heterogeneity of GARS1-associated dominant disease is needed (see Chapter 5).

The majority of the ARS variants associated with dominant phenotypes that we studied demonstrated loss-of-function effects in functional assays. These results are consistent with previously studied ARS variants associated with dominant neuropathy and consistent with a dominant-negative mechanism. A dominant-negative mechanism has been proposed[29] as a mechanism for ARS-mediated dominant disease since all of the ARS enzymes implicated to date function as homodimers in the cytoplasm, and the majority of variants cause reduced function in functional assays[29]. Therefore, mutant subunits may dimerize with wild-type subunits and reduce aminoacylation in patient cells. Moving forward, understanding whether dimerization is necessary for pathogenicity, and if mutations in monomers can cause dominant neuropathy will be important for assessing whether dominant-negative effects are occurring (see Chapter 5). Additionally, a toxic gain-of-function mechanism cannot be ruled out and may be occurring instead of, or in combination with, dominant-negative effects. With a gain-of-function mechanism, variants are proposed to lead to novel, dominantly toxic functions that may particularly affect the peripheral nerve[29]. Neuropathy-associated variants in three ARS loci have been reported to induce conformational openings that permit aberrant interactions with other proteins that affect pathways in the peripheral nervous system (see Chapter 1)[181]. More work is needed to determine whether all neuropathy-associated ARS variants lead to aberrant interactions, and if these interactions result in similar effects for variants in all five of the implicated ARS enzymes (see Chapter 5).

Additionally, despite the fact that ARS enzymes are ubiquitously expressed and essential for every cell, only cells in the peripheral nervous system seem to be affected in dominant peripheral neuropathies caused by ARS variants. This could imply that the peripheral nerve axon is

particularly vulnerable to the dominant negative and/or gain-of-function effects of these mutations and/or that these enzymes have a cell-type-specific function in addition to their canonical function in protein synthesis. Further work is needed to determine the noncanonical functions of ARSs and whether peripheral nerve axons are particularly sensitive to defects in translation.

2.4.2 ARS variants and recessive disease

We identified two siblings with cerebellar ataxia and bi-allelic *AARS2* variants, and our results indicate an expansion of the phenotypic spectrum of *AARS2*-related disorders to include a mitochondrial phenotype with ataxia even in the absence of leukoencephalopathy. Each patient is compound heterozygous for the same two *AARS2* variants, F131del and I328M, both of which affect highly conserved amino acids in the aminoacylation domain. Our yeast complementation assays support a complete loss-of-function effect for F131del and a partial loss-of-function effect for I328M *AARS2* (Figure 2.8); these data are consistent with both the recessive phenotype and the viability of the patients. One of the variants, F131del *AARS2*, was previously reported in a female patient who carried F131del on one allele and R199C and V730M on the other allele[43]. Her presentation differed from the patients described here as she presented with leukoencephalopathy and ovarian failure, which is a well-described feature of *AARS2*-associated neurological disease[43]. Our findings indicate that the F131del *AARS2* variant may not always be associated with leukoencephalopathy or with early ovarian failure depending on what variant (or variants) is on the other allele; the different phenotypes may be related to the functional output of each genotype.

To our knowledge, there has been only one other patient reported who had *AARS2* variants and childhood to adulthood-onset disease without leukodystrophy; Srivastava *et al*. recently reported a 34-year-old man with a family history of tremor, imbalance, and Charcot-Marie-Tooth disease, and his presentation included tremor, polyneuropathy, ataxia, and cognitive and psychiatric decline[242]. He was compound heterozygous for a splice site variant, c.2598 +1G>T, and R199C *AARS2*, and he also carried a paternally inherited *PMP22* duplication[242]. While the patients reported both here and in Srivastava *et al*. presented with ataxia and no leukoencephalopathy, the patients we described here did not have any additional variants in other genes. Additionally, Srivastava *et al*. reported MRI signal abnormalities of the periventricular frontal white matter[242], which differs from the patients we identified who had no signal abnormalities. Our results further indicate that *AARS2* variants can be associated with ataxia without leukoencephalopathy.

Additionally, we identified three patients from two families carrying bi-allelic *HARS1* variants who presented with clinical features including microcephaly, intellectual disability, skeletal deformities, and ataxic broad base gait. Consistent with previously reported *HARS1* variants[97,233], hearing loss was found in HARS1 Patient 3 and peripheral motor nerve involvement occurred in HARS1 Patient 2. Several pieces of evidence support the pathogenicity of the identified *HARS1* variants. All patients presented with an overlapping, recessive multi-system syndrome. Furthermore, functional studies demonstrate that each allele affects HARS1 function: (1) qPCR and western blot analysis revealed a significant reduction of *HARS1* mRNA and protein expression (Figure 2.12); (2) studies in yeast showed that V244Cfs*6 and L305dup cause a loss-of-function effect (Figure 2.13); and (3) aminoacylation assays using patient cells indicated a strong and significant reduction of tRNA charging (Figure 2.14). In sum, based on these genetic,

phenotypic, and functional data, the patients we describe here expand the clinical manifestations of *HARS1* variants and enlarge the set of aminoacyl-tRNA synthetases associated with ataxia and movement disorders.

We also expanded the allelic heterogeneity of *AARS1*-associated recessive disease. We described a three-year-old girl with postnatal onset microcephaly, failure to thrive, seizures, hypomyelination, thin corpus callosum, and developmental delay, and who is compound heterozygous for *AARS1* variants. Multiple pieces of evidence are consistent with the pathogenicity of the *AARS1* variants. The patient's phenotype overlaps with previously reported *AARS1*-related recessive disease phenotypes. Additionally, the variants affect well-conserved residues in important functional domains of the enzyme and display loss-of-function effects in yeast complementation assays (Figure 2.16). Based on these data, we expand the allelic heterogeneity of *AARS1*-associated recessive disease.

A previously unreported *YARS1* variant, I59T, was also described in two sisters with a recessive, multi-system phenotype including neonatal onset cholestatic jaundice, developmental delay, and facial dimorphism. Both patients are homozygous for the variant, which affects a conserved residue in the catalytic domain. When tested in a yeast complementation assay, I59T *YARS1* supported similar yeast cell growth as wild-type *YARS1* (Figure 2.18). Currently it is unclear whether I59T *YARS1* is pathogenic, and future work is needed to further investigate the functional consequences of this variant. *In vitro* aminoacylation assays would be helpful to test the tRNA charging abilities of the mutant enzyme. In fact, a subset of ARS mutations that support yeast cell growth have shown decreased activity in aminoacylation assays[177]. This may

be due to limitations of using yeast, which may not behave like or be as sensitive to decreased ARS activity as human tissues. Importantly, this variant was identified via exome sequencing, and it is possible that splice site or other variants in noncoding regions could be present in this patient and contributing to disease.

The majority of recessive-disease associated variants studied here exhibited complete or partial loss-of-function effects in protein expression studies, yeast complementation assays, and/or aminoacylation assays. These findings are consistent with the extensive previously published genetic and functional data implicating a partial loss-of-function molecular pathology in ARS-associated recessive disease (see Chapter 1 and Chapter 2.1)[29]. It has therefore been proposed that reduced ARS activity may lead to impaired translation in ARS-associated recessive disease, and that these effects on translation contribute to patient phenotypes[29]. Further work is needed to determine the downstream consequences of ARS variants on translation and whether impaired translation contributes to disease pathogenesis (see Chapter 4 and Chapter 5). Additionally, some ARSs have functions unrelated to translation (see Chapter 1)[19]; this raises the possibility that impairment of these non-canonical functions may also contribute to disease pathogenesis. More work is needed to determine the noncanonical functions of ARSs, and whether they are affected by these variants.

Interestingly, as more and more ARS variants are described, we are learning that ARS-associated recessive disease phenotypes can vary depending on which ARS is affected, despite all ARS enzymes being ubiquitously expressed. Here, we reported variants in four ARS that were associated with a variety of recessive disease presentations that had some overlapping features

(*e.g. HARS1*, *AARS1*, and *YARS1* variants were all associated with developmental delay) and some features unique to one of the four ARS (*e.g.* out of the four ARSs, *AARS1* variants were the only variants associated with hypomyelination). How these tissue-specific or tissue-predominant effects arise is not understood and may be due to differing requirements of tissues to translate proteins enriched for specific amino acids. Additionally, the differing availabilities of tRNAs and amino acids across tissues may also affect sensitivity of tissues to the impaired activity of a specific ARS. Uncovering the phenotypic and allelic spectrums of ARS-associated recessive disease phenotypes will improve our understanding of the phenotypes associated with each ARS, increase our ability to diagnose patients with ARS variants, and provide insights into disease mechanisms, which may ultimately inform therapy options.

Chapter 3

CARS1 Variants Cause a Multi-System Recessive Disease Including Microcephaly, Developmental Delay, and Brittle Hair

3.1 Introduction

Aminoacyl-tRNA synthetases (ARSs) are enzymes that charge tRNA molecules with cognate amino acids in the cytoplasm and mitochondria[2]. Consistent with the essential function and ubiquitous expression of these enzymes, mutations in 36 out of the 37 loci encoding an ARS have been implicated in severe, early-onset recessive disease phenotypes involving many different tissues[29]. The genotypes of patients with ARS-mediated recessive disorders suggest a partial loss-of-function mechanism for disease pathogenesis but are also consistent with complete loss of ARS function being lethal. Specifically, patients are either compound heterozygous for one missense mutation and one null allele, compound heterozygous for two missense mutations, or homozygous for a single missense mutation[3,29,30,153,195]. Consistent with this notion, functional studies have revealed that recessive disease-associated ARS mutations cause reduced protein levels in western blot assays, decreased mutant enzyme activity via *in vitro* kinetic assays, and/or diminished ability of the mutated gene to support cellular growth in yeast complementation assays[29,177]. As such, impaired translation as a consequence of decreased tRNA charging is the most likely molecular mechanism of ARS-mediated recessive disease[29], and we predict that with time, patients will be identified with loss-of-function variants in all of the ARS genes.

In this chapter, we evaluate the pathogenicity of variants in cysteinyl-tRNA synthetase 1 (*CARS1* [MIM 123859]), a gene that had not previously been associated with human inherited disease. It is important to note that with advancements in sequencing technologies, variants of uncertain significance are being rapidly discovered in patients with myriad disease phenotypes; before classifying these variants as pathogenic, careful evaluation is necessary[249]. Here, we show that the newly identified *CARS1* variants: (1) segregate with disease in three pedigrees, (2) are absent in the general population, (3) affect conservated amino-acid residues, and (4) reduce the function of CARS1 in several assays. Implicating *CARS1* variants in recessive disease phenotypes expands our understanding of the locus, allelic, and clinical heterogeneity of ARS-associated disease.

A subset of data in this chapter was previously published in the *American Journal of Human Genetics* (Volume 104, Issue 3, pages 520-529, March 7, 2019)[53]. Permission was requested for reproduction of the article through the Copyright Clearance Center; as an author of the article, I retain the right to include the article in a dissertation, and permission is not required. The author of this thesis performed all of the studies in this chapter with the following exceptions: clinical collaborators provided the phenotypic data and performed exome sequencing (Figures 3.1-3.4); Grazia Mancini at the Erasmus Medical Center performed the hair microscopy (Figure 3.5); Desiree Smith, Marisa Mendes, and Gajja Salomons at the Amsterdam University Medical Center performed aminoacylation assays using fibroblast lysates (Figure 3.13); Thomas Christian and Ya-Ming Hou at Thomas Jefferson University performed *in vitro* aminoacylation assays (Figure 3.17); and Sanger sequencing reactions were performed by the University of Michigan Advanced Genomics Core.

3.2 Methods

3.2.1 Patient sample collection and identification of CARS1 *variants*

Three families (referred to as CARS1 Families 1, 2, and 3) were identified with *CARS1* variants (Figure 3.1). Patients with *CARS1* variants were identified at the National Institutes of Health Undiagnosed Diseases Program (Subject 1-3) and the Department of Clinical Genetics at the Erasmus Medical Center (Subject 2-4, Subject 3-3, and Subject 3-4). Informed consent was obtained from all participants, and institute-specific review boards approved all studies. The research groups were connected via GeneMatcher[250]. In CARS1 Family 1, exome sequencing (ES) was performed using the Illumina HiSeq2000 platform and the TrueSeq capture kit as previously described[251-254]. Sequence data were aligned to the human reference genome (hg19), and variants were filtered with VarSifter[255] based on allele frequencies in the NIH-UDP cohort[252,256,257]. In CARS1 Family 2 and CARS1 Family 3, trio ES was performed after DNA enrichment with Agilent Sureselect Exome V4 Capture and using the HiSeq platform (101bp paired-end, Illumina); sequences were mapped with Burrows-Wheeler Aligner (BWA). Variants were detected with the Genome Analysis Toolkit and the variant call format (VCF) files were filtered using the software of Cartagenia (version 3.0.7).

To confirm each variant, Sanger sequencing analysis was performed. Genomic DNA was isolated from subjects' fibroblasts using the Wizard Genomic DNA Purification Kit (Promega). Primers (see Appendix) were designed to PCR amplify around each variant: CARS_Intron_10_F and CARS_Intron_11_R for R341H, S359L, and Q380*; CARS_Intron_11_F and CARS_Intron_12_R for L400Q; and CARS_22F and CARS_22R for S688Qfs*2. PCR products

were column purified and then submitted to the University of Michigan Advanced Genomics Core for Sanger sequencing. Chromatograms were viewed using Sequencher.

The prevalence of each variant in the general population was determined by searching gnomAD[180]. Conservation of each variant was examined by aligning CARS1 protein orthologs from multiple species with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The GenBank accession numbers for the protein sequences used were: human (*Homo sapiens*, NP_001742.1), mouse (*Mus musculus*, NP_001239522.1), zebrafish (*Danio rerio*, NP_001112372.1), worm (*Caenorhabditis elegans*, NP_001293288.1), yeast (*Saccharomyces cerevisiae*, NP_014152.1), and E. coli (*Escherichia coli* 55989, CAU96413.1).

3.2.2 Primary fibroblast cell culture

Primary fibroblasts from Subject 1-3, Subject 2-4, Subject 3-3, and two control (i.e., presumably bearing no pathogenic *CARS1* variants) individuals were obtained and utilized for RT-PCR and western blot analysis. Cells were cultured in T-75 cm^3-treated cell culture flasks (USA Scientific) and maintained at 37°C and 5% CO_2 in complete medium (Dulbecco's modified Eagle medium [Invitrogen]) with 10% fetal bovine serum (ThermoFisher Scientific), 1X L-glutamine (Corning), 1X Penicillin-Streptomycin (final concentration 50 units/ml penicillin and 50 ug/ml streptomycin; ThermoFisher Scientific).

3.2.3 Analysis of the CARS1 *genomic locus*

To investigate the *CARS1* genomic locus and transcript isoforms, we analyzed the human *CARS1* locus on chromosome 11 using the UCSC Human Genome Browser[258] (h38). To visualize

isoform 1 (NM_001751; 748 amino acids) and isoform 2 (NM_139273; 726 amino acids), we used the "RefSeq genes from NCBI track" and adapted data from these tracks for Figures 3.7A and 3.9A. We used the "Multiz Alignments of 100 Vertebrates" track to examine genomic sequence conservation among different species (Figure 3.9A).

3.2.4 RNA isolation and RT-PCR analysis

Fibroblast cells from control and Subject 1-3 were harvested at 90% confluence with 3 ml 0.25% trypsin (ThermoFisher Scientific), quenched with 5 ml complete cell medium, and 3 ml were collected at 2,000 rpm for 2 min. The resulting pellet was used for RNA isolation with the RNeasy Mini Kit (QIAGEN). cDNA was generated using 1 μg RNA and the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). To determine whether both *CARS1* cDNA isoforms 1 and 2 are expressed, previously published primers[259] (CAS1F and CAS1R; Appendix) were used to amplify a 71 bp fragment from cDNA isoform 1 and a 79 bp fragment from cDNA isoform 2. For each reaction, 23 μl of PCR Supermix (Life Technologies) was combined with 0.5 μl of each 20 μM primer solution and 1 μl of cDNA. Standard PCR conditions were used. RT-PCR products (5 μl) were incubated with 2 μl Novex Hi-Density TBE Sample Buffer (5X; Invitrogen) and 3 μl H_2O at 70°C for 10 min and then separated 30 min 120 V on a precast 8% TBE gel (Invitrogen). The gel was stained with SYBR Gold (Invitrogen). To assess the presence of both transcript isoforms in mouse, primers were designed to amplify around the last intron using cDNA from human (hsCARS exon 22F and hsCARS 3' UTR R; see Appendix) and mouse (mmCars around last intron F and mmCars around last intron R; see Appendix). For each reaction, 22.25 μl of PCR Supermix (Life Technologies) was combined with 0.5 μl of each 20 μM primer solution, 0.75 μl of DMSO, and 1 μl of cDNA. Standard PCR

conditions were used. RT-PCR products were separated by gel electrophoresis 60 min 135 V on a 1% TBE agarose gel.

3.2.5 Protein isolation and western blot analysis

Fibroblast cells were harvested at 90% confluence with 3 ml 0.25% trypsin (ThermoFisher Scientific), quenched with 5 ml complete cell medium, and 3 ml were collected at 2,000 rpm for 2 min. The resulting pellet was resuspended in 150 μl cell lysis buffer (RIPA buffer [ThermoFisher Scientific] + 1X EDTA-free protease inhibitor [ThermoFisher Scientific]) and incubated at 4°C for 30 min with gentle shaking. Cell debris was collected at 4°C and 13,000 rpm for 30 min, and the resulting supernatant was subjected to western blot analysis.

Total protein concentration was measured using the Thermo Scientific Pierce BCA Protein Assay Kit, and 20 ug protein/sample was analyzed. Protein samples were prepared with 1X SDS-sample buffer (ThermoFisher Scientific) and 2 μl 2-mercaptoethanol (βME), denatured at 99°C for 5 min and separated on a precast 4-20% or 6% tris-glycine protein gel (ThermoFisher Scientific). Prior to protein transfer, a polyvinylidene difluoride membrane (MilliporeSigma) was prewashed in 100% methanol and soaked in 1X transfer buffer (ThermoFisher Scientific) + 10% methanol with extra thick blot paper (Bio-Rad) and 3 mm chromatography paper (Whatman PLC). Protein was transferred to the membrane overnight using a Mini-Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) at 0.03 A at room temperature and then blocked overnight at 4°C in blocking solution (1 g dry milk powder, 50 ml 1X TBST [tris-buffered saline solution and tween 20, pH 7.5]). The membrane was cut and incubated with either anti-CARS1 antibody (Santa Cruz at 1:250 dilution or Sigma at 1:1,000 dilution), anti-IARS1 antibody

(GeneTex) (1:1,000 dilution), or anti-actin antibody (MilliporeSigma) (at 1:5000 dilution) at 4°C overnight. The membrane was then washed for five minutes three times in 1X TBST and incubated in blocking solution with 10 ng/ml donkey anti-mouse IgG HRP secondary antibody (MilliporeSigma) or 10 ng/ml donkey anti-rabbit IgG HRP secondary antibody (MilliporeSigma) at room temperature for 1 hr. The membrane was washed for five minutes three times in 1X TBST and exposed using SuperSignal West Dura substrate and enhancer (ThermoFisher Scientific) according to manufacturer's instructions.

3.2.6 Steady state aminoacylation reactions on protein lysates from fibroblasts

Desiree Smith, Marisa Mendes, and Gajja Salomons from the Amsterdam University Medical Center performed aminoacylation assays using protein lysates from subject fibroblasts. Assays were performed at 37°C in reaction buffer (50 mmol/L Tris buffer pH 7.5, 12 mmol/L MgCl, 25 mol/L KCl, 1 mg/ml bovine serum albumin, 0.5 mmol/L spermine, 1 mmol/L ATP, 0.2 mmol/L yeast total tRNA, 1 mmol/L dithiotreitol, 0.3 mM [13C4,15N] cysteine, and [D2] glycine). The reactions were terminated with trichloroacetic acid (TCA). After reaction termination and washing, ammonia was added so that [13C4,15N] cysteine and [D2] glycine would be released from the tRNAs. Labeled amino acids were quantified by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Intra-assay variation was <15%. Statistical significance was determined with a one-tailed, unpaired t-test.

3.2.7 Generation of CARS1 *and* CRS1 *expression constructs*

For yeast complementation assays, *CARS1* expression constructs were generated using Gateway cloning technology (Invitrogen). The human *CARS1* (isoform 1; 748 amino acid) open-reading

frame was amplified from human cDNA with primers designed with Gateway sequences attB1 (forward) and attB2 (reverse; primer sequences in Appendix). The yeast *CRS1* locus was amplified from yeast gDNA (primer sequences in Appendix). Entry clones were generated by recombining gel-purified PCR products into the pDONR221 vector using BP clonase (Invitrogen). After transformation into *E. coli*, DNA from individual entry clones was isolated and subjected to Sanger sequencing analysis to confirm the presence of the wild-type *CARS1* or *CRS1* open-reading frame. Since we were unable to amplify *CARS1* transcript isoform 2 (726 amino acids) from human fibroblast cDNA, mutagenesis primers (see Appendix) were designed to insert the 8 bp that are present in isoform 2 into isoform 1. Oligonucleotides containing sequences corresponding to the patient variants (R341H, Q380*, S359L, L400Q, and S688Qfs*2) were also generated (primer sequences in Appendix), and mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent). After transformation into *E. coli*, DNA from individual clones was purified and Sanger sequenced to confirm the presence of each variant and the absence of any errors. Validated wild-type and mutant *CARS1* were LR cloned into the *LEU2*-bearing pYY1 expression construct[232], which we previously made Gateway-compatible[86]. Validated wild-type and mutant *CRS1* were LR cloned into the *LEU2*-bearing pRS315 expression construct. The reactions were transformed into *E. coli*, purified, and digested with *Bsr*GI New England Biosystems (Waltham, MA) to confirm the presence of the *CARS1* open-reading frame or *CRS1* locus.

For *in vitro* aminoacylation assays, the patient variants were introduced into a previously generated wild-type *CARS1* expression construct (pET-24d) using the same mutagenesis primers described above with the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent). After

transformation into *E. coli*, DNA from individual clones was purified and Sanger sequenced to confirm the presence of each variant and the absence of any errors.

3.2.8 Generation of a haploid yeast strain

A haploid Δ*CRS1* strain that also contains the *CRS1*/pRS316 maintenance vector was generated. To generate the maintenance vector, the yeast *CRS1* locus was amplified from yeast DNA (primer sequences in Appendix). Gel-purified PCR products were BP-cloned into the pDONR221 vector. After transformation into *E. coli*, DNA from individual entry clones was isolated and subjected to Sanger sequencing analysis to confirm the presence of the wild-type *CRS1* locus (with one nonsynonymous change due to difficulty with amplification). *CRS1* was then LR-cloned into a URA3-bearing pRS316 vector.

A commercially available diploid heterozygous *crs1*Δ yeast strain (MAT a/α, *his3*Δ*1/his3*Δ*1, leu2*Δ*0/leu2*Δ*0, LYS2/lys2*Δ*0, met15*Δ*0/MET15, ura3*Δ*0/ura3*Δ*0*; YSC6274-201924108, Horizon Discovery) was obtained. Lithium acetate transformations were performed to introduce *CRS1/*pRS316 into the strain; 200 ng of *CRS1*/pRS316, 5 μl 10 mg/ml salmon sperm (Millipore Sigma), and 400 μl PEG solution (40 g polyethylene glycol [Hampton Research] in 43 ml H2O with 7ml 1M lithium acetate [Hampton Research]) were used. Transformation reactions were grown on yeast media lacking uracil (Teknova) at 30°C. One colony from the transformation plate was picked and resuspended in 100 μl yeast extract, peptone, and dextrose (YPD; Fisher Scientific) media, struck out onto a YPD plate, then incubated overnight at 30°C. One colony from that plate was picked and struck out onto a pre-sporulation plate (5% D-glucose, 3% Difco nutrient broth, 2% agar, and 1% Difco yeast extract) and incubated overnight at 30°C. A colony

from that plate was picked, struck out onto a fresh pre-sporulation plate, and incubated overnight at 30°C. A large streak of colonies was inoculated in 2 ml sporulation media and incubated 5 days at 25°C 200 rpm and then 3 days at 30°C 200 rpm. Cells were pelleted by spinning for 2 min at 4000 x g and resuspended in 1 ml 1M D-sorbitol. Lyticase (Sigma-Aldrich) was added (2 µl 10 mg/ml lyticase to 100 µl cells) and incubated for 11 min at room temperature. Sporulated strains were dissected using an MSM 400 dissecting microscope (Singer Instrument) and plated on YPD plates. Resulting spores were individually patched onto solid growth medium containing Geneticin (G418) or 0.1% 5-fluoroorotic acid (5-FOA), or media lacking uracil (Teknova). Two spores that grew on G418 and yeast medium lacking uracil, but did not grow on 5-FOA medium, were selected for use in the yeast viability assays.

3.2.9 Yeast complementation assays

To assess the ability of wild-type and mutant human *CARS1* alleles to support cellular growth, the haploid yeast strain with the endogenous *CRS1* locus deleted and viability maintained via a pRS316 vector bearing wild-type (WT) *CRS1* was transformed with WT *CRS1*, WT *CARS1* (isoform 1), *CARS1* isoform 2, mutant (R341H, Q380*, S359L, pL400Q, and S688Qfs*2) *CARS1,* mutant (D714*) *CRS1*, or empty pYY1 (i.e., bearing no *CARS1* insert). Growth on solid media lacking leucine and uracil selected for transformed cells. Colonies were grown to saturation in liquid media and then spotted onto 0.1% 5-FOA complete solid medium (Teknova) to select for cells that will have spontaneously lost the *URA3*-bearing maintenance vector[246]. Yeast viability was assessed by visual inspection after 5 days of incubation at 30°C. At least two independently generated constructs for wild-type and mutant *CARS1* or *CRS1* were transformed. Two colonies per transformation were assayed.

3.2.10 Yeast growth curves

To generate growth curves, yeast colonies from strains transformed with WT *CARS1,* R341H *CARS1*, S359L *CARS1,* and empty pYY1 were inoculated in liquid medium lacking leucine and uracil and incubated 2 days at 30°C 275 rpm until saturation. Saturated culture (10 µl) was spiked into 4 ml liquid medium lacking leucine and uracil and the OD_{600} was measured in triplicate at 0, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 70, 78, 84, and 103 hours using the Promega GloMax Luminometer. The average of the measurements at each time point was graphed using GraphPad Prism, with error bars representing standard deviation. Colonies from three independent transformations using independently generated constructs were tested.

3.2.11 In-vitro *aminoacylation assays*

In vitro aminoacylation assays were performed by Thomas Christian and Ya-Ming Hou at Thomas Jefferson University. Wild-type and mutant CARS1 were expressed in *E. coli* with a C-terminal His tag and purified with nickel affinity resins (Novagen). The T7 transcript of human $tRNA^{Cys}$/GCA (GCA, anti-codon) was prepared and purified as previously described[259], heat denatured at 85°C for 3 min, and annealed at 37°C for 20 min before use. Steady-state aminoacylation assays were monitored at 37°C in 50 mM HEPES (pH 7.5), 20 mM KCl, 10 mM MgCl2, 4 mM DTT, 2 mM ATP, and 50 µM ^{35}S-cysteine (Perkin Elmer) at a specific activity of 12,100 dpm/pmole. The reaction was initiated by mixing CARS1 enzyme (12 nM) with varying concentrations of tRNA (1.0-16.0 μM). Aliquots of a reaction mixture were quenched by 0.24 M iodoacetic acid and 0.1 M sodium acetate in formamide (pH 5.0) and were spotted on filter paper[260], precipitated by 5% trichloroacetic acid, washed, dried, and measured for radioactivity

using a liquid scintillation counter (LS6000SC; Beckman Coulter Inc). The amount of radioactivity retained on filter pads was corrected for quenching effects to determine the amount of synthesis of Cys-tRNA^{Cys}. Steady-state kinetics was determined by fitting the initial rate of aminoacylation as a function of tRNA concentration to the Michaelis–Menten equation[261].

3.3 Results

3.3.1 Identification of patients with CARS1 *variants*

We identified four patients from three unrelated families (referred to as CARS1 Families 1, 2, and 3) with similar clinical presentations and bi-allelic variants in the cysteinyl-tRNA synthetase 1 (*CARS1*) gene. The appropriate, institute-specific review boards approved all studies, and informed consent was obtained from all subjects. The subject from CARS1 Family 1 (referred to as Subject 1-3; Figure 3.1A) is of mixed European and French-Canadian descent and was enrolled in the NIH Undiagnosed Diseases Program. He was born at 33 weeks gestation to asymptomatic parents and presented with intrauterine growth retardation and microcephaly. In childhood he presented with failure to thrive, non-progressive cognitive delay, peripheral neuropathy, osteoporosis, proportionate short stature, recurrent hernias, mild aortic root dilatation, recurrent elbow dislocation, feeding difficulties, esophagitis requiring Nissen fundoplication, urinary retention, and chronic pain. At age 24 he was diagnosed with hypothalamic hypogonadism and delayed puberty and type II diabetes mellitus. On examination he had: dysmorphic features; barrel shaped trunk with wasting of distal extremities; hypospadias with chordee; hyperextensible joints; myopia; central hypotonia but increase extremity tone; prominent lateral ventricles and sulci with mild cerebral atrophy upon brain MRI (Figure 3.2A-C); and fine brittle hair (Figure 3.3A) and brittle nails (Figure 3.4). Polarized light microscopy of

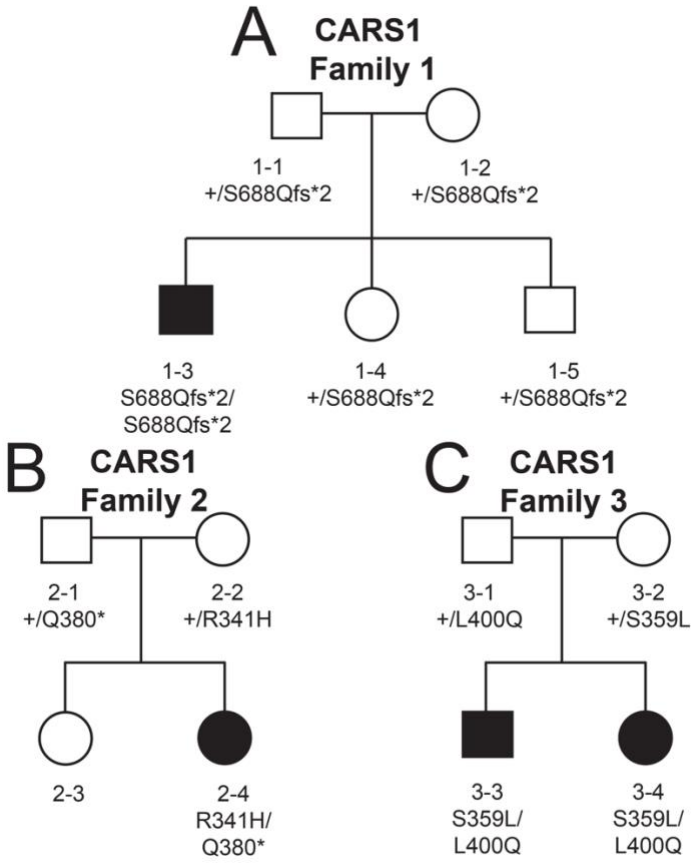


Figure 3.1. Pedigrees harboring *CARS1* variants.

Simplex pedigrees are shown for CARS1 Family 1 (A), CARS1 Family 2 (B), and CARS1 Family 3 (C). Squares represent males, and circles represent females. Subject numbers and genotypes are indicated under each symbol, and filled shapes indicate affected subjects.

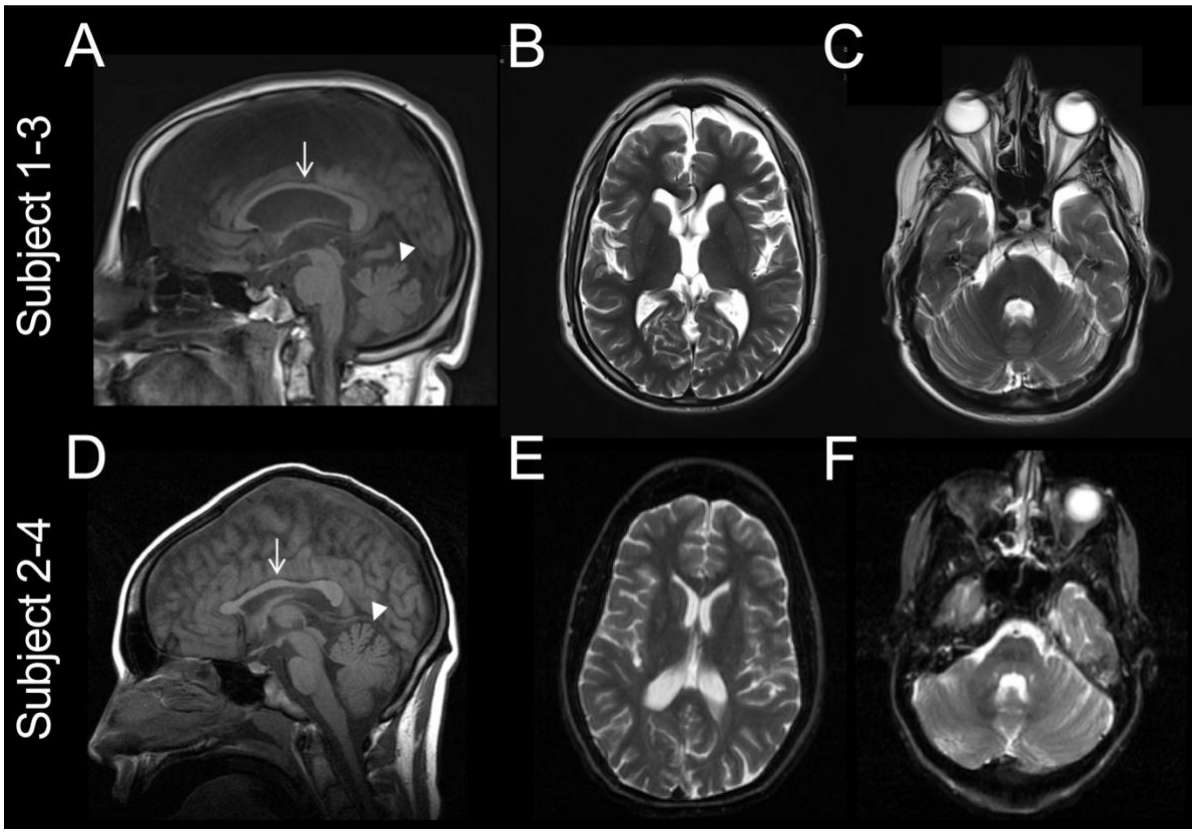


Figure 3.2. Subjects with *CARS1* variants present with central nervous system features.

Sagittal T1 weighted (A) and axial T2 weighted (B and C) imaging on Subject 1-3 shows moderate cerebral atrophy, thin corpus callosum (A; arrow) and mild atrophy of the superior cerebellar vermis (A; arrowhead), thin cerebellar folia, incomplete falx cerebri, incomplete tentorium, and variant anatomy of the circle of Willis. Sagittal T1 weighted imaging on Subject 2-4 shows thin splenium of corpus callosum (D; arrow) and mild atrophy of the vermis (D; arrowhead). T2 weighted axial images of Subject 2-4 show moderate global cerebral atrophy with deep sulci (E), thin cerebellar folia (F), decreased white matter volume, and globally normal myelination (D-F). These data were obtained from clinical collaborators.

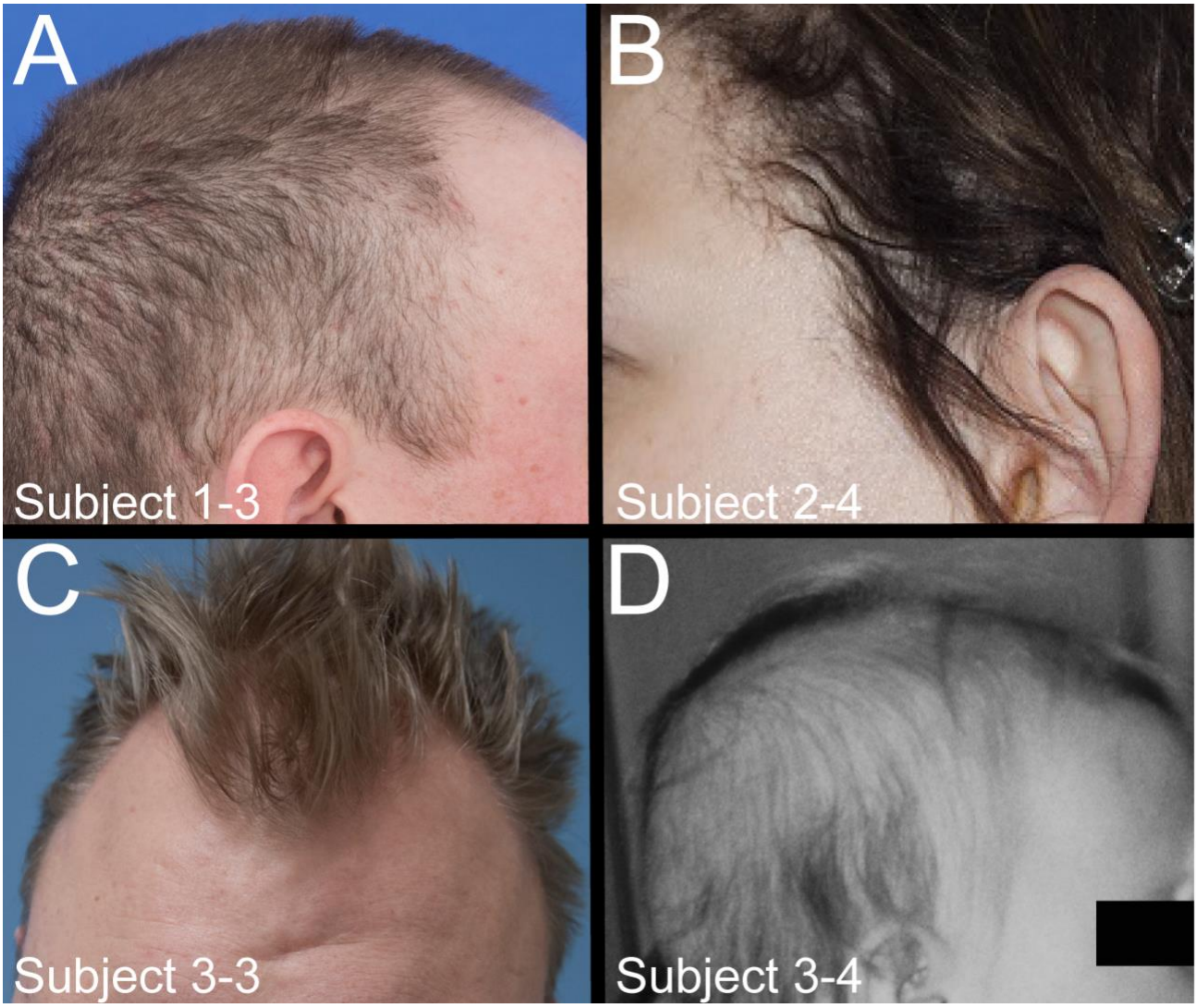


Figure 3.3. Subjects with *CARS1* variants present with brittle hair.

Photographs show the brittle scalp hair in Subject 1-3 (A), Subject 2-4 (B), Subject 3-3 (C), and Subject 3-4 (D). Images were obtained from clinical collaborators.

Figure 3.4. Brittle nails in Subject 1-3.

Photographs showing the left (A) and right (B) foot nail phenotype identified in Subject 1-3. Images were obtained from clinical collaborators.

hair shafts revealed moderate tiger-tail patterns compared to controls (Figure 3.5B). Subject 1-3 has mild motor, language, and cognitive disabilities.

The subject from CARS1 Family 2 (referred to as Subject 2-4; Figure 3.1B) is of Dutch descent and was born to asymptomatic parents at 38 weeks gestation via elective cesarean section due to intrauterine growth retardation. She was hospitalized for two months after birth due to a small atrial septal defect and failure to thrive. In childhood she developed restless behavior, poor sleep, mild dystonia, ataxic hand movements, poor coordination, and developmental delay. At age 30 she was diagnosed with diabetes mellitus. CT scan of the head at 2 years was normal. Brain MRI at 20 years revealed areas of apparent delayed myelination, decreased white matter volume, prominence of cerebellar folia, small corpus striatum, hypoplasia of corpus callosum, and mild, globally diffuse cerebral and cerebellar atrophy (Figure 3.2D-F). The quality of the MRI and the lack of repeat MRI did not allow distinction between hypomyelination and demyelination. On examination she had: microcephaly; fine facial features; fragile nails; and sparse, brittle, and thin hair (Figure 3.3B). Light microscopy of hair shafts revealed trichoschisis and trichorrhexis nodosa, polarized microscopy showed weak but consistent tiger tail banding (Figure 3.5C), and scanning EM showed longitudinal grooves (data not shown). Subject 2-4 has a moderate level of motor, language, and cognitive disability.

CARS1 Family 3 includes two affected siblings (referred to as Subject 3-3 and Subject 3-4; Figure 3.1C) of Dutch descent born to asymptomatic parents. Both subjects presented with: developmental delay, mild spastic ataxia, wide-based gait, pyramidal signs, liver steatosis, and thin hair (Figure 3.3C, D). Polarized microscopy of hair shafts revealed trichorrhexis and

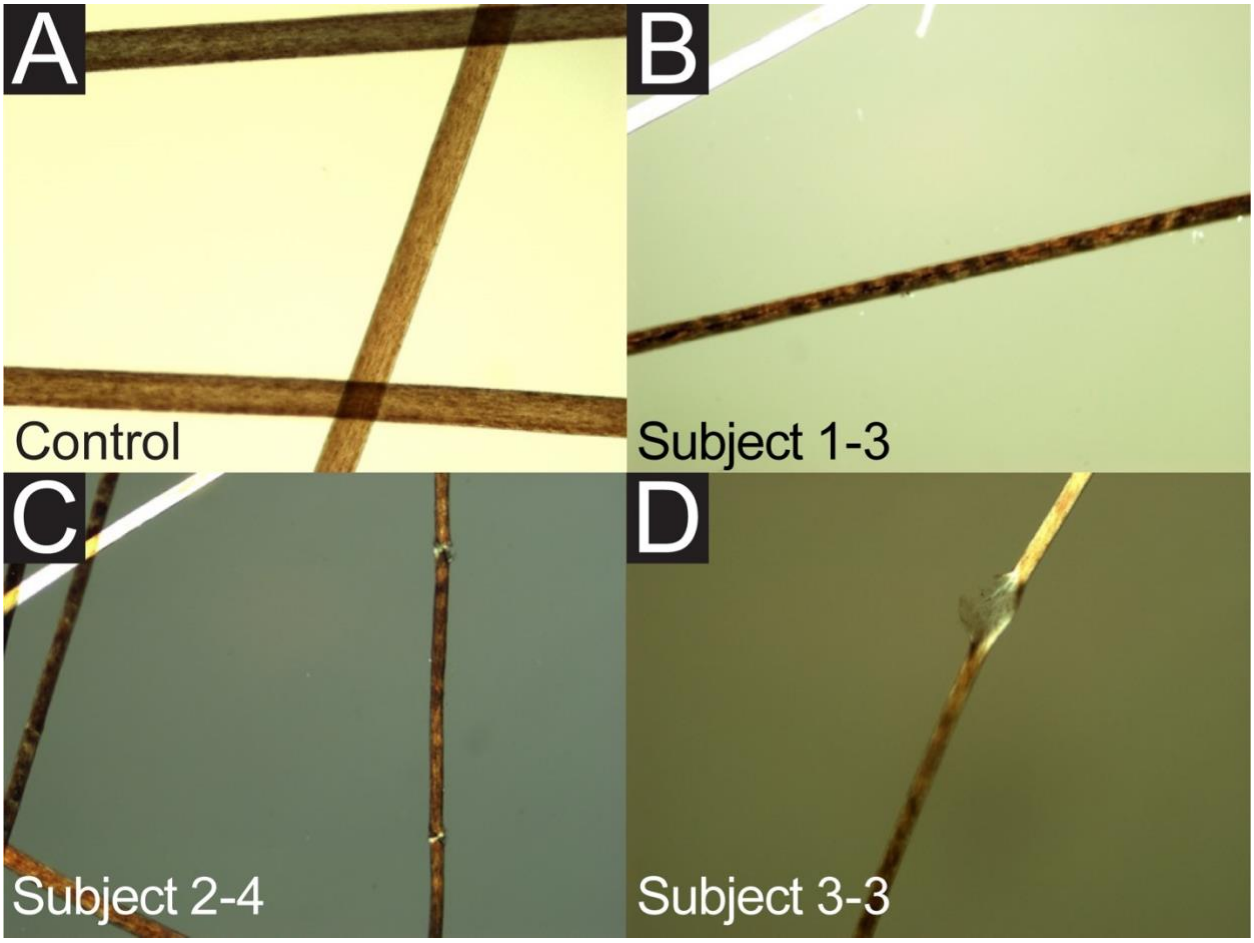


Figure 3.5. Subjects with *CARS1* variants present with hair shaft abnormalities.

Polarized light microscopy of hair shafts at the same magnification. Depicted are: an unrelated, age-matched healthy subject (A); Subject 1-3 showing moderate tiger-tail patterns (B); Subject 2-4, showing trichorrhexis and tiger-tail pattern (C); and Subject 3-3 showing trichorrhexis and moderate tiger-tail patterns (D). Note the difference in shaft diameter between healthy (A) and affected (B-D) subjects. Microscopy was performed by Grazia Mancini at the Erasmus Medical Center.

moderate tiger-tail patterns in Subject 3-3 (Figure 3.5D) and mild tiger-tail patterns in Subject 3-4 (data not shown). Subject 3-3 was observed from infancy due to unexplained failure to thrive, cholestasis, and fat malabsorption, and had complex partial and generalized seizures. Subject 3-4 presented at two months with congenital nystagmus and high myopia; at nine months she had repeated convulsions associated with fever and was diagnosed with hepatomegaly and steatosis. Brain CT of Subject 3-3 revealed hemiatrophy and a wide right ventricle (data not shown). Subject 3-3 is currently 25 years old and has a moderate level of motor, language, and cognitive disability. Subject 3-4 is currently 20 years old and has a mild level of motor, language, and cognitive disability.

3.3.2 Identification of CARS1 *variants*

To identify candidate variants for the unexplained phenotypes observed in our subjects, exome sequencing (ES) was performed on all affected individuals and their unaffected parents (Figure 3.1). Analysis of ES data revealed that all four affected individuals are homozygous or compound heterozygous for *CARS1* variants (Table 3.1 and Figure 3.1). Variant analysis by ES in CARS1 Family 1 revealed additional candidate disease-causing variants in *DEAF1* (MIM 602635), *SOX30* (MIM 606698), and *PTPN13* (MIM 600267). In CARS1 Family 2, ES analysis revealed no additional candidate variants. Finally, in CARS1 Family 3, ES analysis revealed candidate variants in *ZNF185* (MIM 300381). Importantly, the phenotypic similarity among subjects is consistent with all three families carrying disease-associated *CARS1* variants, which were the only common candidates identified via ES; it should be noted that the research groups studying these families were connected via GeneMatcher[250] based on the detection of *CARS1* variants.

Table 3.1. *CARS1* variants identified in subjects with recessive disease

Subject	Nucleotide change[a]	Amino acid change[b]	Detection in gnomAD[c]	Minor allele frequency
1-3	c.2061dup	p.Ser688Glnfs*2	2 / 251,406	0.000008
2-4	c.1022G>A	p.Arg341His	8 / 178,368	0.00004
2-4	c.1138C>T	p.Gln380*	1 / 184,276	0.000005
3-3	c.1076C>T	p.Ser359Leu	8 / 194,406	0.00004
3-3	c.1199T>A	p.Leu400Gln	Not present	0

[a]Human nucleotide positions correspond to GenBank Accession number NM_001751.5
[b]Human amino acid positions correspond to GenBank Accession number NP_001742.1
[c]http://gnomad.broadinstitute.org

Sanger sequencing analysis of genomic DNA isolated from subject fibroblasts confirmed each variant (Figure 3.1 and Figure 3.6). Subject 1-3 is homozygous for S688Qfs*2 (c.2061dup; p.Ser688Glnfs*2; Table 3.1). This variant is present in the gnomAD database[180] only in the heterozygous state (2/251,406). Subject 2-4 is compound heterozygous for two *CARS1* variants. The maternally inherited variant, R341H (c.1022G>A; p.Arg341His), is present in gnomAD (8/178,368) with no homozygous individuals detected (Table 3.1). The paternally inherited variant, Q380* (c.1138C>T; p.Gln380*), is present in gnomAD (1/184,276) with no homozygous individuals detected (Table 3.1). Subject 3-3 and Subject 3-4 are both compound heterozygous for two *CARS1* missense variants. The maternally inherited variant, S359L (c.1076C>T; p.Ser359Leu), is present in the gnomAD database (8/194,406) with no homozygous individuals detected (Table 3.1). The paternally inherited variant, L400Q (c.1199T>A; p.Leu400Gln), is not present in the gnomAD database (Table 3.1).

3.3.3 Characterization of the CARS1 *locus and CARS1 protein*

CARS1 is the sole enzyme responsible for charging $tRNA^{Cys}$ molecules in the cytoplasm of mammalian cells; CARS2 is a nuclear-encoded enzyme responsible for charging $tRNA^{Cys}$ molecules in mitochondria[2]. CARS1 is a class I ARS enzyme containing a catalytic domain that activates and transfers cysteine to tRNA, and an anticodon binding domain for tRNA recognition[262-264]. There are two transcript isoforms of *CARS1* that arise due to alternative splicing of the penultimate exon[259] (Figure 3.7A). It has previously been reported that both isoforms are expressed in a wide range of human tissues[259]. Isoform 1 seems to be the primary isoform; in a previous study, approximately 20% of the mRNA encoding *CARS1* contained the 8

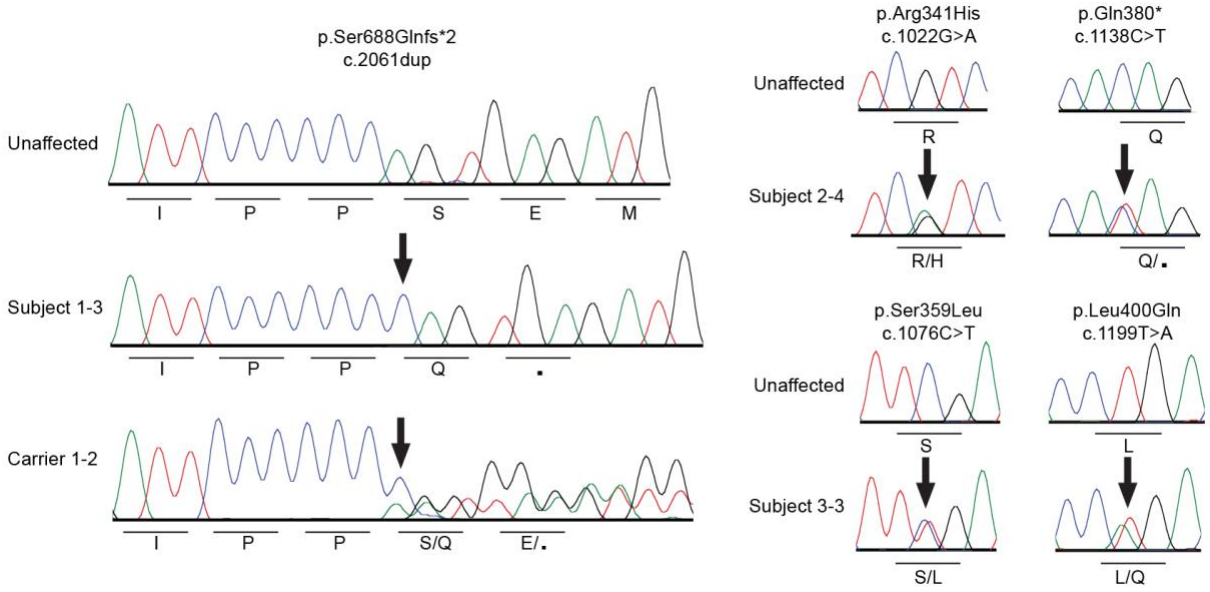


Figure 3.6. Sanger sequencing confirmation of each variant.

Representative sequence chromatograms are shown for the individuals indicated on the left. The protein and cDNA annotations of the identified variants are above and the effect on amino-acid sequence is indicated below. Arrows show the position of each variant.

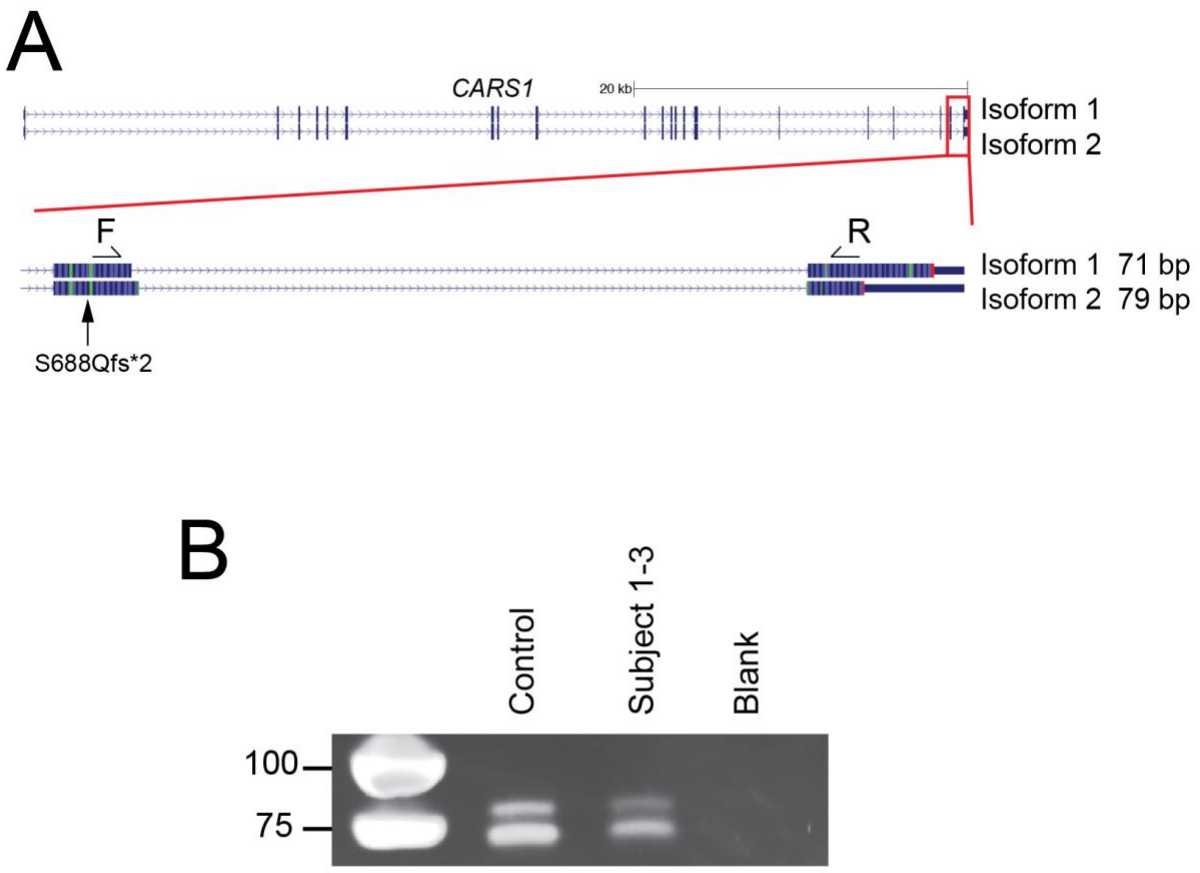


Figure 3.7. Alternative splicing between the last two exons results in two *CARS1* transcript isoforms.

(A) The human *CARS1* isoform 1 (NM_00751; 748 amino acids) and isoform 2 (NM_139273; 726 amino acids) from the UCSC Genome Browser are shown. The red box indicates the region that is zoomed in below, which displays the last two exons and outcomes of alternative splicing. Isoform number is located on the right. (B) RT-PCR assay to test for the presence of transcript isoform 1 (71 bp) and isoform 2 (79 bp). Sample names are across the top, and base pair sizes of markers are provided on the left. The locations of the primers ('F' and 'R') used are depicted in (A).

bp insertion present in isoform 2[259]. We experimentally confirmed the presence of both transcript isoforms in cDNA from human fibroblasts via RT-PCR (Figure 3.7B). These two transcript isoforms are predicted to result in two protein isoforms that differ in their C-terminal amino acid sequences[259] (Figure 3.8A), and we observed the presence of both protein isoforms (isoform 1 is ~85 kDa, and isoform 2 is ~83 kDa) in human fibroblast cells via western blot analysis (Figure 3.8B).

When investigating whether both isoforms are present in other species, we could identify transcripts in other species that were similar to isoform 1, but we were unable to identify any transcripts similar to isoform 2 in many other organisms. To investigate this further, we analyzed the "Multiz Alignments of 100 Vertebrates" in the UCSC Genome Browser[258] (Figure 3.9A) to assess whether expression of both transcripts would be predicted based off of the genomic sequences. Interestingly, the splice donor site in isoform 1 seems to be very well conserved across species; however, the splice donor site for isoform 2 is only present in some primate species (human, chimp, gorilla, orangutan, and gibbon), suggesting isoform 2 may be restricted to only some primate species (Figure 3.9A). We experimentally assessed the presence of both isoforms in human and mouse by amplifying around the last exon-exon junction from cDNA and then analyzing with Sanger sequencing. In chromatograms from human fibroblast cDNA, we see peaks on peaks, consistent with the presence of both transcript isoforms (Figure 3.9B). However, in mouse brain cDNA chromatograms, we do not see any evidence for transcript isoform 2 (Figure 3.9C), suggesting that indeed, isoform 2 may not be expressed in mouse.

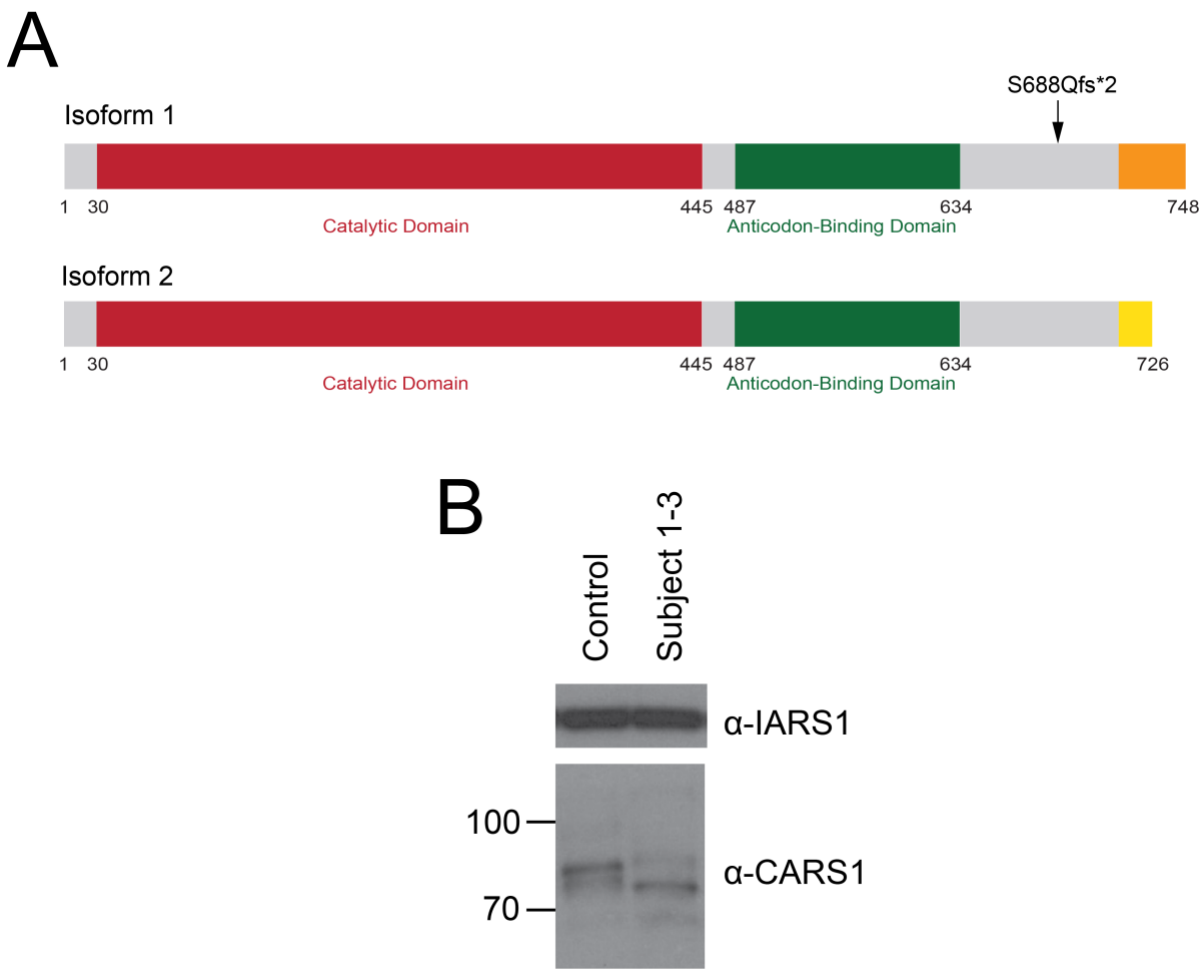


Figure 3.8. CARS1 protein isoforms.

(A) CARS1 functional domains in isoforms 1 and 2 are indicated in red (catalytic domain) and green (anticodon-binding domain), and numbers along the bottom indicate amino-acid positions for the CARS1 protein. (B) Western blot analyses were performed using total protein lysates isolated from control or Subject 1-3 fibroblasts, and an anti-CARS1 or anti-IARS1 antibody was used (indicated on the right). Sample names are across the top and sizes (kDa) are indicated at the left. Analysis revealed the presence of two CARS1 protein isoforms (~83 and ~85 kDa) in control cells and a single truncated CARS1 protein (~78 kDa) in cells from Subject 1-3.

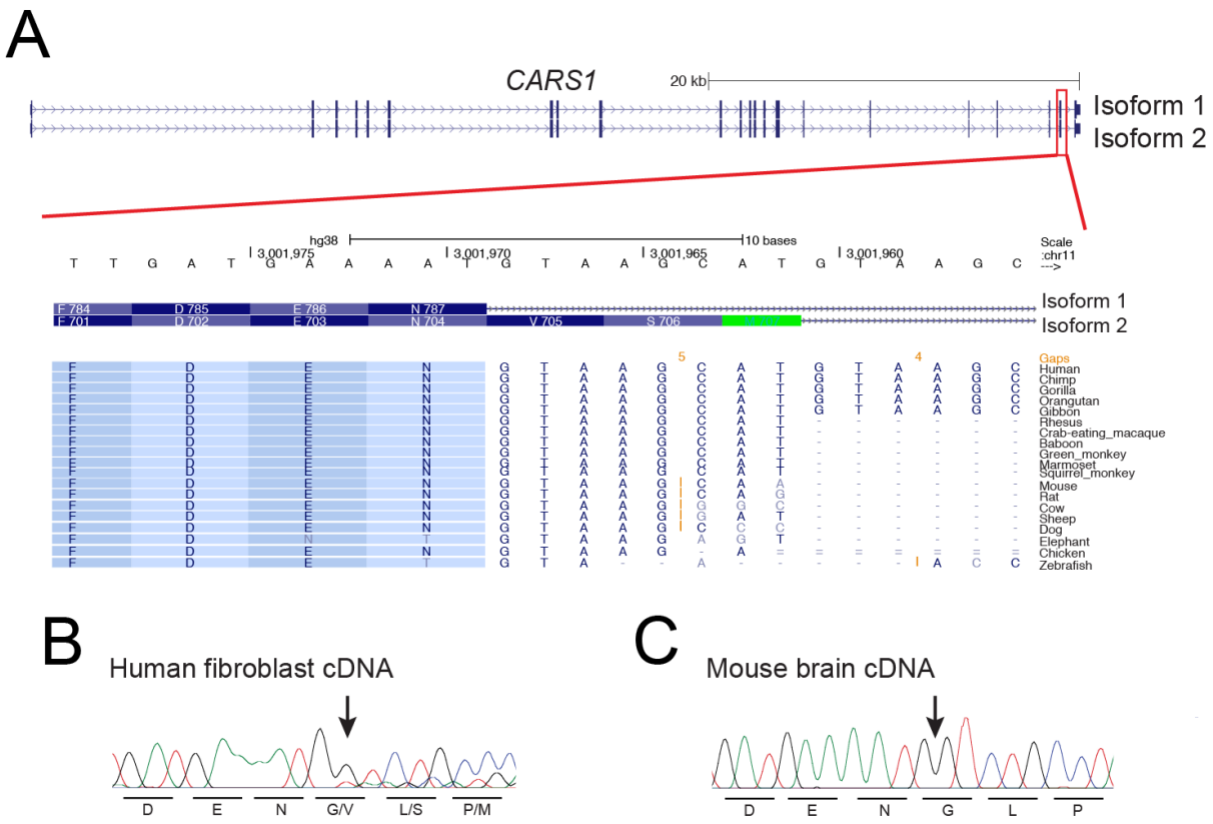


Figure 3.9. *CARS1* transcript isoform 2 may be restricted to some primate species.

(A) The human *CARS1* genomic locus from the UCSC Genome Browser is shown. Isoform 1 (NM_00751; 748 amino acids) and isoform 2 (NM_139273; 726 amino acids) are depicted and labeled on the right. The red box indicates the region that is zoomed in below, which displays the end of the second to last exon and the beginning of the last intron. Sequence conservation data from the Multiz Alignments of 100 Vertebrates are shown for the indicated species on the right. (B) Representative sequence chromatograms are shown for human fibroblast cDNA. The arrow indicates the location where peaks on peaks start, consistent with the presence of both transcript isoforms 1 and 2. The amino acid sequence is indicated below. (C) Representative sequence chromatograms are shown for mouse brain cDNA. The arrow indicates the predicted location where an 8 bp insertion would be present if transcript isoform 2 was present. The amino acid sequence is indicated below.

To evaluate the function of each *CARS1* isoform, we developed a yeast complementation assay. Briefly, a haploid yeast strain was generated with the endogenous yeast *CARS1* ortholog (*CRS1*) deleted and viability maintained with *CRS1* on a *URA3*-bearing maintenance vector. Each *CARS1* isoform was modeled in the human *CARS1* gene; *CARS1* was previously reported to rescue loss of *CRS1*[265]. Each *CARS1* isoform (or wild-type *CRS1*) was transformed, and growth was evaluated on 5-FOA medium, which selects for the spontaneous loss of the maintenance vector[246]. Wild-type *CRS1* rescued yeast viability, while an empty vector did not (Figure 3.10). This is consistent with *CRS1* being an essential gene. *CARS1* isoform 1 supported yeast viability (Figure 3.10), consistent with human *CARS1* isoform 1 supporting loss of the endogenous *CRS1*. *CARS1* isoform 2 did not support any yeast cell growth (Figure 3.10), suggesting that *CARS1* isoform 2 does not support loss of endogenous *CRS1*. This could indicate that *CARS1* isoform 2 is not functional in yeast; it has previously been reported that both proteins were functional in *in vitro* aminoacylation assays[259]. Further work is needed to understand the functional differences between the isoforms.

3.3.4 Localization and conservation of CARS1 *variants*

The R341H, S359L, Q380*, and L400Q *CARS1* variants all affect residues that map to the catalytic domain (Figure 3.11A). The Q380* *CARS1* variant predicts a loss of 368 of the 748 CARS1 (isoform 1) amino-acid residues including complete loss of the anticodon binding domain. The S688Qfs*2 *CARS1* variant maps to the C-terminus in a region that is conserved among eukaryotes and that is important for the binding specificity between CARS1 and $tRNA^{Cys}$[266]. The evolutionary conservation of each affected CARS1 amino-acid residue was assessed by aligning protein sequences of CARS1 orthologs from multiple species. The

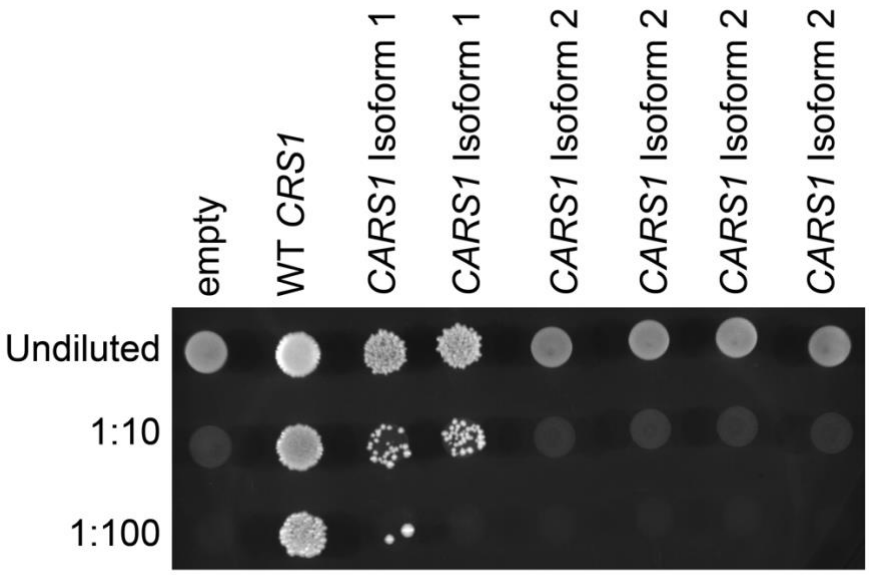


Figure 3.10. *CARS1* isoform 2 does not support yeast growth in yeast complementation assays.

Haploid yeast lacking endogenous *CRS1* (the yeast ortholog of *CARS1*) were transformed with vectors containing *CARS1* isoform 1 or isoform 2, or a vector with no *CARS1* insert ('empty'), which are indicated across the top. Resulting cultures were plated undiluted or diluted (1:10 or 1:100) on solid growth medium containing 5-FOA and grown at 30°C for 5 days.

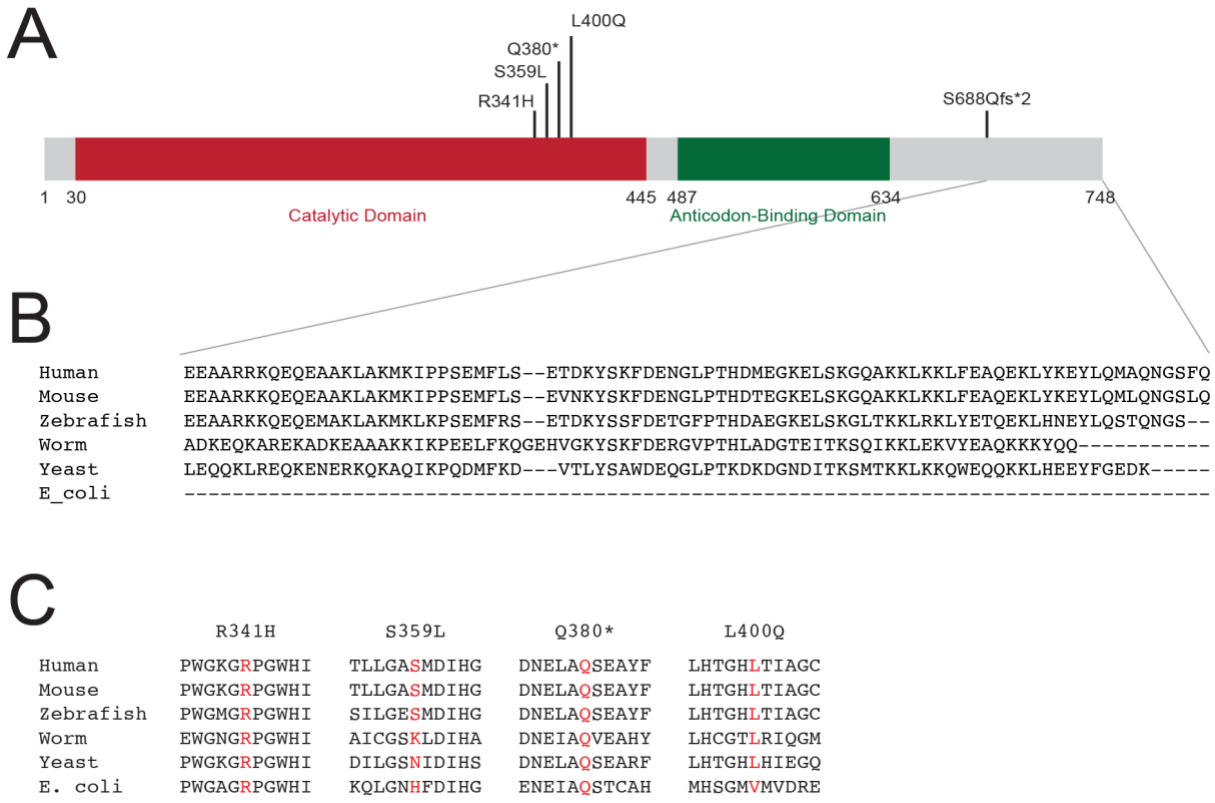


Figure 3.11. Localization and conservation of *CARS1* variants.

(A) CARS1 (isoform 1) functional domains are indicated in red (catalytic domain) and green (anticodon-binding domain), and the position of the variants are shown across the top. Numbers along the bottom indicate amino-acid positions for the CARS1 protein. (B) The CARS1 amino acid residues that S688Qfs*2 *CARS1* is predicted to ablate in multiple, evolutionarily diverse species. (C) The position of each variant is shown along with flanking CARS1 amino acid residues from multiple, evolutionarily diverse species. The protein annotation of the variant is above, and the species names are on the left. The position of the affected residue is shown in red for each species.

S688Q*2 *CARS1* variant is predicted to truncate the (isoform 1) protein by 60 amino acids (Figure 3.11B). Of those 60 amino acids, 28% (17/60) are conserved among all species analyzed except bacteria and 70% (42/60) are conserved among human, mouse, and zebrafish (Figure 3.11B). The R341 residue is conserved among all species analyzed ranging from human to bacteria; the L400 residue is conserved in all species analyzed except bacteria; and the S359 residue is conserved among human, mouse, and zebrafish (Figure 3.11C).

*3.3.5 S688Qfs*2 and Q380** CARS1 *reduce expression of full-length CARS1*

The S688Qfs*2 *CARS1* introduces a termination codon in the penultimate exon; therefore, it is possible that the *CARS1* transcripts may undergo or escape nonsense mediated decay (NMD). NMD is a post transcriptional mechanism that is used by cells to control the quality of mRNA[267]. Studies have informed a model for NMD where a typical target has a stop codon more than 50-55 bp upstream of the 3'-most exon-exon junction[268]. If the stop codon is fewer than 50 bp upstream of the last exon-exon junction, it may escape NMD[268]. S688Qfs*2 *CARS1* introduces a stop codon either 45 bp (isoform 1) or 53 bp (isoform 2) upstream of the last exon-exon junction. Therefore, it is possible that the variant may result in variable NMD of the transcripts (i.e., isoform 1 may escape NMD, and isoform 2 may be subjected to NMD), which could then affect the levels of the truncated protein. We performed RT-PCR using primers[259] to amplify a 71 bp fragment from cDNA isoform 1 and a 79 bp fragment from cDNA isoform 2 (Figure 3.7B). We observed two fragments of expected sizes in both the control and Subject 1-3 fibroblasts (Figure 3.7B). This supports the idea that neither isoform undergoes complete NMD, which is consistent with the viability of the patient. Since S688Qfs*2 *CARS1* introduces a stop codon upstream of the alternative splicing event, we predicted a single truncated protein may be expressed in this

patient (Figure 3.8A). To investigate the expression of CARS1, we performed western blot analysis using control and Subject 1-3 fibroblasts (Figure 3.8B). In fact, we observed a single band (~78 kDa; Figure 3.8B) running lower than either of the wild-type isoforms, consistent with the stable expression of a truncated CARS1 protein and the viability of the patient.

The Q380* *CARS1* variant also encodes a premature stop codon and is therefore predicted to reduce expression of full-length CARS1. To test this and assess CARS1 protein expression in all of the subject fibroblasts, we performed western blot analyses using protein isolated from affected subject fibroblasts and unaffected controls using antibodies against CARS1 and actin. The molecular weights of the two wild-type CARS1 isoforms are predicted to be ~82 kDa and ~85 kDa[259], which appear as a single band when lysates are electrophoresed on a 4-20% tris-glycine gel (Figure 3.12A). In protein isolated from Subject 1-3 fibroblasts, there is a single band at ~78 kDa, consistent with the predicted size of the truncated protein (Figure 3.12A) and with our previously observed results in Figure 3.8B. We therefore conclude that a truncated protein is expressed and stable, which is consistent with the viability of the subject. In protein isolated from Subject 2-4 fibroblasts, full-length CARS1 protein levels were significantly reduced compared to control cells (Figure 3.12A). This could be explained by the Q380* *CARS1* variant ablating the expression of full-length CARS; no truncated protein was detectable upon repeated analysis of an entire blot (Figure 3.12B). In protein isolated from Subject 3-3 fibroblasts, full-length CARS1 protein levels were similar to control cells (Figure 3.12A), which is consistent with this subject carrying two expressed and stable missense CARS1 protein variants.

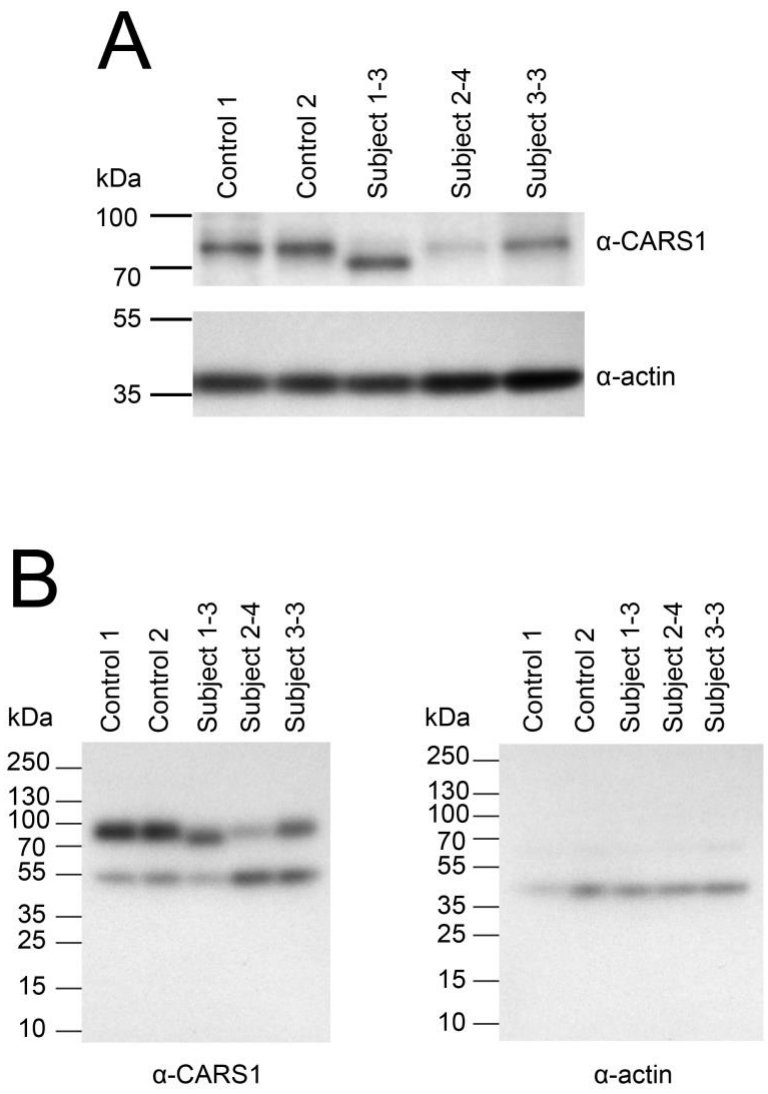


Figure 3.12. Subject 1-3 expresses a truncated CARS1 protein, and Subject 2-4 has reduced CARS1 protein expression.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts using an anti-CARS1 (sc-390230, Santa Cruz, purified mouse antibody, 1:250 dilution) or anti-actin (A5060, Sigma, purified rabbit antibody, 1:5,000 dilution) antibody (indicated on the right). Sample names are across the top and sizes (kDa) are indicated at the left. Analysis revealed a truncated CARS1 protein in cells from Subject 1-3 and reduced CARS1 protein expression in cells from Subject 2-4 compared to controls. (B) Similar western blot as described in (A) but showing the entire blot for CARS1 and actin. An anti-CARS1 (HPA002384, Sigma, 1:1,000 dilution, recognizes amino acids 110-238) antibody was used. This is an independent repeat of the experiment in (A). Please note the presumed non-specific band at ~55kD.

3.3.6 CARS1 enzyme from patient fibroblasts demonstrate decreased aminoacylation activity

To assess the effect of *CARS1* variants in the context of the affected subject genotypes, we performed steady state aminoacylation reactions using protein lysates from subject and control fibroblasts as the source of ARS enzymes. We assessed the ability of these enzymes to charge tRNA with radiolabeled cysteine or glycine, which was measured as an internal control. No decreases were observed for GARS1 (internal control) activity in affected versus unaffected individuals (p-value = 0.624). Compared to control cells and those isolated from the unaffected mother in CARS1 Family 1, each subject showed significantly reduced CARS1 activity (Figure 3.13; p-value < 0.001): ~70% reduction for Subject 1-3, ~80% reduction for Subject 2-4, and ~40% reduction for Subject 3-3. Importantly, these data are consistent with each *CARS1* variant causing a complete or partial loss-of-function effect.

3.3.7 CARS1 *variants result in reduced growth in yeast complementation assays*

To evaluate the functional consequences of each *CARS1* allele in isolation, we performed yeast complementation assays. Each *CARS1* variant was modeled in the human *CARS1* gene (isoform 1), which rescues loss of *CRS1* (Figure 3.10). Wild-type or mutant *CARS1* (or wild-type *CRS1*) was transformed into a haploid yeast strain with the endogenous yeast *CARS1* ortholog (*CRS1*) deleted and viability maintained with *CRS1* on a *URA3*-bearing maintenance vector. Growth was then evaluated on 5-FOA medium, which selects for the spontaneous loss of the maintenance vector[246]. The expression of CARS1 in transformed yeast was assessed via western blot analysis (Figure 3.14). Wild-type *CRS1* and wild-type human *CARS1* rescued yeast viability, while an empty vector did not (Figure 3.15A). This is consistent with our previously observed results in Figure 3.10, and with human *CARS1* supporting loss of the endogenous *CRS1* and with

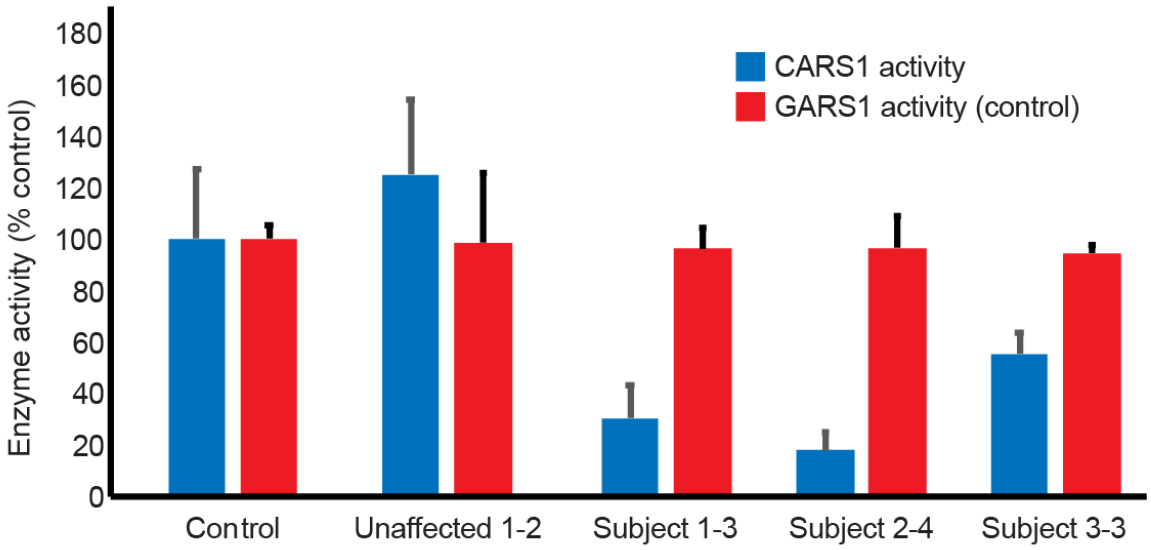


Figure 3.13. CARS1 enzymes from patient fibroblasts exhibit decreased enzyme activity.

The cytosolic fraction of subject fibroblasts was used to determine CARS1 (blue) and GARS1 (red; internal control) steady state aminoacylation activities. CARS1 activity is shown as a percentage of GARS1 activity. Samples are labeled across the bottom, and error bars indicate standard deviations of three technical replicates. These experiments were performed by Desiree Smith, Marisa Mendes, and Gajja Salomons at the Amsterdam University Medical Center.

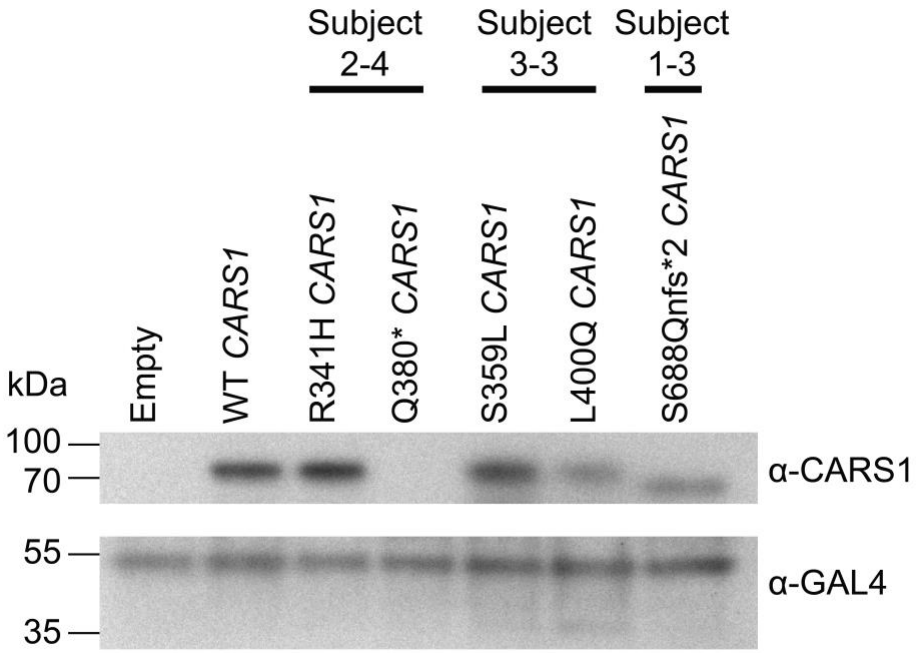


Figure 3.14. Human CARS1 is expressed in transformed yeast.

Western blot analyses were performed with total protein lysates from haploid yeast lacking endogenous *CRS1* that have been transformed with vectors containing the indicated inserts. Sample names are across the top and sizes (kDa) are indicated at the left.

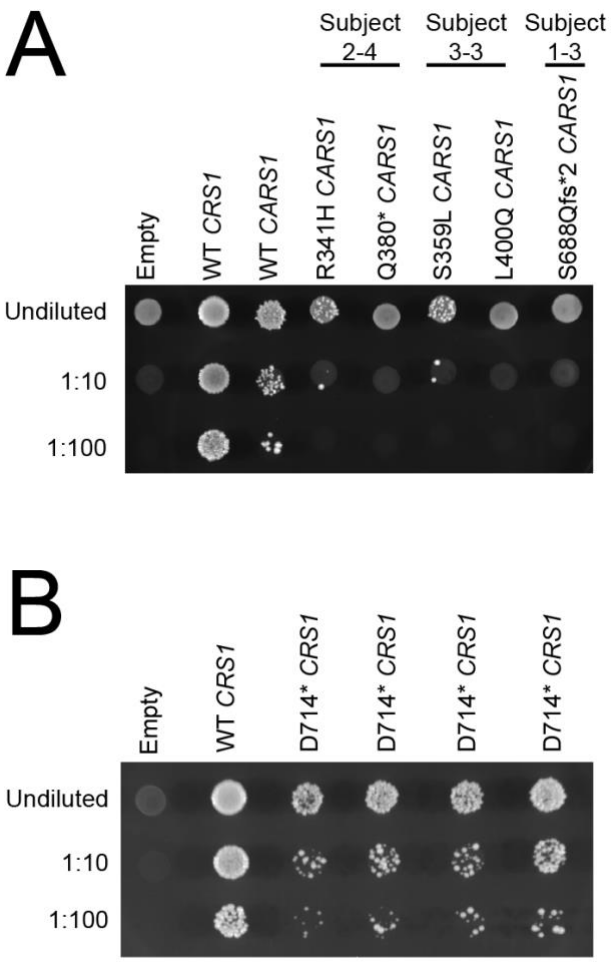


Figure 3.15. All five, disease-associated *CARS1* variants support reduced yeast cell growth in yeast complementation assays.

Haploid yeast lacking endogenous *CRS1* (the yeast ortholog of *CARS1*) were transformed with: (A) vectors containing wild-type *CRS1*, wild-type or mutant *CARS1*, or a vector with no *CARS1* insert ('Empty'); or (B) vectors containing wild-type or D714* *CRS1*, or a vector with no *CRS1* insert ('Empty'). Vector names are indicated across the top. Resulting cultures were plated undiluted or diluted (1:10 or 1:100) on solid growth medium containing 5-FOA and grown at 30°C for 5 days. The corresponding affected subject who carries each variant is listed above.

CRS1 being an essential gene, respectively. The Q380*, L400Q, and S688Qfs*2 *CARS1* variants did not support any yeast cell growth (Figure 3.15A), consistent with a loss-of-function effect. However, S688Qfs*2 *CARS1* not supporting any yeast growth is not consistent with the viability of Subject 1-3, who is homozygous for the variant. This result may be due to testing the variant in human *CARS1* in yeast, as wild-type human *CARS1* supports reduced growth compared to the yeast ortholog *CRS1* (Figure 3.10). Therefore, we tested the variant modeled in yeast *CRS1*; the D714* *CRS1* variant supports reduced yeast cell growth compared to wild-type *CRS1* (Figure 3.15B), consistent with a partial loss-of-function effect and with the viability of the patient. The R341H and S359L *CARS1* variants were associated with severely reduced but not ablated yeast cell growth (Figure 3.15A), consistent with these variants being hypomorphic alleles. To gain a more sensitive assessment of the R341H and S359L alleles, we performed growth curves in liquid 5-FOA media (Figure 3.16). Wild-type *CARS1*-expressing yeast entered mid-log phase at ~ 48 h, while yeast expressing S359L and R341H *CARS1* were delayed; S359L entered mid-log phase at ~61 h, and R341H entered mid-log phase at ~69 h (Figure 3.16). This is consistent with the slower growth observed for R341H and S359L *CARS1* yeast on solid 5-FOA media and suggests that R341H and S359L *CARS1* are hypomorphic alleles.

3.3.8 R341H and S359L CARS1 *reduce aminoacylation activity*

To further assess the hypomorphic R341H and S359L alleles, we performed aminoacylation assays, which evaluate the kinetic properties of ARS variants and have been established as informative for studying disease-associated ARS alleles[177]. We tested recombinant human CARS1 proteins (wild-type, R341H, and S359L) for the ability to charge human cytoplasmic $tRNA^{Cys}$ with ^{35}S-labeled cysteine as previously described[259,269]. Analysis of the initial rate of

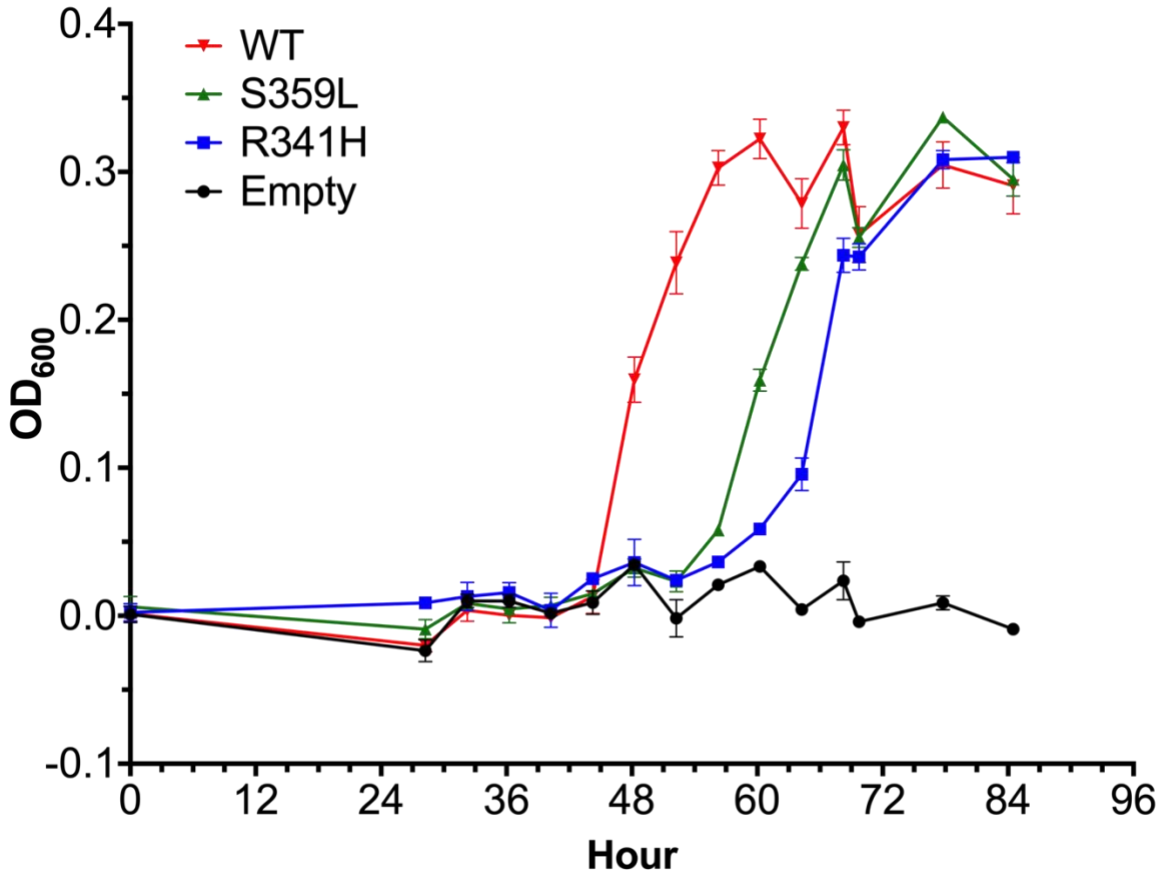


Figure 3.16. Growth curves of yeast expressing S359L and R341H *CARS1*.

Yeast expressing wild-type, S359L, or R341H *CARS1*, or no insert ('Empty') were grown in liquid 5-FOA media. OD_{600} readings were taken at the intervals indicated on the X-axis (hours). Error bars represent standard deviation of the average OD_{600} of three replicates for each strain.

aminoacylation as a function of the tRNA substrate concentration showed that both mutant CARS1 proteins cause a reduction in enzyme activity (Figure 3.17 and Table 3.2): R341H resulted in a 50% reduction in activity compared to wild-type CARS1; and S359L resulted in an 84% reduction in activity compared to wild-type CARS1. These data are consistent with the yeast complementation assay and the skin fibroblasts results and indicate that each of these variants is a hypomorphic allele.

3.4 Discussion

In summary, we present clinical, genetic, and functional data that implicate *CARS1* variants in a multi-system, recessive disease that includes microcephaly, developmental delay, and brittle hair and nails. The genetic and clinical evidence we present supports the pathogenicity of the identified *CARS1* variants in several ways. First, three independent pedigrees displayed an overlapping phenotype with unusual features including brittle hair and nails; these data are consistent with a common underlying genetic etiology among the three families. Second, *CARS1* variants segregate with the presumed recessive phenotype in all three families, and each affected subject is homozygous or compound heterozygous for variants that result in truncated proteins or that affect highly conserved amino acids. Third, our functional studies strongly support the pathogenicity of each *CARS1* variant. Specifically, each variant causes a complete or partial loss-of-function effect on protein expression, tRNA charging, or yeast complementation assays. Indeed, these assays are commonly used to demonstrate a complete or partial loss-of-function effect for disease-associated ARS mutations[177]. Aminoacylation assays have tested the effects of variants in 15 ARS genes implicated in recessive disease and yeast complementation assays have tested the effects of variants in 8 ARS genes implicated in recessive disease[177]. It is important to

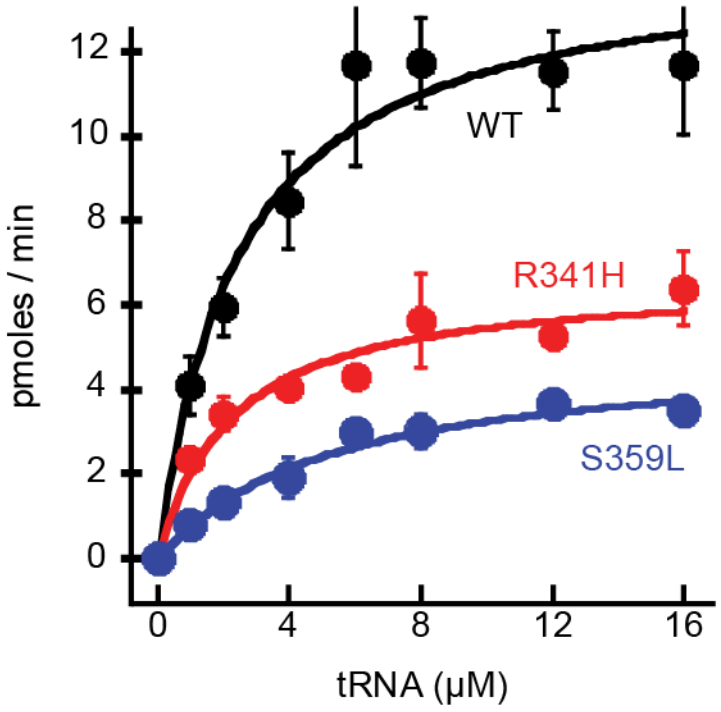


Figure 3.17. R341H and S349L CARS1 have reduced activity in in vitro aminoacylation assays.

Human recombinant wild-type and R341H and S359L CARS1 were assessed for tRNA charging activity *in vitro*. Initial aminoacylation rates (pmole/min) of wild-type (black circles and line), R341H (red circles and lines), and S359L (blue circles and line) CARS1 were plotted against tRNA concentrations and fit to the Michaelis–Menten equation. Error bars indicate standard deviations. These experiments were performed by Thomas Christian and Ya-Ming Hou at Thomas Jefferson University.

Table 3.2. Kinetic properties of mutant CARS1 proteins

Variant	k_{cat} (sec^{-1})	K_m (μM)	Activity compared to wild-type*
Wild-type	0.80	2.4	1.0
R341H	0.37	2.2	0.5
S359L	0.27	5.1	0.16

*The catalytic efficiency (k_{cat}/K_m) is reported.

note that both of these assays test one allele at a time. However, testing aminoacylation activity using fibroblast cell lysates from affected subjects is a technique that determines the effects of ARS mutations in the context of the affected subject genome; these studies are critical for efforts to better classify the effect of disease-associated ARS genotypes on tRNA charging. The tRNA charging assays do not evaluate ARS activity in living cells, which is an important limitation. Yeast complementation assays address this limitation by assessing the effects of ARS variants *in vivo*; however, yeast is a single-celled organism and does not represent multi-cellular tissues in a human. In order to address the limitations of any one assay, we employed multiple approaches, which strengthens our interpretation that each *CARS1* variant causes a complete or partial loss-of-function effect.

Our functional studies on the disease-associated *CARS1* variants also provide insight into the pathogenic mechanism of the clinical phenotype observed in our subjects. That is, the loss-of-function effects suggest that impaired tRNA charging leads to reduced, interrupted, or otherwise altered translation in the affected organ systems. This view is consistent with previous reports of loss-of-function effects of variants in other ARS genes identified in subjects with recessive phenotypes[29]. An alternative—albeit less likely—possibility is that CARS1 has an additional non-canonical function unrelated to tRNA charging that is also affected by the variants described here, and that this function is particularly important for the affected tissues. While CARS1 has been reported to function in cysteine polysulfidation[269] and ferroptosis[270] and may have a role in immune control through an immune-activating domain[271], further work is needed to identify CARS1 secondary functions that might be relevant to the phenotypes in our subjects. Importantly, phenotypic overlap between our subjects and others with ARS-mediated recessive

disease further supports the pathogenicity of the identified *CARS1* variants and indicates a common mechanism for ARS-related recessive disease. Indeed, the combination of clinical features in our subjects is compatible with what has previously been reported in other ARS-related autosomal recessive phenotypes. For example, developmental delay is common to all four subjects with *CARS1* variants, and is also seen in subjects with variants in *DARS1* (MIM 603084), *DARS2* (MIM 610956), *FARS2* (MIM 611592), *KARS1* (MIM 601421), *NARS2* (MIM 612803), *PARS2* (MIM 612036), *QARS1* (MIM 603727), *RARS2* (MIM 611524), *VARS1* (MIM 192150), and *YARS1* (MIM 603623)[29]. Similarly, microcephaly is common to Subject 1-3 and Subject 2-4, and is also seen in subjects with variants in *AARS1* (MIM 601065), *GARS1* (MIM 600287), *IARS1* (MIM 600709), *KARS1*, *NARS2*, *PARS2*, *QARS1*, *RARS2*, *SARS1* (MIM 607529), *VARS1*, *VARS2* (MIM 612802), and *WARS2* (MIM 604733)[29,86]. Finally, liver dysfunction is common to Subject 1-3, Subject 3-3, and Subject 3-4, and is also seen in subjects with variants in *EARS1* (MIML 612799), *FARSB* (MIM: 609690), *FARS2*, *IARS1*, *LARS1* (MIM: 141350), *LARS2* (MIM: 604544), *MARS1* (MIM 156560), and *YARS1*[29,68,74,75,114,224].

In addition to the above phenotypes, Subject 1-3 has an axonal peripheral neuropathy, which has previously been associated with variants in five other ARS loci; variants in *GARS1*, *YARS1*, *AARS1*, *HARS1*, and *WARS1* have been implicated in dominant peripheral neuropathies[29]. The majority of these variants show complete or partial loss-of-function effects in functional assays, suggesting that reduced enzyme function is a common feature of dominant neuropathy-associated ARS variants and may contribute to disease pathogenesis (see Chapter 1)[29,177]. Since reduced function is also a common feature of ARS-associated recessive disease[29], there may be phenotypic overlap with ARS-associated dominant neuropathies. Therefore, severe depletion of

tRNA charging may be responsible for the peripheral neuropathy in this patient, who is homozygous for a partial loss-of-function variant in *CARS1*.

Interestingly, all four affected individuals described here have some level of fine, brittle hair with microscopic shaft abnormalities. At the time of publication, this phenotype had not previously been associated with any disease-causing ARS variants, suggesting a potential *CARS1*-specific phenotype. Hair and nails may be particularly susceptible to impairments in CARS1 function due to the high cysteine content of keratins expressed in these tissues. Recently, brittle hair has also been associated with variants in *TARS1*[153], raising the possibility that brittle hair might be associated with more genes but is more severe and prominent in patients with *CARS1* mutations. Additionally, CARS1 deficiency renders individuals susceptible to other characteristic phenotypes including central nervous system impairment and liver steatosis. As described above, these features have also been identified in other ARS-associated disease, suggesting the possibility of a global effect on translation that occurs when any ARS is impaired. Further studies in appropriate animal models are required to confirm our observations and to determine if the translation of cysteine-rich proteins is affected by impaired CARS1 activity (see Chapters 4 and 5).

In conclusion, we present clinical, genetic, and functional data that implicate *CARS1* variants in a multi-system, recessive disease including microcephaly, developmental delay, and brittle hair and nails. These findings expand the locus, allelic, and clinical heterogeneity of ARS-related disease, and further support the model that impaired tRNA charging followed by translation defects is the primary pathogenic mechanism of ARS-related recessive disease.

Chapter 4

Investigating the Downstream Consequences of *CARS1* Variants on Translation

4.1 Introduction

Variants in 19 of the 20 aminoacyl-tRNA synthetase (ARS) genes encoding cytoplasmic enzymes have been implicated in multi-system recessive diseases (see Chapters 1-3)[29]. Most patients present with similar clinical phenotypes including developmental delay and microcephaly[29]; however, some tissues appear to be specifically sensitive to defects in a particular ARS gene. For example, only *CARS1* and *TARS1* variants are known to cause a highly penetrant brittle hair phenotype[53,153]. Data from genetic and functional studies suggest a partial-loss-of-function molecular pathology for ARS-associated recessive disease. First, the genotypes of patients suggest a partial-loss-of-function mechanism; patients are either compound heterozygous for one missense mutation and one null allele, compound heterozygous for two missense mutations, or homozygous for a single missense mutation[3,29,30,153,195]. Second, functional studies indicate that recessive disease-associated ARS mutations cause reduced ARS protein levels and/or decreased enzyme activity[29,177]. Though there is abundant evidence that a partial-loss-of-function molecular pathology is common to recessive disease-associated ARS variants, the downstream effects on cellular, tissue, and organism function are poorly defined. Furthermore, it is currently unclear why some tissues are more sensitive to impairments of particular ARSs.

As described in Chapter 3, cysteinyl-tRNA synthetase 1 (CARS1) charges tRNA with cysteine in the cytoplasm, and data from functional assays indicate that disease-associated *CARS1* variants cause loss-of-function or partial loss-of-function effects[53]. Importantly, the genotype of each patient indicates that some CARS1 activity is retained, which is consistent with the idea that a complete loss of CARS1 function would be lethal. Since our results suggest that CARS1 function is reduced, charged $tRNA^{Cys}$ abundance is likely reduced, which may lead to defects in translation. We hypothesize that the translation of cysteine-rich proteins may be particularly affected by impaired CARS1 function, which then contributes to disease pathogenesis. This is consistent with the brittle hair and nail phenotype of patients with *CARS1* variants since hair and nails are abundant in cysteine-rich keratins and keratin-associated proteins.

Impaired CARS1 activity could also lead to a general or global effect on translation, which may be mediated through the integrated stress response. General control non-derepressible-2 (GCN2) recognizes uncharged tRNA and phosphorylates translation initiation factor eIF2α, which then represses translation globally (see Chapter 1)[1]. Amino acid deprivation, including cysteine deprivation, triggers this response[1,272]. Since *CARS1* variants lead to decreased aminoacylation activity, impaired tRNA charging may result in increased uncharged tRNA that GCN2 can recognize and bind. A recent publication supports the hypothesis that loss-of-function and partial-loss-of-function disease-associated ARS variants can result in global effects on translation through the integrated stress response; Mullen et al. reported that the overexpression of histidyl-tRNA synthetase 1 (*HARS1*) dominant disease-associated variants in a neurite outgrowth model in rat pheochromocytoma cells (PC12) led to the attenuation of translation and increased phosphorylation of eIF2α[273].

In this chapter, we describe studies investigating the effects of *CARS1* variants on translation in cell, yeast, and mouse models. The author of this thesis performed all of the studies in this chapter with the following exceptions: Sanger sequencing, library preparation, and RNA sequencing reactions were performed by the University of Michigan Advanced Genomics Core; the indexed human reference genome (GRCh38) used to map RNA sequencing reads was generated by Dr. Jacob Kitzman (University of Michigan); Matthew Pun and Jonathon Kuo assisted with the computational analysis of the human proteome (Figures 4.13B, 4.14B, and 4.15B); the yeast dual luciferase reporter construct was obtained and made Gateway compatible by Rebecca Meyer-Schuman (Antonellis lab; Figure 4.16); the S688Qfs*2 *Cars1* mouse model design and transgenesis was performed by the Transgenic Animal Model Core of the University of Michigan's Biomedical Research Core Facilities, Wanda Filipiak, and Galina Gavrilina; complete blood count, full chemistry panel, and necropsies were performed by Dr. Robert Sigler and the University of Michigan Unit for Laboratory Animal Medicine (ULAM) In-Vivo Animal Core (Tables 4.3, 4.4, 4.5, and 4.6); and quantitative proteomics was performed by the University of Michigan Proteomics and Peptide Synthesis Core (Figure 4.27 and Tables 4.7 and 4.8). The S688Qfs*2 *Cars1* mouse was generated in collaboration with Dr. Miriam Meisler's lab at the University of Michigan; Dr. Miriam Meisler, Dr. Young Park, Dr. Guy Lenk, Sydney Musser, and Pooja Varanasi were instrumental for performing guidance on design, husbandry, and experimentation.

4.2 Methods

4.2.1 Fibroblast cells and cell culture

Primary fibroblasts from Subject 1-3 (S688Qfs*2/S688Qfs*2 *CARS1*), Unaffected 1-2 (S688Qfs*2/+ *CARS1*), Subject 2-4 (R341H/Q380* *CARS1*), and Subject 3-3 (S359L/L400Q *CARS1*) were obtained by collaborating physicians. Control primary fibroblasts (presumably bearing no pathogenic *CARS1* variants) were obtained from ATCC (PCS-201-012). Cells were maintained at 37°C and 5% CO2 in complete medium (Dulbecco's modified Eagle medium [ThermoFisher Scientific]) with 10% fetal bovine serum (ThermoFisher Scientific), 1X L-glutamine (Corning), 1X penicillin-streptomycin (final concentration 50 units/ml penicillin and 50 ug/mL streptomycin; ThermoFisher Scientific), unless otherwise specified. For cysteine and methionine deprivation experiments, cells were expanded in complete media and then cultured for 3 h in medium lacking cysteine and methionine (ThermoFisher Scientific)[274]. For experiments with only cysteine deprivation, medium lacking cysteine and methionine was supplemented with methionine (30 mg/L). For dithiothreitol (DTT) experiments, cells were cultured for 30 min or 1 h in medium with 2 mM DTT (Fisher Scientific)[275]. Cells were cultured in T-75cm^3-treated cell culture flasks (CytoOne) or 6-well plates (Fisher Scientific).

4.2.2 Puromycin incorporation assays

To investigate fibroblast puromycin incorporation, surface sensing of translation (SUnSET) assays were performed as previously described[276]. Control and patient cells were plated in 6-well plates (200,000 cells/well). The following day, cells were incubated with 10 μg/ml puromycin (Sigma-Aldrich) for 10 min at 37°C. For dithiothreitol (DTT) experiments, cells were incubated with 2 mM DTT (ThermoFisher Scientific) for 30 minutes or 1 hour prior to puromycin

treatment according to previously published methods[275]. For cycloheximide experiments, cells were incubated with 50 μg/ml cycloheximide (Sigma-Aldrich) for 30 minutes prior to puromycin treatment.

4.2.3 Fibroblast cell protein isolation and western blot analyses

Patient fibroblasts were harvested at approximately 90% confluency with 0.25% trypsin (ThermoFisher Scientific), quenched with complete cell medium, and collected at 2,000 rpm for 2 minutes. For experiments to calculate the amount of protein isolated per cell, cell number was measured using trypan blue (Invitrogen) and the Countess automated cell counter (Invitrogen) prior to collection. The resulting pellet was resuspended in cell lysis buffer (RIPA buffer [ThermoFisher Scientific] + 1X EDTA-free protease inhibitor [ThermoFisher Scientific] + 1X phosphatase inhibitor [ThermoFisher Scientific]) and incubated at 4°C for 30 min with gentle rocking. Cell debris was collected at 4°C and 13,000 rpm for at least 10 minutes, and total protein concentration was measured using the Thermo Scientific Pierce BCA Protein Assay Kit. To calculate the amount of protein isolated per cell, the total concentration was divided by the number of cells collected.

Protein samples (10 μg) were prepared with 1X SDS-sample buffer (ThermoFisher Scientific) and 2-mercaptoethanol (βME), denatured at 99°C for 5 min, and separated on a precast 4-20% tris-glycine gel (ThermoFisher Scientific) at 150 V for 75 minutes at room temperature. A polyvinylidene difluoride membrane (MilliporeSigma, Burlington, MA) was prewashed in 100% methanol and soaked in 1X transfer buffer (ThermoFisher Scientific) + 10% methanol with blot paper (Bio-Rad). Protein was transferred to the membrane overnight using a Mini-Trans-Blot

Electrophoretic Transfer Cell (Bio-Rad) for 1 hour at 100 V at room temperature and then incubated for 1 hour in blocking solution (5% dry milk powder + 1X TBST [tris-buffered saline solution and tween 20, pH 7.5]) at room temperature. For eIF2α phosphorylation experiments, the membrane was incubated with anti-phospho-eIF2α (Ser51; Cell Signaling Technology; 3398; 1:1,000) or anti-eIF2α (Cell Signaling Technology; 5324; 1:25,000) at 4°C overnight. For puromycin incorporation experiments, anti-puromycin (Sigma 12D10; MABE343; 1:1,000) and anti-actin (Sigma; A5060; 1:10,000) antibodies were used. The membrane was then washed three times for 5 minutes in 1X TBST and incubated in blocking solution with IRDye 680RD goat anti-rabbit IgG secondary antibody (LI-COR; 925-68071) and/or IRDye 800CW goat anti-mouse IgG secondary antibody (LI-COR; 925-32210) at room temperature for 1 hour. The membrane was washed three times for 5 minutes in 1X TBST and imaged using a LI-COR Biosciences Odyssey CLx imager. Band or lane intensity was measured using ImageJ. Each experiment was performed three times. Statistical significance was calculated using one-way ANOVA (GraphPad Prism).

4.2.4 RNA isolation and RNA sequencing analysis

Fibroblast cells from patients (Subject 1-3 [S688Qfs*2/S688Qfs*2 *CARS1*], Subject 2-4 [R341H/Q380* *CARS1*], and Subject 3-3 [S359L/L400Q *CARS1*]) and unaffected individuals (control [primary dermal fibroblasts from normal human; PCS-201-012 ATCC] and Unaffected 1-2 [S688Qfs*2/+ *CARS1*]) were grown in standard conditions and harvested at approximately 90% confluency with 0.25% trypsin (ThermoFisher Scientific), quenched with complete cell medium, and collected at 2,000 rpm for 2 minutes. Total RNA was then isolated using the RNeasy Mini Kit (QIAGEN) following manufacturer's instructions from at least three flasks per

cell line. RNA concentration and purity were evaluated using a NanoDrop Lite (Thermo Fisher Scientific), and the two samples per cell line with the highest 260/230 ratios were submitted to the UM Advanced Genomics core for stranded RNA library preparation, which included cDNA synthesis from polyadenylated mRNA and subsequent attachment of Illumina adapters to the cDNA. cDNA libraries were then sequenced on the Illumina HiSeq-4000 (paired-end 150 cycles). Quality of the sequencing data was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to the human reference genome (GRCh38) using STAR[277], and SAM files were converted to BAM files using SAMtools[278]. Reads were counted using featureCounts[279], and differentially expressed genes were determined (FDR < 0.05) using edgeR[280]. Mean-difference plots were generated (edgeR).

To determine whether transcripts with high cysteine codon content were downregulated in patient cells compared to controls, the cysteine content was determined for the proteins in the human proteome, which was obtained from UniProt (Proteome ID UP000005640; accessed November 3, 2019). Proteins from the Swiss-Prot UniProt database were used and uncharacterized proteins were removed. Biostrings in R was used to determine the number of each amino acid in each protein. For each protein, the percent cysteine was calculated by dividing the number of cysteines by the total number of amino acids. The proportion of cysteine of proteins encoded by the downregulated, upregulated, or unchanged protein-coding genes was then determined and plotted (GraphPad Prism). Statistical significance was determined using one-way ANOVA (GraphPad Prism).

Genes that were downregulated, upregulated, or unchanged were analyzed using geneontology.org[204,205] to identify biological processes that were significantly enriched in each group. PANTHER Overrepresentation Test (Released 20210224)[206] using the GO Ontology database (Released 2021-02-01) was used with Fisher's Exact test and Bonferroni correction for multiple testing. In Figures 4.13-4.15, the top 12 biological processes (or all of the processes if there were fewer than 12) with a p-value < 0.05 are listed and ranked by fold enrichment (graphed using GraphPad Prism).

4.2.5 Generation of yeast dual luciferase reporter constructs

We obtained a yeast plasmid from Dr. Kristin Koutmou (Department of Chemistry, University of Michigan), pR1F1, which contains the renilla and firefly luciferase genes under the control of the alcohol dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase promoters, respectively. Rebecca Meyer-Schuman (Antonellis lab) made the region of the plasmid containing the renilla and firefly luciferase genes Gateway-compatible and cloned it into pDONR. N-terminal cysteine-rich coding sequences (or a non-cysteine containing sequence as a control) were then cloned in frame with the renilla coding sequence using restriction enzyme cloning (*Sac*I and *Not*I; New England BioLabs). The human metallothionein 1B (*MT1B*; NM_005947.3; 34% cysteine), yeast copper metallothionein (*CUP1*; NM_001179185.1; 20% cysteine), and yeast translation machinery-associated 7 (*TMA7*; NM_001184339.1; 0% cysteine) open reading frames were obtained as gBlocks (IDT) with *Sac*I and *Not*I restriction enzyme digest sites added to the 5' and 3' ends, respectively. Resulting constructs were screened for the appropriately sized insert by digesting with *Sac*I and *Not*I. Plasmids were then submitted for Sanger sequencing (UM Advanced Genomics Core) to confirm the inserted gBlock sequence in frame with the renilla

coding sequence. The sequences were then LR-cloned into p413, a yeast expression construct carrying the *HIS3* gene, and verified by restriction enzyme digests.

4.2.6 Dual luciferase reporter assay in yeast

To investigate the effects of *CARS1* variants on translating cysteine-rich sequences in a yeast model, we used the haploid yeast strain with the endogenous *CRS1* locus deleted and viability maintained via a pRS316 vector bearing wild-type (WT) *CRS1* (see Chapter 3). Yeast cells were transformed with empty R1F1/p413, TMA7 R1F1/p413, CUP1 R1F1/p413, or MT1B R1F1/p413. Growth on solid medium lacking histidine and uracil (Takara) selected for transformed cells. Yeast cells were then transformed with WT CRS1/pRS315 or D714* CRS1/pRS315. Growth on solid medium lacking histidine, leucine, and uracil (Takara) was used to select for transformed cells. Colonies were grown for 2 days at 30°C to saturation in liquid medium, and then 40 µl was spiked into 0.1% 5-FOA liquid medium lacking histidine and leucine (Takara) to select for cells that spontaneously lost the *URA3*-bearing maintenance vector[246]. After 20 hours, 10 µl of each sample was lysed in 50 µl passive lysis buffer (Promega) in a white polystyrene round bottom plate (Corning) for 1 hour shaking at room temperature. Renilla and firefly luciferase activities were then determined using the Dual Luciferase Reporter 1000 Assay System (Promega) and a Glomax Multi-Detection System (Promega). Each reaction was performed at least 8 times. The ratio of renilla to firefly activity was calculated. Statistical significance was calculated using one-way ANOVA or multiple t-tests, as indicated in figure legend.

*4.2.7 Generation of the S688Qfs*2* Cars1 *mouse model*

To model the S688Qfs*2 *CARS1* patient variant in mouse, we collaborated with Dr. Miriam Meisler, Dr. Young Park, the Transgenic Animal Model Core (TAMC) of the University of Michigan Biomedical Research Core Facilities, Wanda Filipiak, and Galina Gavrilina to edit the mouse genome to harbor the S771Qfs*2 *Cars1* variant (ENSMUSE00000206281), which is the equivalent of human S688Qfs*2 *CARS1*; to avoid confusion, we will refer to this mutation as S688Qfs*2 *Cars1*. The TAMC designed an sgRNA to target exon 22 for CRISPR/Cas9-mediated mutagenesis. To assess the ability of the sgRNA to direct Cas9 digestion of the target DNA sequence, the sgRNA (Synthego; 30 ng/ul) and Cas9 protein (Sigma; 50 ng/ul) were microinjected into mouse zygotes, and genomic DNA was prepared from the resulting blastocysts. A DNA fragment spanning the expected Cas9 cut site was PCR amplified and analyzed with Sanger sequencing. Five of the six test blastocysts showed evidence of an active sgRNA, so we proceeded with this sgRNA. The CRISPR/Cas9 reagents and a DNA donor that introduces S688Qfs*2 upon homology-directed repair were microinjected into fertilized eggs obtained by mating (C57BL/6J X SJL)F1 or C57BL/6J female mice with (C57BL/6J X SJL)F1 male mice. Pronuclear microinjection was performed as described[281]. Microinjected eggs were then surgically transferred to recipients. Tail biopsies from potentially transgenic mice were obtained 5 weeks after injecting eggs (three weeks gestation time and two weeks of post-natal growth) and assessed for the presence of the S688Qfs*2 *Cars1* mutation using a restriction enzyme digest genotyping strategy (see Chapter 4.2.8) followed by Sanger sequencing (UM Advanced Genomics Core). The genotyping PCR products were also utilized for TOPO TA cloning (Invitrogen) followed by Sanger sequencing (UM Advanced Genomics Core) to confirm the presence of the desired allele. Three mice (ear tags 627, 672, and 702) showed evidence of

the S688Qfs*2 *Cars1* mutation and were subsequently mated with C57BL/6J mice to test transmission of the allele; we designated these mice as the N1 generation. Offspring from all three matings showed evidence of transmission of the allele; two lines (627 [A] and 672 [B]) were maintained via mating with C57BL/6J. Heterozygous mice were intercrossed to generate homozygous mice; when N1 heterozygous mice were intercrossed, we designated those offspring N1F1.

*4.2.8 S688Qfs*2* Cars1 *genotyping strategy*

Tail biopsies were lysed in 250 µl (or 125 µl for smaller tail biopsies) 50 mM NaOH for at least 15 min at 95°C and then neutralized with 50 µl (or 25 µl for smaller tail biopsies) 1M Tris-HCl pH 6.8. After 30 s centrifugation at 15,000 rpm, 1 µl of the lysis reaction was used for PCR amplification using PCR SuperMix (Invitrogen). PCR amplification was confirmed by subjecting 5 µl to gel electrophoresis. The products were column-purified and then digested with *Dra*III-HF (NEB) overnight at 37°C. The entire digest reaction was subjected to gel electrophoresis and analysis. The wild-type PCR product contains one *Dra*III cut site, and the expected band sizes are 142 bp and 193 bp. The *Dra*III cute site is ablated by the S688Qfs*2 *Cars1* mutation, and the expected band size for PCR products containing the mutation is 335 bp. The genotypes of mice selected for mating were confirmed using Sanger sequencing (UM Advanced Genomics Core).

4.2.9 Mouse dissections

The author assisted Dr. Guy Lenk (Meisler lab, University of Michigan), who performed dissections on three wild-type and three homozygous 3-month-old male mutant mice (S688Qfs*2/S688Qfs*2 *Cars*; N1F1). One wild-type and one homozygous mutant mouse were

from line A, and the other four mice were from line B. Total body weight was measured, and mice were observed walking in a serological pipet holder for any overt behavioral phenotypes. Mice were euthanized with CO_2, and the paw was pinched to ensure no reflex response was present before proceeding. Mice were immediately perfused by injecting 1X PBS into the left side of the heart. Liver and brain were dissected, and total organ weight was measured. Half of each tissue was stored immediately on dry ice for submission to the University of Michigan Proteomics and Peptide Synthesis Core for quantitative proteomics (see Chapter 4.2.14). The other half of each tissue was homogenized in 500 µl 1X PBS using a Dounce homogenizer. Half (250 µl) was used for protein isolation and homogenized with 450 µl RIPA buffer (ThermoFisher Scientific) with 1X EDTA-free protease inhibitor (ThermoFisher Scientific) and 1X Halt phosphatase inhibitor (ThermoFisher Scientific). The samples were then centrifuged at 4,500 x g for 5 min at 4°C.

4.2.10 Mouse western blot analyses

Total protein concentration was measured using the Thermo Scientific Pierce BCA Protein Assay Kit, and 25 µg protein/sample was analyzed. Protein samples were prepared with 1X SDS-sample buffer (ThermoFisher Scientific) and 2-mercaptoethanol (βME), denatured at 99°C for 5 min, and separated on a precast 4-20% tris-glycine gel (ThermoFisher Scientific) at 150 V for 75 minutes at room temperature. A polyvinylidene difluoride membrane (MilliporeSigma, Burlington, MA) was prewashed in 100% methanol and soaked in 1X transfer buffer (ThermoFisher Scientific) + 10% methanol with blot paper (Bio-Rad). Protein was transferred to the membrane overnight using a Mini-Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) for 1 hour at 100 V at room temperature and then incubated for 1 hour in blocking solution (5% dry

milk powder + 1X TBST [tris-buffered saline solution and tween 20, pH 7.5]) at room temperature. The membrane was incubated with anti-CARS1 (Sigma; 1:1,000) or anti-actin (MilliporeSigma; 1:5,000) at 4°C overnight. The membrane was then washed three times for 5 minutes in 1X TBST and incubated in blocking solution with IRDye 680RD goat anti-rabbit IgG secondary antibody (LI-COR; 925-68071) and/or IRDye 800CW goat anti-mouse IgG secondary antibody (LI-COR; 925-32210) at room temperature for 1 hour. The membrane was washed three times for 5 minutes in 1X TBST and imaged using a LI-COR Biosciences Odyssey CLx imager. Band or lane intensity was measured using ImageJ. Western blots were repeated three times each, and statistical significance was determined using one-way ANOVA (GraphPad Prism).

4.2.11 Mouse RT-PCR analyses

Wild-type mouse brain cDNA (postnatal day 3) was obtained from Elizabeth Fogarty (Antonellis lab). To determine whether *Cars1* transcripts containing exon 2 are expressed in mouse brain, primers were designed to amplify a sequence surrounding exon 2 (mmCars Exon 3F and mmCars Exon 1R; see Appendix) The expected PCR product sizes are a 382-bp fragment from transcripts containing exon 2 and a 133-bp fragment from transcripts without exon 2. For each reaction, 22 μl of PCR Supermix (Life Technologies) was combined with 0.5 μl of each 20 μM primer solution, 0.75 μl of DMSO, and 1 μl of cDNA. Standard PCR conditions were used. RT-PCR products were separated by gel electrophoresis 60 min 135 V on a 1% TBE agarose gel.

4.2.12 Mouse Mendelian ratios and body weight measurements

Mice were sexed and genotyped at weaning (around 3 weeks old), and Mendelian ratios were calculated for all N1F1 mice (n = 151). The chi-square test was performed using the GraphPad calculator (https://www.graphpad.com/quickcalcs/chisquared1.cfm). Mouse total body weight was measured using a plastic beaker and scale. For N1F1 mice, weight at 10 weeks old was recorded (n = 76). Statistical significance was determined using a Student's t-test (GraphPad Prism).

4.2.13 Mouse phenotyping necropsies

Two wild-type and two homozygous mutant animals were submitted to the ULAM In-Vivo Animal Core and phenotypically examined by gross and histopathology. All animals were born on October 4, 2019, and necropsy was performed November 19, 2019. At necropsy, body weight, tibial length, and liver weight was collected. Tissues were embedded and slides were sectioned at 4 microns. Tissues collected included cervical skin, brain, spinal cord, liver, kidneys, spleen, heart, skeletal muscle, lung, adrenal, pancreas, stomach, small intestine, colon, cecum, eyes, femur, forepaw, sternum and testes. Slides were stained with hematoxylin and eosin. Blood was obtained for complete blood count and full chemistry panel.

4.2.14 Quantitative proteomics

Liver and brain were dissected from three wild-type and three homozygous 3-month-old male mutant mice (S688Qfs*2/S688Qfs*2 *Cars*; N1F1) as described above (see Chapter 4.2.9). The total organ weights were measured, and half of each tissue was stored immediately on dry ice for submission to the University of Michigan Proteomics and Peptide Synthesis Core for quantitative

proteomics. Tissues were then extracted using modified RIPA buffer (2% SDS, 150 mM NaCl, 50 mM Tris pH 8, 1X Roche cOmplete protease inhibitor). Extracts were concentrated by trichloroacetic acid according to previously published methods[282]. Protein pellets were washed with ice cold acetone and solubilized in urea buffer (8M urea, 150 mM NaCl, 50 mM Tris-HCl pH 8, 1X Roche cOmplete protease inhibitor). The protein concentration of the processed material was determined by fluorometry on a Qubit (Invitrogen). Protein from each sample (50 µg) was digested by: (1) reduction with 14mM dithiothreitol at 25°C for 30 minutes followed by alkylation with 14 mM iodoacetamide at 25°C for 45 minutes in the dark; (2) digestion with 25 µg sequencing grade trypsin (Promega) at 37°C overnight; (3) termination with 5 µL of formic acid, centrifugation at 10,000 g for 10 minutes, and desalting using an Empore SD solid phase extraction plate; and (4) lyophilization and reconstitution in 0.1% TFA for analysis.

Each digested sample (3 µg) was analyzed by nano LC-MS/MS (4 h LC gradient) with a Waters M-Class LC system interfaced to a ThermoFisher Fusion Lumos mass spectrometer. Peptides were loaded on a trapping column and eluted over a 75 µm analytical column at 350 nL/min; both columns were packed with XSelect CSH C18 resin (Waters); the trapping column contained a 3.5µm particle, the analytical column contained a 2.4 µm particle. The column was heated to 55°C using a column heater (Sonation). The mass spectrometer was operated in data-dependent mode, with the Orbitrap operating at 15,000 FWHM for MS/MS. The instrument was run with a 3 second cycle for MS/MS.

To determine protein identities, the data were searched using a local copy of Mascot (Matrix Science). Mascot DAT files were parsed into Scaffold (Proteome Software) for validation and

filtering and to create a non-redundant list per sample. Data were filtered using a 1% protein and peptide false discovery rate (FDR) and requiring at least two unique peptides per protein. The total spectrum count was used to calculate the normalized spectral abundance factor (NSAF)[283] for each protein in each sample to assess protein up- and down-regulation (fold change) of identical proteins among the different preparations and samples. Proteins with fold change (mutant/wild-type) > 2 or < 0.5 and p-value < 0.05 were determined to be significantly differentially expressed.

4.3 Results

4.3.1 Puromycin incorporation and phosphorylation of eIF2α in patient fibroblasts are not significantly different from controls in standard growth conditions

Disease-associated *CARS1* variants may affect global translation, and this may be mediated through the integrated stress response, a mechanism of translational control involving the phosphorylation of eukaryotic translation initiation factor eIF2α, which results in inhibition of protein synthesis[1]. One method of activating the integrated stress response is through general control nonderepressible 2 (GCN2), a eukaryotic translation initiation factor 2 subunit 1 (eIF2α) kinase that phosphorylates eIF2α upon binding to uncharged tRNAs[1]. We hypothesized that impaired CARS1 activity results in increased uncharged $tRNA^{Cys}$, which can then activate GCN2 and lead to inhibition of translation globally. To determine if global translation is impaired in the context of pathogenic *CARS1* variants, we measured puromycin incorporation using surface sensing of translation (SUnSET[276]) in patient and control fibroblast cells. We incubated the cells with puromycin, isolated whole-cell protein extracts, and performed western blot analyses using an antibody to puromycin. No difference in puromycin incorporation was observed in patient

cells compared to controls (n = 3; Figure 4.1). As a control to test that the puromycin signal is reflecting the rate of translation, we treated cells with cycloheximide, a translation inhibitor that interferes with elongation, and observed an ablation of the puromycin signal on western blot analysis (Figure 4.2). Additionally, there was no significant difference in the total amount of protein isolated per cell from patient cells compared to controls (Figure 4.3). Furthermore, we assessed eIF2α phosphorylation via western blot analysis using an antibody that recognizes the phosphorylation of serine 51 of eIF2α; we also observed no significant difference in eIF2α phosphorylation in patient cells compared to controls (n = 3; Figure 4.4). These results suggest that patient fibroblasts—under these specific experimental conditions—are not activating the integrated stress response, and that translation is not affected globally.

4.3.2 No differences in puromycin incorporation or eIF2α phosphorylation are shared by all patient cells after DTT treatment

Since patient fibroblasts in standard culture conditions did not show any differences in global translation compared to controls, we tested if stress may exacerbate any effects on translation in patient cells that were undetectable in standard conditions. We stressed the cells by treating with dithiothreitol (DTT), a strong reducing agent that prevents disulfide bond formation[284]. DTT treatment leads to the accumulation of misfolded proteins and the unfolded protein response, which is another trigger of eIF2α phosphorylation and the integrated stress response[284]. We incubated patient and control cells with 2 mM DTT for 30 or 60 minutes[275] prior to puromycin treatment. After 30 minutes of DTT treatment, Unaffected 1-2, Subject 2-4, and Subject 3-3 had significantly reduced puromycin incorporation compared to control cells, and

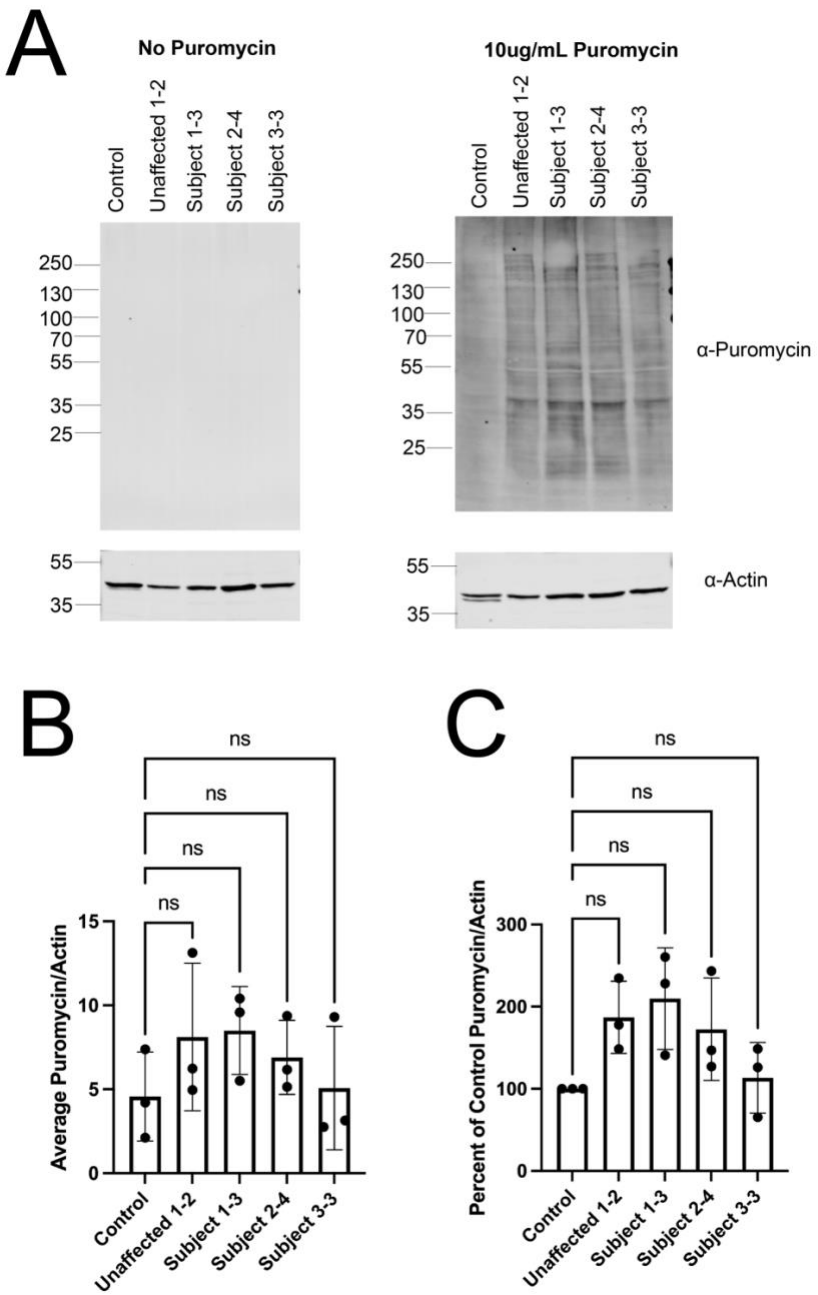


Figure 4.1. Patient fibroblast cells show similar levels of puromycin incorporation compared to control fibroblasts, indicating similar levels of global translation.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-puromycin antibody or anti-actin antibody was used (indicated on the right). Treatment and sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of puromycin signal was quantified using ImageJ and normalized to actin expression. (C) Normalized puromycin signal in patient cells is shown as a percent of the normalized puromycin signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the puromycin/actin from each individual experiment, and bars represent the average (n = 3). Error bars represent the standard deviation.

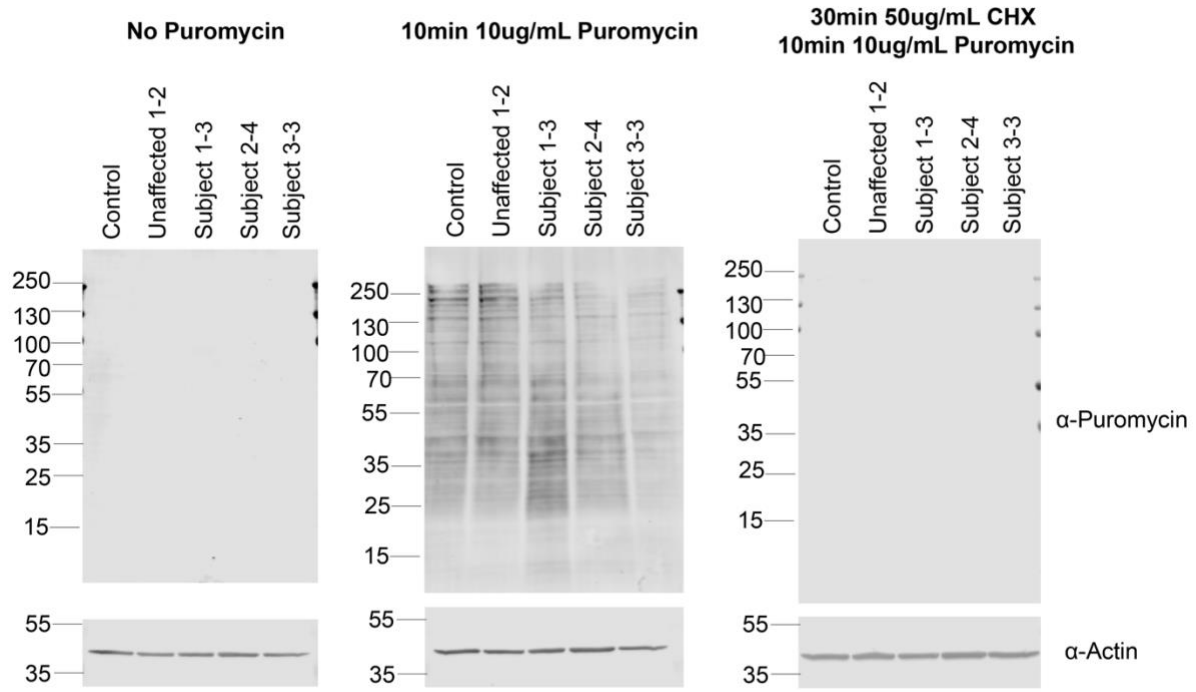


Figure 4.2. Cycloheximide treatment ablates puromycin incorporation.

Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-puromycin or anti-actin antibody was used (indicated on the right). Treatment and sample names are across the top, and sizes (kDa) are indicated at the left.

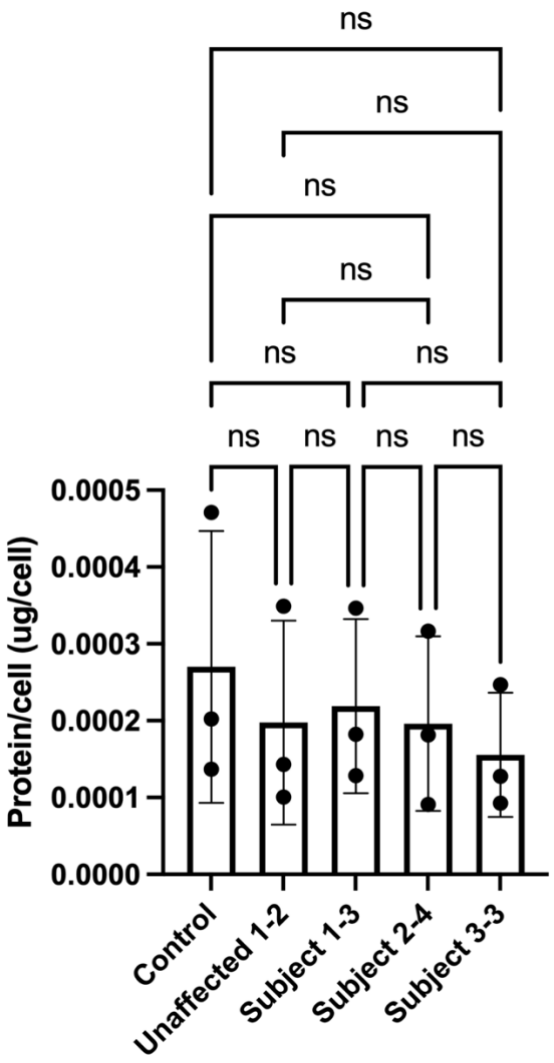


Figure 4.3. Patient fibroblasts show similar amounts of protein per cell compared to control fibroblasts.

Total protein from each cell line was isolated and quantified and then divided by the number of cells. Statistical significance was calculated using one-way ANOVA. The dots represent the protein/cell from each individual experiment, and bar heights represent the average (n = 3). Error bars represent the standard deviation.

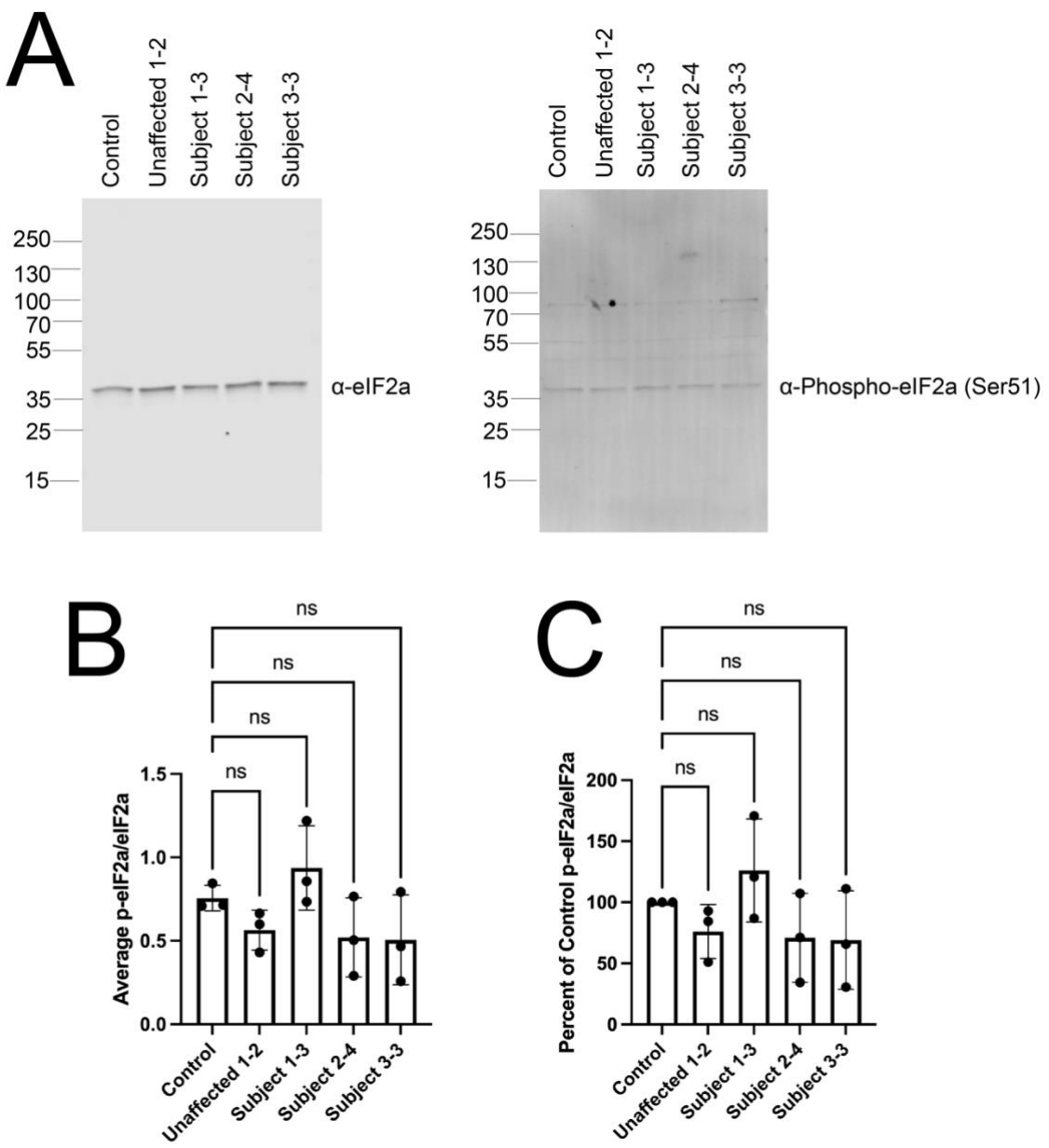


Figure 4.4. Patient fibroblast cells show similar levels of eIF2α phosphorylation compared to control cells.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-phospho-eIF2α or anti-eIF2α antibody was used (indicated on the right). Sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of phosphorylated eIF2α signal was quantified using ImageJ and normalized to total eIF2α expression. (C) Normalized p-eIF2α signal in patient cells is shown as a percent of the normalized p-eIF2α signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the p-eIF2α/eIF2α of each individual experiment, and the bars represent the average (n = 3). Error bars represent the standard deviation.

Subject 1-3 did not have significantly different puromycin incorporation compared to control (Figure 4.5). Phosphorylation of eIF2α was not significantly different between any patient and control cells except for Subject 3-3, which had reduced phosphorylation of eIF2α (Figure 4.6), suggesting that the differences in translation may not be mediated through the integrated stress response. Cells were also incubated with DTT for 60 minutes, and we observed no difference in puromycin incorporation (Figure 4.7) or phosphorylation of eIF2α (Figure 4.8) between patient and control cells. In sum, no differences in puromycin incorporation were shared by all patient cells after DTT treatment. Additionally, DTT treatment did not result in an increase in the phosphorylation of eIF2α in patient cells compared to controls, suggesting that the integrated stress response is not responsible for the decreased puromycin incorporation observed and/or that this response is not activated in the presence of *CARS1* mutations in fibroblast cells.

4.3.3 Puromycin incorporation in patient cells is not significantly decreased compared to control cells after cysteine and methionine deprivation

In the above experiments, cells were cultured in complete media and supplied with ample nutrients including amino acids. We investigated whether depleting the media of cysteine may exacerbate effects on translation in the patient cells since cysteine is a component of the aminoacylation reaction. We cultured patient and control cells with medium containing all amino acids except cysteine for 3 h prior to puromycin treatment and observed no difference in puromycin incorporation between patient and control cells (Figure 4.9). Phosphorylation of eIF2α was also not significantly different between patient and control cells cultured without cysteine except for Subject 3-3, which had reduced phosphorylation of eIF2α (Figure 4.10). Since cysteine can be generated from methionine[285], we also cultured fibroblasts with medium

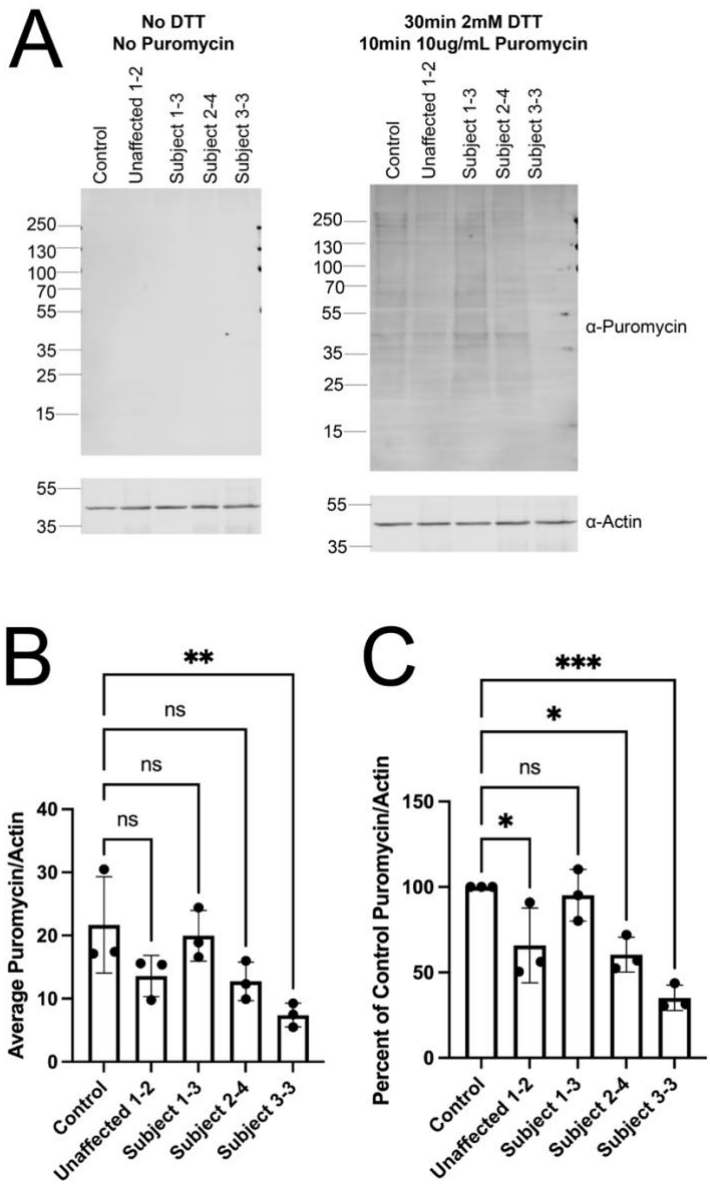


Figure 4.5. Unaffected 1-2, Subject 2-4, and Subject 3-3 fibroblasts show reduced levels of puromycin incorporation compared to control fibroblasts after 30 min DTT treatment.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-puromycin antibody or anti-actin antibody was used (indicated on the right). Treatment and sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of puromycin signal was quantified using ImageJ and normalized to actin expression. (C) Normalized puromycin signal in patient cells is shown as a percent of the normalized puromycin signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the puromycin/actin from each individual experiment, and bars represent the average (n = 3). Error bars represent the standard deviation. * p- value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

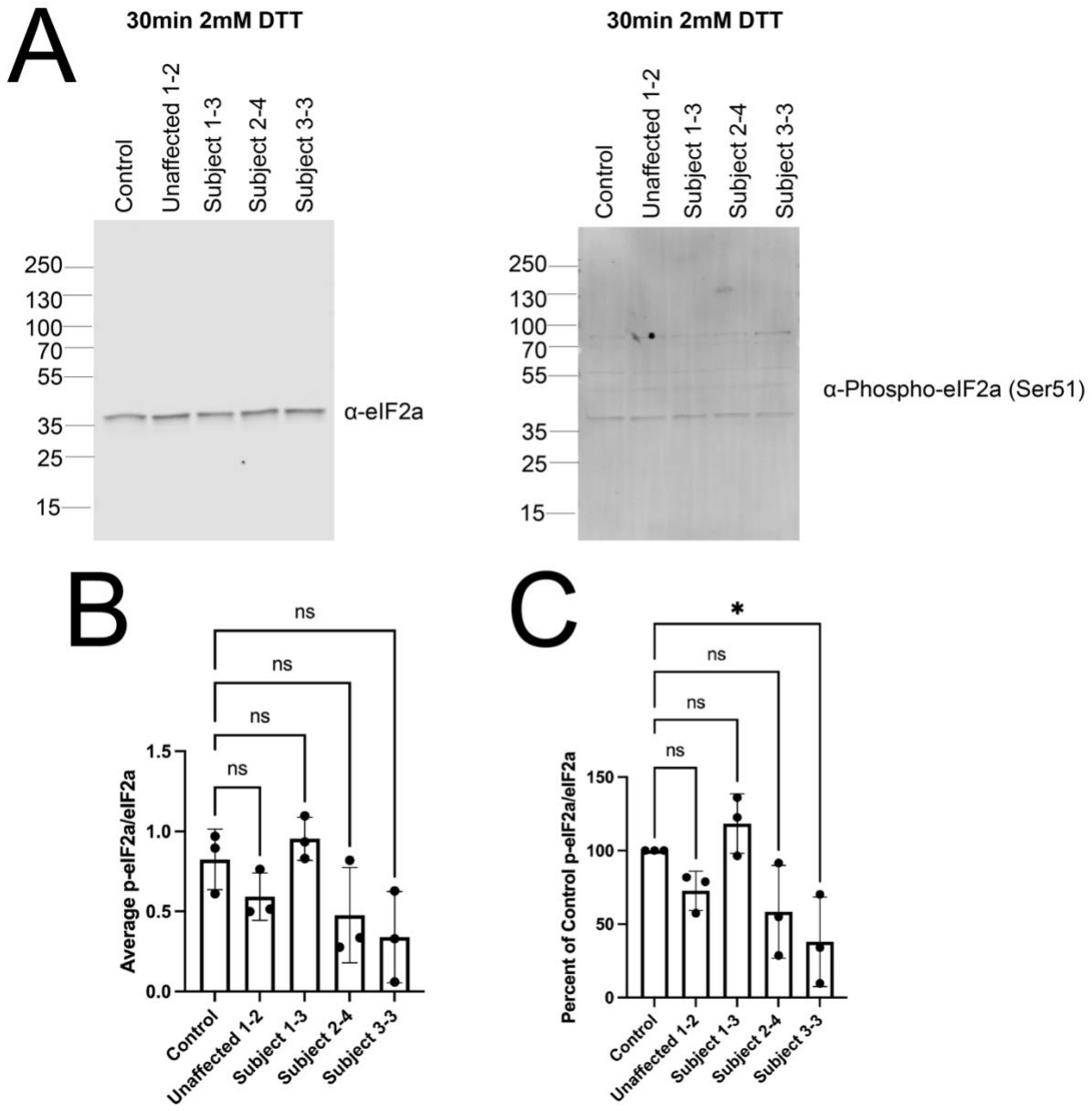


Figure 4.6. Patient fibroblast cells show similar levels of eIF2a phosphorylation compared to control cells after 30 min DTT treatment, except Subject 3-3.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-phospho-eIF2α or anti-eIF2α antibody was used (indicated on the right). Sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of phosphorylated eIF2α signal was quantified using ImageJ and normalized to total eIF2α expression. (C) Normalized p-eIF2α signal in patient cells is shown as a percent of the normalized p-eIF2α signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the p-eIF2α/eIF2α of each individual experiment, and the bars represent the average (n = 3). Error bars represent the standard deviation. * p-value < 0.05.

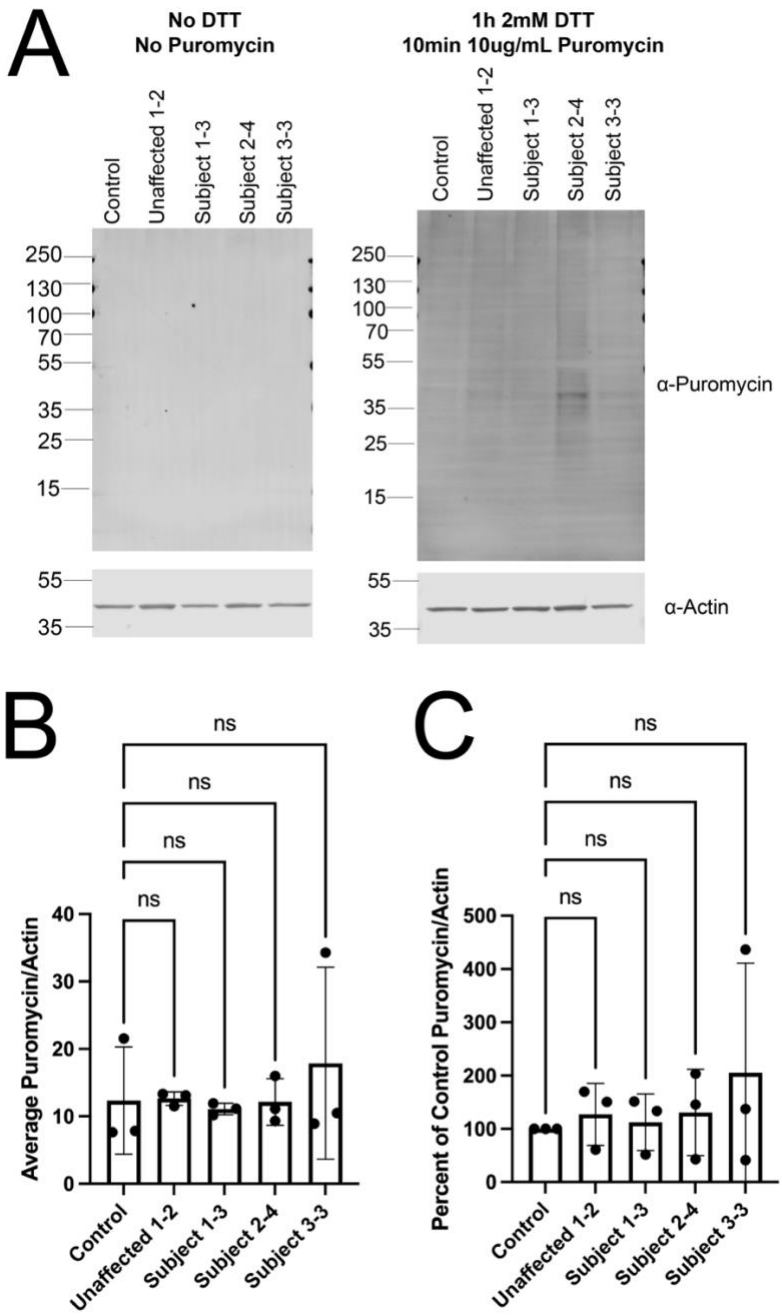


Figure 4.7. Patient fibroblast cells show similar levels of puromycin incorporation compared to control fibroblasts after 1h DTT treatment.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-puromycin antibody or anti-actin antibody was used (indicated on the right). Treatment and sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of puromycin signal was quantified using ImageJ and normalized to actin expression. (C) Normalized puromycin signal in patient cells is shown as a percent of the normalized puromycin signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the puromycin/actin from each individual experiment, and bars represent the average (n = 3). Error bars represent the standard deviation.

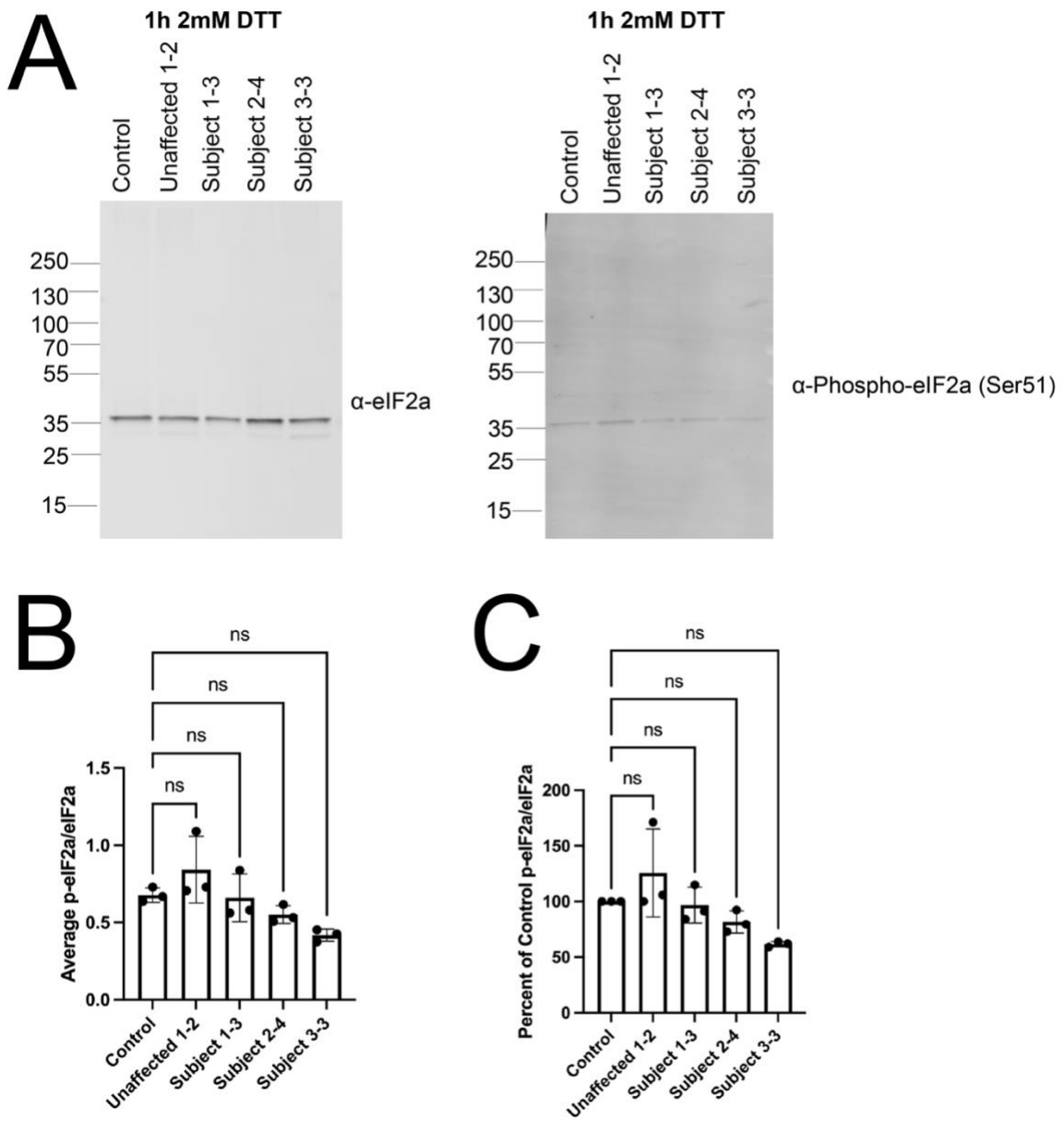


Figure 4.8. Patient fibroblast cells show similar levels of eIF2a phosphorylation compared to control cells after 1 h DTT treatment.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-phospho-eIF2α or anti-eIF2α antibody was used (indicated on the right). Sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of phosphorylated eIF2α signal was quantified using ImageJ and normalized to total eIF2α expression. (C) Normalized p-eIF2α signal in patient cells is shown as a percent of the normalized p-eIF2α signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the p-eIF2α/eIF2α of each individual experiment, and the bars represent the average (n = 3). Error bars represent the standard deviation.

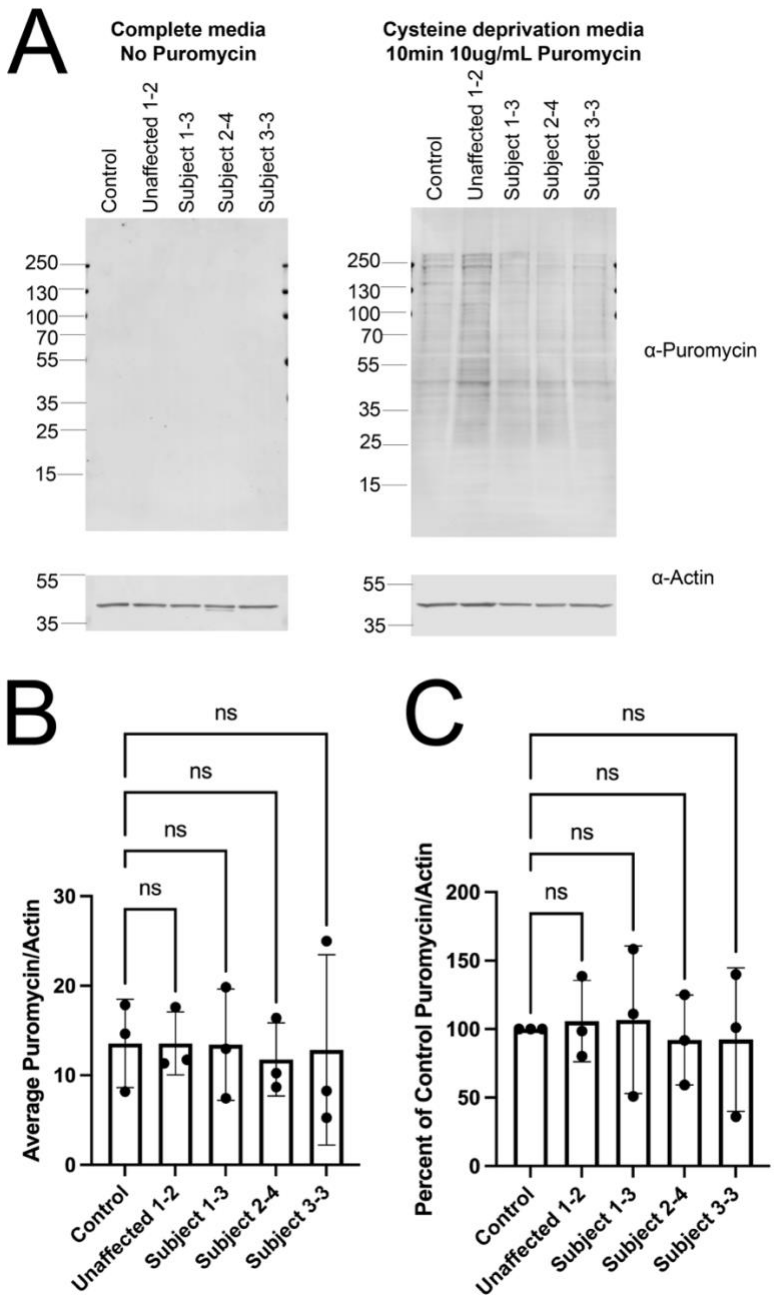


Figure 4.9. Patient fibroblast cells show similar levels of puromycin incorporation compared to control fibroblasts in media lacking cysteine.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-puromycin antibody or anti-actin antibody was used (indicated on the right). Treatment and sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of puromycin signal was quantified using ImageJ and normalized to actin expression. (C) Normalized puromycin signal in patient cells is shown as a percent of the normalized puromycin signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the puromycin/actin from each individual experiment, and bars represent the average (n = 3). Error bars represent the standard deviation.

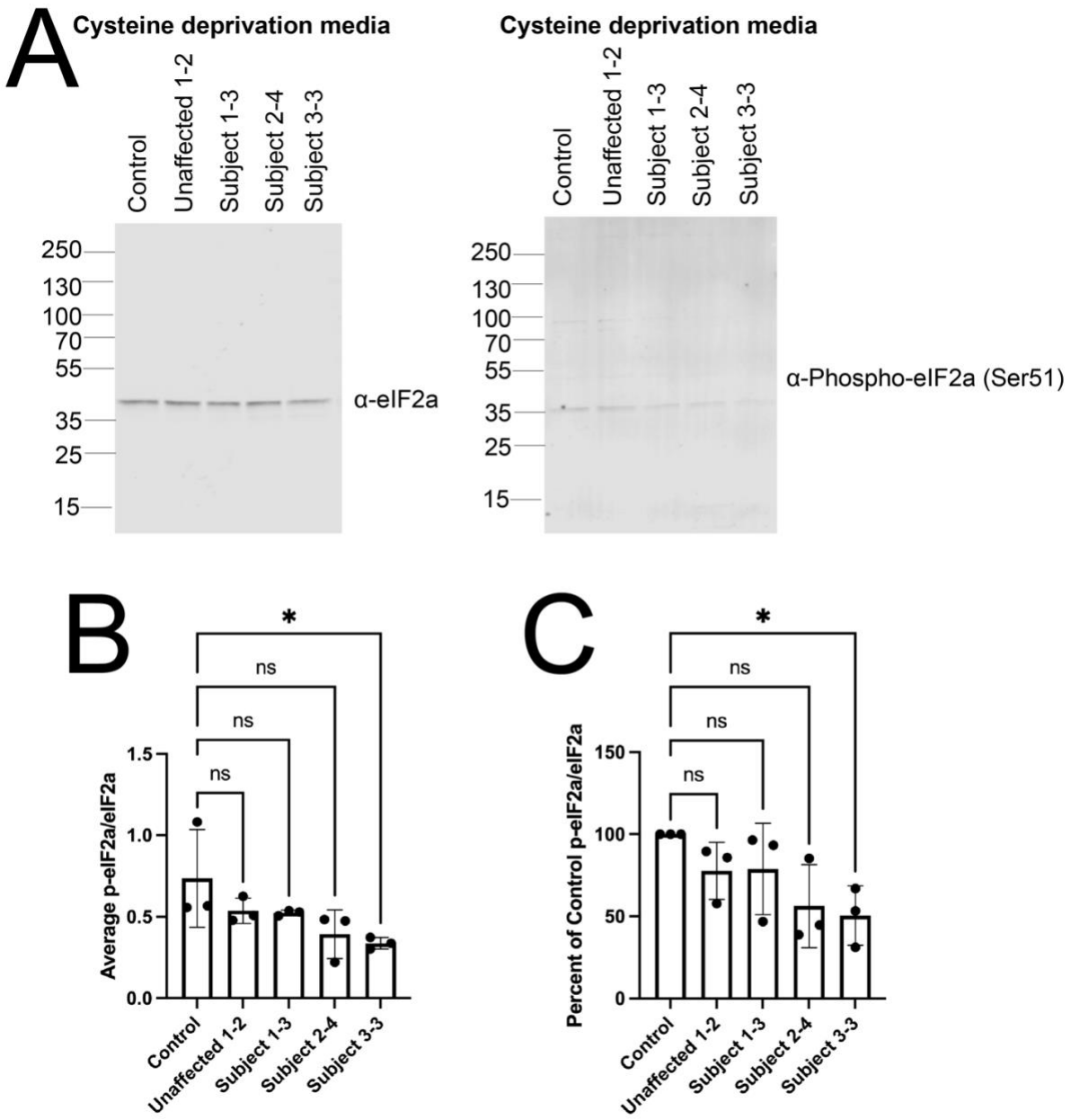


Figure 4.10. Patient fibroblast cells show similar levels of eIF2a phosphorylation compared to control cells in media lacking cysteine, except Subject 3-3.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-phospho-eIF2α or anti-eIF2α antibody was used (indicated on the right). Sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of phosphorylated eIF2α signal was quantified using ImageJ and normalized to total eIF2α expression. (C) Normalized p-eIF2α signal in patient cells is shown as a percent of the normalized p-eIF2α signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the p-eIF2α/eIF2α of each individual experiment, and the bars represent the average (n = 3). Error bars represent the standard deviation.

containing all amino acids except both cysteine and methionine. After subsequent puromycin treatment, surprisingly, the raw intensity of puromycin signal normalized to actin in Unaffected 1-2 and Subject 1-3 cells was increased compared to control cells (Figure 4.11B); however, when values were normalized to the control lane on each blot, the differences are no longer statistically significant (Figure 4.11C). Phosphorylation of eIF2α was significantly reduced in Subjects 1-3 and 3-3 after cysteine and methionine deprivation compared to control fibroblasts (Figure 4.12). In sum, our results suggest that patient fibroblasts are not experiencing significant reductions in global translation or increased activation of the integrated stress response compared to control cells in response to cysteine and methionine deprivation.

4.3.4 Both downregulated and upregulated genes have higher cysteine codon content than unchanged genes in patient cells compared to controls

Since disease-associated *CARS1* variants cause complete or partial loss-of-function effects (see Chapter 3), tRNA charging is likely impaired. A lack of available charged $tRNA^{Cys}$ may then lead to ribosome stalling at cysteine codons and the subsequent degradation of those transcripts, which we predict may occur more frequently with high cysteine-codon-containing transcripts. To investigate this possibility, we isolated total RNA in duplicate from fibroblast cells from patients (Subject 1-3 [S688Qfs*2/S688Qfs*2 *CARS1*], Subject 2-4 [R341H/Q380* *CARS1*], and Subject 3-3 [S359L/L400Q *CARS1*]) and unaffected individuals (control [primary dermal fibroblasts from normal human; no *CARS1* mutations; PCS-201-012 ATCC] and Subject 1-2 [S688Qfs*2/+ *CARS1*]). The University of Michigan Advanced Genomics Core performed stranded library preparation and paired-end RNA sequencing (Illumina HiSeq-4000). After quality control, mapping reads, and counting reads, we identified 358 differentially expressed genes between

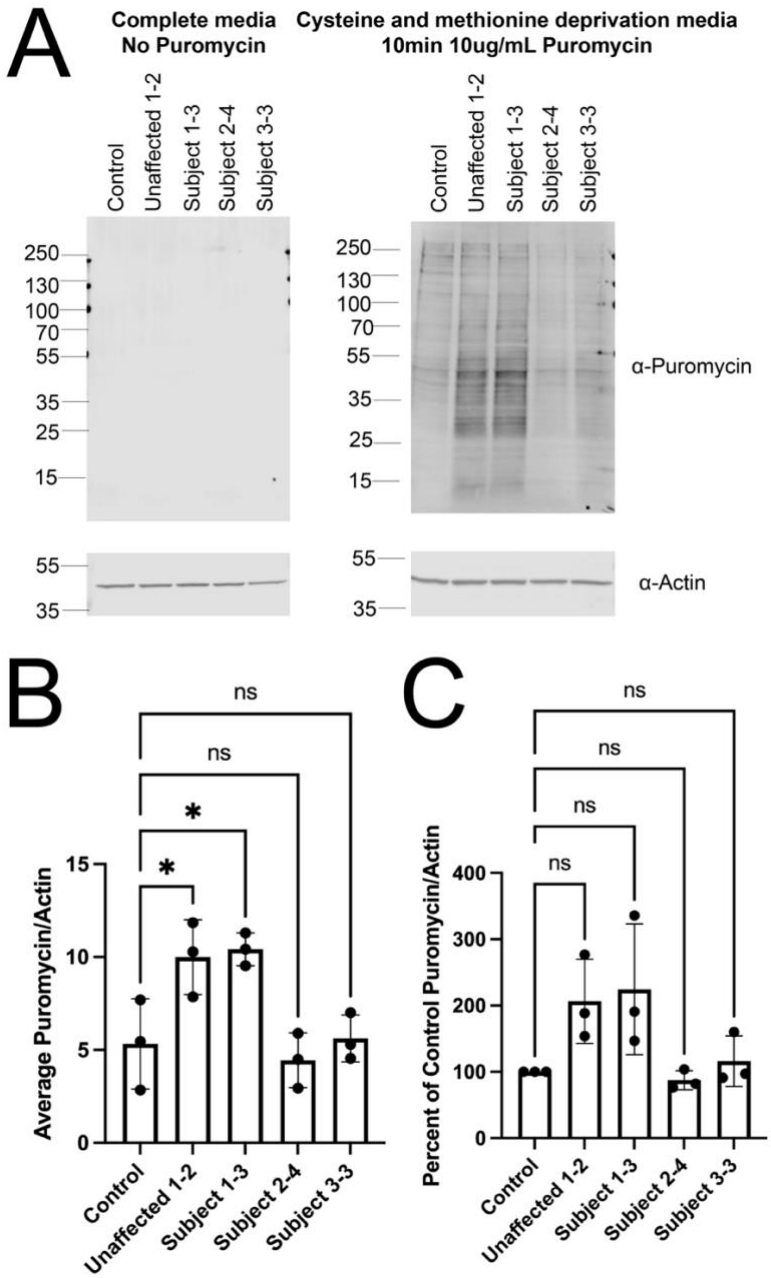


Figure 4.11. Patient fibroblast cells show similar levels of puromycin incorporation compared to control fibroblasts in media lacking cysteine and methionine.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-puromycin antibody or anti-actin antibody was used (indicated on the right). Treatment and sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of puromycin signal was quantified using ImageJ and normalized to actin expression. (C) Normalized puromycin signal in patient cells is shown as a percent of the normalized puromycin signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the puromycin/actin from each individual experiment, and bars represent the average (n = 3). Error bars represent the standard deviation. * p-value < 0.05.

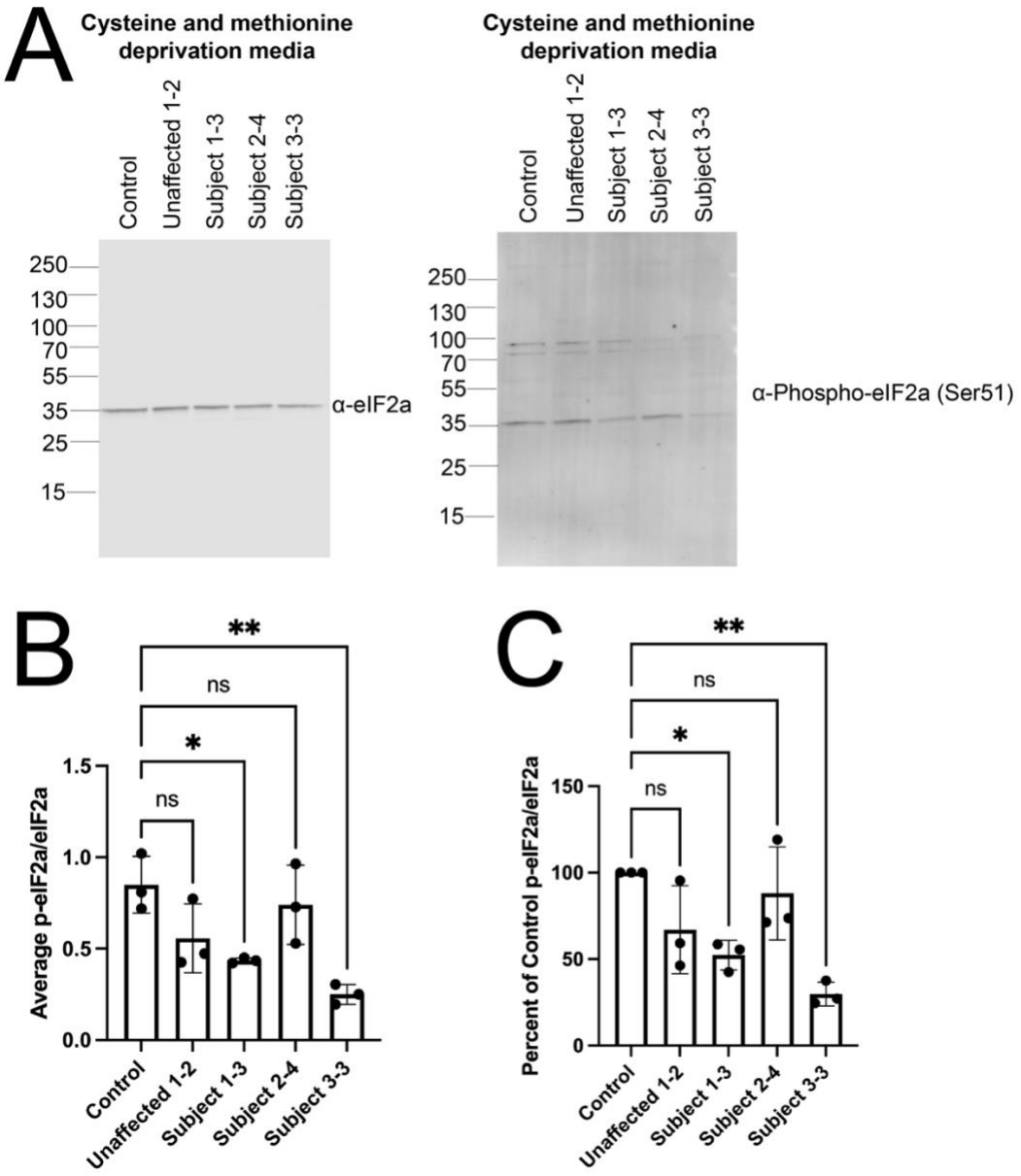


Figure 4.12. Subject 1-3 and Subject 3-3 show reduced levels of eIF2a phosphorylation compared to control cells in media lacking cysteine and methionine.

(A) Western blot analyses were performed using total protein lysates isolated from fibroblasts, and an anti-phospho-eIF2α or anti-eIF2α antibody was used (indicated on the right). Sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of phosphorylated eIF2α signal was quantified using ImageJ and normalized to total eIF2α expression. (C) Normalized p-eIF2α signal in patient cells is shown as a percent of the normalized p-eIF2α signal of the control cells on the same blot. Statistical significance in (B) and (C) was calculated using one-way ANOVA. The dots represent the p-eIF2α/eIF2α of each individual experiment, and the bars represent the average (n = 3). Error bars represent the standard deviation. * p- value < 0.05; ** p-value < 0.01.

patient (Subjects 1-3, 2-4, and 3-3) and control cells (presumably bearing no pathogenic *CARS1* mutations); in patient cells, 188 genes were upregulated, and 170 genes were downregulated (Figure 4.13A). To assess whether downregulated genes were cysteine-rich, we determined the proportion of cysteine in the protein resulting from each protein-coding gene (Figure 4.13B). We observed an increased average cysteine proportion in the downregulated genes; the average cysteine proportion (0.03) was significantly higher than the cysteine proportion among the genes that were not differentially expressed (no change; 0.02); however, the upregulated genes also had a higher average cysteine proportion (0.03) compared to the unchanged genes (Figure 4.13B). Since the difference in cysteine proportion between downregulated and unchanged (and upregulated and unchanged) are small (1%), the biological significance of the increased cysteine proportion is unclear. Gene ontology analysis of downregulated genes showed an enrichment for 'positive regulation of multicellular organismal process' (GO:0051240; p-value = 1.75e-2; Figure 4.13C). Upregulated genes showed an enrichment for 'regulation of smooth muscle cell migration' (GO:0014910; p-value = 4.45e-2; Figure 4.13D). More work is needed to determine if and how these processes are related to disease pathogenesis.

Since Unaffected 1-2 is heterozygous for S688Qfs*2 *CARS1* and is the unaffected parent of Subject 1-3 (who is homozygous for S688Qfs*2/S688Qfs*2 *CARS1*), we also compared the gene expression of patient cells to a control group including both control cells (presumably bearing no pathogenic *CARS1* mutations) and cells from Unaffected 1-2 (S688Qfs*2/+ *CARS1)*. We identified 400 differentially expressed genes between these two groups; 256 were upregulated, and 144 were downregulated in the patient cells (Figure 4.14A). Similar to the analysis above comparing patients to only the control cells, we observed an increased average cysteine

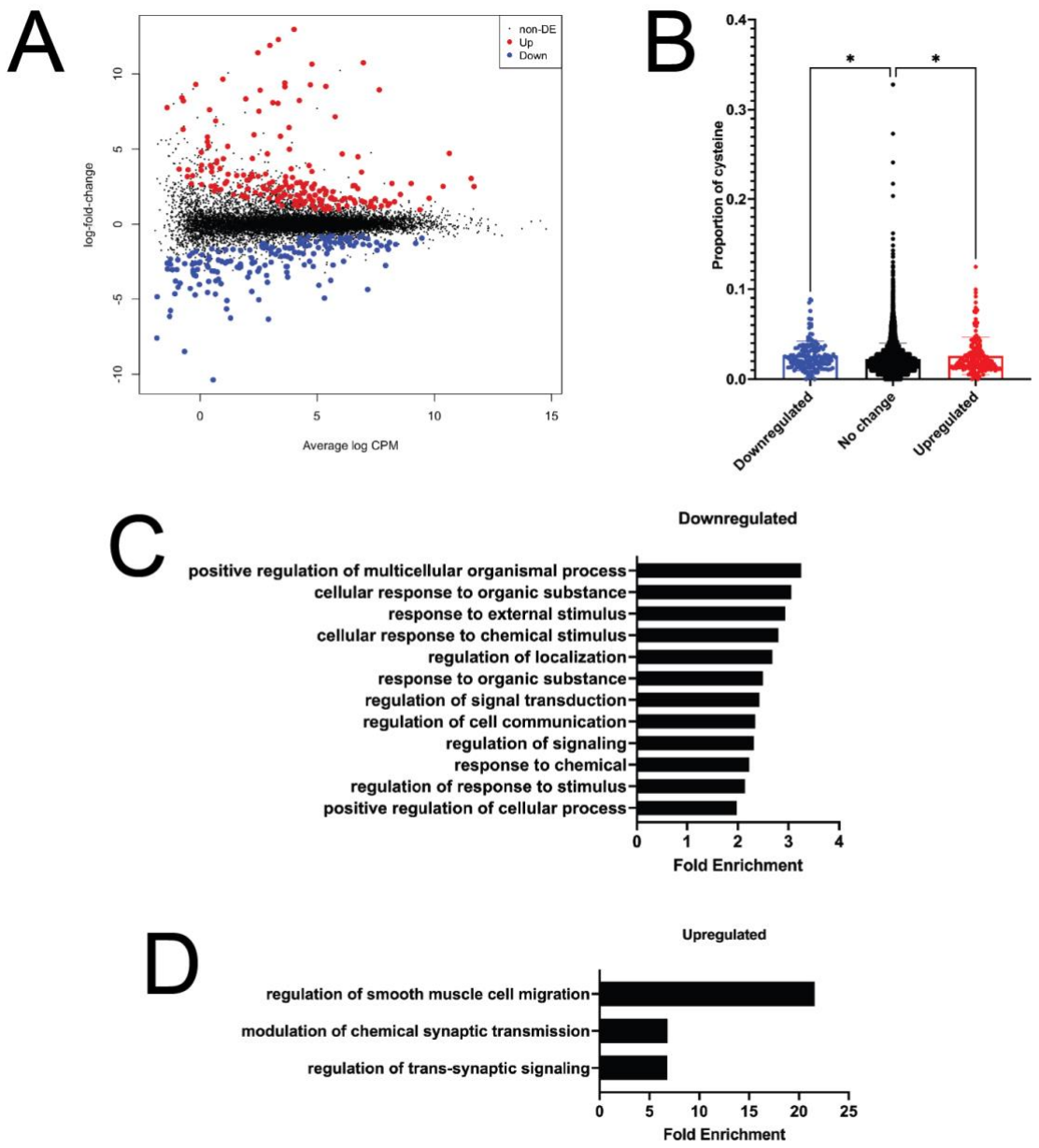


Figure 4.13. Both downregulated and upregulated genes have higher cysteine content than unchanged genes in patient cells compared to control cells.

(A) Mean-difference plot depicting whether genes are upregulated (red dots), downregulated (blue dots), and non-differentially-expressed (black dots) in patient cells compared to control. Log_2(fold change) is on the y-axis, and the average log_2(counts per million) is on the x-axis. (B) Genes were grouped based on differential expression (downregulated, unchanged, or upregulated), and the proportion of cysteine for each gene was determined and plotted. Bar heights represent mean, and error bars represent standard deviation. Statistical significance was calculated using one-way ANOVA. * p-value < 0.05. Gene ontology analysis on downregulated (C) and upregulated (D) genes identified biological processes that are significantly enriched. The biological processes with a p-value < 0.05 are listed and ranked by fold enrichment.

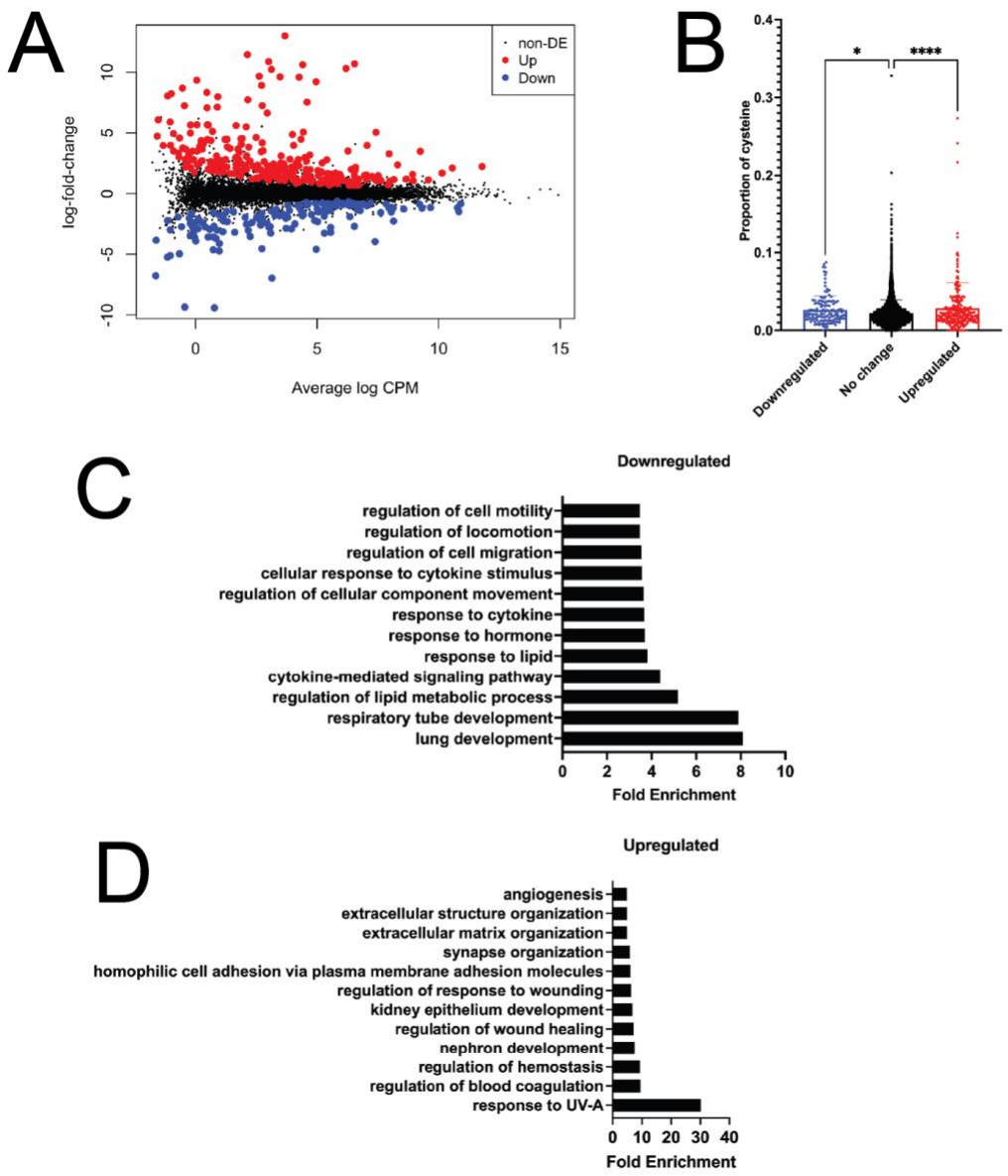


Figure 4.14. Both downregulated and upregulated genes have higher cysteine content than unchanged genes in patient cells compared to fibroblasts from unaffected individuals.

(A) Mean-difference plot depicting whether genes are upregulated (red dots), downregulated (blue dots), and non-differentially-expressed (black dots) in patient cells compared to control and Unaffected 1-2 cells. Log_2(fold change) is on the y-axis, and the average log_2(counts per million) is on the x-axis. (B) Genes were grouped based on differential expression (downregulated, unchanged, or upregulated), and the proportion of cysteine for each gene was determined and plotted. Bar heights represent mean, and error bars represent standard deviation. Statistical significance was calculated using one-way ANOVA. * p-value < 0.05. Gene ontology analysis on downregulated (C) and upregulated (D) genes identified biological processes that are significantly enriched. The top 12 biological processes with a p-value < 0.05 are listed and ranked by fold enrichment.

proportion in downregulated genes (0.03) compared to unchanged genes (0.02), but the upregulated genes also had increased average cysteine proportion (0.03) compared to unchanged genes (Figure 4.14B). Gene ontology analysis of downregulated genes showed an enrichment for 'lung development' (GO:0030324; p-value = 2.41e-2; Figure 4.14C), and upregulated genes showed an enrichment for 'response to UV-A' (GO:0070141; p-value = 1.94e-2; Figure 4.14D). Again, the disease relevance of these results and the associated processes is unclear.

Finally, we compared the gene expression of cells from Subject 1-3 (S688Qfs*2/S688Qfs*2 *CARS1*) to cells from his parent, Unaffected 1-2 (S688Qfs*2/+ *CARS1*). We identified 173 differentially expressed genes; 86 were upregulated, and 87 were downregulated (Figure 4.15A). There was no statistically significant difference in cysteine proportion between upregulated and unchanged genes or between downregulated genes and unchanged genes (Figure 4.15B). Gene ontology analysis of downregulated genes showed the greatest enrichment for 'collagen fibril organization' (GO:0030199; p-value = 2.08e-3; Figure 4.15C), which may be relevant to the patient clinical presentation since the patient has connective tissue-related phenotypes including poor wound healing and recurrent hernias. Gene ontology analysis of upregulated genes did not reveal enrichment for any biological processes. In summary, our analyses indicate that cysteine-codon-rich transcripts are not preferentially downregulated in patient cells in standard growth conditions.

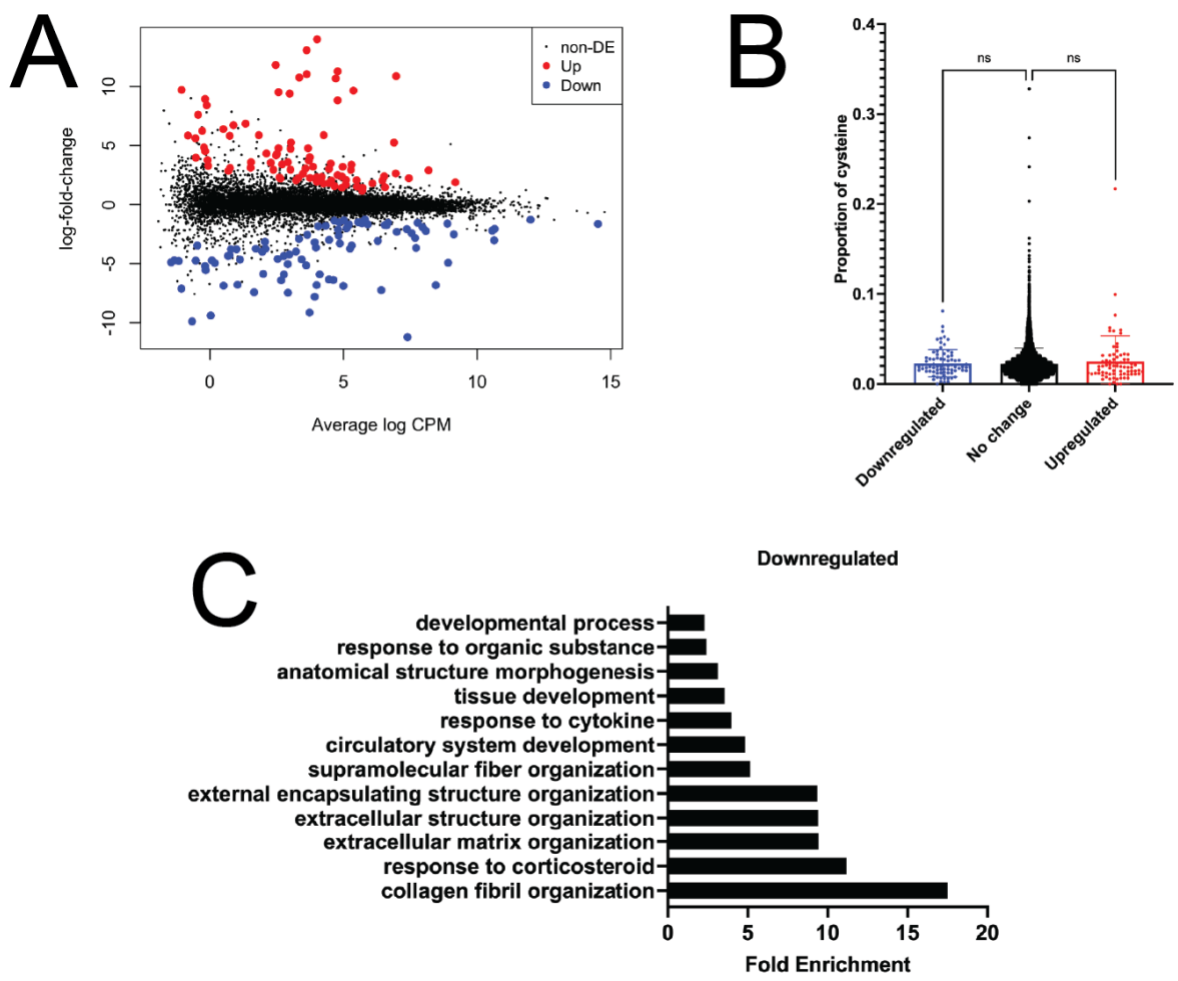


Figure 4.15. Transcripts with high cysteine codon content are not downregulated in cells from Subject 1-3 compared to fibroblasts from his parent, Unaffected 1-2.

(A) Mean-difference plot depicting whether genes are upregulated (red dots), downregulated (blue dots), and non-differentially-expressed (black dots) in Subject 1-3 cells compared to Unaffected 1-2 cells. Log_2(fold change) is on the y-axis, and the average log_2(counts per million) is on the x-axis. (B) Genes were grouped based on differential expression (downregulated, unchanged, or upregulated), and the proportion of cysteine for each gene was determined and plotted. Bar heights represent mean, and error bars represent standard deviation. Statistical significance was calculated using one-way ANOVA. * p-value < 0.05. (C) Gene ontology analysis on downregulated genes identified biological processes that are significantly enriched. The top 12 biological processes with a p-value < 0.05 are listed and ranked by fold enrichment.

4.3.5 Yeast expressing mutant CRS1 *show reduced activity of renilla fused to cysteine-rich protein sequences*

To test if disease-associated *CARS1* variants impair translation of cysteine-rich proteins in our yeast model, we developed a dual luciferase reporter assay in yeast. In this assay, an N-terminal cysteine-rich coding sequence was cloned in frame with the renilla coding sequence on a plasmid that contains the firefly luciferase gene as a control (Figure 4.16). For the N-terminal sequences, we tested two proteins of the same size (both are 61 amino acids): human metallothionein 1B (*MT1B*) and yeast copper metallothionein (*CUP1*), which are 34% and 20% cysteine, respectively. We used a yeast strain that has the endogenous *CRS1* (the yeast ortholog of *CARS1*) deleted, with viability maintained by a copy of wild-type (WT) *CRS1* on a *URA3*-bearing vector. Dual luciferase reporter constructs containing *MT1B* or *CUP1* were transformed into the yeast strain. Subsequently, constructs containing either wild-type (WT or mutant D714* *CRS1* (corresponds to a disease-associated variant) were transformed. The yeast were then grown in 5-FOA medium, which selects for cells that have spontaneously lost the maintenance vector. We measured renilla luciferase activity as a proxy for the capacity for yeast expressing WT or D714* *CRS1* to translate through the cysteine-rich sequence. Controls included a plasmid with no insert and a plasmid with a similarly-sized coding sequence (64 amino acids) with no cysteine codons (yeast *TMA7*).

For all constructs in both WT and D714* *CRS1*-expressing yeast, firefly luciferase and renilla luciferase activity is observed (Figure 4.17A-D), indicating that we can successfully transform and measure luciferase activity using these constructs in our yeast models. To evaluate renilla luciferase activity, we normalized to the control firefly luciferase activity (Figure 4.17E-F). In

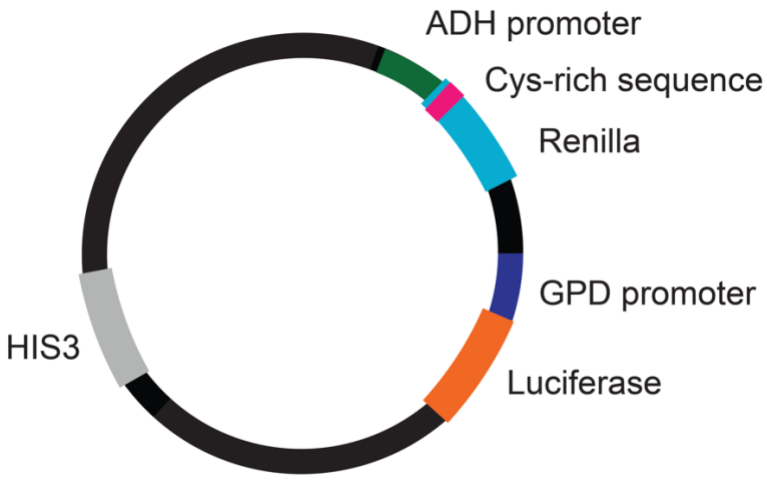


Figure 4.16. Cartoon of dual luciferase reporter construct.

Cysteine-rich sequences were cloned in frame with the renilla coding sequence on a plasmid that contains the firefly luciferase gene as a control. ADH = alcohol dehydrogenase; GPD = glyceraldehyde-3-phosphate dehydrogenase.

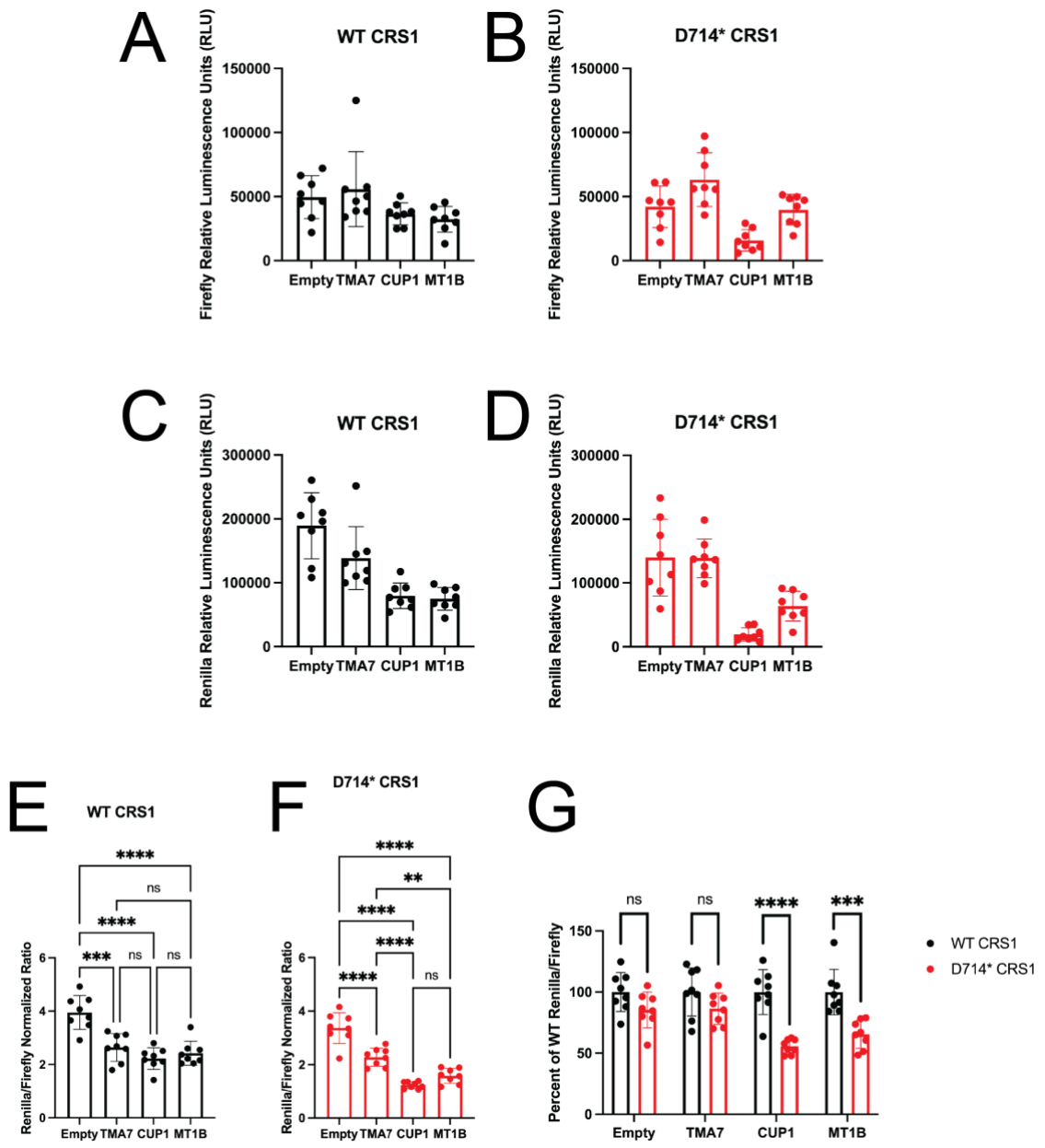


Figure 4.17. Yeast expressing mutant *CRS1* show reduced renilla that is fused to cysteine-rich protein sequences.

Firefly luciferase activity (A and C) and renilla luciferase activity (B and C) were measured from yeast expressing wild-type (WT; A and B) or D714* *CRS1* (C and D) and containing reporter constructs with no insert ('Empty'), an insert with no cysteine ('TMA7'), or a cysteine-rich insert ('CUP1' and 'MT1B'). Renilla luciferase activity was normalized to the control firefly luciferase activity (E and F). Normalized renilla activity in D714* *CRS1* yeast is shown as a percent of the normalized renilla activity of the corresponding construct in WT *CRS1* yeast (G). Statistical significance was calculated using one-way ANOVA (E, F) or multiple t-tests (G) (**p-value < 0.01, ***p-value < 0.001, ****p-value < 0.0001). Bars represent the average of 8 replicate measurements, which are depicted as individual dots. Error bars represent the standard deviation.

both WT and D714* *CRS1*-expressing yeast, all constructs with an insert (cysteine-rich or no cysteine) had significantly reduced renilla activity compared to the construct with no insert (Figure 4.17E-F). In D714* *CRS1*-expressing yeast, renilla activity from constructs carrying cysteine-rich sequences (*CUP1* and *MT1B*) was also significantly reduced compared to the *TMA7* construct with no cysteines (Figure 4.17F). To directly compare normalized renilla activity of the same construct between WT and D714* *CRS1*-expressing yeast, we set the activity in the WT *CRS1*-expressing yeast to 100% and calculated the percent decreased activity in the mutant *CRS1*-expressing yeast (Figure 4.17G). There was no significant difference between WT and D714* *CRS1* yeast for the empty or control (*TMA7*) constructs (Figure 4.17G). In contrast, a significant decrease in renilla activity was observed for the D714* *CRS1*-expressing yeast compared to WT *CRS1*-expressing yeast, specifically in the presence of the cysteine-rich sequences *CUP1* or *MT1B* (Figure 4.17G). These results suggest that translation of the cysteine-rich sequences may be impaired in the mutant *CRS1* yeast and are consistent with our hypothesis that impaired CARS1 activity affects translation of cysteine-rich proteins.

4.3.6 Characterizing a mouse model of CARS1-*associated recessive disease*

To investigate the effects of disease-associated *CARS1* variants on translation in a multicellular, mammalian organism, we generated a mouse model in collaboration with Dr. Miriam Meisler's laboratory (University of Michigan) and the University of Michigan Transgenic Animal Model Core. Using CRISPR/Cas9-mediated mutagenesis, we modeled S688Qfs*2 *CARS1*, which was identified in the homozygous state in a patient (Subject 1-3; see Chapter 3). We obtained three founder mice carrying S688Qfs*2 *Cars1* (Figure 4.18A), and we proceeded with two independent lines of mice heterozygous for S688Qfs*2 *Cars1*. We crossed

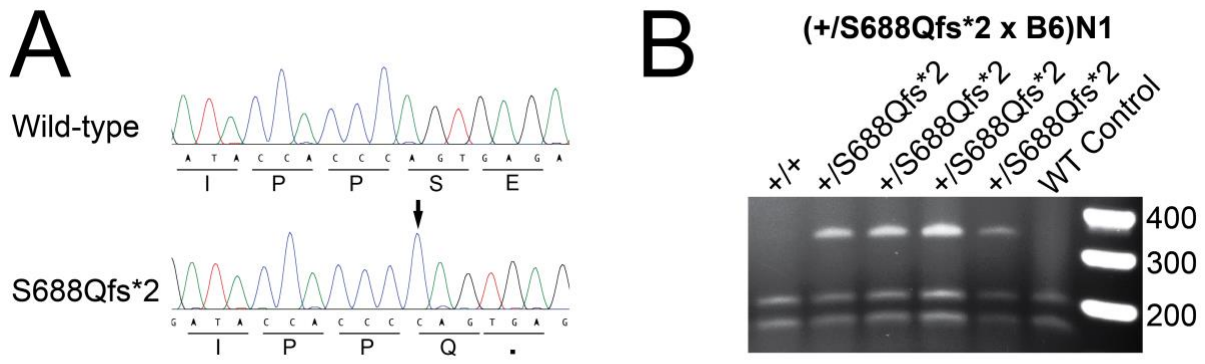


Figure 4.18. Evidence for mutagenesis and transmission of S688Qfs*2 *Cars1* in mice.

(A) Sanger sequencing chromatograms of cloned PCR products from a +/S688Qfs*2 *Cars1* founder mouse. The arrow shows the single base pair insertion, and the effect on amino-acid sequence is indicated below. (B) *Dra*III restriction enzyme digest reactions were performed to genotype mice resulting from the mating of +/S688Qfs*2 founder mice to wild-type B6 mice. S688Qfs*2 ablates a *Dra*III cut site (results in a 335 bp fragment). Wild-type and heterozygous mice were observed, indicating transmission of the mutant allele.

each line to B6 mice (obtained from Dr. Miriam Meisler's laboratory) to establish germline transmission (Figure 4.18B). To model recessive disease and recapitulate the genotype of the patient, who is homozygous for S688Qfs*2 *CARS1*, we intercrossed mice heterozygous for S688Qfs*2 *Cars1* to generate homozygous mice (Figure 4.19A and B). The numbers of each genotype at weaning (~3 weeks old) were observed at Mendelian ratios (Table 4.1); we observed 39 +/+ *Cars1*, 83 +/S88Qfs*2 *Cars1*, and 29 S688Qfs*2/S688Qfs*2 *Cars1* animals ($p = 0.2448$). These data indicate that the S688Qfs*2/S688Qfs*2 *Cars1* genotype does not reduce viability before this timepoint.

From our assessment of Subject 1-3 fibroblasts, we observed that S688Qfs*2 *CARS1* transcripts do not undergo complete nonsense mediated decay and that a truncated CARS1 protein is expressed at levels comparable to cells from unaffected individuals (see Chapter 3); however, these cells may not be severely affected in patients. To examine CARS1 protein expression in mouse tissues relevant to the human phenotype, we obtained brain and liver from three wild-type and three homozygous mutant mice and performed western blot analyses. Consistent with our assessment of patient cell lines and with the viability of the mice, homozygous mice express a truncated CARS1 protein (Figure 4.20A and 4.21A) in both tissues. Interestingly, we observed increased CARS1 protein expression in homozygous mutant mouse brains compared to wild-type mice (Figure 4.20B-C). However, there was no significant difference in liver CARS1 expression between mutant and wild-type mice (Figure 4.21B-C). Further work is needed to determine the mechanism by which CARS1 protein levels are increased in mutant mouse brain samples and to determine if this effect is observed in samples from other tissues.

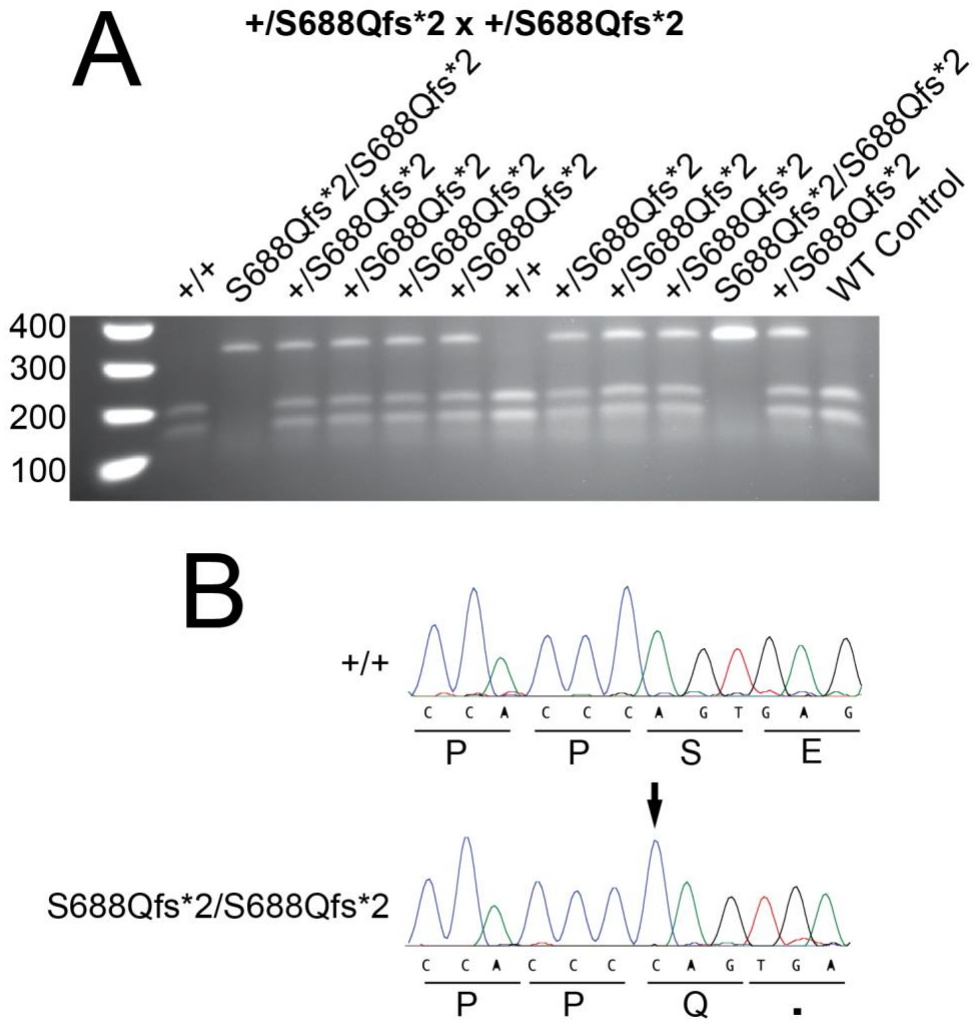


Figure 4.19. S688Qfs*2/S688Qfs*2 *Cars1* mice were obtained from intercrossing +/S688Qfs*2 *Cars1* mice.

(A) *Dra*III restriction enzyme digest reactions were performed to genotype mice resulting from the intercross of +/S688Qfs*2 mice. S688Qfs*2 ablates a *Dra*III cut site (results in a 335 bp fragment). Wild-type, heterozygous, and homozygous mice were observed; see numbers of each genotype in Table 4.1. (B) Representative sanger sequencing chromatograms of PCR products from a wild-type and S688Qfs*2/S688Qfs*2 mouse. The genotype is indicated on the left, and the effect on amino-acid sequence is indicated below. Arrow shows the S688Qfs*2 *Cars1* single base pair insertion.

Table 4.1. Mendelian ratios of N1F1 mice that were genotyped at weaning (~3 weeks old)

	Genotype	Observed	Expected	P-value
Total	+/+	39	37.75	0.2448
	+/S688Qfs*2	83	75.5	
	S688Qfs*2/S688Qfs*2	29	37.75	

	Genotype	Observed	Expected	P-value
Female	+/+	16	18.5	0.1570
	+/S688Qfs*2	45	37	
	S688Qfs*2/S688Qfs*2	13	18.5	

	Genotype	Observed	Expected	P-value
Male	+/+	23	19.25	0.5258
	+/S688Qfs*2	38	38.5	
	S688Qfs*2/S688Qfs*2	16	19.25	

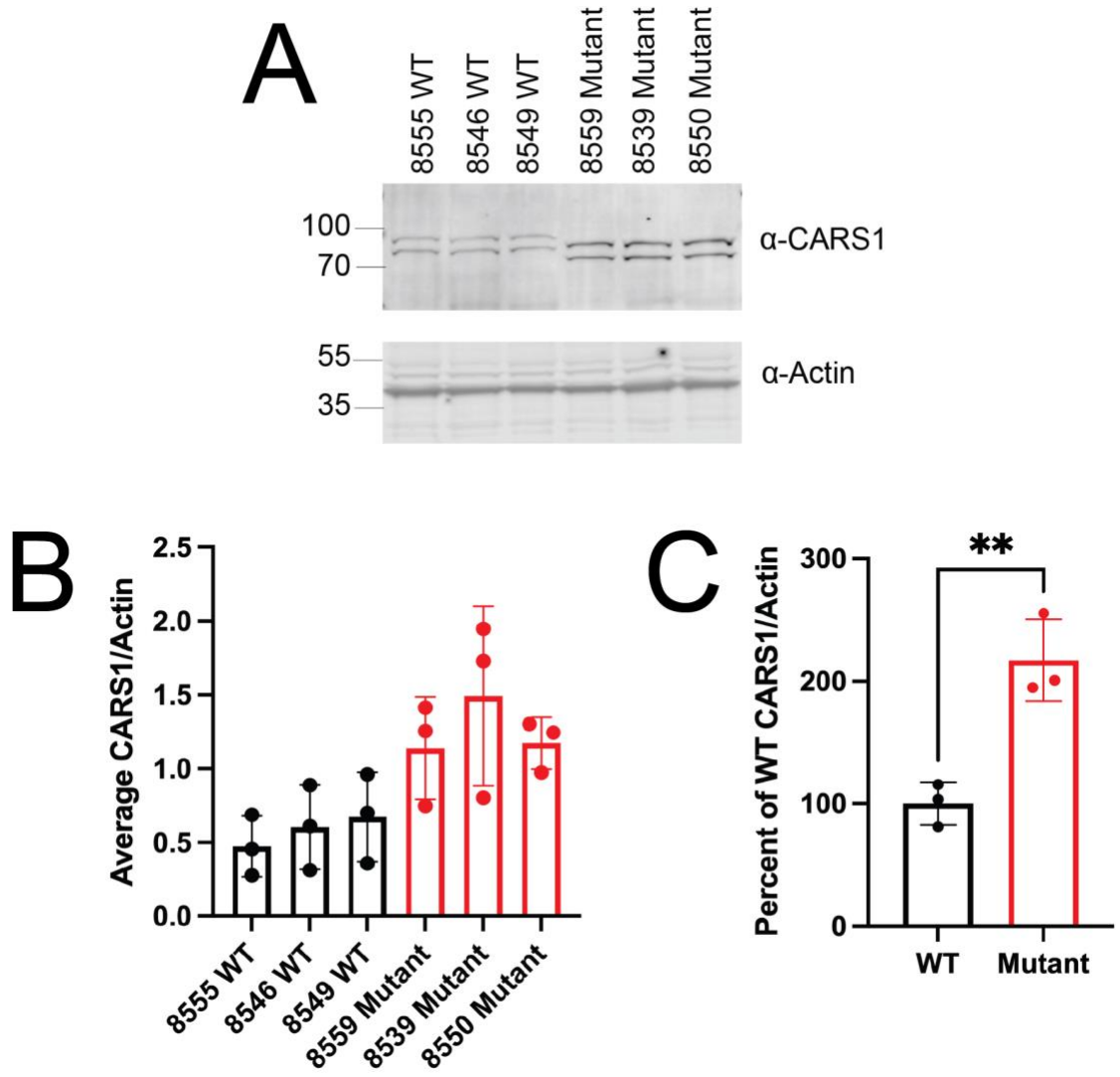


Figure 4.20. Homozygous mutant mice express truncated CARS1 protein in the brain.

(A) Western blot analyses were performed using total protein lysates isolated from mouse brain, and an anti-CARS1 antibody or anti-actin antibody was used (indicated on the right). Sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of CARS1 signal was quantified using ImageJ and normalized to actin expression. The dots represent the CARS1/actin from each individual experiment, and bars represent the average (n = 3). Error bars represent the standard deviation. (C) Normalized CARS1 signal in mutant animals is shown as a percent of the normalized CARS1 signal of wild-type animals. Statistical significance was calculated using a t-test. The dots represent the average CARS1/actin for each animal, and bars represent the average (n = 3). Error bars represent the standard deviation. ** $p < 0.01$.

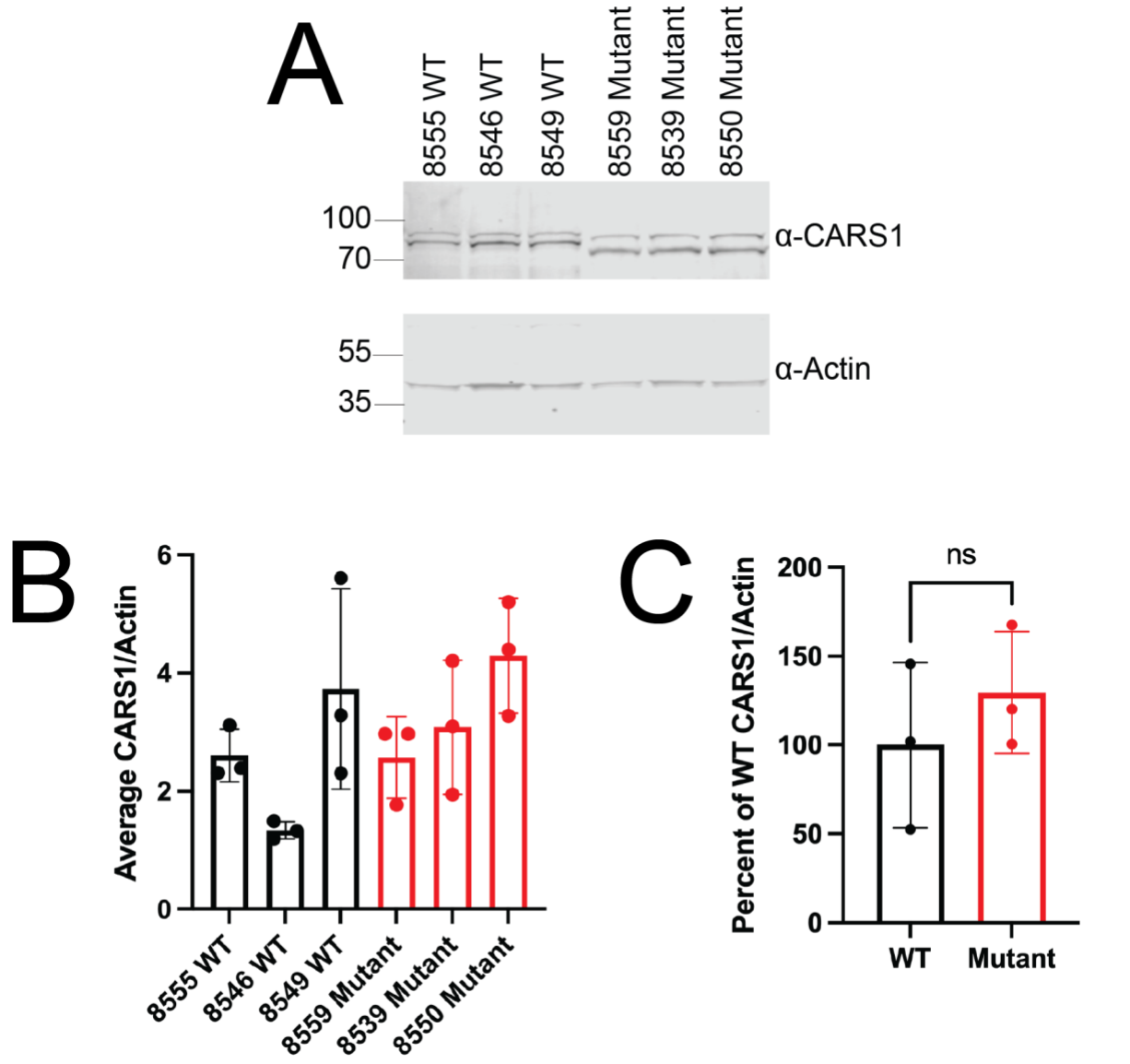


Figure 4.21. Homozygous mutant mice express truncated CARS1 protein in the liver.

(A) Western blot analyses were performed using total protein lysates isolated from mouse liver, and an anti-CARS1 antibody or anti-actin antibody was used (indicated on the right). Sample names are across the top, and sizes (kDa) are indicated at the left. (B) The intensity of CARS1 signal was quantified using ImageJ and normalized to actin expression. The dots represent the CARS1/actin from each individual experiment, and bars represent the average (n = 3). Error bars represent the standard deviation. (C) Normalized CARS1 signal in mutant animals is shown as a percent of the normalized CARS1 signal of wild-type animals. Statistical significance was calculated using a t-test. The dots represent the average CARS1/actin from each animal, and bars represent the average (n = 3). Error bars represent the standard deviation. ** $p < 0.01$.

During the assessment of CARS1 protein expression in mouse brain and liver, we observed that two CARS1 protein isoforms (~86 kDa and ~95 kDa) are present in both tissues (Figures 4.20 and 4.21). The smaller protein isoform is consistent with the expected size of the mouse CARS1 protein (NM_001252593; 748 amino acids; equivalent to human CARS1 isoform 1 [see Chapter 3]), which is predicted to be ~86 kDa. We suspect that the larger protein isoform may represent an isoform of CARS1 that contains an additional in-frame exon, exon 2, which is predicted to be ~95 kDa (NM_013742; 831 amino acids). In humans, transcripts containing exon 2 have also been identified (NM_001014437 and NM_001194997). Inclusion of exon 2, as a result of alternative splicing, leads to the addition of 83 amino acids that are not essential for aminoacylation and that interact with elongation factor-1; however, the biological significance of this interaction is not fully understood[286]. Additionally, in humans, the isoform with exon 2 was reported to only be expressed in the testis[286]. Consistent with this, we did not observe expression of this isoform in our patient fibroblast assays (see Chapter 3). However, an isoform with exon 2 may be more widely expressed in mouse; publicly available brain RNA sequencing data suggest that transcripts containing exon 2 are expressed in the mouse brain[287]. We experimentally assessed the presence of exon 2-containing transcripts in wild-type mouse brain using RT-PCR (Figure 4.22). We observed two RT-PCR products, consistent with transcripts containing (382 bp) or lacking (133 bp) exon 2 (Figure 4.22B). Sanger sequencing analysis confirmed the presence of RT-PCR products with and without exon 2 (Figure 4.22C). These results suggest that in mouse brain, *Cars1* transcripts both with and without exon 2 are expressed. Further work is needed to determine the expression of the isoform containing exon 2 in different tissues, the functions of the exon 2-containing protein isoform, and how the S688Qfs*2 mutation may affect its functions.

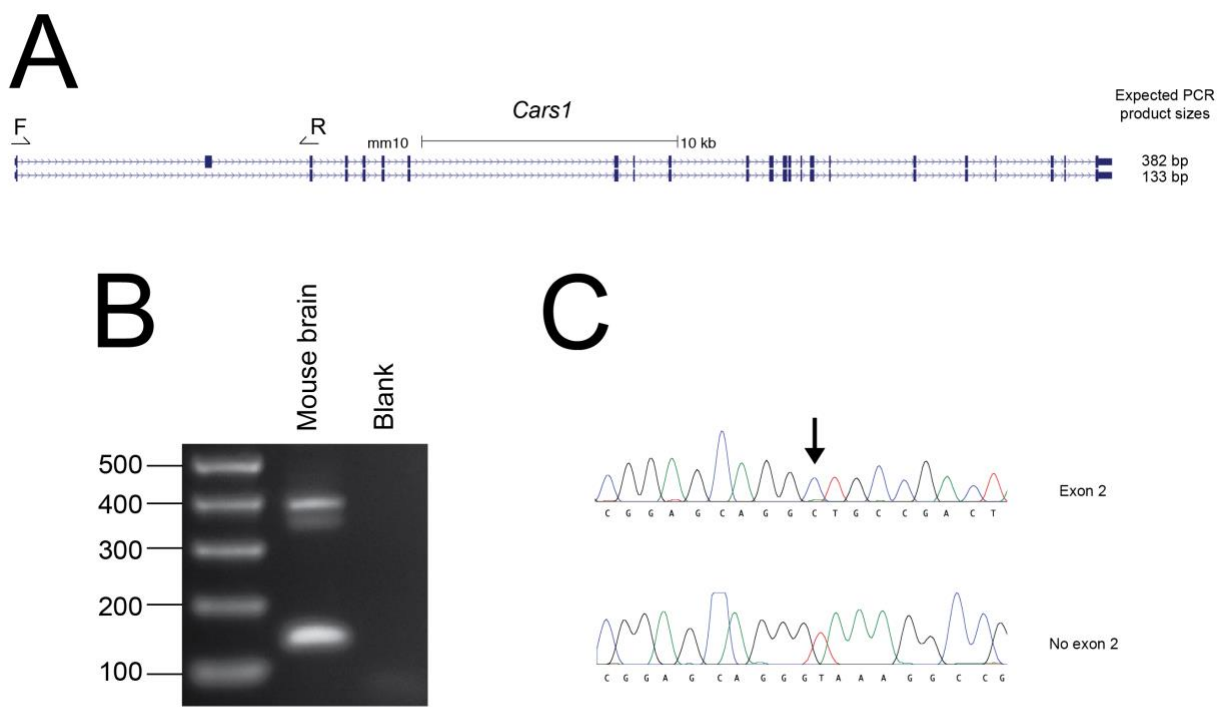


Figure 4.22. *Cars1* transcripts containing exon 2 are expressed in mouse brain.

(A) The mouse *Cars1* genomic locus from the UCSC Genome Browser is shown. Isoforms containing exon 2 (NM_013742; 831 amino acids) and lacking exon 2 (NM_001252593; 748 amino acids) are depicted. Location of RT-PCR primers used to test for exon 2 are depicted with arrows ('F' and 'R'), and the expected RT-PCR product sizes for each isoform transcript is labeled on the right. (B) RT-PCR assay to test for the presence of exon 2-containing transcripts in mouse brain. Base pair sizes of markers are on the left. (C) Sequence chromatograms are shown for the corresponding RT-PCR products in (B). Arrow indicates the beginning of exon 2.

4.3.7 Homozygous mutant Cars1 *mice exhibit reduced body weight*

While establishing the *Cars1* mouse model, we observed that homozygous mice exhibit reduced body weight compared to wild-type mice (Figure 4.23 and 4.24). To address this, we measured the weights of wild-type and homozygous mutant male and female animals at 10 weeks of age. At 10 weeks of age, male homozygous mutant mice have an average weight of 21.6 g (n = 11), which is significantly less (p-value < 0.001) than male wild-type mice (average weight = 26.2 g; n = 9) (Figure 4.23C). Similarly, at 10 weeks of age, female homozygous mice (average weight = 16.87 g; n=10) are significantly smaller (p-value < 0.001) than wild-type female mice (average weight = 19.1 g; n = 6) (Figure 4.23D). Interestingly, this finding is consistent with the growth restriction phenotype observed in patients with bi-allelic *CARS1* variants.

We additionally assessed appearance and behavior of three wild-type and three homozygous mutant male mice at 12 weeks old by monitoring them in an enclosed container. No differences in overall appearance (aside from body size) and no overt behavioral or motor phenotypes were observed by the author or Dr. Guy Lenk (Meisler laboratory, University of Michigan; Figure 4.25). We used a USB microscope camera connected to a computer to image hair and nails of each mouse (Figure 4.25), and again, no gross differences were noted between wild-type and mutant animals.

To further investigate mutant-associated phenotypes, we submitted two wild-type (ear tags 8867 and 8869) and two homozygous mutant (ear tags 8870 and 8871) *Cars1* mice at 46 days old to

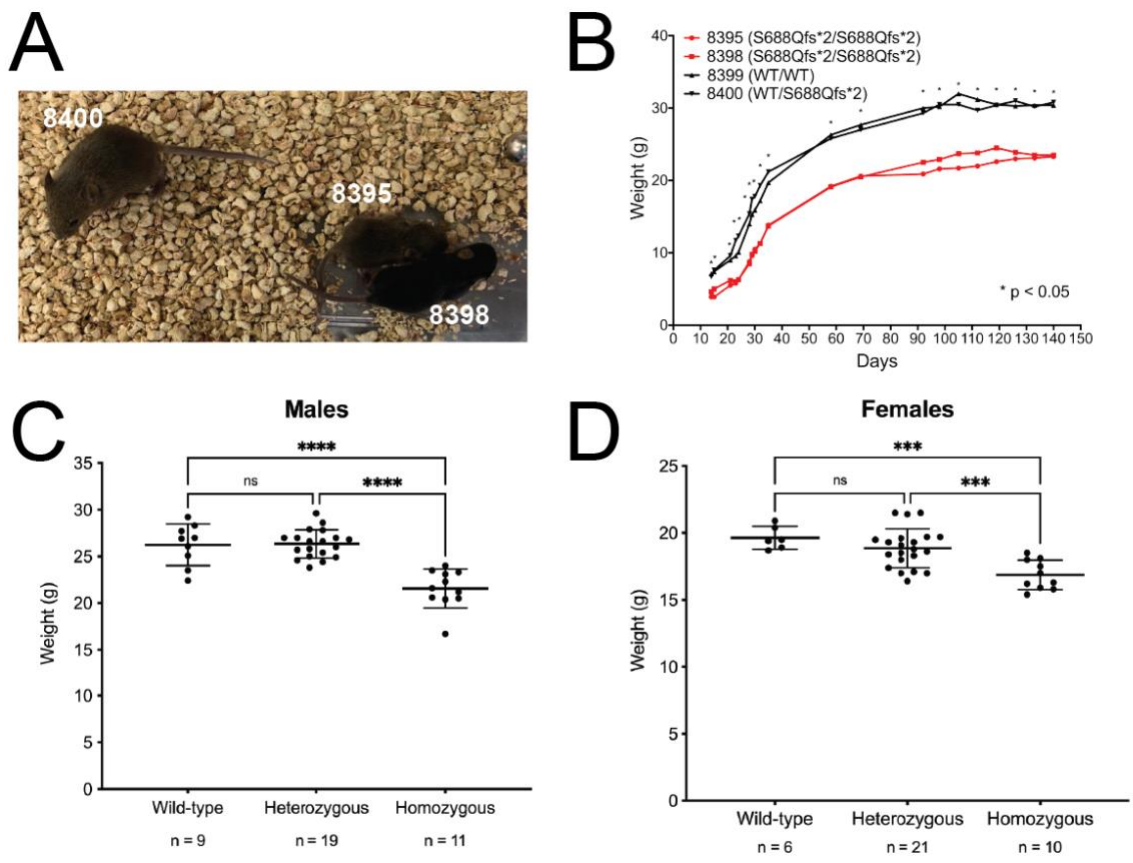


Figure 4.23. Homozygous mutant mice have reduced body weight compared to wild-type mice.

(A) Image of S688Qfs*2/S688Qfs*2 mice (ear tags 8395 and 8398) and a wild-type litter mate (ear tag 8400) at 14 days old. (B) Body weights of an N1F1 male litter. Weight is on the y-axis, and days is on the x-axis. Weights of homozygous mutant animals are plotted in red, and wild-type or heterozygous animal weights are plotted in black. Statistical significance was determined by t-test. (C) Male homozygous mice at 10 weeks of age are significantly smaller than wild-type and heterozygous mice. (D) Female homozygous mice at 10 weeks of age are significantly smaller than wild-type and heterozygous mice. Statistical significance as determined by one-way ANOVA. * p-value < 0.05; *** p-value < 0.001; **** p-value < 0.0001.

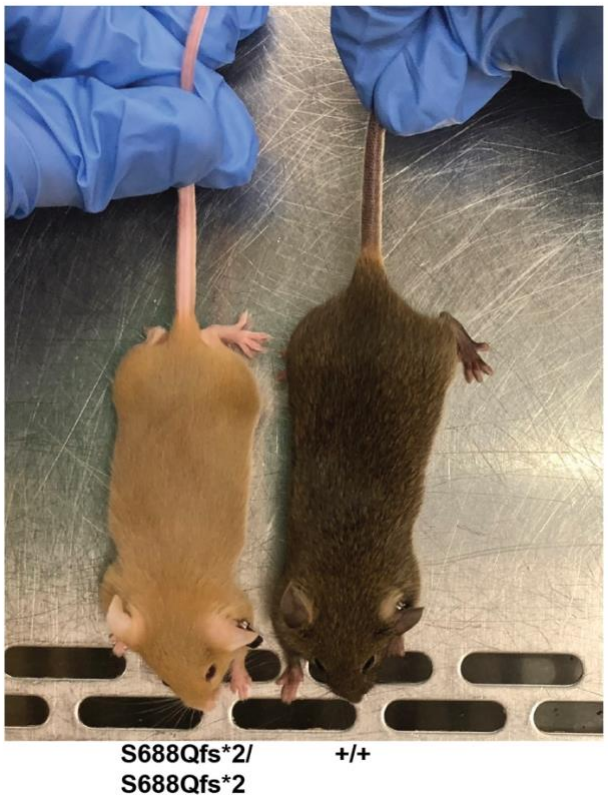


Figure 4.24. Homozygous mutant mice are smaller than wild-type mice.

Mice were imaged at 39 days old and then submitted to the University of Michigan In Vivo Animal Core for phenotyping necropsies. Note that different coat colors were consistent with coat color variants in the mixed C57BL/6J and SJL background and were not correlated with the *Cars1* genotypes.

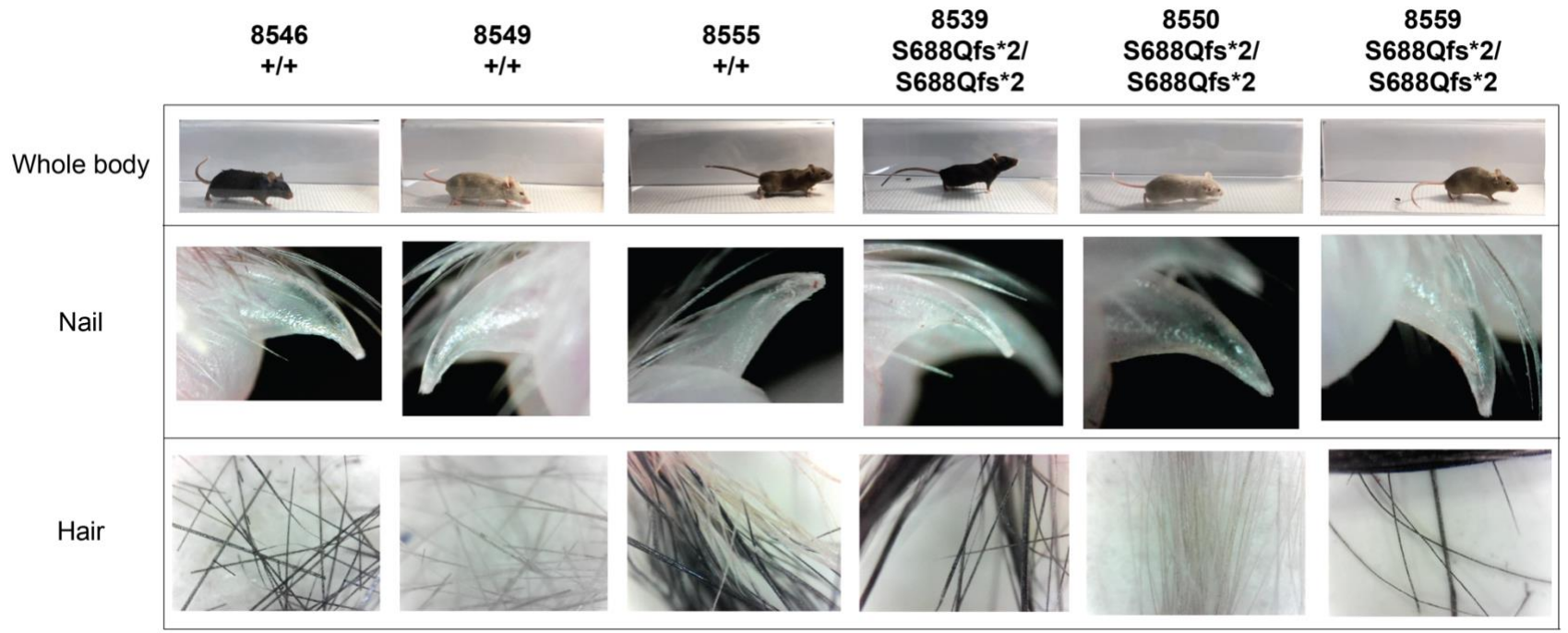


Figure 4.25. No obvious differences were observed between mutant and wild-type animals in overall appearance, nails, or hair.

Images were taken of mouse whole body (first row), nails (second row), and hair (third row). Ear tag and genotypes are labeled across the top. Note that different coat colors were consistent with the coat color variants in the mixed C57BL/6J and SJL background and were not correlated with the *Cars1* genotypes.

the In-Vivo Animal Core (IVAC) run by the Unit for Laboratory Animal Medicine at the University of Michigan for phenotyping necropsies and bloodwork (Table 4.2). As previously observed, the homozygous mutant mice (average 16 g) were smaller than wild-type mice (average 21 g; Figure 4.24 and Table 4.2 and 4.3), but no other differences were noted on gross examination. Liver weights were not significantly different between mutant and wild-type mice, and tibial length was 20-21 mm for all of the mice (Table 4.3). Complete blood count with automated differential revealed that each mutant animal had only one value outside of normal limits (Table 4.4); the mutant mouse with ear tag 8870 was just under the normal range for mean platelet volume (4.9 fL; normal range is 5.0-20.0 fL), and the mutant mouse with ear tag 8871 had just higher than the normal range monocyte number (0.44 K/ul; normal range is 0.0-0.4). Both wild-type animals also had high monocyte numbers (Table 4.4). Hemoglobin was lower in mutant animals (average 13 g/dL) compared to wild-type animals (average 16 g/dL), and platelet count was higher in mutant animals (average 745) compared to wild-type animals (average 391); however, further work is needed to determine the reproducibility and significance of these findings. Full chemistry panels were also performed, and there were no abnormal values that were shared by both mutant animals and not also shared with the wild-type animals (Table 4.5). Histologic examination was performed on multiple organs to test for pathological changes, paying particular attention to the skin, liver, central nervous system, and long bone growth plates. One mutant (8870) had lateral thinning of the spinal cord (Table 4.6), and the other mutant (8871) had lymphocytic peritonitis in peripancreatic fat, which was not considered clinically significant (Table 4.6). The biological significance of these changes is unclear. Other findings were considered sporadic background changes. Overall, homozygous mutant mice are smaller than wild-type, but have no other observable differences in the assays employed above.

Table 4.2. Animals submitted to IVAC for phenotyping necropsy

Ear Tag	Sex	Age at Necropsy	*Cars1* Genotype	Body Weight*
8867	Male	46 days	+/+	19.9g
8869	Male	46 days	+/+	22.4g
8870	Male	46 days	S688Qfs*2/S688Qfs*2	15.9g
8871	Male	46 days	S688Qfs*2/S688Qfs*2	16.2g

* Measured by the author prior to submission to IVAC.

Table 4.3. Parameters collected at necropsy by IVAC

Ear Tag	8867	8869	8870	8871
***Cars1* Genotype**	+/+	+/+	S688Qfs*2/ S688Qfs*2	S688Qfs*2/ S688Qfs*2
Terminal Body Weight**	20.78g	--*	17.11g	--*
Tibial Length (mm)	20	21	20	21
Liver Weight	0.986g	1.475g	0.978g	0.883g

***** Data inadvertently not collected at study termination.
** Measured by IVAC.

Table 4.4. Complete blood count with automated differential of wild-type and homozygous mutant mice

		8867	8869	8870	8871
		+/+	+/+	S688Qfs*2/ S688Qfs*2	S688Qfs*2/ S688Qfs*2
White blood cell	K/uL	10.26	4.32	4.34	5.72
Neutrophil #	K/uL	1.98	0.79	0.96	1.01
Lymphocyte #	K/uL	7.52	3.08	3.15	4.09
Monocyte #	K/uL	0.62 H	0.42 H	0.21	0.44 H
Eosinophil #	K/uL	0.13	0.03	0.02	0.14
Basophil #	K/uL	0.02	0.01	0	0.03
Neutrophil %	%	19.25	18.23	22.03	17.68
Lymphocyte %	%	73.3	71.29	72.55	71.59
Monocyte %	%	6.05	9.62	4.9	7.77
Eosinophil %	%	1.23	0.59	0.48	2.38
Basophil %	%	0.17	0.27	0.04	0.59
Red blood cell	M/uL	9.5 H	10.2 H	7.87	8.13
Hemoglobin	g/dL	15.1	16.1	12.3	13.5
Hematocrit	%	47.5 H	50.7 H	39.3	42
Mean corpuscular volume	fL	50	49.7	49.9	51.6
Mean corpuscular hemoglobin	Pg	15.9	15.8	15.6	16.6
Mean corpuscular hemoglobin concentration	g/dL	31.8	31.8	31.3	32.1
Red cell distribution width	%	17.4	17.6	18.7	17
Platelet	K/uL	414 L	368 L	592	897
Mean platelet volume	fL	5.6	6.1	4.9 L	5.2

Assays were performed by IVAC. H = high; L = low.

Table 4.5. Full chemistry panel of wild-type and homozygous mutant mice

		8867	**8869**	**8870**	**8871**
		+/+	**+/+**	**S688Qfs*2/ S688Qfs*2**	**S688Qfs*2/ S688Qfs*2**
Triglycerides	mg/dL	172	168	114	60 L
Cholesterol	mg/dL	70	93	71	59 L
Aspartate aminotransferase	U/L	799 H	716 H	775 H	78
Glucose	mg/dL	452 H	315	285	201
Alanine aminotransferase	U/L	**	**	110	37
Creatine phosphokinase	U/L	**	**	7154	295
Total bilirubin	mg/dL	0.07 L	0.06 L	0.07 L	0.06 L
Blood urea nitrogen	mg/dL			35 H	21
Alkaline phosphatase	U/L			175	206

Assays were performed by IVAC. H = high; L= low; ** = not enough serum to run test.

Table 4.6. Histology analysis of wild-type and homozygous mutant mice

	8867	**8869**	**8870**	**8871**
	+/+	**+/+**	**S688Qfs*2/ S688Qfs*2**	**S688Qfs*2/ S688Qfs*2**
Skin, haired, cervical area	N	*	N	N
Fibrosis, periadnexal	-	2-f	-	-
Decreased density of follicles	-	3-mf	-	-
Brain	N	N	N	N
Spinal cord	N	N	*	N
Hypoplasia, lateral white tracts	-	-	2-mf	-
Kidney	N	N	N	N
Liver	N	N	N	*
Liver, hepatocellular vacuolation, centrilobular	-	-	-	1-mf
Spleen	*	*	*	N
Prominent extramedullary hematopoiesis, red pulp	1-mf	2-mf	1-mg	2-mf
Heart	N	N	N	N
Skeletal muscle	N	N	N	N
Lung	N	*	*	*
Alveolar hemorrhage	-	2-mf	2-f	2-f
Coagulative necrosis, alveolar wall	-	2-f	-	-
Adrenal	N	N	N	N
Pancreas	N	N	N	*
Lymphocytic peritonitis, peripancreatic mesentery	-	-	-	1-mf
Small intestine	N	N	N	N
Stomach	N	N	N	N
Colon and cecum	N	N	N	N
Eyes	N	*	*	N
Vacuolation, corneal stroma	-	2-mf	2-mf	-
Femur	N	N	N**	N
Forepaw	N	N	N	N
Sternum	N	N	N	N
Testes	N	N	N	N
Miscellaneous gross lesions	NA	NA	NA	NA

Assays were performed by IVAC. N = normal tissue morphology; * = changes noted in tissue; mf = multifocal; f = focal; 1 = minimal; 2 = mild severity; ** no growth plate in section.

4.3.8 Few proteins are differentially expressed in mutant brains and livers compared to wild-type tissues

To determine if expression of cysteine-rich proteins is affected in the mouse, we collected brain and liver samples from three wild-type and three homozygous mutant male mice at 3 months. The weight of each organ was measured, and the brain weights were significantly reduced in homozygous mutant mice compared to wild-type mice (Figure 4.26A); however, when the brain weights were normalized to total body weight; there was no statistically significant difference (Figure 4.26B). There was no significant difference in liver weights of wild-type or homozygous mutant mice (Figures 4.26C-D). We submitted half of each brain and liver sample from each animal to the University of Michigan Proteomics and Peptide Synthesis Core. Samples were subjected to protein extraction, trypsin digestion, and liquid chromatography with tandem mass spectrometry (LC-MS/MS). Data were analyzed using the Scaffold proteome software and filtered using a 1% protein and peptide FDR, and at least two unique peptides were required per protein. In the brain, 3,630 total proteins were identified (Figure 4.27A); three proteins were down-regulated (fold change < 0.5; p-value < 0. 05; Table 4.7), and one protein was up-regulated (fold change > 2; p-value < 0. 05; Table 4.7) in mutant animals compared to wild-type animals. In the liver 2,748 total proteins were identified (Figure 4.27B); five proteins were down-regulated (fold change < 0.5; p-value < 0. 05; Table 4.8), and nine proteins were up-regulated (fold change > 2; p-value < 0. 05; Table 4.8) in the mutant animals compared to wild-type animals. No obvious trends were observed among these small groups of proteins. Due to low numbers of spectral counts overall and specifically of peptides covering cysteine-rich proteins across all samples, we cannot conclude if cysteine-rich proteins are expressed at lower levels in homozygous mutant animals.

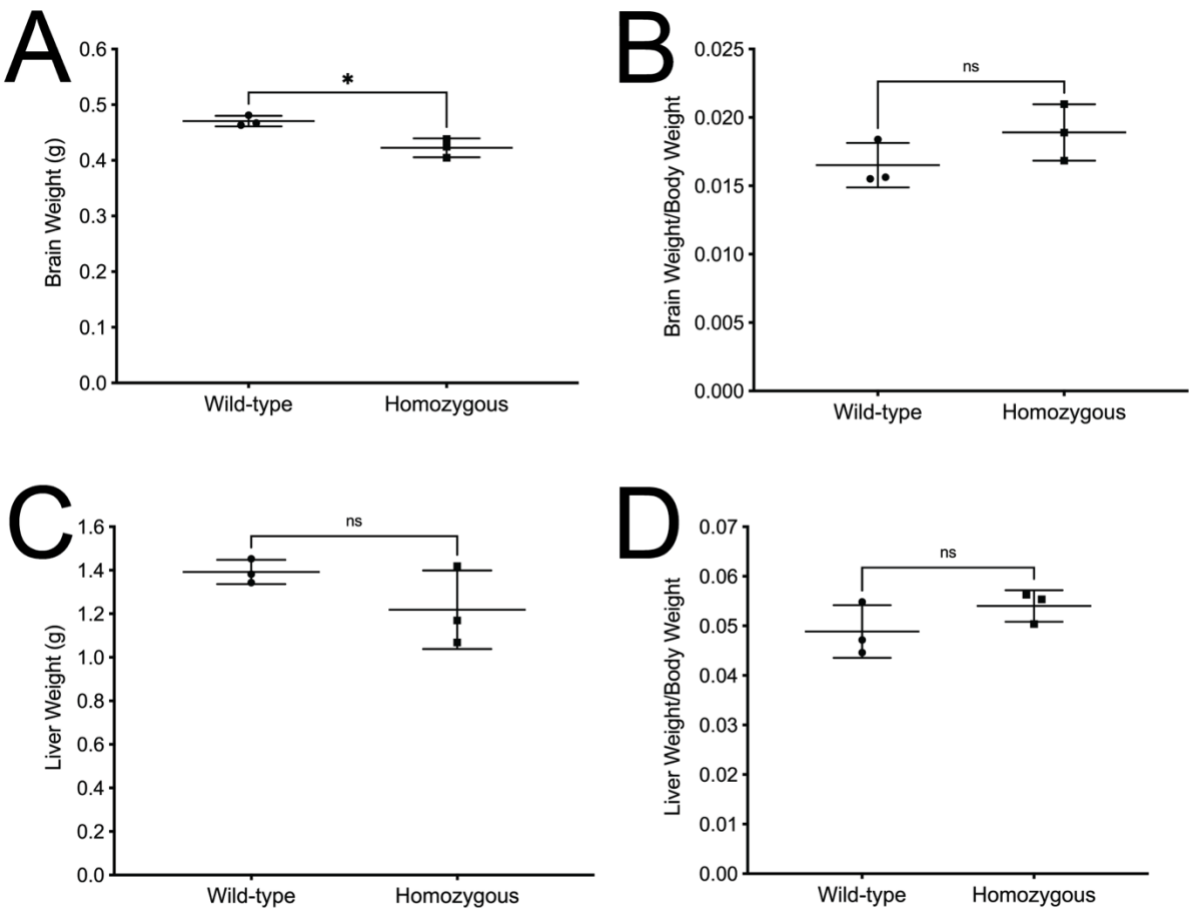


Figure 4.26. Homozygous mutant mice have similar brain and liver weights as wild-type mice.

(A) Brain was dissected and weighed from three wild-type and three homozygous mutant animals at 3 months old. (B) Brain weights in (A) were normalized to total body weight. (C) Liver was dissected and weighed from three wild-type and three homozygous mutant animals at 3 months old. (D) Liver weights in (C) were normalized to total body weight. Dots represent each mouse. Statistical significance was determined using t-test. * p-value < 0.05.

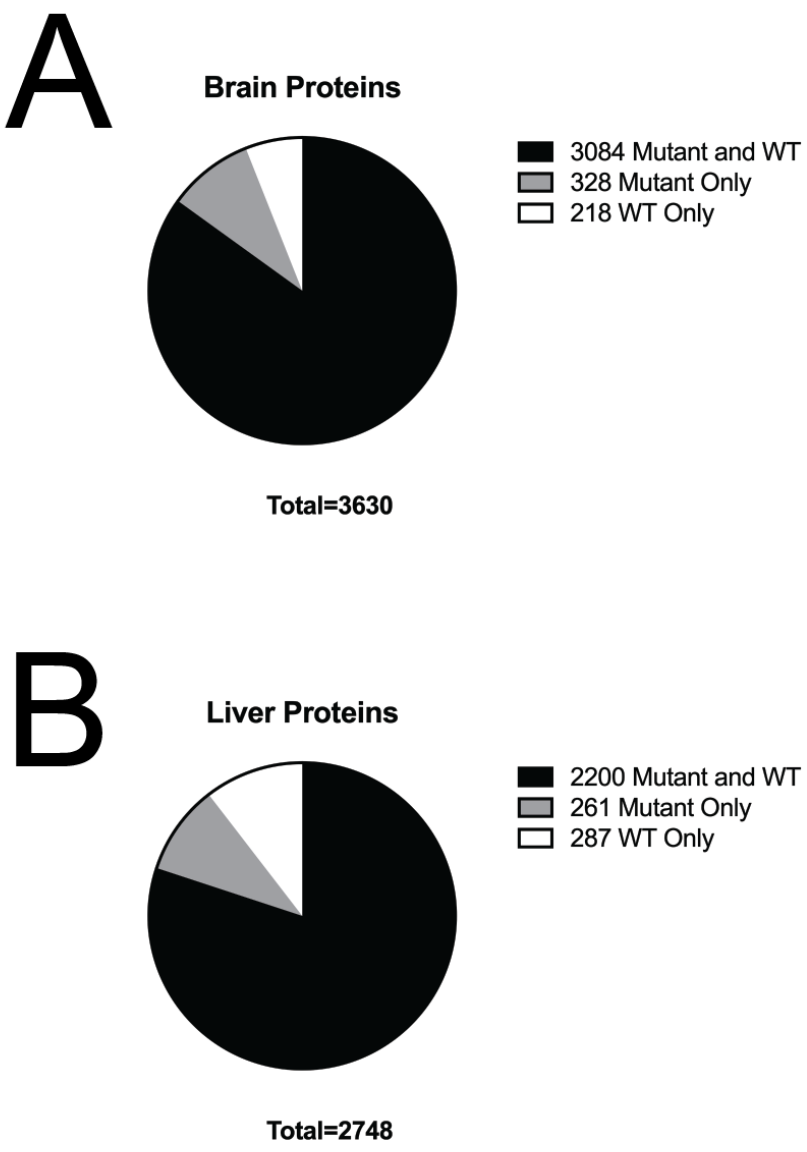


Figure 4.27. Total number of proteins identified in brain and liver through quantitative proteomics experiments.

Brain and liver were dissected from three wild-type and three homozygous mutant (S688Qfs*2/S688Qfs*2 *Cars*) 3-month-old male mice, and protein was isolated and subjected to quantitative proteomics analysis (LC-MS/MS). The number of proteins identified in brain (A) and liver (B) from wild-type and homozygous mutant mice are depicted. The quantitative proteomics experiments were performed by the University of Michigan Proteomics and Peptide Synthesis Core.

Table 4.7. Differentially expressed proteins in homozygous mutant mouse brains

Protein	Fold Change (Mutant/WT)	P-Value
28S ribosomal protein S22, mitochondrial	0.25	0.0018
Guanidinoacetate N-methyltransferase	0.36	0.034
V-type proton ATPase subunit F	0.46	0.013
Laminin subunit alpha-5	4.25	0.034

Fold change for each protein was calculated by dividing the average normalized spectral abundance factor (NSAF) of the three mutant animals by the average NSAF of the three wild-type animals. Statistical significance was determined by t-test. Proteins with fold change (mutant/wild-type) > 2 or < 0.5 and p-value < 0.05 were determined to be significantly differentially expressed. These experiments were performed by the University of Michigan Proteomics and Peptide Synthesis Core.

Table 4.8. Differentially expressed proteins in homozygous mutant mouse livers

Protein	Fold Change (Mutant/WT)	P-Value
Inter alpha-trypsin inhibitor, heavy chain 4	0.17	0.0012
Protein transport protein Sec16B	0.18	0.00092
Cysteine and glycine-rich protein 3	0.22	0.0095
Phosphatidylinositol-binding clathrin assembly protein	0.39	0.039
Stromal cell-derived factor 2-like protein 1	0.40	0.015
Opioid growth factor receptor	2.26	0.032
Haloacid dehalogenase-like hydrolase domain-containing protein 3	2.58	0.030
GTP cyclohydrolase 1	3.29	0.028
Elongation factor G, mitochondrial	3.60	0.015
Protein AMBP	4.36	0.043
Creatine kinase B-type	4.54	0.0071
Serine/threonine-protein kinase PAK 2	5.06	0.022
V-type proton ATPase 16 kDa proteolipid subunit	5.63	0.0035
39S ribosomal protein L1, mitochondrial	5.71	0.038

Fold change for each protein was calculated by dividing the average normalized spectral abundance factor (NSAF) of the three mutant animals by the average NSAF of the three wild-type animals. Statistical significance was determined by t-test. Proteins with fold change (mutant/wild-type) > 2 or < 0.5 and p-value < 0.05 were determined to be significantly differentially expressed. These experiments were performed by the University of Michigan Proteomics and Peptide Synthesis Core.

4.4 Discussion

In summary, we present studies using patient fibroblast cells and yeast and mouse models to investigate the downstream consequences of *CARS1* variants on translation. Since CARS1 charges tRNA with cysteine for use in translation, we hypothesized that impaired CARS1 activity may lead to defects in global translation through the integrated stress response that is triggered by uncharged tRNA We also predicted that translation of cysteine-rich proteins may be affected due to reduced abundance of charged $tRNA^{Cys}$.

4.4.1 The effects of CARS1 *variants on global translation*

To test if *CARS1* variants affect protein synthesis globally, we tested puromycin incorporation of patient and control fibroblasts. To specifically interrogate whether a general effect on translation may be occurring through the integrated stress response, we also assessed phosphorylation of eIF2α. In standard growth conditions, there were no differences in puromycin incorporation or phosphorylation of eIF2α, indicating that global translation was not affected in these cells under these conditions (Figures 4.1 and 4.4). We also investigated if translation stress through DTT treatment and through cysteine and methionine deprivation would reveal any differences between patient and control cells (Figure 4.5-4.12), and while some differences were observed, no differences were shared by all patient cells. Most notably, after 30 minutes of DTT treatment, two of the patient cells lines and one of the unaffected cell lines showed decreased puromycin incorporation (Figure 4.5), and interestingly, this was not accompanied by increased eIF2α phosphorylation (Figure 4.6), which would be expected if the integrated stress response was increased. In fact, Subject 3-3 showed a decrease in eIF2α phosphorylation, suggesting that the

integrated stress response was not responsible for the decreased puromycin incorporation observed. Puromycin incorporation was not significantly different between patient and control cells after 60 minutes DTT, cysteine deprivation, or cysteine and methionine deprivation (Figure 4.7-4.12), indicating that global translation was not affected in these patient fibroblasts compared to control cells, even when stressed. Overall, our assays do not indicate significant reductions of global translation in fibroblasts from patients carrying pathogenic *CARS1* variants. These findings are consistent with previously reported data showing that fibroblasts from patients with *FARSB*-associated recessive disease also do not exhibit reduced puromycin incorporation[75]. However, it is possible that there are small differences in global protein synthesis that these assays are not sensitive enough to detect. Additionally, these studies were only performed on fibroblast cells in culture, which is not an ideal model for multi-system disease. Studying cells from tissues relevant to the human phenotypes (like brain and liver) may yield different results. Additionally, cells in culture may not reflect the behavior of cells in the multicellular environment of tissues; therefore, model organisms, like the mouse model, should be employed. Therefore, it is still possible that a global translation effect may contribute to disease pathogenesis in some tissues and cell types.

4.4.2 The effects of CARS1 *variants on the translation of cysteine-rich proteins*

We hypothesized that impaired CARS1 activity would result in decreased available charged $tRNA^{Cys}$, and that this might lead to ribosome stalling at cysteine codons. Stalling may occur more frequently on transcripts containing many cysteines and lead to degradation of the transcripts. To investigate if high cysteine-containing transcripts had decreased expression in patients with *CARS1* variants, we performed RNA sequencing on fibroblast cells from patients

and unaffected individuals (Figure 4.13 and 4.14). When we compared gene expression of patient cells with control cells (or with control and Unaffected 1-2 cells), the downregulated genes have higher proportion of cysteine compared to the genes with no difference in expression; however, the upregulated genes also have a higher proportion of cysteine compared to the unchanged genes (Figure 4.13B and 4.14B). Additionally, the increase in average proportion of cysteine is very small (0.03 compared to 0.02), and so the biological relevance of this increase is unclear. Overall, it seems that cysteine-codon-rich transcripts are not preferentially downregulated in patient fibroblast cells in these conditions since upregulated genes also have higher cysteine codon content compared to unchanged genes; however, there are limitations with these studies that may affect interpretations, such as the limitations of studying fibroblasts described above in Chapter 4.4.1. Additionally, the cells were cultured in standard growth conditions (see Chapter 4.2.1), and we suspect that this might not be the most appropriate model since the cells have an abundance of resources and may have found ways to adapt to having impaired CARS1 activity in this culture setting. Consistent with this notion, fibroblasts were expanded in culture without difficulty and have no overt growth or morphological phenotype. Furthermore, to specifically examine the effects of the *CARS1* mutations on gene expression, patient cells with *CARS1* mutations corrected may be better controls, since the control cells used (normal human fibroblasts purchased from ATCC) may have other genetic differences from patient cells that affect translation. Since we had cells from the parent of one of the patients, we compared Subject 1-3 with his mother Subject 1-2, whose genome should be more similar to Subject 1-3 than the control cells (Figure 4.15). This comparison interestingly revealed that genes involved in collagen fibril organization were downregulated in the patient, who presented with connective tissue phenotypes. However, these transcripts are not cysteine-codon-rich; the

average percent cysteine content is 0.8%, which is below the proteome average of 3%. More work is needed to determine what causes these gene expression changes and if they may contribute to the poor wound healing and recurrent hernia phenotypes.

To address some of the limitations of the fibroblast model, we investigated the effect on translation of cysteine-rich proteins in a yeast model, which allows us to examine differences between a yeast strain expressing wild-type and mutant *CARS1* or *CRS1*, the yeast ortholog of *CARS1*. We previously showed that yeast strains that express patient variants display a reduced yeast growth phenotype compared to wild-type yeast (see Chapter 3). To test for effects on the translation of cysteine-rich proteins using this system, we developed a dual luciferase assay in which cysteine-rich sequences were cloned in frame with renilla. Then, renilla luciferase activity was measured as a proxy for the capacity for yeast expressing wild-type or D714* *CRS1* to translate through the cysteine-rich sequence yeast. Yeast expressing mutant *CRS1* had significantly reduced renilla activity for constructs containing cysteine-rich sequences compared to wild-type *CRS1*-expressing yeast (Figure 4.17). These results suggest that translation of the cysteine-rich sequences may be impaired in the mutant *CRS1*-expressing yeast and are consistent with our hypothesis that impaired CARS1 activity affects translation of cysteine-rich proteins. We have observed similar results in replicate experiments, but there is variability in the measurements; to address this, we are generating yeast with the D714* mutation knocked into the genome instead of expressed from an exogenous plasmid, which can result in variable expression due to differences in plasmid copy number. Importantly, future work is also needed to rule out that the observed reduction in renilla activity is due to decreased transcript levels.

To assess the effects of *CARS1* variants on translation in a multicellular organism, we generated a mouse model of a patient mutation, S688Qfs*2 *CARS1*. Homozygous mutant mice presented with reduced body weight compared to wild-type and heterozygous animals (Figure 4.23 and 4.24), which appears to recapitulate the growth restriction phenotype observed in Subject 1-3. However, behavioral and histological phenotyping studies revealed no other gross differences between homozygous mutant and wild-type mice (Figure 4.25 and Table 4.6). We analyzed protein expression in brain and liver samples from homozygous mutant and wild-type mice through quantitative proteomics and observed very few differentially expressed proteins (Table 4.7 and 4.8). These findings may be affected by the low spectral counts across all proteins, so we cannot conclude whether cysteine-rich proteins are differentially expressed. Despite recapitulating the genotype of a patient with severe disease, the mouse model does not appear to be as severely affected as the patient. Interestingly, CARS1 protein expression in the mouse brain (but not liver) was significantly increased in homozygous mutant animals compared to wild-type animals on western blot analysis (Figure 4.20). The explanation for this observation is currently unclear; perhaps there is a feedback mechanism that results in upregulation of the CARS1 protein expression in the brain in response to impaired CARS1 activity. If mutant CARS1 protein is indeed overexpressed in the brain such that there is sufficient tRNA charging capability, it may explain the lack of observable brain phenotypes in the homozygous mutant animals. Another possibility could be that certain cell types express CARS1 at different levels, and the number and proportion of different cell types may vary between mutant and wild-type animals. Future work is needed to understand CARS1 protein expression levels across many tissues (to determine if this is a brain-specific phenomenon) and whether *Cars1* mRNA is differentially expressed in brain and other tissues.

Additionally, we observed that two *CARS1* transcript and protein isoforms (~86 kDa and ~95 kDa) are present in mouse brain and liver (Figures 4.20-4.21). Our data suggest that the isoforms are the result of alternative splicing of exon 2, an in-frame exon that encodes 83 amino acids. While previous work indicated that the isoform containing exon 2 is only expressed in the testis in humans[286], this isoform may be more widely expressed in mouse; further work is needed to understand the expression of this isoform in additional mouse tissues. Furthermore, the functions of the isoform with exon 2 are not clear. It was previously reported that isoforms with and without exon 2 have aminoacylation activity[286]. The isoform with exon 2 was shown to interact with elongation factor-1 (EF-1), and this interaction may facilitate transfer of aminoacylated-tRNAs to the translation elongation machinery[286]. If the isoform with exon 2 has increased ability to supply cysteinylated tRNAs to translating ribosomes due to the EF-1 interaction, this may indicate enhanced capability to translate cysteine codons in the mouse tissues compared to human tissues that do not express this isoform. This may then contribute to the lack of observable brain and liver phenotypes in the mouse model compared to the patients. Since the S688Qfs*2 *Cars1* variant affects both isoforms, it will be important to understand if the interaction between the isoform with exon 2 and EF-1 results in increased translation capability and if the variant affects this interaction.

It is possible that the patient genotype (S688Qfs*2/S688Qfs*2 *CARS1*) may result in differing phenotypes in human and mouse, and that the mouse is not as sensitive to this level of reduced CARS1 function. Therefore, to further repress CARS1 function, we are establishing a mouse model that is heterozygous for a *Cars1* null allele, with plans to study animals that are compound

heterozygous for a null allele and S688Qfs*2 *Cars1*. It is important to note that since *CARS1* is an essential gene, we do not anticipate that homozygosity for two null alleles would be compatible with life. Future work will investigate if this presumably more severe genotype results in a more severe phenotype with detectable effects on translation.

Chapter 5

Conclusions and Future Directions

5.1 Summary

Aminoacyl-tRNA synthetases (ARSs) are ubiquitously expressed, essential enzymes that ligate tRNA molecules with cognate amino acids, the first step in translation[2]. In 2003, variants in an ARS locus were implicated in disease for the first time; variants in glycyl-tRNA synthetase 1 (*GARS1*) were linked to dominant peripheral neuropathy[24]. Since then, variants in all 37 human ARS loci have been implicated in genetic disease[3,30,31]. To date, variants in five ARS loci have been implicated in dominant peripheral neuropathies, and variants in 36 ARS loci have been implicated in recessive disease including overlapping features (*e.g.*, microcephaly and developmental delay)[29]. However, some tissues seem to be particularly sensitive to defects in a particular ARS enzyme and this tissue-predominance is not understood[3]. Additionally, while the number of identified variants is growing, the allelic and clinical spectrum of ARS-associated disease is incomplete. In this thesis, we aimed to address some of these topics and the results are summarized here.

In Chapter 2, we investigated the evidence for pathogenicity of newly identified variants in ARS genes by carefully assessing: (1) segregation of the variants with disease, (2) frequency of the variants in the general population, (3) conservation of the affected residues, and (4) functional

consequences of the variants. This work improved our understanding of ARS-associated phenotypes, the allelic spectrum of ARS-mediated disease, and disease mechanisms.

We studied variants in *AARS1* and *GARS1* associated with dominant disease. Three *AARS1* variants were identified in patients with dominant Charcot-Marie-Tooth (CMT) disease. Two of the variants, R326W and S627L *AARS1*, displayed loss-of-function or partial-loss-of-function effects in yeast complementation assays[38], which is consistent with the majority of previously published CMT-associated ARS variants. However, E337K *AARS1* resulted in increased cell growth in yeast complementation assays and increased tRNA charging in *in vitro* aminoacylation assays[38], indicating hypermorphic effects. How increased aminoacylation could contribute to disease pathogenesis is currently unclear; however, the studies tested the mutant allele in isolation and not in the presence of wild-type enzyme, and the mutant enzyme may behave differently in a dimer. Furthermore, we presented analyses of a *GARS1* variant, I334N, in dominant infantile spinal muscular atrophy (iSMA). We identified two patients with early-onset neuropathy and respiratory distress, and who were both heterozygous for I334N *GARS1*, which displayed loss-of-function effects in yeast complementation assays[95]. This work was consistent with, and contributed to, the growing body of literature suggesting that there is a clinical spectrum of *GARS1*-associated dominant disease that ranges from later-onset, upper-limb predominant peripheral neuropathy to severe, early-onset SMA-like phenotypes[95].

Additionally, we described *AARS2*, *HARS1*, *AARS1*, and *YARS1* variants associated with recessive disease, increasing the allelic and phenotypic heterogeneity of ARS-associated recessive disease. Specifically, our assessments of *AARS2*[52] and *HARS1*[98] variants indicated an

expansion of the phenotypes associated with these genes to include ataxia. The majority of the variants exhibited complete or partial loss-of-function effects in functional assays, which is consistent with previous literature[29] and suggests a partial loss-of-function molecular pathology in ARS-associated recessive disease.

In Chapter 3, we described clinical, genetic, and functional data that for the first time implicated *CARS1* variants in human disease. We identified four individuals from three families with overlapping phenotypes including microcephaly, developmental delay, liver disease, and brittle hair and nails. The patients carried bi-allelic *CARS1* variants that: (1) segregated with disease consistent with recessive disease, (2) were rare and not seen in the homozygous state in the general population, and (3) affected conserved residues[53]. We performed protein expression studies, aminoacylation assays, and yeast complementation assays that indicated that each *CARS1* variant caused a loss-of-function or partial loss-of-function effect[53]. There are multiple conclusions and implications of this work. First, this work contributes to the understanding that mutations in any ARS enzyme can cause human inherited disease, highlighting the importance of this gene family to biological processes and disease. Second, these data are consistent with, and add to, the extensive genetic and functional evidence that point to a partial loss of enzyme function as the molecular mechanism of ARS-mediated recessive disease. Additionally, this work highlighted the importance of collaborations to genetic research; GeneMatcher[250] was utilized to connect researchers and clinicians who had identified patients with *CARS1* variants, which then sparked collaboration among researchers within the ARS field to thoroughly investigate the variants.

Since our data indicated that disease-associated *CARS1* variants cause a loss-of-function effect, $tRNA^{Cys}$ charging is likely impaired in affected patient cells, which may lead to defects in translation that contribute to disease pathogenesis. In Chapter 4, we investigated the downstream consequences of *CARS1* variants on translating cysteine-rich proteins and on global protein synthesis using cell, yeast, and mouse models. Fibroblast cells from patients did not show major changes in gene expression or global translation compared to control cells; however, we cannot rule out the possibility that the cells may behave differently in the context of multicellular tissues in a human body compared to in culture. Additionally, only fibroblasts were tested; other cell types that are relevant to disease pathogenesis, such as neurons or liver cells, may be more responsive to impaired CARS1 activity. In yeast, we developed a dual luciferase assay to investigate the effects of *CARS1* variants on the translation of cysteine-rich sequences. Cysteine-rich sequences were cloned in frame with renilla, and renilla luciferase activity was measured as a proxy for the ability of yeast expressing wild-type or mutant *CRS1*, the yeast ortholog of *CARS1*, to translate the cysteine-rich sequences. Yeast expressing mutant *CRS1* had significantly reduced renilla activity for constructs containing cysteine-rich sequences compared to wild-type *CRS1*-expressing yeast (Figure 4.17). These results suggest that the translation of the cysteine-rich sequences may be impaired in the mutant *CRS1*-expressing yeast and are consistent with our hypothesis that impaired CARS1 activity affects translation of cysteine-rich proteins. Finally, we generated a mouse modeling S688Qfs*2 *CARS1*, a patient mutation. Homozygous mice presented with a growth restriction phenotype consistent with some of the patient phenotypes. Quantitative proteomics on liver and brain revealed very few differentially expressed proteins between mutant and wild-type animals; however, overall low spectral counts complicated interpretations of these data. Future work investigating a mouse that is compound heterozygous

for a null allele and S688Qfs*2 *Cars1* will reveal whether a mouse with a more severe genotype has a more severe phenotype with detectable effects on translation.

In sum, this work expanded the allelic, locus, and clinical heterogeneity of ARS-associated disease and contributed to our understanding of the role of ARS variants in disease. Additionally, these data raised several outstanding questions, which are outlined in this chapter. Parts of this chapter are based on a review article published by the author in *Trends in Genetics* (Volume 36, Issue 2, pages 105-117, February 1, 2020)[3]. Permission was requested for reproduction of the article through the Copyright Clearance Center; as an author of the article, I retain the right to include the article in a dissertation, and permission is not required.

5.2 Outstanding questions and future directions

5.2.1 Understanding the complete allelic and clinical spectrum of ARS-associated disease

Identifying the full panel of pathogenic ARS variants and their associated patient phenotypes would increase our understanding of the molecular mechanisms of ARS-associated disease and improve diagnosis of patients. Three areas of investigation would greatly improve our efforts towards generating this full panel: (1) continued clinical evaluation of previously reported patients with ARS variants; (2) identification of individuals with ARS variants in available databases; and (3) large scale functional assays to determine the functional consequences of all possible ARS variants.

First, thorough and longitudinal clinical phenotyping of all patients identified with ARS variants would expand our knowledge of ARS-associated phenotypes. Reevaluating previously reported

patients may yield insights into later-onset phenotypes and phenotypes that were previously overlooked, such as milder phenotypes in tissues not originally assessed. This would be particularly important if mouse or other models of ARS-associated disease reveal phenotypes not originally detected in patients. Increased phenotyping data would be helpful towards understanding the complete set of phenotypes associated with ARS enzymes, in addition to identifying phenotypes that may be specific to a particular ARS enzyme. These data would not only aid diagnosis of patients but would also expand our knowledge of tissue-predominant phenotypes associated with variants in specific ARS loci. Additionally, relatives of patients who have bi-allelic ARS variants causing recessive disease, and who are carriers of ARS variants should be further evaluated, specifically for peripheral neuropathies using detailed neurological examinations and electrophysiological studies. Even if previous clinical evaluations were performed on carriers indicating no evidence of neurological disease, it is possible that increased age may make such individuals more prone to a late-onset peripheral neuropathy. These data would increase our understanding of which loss-of-function and partial loss-of-function variants can cause dominant disease and which cannot.

Accumulating clinical data from previously reported patients would require significant coordination of effort across clinicians, researchers, and patients and would be limited by many factors such as willingness of patients to participate and ability to contact them. Therefore, other resources should be utilized to learn more about ARS variants and disease, such as large datasets containing clinical and genetic information. A systematic evaluation of all ARS variants identified in biomedical databases and biobanks that contain genetic and clinical information may additionally contribute to our knowledge about phenotypes associated with ARS variants.

The UK Biobank should be prioritized as it contains data from 500,000 participants, including genetic information from whole exome sequencing and whole genome sequencing[288]. Any variants that are identified in individuals with undiagnosed disease should be further assessed for evidence for pathogenicity. This would include determining whether: the variants segregate with disease (when information about relatives is known), the affected residues are conserved, the variants are present in unaffected populations, and the variants have loss-of-function effects.

Finally, determining the functional consequences of ARS variants is critical towards building or refuting arguments of pathogenicity. ARS variants implicated in both dominant and recessive disease display loss-of-function or partial loss-of-function effects and investigating the effects of variants on gene function is a necessary step when evaluating the evidence for pathogenicity of newly identified ARS variants. With increased use of genetic sequencing in the clinical setting, more and more variants of uncertain significance are identified, and the relevance to disease is unclear; thus, determining the effects on function is a bottleneck for ARS variants. Currently, the investigation of the functional consequences of an ARS variant is low-throughput, time-consuming, and labor-intensive, requiring yeast assays and/or aminoacylation experiments for one variant at a time. Large-scale functional studies should be employed to determine the effects of many ARS variants in one experiment. The Antonellis laboratory is currently working with Jacob Kitzman's laboratory to catalog *GARS1* variants with deleterious effects on gene function using a saturation genome editing approach. All possible nucleotide changes are knocked into the endogenous genome of haploid mammalian cells using CRISPR-Cas9 mediated homology-directed repair, and the functional consequences are then determined by evaluating cell viability. In addition to exonic sequences, flanking intronic sequences can be assessed to detect splice-

disruptive mutations. A similar approach could be expanded to all ARS loci to generate a catalog of all possible variants. Then, as variants are identified in patients, clinicians and researchers could combine the clinical and genetic information with these functional data to evaluate the evidence for pathogenicity of variants. Additionally, these data would highlight important sequences and domains of the enzymes that may provide insights into ARS structure and biology.

5.2.2 Defining the molecular mechanisms of ARS-associated dominant disease

Variants in five ARSs have been implicated in dominant peripheral neuropathy, and both dominant-negative and gain-of-function mechanisms have been proposed to explain the peripheral nerve phenotype[29]. The five ARS enzymes associated with dominant neuropathy function as homodimers and charge tRNA in the cytoplasm, and since the variants show loss-of-function effects, a dominant-negative mechanism may explain the phenotype[29]. In a patient cell, the mutant subunits may dimerize with wild-type subunits and reduce aminoacylation. This hypothesis raises questions about whether variants in every dimeric, cytoplasmic ARS can cause dominant neuropathy, or if there are other unique features about the five enzymes implicated to date. One ARS to investigate for a potential link with dominant neuropathy is CARS1, which is a homodimeric, cytoplasmic enzyme[259,266]. Importantly, the recessive phenotypic spectrum we associated with *CARS1* variants includes axonal neuropathy[53]; thus, reduced CARS1 function (caused by carrying two *CARS1* mutant alleles or carrying one dominant negative *CARS1* allele) may lead to peripheral nerve dysfunction. Yeast and *C. elegans* models have been used to accurately predict the pathogenicity of neuropathy-associated ARSs[177] and should be utilized to further investigate additional ARS variants. For example, *CARS1* variants should be engineered

to share characteristics with neuropathy-associated variants at other ARS loci[177]; these characteristics include: (1) conservation among human, worm, and yeast; (2) missense changes; (3) glycine to arginine substitutions; and (4) residues near the amino acid binding pocket of CARS1. Each engineered variant should be tested for loss-of-function effects in yeast complementation assays. Prioritized loss-of-function variants should then be tested for dominant neurotoxicity in *C. elegans* by modeling variants in the endogenous worm ortholog using CRISPR Cas9-mediated mutagenesis and then testing for effects on viability, behavior, and axon morphology[177]. A similar approach could be designed to determine whether variants in monomeric ARS enzymes can cause dominant neuropathy; this is important to investigate since dimerization is a requirement for a dominant-negative mechanism. Additionally, it will be important to use co-immunoprecipitation assays[28] to determine whether all ARS variants implicated in dominant disease impair dimerization. Rebecca Meyer-Schuman in the Antonellis lab is currently testing whether dimer-ablating mutations can rescue cell growth of yeast expressing dominant neuropathy-associated *AARS1* alleles in the presence of wild-type *AARS1*. In the future, these experiments could be expanded to test other ARS variants associated with dominant disease to determine if ablating dimerization rescues yeast growth for all neuropathy-associated ARS variants. These dimerization experiments would be particularly important to test E337K AARS1, which exhibit hypermorphic effects in yeast and *in vitro* aminoacylation assays in isolation[38], but may behave differently in a dimer with wild-type AARS1.

In addition to a dominant-negative mechanism, a gain-of-function mechanism has been proposed stating that variants result in new functions that contribute to disease pathogenesis. For *AARS1*[186], *GARS1*[182,183], and *YARS1*[187], it has been suggested that variants induce conformational

openings that allow aberrant interactions with other proteins that affect pathways in the peripheral nervous system (see Chapter 1). To determine whether variants in all five ARS enzymes result in new aberrant interactions, immunoprecipitation and mass spectrometry studies comparing interacting partners of wild-type and mutant ARSs are needed. Additionally, it will be important to determine if these interactions result in similar effects for multiple variants in a given ARS and for variants in all five of the implicated ARS enzymes.

5.2.3 Investigating the downstream consequences of ARS variants on translation and the tissue-predominant recessive phenotypes associated with ARS variants

Given that genetic and functional data point to a loss of enzyme function as the molecular mechanism of ARS-mediated recessive disease, impaired ARS activity likely leads to impaired translation. Since many ARS-associated recessive phenotypes include overlapping features, an underlying global effect on translation may be the common defect of pathogenic mutations in any ARS enzyme. Our analyses in Chapter 4 indicate that fibroblasts from patients with bi-allelic *CARS1* variants do not exhibit effects on global protein synthesis. However, these experiments were performed on fibroblast cells in culture, and future experiments should investigate global translation in different cell types and in multicellular organisms. In addition to the puromycin incorporation assay employed in this thesis (SUnSET), there are approaches to investigate protein synthesis that should be explored such as fluorescent non-canonical amino acid tagging (FUNCAT), which uses click chemistry to conjugate a fluorophore to nascent peptides with azidohomoalanine (AHA)[289,290]. Translation can then be quantitated using fluorescence-activated cell sorting (FACS)[290], which may reflect a more accurate measurement of, and detect more subtle differences in, translation than western blot approaches.

While variants in cytoplasmic ARSs cause recessive phenotypes often affecting a wide array of tissues, these variants have also been linked with tissue-predominant effects associated with variants in a particular ARS enzyme. One potential explanation for the observed tissue predominance of ARS-associated recessive disease is that tissues may have different demands for different amino acids. For example, patients with *CARS1* variants present with brittle hair and nails[53], which both express cysteine-rich keratins and keratin-associated proteins. Therefore, if $tRNA^{Cys}$ charging with cysteine is impaired, hair and nails may be more susceptible to aminoacylation defects because of the high cysteine demand for keratin translation. Data in this thesis start to address this hypothesis, but further investigation is needed to fully elucidate whether translation of cysteine-rich proteins is specifically impaired. We developed a dual luciferase assay in yeast to test translation of cysteine-rich sequences, and our current efforts involve improving the assay by knocking in variants into the endogenous *CRS1* locus instead of expressing *CRS1* from a plasmid. Additional sequences with varying amounts of cysteine should be tested to determine whether there is a threshold at which translation is impaired and/or if there is an association between cysteine content and the reduction in translation. Furthermore, we performed quantitative proteomics studies using label-free quantitative mass spectrometry on brain and liver tissue samples in mice homozygous for the disease-associated *Cars1* variant, S688Qfs*2; however, low spectral counts limited our ability to determine whether expression of cysteine-rich proteins were reduced. Additionally, the mouse model showed no obvious phenotypes in the relevant tissues, suggesting that the variants may not affect mouse physiology the same ways they affect humans. Ongoing work in the Antonellis laboratory is investigating whether a mouse compound heterozygous for S688Qfs*2 and a null allele will result in a more

severe phenotype with detectable effects on translation. If a more severe phenotype is observed, targeted proteomics approaches[291] should be employed to specifically assess cysteine-rich protein expression in affected tissues with high sensitivity.

Two additional experiments would be helpful towards testing the hypothesis that translation of cysteine-rich proteins is affected. First, ribosome profiling studies[292] would be useful to determine if ribosomes are stalling at cysteine codons, in particular along transcripts with many cysteine codons. Additionally, our analyses revealed a notable group of proteins that have high cysteine content: metallothioneins (MTs), which are metal-binding proteins that function in detoxification of heavy metals such as cadmium and copper[293]. If MTs are affected by disease associated *CARS1* variants, patients may be more susceptible to heavy metal toxicity. Experiments in relevant models with depleted CARS1 function are needed to evaluate metallothionein protein levels and to test for differential sensitivity to heavy metal exposure. For example, in the *Cars1* mouse model, mice could be exposed to copper and then evaluated for any effects on survival, induction of metallothionein expression using western blot analysis, and degeneration in the brain using histopathology. Any relevant findings of enhanced sensitivity should be communicated to clinicians who can inform patients and families.

An additional possibility to explain tissue-specific phenotypes associated with ARS variants relates to the availability of tRNA[194,196] and amino acids[197,198], which may vary among tissues—due to differences in pathways including biosynthesis and salvage—and influence the cellular response to impaired ARS activity. For example, a mutant CARS1 enzyme with reduced activity would likely charge more $tRNA^{Cys}$ molecules in a tissue with abundant $tRNA^{Cys}$ and cysteine

compared to a tissue with low availability of these molecules. Studies investigating tissue-specific $tRNA^{Cys}$ expression and cysteine abundance are needed to address this hypothesis. First, to quantify tRNA abundance, tRNA sequencing techniques[294,295] or tRNA microarrays[194] could be applied to different tissues, which would indicate if patient phenotypes occur in tissues with low levels of $tRNA^{Cys}$. Second, amino acid content in different tissues could be determined using chromatography[153,296] to determine if patient phenotypes occur in tissues with low levels of cysteine.

A final consideration regarding the mechanisms behind tissue-predominant ARS-associated phenotypes is that secondary functions have been reported for some ARSs. More basic science research is needed to uncover secondary functions of ARSs and to determine how these functions may relate to ARS-associated disease. There are two relevant areas of investigation related to CARS1 secondary functions: (1) CARS1 functions in cysteine polysulfidation[269] and ferroptosis[270]; and (2) alternative splicing results in two widely expressed CARS1 protein isoforms that contain different C-terminal amino-acid sequences (see Chapter 3 for further discussion of these isoforms)[259]. Studies measuring cysteine hydropersulfide[297] in relevant patient tissues would indicate if cysteine polysulfidation is affected by impaired CARS1. Additionally, knockdown of *CARS1* has been reported to inhibit ferroptosis induced by cysteine deprivation and to induce the transsulfuration pathway[270]; studies investigating whether cystathionine, a transsulfuration pathway metabolite, accumulates in the environment of pathogenic *CARS1* variants would be informative. To further investigate the two widely expressed CARS1 protein isoforms and their potential functions, subcellular fractionation[298] and fluorescence imaging[299] experiments would indicate if the isoforms have different subcellular

localizations. Co-immunoprecipitation and mass spectrometry experiments[300] would also indicate if the isoforms have differential binding partners. Additionally, the isoforms were reported to be differentially expressed (80% isoform 1 vs. 20% isoform 2) at the RNA level in osteosarcoma cells[259]; further work is needed to determine whether differential isoform expression is consistent across tissues and whether this correlates with protein expression levels. By defining secondary functions of ARS enzymes, the community may then determine if—and how—pathogenic variants affect these activities.

5.2.4 Understanding phenotypic variability associated with variants in the same ARS gene

Different variants in the same gene can present with different phenotypes, such as the spectrum of *GARS1*-associated dominant phenotypes that includes later-onset upper limb predominant peripheral neuropathy and early-onset SMA-like phenotypes[95]. Differing levels of impairment of tRNA charging may contribute the wide array of phenotype severity, and there may be other genetic and/or environmental factors impacting the phenotypes. Careful assessment of genomic sequencing data from patients with *GARS1*-associated dominant phenotypes may reveal variants at other loci that act as genetic modifiers and alter the phenotypic outcome of the *GARS1* variants. Given the limited number of patients and available sequencing data, the ability to identify and prioritize such variants may be limited. Therefore, modifier screens in model organisms may also be useful approaches to investigate genes and pathways that are important for *GARS1* function and disease. Ongoing work by Sheila Marte, a graduate student in the Antonellis lab, is focused on developing screens to identify genes that alter the viability of yeast expressing mutant *GARS1*. Results from these experiments will provide insights into disease pathogenesis and may suggest potential therapeutic targets.

Additionally, individuals with the same ARS variants may present with variable phenotypes. In Chapter 2 we described two siblings (a brother and a sister) who were both compound heterozygous for the same variants in *AARS2* (F131del and I328M), and both presented with an ataxia phenotype. However, in addition to impaired coordination, the brother also had vision problems and cognitive impairment[52]. Additionally, the brother developed symptoms in childhood, while the sister presented at age 26[52]. The difference in their presentations is not understood and may reflect environmental and/or genetic contributions. Importantly, the variants identified in the brother were revealed via commercial genetic testing specifically for *AARS2*, and he may have additional variants in other genes that are not present in his sister and that contribute to his disease. Though our data indicate that his AARS2-related ataxia phenotype is recessive, his vision and cognitive impairment phenotypes may be dominant and due to a *de novo* mutation in a different gene; dominant developmental disorders caused by *de novo* mutations have a prevalence of 1 in 213 to 1 in 448 births[301]. The phenotypes could also be recessive and caused by bi-allelic mutations in another gene (and his sister could be a carrier or have no variants in that gene). To determine whether his cognitive impairment phenotype may be caused by variants outside of *AARS2*, sequencing a panel of genes associated with intellectual disability could be useful. Exome or genome sequencing may also reveal additional variants and could be compared with the sequencing data from his sister to gain insight into genetic differences between the siblings. Understanding the mechanisms underlying the differences in phenotype severity may yield insights into disease pathogenesis, patient prognosis, and potential therapeutic targets.

5.2.5 Developing therapies for ARS-mediated disease

In ARS-mediated dominant disease, the mutant allele is dominantly toxic; therefore, one avenue for developing treatments could be to target the expression of the mutant allele and reduce or ablate expression of the allele while leaving the wild-type allele unaffected. Importantly, heterozygosity for one null allele does not cause disease[29], demonstrating that the activity from one wild-type allele is sufficient. Reducing expression of the mutant allele has recently been tested in a mouse model of *GARS1*-associated peripheral neuropathy modeling a variant (ΔETAQ) associated with severe peripheral neuropathy[96]. Allele-specific RNAi sequences reduced mutant *Gars1* transcripts in mice and resulted in prevention of the onset of peripheral neuropathy when the mice were treated at birth: treatment after the onset of symptoms was also beneficial[96]. Another promising strategy for reducing expression of allele-specific transcripts involves utilizing antisense oligonucleotides (ASOs), which have been developed for neurological disorders like spinal muscular atrophy[302]. ASOs can act through multiple mechanisms to decrease translation of an mRNA, such as binding RNA and targeting it for degradation or binding splice sites to skip exons[302]. These data suggest that RNAi-based and ASO gene therapies could be treatment options and should also be explored with other ARS variants associated with dominant disease. For both approaches, it is critical that the therapy only reduce expression of the mutant transcript and leaves the wild-type allele unaffected; if both alleles are down regulated, aminoacylation would be further impaired and may lead to worse disease. If the molecular mechanism for ARS-associated dominant disease is a dominant negative mechanism, then increasing the amount of wild-type enzyme should improve the patient phenotype due to increased wild-type-wild-type dimers that are fully functional. Further work is needed to investigate the effects of overexpressing wild-type ARSs in dominant disease models;

Jennifer Pierluissi in the Antonellis laboratory is currently testing this hypothesis in a *C. elegans* model of GARS1-mediated dominant disease.

In ARS-mediated recessive disease, extensive evidence suggests reduced function is the molecular mechanism of disease pathogenesis; therefore, increasing the amount of wild-type enzyme should improve the phenotype. In support of this, overexpression of wild-type *GARS1* rescued the reduced viability of mice that were homozygous or compound heterozygous for *Gars1* mutations[179]. Further work is needed to determine whether rescue can be achieved in models of additional ARS variants, including the *Cars1* mouse model. Additionally, some patients are bi-allelic for ARS variants that reduce ARS activity but do not lead to the complete ablation of ARS function. Because of this, each patient has at least one hypomorphic allele that retains some activity. Therefore, increasing the expression of a hypomorphic allele may increase the ARS function enough to improve the phenotype. This may be tested by overexpressing the mutant mRNAs in relevant models and assessing phenotypic improvement. Furthermore, antisense oligos (ASOs) have been used to increase expression of mRNAs with upstream open reading frames (uORFs); multiple ARSs have reported uORFs[303,304]. uORFs reduce translation from the downstream primary ORF, so ASOs binding to the uORFs lead to increased translation from the downstream primary ORF[302,305,306].

Increasing the availability of other components of the aminoacylation reaction, such as amino acids and tRNA, also warrant further investigation as interventions for ARS-associated recessive disease. As mentioned above, the concentration of tRNA[194,196] and amino acids[197,198] differ among tissues. Amino acids and tRNA are critical substrates of the aminoacylation reaction, and

increased supply may increase the rate of the reaction and subsequently lead to clinical improvement. In fact, amino acid supplementation has already been tested in a few patients. Two brothers who were homozygous for *MARS1* variants presented with pulmonary alveolar proteinosis and interstitial lung and liver disease, including multiorgan failure and respiratory insufficiency requiring six whole lung lavages in one brother[119]. Both brothers were started on oral methionine supplement and high protein intake, which led to complete respiratory recovery and improvement in growth and neurodevelopmental state[119]. Additionally, a patient with bi-allelic *IARS1* variants who presented with growth restriction and infantile hepatopathy had improved development after isoleucine supplementation[102]. *In vitro* aminoacylation assays testing mutant ARS enzymes with varying amounts of amino acid substrates should be performed to determine if tRNA charging is improved with increasing amino acids. Yeast models could also be utilized to determine whether increased amino acid concentration in media improves cell growth of yeast expressing mutant ARS genes. Further investigation, including controlled clinical trials, is needed to understand whether amino acid supplementation may be a viable therapy option for all patients with ARS-associated recessive disease and to determine the appropriate dosages with which to treat patients, since some amino acids are toxic at high levels[307,308].

5.3 Concluding remarks

In this thesis, we present studies implicating newly identified ARS variants in human disease, expanding the clinical, locus, and allelic spectrum of ARS-associated disease. The data implicated variants in *CARS1* in disease for the first time and further investigated the downstream consequences of these variants on translation. This work raised several additional

questions and opportunities for further study of the mechanisms of ARS-related disease and potential treatments. The studies presented here will provide insight into the pathogenic mechanisms of ARS variants and may lead to potential avenues for therapeutic development.

Appendix

Table A.1. Gateway cloning primers for yeast complementation constructs

Primer	Sequence
AARS2 ORF GW F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCAGCGTCAGTGGCAGCTGC
AARS2 ORF GW R	GGGGACCACTTTGTACAAGAAAGCTGGGTTCAGAGCTGGCTGAGGGCATAGG
CARS GW F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCAGATTCCTCCGGGCAGC
CARS GW R2	GGGGACCACTTTGTACAAGAAAGCTGGGTTCACTGGAAGCTTCCATTCTGGGC
CRS1 GW F2	GGGGACAAGTTTGTACAAAAAAGCAGGCTAAAAGGTTGATAACGACTGCAAT
CRS1 GW R2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCCCTTTTCTCACAAGCGC

Table A.2. Mutagenesis primers

Primer	Sequence
R326W AARS mutF	GGATATGTGTTGAGATGGATTCTCCGCCGAG
R326W AARS mutR	CTCGGCGGAGAATCCATCTCAACACATATCC
E337K AARS mutF	GTCCGATACGCCCATAAAAAGCTCAATGCCA
E337K AARS mutR	TGGCATTGAGCTTTTTATGGGCGTATCGGAC
S627L AARS mutF	CTGACCAGAAAGGCTTATTGGTTGCTCCTGA
S627L AARS mutR	TCAGGAGCAACCAATAAGCCTTTCTGGTCAG
C115R AARS mutF	TTTAAGGAATTGGCACGTAAGATGGCTCTGG
C115R AARS mutR	CCAGAGCCATCTTACGTGCCAATTCCTTAAA
R750W AARS mutF	ATTGCCAAGGGTATCTGGAGGATTGTGGCTG
R750W AARS mutR	CAGCCACAATCCTCCAGATACCCTTGGCAAT
Y690Lfs*2 AARS mutF	GTGTTTGATGAGACCCTATCCTGACCCTGTG
Y690Lfs*2 AARS mutR	CACAGGGTCAGGATAGGGTCTCATCAAACAC
G913D AARS mutF	ATGCAGCCAATCGGGACTTAAAAGCCAGCGA
G913D AARS mutR	TCGCTGGCTTTTAAGTCCCGATTGGCTGCAT
K337E scALA1 mutF	GCCCGTTACGCCCGTGAATACATGAATTACC
K337E scALA1 mutR	GGTAATTCATGTATTCACGGGCGTAACGGGC
I334N GARS seq F	CTTTTAGAAATGAGAACTCCCCTCGATCTGG
I334N GARS seq F	CCAGATCGAGGGGAGTTCTCATTTCTAAAAG
F102del ALA1 mutF	TATCATCATACCTTTGAAATGCTGGGTAAC
F102del ALA1 mutR	GTTACCCAGCATTTCAAAGGTATGATGATA
V306M ALA1 mutF	GTTTTAGCTGATCATATGCGTACATTGACTTTT
V306M ALA1 mutR	AAAAGTCAATGTACGCATATGATCAGCTAAAAC
F98del AARS mutF	TATCATCACACCTTCGAGATGCTGGGCTCT
F98del AARS mutR	AGAGCCCAGCATCTCGAAGGTGTGATGATA
I302M AARS mutF	GTGCTGGCTGACCACATGCGGACCATCACTGTG
I302M AARS mutR	CACAGTGATGGTCCGCATGTGGTCAGCCAGCAC
D206Y HARS mutF	TTCACTTCAGATAGGCTACTTCCTGGTCAAGG
D206Y HARS mutR	CCTTGACCAGGAAGTAGCCTATCTGAAGTGAA
I465L HARS mutF	ATCCCACTGGTGGCTCTCATCGGCGAGCAGG
I465L HARS mutR	CCTGCTCGCCGATGAGAGCCACCAGTGGGAT
L305dup HARS mutF	GGAGACCTGAAGTTGTTGCTCTTTGAGTACCTG
L305dup HARS mutR	CAGGTACTCAAAGAGCAACAACTTCAGGTCTCC
V244Cfs*6 HARS mutF	GACAAGCTGGACAAGTGTCCTGGGAAGAGG
V244Cfs*6 HARS mutR	CCTCTTCCCAGGACACTTGTCCAGCTTGTC
I59T YARS mutF	TGCCCATGTCAAAGACTGCAGACTTCTTAAA
I59T YARS mutR	TTTAAGAAGTCTGCAGTCTTTGACATGGGCA
R341H CARS mutF	CTTGGGGAAAGGGTCATCCGGGCTGGCATAT
R341H CARS mutR	ATATGCCAGCCCGGATGACCCTTTCCCCAAG
S359L CARS mutF	CCCTCCTAGGGGCTTTGATGGACATTCACGG
S359L CARS mutR	CCGTGAATGTCCATCAAAGCCCCTAGGAGGG
Q380* CARS mutF	GACAATGAGCTGGCATAGTCGGAGGCCTACT
Q380* CARS mutR	AGTAGGCCTCCGACTATGCCAGCTCATTGTC
L400Q CARS mutF	TGCACACAGGCCACCAGACCATTGCAGGCTG
L400Q CARS mutR	CAGCCTGCAATGGTCTGGTGGCCTGTGTGCA
S688Qfs*2 CARS mutF	ATGAAGATTCCCCCCCAGTGAGATGTTCTTG
S688Qfs*2 CARS mutR	CAAGAACATCTCACTGGGGGGGAATCTTCAT
CARS alt mut F	CTCCAAGTTTGATGAAAATGTAAGCATGGTCTGCCCACACATGACATG
CARS alt mut R	CATGTCATGTGTGGGCAGACCATGCTTACATTTTCATCAAACTTGGAG
D714* CRS1 mutF	CAAATTAAACCACAATAAATGTTCAAGGATGTC
D714* CRS1 mutR	GACATCCTTGAACATTTATTGTGGTTTAATTTG

Table A.3. Primers to amplify around *HARS1* variants for Sanger sequencing

Primer	Sequence
I465L HARS seq F	GTCCCTGAGGACAGCAGAAG
I465L HARS seq R	TCAATGCCAGGCTTTTTCTC
L305dup HARS seq F	CCTTGTCAGCCTCACTCACA
L305dup HARS seq R	CAAGGCTCAGGTCAAAGGAG
D206Y HARS seq F	CCCCCTTCAGAAAGGTGAAT
D206Y HARS seq R	TACCTGCAAGGAACCAATGC
V244Cfs*6 HARS seq F	TAGGCTCATGGGCAGCTACT
V244Cfs*6 HARS seq R	TTTAGGACTGAGGGCAGTGG

Table A.4. Primers to amplify around *CARS1* variants for Sanger sequencing

Primer	Sequence
CARS_Intron_10_F	GTCTTTTCATGTGCGGGATT
CARS_Intron_11_R	CCCGATGAAGGTGTCAAGTT
CARS_Intron_11_F	CTGTGGACAGCACCAACAAG
CARS_Intron_12_R	CCCTTATTCTCCCGAGTGCT
CARS_22F	CAAGTTGAGGGCTGGAGGTA
CARS_22R	TGAAAATTCCTGTCCTCCCAC

Table A.5. RT-PCR primers to amplify around alternative splicing in *CARS1*

Primer	**Sequence**
CAS1F	CTTGTCAGAAACCGACAAATACTCC
CAS1R	CTTTGCCCTCCATGTCATGTGTGG
hsCARS exon 22 F	GCAAAGCTGGCCAAGATGAA
hsCARS 3' UTR R	AACATGATAGGAGCGCTGGG
mmCars around last intron F	GGAGGCTGCCATTCTGTAAC
mmCars around last intron R	AAAGCTGGCCAAGATGAAGA
mmCars Exon 3F	TGTTCCGGGTGAGGCTATTA
mmCars Exon 1R	GGGTCTTGATTCCCAGTGAT

Bibliography

1 Craig, N. L. *Molecular biology : principles of genome function*. (Oxford University Press, 2010).

2 Antonellis, A. & Green, E. D. The role of aminoacyl-tRNA synthetases in genetic diseases. *Annu Rev Genomics Hum Genet* **9**, 87-107, doi:10.1146/annurev.genom.9.081307.164204 (2008).

3 Kuo, M. E. & Antonellis, A. Ubiquitously Expressed Proteins and Restricted Phenotypes: Exploring Cell-Specific Sensitivities to Impaired tRNA Charging. *Trends Genet* **36**, 105-117, doi:10.1016/j.tig.2019.11.007 (2020).

4 Gebauer, F. & Hentze, M. W. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5**, 827-835, doi:10.1038/nrm1488 (2004).

5 Donnelly, N., Gorman, A. M., Gupta, S. & Samali, A. The eIF2alpha kinases: their structures and functions. *Cell Mol Life Sci* **70**, 3493-3511, doi:10.1007/s00018-012-1252-6 (2013).

6 Kapur, M. & Ackerman, S. L. mRNA Translation Gone Awry: Translation Fidelity and Neurological Disease. *Trends Genet* **34**, 218-231, doi:10.1016/j.tig.2017.12.007 (2018).

7 Sasikumar, A. N., Perez, W. B. & Kinzy, T. G. The many roles of the eukaryotic elongation factor 1 complex. *Wiley Interdiscip Rev RNA* **3**, 543-555, doi:10.1002/wrna.1118 (2012).

8 Ling, J., Reynolds, N. & Ibba, M. Aminoacyl-tRNA synthesis and translational quality control. *Annu Rev Microbiol* **63**, 61-78, doi:10.1146/annurev.micro.091208.073210 (2009).

9 Da Costa, L., Narla, A. & Mohandas, N. An update on the pathogenesis and diagnosis of Diamond-Blackfan anemia. *F1000Res* **7**, doi:10.12688/f1000research.15542.1 (2018).

10 Da Costa, L. *et al.* Molecular approaches to diagnose Diamond-Blackfan anemia: The EuroDBA experience. *Eur J Med Genet* **61**, 664-673, doi:10.1016/j.ejmg.2017.10.017 (2018).

11 Leegwater, P. A. *et al.* Subunits of the translation initiation factor eIF2B are mutant in leukoencephalopathy with vanishing white matter. *Nat Genet* **29**, 383-388, doi:10.1038/ng764 (2001).

12 Ribas de Pouplana, L. & Schimmel, P. Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. *Cell* **104**, 191-193, doi:10.1016/s0092-8674(01)00204-5 (2001).

13 Mathews, M., Sonenberg, N. & Hershey, J. W. B. *Translational control in biology and medicine*. 3rd edn, (Cold Spring Harbor Laboratory Press, 2007).

14 Pak, D., Kim, Y. & Burton, Z. F. Aminoacyl-tRNA synthetase evolution and sectoring of the genetic code. *Transcription* **9**, 205-224, doi:10.1080/21541264.2018.1467718 (2018).

15 Giege, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* **26**, 5017-5035, doi:10.1093/nar/26.22.5017 (1998).

16 Boczonadi, V. *et al.* Mutations in glycyl-tRNA synthetase impair mitochondrial metabolism in neurons. *Hum Mol Genet* **27**, 2187-2204, doi:10.1093/hmg/ddy127 (2018).

17 Tolkunova, E., Park, H., Xia, J., King, M. P. & Davidson, E. The human lysyl-tRNA synthetase gene encodes both the cytoplasmic and mitochondrial enzymes by means of an unusual alternative splicing of the primary transcript. *J Biol Chem* **275**, 35063-35069, doi:10.1074/jbc.M006265200 (2000).

18 Echevarria, L. *et al.* Glutamyl-tRNAGln amidotransferase is essential for mammalian mitochondrial translation in vivo. *Biochem J* **460**, 91-101, doi:10.1042/BJ20131107 (2014).

19 Guo, M. & Schimmel, P. Essential nontranslational functions of tRNA synthetases. *Nat Chem Biol* **9**, 145-153, doi:10.1038/nchembio.1158 (2013).

20 Kim, S., You, S. & Hwang, D. Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nat Rev Cancer* **11**, 708-718, doi:10.1038/nrc3124 (2011).

21 Guo, M., Schimmel, P. & Yang, X. L. Functional expansion of human tRNA synthetases achieved by structural inventions. *FEBS Lett* **584**, 434-442, doi:10.1016/j.febslet.2009.11.064 (2010).

22 Wakasugi, K. *et al.* A human aminoacyl-tRNA synthetase as a regulator of angiogenesis. *Proc Natl Acad Sci U S A* **99**, 173-177, doi:10.1073/pnas.012602099 (2002).

23 Lo, W. S. *et al.* Human tRNA synthetase catalytic nulls with diverse functions. *Science* **345**, 328-332, doi:10.1126/science.1252943 (2014).

24 Antonellis, A. *et al.* Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. *Am J Hum Genet* **72**, 1293-1299, doi:10.1086/375039 (2003).

25 Jordanova, A. *et al.* Disrupted function and axonal distribution of mutant tyrosyl-tRNA synthetase in dominant intermediate Charcot-Marie-Tooth neuropathy. *Nat Genet* **38**, 197-202, doi:10.1038/ng1727 (2006).

26 Latour, P. *et al.* A major determinant for binding and aminoacylation of tRNA(Ala) in cytoplasmic Alanyl-tRNA synthetase is mutated in dominant axonal Charcot-Marie-Tooth disease. *Am J Hum Genet* **86**, 77-82, doi:10.1016/j.ajhg.2009.12.005 (2010).

27 Vester, A. *et al.* A loss-of-function variant in the human histidyl-tRNA synthetase (HARS) gene is neurotoxic in vivo. *Hum Mutat* **34**, 191-199, doi:10.1002/humu.22210 (2013).

28 Tsai, P. C. *et al.* A recurrent WARS mutation is a novel cause of autosomal dominant distal hereditary motor neuropathy. *Brain* **140**, 1252-1266, doi:10.1093/brain/awx058 (2017).

29 Meyer-Schuman, R. & Antonellis, A. Emerging mechanisms of aminoacyl-tRNA synthetase mutations in recessive and dominant human disease. *Hum Mol Genet* **26**, R114-R127, doi:10.1093/hmg/ddx231 (2017).

30 Manole, A. *et al.* De Novo and Bi-allelic Pathogenic Variants in NARS1 Cause Neurodevelopmental Delay Due to Toxic Gain-of-Function and Partial Loss-of-Function Effects. *Am J Hum Genet* **107**, 311-324, doi:10.1016/j.ajhg.2020.06.016 (2020).

31 Krenke, K. *et al.* FARSA mutations mimic phenylalanyl-tRNA synthetase deficiency caused by FARSB defects. *Clin Genet* **96**, 468-472, doi:10.1111/cge.13614 (2019).

32 Simons, C. *et al.* Loss-of-function alanyl-tRNA synthetase mutations cause an autosomal-recessive early-onset epileptic encephalopathy with persistent myelination defect. *Am J Hum Genet* **96**, 675-681, doi:10.1016/j.ajhg.2015.02.012 (2015).

33 Nakayama, T. *et al.* Deficient activity of alanyl-tRNA synthetase underlies an autosomal recessive syndrome of progressive microcephaly, hypomyelination, and epileptic encephalopathy. *Hum Mutat* **38**, 1348-1354, doi:10.1002/humu.23250 (2017).

34 Marten, L. M. *et al.* Recurrent acute liver failure in alanyl-tRNA synthetase-1 (AARS1) deficiency. *Mol Genet Metab Rep* **25**, 100681, doi:10.1016/j.ymgmr.2020.100681 (2020).

35 McLaughlin, H. M. *et al.* A recurrent loss-of-function alanyl-tRNA synthetase (AARS) mutation in patients with Charcot-Marie-Tooth disease type 2N (CMT2N). *Hum Mutat* **33**, 244-253, doi:10.1002/humu.21635 (2012).

36 Lin, K. P. *et al.* The mutational spectrum in a cohort of Charcot-Marie-Tooth disease type 2 among the Han Chinese in Taiwan. *PLoS One* **6**, e29393, doi:10.1371/journal.pone.0029393 (2011).

37 Motley, W. W. *et al.* A novel AARS mutation in a family with dominant myeloneuropathy. *Neurology* **84**, 2040-2047, doi:10.1212/WNL.0000000000001583 (2015).

38 Weterman, M. A. J. *et al.* Hypermorphic and hypomorphic AARS alleles in patients with CMT2N expand clinical and molecular heterogeneities. *Hum Mol Genet* **27**, 4036-4050, doi:10.1093/hmg/ddy290 (2018).

39 Zhao, Z. *et al.* Alanyl-tRNA synthetase mutation in a family with dominant distal hereditary motor neuropathy. *Neurology* **78**, 1644-1649, doi:10.1212/WNL.0b013e3182574f8f (2012).

40 Lynch, D. S. *et al.* Analysis of Mutations in AARS2 in a Series of CSF1R-Negative Patients With Adult-Onset Leukoencephalopathy With Axonal Spheroids and Pigmented Glia. *JAMA Neurol* **73**, 1433-1439, doi:10.1001/jamaneurol.2016.2229 (2016).

41 Lee, J. M. *et al.* Two Korean siblings with recently described ovarioleukodystrophy related to AARS2 mutations. *Eur J Neurol* **24**, e21-e22, doi:10.1111/ene.13245 (2017).

42 Hamatani, M. *et al.* The first Japanese case of leukodystrophy with ovarian failure arising from novel compound heterozygous AARS2 mutations. *J Hum Genet* **61**, 899-902, doi:10.1038/jhg.2016.64 (2016).

43 Dallabona, C. *et al.* Novel (ovario) leukodystrophy related to AARS2 mutations. *Neurology* **82**, 2063-2071, doi:10.1212/WNL.0000000000000497 (2014).

44 Szpisjak, L. *et al.* Novel AARS2 gene mutation producing leukodystrophy: a case report. *J Hum Genet* **62**, 329-333, doi:10.1038/jhg.2016.126 (2017).

45 Tang, Y. *et al.* AARS2 leukoencephalopathy: A new variant of mitochondrial encephalomyopathy. *Mol Genet Genomic Med* **7**, e00582, doi:10.1002/mgg3.582 (2019).

46 Wang, J. Y. *et al.* A homozygous mutation of alanyl-transfer RNA synthetase 2 in a patient of adult-onset leukodystrophy: A case report and literature review. *Brain Behav* **9**, e01313, doi:10.1002/brb3.1313 (2019).

47 Gotz, A. *et al.* Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. *Am J Hum Genet* **88**, 635-642, doi:10.1016/j.ajhg.2011.04.006 (2011).

48 Nielsen, S. K. *et al.* Recessive Inheritance of a Rare Variant in the Nuclear Mitochondrial Gene for AARS2 in Late-Onset Dilated Cardiomyopathy. *Circ Genom Precis Med* **13**, 560-562, doi:10.1161/CIRCGEN.120.003086 (2020).

49 Taylor, R. W. *et al.* Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. *JAMA* **312**, 68-77, doi:10.1001/jama.2014.7184 (2014).

50 Peragallo, J. H., Keller, S., van der Knaap, M. S., Soares, B. P. & Shankar, S. P. Retinopathy and optic atrophy: Expanding the phenotypic spectrum of pathogenic variants in the AARS2 gene. *Ophthalmic Genet* **39**, 99-102, doi:10.1080/13816810.2017.1350723 (2018).

51 Kiraly-Borri, C. *et al.* Siblings with lethal primary pulmonary hypoplasia and compound heterozygous variants in the AARS2 gene: further delineation of the phenotypic spectrum. *Cold Spring Harb Mol Case Stud* **5**, doi:10.1101/mcs.a003699 (2019).

52 Kuo, M. E., Antonellis, A. & Shakkottai, V. G. Alanyl-tRNA Synthetase 2 (AARS2)-Related Ataxia Without Leukoencephalopathy. *Cerebellum*, doi:10.1007/s12311-019-01080-y (2019).

53 Kuo, M. E. *et al.* Cysteinyl-tRNA Synthetase Mutations Cause a Multi-System, Recessive Disease That Includes Microcephaly, Developmental Delay, and Brittle Hair and Nails. *Am J Hum Genet* **104**, 520-529, doi:10.1016/j.ajhg.2019.01.006 (2019).

54 Coughlin, C. R., 2nd *et al.* Mutations in the mitochondrial cysteinyl-tRNA synthase gene, CARS2, lead to a severe epileptic encephalopathy and complex movement disorder. *J Med Genet* **52**, 532-540, doi:10.1136/jmedgenet-2015-103049 (2015).

55 Hallmann, K. *et al.* A homozygous splice-site mutation in CARS2 is associated with progressive myoclonic epilepsy. *Neurology* **83**, 2183-2187, doi:10.1212/WNL.0000000000001055 (2014).

56 Taft, R. J. *et al.* Mutations in DARS cause hypomyelination with brain stem and spinal cord involvement and leg spasticity. *Am J Hum Genet* **92**, 774-780, doi:10.1016/j.ajhg.2013.04.006 (2013).

57 Wolf, N. I. *et al.* DARS-associated leukoencephalopathy can mimic a steroid-responsive neuroinflammatory disorder. *Neurology* **84**, 226-230, doi:10.1212/WNL.0000000000001157 (2015).

58 Lan, M. Y., Chang, Y. Y., Yeh, T. H., Lin, T. K. & Lu, C. S. Leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL) with a novel DARS2 mutation and isolated progressive spastic paraparesis. *J Neurol Sci* **372**, 229-231, doi:10.1016/j.jns.2016.11.058 (2017).

59 Kohler, C. *et al.* Early-onset leukoencephalopathy due to a homozygous missense mutation in the DARS2 gene. *Mol Cell Probes* **29**, 319-322, doi:10.1016/j.mcp.2015.06.005 (2015).

60 Isohanni, P. *et al.* DARS2 mutations in mitochondrial leucoencephalopathy and multiple sclerosis. *J Med Genet* **47**, 66-70, doi:10.1136/jmg.2009.068221 (2010).

61 Martikainen, M. H., Ellfolk, U. & Majamaa, K. Impaired information-processing speed and working memory in leukoencephalopathy with brainstem and spinal cord involvement and elevated lactate (LBSL) and DARS2 mutations: a report of three adult patients. *J Neurol* **260**, 2078-2083, doi:10.1007/s00415-013-6940-0 (2013).

62 Sharma, S. *et al.* Leukoencephalopathy with brain stem and spinal cord involvement and high lactate: a genetically proven case without elevated white matter lactate. *J Child Neurol* **26**, 773-776, doi:10.1177/0883073810390695 (2011).

63 Lin, J. *et al.* Leukoencephalopathy with brainstem and spinal cord involvement and normal lactate: a new mutation in the DARS2 gene. *J Child Neurol* **25**, 1425-1428, doi:10.1177/0883073810370897 (2010).

64 Yelam, A., Nagarajan, E., Chuquilin, M. & Govindarajan, R. Leucoencephalopathy with brain stem and spinal cord involvement and lactate elevation: a novel mutation in the DARS2 gene. *BMJ Case Rep* **12**, doi:10.1136/bcr-2018-227755 (2019).

65 Yahia, A. *et al.* Intra-familial phenotypic heterogeneity in a Sudanese family with DARS2-related leukoencephalopathy, brainstem and spinal cord involvement and lactate elevation: a case report. *BMC Neurol* **18**, 175, doi:10.1186/s12883-018-1180-7 (2018).

66 Gungor, O. *et al.* A compound heterozygous EARS2 mutation associated with mild leukoencephalopathy with thalamus and brainstem involvement and high lactate (LTBL). *Brain Dev* **38**, 857-861, doi:10.1016/j.braindev.2016.04.002 (2016).

67 Biancheri, R. *et al.* Expanding the Clinical and Magnetic Resonance Spectrum of Leukoencephalopathy with Thalamus and Brainstem Involvement and High Lactate (LTBL) in a Patient Harboring a Novel EARS2 Mutation. *JIMD Rep* **23**, 85-89, doi:10.1007/8904_2015_434 (2015).

68 Oliveira, R. *et al.* Lethal Neonatal LTBL Associated with Biallelic EARS2 Variants: Case Report and Review of the Reported Neuroradiological Features. *JIMD Rep* **33**, 61-68, doi:10.1007/8904_2016_581 (2017).

69 Sahin, S. *et al.* Leukoencephalopathy with thalamus and brainstem involvement and high lactate caused by novel mutations in the EARS2 gene in two siblings. *J Neurol Sci* **365**, 54-58, doi:10.1016/j.jns.2016.04.008 (2016).

70 Steenweg, M. E. *et al.* Leukoencephalopathy with thalamus and brainstem involvement and high lactate 'LTBL' caused by EARS2 mutations. *Brain* **135**, 1387-1394, doi:10.1093/brain/aws070 (2012).

71 Danhauser, K. *et al.* EARS2 mutations cause fatal neonatal lactic acidosis, recurrent hypoglycemia and agenesis of corpus callosum. *Metab Brain Dis* **31**, 717-721, doi:10.1007/s11011-016-9793-2 (2016).

72 Mendes, M. I. *et al.* Bi-allelic Mutations in EPRS, Encoding the Glutamyl-Prolyl-Aminoacyl-tRNA Synthetase, Cause a Hypomyelinating Leukodystrophy. *Am J Hum Genet* **102**, 676-684, doi:10.1016/j.ajhg.2018.02.011 (2018).

73 Schuch, L. A. *et al.* FARS1-related disorders caused by bi-allelic mutations in cytosolic phenylalanyl-tRNA synthetase genes: Look beyond the lungs! *Clin Genet*, doi:10.1111/cge.13943 (2021).

74 Antonellis, A. *et al.* Compound heterozygosity for loss-of-function FARSB variants in a patient with classic features of recessive aminoacyl-tRNA synthetase-related disease. *Hum Mutat* **39**, 834-840, doi:10.1002/humu.23424 (2018).

75 Xu, Z. *et al.* Bi-allelic Mutations in Phe-tRNA Synthetase Associated with a Multi-system Pulmonary Disease Support Non-translational Function. *Am J Hum Genet* **103**, 100-114, doi:10.1016/j.ajhg.2018.06.006 (2018).

76 Zadjali, F. *et al.* Homozygosity for FARSB mutation leads to Phe-tRNA synthetase-related disease of growth restriction, brain calcification, and interstitial lung disease. *Hum Mutat* **39**, 1355-1359, doi:10.1002/humu.23595 (2018).

77 Yang, Y. *et al.* A Newly Identified Missense Mutation in FARS2 Causes Autosomal-Recessive Spastic Paraplegia. *Hum Mutat* **37**, 165-169, doi:10.1002/humu.22930 (2016).

78 Elo, J. M. *et al.* Mitochondrial phenylalanyl-tRNA synthetase mutations underlie fatal infantile Alpers encephalopathy. *Hum Mol Genet* **21**, 4521-4529, doi:10.1093/hmg/dds294 (2012).

79 Almalki, A. *et al.* Mutation of the human mitochondrial phenylalanine-tRNA synthetase causes infantile-onset epilepsy and cytochrome c oxidase deficiency. *Biochim Biophys Acta* **1842**, 56-64, doi:10.1016/j.bbadis.2013.10.008 (2014).

80 Walker, M. A. *et al.* Novel Compound Heterozygous Mutations Expand the Recognized Phenotypes of FARS2-Linked Disease. *J Child Neurol* **31**, 1127-1137, doi:10.1177/0883073816643402 (2016).

81 Cho, J. S. *et al.* FARS2 mutation and epilepsy: Possible link with early-onset epileptic encephalopathy. *Epilepsy Res* **129**, 118-124, doi:10.1016/j.eplepsyres.2016.11.022 (2017).

82 Raviglione, F. *et al.* Clinical findings in a patient with FARS2 mutations and early-infantile-encephalopathy with epilepsy. *Am J Med Genet A* **170**, 3004-3007, doi:10.1002/ajmg.a.37836 (2016).

83 Almannai, M. *et al.* FARS2 deficiency; new cases, review of clinical, biochemical, and molecular spectra, and variants interpretation based on structural, functional, and evolutionary significance. *Mol Genet Metab* **125**, 281-291, doi:10.1016/j.ymgme.2018.07.014 (2018).

84 Vernon, H. J., McClellan, R., Batista, D. A. & Naidu, S. Mutations in FARS2 and non-fatal mitochondrial dysfunction in two siblings. *Am J Med Genet A* **167A**, 1147-1151, doi:10.1002/ajmg.a.36993 (2015).

85 McMillan, H. J. *et al.* Compound heterozygous mutations in glycyl-tRNA synthetase are a proposed cause of systemic mitochondrial disease. *BMC Med Genet* **15**, 36, doi:10.1186/1471-2350-15-36 (2014).

86 Oprescu, S. N. *et al.* Compound heterozygosity for loss-of-function GARS variants results in a multisystem developmental syndrome that includes severe growth retardation. *Hum Mutat* **38**, 1412-1420, doi:10.1002/humu.23287 (2017).

87 Del Bo, R. *et al.* Coexistence of CMT-2D and distal SMA-V phenotypes in an Italian family with a GARS gene mutation. *Neurology* **66**, 752-754, doi:10.1212/01.wnl.0000201275.18875.ac (2006).

88 Abe, A. & Hayasaka, K. The GARS gene is rarely mutated in Japanese patients with Charcot-Marie-Tooth neuropathy. *J Hum Genet* **54**, 310-312, doi:10.1038/jhg.2009.25 (2009).

89 James, P. A. *et al.* Severe childhood SMA and axonal CMT due to anticodon binding domain mutations in the GARS gene. *Neurology* **67**, 1710-1712, doi:10.1212/01.wnl.0000242619.52335.bc (2006).

90 Eskuri, J. M., Stanley, C. M., Moore, S. A. & Mathews, K. D. Infantile onset CMT2D/dSMA V in monozygotic twins due to a mutation in the anticodon-binding domain of GARS. *J Peripher Nerv Syst* **17**, 132-134, doi:10.1111/j.1529-8027.2012.00370.x (2012).

91 Rohkamm, B. *et al.* Further evidence for genetic heterogeneity of distal HMN type V, CMT2 with predominant hand involvement and Silver syndrome. *J Neurol Sci* **263**, 100-106, doi:10.1016/j.jns.2007.06.047 (2007).

92 Lee, H. J. *et al.* Two novel mutations of GARS in Korean families with distal hereditary motor neuropathy type V. *J Peripher Nerv Syst* **17**, 418-421, doi:10.1111/j.1529-8027.2012.00442.x (2012).

93 Sivakumar, K. *et al.* Phenotypic spectrum of disorders associated with glycyl-tRNA synthetase mutations. *Brain* **128**, 2304-2314, doi:10.1093/brain/awh590 (2005).

94 Lee, D. C. *et al.* A recurrent GARS mutation causes distal hereditary motor neuropathy. *J Peripher Nerv Syst* **24**, 320-323, doi:10.1111/jns.12353 (2019).

95 Markovitz, R. *et al.* GARS-related disease in infantile spinal muscular atrophy: Implications for diagnosis and treatment. *Am J Med Genet A* **182**, 1167-1176, doi:10.1002/ajmg.a.61544 (2020).

96 Morelli, K. H. *et al.* Allele-specific RNA interference prevents neuropathy in Charcot-Marie-Tooth disease type 2D mouse models. *J Clin Invest* **129**, 5568-5583, doi:10.1172/JCI130600 (2019).

97 Puffenberger, E. G. *et al.* Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PLoS One* **7**, e28936, doi:10.1371/journal.pone.0028936 (2012).

98 Galatolo, D. *et al.* Bi-allelic mutations in HARS1 severely impair histidyl-tRNA synthetase expression and enzymatic activity causing a novel multisystem ataxic syndrome. *Hum Mutat* **41**, 1232-1237, doi:10.1002/humu.24024 (2020).

99 Safka Brozkova, D. *et al.* Loss of function mutations in HARS cause a spectrum of inherited peripheral neuropathies. *Brain* **138**, 2161-2172, doi:10.1093/brain/awv158 (2015).

100 Pierce, S. B. *et al.* Mutations in mitochondrial histidyl tRNA synthetase HARS2 cause ovarian dysgenesis and sensorineural hearing loss of Perrault syndrome. *Proc Natl Acad Sci U S A* **108**, 6543-6548, doi:10.1073/pnas.1103471108 (2011).

101 Orenstein, N. *et al.* Bi-allelic IARS mutations in a child with intra-uterine growth retardation, neonatal cholestasis, and mild developmental delay. *Clin Genet* **91**, 913-917, doi:10.1111/cge.12930 (2017).

102 Kopajtich, R. *et al.* Biallelic IARS Mutations Cause Growth Retardation with Prenatal Onset, Intellectual Disability, Muscular Hypotonia, and Infantile Hepatopathy. *Am J Hum Genet* **99**, 414-422, doi:10.1016/j.ajhg.2016.05.027 (2016).

103 Vona, B. *et al.* Expanding the clinical phenotype of IARS2-related mitochondrial disease. *BMC Med Genet* **19**, 196, doi:10.1186/s12881-018-0709-3 (2018).

104 Moosa, S. *et al.* Confirmation of CAGSSS syndrome as a distinct entity in a Danish patient with a novel homozygous mutation in IARS2. *Am J Med Genet A* **173**, 1102-1108, doi:10.1002/ajmg.a.38116 (2017).

105 Schwartzentruber, J. *et al.* Mutation in the nuclear-encoded mitochondrial isoleucyl-tRNA synthetase IARS2 in patients with cataracts, growth hormone deficiency with short stature, partial sensorineural deafness, and peripheral neuropathy or with Leigh syndrome. *Hum Mutat* **35**, 1285-1289, doi:10.1002/humu.22629 (2014).

106 Santos-Cortez, R. L. *et al.* Mutations in KARS, encoding lysyl-tRNA synthetase, cause autosomal-recessive nonsyndromic hearing impairment DFNB89. *Am J Hum Genet* **93**, 132-140, doi:10.1016/j.ajhg.2013.05.018 (2013).

107 McLaughlin, H. M. *et al.* Compound heterozygosity for loss-of-function lysyl-tRNA synthetase mutations in a patient with peripheral neuropathy. *Am J Hum Genet* **87**, 560-566, doi:10.1016/j.ajhg.2010.09.008 (2010).

108 McMillan, H. J. *et al.* Congenital Visual Impairment and Progressive Microcephaly Due to Lysyl-Transfer Ribonucleic Acid (RNA) Synthetase (KARS) Mutations: The Expanding Phenotype of Aminoacyl-Transfer RNA Synthetase Mutations in Human Disease. *J Child Neurol* **30**, 1037-1043, doi:10.1177/0883073814553272 (2015).

109 Sun, C. *et al.* Loss-of-function mutations in Lysyl-tRNA synthetase cause various leukoencephalopathy phenotypes. *Neurol Genet* **5**, e565, doi:10.1212/NXG.0000000000000316 (2019).

110 Casey, J. P. *et al.* Identification of a mutation in LARS as a novel cause of infantile hepatopathy. *Mol Genet Metab* **106**, 351-358, doi:10.1016/j.ymgme.2012.04.017 (2012).

111 Casey, J. P. *et al.* Clinical and genetic characterisation of infantile liver failure syndrome type 1, due to recessive mutations in LARS. *J Inherit Metab Dis* **38**, 1085-1092, doi:10.1007/s10545-015-9849-1 (2015).

112 Pierce, S. B. *et al.* Mutations in LARS2, encoding mitochondrial leucyl-tRNA synthetase, lead to premature ovarian failure and hearing loss in Perrault syndrome. *Am J Hum Genet* **92**, 614-620, doi:10.1016/j.ajhg.2013.03.007 (2013).

113 Solda, G. *et al.* First independent replication of the involvement of LARS2 in Perrault syndrome by whole-exome sequencing of an Italian family. *J Hum Genet* **61**, 295-300, doi:10.1038/jhg.2015.149 (2016).

114 Riley, L. G. *et al.* LARS2 Variants Associated with Hydrops, Lactic Acidosis, Sideroblastic Anemia, and Multisystem Failure. *JIMD Rep* **28**, 49-57, doi:10.1007/8904_2015_515 (2016).

115 Kosaki, R., Horikawa, R., Fujii, E. & Kosaki, K. Biallelic mutations in LARS2 can cause Perrault syndrome type 2 with neurologic symptoms. *Am J Med Genet A* **176**, 404-408, doi:10.1002/ajmg.a.38552 (2018).

116 van Meel, E. *et al.* Rare recessive loss-of-function methionyl-tRNA synthetase mutations presenting as a multi-organ phenotype. *BMC Med Genet* **14**, 106, doi:10.1186/1471-2350-14-106 (2013).

117 Sun, Y. *et al.* Mutations in methionyl-tRNA synthetase gene in a Chinese family with interstitial lung and liver disease, postnatal growth failure and anemia. *J Hum Genet* **62**, 647-651, doi:10.1038/jhg.2017.10 (2017).

118 Rips, J. *et al.* MARS variant associated with both recessive interstitial lung and liver disease and dominant Charcot-Marie-Tooth disease. *Eur J Med Genet* **61**, 616-620, doi:10.1016/j.ejmg.2018.04.005 (2018).

119 Lenz, D. *et al.* Rescue of respiratory failure in pulmonary alveolar proteinosis due to pathogenic MARS1 variants. *Pediatr Pulmonol* **55**, 3057-3066, doi:10.1002/ppul.25031 (2020).

120 Hadchouel, A. *et al.* Biallelic Mutations of Methionyl-tRNA Synthetase Cause a Specific Type of Pulmonary Alveolar Proteinosis Prevalent on Reunion Island. *Am J Hum Genet* **96**, 826-831, doi:10.1016/j.ajhg.2015.03.010 (2015).

121 Gonzalez, M. *et al.* Exome sequencing identifies a significant variant in methionyl-tRNA synthetase (MARS) in a family with late-onset CMT2. *J Neurol Neurosurg Psychiatry* **84**, 1247-1249, doi:10.1136/jnnp-2013-305049 (2013).

122 Hyun, Y. S. *et al.* Rare variants in methionyl- and tyrosyl-tRNA synthetase genes in late-onset autosomal dominant Charcot-Marie-Tooth neuropathy. *Clin Genet* **86**, 592-594, doi:10.1111/cge.12327 (2014).

123 Gillespie, M. K. *et al.* A Novel Mutation in MARS in a Patient with Charcot-Marie-Tooth Disease, Axonal, Type 2U with Congenital Onset. *J Neuromuscul Dis* **6**, 333-339, doi:10.3233/JND-190404 (2019).

124 Webb, B. D. *et al.* Novel, compound heterozygous, single-nucleotide variants in MARS2 associated with developmental delay, poor growth, and sensorineural hearing loss. *Hum Mutat* **36**, 587-592, doi:10.1002/humu.22781 (2015).

125 Bayat, V. *et al.* Mutations in the mitochondrial methionyl-tRNA synthetase cause a neurodegenerative phenotype in flies and a recessive ataxia (ARSAL) in humans. *PLoS Biol* **10**, e1001288, doi:10.1371/journal.pbio.1001288 (2012).

126 Wang, L. *et al.* Author Correction: Loss of NARS1 impairs progenitor proliferation in cortical brain organoids and leads to microcephaly. *Nat Commun* **12**, 1192, doi:10.1038/s41467-021-21448-1 (2021).

127 Sofou, K. *et al.* Whole exome sequencing reveals mutations in NARS2 and PARS2, encoding the mitochondrial asparaginyl-tRNA synthetase and prolyl-tRNA synthetase, in patients with Alpers syndrome. *Mol Genet Genomic Med* **3**, 59-68, doi:10.1002/mgg3.115 (2015).

128 Mizuguchi, T. *et al.* PARS2 and NARS2 mutations in infantile-onset neurodegenerative disorder. *J Hum Genet* **62**, 525-529, doi:10.1038/jhg.2016.163 (2017).

129 Seaver, L. H. *et al.* Lethal NARS2-Related Disorder Associated With Rapidly Progressive Intractable Epilepsy and Global Brain Atrophy. *Pediatr Neurol* **89**, 26-30, doi:10.1016/j.pediatrneurol.2018.07.014 (2018).

130 Vanlander, A. V. *et al.* Two siblings with homozygous pathogenic splice-site variant in mitochondrial asparaginyl-tRNA synthetase (NARS2). *Hum Mutat* **36**, 222-231, doi:10.1002/humu.22728 (2015).

131 Simon, M. *et al.* Mutations of human NARS2, encoding the mitochondrial asparaginyl-tRNA synthetase, cause nonsyndromic deafness and Leigh syndrome. *PLoS Genet* **11**, e1005097, doi:10.1371/journal.pgen.1005097 (2015).

132 Ciara, E. *et al.* Clinical and molecular characteristics of newly reported mitochondrial disease entity caused by biallelic PARS2 mutations. *J Hum Genet* **63**, 473-485, doi:10.1038/s10038-017-0401-z (2018).

133 Zhang, X. *et al.* Mutations in QARS, encoding glutaminyl-tRNA synthetase, cause progressive microcephaly, cerebral-cerebellar atrophy, and intractable seizures. *Am J Hum Genet* **94**, 547-558, doi:10.1016/j.ajhg.2014.03.003 (2014).

134 Salvarinova, R. *et al.* Expansion of the QARS deficiency phenotype with report of a family with isolated supratentorial brain abnormalities. *Neurogenetics* **16**, 145-149, doi:10.1007/s10048-014-0432-y (2015).

135 Leshinsky-Silver, E. *et al.* Severe growth deficiency, microcephaly, intellectual disability, and characteristic facial features are due to a homozygous QARS mutation. *Neurogenetics* **18**, 141-146, doi:10.1007/s10048-017-0516-6 (2017).

136 Wolf, N. I. *et al.* Mutations in RARS cause hypomyelination. *Ann Neurol* **76**, 134-139, doi:10.1002/ana.24167 (2014).

137 Nafisinia, M. *et al.* Mutations in RARS cause a hypomyelination disorder akin to Pelizaeus-Merzbacher disease. *Eur J Hum Genet* **25**, 1134-1141, doi:10.1038/ejhg.2017.119 (2017).

138 Cassandrini, D. *et al.* Pontocerebellar hypoplasia type 6 caused by mutations in RARS2: definition of the clinical spectrum and molecular findings in five patients. *J Inherit Metab Dis* **36**, 43-53, doi:10.1007/s10545-012-9487-9 (2013).

139 Edvardson, S. *et al.* Deleterious mutation in the mitochondrial arginyl-transfer RNA synthetase gene is associated with pontocerebellar hypoplasia. *Am J Hum Genet* **81**, 857-862, doi:10.1086/521227 (2007).

140 Glamuzina, E. *et al.* Further delineation of pontocerebellar hypoplasia type 6 due to mutations in the gene encoding mitochondrial arginyl-tRNA synthetase, RARS2. *J Inherit Metab Dis* **35**, 459-467, doi:10.1007/s10545-011-9413-6 (2012).

141 Li, Z. *et al.* A novel mutation in the promoter of RARS2 causes pontocerebellar hypoplasia in two siblings. *J Hum Genet* **60**, 363-369, doi:10.1038/jhg.2015.31 (2015).

142 Ngoh, A. *et al.* RARS2 mutations in a sibship with infantile spasms. *Epilepsia* **57**, e97-e102, doi:10.1111/epi.13358 (2016).

143 Rankin, J. *et al.* Pontocerebellar hypoplasia type 6: A British case with PEHO-like features. *Am J Med Genet A* **152A**, 2079-2084, doi:10.1002/ajmg.a.33531 (2010).

144 Namavar, Y. *et al.* Clinical, neuroradiological and genetic findings in pontocerebellar hypoplasia. *Brain* **134**, 143-156, doi:10.1093/brain/awq287 (2011).

145 Nishri, D. *et al.* RARS2 mutations cause early onset epileptic encephalopathy without ponto-cerebellar hypoplasia. *Eur J Paediatr Neurol* **20**, 412-417, doi:10.1016/j.ejpn.2016.02.012 (2016).

146 Luhl, S. *et al.* Novel homozygous RARS2 mutation in two siblings without pontocerebellar hypoplasia - further expansion of the phenotypic spectrum. *Orphanet J Rare Dis* **11**, 140, doi:10.1186/s13023-016-0525-9 (2016).

147 Alkhateeb, A. M., Aburahma, S. K., Habbab, W. & Thompson, I. R. Novel mutations in WWOX, RARS2, and C10orf2 genes in consanguineous Arab families with intellectual disability. *Metab Brain Dis* **31**, 901-907, doi:10.1007/s11011-016-9827-9 (2016).

148 Mathew, T., Avati, A., D'Souza, D. & Therambil, M. Expanding spectrum of RARS2 gene disorders: Myoclonic epilepsy, mental retardation, spasticity, and extrapyramidal features. *Epilepsia Open* **3**, 270-275, doi:10.1002/epi4.12108 (2018).

149 Musante, L. *et al.* Mutations of the aminoacyl-tRNA-synthetases SARS and WARS2 are implicated in the etiology of autosomal recessive intellectual disability. *Hum Mutat* **38**, 621-636, doi:10.1002/humu.23205 (2017).

150 Rivera, H. *et al.* A new mutation in the gene encoding mitochondrial seryl-tRNA synthetase as a cause of HUPRA syndrome. *BMC Nephrol* **14**, 195, doi:10.1186/1471-2369-14-195 (2013).

151 Belostotsky, R. *et al.* Mutations in the mitochondrial seryl-tRNA synthetase cause hyperuricemia, pulmonary hypertension, renal failure in infancy and alkalosis, HUPRA syndrome. *Am J Hum Genet* **88**, 193-200, doi:10.1016/j.ajhg.2010.12.010 (2011).

152 Linnankivi, T., Neupane, N., Richter, U., Isohanni, P. & Tyynismaa, H. Splicing Defect in Mitochondrial Seryl-tRNA Synthetase Gene Causes Progressive Spastic Paresis Instead of HUPRA Syndrome. *Hum Mutat* **37**, 884-888, doi:10.1002/humu.23021 (2016).

153 Theil, A. F. *et al.* Bi-allelic TARS Mutations Are Associated with Brittle Hair Phenotype. *Am J Hum Genet* **105**, 434-440, doi:10.1016/j.ajhg.2019.06.017 (2019).

154 Diodato, D. *et al.* VARS2 and TARS2 mutations in patients with mitochondrial encephalomyopathies. *Hum Mutat* **35**, 983-989, doi:10.1002/humu.22590 (2014).

155 Karaca, E. *et al.* Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. *Neuron* **88**, 499-513, doi:10.1016/j.neuron.2015.09.048 (2015).

156 Stephen, J. *et al.* Loss of function mutations in VARS encoding cytoplasmic valyl-tRNA synthetase cause microcephaly, seizures, and progressive cerebral atrophy. *Hum Genet* **137**, 293-303, doi:10.1007/s00439-018-1882-3 (2018).

157 Okur, V., Ganapathi, M., Wilson, A. & Chung, W. K. Biallelic variants in VARS in a family with two siblings with intellectual disability and microcephaly: case report and review of the literature. *Cold Spring Harb Mol Case Stud* **4**, doi:10.1101/mcs.a003301 (2018).

158 Friedman, J. *et al.* Biallelic mutations in valyl-tRNA synthetase gene VARS are associated with a progressive neurodevelopmental epileptic encephalopathy. *Nat Commun* **10**, 707, doi:10.1038/s41467-018-07067-3 (2019).

159 Siekierska, A. *et al.* Biallelic VARS variants cause developmental encephalopathy with microcephaly that is recapitulated in vars knockout zebrafish. *Nat Commun* **10**, 708, doi:10.1038/s41467-018-07953-w (2019).

160 Begliuomini, C. *et al.* VARS2-linked mitochondrial encephalopathy: two case reports enlarging the clinical phenotype. *BMC Med Genet* **20**, 77, doi:10.1186/s12881-019-0798-7 (2019).

161 Ruzman, L. *et al.* A novel VARS2 gene variant in a patient with epileptic encephalopathy. *Ups J Med Sci*, 1-5, doi:10.1080/03009734.2019.1670297 (2019).

162 Baertling, F. *et al.* Neonatal encephalocardiomyopathy caused by mutations in VARS2. *Metab Brain Dis* **32**, 267-270, doi:10.1007/s11011-016-9890-2 (2017).

163 Li, J. Q., Dong, H. L., Chen, C. X. & Wu, Z. Y. A novel WARS mutation causes distal hereditary motor neuropathy in a Chinese family. *Brain* **142**, e49, doi:10.1093/brain/awz218 (2019).

164 Theisen, B. E. *et al.* Deficiency of WARS2, encoding mitochondrial tryptophanyl tRNA synthetase, causes severe infantile onset leukoencephalopathy. *Am J Med Genet A* **173**, 2505-2510, doi:10.1002/ajmg.a.38339 (2017).

165 Wortmann, S. B. *et al.* Biallelic variants in WARS2 encoding mitochondrial tryptophanyl-tRNA synthase in six individuals with mitochondrial encephalopathy. *Hum Mutat* **38**, 1786-1795, doi:10.1002/humu.23340 (2017).

166 Burke, E. A. *et al.* Biallelic mutations in mitochondrial tryptophanyl-tRNA synthetase cause Levodopa-responsive infantile-onset Parkinsonism. *Clin Genet* **93**, 712-718, doi:10.1111/cge.13172 (2018).

167 Nowaczyk, M. J. *et al.* A novel multisystem disease associated with recessive mutations in the tyrosyl-tRNA synthetase (YARS) gene. *Am J Med Genet A* **173**, 126-134, doi:10.1002/ajmg.a.37973 (2017).

168 Gonzaga-Jauregui, C. *et al.* Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy. *Cell Rep* **12**, 1169-1183, doi:10.1016/j.celrep.2015.07.023 (2015).

169 Riley, L. G. *et al.* Mutation of the mitochondrial tyrosyl-tRNA synthetase gene, YARS2, causes myopathy, lactic acidosis, and sideroblastic anemia--MLASA syndrome. *Am J Hum Genet* **87**, 52-59, doi:10.1016/j.ajhg.2010.06.001 (2010).

170 Shahni, R. *et al.* A distinct mitochondrial myopathy, lactic acidosis and sideroblastic anemia (MLASA) phenotype associates with YARS2 mutations. *Am J Med Genet A* **161A**, 2334-2338, doi:10.1002/ajmg.a.36065 (2013).

171 Sasarman, F., Nishimura, T., Thiffault, I. & Shoubridge, E. A. A novel mutation in YARS2 causes myopathy with lactic acidosis and sideroblastic anemia. *Hum Mutat* **33**, 1201-1206, doi:10.1002/humu.22098 (2012).

172 Nakajima, J. *et al.* A novel homozygous YARS2 mutation causes severe myopathy, lactic acidosis, and sideroblastic anemia 2. *J Hum Genet* **59**, 229-232, doi:10.1038/jhg.2013.143 (2014).

173 Riley, L. G. *et al.* Phenotypic variability and identification of novel YARS2 mutations in YARS2 mitochondrial myopathy, lactic acidosis and sideroblastic anaemia. *Orphanet J Rare Dis* **8**, 193, doi:10.1186/1750-1172-8-193 (2013).

174 Ardissone, A. *et al.* A Novel Homozygous YARS2 Mutation in Two Italian Siblings and a Review of Literature. *JIMD Rep* **20**, 95-101, doi:10.1007/8904_2014_397 (2015).

175 Sagi-Dain, L. *et al.* Whole-exome sequencing reveals a novel missense mutation in the MARS gene related to a rare Charcot-Marie-Tooth neuropathy type 2U. *J Peripher Nerv Syst* **23**, 138-142, doi:10.1111/jns.12264 (2018).

176 Nam, S. H. *et al.* Identification of Genetic Causes of Inherited Peripheral Neuropathies by Targeted Gene Panel Sequencing. *Mol Cells* **39**, 382-388, doi:10.14348/molcells.2016.2288 (2016).

177 Oprescu, S. N., Griffin, L. B., Beg, A. A. & Antonellis, A. Predicting the pathogenicity of aminoacyl-tRNA synthetase mutations. *Methods* **113**, 139-151, doi:10.1016/j.ymeth.2016.11.013 (2017).

178 Seburn, K. L., Nangle, L. A., Cox, G. A., Schimmel, P. & Burgess, R. W. An active dominant mutation of glycyl-tRNA synthetase causes neuropathy in a Charcot-Marie-Tooth 2D mouse model. *Neuron* **51**, 715-726, doi:10.1016/j.neuron.2006.08.027 (2006).

179 Motley, W. W. *et al.* Charcot-Marie-Tooth-linked mutant GARS is toxic to peripheral neurons independent of wild-type GARS levels. *PLoS Genet* **7**, e1002399, doi:10.1371/journal.pgen.1002399 (2011).

180 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).

181 Wei, N., Zhang, Q. & Yang, X. L. Neurodegenerative Charcot-Marie-Tooth disease as a case study to decipher novel functions of aminoacyl-tRNA synthetases. *J Biol Chem* **294**, 5321-5339, doi:10.1074/jbc.REV118.002955 (2019).

182 He, W. *et al.* CMT2D neuropathy is linked to the neomorphic binding activity of glycyl-tRNA synthetase. *Nature* **526**, 710-714, doi:10.1038/nature15510 (2015).

183 Sleigh, J. N. *et al.* Trk receptor signaling and sensory neuron fate are perturbed in human neuropathy caused by Gars mutations. *Proc Natl Acad Sci U S A* **114**, E3324-E3333, doi:10.1073/pnas.1614557114 (2017).

184 Mo, Z. *et al.* Aberrant GlyRS-HDAC6 interaction linked to axonal transport deficits in Charcot-Marie-Tooth neuropathy. *Nat Commun* **9**, 1007, doi:10.1038/s41467-018-03461-z (2018).

185 Benoy, V. *et al.* HDAC6 is a therapeutic target in mutant GARS-induced Charcot-Marie-Tooth disease. *Brain* **141**, 673-687, doi:10.1093/brain/awx375 (2018).

186 Sun, L. *et al.* CMT2N-causing aminoacylation domain mutants enable Nrp1 interaction with AlaRS. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2012898118 (2021).

187 Blocquel, D. *et al.* Alternative stable conformation capable of protein misinteraction links tRNA synthetase to peripheral neuropathy. *Nucleic Acids Res* **45**, 8091-8104, doi:10.1093/nar/gkx455 (2017).

188 Morelli, K. H. *et al.* Allele-specific RNA interference prevents neuropathy in Charcot-Marie-Tooth disease type 2D mouse models. *J Clin Invest*, doi:10.1172/JCI130600 (2019).

189 Zhang, J., Zhang, Z., Zhang, Y. & Wu, Y. Distinct magnetic resonance imaging features in a patient with novel RARS2 mutations: A case report and review of the literature. *Exp Ther Med* **15**, 1099-1104, doi:10.3892/etm.2017.5491 (2018).

190 Scheper, G. C. *et al.* Mitochondrial aspartyl-tRNA synthetase deficiency causes leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. *Nat Genet* **39**, 534-539, doi:10.1038/ng2013 (2007).

191 Talim, B. *et al.* Multisystem fatal infantile disease caused by a novel homozygous EARS2 mutation. *Brain* **136**, e228, doi:10.1093/brain/aws197 (2013).

192 Sommerville, E. W. *et al.* Clinical Features, Molecular Heterogeneity, and Prognostic Implications in YARS2-Related Mitochondrial Myopathy. *JAMA Neurol* **74**, 686-694, doi:10.1001/jamaneurol.2016.4357 (2017).

193 Gonzalez-Serrano, L. E., Chihade, J. W. & Sissler, M. When a common biological role does not imply common disease outcomes: Disparate pathology linked to human mitochondrial aminoacyl-tRNA synthetases. *J Biol Chem* **294**, 5309-5320, doi:10.1074/jbc.REV118.002953 (2019).

194 Dittmar, K. A., Goodenbour, J. M. & Pan, T. Tissue-specific differences in human transfer RNA expression. *PLoS Genet* **2**, e221, doi:10.1371/journal.pgen.0020221 (2006).

195 Wang, L. *et al.* Loss of NARS1 impairs progenitor proliferation in cortical brain organoids and leads to microcephaly. *Nat Commun* **11**, 4038, doi:10.1038/s41467-020-17454-4 (2020).

196 Sagi, D. *et al.* Tissue- and Time-Specific Expression of Otherwise Identical tRNA Genes. *PLoS Genet* **12**, e1006264, doi:10.1371/journal.pgen.1006264 (2016).

197 Barle, H. *et al.* The concentrations of free amino acids in human liver tissue obtained during laparoscopic surgery. *Clin Physiol* **16**, 217-227 (1996).

198 Bergstrom, J., Furst, P., Noree, L. O. & Vinnars, E. Intracellular free amino acid concentration in human muscle tissue. *J Appl Physiol* **36**, 693-697, doi:10.1152/jappl.1974.36.6.693 (1974).

199 Shimomura, Y. & Ito, M. Human hair keratin-associated proteins. *J Investig Dermatol Symp Proc* **10**, 230-233, doi:10.1111/j.1087-0024.2005.10112.x (2005).

200 Rice, R. H., Xia, Y., Alvarado, R. J. & Phinney, B. S. Proteomic analysis of human nail plate. *J Proteome Res* **9**, 6752-6758, doi:10.1021/pr1009349 (2010).

201 Rice, R. H. Proteomic analysis of hair shaft and nail plate. *J Cosmet Sci* **62**, 229-236 (2011).

202 Chamcheu, J. C. *et al.* Keratin gene mutations in disorders of human skin and its appendages. *Arch Biochem Biophys* **508**, 123-137, doi:10.1016/j.abb.2010.12.019 (2011).

203 Schweizer, J., Langbein, L., Rogers, M. A. & Winter, H. Hair follicle-specific keratins and their diseases. *Exp Cell Res* **313**, 2010-2020, doi:10.1016/j.yexcr.2007.02.032 (2007).

204 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

205 The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338, doi:10.1093/nar/gky1055 (2019).

206 Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**, D419-D426, doi:10.1093/nar/gky1038 (2019).

207 Cabrijan, L. & Lipozencic, J. Adhesion molecules in keratinocytes. *Clin Dermatol* **29**, 427-431, doi:10.1016/j.clindermatol.2011.01.012 (2011).

208 Yaeger, D. *et al.* Outcomes of clinical examination and genetic testing of 500 individuals with hearing loss evaluated through a genetics of hearing loss clinic. *Am J Med Genet A* **140**, 827-836, doi:10.1002/ajmg.a.31179 (2006).

209 Zhou, X. L. *et al.* Mutations in KARS cause early-onset hearing loss and leukoencephalopathy: Potential pathogenic mechanism. *Hum Mutat* **38**, 1740-1750, doi:10.1002/humu.23335 (2017).

210 Scheidecker, S. *et al.* Mutations in KARS cause a severe neurological and neurosensory disease with optic neuropathy. *Hum Mutat*, doi:10.1002/humu.23799 (2019).

211 Colantonio, J. R. *et al.* The dynein regulatory complex is required for ciliary motility and otolith biogenesis in the inner ear. *Nature* **457**, 205-209, doi:10.1038/nature07520 (2009).

212 Olbrich, H. *et al.* Loss-of-Function GAS8 Mutations Cause Primary Ciliary Dyskinesia and Disrupt the Nexin-Dynein Regulatory Complex. *Am J Hum Genet* **97**, 546-554, doi:10.1016/j.ajhg.2015.08.012 (2015).

213 Naz, S. *et al.* Mutations in a novel gene, TMIE, are associated with hearing loss linked to the DFNB6 locus. *Am J Hum Genet* **71**, 632-636, doi:10.1086/342193 (2002).

214 Hansen, J. E., Ampaya, E. P., Bryant, G. H. & Navin, J. J. Branching pattern of airways and air spaces of a single human terminal bronchiole. *J Appl Physiol* **38**, 983-989, doi:10.1152/jappl.1975.38.6.983 (1975).

215 Butler, J. P. & Tsuda, A. Transport of gases between the environment and alveoli--theoretical foundations. *Compr Physiol* **1**, 1301-1316, doi:10.1002/cphy.c090016 (2011).

216 Abuduxikuer, K. *et al.* Novel methionyl-tRNA synthetase gene variants/phenotypes in interstitial lung and liver disease: A case report and review of literature. *World J Gastroenterol* **24**, 4208-4216, doi:10.3748/wjg.v24.i36.4208 (2018).

217 Martinez, F. J. *et al.* Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* **3**, 17074, doi:10.1038/nrdp.2017.74 (2017).

218 Devine, M. S. & Garcia, C. K. Genetic interstitial lung disease. *Clin Chest Med* **33**, 95-110, doi:10.1016/j.ccm.2011.11.001 (2012).

219 Rosen, S. H., Castleman, B. & Liebow, A. A. Pulmonary alveolar proteinosis. *N Engl J Med* **258**, 1123-1142, doi:10.1056/NEJM195806052582301 (1958).

220 Enaud, L. *et al.* Pulmonary alveolar proteinosis in children on La Reunion Island: a new inherited disorder? *Orphanet J Rare Dis* **9**, 85, doi:10.1186/1750-1172-9-85 (2014).

221 Comisso, M., Hadchouel, A., de Blic, J. & Mirande, M. Mutations in MARS identified in a specific type of pulmonary alveolar proteinosis alter methionyl-tRNA synthetase activity. *FEBS J* **285**, 2654-2661, doi:10.1111/febs.14510 (2018).

222 Nogee, L. M. *et al.* A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med* **344**, 573-579, doi:10.1056/NEJM200102223440805 (2001).

223 Lin, W. X., Zheng, Q. Q., Guo, L., Cheng, Y. & Song, Y. Z. [Clinical feature and molecular diagnostic analysis of the first non-caucasian child with infantile liver failure syndrome type 1]. *Zhongguo Dang Dai Er Ke Za Zhi* **19**, 913-920 (2017).

224 Peroutka, C. *et al.* Severe Neonatal Manifestations of Infantile Liver Failure Syndrome Type 1 Caused by Cytosolic Leucine-tRNA Synthetase Deficiency. *JIMD Rep* **45**, 71-76, doi:10.1007/8904_2018_143 (2019).

225 Haack, T. B. *et al.* Biallelic Mutations in NBAS Cause Recurrent Acute Liver Failure with Onset in Infancy. *Am J Hum Genet* **97**, 163-169, doi:10.1016/j.ajhg.2015.05.009 (2015).

226 Han, J. M. *et al.* Leucyl-tRNA synthetase is an intracellular leucine sensor for the mTORC1-signaling pathway. *Cell* **149**, 410-424, doi:10.1016/j.cell.2012.02.044 (2012).

227 Bockaert, J. & Marin, P. mTOR in Brain Physiology and Pathologies. *Physiol Rev* **95**, 1157-1187, doi:10.1152/physrev.00038.2014 (2015).

228 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).

229 Sokabe, M. *et al.* The structure of alanyl-tRNA synthetase with editing domain. *Proc Natl Acad Sci U S A* **106**, 11028-11033, doi:10.1073/pnas.0904645106 (2009).

230 Garin, S., Levi, O., Cohen, B., Golani-Armon, A. & Arava, Y. S. Localization and RNA Binding of Mitochondrial Aminoacyl tRNA Synthetases. *Genes (Basel)* **11**, doi:10.3390/genes11101185 (2020).

231 Chang, C. Y., Chien, C. I., Chang, C. P., Lin, B. C. & Wang, C. C. A WHEP Domain Regulates the Dynamic Structure and Activity of Caenorhabditis elegans Glycyl-tRNA Synthetase. *J Biol Chem* **291**, 16567-16575, doi:10.1074/jbc.M116.730812 (2016).

232 Chien, C. I. *et al.* Functional substitution of a eukaryotic glycyl-tRNA synthetase with an evolutionarily unrelated bacterial cognate enzyme. *PLoS One* **9**, e94659, doi:10.1371/journal.pone.0094659 (2014).

233 Abbott, J. A. *et al.* Substrate interaction defects in histidyl-tRNA synthetase linked to dominant axonal peripheral neuropathy. *Hum Mutat* **39**, 415-432, doi:10.1002/humu.23380 (2018).

234 Antonellis, A. *et al.* Functional analyses of glycyl-tRNA synthetase mutations suggest a key role for tRNA-charging enzymes in peripheral axons. *J Neurosci* **26**, 10397-10406, doi:10.1523/JNEUROSCI.1671-06.2006 (2006).

235 Griffin, L. B. *et al.* Impaired function is a common feature of neuropathy-associated glycyl-tRNA synthetase mutations. *Hum Mutat* **35**, 1363-1371, doi:10.1002/humu.22681 (2014).

236 Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**, 19-27 (1989).

237 Boeke, J. D., Trueheart, J., Natsoulis, G. & Fink, G. R. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol* **154**, 164-175, doi:10.1016/0076-6879(87)54076-9 (1987).

238 Hou, Y. M., Westhof, E. & Giege, R. An unusual RNA tertiary interaction has a role for the specific aminoacylation of a transfer RNA. *Proc Natl Acad Sci U S A* **90**, 6776-6780, doi:10.1073/pnas.90.14.6776 (1993).

239 Ripmaster, T. L., Shiba, K. & Schimmel, P. Wide cross-species aminoacyl-tRNA synthetase replacement in vivo: yeast cytoplasmic alanine enzyme replaced by human polymyositis serum antigen. *Proc Natl Acad Sci U S A* **92**, 4932-4936, doi:10.1073/pnas.92.11.4932 (1995).

240 Klein, C. J. *et al.* Application of whole exome sequencing in undiagnosed inherited polyneuropathies. *J Neurol Neurosurg Psychiatry* **85**, 1265-1272, doi:10.1136/jnnp-2013-306740 (2014).

241 Turner, R. J., Lovato, M. & Schimmel, P. One of two genes encoding glycyl-tRNA synthetase in Saccharomyces cerevisiae provides mitochondrial and cytoplasmic functions. *J Biol Chem* **275**, 27681-27688, doi:10.1074/jbc.M003416200 (2000).

242 Srivastava, S. *et al.* Expansion of the clinical spectrum associated with AARS2-related disorders. *Am J Med Genet A*, doi:10.1002/ajmg.a.61188 (2019).

243 Wang, J. Y. *et al.* A homozygous mutation of alanyl-transfer RNA synthetase 2 in a patient of adult-onset leukodystrophy: A case report and literature review. *Brain Behav*, e01313, doi:10.1002/brb3.1313 (2019).

244 Tang, H. L. *et al.* Translation of a yeast mitochondrial tRNA synthetase initiated at redundant non-AUG codons. *J Biol Chem* **279**, 49656-49663, doi:10.1074/jbc.M408081200 (2004).

245 O'Rourke, T. W., Doudican, N. A., Mackereth, M. D., Doetsch, P. W. & Shadel, G. S. Mitochondrial dysfunction due to oxidative mitochondrial DNA damage is reduced through cooperative actions of diverse proteins. *Mol Cell Biol* **22**, 4086-4093, doi:10.1128/MCB.22.12.4086-4093.2002 (2002).

246 Boeke, J. D., LaCroute, F. & Fink, G. R. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet* **197**, 345-346, doi:10.1007/BF00330984 (1984).

247 Tracewska-Siemiatkowska, A. *et al.* An Expanded Multi-Organ Disease Phenotype Associated with Mutations in YARS. *Genes (Basel)* **8**, doi:10.3390/genes8120381 (2017).

248 Williams, K. B. *et al.* Homozygosity for a mutation affecting the catalytic domain of tyrosyl-tRNA synthetase (YARS) causes multisystem disease. *Hum Mol Genet* **28**, 525-538, doi:10.1093/hmg/ddy344 (2019).

249 MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-476, doi:10.1038/nature13127 (2014).

250 Sobreira, N., Schiettecatte, F., Boehm, C., Valle, D. & Hamosh, A. New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. *Hum Mutat* **36**, 425-431, doi:10.1002/humu.22769 (2015).

251 Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59, doi:10.1038/nature07517 (2008).

252 Gahl, W. A. *et al.* The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet Med* **14**, 51-59, doi:10.1038/gim.0b013e318232a005 (2012).

253 Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182-189, doi:10.1038/nbt.1523 (2009).

254 Teer, J. K. *et al.* Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res* **20**, 1420-1431, doi:10.1101/gr.106716.110 (2010).

255 Teer, J. K., Green, E. D., Mullikin, J. C. & Biesecker, L. G. VarSifter: visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics* **28**, 599-600, doi:10.1093/bioinformatics/btr711 (2012).

256 Gahl, W. A. & Tifft, C. J. The NIH Undiagnosed Diseases Program: lessons learned. *JAMA* **305**, 1904-1905, doi:10.1001/jama.2011.613 (2011).

257 Markello, T. C. *et al.* Recombination mapping using Boolean logic and high-density SNP genotyping for exome sequence filtering. *Mol Genet Metab* **105**, 382-389, doi:10.1016/j.ymgme.2011.12.014 (2012).

258 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).

259 Davidson, E., Caffarella, J., Vitseva, O., Hou, Y. M. & King, M. P. Isolation of two cDNAs encoding functional human cytoplasmic cysteinyl-tRNA synthetase. *Biol Chem* **382**, 399-406, doi:10.1515/BC.2001.049 (2001).

260 Hou, Y. M., Shiba, K., Mottes, C. & Schimmel, P. Sequence determination and modeling of structural motifs for the smallest monomeric aminoacyl-tRNA synthetase. *Proc Natl Acad Sci U S A* **88**, 976-980, doi:10.1073/pnas.88.3.976 (1991).

261 Schreier, A. A. & Schimmel, P. R. Transfer ribonucleic acid synthetase catalyzed deacylation of aminoacyl transfer ribonucleic acid in the absence of adenosine monophosphate and pyrophosphate. *Biochemistry* **11**, 1582-1589, doi:10.1021/bi00759a006 (1972).

262 Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res* **45**, D190-D199, doi:10.1093/nar/gkw1107 (2017).

263 Liu, C., Sanders, J. M., Pascal, J. M. & Hou, Y. M. Adaptation to tRNA acceptor stem structure by flexible adjustment in the catalytic domain of class I tRNA synthetases. *RNA* **18**, 213-221, doi:10.1261/rna.029983.111 (2012).

264 Marchler-Bauer, A. *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**, D200-D203, doi:10.1093/nar/gkw1129 (2017).

265 Kachroo, A. H. *et al.* Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **348**, 921-925, doi:10.1126/science.aaa0769 (2015).

266 Liu, C. *et al.* Kinetic quality control of anticodon recognition by a eukaryotic aminoacyl-tRNA synthetase. *J Mol Biol* **367**, 1063-1078, doi:10.1016/j.jmb.2007.01.050 (2007).

267 Amrani, N., Sachs, M. S. & Jacobson, A. Early nonsense: mRNA decay solves a translational problem. *Nat Rev Mol Cell Biol* **7**, 415-425, doi:10.1038/nrm1942 (2006).

268 Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* **5**, 89-99, doi:10.1038/nrm1310 (2004).

269 Akaike, T. *et al.* Cysteinyl-tRNA synthetase governs cysteine polysulfidation and mitochondrial bioenergetics. *Nat Commun* **8**, 1177, doi:10.1038/s41467-017-01311-y (2017).

270 Hayano, M., Yang, W. S., Corn, C. K., Pagano, N. C. & Stockwell, B. R. Loss of cysteinyl-tRNA synthetase (CARS) induces the transsulfuration pathway and inhibits ferroptosis induced by cystine deprivation. *Cell Death Differ* **23**, 270-278, doi:10.1038/cdd.2015.93 (2016).

271 Cho, S. *et al.* Endogenous TLR2 ligand embedded in the catalytic region of human cysteinyl-tRNA synthetase 1. *J Immunother Cancer* **8**, doi:10.1136/jitc-2019-000277 (2020).

272 Lee, J. I. *et al.* HepG2/C3A cells respond to cysteine deprivation by induction of the amino acid deprivation/integrated stress response pathway. *Physiol Genomics* **33**, 218-229, doi:10.1152/physiolgenomics.00263.2007 (2008).

273 Mullen, P. *et al.* Neuropathy-associated histidyl-tRNA synthetase variants attenuate protein synthesis in vitro and disrupt axon outgrowth in developing zebrafish. *FEBS J* **288**, 142-159, doi:10.1111/febs.15449 (2021).

274 Darnell, A. M., Subramaniam, A. R. & O'Shea, E. K. Translational Control through Differential Ribosome Pausing during Amino Acid Limitation in Mammalian Cells. *Mol Cell* **71**, 229-243 e211, doi:10.1016/j.molcel.2018.06.041 (2018).

275 Moon, S. L. & Parker, R. EIF2B2 mutations in vanishing white matter disease hypersuppress translation and delay recovery during the integrated stress response. *RNA* **24**, 841-852, doi:10.1261/rna.066563.118 (2018).

276 Schmidt, E. K., Clavarino, G., Ceppi, M. & Pierre, P. SUnSET, a nonradioactive method to monitor protein synthesis. *Nat Methods* **6**, 275-277, doi:10.1038/nmeth.1314 (2009).

277 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

278 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

279 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

280 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

281 Becker, K. & Jerchow, B. in *Advanced Protocols for Animal Transgenesis: An ISTT Manual* (eds Shirley Pease & Thomas L. Saunders) 99-115 (Springer Berlin Heidelberg, 2011).

282 K. Jessie, O. H. H. a. Z. H. A. R. Protein Precipitation Method for Salivary Proteins and Rehydration Buffer for Two-Dimensional Electrophoresis. *Biotechnology* **7**, 686-693, doi:10.3923/biotech.2008.686.693 (2008).

283 Zybailov, B. *et al.* Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J Proteome Res* **5**, 2339-2347, doi:10.1021/pr060161n (2006).

284 Kaufman, R. J. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. *Genes Dev* **13**, 1211-1233, doi:10.1101/gad.13.10.1211 (1999).

285 Yin, J. *et al.* L-Cysteine metabolism and its nutritional implications. *Mol Nutr Food Res* **60**, 134-146, doi:10.1002/mnfr.201500031 (2016).

286 Kim, J. E. *et al.* An elongation factor-associating domain is inserted into human cysteinyl-tRNA synthetase by alternative splicing. *Nucleic Acids Res* **28**, 2866-2872, doi:10.1093/nar/28.15.2866 (2000).

287 Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929-11947, doi:10.1523/JNEUROSCI.1860-14.2014 (2014).

288 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779, doi:10.1371/journal.pmed.1001779 (2015).

289 tom Dieck, S. *et al.* Direct visualization of newly synthesized target proteins in situ. *Nat Methods* **12**, 411-414, doi:10.1038/nmeth.3319 (2015).

290 Iwasaki, S. & Ingolia, N. T. The Growing Toolbox for Protein Synthesis Studies. *Trends Biochem Sci* **42**, 612-624, doi:10.1016/j.tibs.2017.05.004 (2017).

291 Borras, E. & Sabido, E. What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics* **17**, doi:10.1002/pmic.201700180 (2017).

292 Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**, 1534-1550, doi:10.1038/nprot.2012.086 (2012).

293 Thirumoorthy, N. *et al.* A review of metallothionein isoforms and their role in pathophysiology. *World J Surg Oncol* **9**, 54, doi:10.1186/1477-7819-9-54 (2011).

294 Gogakos, T. *et al.* Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep* **20**, 1463-1475, doi:10.1016/j.celrep.2017.07.029 (2017).

295 Goodarzi, H. *et al.* Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* **165**, 1416-1427, doi:10.1016/j.cell.2016.05.046 (2016).

296 Cheng, S., Stone, J. & de Berker, D. Trichothiodystrophy and fragile hair: the distinction between diagnostic signs and diagnostic labels in childhood hair disease. *Br J Dermatol* **161**, 1379-1383, doi:10.1111/j.1365-2133.2009.09403.x (2009).

297 Doka, E. *et al.* A novel persulfide detection method reveals protein persulfide- and polysulfide-reducing functions of thioredoxin and glutathione systems. *Sci Adv* **2**, e1500968, doi:10.1126/sciadv.1500968 (2016).

298 La Bella, V., Kallenbach, S. & Pettmann, B. Expression and subcellular localization of two isoforms of the survival motor neuron protein in different cell types. *J Neurosci Res* **62**, 346-356, doi:10.1002/1097-4547(20001101)62:3<346::AID-JNR4>3.0.CO;2-D (2000).

299 Kohnhorst, C. L., Schmitt, D. L., Sundaram, A. & An, S. Subcellular functions of proteins under fluorescence single-cell microscopy. *Biochim Biophys Acta* **1864**, 77-84, doi:10.1016/j.bbapap.2015.05.014 (2016).

300 Maccarrone, G., Bonfiglio, J. J., Silberstein, S., Turck, C. W. & Martins-de-Souza, D. Characterization of a Protein Interactome by Co-Immunoprecipitation and Shotgun Mass

Spectrometry. *Methods Mol Biol* **1546**, 223-234, doi:10.1007/978-1-4939-6730-8_19 (2017).

301 Deciphering Developmental Disorders, S. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433-438, doi:10.1038/nature21062 (2017).

302 Rinaldi, C. & Wood, M. J. A. Antisense oligonucleotides: the next frontier for treatment of neurological disorders. *Nat Rev Neurol* **14**, 9-21, doi:10.1038/nrneurol.2017.148 (2018).

303 Young, S. K., Baird, T. D. & Wek, R. C. Translation Regulation of the Glutamyl-prolyl-tRNA Synthetase Gene EPRS through Bypass of Upstream Open Reading Frames with Noncanonical Initiation Codons. *J Biol Chem* **291**, 10824-10835, doi:10.1074/jbc.M116.722256 (2016).

304 Alexandrova, J., Paulus, C., Rudinger-Thirion, J., Jossinet, F. & Frugier, M. Elaborate uORF/IRES features control expression and localization of human glycyl-tRNA synthetase. *RNA Biol* **12**, 1301-1313, doi:10.1080/15476286.2015.1086866 (2015).

305 Liang, X. H. *et al.* Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nat Biotechnol* **34**, 875-880, doi:10.1038/nbt.3589 (2016).

306 Liang, X. H., Shen, W. & Crooke, S. T. Specific Increase of Protein Levels by Enhancing Translation Using Antisense Oligonucleotides Targeting Upstream Open Frames. *Adv Exp Med Biol* **983**, 129-146, doi:10.1007/978-981-10-4310-9_9 (2017).

307 Schwahn, B. C. *et al.* Cystathionine beta synthase deficiency and brain edema associated with methionine excess under betaine supplementation: Four new cases and a review of the evidence. *JIMD Rep* **52**, 3-10, doi:10.1002/jmd2.12092 (2020).

308 Olney, J. W., Zorumski, C., Price, M. T. & Labruyere, J. L-cysteine, a bicarbonate-sensitive endogenous excitotoxin. *Science* **248**, 596-599, doi:10.1126/science.2185543 (1990).