INVESTIGATING THE ROLE OF HISTIDYL-TRNA SYNTHETASE (*HARS*) AND GLYCYL-TRNA SYNTHETASE (*GARS*) MUTATIONS IN CHARCOT-MARIE-TOOTH DISEASE

by

Aimée Vester

A senior honors thesis submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science
(Neuroscience)
in the University of Michigan
2012

Assistant Professor Anthony Antonellis, Sponsor
Department of Human Genetics
Department of Neurology

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor, Tony Antonellis. I have learned an immeasurable amount while working in the Antonellis lab, and thank Tony for his support, guidance, and patience. Doing research has been one of the most formative activities in my three years at Michigan, and it will definitely influence my future plans. Tony mentored me in all my experiments, and also helped with the figures and in editing the thesis. Second, I would like to thank Heather McLaughlin. I am very grateful that I was able to work with her these past three years, and for her patience with all of my questions. Heather, in addition to helping me throughout experiments, designed the bisulfite sequencing primers and performed the bisulfite sequencing analysis. I would like to thank Chani Hodonsky for all of her help in answering lab questions and for her encouragement as well. Also, thank you to Megan Brewer, Chetna Gopinath, and Bill Law for giving me a hand whenever I needed it, and for lots of good advice.

Furthermore, I would like to thank my co-sponsor, Cathy Collins. I appreciate her willingness to let me learn new techniques in her lab this semester, her many suggestions, and her support. I would also like to thank Susan Klinedinst for showing me everything from larval dissections to imaging. In addition teaching me, Susan collected virgin flies and set up crosses for my fly experiments. Thank you to Xin, Jiaxing, Bibhu, Yan, Pushpa, Nicolette, Jen, and Dhwani for welcoming me to their lab, too.

In addition, I would like to thank Laura Olsen for her advice and support, and for acting as my thesis reader. It was not until Professor Olsen's writing class that I started to seriously consider writing a thesis.

I also thank Miriam Meisler and members of the Meisler lab for numerous helpful suggestions and comments during lab meetings.

Lastly, I thank my family and friends for their love and support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS i

ABSTRACT iii

Introduction 4
I. Charcot-Marie-Tooth disease 4
II. Aminoacyl-tRNA synthetases 6
III. Proposed studies 9

CHAPTER TWO 12
I. Introduction 12
II. Methods 17
III. Results 28
IV. Discussion 32

CHAPTER THREE 37
I. Introduction 37
II. Methods 39
III. Results 40
IV. Discussion 42

CHAPTER FOUR 45
Conclusions and Future Directions 45
I. Implicating Histidyl-tRNA Synthetase (HARS) in CMT Disease Pathogenesis 45
II. Evaluating Glycyl-tRNA Synthetase (GARS) Protein Localization in a Drosophila Model 48

TABLES AND FIGURES 51
Table 1: Variants found in HARS 51
Figure 1: Verification and mapping of HARS variants 52
Figure 2: Conservation of HARS variants 53
Figure 3: Localization of HARS protein in vitro 54
Figure 4: HARS yeast complementation assays 55
Figure 5: HARS dimerization assays 57
Figure 6: HARS bisulfite sequencing 59
Figure 7: GARS protein localization in Drosophila motor neurons 60
Figure 8: Localization of mitochondria and GARS in sensory neurons of Drosophila 61
Figure 9: Wild-type and mutant GARS protein localization in Drosophila sensory neurons 63

REFERENCES 65

ABSTRACT

INVESTIGATING THE ROLE OF HISTIDYL-TRNA SYNTHETASE (*HARS*) AND GLYCYL-TRNA SYNTHETASE (*GARS*) MUTATIONS IN CHARCOT-MARIE-TOOTH DISEASE

by

Aimée Vester

Charcot-Marie-Tooth disease (CMT) is the most common type of inherited peripheral neuropathy, with a prevalence of 1 in 2,500 individuals world wide. Charcot-Marie-Tooth disease encompasses a family of neurodegenerative disorders that mainly affect motor and sensory function in the extremities. Four genes implicated in CMT disease thus far encode aminoacyl-tRNA synthetases (ARSs). ARSs are essential, ubiquitously expressed enzymes responsible for charging tRNA molecules, and are therefore necessary for protein translation in the cytoplasm and mitochondria. ARS enzymes are responsible for the first step of protein synthesis—they facilitate charging of the correct tRNA molecule onto the corresponding amino acid. Considering the critical role of ARSs in normal cellular function, and the frequency with which mutations in ARS genes associate with CMT, it is of interest whether any of the remaining ARS genes lead to CMT disease. In this study, variants of the histidyl-tRNA synthetase gene (*HARS*) were identified in patients diagnosed with CMT disease. The pathogenicity of these variants in *HARS* was assessed computationally and experimentally. In addition, previously identified pathogenic mutations in the *GARS* gene were studied functionally in a fly model system. Hopefully, functional data will provide further evidence for possible CMT disease mechanisms. This research will contribute to understanding of ARS protein functionality in peripheral neurons, and any role ARS proteins may have in CMT disease onset.

CHAPTER ONE

Introduction

I. Charcot-Marie-Tooth disease

Charcot-Marie-Tooth (CMT) disease is the most common type of inherited peripheral neuropathy, affecting about 1 in 2,500 individuals. The disease affects the extremities of the body, meaning the hands and feet, and leads to degeneration of the neurons in those structures. CMT is commonly organized into type 1 (CMT1) and type 2 (CMT2), based on electrophysiological symptoms. Patients with CMT1 exhibit lowered motor neuron conduction velocities (MNCV's). Normal MNCV's are usually 38 m/s (Reilly et al. 2011). CMT1 disease mostly affects Schwann cells, and patients with CMT1 show peripheral nerve demyelination. Schwann cells myelinate motor and sensory neurons in the peripheral nervous system; this myelination helps increase the speed of action potentials along the neurons (Purves et al. 2001). In contrast, patients with CMT2 exhibit normal MNCV's—above 38 m/s (Reilly et al. 2011). Usually, Schwann cells are not affected and peripheral nerve demyelination does not occur. However, axonal damage is observed in peripheral nerves, as well as a consequent decrease in nerve amplitudes (Reilly et al. 2011).

CMT leads to a multitude of symptoms that range in severity. These symptoms can include pes cavus, a steppage gait, and distal muscular weakness (Dyck and Lambert 1968). A patient with pes cavus exhibits a high-arched foot deformity (Reilly et al. 2011). Steppage gait symptoms cause a patient to walk in a high-stepping manner, due to weakening of the foot muscles and peroneal nerves (Casasnovas et al. 2008). Generally, though, symptoms are not life-shortening and patients remains ambulatory (Patzko et al. 2011). Current treatments for CMT are palliative, because there are no effective, CMT-specific medications (Patzko et

al. 2011). Examples of treatment include orthotic support, which is often used to help stabilize an individual when walking, and orthotic surgery, which may be elected to correct the pes cavus foot (Casasnovas et al. 2008). Although CMT disease is not fatal, CMT impacts quality of life. Given the prevalence of CMT, and the impact that the disease can have on an individual's functionality, finding effective clinical treatments is of importance.

Though a few clinical trials are underway, most helpful thus far are modern genetic discoveries that help clinicians make more accurate CMT diagnoses (Reilly et al. 2011). CMT was discovered in 1886, but it was not until 1991 that a gene was first implicated with the disease. At that time, a peripheral myelin protein 22 gene (*PMP22*) duplication was found on chromosome 17 in multiple families affected by CMT. The discovery led to association of *PMP22* with CMT type 1A (Lupski et al. 1991). Since then, forty genes have been implicated with CMT disease. Not surprisingly, CMT-associated genes encode proteins with critical functions in Schwann cells or axons (Patzko et al. 2011).

Genes associated with CMT1 are specific to Schwann cell function, whereas genes associated with CMT2 are specific to affect axon function. For example, *PMP22* encodes a glycoprotein expressed in compact myelin of peripheral nerves (Verhamme et al. 2011). A mouse modeling the *PMP22* duplication exhibited both delayed myelination and loss of myelinated fibers in peripheral neurons (Verhamme et al. 2011). In a second example, mutations in the neurofilament light subunit (*NEFL*) gene are associated with CMT2E. Neurofilament proteins are critical structural components of neurons, necessary for radial axon growth, and compose most of a mature axon's cytoskeleton (Lee and Cleveland 1994). Expression of CMT-associated *NEFL* mutations in motor neurons shows degeneration of motor neurons, as well as a loss of motor neuron viability that starts at the neurite projections

(Zhai et al. 2007). Thus, modern genetic studies provide promising mechanistic data regarding CMT disease. Such studies will inform later clinical solutions for patients with CMT.

II. Aminoacyl-tRNA synthetases

Four of the forty genes associated with CMT to date encode aminoacyl tRNA-synthetases. Glycyl-tRNA synthetase (*GARS*), tyrosyl-tRNA synthetase (*YARS*), alanyl-tRNA synthetase (*AARS*), and lysyl-tRNA synthetase (*KARS*) are all associated with various forms of the disease (Antonellis et al. 2003; Jordanova et al. 2006; Latour et al. 2010; McLaughlin et al. 2010). Notably, aminoacyl-tRNA synthetases (ARSs) function in both Schwann cells and axons. ARSs are ubiquitously expressed, essential enzymes, critical to the first step of protein synthesis. The enzymes aminoacylate, or charge, tRNAs in the mitochondria or cytoplasm with the corresponding amino acid. The tRNA: amino acid complex can then be transported to the ribosome for consequent polypeptide chain elongation. ARSs facilitate translation in the mitochondria or in the cytoplasm, though localization in these areas differs per enzyme (Antonellis et al. 2008).

As previously discussed, there are clear correlations between previously studied, CMT-associated genes and the type of CMT. It is interesting, therefore, that ubiquitously-expressed ARS genes have been implicated with CMT. *GARS* and *AARS* mutations are associated with CMT2. CMT2 does not affect MNCVs and exhibits primary peripheral axon degeneration. Appropriately, patients with *GARS* mutations exhibited normal MNCVs and degeneration of peripheral nerves (Ionasescu et al. 1996, Sivakumar et al. 2005). Patients with *AARS* mutations also showed normal MNCVs, and reduced nerve amplitudes (Latour et

al 2010). *YARS* mutations are associated with autosomal dominant intermediate CMT type C (DI-CMTC). DI-CMTC cases report intermediate MNCVs, as well as both axonal and demyelinating symptoms (Jordanova et al. 2006). *KARS* mutations are also associated with intermediate CMT. A single patient compound heterozygous for *KARS* mutations has intermediate CMT and exhibits intermediate MNCVs and decreased nerve amplitudes (McLaughlin et al. 2010). Thus, the majority of CMT cases reported in patients with ARS mutations show axonal deficits, not Schwann cell deficits. Questions now arise about the role of ARSs in normal axon function, and also about the neurotoxicity of variant ARS proteins.

The majority of CMT-associated ARS mutations are dominant missense mutations, present in a heterozygous state. Given the essential function of ARS enzymes to translation, several disease mechanisms have been proposed. These include: a) a loss of enzymatic activity; b) altered protein interactions; c) haploinsufficiency effects; and d) a toxic gain-of-function of the mutant protein (Antonellis et al. 2008, Antonellis et al. 2006). Current studies aim to clarify whether any of these mechanisms are in effect.

For many of the disease-associated mutations, a loss of enzymatic activity is detected in aminoacylation assays. A haploinsufficiency mechanism, in which insufficient charging of the amino acid onto the corresponding tRNA occurs, is one theory for impaired cell function. As shown in *in vitro* aminoacylation studies and yeast assay experiments, such a decrease in activity is possible. Several disease-associated mutations impair enzymatic activity or cause a loss-of-function in yeast cells, or both (Antonellis et al. 2006, Jordanova et al. 2006, McLaughlin et al. 2010, McLaughlin et al. 2011). Some, but not all disease-associated *GARS* mutations caused aminoacylation defects (Nangle et al. 2007). A similar finding occurred in *Drosophila* experiments modeling *YARS* mutations: impaired enzyme activity occurred in

two out of three studied DI-CMT-associated *YARS* variants (Storkebaum et al. 2009). In *AARS* studies, however, both disease-causing mutations exhibit reduced aminoacylation activity (McLaughlin et al. 2011). Additionally, studies in *KARS* showed that one disease-associated mutation impaired aminoacylation, while another caused a loss-of-function in yeast (McLaughlin et al. 2010).

The haploinsufficiency mechanism was studied *in vivo* in *GARS* mouse models. The mouse had a dominant *GARS* mutation and exhibited a CMT2D phenotype (Seburn et al. 2006). In the mouse, the variant *GARS* mRNA levels were equal to wild-type *GARS* mRNA levels in wild-type mice. Further studies in mice utilized a gene trap insertion that reduced *GARS* mRNA levels 50%. This mouse model showed that a 50% loss of GARS protein does not lead to a CMT phenotype (Seburn et al. 2006). The studies led to the conclusion that the CMT2D-causing *GARS* mutation does not cause disease via a haploinsufficiency mechanism, but rather via a gain of pathogenic function of the variant protein (Seburn et al. 2006).

Second, interactions of ARS proteins with each other and with other proteins in neurons could lead to altered neuronal functionality. For example, most ARS enzymes function as homodimers. Altered dimerization, then, could have mislocalization effects. Mislocalization of GARS granules was seen in cultured mouse motor neuron (MN-1) cells. This mislocalization could have occurred via impaired dimerization and aggregation of mutant proteins, leading to affected axonal transport (Antonellis et al. 2006).

Furthermore, mapping *GARS* mutations to the GARS protein showed that many of the disease-associated mutations are located along the GARS dimer interface, and that most mutations affect dimerization activity in some way (Nangle et al. 2007). Additionally, KARS functions as both a dimer and a tetramer, and disease-associated *KARS* mutations reside

within either the anticodon-binding or catalytic domain. One disease-associated *KARS* mutation, L133H, resides within the tetramer's dimer-dimer interface (McLaughlin et al. 2010).

In addition to functional studies characterizing variant ARS protein behaviors, computational studies in all ARS genes have shown strong correlations with disease. Conservation analysis, for example, shows the disease-associated ARS gene mutations often affect amino-acid residues highly conserved among evolutionarily-divergent species. An *AARS* mutation found in two separate French families affects a residue conserved in all species studied, including bacteria (Latour et al. 2010). Variations in highly conserved amino-acid residues in the remaining *GARS*, *YARS*, and *KARS* genes are also often pathogenic (Antonellis et al. 2003, Jordanova et al. 2006, McLaughlin et al. 2010).

Thus, studies of ARS genes and their role in CMT disease performed to date motivate further studies in other ARS genes, and more in-depth studies in ARS-CMT disease mechanisms. These studies have also provided valuable information about biochemical pathways of disease. Understanding biochemical CMT mechanisms is vital to further translational studies and, ultimately, the development of viable, CMT-specific cures. Investigating mutations of ARS enzymes, and the effects of these mutations, also elucidates the influence of ARS enzymes on normal peripheral nerve cell development and maintenance.

III. Proposed studies

The majority of my thesis focuses on mutations identified in patients with CMT in the histidyl-tRNA synthetase (*HARS*) gene (see Table 1). *HARS* encodes the HARS protein, an

enzyme responsible for ligating histidyl onto corresponding tRNAs in the cytoplasm. The HARS2 protein performs the analogous mitochondrial function (Freist et al. 1999). My initial studies identified candidate pathogenic mutations, using conservation sequence analysis and patient sequencing.

Though no previous studies have investigated the role of *HARS* in CMT disease, past biochemical, cell culture, and animal studies in other CMT-associated, ARSs led to significant findings about the proposed CMT disease mechanisms (Motley, Talbot and Fischbeck 2010, Antonellis et al. 2008). Now, it is of interest to look for analogous variant behaviors with mutant HARS proteins. Perhaps, we could then propose analogous disease mechanisms. My studies include localization of mutant proteins *in vitro,* yeast viability assays to test for a loss-of-function effect, and dimerization studies. It is anticipated that because HARS is also an essential, ubiquitous protein with structure and function similar to that of the other ARS enzymes, mutations in the gene will produce detrimental effects similar to CMT-associated mutations in the *GARS*, *YARS*, *AARS*, and *KARS* genes.

A second component of my thesis studies the localization of mutant GARS protein *in vivo.* Previous *in vitro* studies in MN-1 cells showed that disease-associated *GARS* mutations impair localization (Antonellis et al. 2006). Localization of GARS has never been studied in an *in vivo* model, however. In collaboration with the Collins laboratory in the Molecular, Cellular, and Developmental Biology Department, I utilized a *Drosophila* model system to localize and characterize *GARS* mutations. The *Drosophila* model system allowed me to investigate whether mutant GARS has a dominant negative effect on GARS localization *in vivo*. A dominant negative effect would cause mutant GARS to impair localization of the wild-type protein. I expressed wild-type *GARS* in motor neurons, and expressed both wild-

type and variant *GARS* in sensory neurons. The variant studied was *GARS* G240R—a mutation associated with CMT disease in humans (Antonellis et al. 2003), and associated with mislocalization *in vitro* (Antonellis et al. 2006).

CHAPTER TWO

Implicating Histidyl-tRNA Synthetase (*HARS*) in CMT Pathogenesis

I. Introduction

The human *HARS* gene resides on the long arm of chromosome 5, and encodes the histidyl-tRNA synthetase enzyme (HARS). HARS is an enzyme responsible for ligating histidine amino acid onto corresponding tRNA molecules during the translation process (Freist et. al 1999). Unlike the bi-functional GARS and KARS proteins, human HARS is a cytoplasm-specific protein. The mitochondria-specific counterpart of human *HARS* is *HARS2* (Antonellis et al. 2008). To date, there are no published studies associating *HARS* mutations with CMT, though the role of human HARS as an antigen in rheumatic arthritis and myositis (Freist et. al 1999; Soejima et. al 2011) has been investigated. This study focuses on seven *HARS* mutations identified in patients with CMT disease and no other known mutations. Specifically, our goal was to identify *HARS* mutations associated with a CMT phenotype.

To begin isolating candidate *HARS* mutations, we first confirmed all of the *HARS* mutations previously identified in a large-scale mutation analysis, which was performed by the National Institutes of Health Intramural Sequencing Center (NISC). Patient sequencing is an essential component in studying *HARS*, because identified variants may represent an artifact of DNA sequencing. We therefore confirmed each variant ourselves, prior to our assessments.

Next, we studied the conservation of HARS protein sequences among various species. Evolutionarily important amino-acid sequences are more likely to be conserved among many species—including eukaryotes and prokaryotes. A mutation in one of these well-conserved sequences, then, might be more likely to be disease-causing. Previous

comparative sequence analysis studies in the *GARS*, *YARS*, *AARS*, and *KARS* genes showed that the affected amino-acid residues of many disease-associated mutations are conserved well among vertebrate and invertebrate species. G240R in *GARS* for example, affects a residue that is conserved in many species, including fruit fly (*Drosophila melanogaster*) and baker's yeast (*Saccharomyces cerevisiae*), and is also associated with CMT2D (Antonellis et al. 2003).

After these computational studies, we employed functional studies to look for potential loss-of-function or neurotoxic effects of the *HARS* variants. First, we examined whether the *HARS* variants affected localization of HARS proteins in neurons. Previous studies of GARS protein localization in motor neurons showed distinct localization between wild-type and variant GARS proteins. Wild-type GARS shows unique, granular localization to the cell body and neurite projections. Many variant forms of GARS did not show this granular phenotype, and localization was instead more diffuse (Antonellis et al. 2006). To investigate whether HARS protein localization follows a similar pattern, we expressed both wild-type or variant HARS in MN-1 cells, using green fluorescent protein (GFP) to visualize proteins of interest.

Second, we assessed whether the *HARS* variants would lead to a loss of function in a yeast complementation assays. Yeast assays helped identify candidate mutations in previously studied ARS genes (Antonellis et al. 2006, Jordanova et al. 2006, McLaughlin et al. 2010). The yeast cell is a quickly growing model that facilitates good complementation testing because it is a diploid organism. The assays allow for complementation with either wild-type or variant ARS enzymes, so that viability of the variants in yeast cells can be determined. If there is a loss of viability, the variant protein could either have toxic effects to

the cell, or one copy of the normal protein could be insufficient for cell function. The yeast ortholog of *HARS* is *HTS1*, and *HTS1* is also cytoplasm-specific. Only human *HARS* variants able to be modeled in yeast were analyzed (R137Q, G205D, V238A, K376R, P505S). The A5E *HARS* variant could not be mapped to the orthologous sequence (see Figure 2). To determine whether presence of the mutant protein, alongside the wild-type protein, is enough to interfere with wild-type protein and normal yeast cell function, yeast growth curve analysis was also employed. Growth curve analysis was performed for the V238A and P505S *HARS* mutations, because these variants did not appear in public variant databases. This was discovered in earlier population screening work performed by Heather McLaughlin, in our lab.

The crystal structure of HARS revealed that the enzyme forms a homodimer in two prokaryotes (*E. coli* and *T. thermophilus*), and we might thus expect HARS to form a homodimer in cultured mammalian cells (Freist et al. 1999). A possible mechanism of disease is diminished dimerization activity (see Chapter 1, Section 2), and we sought to analyze both wild-type and variant HARS dimerization. Mapping of ARS gene mutations in previous studies (Nangle et al. 2007, McLaughlin et al. 2010, McLaughlin et al. 2011) showed that many disease-associated ARS gene mutations reside within a catalytic or anticodon-binding domain. As such, the identified *HARS* mutations were mapped to the HARS protein, with domains annotated. A mutation residing within a catalytic domain might be more likely to impair aminoacylation; a mutation residing within a dimerization domain might be more likely to impair dimerization.

We first studied dimerization using a mammalian two-hybrid assay. This assay utilizes luciferase reporter genes to study protein interactions *in vitro*. This two-hybrid

system uses three vectors: pACT, pBIND, and pG5*luc*. Constructs containing the proteins of interest, plus either pACT or pBIND are created. Interactions between the proteins of interest lead to interactions between pACT and pBIND and subsequently, luciferase gene transcription and expression. This is mediated by the VP16 activation domain of pACT, which interacts with the GAL4 binding of pBIND upon dimerization of the proteins of interest. pG5*luc* contains five GAL4 binding sites and a firefly luciferase gene. The activated pBIND construct binds to the GAL4 binding sites of pG5*luc,* initiating firefly luciferase gene expression. To normalize for different transfection efficiencies, pBIND also contains a *Renilla reniformis* luciferase gene. The ratio of *Renilla* and firefly luciferase expression is then measured in MN-1 cell lysates by a luminometer and analyzed in the assay.

Secondly, I attempted to study the dimerization of the wild-type HARS protein and R137Q protein with a DSS crosslinker assay. This assay uses disuccinimidyl suberate (DSS). DSS contains esters that react with amines in the proteins of interest and form amide bonds. These bonds crosslink any protein interactions, so that the proteins remain dimerized throughout lysis and denaturation steps of analysis. The DSS-treated samples are then compared to untreated samples via Western blot, and it is determined whether wild-type or variant HARS dimers have formed.

Besides studies to characterize interactions of the variant protein *in vitro*, it was also of interest to look at the genetic mechanisms by which variants in *HARS* could contribute to disease. The R137Q *HARS* mutation was not only found in a patient affect by CMT, but also in three additional individuals in the general population. These databases consist of genomic DNA from patients presumably unaffected by CMT disease. Various environmental influences could lead some individuals with the R137Q mutation, and not others, to express

CMT. However, it is of interest to determine whether this particular amino acid region is vulnerable to mutation, and whether the R137Q mutation is a recurrent mutation. Notably, the mutation affects a cytosine that resides within a CpG dinucleotide, and causes a cytosine to thymidine base change. Similarly, one of the disease-associated *AARS* variants, R329H, was found in three separate families. The mutation affects a cytosine residing in a CpG dinucleotide, and causes a cytosine to thymidine base change also. When methylated, CpG locations are particularly susceptible to mutation because the cytosine bases can be deaminated into thymidines. With the *AARS* studies, bisulfite sequencing was used to determine the level of methylation surrounding the R329H mutation, and whether the CpG affected by the R329H mutation was methylated (McLaughlin et al. 2011). Sodium bisulfite treatment of DNA results in the conversion of unmethylated cytosines to uracil residues. This treatment leaves methylated cytosines unchanged, and these changes can then be identified by DNA sequencing. The *AARS* studies revealed that there are multiple, methylated cytosines in the exon encompassing the R329H mutation. Furthermore, the R329H mutation resides on three different haplotypes in three distinct families. The *AARS* R329H mutation is, therefore, a recurrent mutation in the population (McLaughlin et al. 2011). Perhaps, *HARS* R137Q is also a recurrent mutation—it was found in four individuals. A bisulfite sequencing study was performed to determine if the mutation resides in a particularly susceptible area.

The computational and functional studies described herein aim to characterize *HARS* variants and determine if the identified mutations in *HARS* could contribute to CMT disease. Computational studies help to predict each mutation's potential for pathogenicity, and functional studies elucidate possible mechanisms of disease. Importantly, our results will

contribute to understanding of HARS protein's role in normal neuron function, and pinpoint loss-of-function, neurotoxic *HARS* mutations.

II. Methods

Patient sequencing

Primers were designed to amplify every *HARS* protein-encoding exon that encompassed each observed variant. 100 ng of DNA from patients with CMT was amplified with 125 ng of the corresponding primers, using Taq Polymerase (Invitrogen). PCR reaction mixtures contained 1 µL each of forward and reverse primers, 1 µL H_2O, sufficient genomic DNA to reach the 100 ng DNA amount, and sufficient Taq Polymerase to reach a final reaction volume of 25 µL. Reactions were cycled by a thermal-top cycler for approximately 25 cycles. Cycle conditions varied per amplicon. Primer melting temperatures were used to determine optimal annealing temperature ($T_{melting}$ - 5°C = $T_{annealing}$). PCR products were analyzed by electrophoresis through a 1% agarose gel stained with ethidium bromide. Products were column purified and sequenced by the UM DNA Sequencing Core for verification. Sequencing results were analyzed using Sequencher software (Gene Codes).

Comparative sequence analysis

Ortholog protein sequences to the human HARS protein sequence were obtained using the NCBI Database's BLAST tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequences obtained from diverse species were aligned using the CLUSTALW2 Multiple Sequence Alignment program (Larkin et al. 2007). The following species and GenBank accession numbers were used: human (*Homo sapiens* NP_002100.2), chimp (*Pongo abelii* XP_517980.2), mouse (*Mus musculus* NP_032240.3), chicken (*Gallus gallus*

NP_001006144.1), frog (*Xenopus laevis* AAH76748.1), fish (*Danio rerio* NP_001004586.1), fly (*Drosophila melanogaster* NP_728180.1), worm (*Caenorhabditis elegans* NP_001023373.1), plant (*Arabidopsis thaliana* NP_190196.1), baker's yeast (*Saccharomyces cerevisiae* AAA34696.1), and bacteria (*Escherichia coli* NP_289067.1). Each amino acid was analyzed via visual inspection.

Development of HARS expression constructs

The wild-type *HARS* open reading frame was previously cloned in-frame into pDONR221 (Invitrogen) (Anthony Antonellis). To generate *HARS* constructs expressing the desired amino-acid changes within the *HARS* open reading frame, a QuickChange II XL Site-Directed Mutagenesis kit (Stratagene) was used. Primers were designed, utilizing the QuickChange Primer Design Program (www.agilent.com/genomics/qcpd), to express each *HARS* mutation. 125 ng of primer was used in each mutagenic reaction, along with 5 µL of 10x reaction buffer (Stratagene), 3 µL of QuikSolution reagent (Stratagene), 1 µL of dNTP mix (Stratagene), and 1 µL of the DNA template (Wt. *HARS* in pDONR221). The final reaction volume was brought to 50µL with ddH_2O, and mixed. 1 µL of *PfuUltra* HF DNA polymerase (Stratagene) was added. A thermal cycler was used to cycle reactions at standard mutagenic parameters for 18 cycles. After cycling, parental DNA was digested with 1 µL *Dpn*I restriction enzyme (Stratagene). Reaction mixtures were digested for 1 hour at 37°C. DNA was transformed into competent *E. coli* cells (Invitrogen). 5µL of DNA was placed into 12µL of cells and incubated on ice for 30 minutes. After incubation, heat shock was performed at 42°C for 30 seconds. 63 µL of S.O.C. medium (Invitrogen) was added immediately afterwards. Cells were incubated at 37°C for one hour. 50µL of the reaction

mixture was spread onto agar medium containing kanamycin and incubated overnight at 37°C.

HARS protein localization in cultured neurons

DNA constructs containing either wild-type or variants *HARS* open reading frames, in-frame with a green fluorescent protein (GFP) tag, were developed. To do so, the *HARS* open reading frame, originally in the pDONR221 vector (Invitrogen), was cloned into pEGFP-N2 (Clontech) using Gateway LR Clonase Enzyme II (Invitrogen). DNA was transformed into competent cells, transformed onto agar plates containing kanamycin, and incubated overnight at 37°C. Bacterial colonies were grown in bacterial media containing kanamycin and incubated overnight at 37°C. Plasmid DNA was then extracted with a spin miniprep kit (Qiagen). DNA was digested with *Bsr*GI restriction enzyme (New England Biosystems), to ensure that *HARS* had been cloned into peGFP-N2 successfully. The digestion cocktail contained 2 µL 10X Bovine Serum Albumin (New England BioLabs), 2 µL Buffer 2 (New England BioLabs, 0.2 µL *Bsr*GI (New England BioLabs) 0.5 µL DNA, and 15.3 µL diH_2O. DNA was digested at 37°C for one hour. DNA was sequenced to ensure correct insertion into pEGFP-N2. The University of Michigan DNA Sequencing Core sequenced all DNA and Sequencher software (Gene Codes) was used to analyze sequencing results.

MN-1 cells used for transfections were cultured in DMEM growth medium (Invitrogen) supplemented with 10% fetal bovine serum, 100 U/mL penicillin, 50 µg/mL streptomycin, and 2 mM L-glutamine. MN-1 cells were grown at 37 °C in 5% CO_2, and were counted using a Countess Automated Cell Counter (Invitrogen) for all experiments. Approximately 1.25 x 10^5 cells were added per well, to a 4-well polysterene tissue culture

treated glass slide (BD Biosciences) for each transfection. Constructs containing wild-type or mutant *HARS* and GFP were transfected into MN-1 cells using Lipofectamine 2000 reagent. A mixture of 1.25 µL Lipofectamine 2000 reagent (Invitrogen) and 125 µL OptiMEM growth medium (Invitrogen) was incubated at room temperature for 10 minutes. 1.5 µg of purified plasmid DNA was diluted in 125 µL OptiMEM growth medium, and added to the first Lipofectamine-OptiMEM mixture. This cocktail was incubated at room temperature for 20 minutes. MN-1 cells were washed once with 1 x PBS, and 250 µL of the reaction cocktail was added. Cells were incubated for 4 hours at 37°C, and 500 µL DMEM growth medium (described above) was added for incubation. Cells were incubated 48 hours before fixation (McLaughlin et al. 2011).

To fix, MN-1 cells were washed once with 1x PBS for 5 minutes, incubated in 4% paraformaldehyde for 10 minutes, and washed three times with 1x PBS for 5 minutes. Nuclei were stained with 300 nM DAPI diluted in 1X PBS for 5 minutes; cells were washed three times with 1X PBS for 5 minutes after staining. All steps were performed at room temperature. The well apparatus was removed, and ProLong anti-fade reagent (Invitrogen) added. Slides were covered and sealed with nail polish. Images were taken using a fluorescent microscopy. A Carl Zeiss Lumar V12 Stereo microscope was used to image cells, and AxioVision version 4.5 software was used for analysis.

Yeast complementation assays

HARS and *HTS1* protein sequences were aligned using CLUSTALW software (Larkin et al. 2007) as described above. Mutagenic primers were designed to mutate the *HTS1* open reading frame, so that it would harbor each human amino acid change (see Figure 4a).

To feasibly knock out one of the endogenous *HTSI* alleles in the assay but maintain the diploid yeast (*Saccharomyces cerevisiae*) strain, a maintenance vector (described below) was introduced via tetrad dissection. Tetrad dissections selected for strains that did not have the endogenous *HTS1* allele (*hts1*Δ), but that did have the maintenance vector. The maintenance vector used contained wild-type *HTSI* in pRS316. The pRS316 vector (American Type Culture Collection) encodes URA3, which was used for selection. The *URA3* gene encodes orotidine 5-phosphate decarboxylase, an enzyme necessary for pyrimidine ribonucleotide synthesis.

A construct containing wild-type *HTSI* in pRS316 was made from wild-type *HTSI* in pDONR221, using the Gateway LR Clonase Enzyme II (Invitrogen). A 10μL reaction mixture was used, containing 7 μL of the entry clone (Wt.*HTSI* in pDONR221), 1 μL pRS316, and 2 μL of Gateway LR Clonase Enzyme II (Invitrogen). After a 1 hour incubation period at room temperature, 1 μL of Proteinase K solution was added, and the mixture was incubated for another 10 minutes at 37°C. 4 μL of the reaction was then transformed into competent *E. Coli* cells. 50 μL of the transformation cocktail was plated onto agar medium containing ampicillin and incubated overnight at 37°C. Bacterial colonies were placed into medium containing ampicillin and incubated at 37°C overnight. Plasmid DNA was extracted using a QIAPrep Miniprep Kit (Qiagen). The DNA was digested with *Bsr*GI restriction enzyme (New England Biosystems) to verify cloning had occurred. The digestion reaction was incubated at 37°C for one hour. The reaction mixture used was the same as the *Bsr*GI reaction described in the previous section.

Next, constructs containing either wild-type or variant *HTSI* and the pRS315 vector (American Type Culture Collection) were created. *HTSI* variants were modeled in

pDONR221 (Invitrogen) using a QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene) and the appropriate mutant-harboring oligonucleotides. The modified *HTS1* open reading frame was then cloned into pRS315 using Gateway LR Clonase Enzyme II (Invitrogen). The same method to clone and purify DNA was used as in the previous section. pRS315 is a LEU2-bearing vector—*LEU2* encodes an enzyme necessary for leucine synthesis (beta-isopropylmalate dehydrogenase).

The strain containing *hts1*Δ and the maintenance vector was transformed with wild-type or variant *HTSI*-pRS315. A transformation with empty pRS315 was also performed as a control. Lithium acetate transformations used 200 ng of DNA, and were performed at 30°C. Reactions were then plated onto yeast SD medium lacking leucine and uracil (Teknova) and incubated at 30°C for 2-3 days. Yeast colonies were grown in yeast SD medium lacking leucine and uracil (Teknova) at 30°C overnight. 1:1, 1:10, and 1:50 dilutions were plated on 5-FOA plates (Teknova) and incubated at 30°C for 2 days. The *URA3* gene encodes a product that reacts with 5-FOA and produces fluorouracil, which is toxic to cells. Cells containing the pRS316 vector therefore died, and those containing the pRS315 vector did not. This allows for the selection of yeast cells that contain wild-type or variant *HTS1* in pRS315, and that have spontaneously lost the maintenance vector. Viability was determined by inspection after 2-3 days of incubation at 30°C.

Yeast growth curve analysis

The aforementioned yeast strains, containing *hts1*Δ and wild-type or variant *HTS1* in pRS315, were grown overnight in yeast SD medium lacking leucine (Teknova) at 30°C. Optical density measurements were taken the next morning. Dilutions were made with SD medium lacking leucine, to obtain an equal starting concentration of 0.05 OD for each

sample. After reaching equal concentrations, optical density measurements were taken every hour for eight hours, and once after 25 hours. Eight measurements were taken for all samples. Data were analyzed using Microsoft Excel software. The average OD value was plotted at each time point for samples wild-type or variant *HTS1,* and standard deviations were indicated.

Dimerization studies: Mammalian two-hybrid assays

To test wild-type and mutant HARS protein interactions, we created constructs expressing wild-type or mutant *HARS* in pACT or pBIND (Promega). The *HARS* open reading frame was cloned into each vector, in frame. This was achieved by cloning wild-type or variant *HARS* in pDONR221 into pACT or pBIND with Gateway LR Clonase Enzyme II (Invitrogen). Target constructs created contained one of four combinations: wild-type *HARS*-pACT, *HARS*.R137Q-pACT, wild-type *HARS*-pBIND, or *HARS*.R137Q-pBIND. A 10µL reaction mixture was used, containing 7 µL of wild-type or variant *HARS* in pDONR221, 1 µL of pACT or pBIND, and 2 µL of Gateway LR Clonase Enzyme II (Invitrogen). After a 1 hour incubation period at room temperature, 1 µL of Proteinase K solution was added, and the mixture was incubated for 10 minutes at 37°C. 4 µL of the reaction was transformed into competent *E. Coli* cells. 50 µL of the transformation cocktail was plated onto agar medium containing ampicillin and incubated overnight at 37°C. Bacterial colonies were placed into medium containing ampicillin and incubated at 37°C overnight. Plasmid DNA was extracted using a QIAPrep Miniprep Kit (Qiagen), and digested with *Xba*I restriction enzyme (New England Biosystems) to verify cloning had occurred. The digest reaction mixture contained 15.5 µL H_2O, 2 µL 10X Buffer 2 (New England Biosystems), 0.5 µL *Xba*I restriction enzyme, and 2 µL DNA. The digestion reaction was incubated at 37°C for 1 hour. DNA was

sent to the UM Sequencing Core to verify correct insertion, and sequencing results were analyzed with Sequencher software (Gene Codes).

Wild-type or variant *HARS* constructs were transfected into MN-1 cells using Lipofectamine 2000 reagent (Invitrogen). Transfections were performed in polysterene 96-well tissue culture dishes (Corning), with approximately 10,000 cells per well. MN-1 cells were cultured as previously described (see "HARS protein localization in cultured neurons"). 0.25 μL of Lipofectamine 2000 was added to 25 μL OptiMEM medium (Invitrogen) and incubated for 10 minutes at room temperature. 300 ng of DNA was diluted in 25 μL of OptiMEM medium and added to the Lipofectamine-OptiMEM mixture. This was incubated for 20 minutes at room temperature. Cells were washed with 1X PBS and 50 μL of the transfection mixture was added. After incubation at 37°C for 4 hours, cells were aspirated and 75 μL of DMEM growth medium (Invitrogen) was added. Cells were incubated for 48 hours at 37°C prior to luciferase assays.

To perform luciferase assays, MN-1 cells were washed with 1X PBS and lysed for 1 hour at room temperature with 1X Passive Lysis Buffer (Promega). 10 μL of lysed cells were transferred to a polysterene 96-well assay plate (Corning). Firefly luciferase and *Renilla* luminescence of the cells was analyzed using the Dual Luciferase Reporter 1000 Assay System (Promega) and a Glomax Multi-Detection System (Promega) (Hodonsky et al. 2011). The ratio of firefly to *Renilla* luminescence was compared between each of the following groups: wild-type protein, mutant protein, positive control, and negative control. Data were analyzed using Microsoft Excel software, and the mean and standard deviation values calculated (see Figure 5a).

Dimerization studies: DSS crosslinker assays

Constructs expressing wild-type or R137Q *HARS* and a C-terminal V5 tag for detection were developed. The *HARS* open reading frame, either wild-type or with the R137Q variant, was cloned, in frame, into pDEST40 (Invitrogen). The cloning reaction contained 7 µL of wild-type *HARS* in pDONR221, 1 µL of pDEST40 vector, and 2 µL of Gateway LR Clonase Enzyme II (Invitrogen). After incubating for 1 hour at room temperature, 1 µL of Proteinase K solution was added. The mixture was incubated for 10 minutes at 37°C, and 4 µL of the reaction mixture was transformed into competent *E. Coli* cells. 50 µL of the transformation cocktail was plated onto agar medium containing ampicillin and incubated overnight at 37°C. Bacterial colonies were placed into medium containing ampicillin and incubated at 37°C overnight. Plasmid DNA was extracted using a QIAPrep Miniprep Kit (Qiagen), and digested with *Bsr*GI restriction enzyme (New England Biosystems) to ensure the cloning reaction had occurred. The digestion reaction was incubated at 37°C for 1 hour, and contained 2 µL 10X Bovine Serum Albumin (New England BioLabs), 2 µL Buffer 2 (New England BioLabs, 0.2 µL *Bsr*GI (New England BioLabs) 0.5 µL DNA, 15.3 µL diH_2O.

MN-1 cells were transfected with 4 µg of DNA, using Lipofectamine 2000 reagent (Invitrogen) and OptiMEM medium (Invitrogen). Cells were incubated for 48 hours at 37ºC before treatment. Samples were washed with 1X PBS and resuspended in 50 mM HEPES buffer. To crosslink HARS protein, cells were treated with 50 mM DSS (Pierce) to achieve a final crosslinking concentration of 5mM. After incubation at 4ºC for 1 hour, the reaction was quenched with 3 µL of 1M Tris-HCl buffer (pH 7.5) for 15 minutes. Cells were collected and lysed with 1X protease inhibitor in 1% NP-40 (Pierce) for 20 minutes on ice.

Protein lysates were added to 25 µL of 2X Novex® Tris-Glycine Sample Buffer (Invitrogen) and 1 µL β-mercaptoethanol and boiled at 99 ºC for 5 minutes. Denatured protein samples were placed on ice and then electrophoresed through a Novex® 4-20% Tris-Glycine gel (Invitrogen) at 200V for 1 hour. Protein was blotted onto transfer membrane using the iBlot Dry Blotting System (Invitrogen) for 7 minutes. The membrane was blocked in 100 mL of 1X PBS, 50 µL of Tween-20, and 5 g of nonfat dry milk at 4ºC overnight. The membrane was washed in 1X PBS plus 0.05% Tween-20, prior to incubation in a 1:2000 dilution of anti-v5 antibody in blocking solution for 1 hour, all at room temperature. The membrane was washed three times in 1X PBS plus 0.05% Tween-20, and then incubated in secondary antibody conjugated against mouse for 1 hour at room temperature. The secondary antibody was diluted 1:2000 in blocking solution, and the membrane was washed three times in 1X PBS plus 0.05% Tween-20 afterwards. A detection solution containing 5 mL of ECL Western Blotting Reagent 1 (Pierce) and 5 mL of ECL Western Blotting Reagent 2 (Pierce) was placed on the membrane for 1 minute and excess solution was removed. The membrane was wrapped in plastic, placed into a cassette, and exposed to film for development.

HARS Bisulfite sequencing

Two samples of unaffected genomic DNA were treated with sodium bisulfite, following the protocol of an EpiTect Bisulfite Treatment kit (Qiagen). Primers, designed by Heather McLaughlin, surrounded the sodium bisulfite-converted region of *HARS* exon 5 with the R137Q mutation: (forward: [5'-TTTTTGTTATTAGGTTTTTTTTGTT A-3'] reverse: [5'-AAACCCCAAATAACTCATCCTAC-3']). The reaction mixture contained 5 µL of 10X Jumpstart reaction buffer (Sigma), 1 µL of dNTP mix (New England Biolabs), 0.5 µL of forward and reverse primers (125 ng total), 2 µL of genomic DNA, and 40 µL of H_2O. The

reactions were cycled using a thermal-top cycler, for 40 cycles. Reaction cycling conditions were: initial denaturation at 94ºC for 2 minutes, denaturation at 94ºC for 30 seconds, annealing at 50ºC for 30 seconds, extension at 72ºC for 30 seconds, and a final extension step at 72ºC for 7 minutes. Reactions were kept at 4ºC until samples were electrophoresed through a 1% low-melt agarose gel containing ethidium bromide. A QIAQuick Gel Extraction Kit (Qiagen) was used to purify DNA, once the correctly-sized DNA band was cut out of the gel. Using a TOPO TA Cloning Kit for Sequencing (Invitrogen), DNA was cloned into a pCR4-TOPO TA vector (Invitrogen) to be sequenced. The cloning reaction contained 1 µL of purified DNA, 1 µL of salt solution (Invitrogen), 1 µL of the pCR4TOPO vector (Invitrogen), and 3 µL of diH_2O (Invitrogen). The reaction was incubated at room temperature for 5 minutes and placed on ice prior to transformation into competent *E. coli* cells. 4 µL DNA and 12 µL *E. coli* cells were used in the transformation. 50 µL of the transformation cocktail was plated on agar medium containing kanamycin and incubated overnight at 37 ºC. Colonies were picked and grown overnight in bacterial medium containing kanamycin, at 37 ºC. DNA was purified using a QIAPrep Miniprep Kit (Qiagen) and digested with *Bsr*GI restriction enzyme to ensure the cloning reaction had occurred. The digestion reaction contained 1 µL of 10X Bovine Serum Albumin (New England BioLabs), 1 µL of Buffer 2 (New England BioLabs, 0.1 µL of *Bsr*GI (New England BioLabs) 0.5 µL of DNA, and 7.4 µL of diH_2O and was incubated at 37ºC for 1 hour. DNA was sent to the UM Sequencing Core for sequencing, and results were analyzed using Sequencher software (Gene Codes).

III. Results

HARS variant validation in patients with CMT

To confirm that the *HARS* variants identified by the aforementioned, large-scale mutation analysis (see Chapter 2, Section 1) were not artifacts of DNA sequencing, we sequenced DNA of patients with CMT. Patient sequencing results, shown in Figure 1a, show that mutations A5E, R137Q, G205D, V238A, K376R, and P505S are present in corresponding patients with CMT. These patients are all heterozygotes and therefore display both a normal and mutant allele in sequencing chromatogram results, as indicated by the arrow. One mutation identified by the computational screen, T406N, was not identified in the corresponding patients with CMT. It is possible that T406N was an artifact of DNA sequencing in the original NISC screen. Thus, the T406N variant would not be expected to associate with CMT disease because the mutation is not present in patients with CMT. Confirming that the other *HARS* variants exist in patients with CMT disease, though, is an important first step in assessing each variant because we want to show a relationship between each variant and CMT disease.

HARS protein sequence conservation in multiple species

The conservation of the original amino-acid residue in multiple species was examined for all variants, in hopes of finding residues that were conserved well among evolutionarily divergent species. Well-conserved amino-acid residues could be more likely to lead to detrimental effects if mutated. As shown in Figure 2, the alanine variant, A5E, was conserved among human, chimpanzee, mouse, frog, and worm species. The glycine variant, G205D, was conserved among all vertebrate species studied, in addition to fruit fly and worm species. The valine variant, V238A, was conserved among all vertebrate species studied except

chimpanzee, plant, and bacteria. The lysine variant, K376R, was conserved among all vertebrate species studied. The proline variant, P505S, was conserved only among human and mouse species. Notably, the arginine variant, R137Q, was conserved among all vertebrate and invertebrate species analyzed. The arginine 137 amino acid was conserved in more species than any of the other amino-acid residues studied. This might indicate that this particular arginine resides at a more functionally important region of HARS, because the amino acid was conserved well evolutionarily.

Variant HARS does not mislocalize in cultured neurons

To characterize localization of wild-type HARS within neurons and determine whether mutant HARS exerts a dominant negative effect on wild-type HARS, we studied the localization of HARS in cultured neurons. Perhaps, mutant HARS impairs localization of wild-type HARS within cells. Initial results suggest that wild-type HARS protein localizes diffusely to the nucleus and throughout the axons, when examined using fluorescent microscopy (Figure 3). There was no significant difference in gross localization of wild-type and variant HARS protein in MN-1 cells, at this resolution. This differs from the localization patterns seen with GARS protein, where granular wild-type localization differed noticeably from the more diffuse variant protein localization (Antonellis et al. 2006). Differential localization was also seen with the YARS proteins, though not with the AARS and KARS proteins (Jordanova et al. 2006, McLaughlin et al. 2010, McLaughlin et al. 2011). It appears that mutant HARS does not impair localization of wild-type HARS *in vitro*, and does not act in a dominant-negative manner. Perhaps, HARS acts through a different pathogenic mechanism than GARS or YARS.

The R137Q HARS variant causes a loss-of-function in yeast

HARS mutations were modeled in yeast to test whether each mutation would complement deletion of *HTS1* (the *HARS* yeast ortholog). Mutations unable to complement deletion of *HTS1* could affect yeast cell function via toxicity or haploinsufficiency mechanisms. Previous studies in *GARS*, *YARS*, *KARS*, and *AARS* modeled disease-associated ARS gene mutations in yeast and determined that these mutations were not able to complement deletion of the yeast orthologs (Antonellis et al. 2006, Jordanova et al. 2006, McLaughlin et al. 2010, McLaughlin et al. 2011). The wild-type *HTS1* construct led to full yeast cell viability, indicating that a successful, viable experimental construct was created in the study. The empty pRS315 vector did not confer yeast viability when transformed. This result was predicted because *HTS1* is essential to normal yeast cell function, but confirmed our predictions. Yeast cell viability similar to that of the wild-type *HTS1* construct occurred in experiments with the V238A, K376R, and P505S *HARS* variants. Conversely, the R137Q *HARS* variant did not confer yeast cell viability, indicating that it did not complement the knocked-out *HTS1* allele. The inability of R137Q *HARS* to complement yeast cell function and confer viability suggests a loss-of-function effect, which is important in determining which *HARS* mutations could be potentially disease-causing. All yeast complementation results were replicated at various dilutions, as indicated in Figure 4b.

V238A and P505S HARS variants do not affect yeast growth

We wanted to determine if mutant HTS1 impairs wild-type HTS1 and subsequently, yeast cell growth via a dominant negative mechanism. We tested this with growth curve analyses. Yeast growth was assessed over 25 hours for the V238A and P505S *HTS1* variants (see Figure 4c). No yeast growth showed a marked decline when variant HTS1 was present,

indicating that there were no measurable toxic effects of the variant protein within the measured time frame. This suggests that the two variant HTS1 proteins do not exert dominant negative, toxic effects on wild-type proteins in the yeast cell. These variants also do not cause a loss-of-function, and therefore appear to be less detrimental to cell function in this *in vivo* model.

No HARS dimerization occurs in mammalian two-hybrid assays

We know that HARS forms a dimer in two bacterial species (Freist et al. 1999), and therefore wanted to test whether mutant HARS impairs dimerization activity of the protein. The provided positive controls showed markedly high luminescence ratios, as expected (see Figure 5a). However, the luminescence ratio of wild-type HARS dimerization activity was close to 1, indicating that dimerization of the wild-type HARS protein was not achieved. Thus, differential dimerization activity between the wild-type and variant HARS proteins could not be assessed with this assay.

Mutant HARS does not impair dimerization

Since I was unable to achieve the expected wild-type HARS dimerization in the two-hybrid assay, I performed crosslinker analysis to determine if variant HARS impaired HARS dimerization activity. Our initial results indicate that dimerization of HARS R137Q is not impaired, though there is a high amount of background in the results (see Figure 5b). This indicates that R137Q HARS is still able to dimerize. This is an important finding for the dominant negative pathogenicity model, because R137Q HARS protein interactions have to occur for the mutant protein to cause toxic effects.

HARS Bisulfite Sequencing

The R137Q *HARS* variant was found in one patient with CMT and three additional individuals in the general population, and we therefore tested if the variant resides in a region susceptible to mutation, and could possibly represent a recurrent mutation. DNA sequencing results from two individuals without CMT disease were analyzed to determine the amount of methylated cytosines at each CpG island (see Figure 6). Methylated CpG islands are susceptible to mutation because the cytosine can be deaminated into a thymidine (McLaughlin et al. 2011). For the first control individual, three clones were analyzed, and four clones were analyzed for the second control individual. There are eight CpG islands within exon 5 of *HARS*, which encompasses the R137Q variant. In the seven clones analyzed, 93.8% of the eight CpG islands in exon 5 were methylated. Our initial results suggest that the R137Q mutation resides within an area of *HARS* that is highly methylated, and would probably be more susceptible to base changes.

IV. Discussion

Our analyses suggest that the A5E, G205D, V238A, K376R, T406N, and P505S *HARS* variants are less likely to be pathogenic. This is indicated both by initial computational studies and subsequent functional studies. The computational studies showed that while some of the variants indeed reside at amino acid residues conserved among diverse species, none are conserved in as many species as arginine 137. Functional studies showed that the other variants do not display gross mislocalization *in vitro*, and that variants complement yeast cell function as well as the wild-type protein.

Computational and functional studies presented here suggest that the R137Q *HARS* variant is a loss-of-function mutation. The conservation of arginine 137 in various vertebrate and invertebrate species provides the first piece of evidence that the variant resides in an important protein region. Highly conserved variants in previously-studied ARS genes have also been associated with CMT disease (Antonellis et al. 2003, Jordanova et al. 2006, McLaughlin et al. 2011).

Perhaps most importantly, the inability of the R137Q *HARS* variant to complement yeast cell function in viability assays indicates that the variant causes a loss of function. Previous yeast complementation studies with other ARS genes indicate the reliability of yeast assays. Two variants in *AARS* that did not complement yeast cell viability also impaired aminoacylation activity *in vitro*, and one of these two variants was found in three separate families affected by CMT (McLaughlin et al. 2011). The complementation assay itself does not necessarily determine whether the variant protein acts in a dominant-negative or haploinsufficient manner. It is possible that, via a dominant-negative mechanism, the variant protein confers toxicity in yeast cells through added functions or protein aggregation effects. It is also possible, via a haploinsufficient mechanism, that the variant protein does not allow for sufficient enzymatic activity, or mischarges tRNAs in yeast cells.

Gross localization of HARS R137Q protein did not differ from that of the wild-type protein, suggesting that protein mislocalization might not be responsible for pathogenicity in cells. However, the diffuse localization of wild-type HARS also contrasts with the granular pattern of GARS and YARS proteins *in vitro* (Antonellis et al. 2006, Jordanova et al. 2006). Perhaps mislocalization of variant HARS protein does occur, but at a resolution not able to be examined via our fluorescent microscopy. It is also possible that the protein does not fully

localize throughout neurite projections, and this was not observed because the full length of the neurite projections was not visualized.

Based on our initial dimerization experiments, R137Q HARS dimerization is not affected. However, more conclusive results, perhaps obtained with a co-immunoprecipitation assay, would be beneficial in evaluating disease mechanisms. The mammalian two-hybrid assay was unsuccessful in studying HARS dimerization. We expected that the positive control would show markedly higher luminescence, because the positive control constructs expressed proteins that showed strong dimer interactions. The negative control is determined when the only proteins present in the transfection are the pACT and pBIND vectors, which does not result in increased luminescence of either firefly or *Renilla* luciferase in the absence of target protein interactions. The firefly: *Renilla* luminescence ratio is set at one, and all other values are expressed relative to this control ratio. We expected the wild-type HARS dimerization interactions to produce a luminescence ratio somewhere above that of the negative control, because we know that wild-type HARS forms a homodimer in two bacterial species (Freist et al. 1999). We did not know what to expect for the R137Q variant proteins, but would expect that any decrease in dimerization caused by the mutation would result in a lowered luminescence ratio. Repeated results showed, however, that both the wild-type and variant luminescence ratios were close to one, indicating that there were no dimerization interactions. Furthermore, the DSS crosslinker assay showed preliminary, but not conclusive dimerization results either. Though dimerization was observed in some of the experiments, the protocol should be improved to reduce Western blot background, and ensure that all HARS protein bands can be visualized.

Impaired HARS dimerization could alter the previously described aminoacylation process, impair localization of the ARS proteins to the extremities, or lead to protein aggregation. A malformed dimer could change the shape of the site responsible for tRNA charging and for carrying the corresponding amino acid during the aminoacylation process. Histidine, for example, is one of few amino acids with an imidazole ring side chain (Freist et al. 1999), and this relatively unique side chain could have important binding site shape requirements that are easily disrupted by altered dimerization. Furthermore, changes in dimerization could result in protein localization differences in the long axons of the periphery. A monomer or partial dimer could interact sub-optimally with transport machinery. Lastly, protein aggregation of non-dimerizing (and potentially misfolded) proteins could occur, leading to build-up of otherwise unusable proteins in parts of the neuron. This could then have toxic effects on the cell.

Previous studies in other CMT-associated, ARS proteins have led to significant advances in proposed CMT disease mechanisms (Motley, Talbot and Fischbeck 2010, Antonellis et al. 2008). Though the mechanism of action of previously studied ARS mutations is still unknown, various mechanisms are under consideration. Current research favors a dominant-negative mechanism (McLaughlin et al. 2011, Stum et al. 2011). This is because the ARS mutations are heterozygous, missense mutations usually resulting in a dominant phenotype (McLaughlin et al. 2011). Dominant-negative mutations would interfere with wild-type protein function. Altered localization of the protein might also hamper normal cell processes. Haploinsufficiency in aminoacylation pathways seems less likely a mechanism for ARS mutations, because mice containing a *Gars* null allele did not develop neuropathy (Seburn et al. 2006). *Gars* is the mouse ortholog to human GARS. Given these

proposed mechanisms, it is now of interest to further investigate functionality of *HARS* variants and determine whether disease mechanisms analogous to those proposed in other ARS mutations exist. For one, the *HARS* R137Q mutation is a heterozygous, missense mutation, and could also operate in a dominant-negative way.

CHAPTER THREE

Evaluating Glycyl-tRNA Synthetase (GARS) Protein Localization in a *Drosophila* Model System

I. Introduction

Prior *in vitro* studies indicate that certain disease-associated *GARS* mutations mislocalize when expressed in cultured neurons (Antonellis et al. 2006). It is still unclear what the physiological significance of this *in vitro* finding is, however, and an *in vivo* model would be beneficial. Perhaps, the mutant GARS protein has a dominant negative effect on wild-type GARS, and mutant GARS proteins undermine the ability of the wild-type proteins to localize in axons. In collaboration with the Collins laboratory, I examined *GARS* mutations in flies. These studies allowed me to localize GARS in an *in vivo* model, to determine whether localization of GARS variants differs from localization of wild-type GARS, as was seen in the previous *in vitro* studies (Antonellis et al. 2006). The fly model's *in vivo* data helps contextualize the *in vitro* findings. Mislocalization of mutant proteins *in vivo*, would help validate the *in vitro* system as a localization model for *GARS* mutations. *In vivo* experiments would also determine whether mutant proteins change localization of endogenous fly proteins. This would support the idea that mutant GARS acts through a dominant negative mechanism.

Previous studies of *YARS* mutations in *Drosophila* indicate the feasibility of modeling human CMT disease-associated ARS mutations in a fly model. Biochemical pathways are conserved well between humans and flies, and flies are thus useful models of neurodegeneration. Furthermore, there is strong conservation between human and *Drosophila* ARS proteins (Storkebaum et al. 2009). Modeling of three disease-associated

YARS mutations in *Drosophila* resulted in a neurodegenerative phenotype (Storkebaum et al. 2009). The mutant flies exhibited impaired motor performance when mutant YARS was expressed both ubiquitously and specifically in neurons. Electrophysiological and morphological neuronal defects were also evident, as well as terminal axonal degeneration (Storkebaum et al. 2009).

A previous study of *GARS* in *Drosophila* indicates that CMT-like phenotypes can be modeled in flies, and that human GARS can act on neuronal fly processes (Chihara et al. 2007). In the study, wild-type human GARS was over-expressed in *Drosophila* modeling a homozygous *Aats-gly* mutation. Human GARS was able to fully rescue the neuronal projection defects in dendritic and axonal arborization (Chihara et al. 2007). Two disease-associated mutations in human *GARS* were also over-expressed in the fly model. These mutations were not able to rescue neuron function. This indicates that the two mutations cause a loss of function, and that disease-associated *GARS* mutations can be modeled in the *Drosophila* model system with a subsequent loss of function (Chihara et al. 2007).

To mediate expression of *Aats-gly* in *Drosophila,* we used a GAL4-UAS system to express the protein in target locations. GAL4 is a yeast transcription factor, not endogenously functional in *Drosophila*. GAL4 recognizes a UAS (Upstream Activating Sequence) element. When constructs containing the UAS element and gene of interest are developed, expression of the gene is directed to target areas of the animal using specific promoters (Jones et al. 2009). Two different promoters were used to mediate GARS expression in the flies. The BG380 promoter mediates expression in motor neurons, whereas the Pickpocket (ppk) promoter mediates expression in sensory neurons.

Wild-type GARS shows distinct, granular localization in cultured motor neuron cells, and this localization differs in variant GARS forms (Antonellis et al. 2006). Our first goal was to determine whether similar, granular localization patterns emerge *in vivo*. To do so, we analyzed *Aats-gly* expression in the motor and sensory neurons, and also the neuromuscular junctions of *Drosophila*. Secondly, we preliminarily tested the co-localization of mitochondria and a wild-type *Aats-gly* construct lacking the mitochondrial targeting sequence (Chihara et al. 2007). Lastly, we studied the localization of mutant *Aats-gly* and determined if it localizes differently than wild-type *Aats-gly* proteins. Differential localization might suggest that the mutant protein interferes with wild-type protein localization.

II. Methods

Genetics

The strains used were: BG380-Gal4 (Budnik et al. 1996), UAScyto*gars*-myc (Chihara et al. 2007), UAS-eGFP, UAScyto*gars*G240R, UAS-CD8RFP, ppk-Gal4, and UAS-mitoGFP.

Immunocytochemistry

Wandering third instar larvae were dissected with 1x PBS. Upon dissection, animals were fixed with 4% paraformaldehyde for 30 minutes, and washed with 1x PBS plus 0.3% Triton (Xiong et al. 2010). Animals were blocked in 1X PBS, 0.3% Triton, and 5% normal goat serum for about an hour at room temperature, and then incubated in mouse anti-myc antibody, diluted to 1:50 in 1X PBS, 0.3% Triton, and 5% normal goat serum at 4°C overnight. After 3 washes in 1X PBS plus 0.3% Triton of 5 minutes each, animals were

incubated in secondary antibody conjugated against A488 for mouse, and Cy3 antibody conjugated against horseradish peroxidase (HRP). HRP is a plant glycoprotein that specifically stains neural tissue in flies (Paschinger et al. 2009). Both secondary antibodies were diluted to 1:1000 in 1X PBS, 0.3% Triton, and 5% normal goat serum, and animals were incubated for a few hours at room temperature. Animals were washed 3 times in 1X PBS, 0.3% Triton, and 5% normal goat serum, and placed in 50% glycerol. Animals were mounted onto glass slides, and Vectashield reagent (Vector Labs) was added. Slides were covered and sealed with nail polish, and kept at 4°C prior to imaging.

Imaging

All images were taken at room temperature, using a spinning-disk confocal microscope (PerkinElmer). The confocal system contained a scanner, an electron microscopy charge-coupled device camera mounted on an inverted microscope, and 100x1.46 NA oil objectives. Volocity software (PerkinElmer) was used for all imaging, and similar imaging settings were used for each experiment (Xiong et al. 2010).

III. Results

GARS protein localization in Drosophila motor neurons

To determine whether GARS localizes in a granular pattern, as observed in cultured neurons (Antonellis et al. 2006), we localized *Aats-gly* in fly motor neurons *in vivo.* Our initial results showed rather diffuse localization of wild-type GARS throughout the neuromuscular junction (see Figure 7). These results suggest that localization of GARS in an *in vivo* system does not follow the granular staining of GARS *in vitro*, and that further studies *in vivo* might therefore show mutant protein interactions that differ from previous

experiments. The initial results might also suggest that localization functions differently in cultured cells than it does in live cells.

GARS protein localization in Drosophila sensory neurons

To characterize localization of wild-type *Aats-gly* in the ventral nerve cord and projecting cell bodies, and determine whether mutant *Aats-gly* localizes differently in those structures, we expressed the protein in *Drosophila* using a sensory neuron-specific promoter. In the ventral nerve cord and projecting cell bodies, localization of wild-type *Aats-gly* appeared punctate. Individual neurons also showed punctate *Aats-gly* staining in axonal projections (see Figure 9a). Next, we tested whether the cytoplasm-specific *Aats-gly* aggregates localized with mitochondria. When mitochondria were co-stained, the cytoplasm-specific *Aats-gly* expression appeared more diffuse than mitochondrial staining (see Figure 8), and only some of the large *Aats-gly* protein aggregates appeared to co-localize with the mitochondrial structures. This indicates that the observed cytoplasm-specific *Aats-gly* puncta are most likely not protein aggregates solely within mitochondria.

Co-expressing the G240R *GARS* mutation in *Aats-gly* with wild-type, cytoplasmic *Aats-gly* does not appear to cause significant localization defects of wild-type cytoplasmic *Aats-gly* protein in *Drosophila*. Qualitative inspection shows similar granular localization patterns of myc-stained *Aats-gly* in the ventral nerve cord and in cell bodies (see Figure 9), as was observed when solely myc-stained *Aats-gly* was expressed in those structures. This indicates that the mutant protein might not inhibit the ability of wild-type GARS to localize in neurons. However, we were unable to measure levels of the mutant protein's expression, so we do not know if the level of mutant protein expression was perhaps too low to cause mislocalization effects.

IV. Discussion

With an early fly model of *Aats-gly* protein localization, we explored whether GARS localization would mimic GARS localization seen in earlier *in vitro* studies (Antonellis et al. 2006). Results from our localization studies in motor and sensory neurons indicate that the *in vivo* fly model does show some localization characteristics exhibited in previous experiments. Specifically, granular staining appears to occur in sensory neurons, similar to granular staining in earlier *in vitro* studies in cultured MN-1 cells (Antonellis et al. 2006). We observed rather large protein aggregates of the cytoplasm-specific *Aats-gly* protein. The expression construct used was developed in an earlier *GARS* fly model, and lacks the mitochondrial targeting sequence of *Aats-gly* (Chihara et al. 2007). The large protein aggregates we observed in sensory neurons appeared similar in size to mitochondria. Since we were expressing cytoplasm-specific *Aats-gly*, it would be an interesting finding if the cytoplasm-specific protein was also expressed within mitochondria. Co-staining of mitochondria and cytoplasmic *Aats-gly* to identify these protein aggregates showed that *Aats-gly* protein did localize with some, but not solely, mitochondria.

Utilizing the fly model, we also tried to determine whether mutant *Aats-gly*, expressing the G240R *GARS* mutation, localizes differently in sensory neurons. We did not see a significant difference in localization of wild-type *Aats-gly* when G240R *GARS* was expressed, however. This suggests that, qualitatively, the *in vivo* study does not replicate the results of the previous *in vitro* studies, in which mislocalization of some mutant GARS protein occurs. It should be noted, however, that we did not test the expression levels of mutant GARS. Possibly, the levels of wild-type *Aats-gly* were sufficient for normal localization, and there was not enough mutant protein present to interfere with normal protein

function. We might want to test other fly strains with the G240R *GARS* insertion, because they could express the mutant protein at different levels and lead to a mislocalization phenotype.

Additionally, we could perhaps also model the mutant protein's effects in a more quantitative manner, to test for more subtle effects the mutant protein could have on endogenous protein localization and function. A dominant-negative mechanism would imply that localization of the endogenous protein is somehow altered by presence of the mutant protein. The experiments shown cannot account for alterations that occur at a much higher resolution. Live imaging in fly axons could possibly be used to visualize transport of *Aats-gly* within the axon, and determine whether any transport defects or alterations occur when *GARS* mutations are expressed.

The ability to visualize wild-type and variant GARS in motor and sensory neurons informs hypotheses about the role of GARS protein and ARSs in axons. A dominant negative mechanism seems less likely for the G240R *GARS* mutant, based on our very early *in vivo* studies. However, subtler interactions of mutant GARS protein are possible, and could still point to a gain of function of the mutant proteins in patients with CMT. In addition, other *GARS* mutations could lead to disease through a different mechanism than that of the G240R *GARS* variant. Perhaps, other mutations in *GARS* do cause mislocalization *in vivo*, and better illustrate the previously proposed dominant-negative mechanism. Expressing other disease-associated *GARS* mutations in the described fly model would test this. Also, expressing the G240R *GARS* mutation and other disease-associated *GARS* mutations in motor neurons, instead of sensory neurons, might lead to different results. The previous studies resulting in mislocalization of mutant GARS were performed in mouse motor neuron cells (Antonellis et

al. 2006). Lastly, live imaging to visualize any transport of *Aats-gly* protein that occurs in fly neurons would also test whether localization impairment occurs more subtly. The mutant protein could slow movement of wild-type *Aats-gly* to certain locations within cells; this effect cannot be visualized by our still-image assays, because wild-type *Aats-gly* would still localize to the affected locations in time.

CHAPTER FOUR

Conclusions and Future Directions

I. Implicating Histidyl-tRNA Synthetase (HARS) in CMT Disease Pathogenesis

Mutations were previously identified in *HARS,* in patients with CMT. Given the prevalence of CMT-associated ARS genes, our main goals were to isolate *HARS* mutations most likely to associate with CMT disease and study possible mechanisms by which *HARS* mutations could cause CMT disease. The study was directed by earlier investigations of the disease-associated mutations in *GARS*, *YARS*, *AARS*, and *KARS* (Antonellis et al. 2003, Antonellis et al. 2006, Jordanova et al. 2006, Latour et al. 2010, McLaughlin et al. 2011, McLaughlin et al. 2010).

First, we confirmed that the mutations identified in the large-scale mutation analysis performed by the NISC were indeed present in patients with CMT. We examined the evolutionary conservation of the HARS protein sequence in diverse species and inspected the affected amino-acid residue of each confirmed *HARS* variant. These initial studies helped identify those mutations not likely to be associated with CMT disease because they were not present in patients with CMT, or because they were not conserved well among other species.

Many previously studied ARS gene mutations associate with axonal CMT disease, suggesting that the pathogenicity of ARS gene mutations affects axon function. Therefore, we examined ways in which mutant HARS could affect axon function. We looked for a loss-of-function in yeast complementation assays, studied mutant HARS localization in cultured neurons, and analyzed mutant HARS dimerization. We also initially determined whether the R137Q *HARS* variant, found in three individuals in the general population, resided in a region of the gene with high levels of CpG methylation.

The studies described herein help determine possible ways that *HARS* mutations lead to axon dysfunction in CMT. One of the variants, R137Q *HARS*, caused a loss of function in the yeast complementation assay, indicating that the mutant protein was not able to complement deletion of the endogenous allele. This suggests that the mutant protein either exerts toxic effects on the wild-type protein, or does not produce sufficient activity for normal yeast cell function. The finding that dimerization of mutant HARS is not affected indicates that a haploinsufficient mechanism is less likely. A lack of dimerization might lead to mutant protein aggregation or mislocalization of non-dimerized proteins in the axon. Gross localization of wild-type and mutant HARS was not different in *in vitro* studies of cultured neurons, however.

Appearance of the R137Q *HARS* variant in the general population databases suggests that the mutation could have incomplete disease penetrance and recur in the population (Vester et al. 2012). Initial bisulfite sequencing analysis indicates that the mutation resides within an area containing significant CpG island methylation. The previously studied R329H mutation in *AARS* also showed significant methylation in the area encompassing the R329H mutation. Notably, the R329H mutation was found in three separate families with CMT disease, and was determined to be a recurrent mutation (McLaughlin et al. 2011, Latour et al. 2009).

The described studies of *HARS* mutations isolate a variant likely to associate with CMT disease—the studies suggest that the R137Q *HARS* variant represents a loss of function allele. This is indicated by the conservation of the arginine 137 residue among various species, and the loss of function in yeast. In further support of this, studies by Asim Beg in

the Department of Pharmacology model the R137Q variant in *C. elegans*, and indicate that the variant is neurotoxic *in vivo* (Vester et al. 2012).

Future studies ought to expand and further validate the aforementioned studies. First, studies to more conclusively determine effects of mutant HARS on dimerization would confirm initial dimerization activity results. Clearer dimerization results could perhaps be achieved through co-immunoprecipitation experiments, or the use of a different antibody in the aforementioned DSS assay protocol. To further validate the initial bisulfite sequencing results, we will need to sequence other exons of *HARS*, to ensure that they do not also exhibit high levels of methylation. High levels of methylation in other areas of the gene would not support the idea that the R137Q resides within a particularly mutable region of *HARS*. Also, we would want to bisulfite-treat a region with an identified CpG island, as a control for our assay. The *AARS* studies performed such a control by amplifying a *SOX3* region containing a CpG island (McLaughlin et al. 2011).

Information about the R137Q *HARS* mutation in additional patients with CMT would also help to associate the variant with the disease. Though data is available for one of three individuals with the R137Q *HARS* mutation found in the general population, data is not available for the other two individuals. This data would be helpful in studying the incomplete penetrance of the R137Q *HARS* mutation. There was relatively little data about family members of the individual carrying the R137Q *HARS* mutation, making it difficult to conclusively implicate the variant with disease. This illustrates the challenges of associating genes with disease when little familial data is available. Advances in genetics allow us to screen genomes at a much faster rate. The data available, however, are not necessarily easily

associated with disease and subsequent translational research is often far out of reach. Thus, cost and time-effective methods of testing ARS mutations for pathogenicity are desirable.

II. Evaluating Glycyl-tRNA Synthetase (GARS) Protein Localization in a Drosophila Model

Our *GARS* studies in *Drosophila* aimed to a) compare GARS localization *in vivo* to GARS localization exhibited in the previous *in vitro* model, and b) determine whether mutant GARS localizes differently from wild-type GARS—perhaps through a dominant negative mechanism that impairs localization of the wild-type protein. As discussed previously, a dominant-negative disease mechanism is the favored pathogenic mechanism for ARS mutations, based on current studies (McLaughlin et al. 2011, Stum et al. 2011). Dominant-negative interactions mandate that mutant proteins interfere with wild-type protein function. ARS mutations are heterozygous, missense mutations that usually render a dominant phenotype (McLaughlin et al. 2011), suggesting that such interactions are quite possible.

The G240R *GARS* mutation segregates with CMT2D in a large pedigree (Motley, Talbot and Fischbeck. 2010). In addition, the mutant protein was inactive in aminoacylation assays (Nangle et al. 2007), and mislocalizes when expressed in cultured motor neurons (Antonellis et al. 2008). Given the strong association with CMT of the G240R *GARS* mutation, the mutation's inactivity in enzyme activity studies, and the mutation's mislocalization *in vitro,* we hypothesized that the mutant protein would lead to altered localization *in vivo*. The G240R *GARS* mutation has not been tested in fly or mouse models expressing *GARS* (Chihara et al. 2007, Seburn et al. 2006). Therefore, the described experiments provide the first *in vivo* model of the G240R *GARS* mutation.

To establish the initial model of GARS localization in *Drosophila*, we first characterized expression of wild-type *Aats-gly* in motor and sensory neurons. *Aats-gly* is the fly ortholog to human *GARS*; *Aats-gly* is sometimes also referred to as *gars* (Chihara et al. 2007). The *in vitro* model exhibited characteristic granular staining in cultured neurons (Antonellis et al. 2006), and we tested whether this also occurred in flies. In addition, co-staining of mitochondria and wild-type *Aats-gly* in sensory neurons further characterized the distribution of cytoplasm-specific, wild-type *Aats-gly* within sensory neurons. We observed co-staining of some, but not all, *Aats-gly* protein aggregates with mitochondria (see Figure 8). The co-staining was an interesting preliminary finding, given that the cytoplasm-specific *Aats-gly* construct lacks the mitochondrial targeting sequence normally found in the *Aats-gly* open reading frame (Chihara et al. 2007). Lastly, we assayed whether the G240R *GARS* mutation expressed in the *Aats-gly* fly ortholog would alter localization of wild-type *Aats-gly* protein. This tested whether the mutant proteins altered localization of wild-type protein. Mislocalization of mutant GARS expressed in cultured neurons occurred previously (Antonellis et al. 2006), and suggested that mutant proteins altered localization of endogenous GARS—perhaps via dominant negative mechanism. Our initial results showed that expression of the G240R *GARS* mutation did not significantly alter localization of wild-type *Aats-gly* (see Figure 9), though we would want to replicate our experiments with different strains to ensure optimal expression levels of G240R *GARS*.

To continue investigating possible dominant-negative mechanisms of disease in CMT, we could expand the *Drosophila* studies. More quantitative experiments might include measuring intensity of wild-type *Aats-gly* expression, and comparing that to the intensity of wild-type *Aats-gly* expression when mutant G240R *GARS* in *Aats-gly* is also present. We

could also investigate transport defects via live-imaging studies. *Aats-gly* exhibited granular localization in *Drosophila* sensory neurons, and sometimes associated with mitochondria. Testing transport of *Aats-gly* in the mitochondria and along axons could reveal more minute toxic effects the mutant protein has on wild-type *Aats-gly*. Furthermore, we tested interactions of mutant protein with wild-type *Aats-gly* by visualizing the wild-type protein and assessing whether there are localization differences. Developing transgenic flies that overexpress myc-tagged, mutant *Aats-gly* would test for mislocalization of mutant *Aats-gly* protein, and would allow for mutant protein visualization in a more direct way than the current experiment.

In vivo models are important for associating ARSs with CMT, and for investigating pathogenic mechanisms of the disease. The G240R *GARS* mutation did not show a loss of function in *in vivo* yeast complementation assays (Antonellis et al. 2006), but did show enzyme activity and localization defects (Nangle et al. 2007, Antonellis et al. 2006). *In vivo* studies could help clarify whether the mutant protein operates via a haploinsufficient or dominant negative mechanism. In contrast, the R137Q *HARS* mutation does not exhibit mislocalization in cultured neurons, but does exhibit a loss of function in yeast complementation assays (see Chapter 2, Section 3). *In vivo* studies in a worm model show that the mutation is neurotoxic, suggesting a dominant negative, gain-of-function effect of the mutant protein (Vester et al. 2012). These studies exemplify that the effects of various ARS gene mutations are not consistent in similar studies. Diverse, physiologically relevant models can therefore be useful in assessing these mutations.

TABLES AND FIGURES

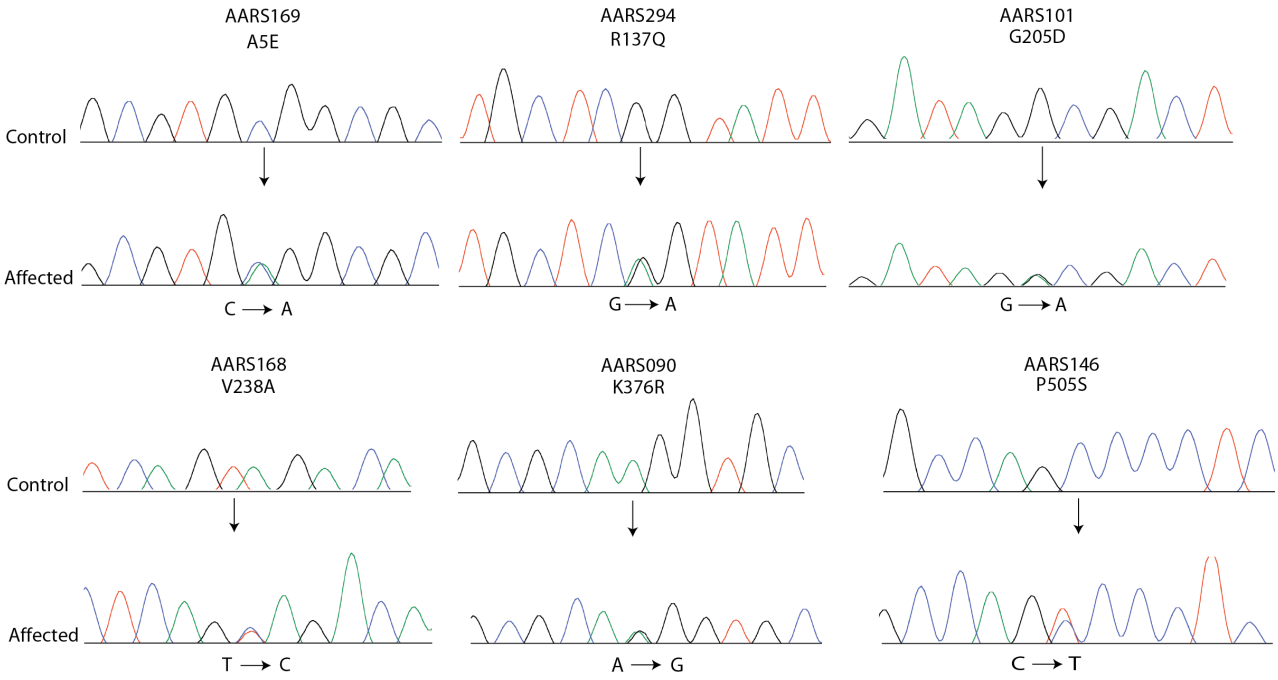
Table 1: Variants found in HARS

Mutation	Wt. Amino Acid	Codon	Wt. Group	Mut. Amino Acid	Mut. Codon	Mut. Group
A5E	Ala	GCG	Non-polar	Glu	G*A*G	Neg. charged
R137Q	Arg	CGG	Pos. charged	Gln	C*A*G	Polar
G205D	Gly	GGC	Polar	Asp	G*A*C	Neg. charged
V238A	Val	GTA	Non-polar	Ala	GCA	Non-polar
K376R	Lys	AAG	Pos. charged	Arg	A*G*G	Pos. charged
P505S	Pro	CCC	Non-polar	Ser	*T*CC	Polar

***HARS* variants identified in a large-scale screen of all ARS genes and verified in patients**. Under the "Mutation" column, the original amino-acid residue is given, then the affected amino-acid position, and then the mutated residue. The amino acids are shortened using standard three-letter amino acid nomenclature. In the "Mut Codon" column, affected bases are italicized. Throughout the table, "Wt." indicates wild-type, and "Mut." indicates mutant.

Figure 1: Verification and mapping of HARS variants

a.) **Verification of identified *HARS* mutations in patients with CMT.** Sequencing chromatograms show the area around each labeled variant, with the location of the variant indicated by the arrow. All patients are heterozygous for the *HARS* variants, and therefore both the mutated and wild-type nucleotides are shown.



b.) **Mapping of *HARS* mutations to HARS protein domains.** All variants were mapped to a depiction of the HARS protein. Identified functional domains of the protein are shown: in green is the WHEP-TRS domain, in red the catalytic domain, and in blue the tRNA-binding domain. Position numbers of the domains are indicated below the protein depiction.

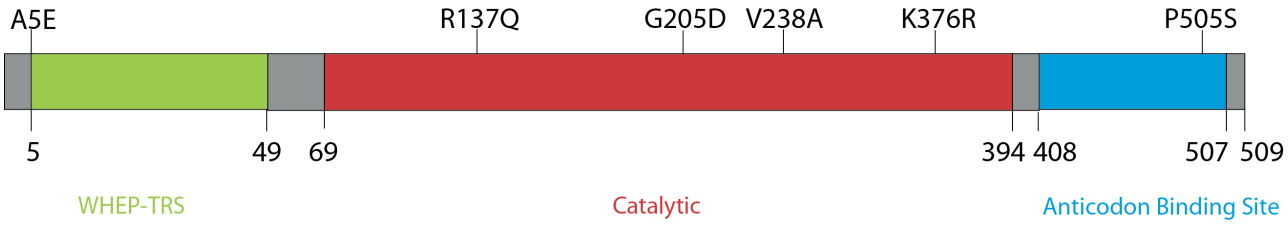


Figure 2: Conservation of HARS variants

Conservation of affected amino-acid residues in multiple species. Conservation of all variants was investigated at each variant amino-acid position, using multiple-species protein sequence alignments. The species that each protein sequence originates from is listed at the left. The amino acid change for each variant is listed above the sequence regions. Sequence regions flank the original amino acid, with red indicating where the original amino acid is conserved.

	A5E	R137Q	V238A	K376R	P505S
human	-MAER**A**ALEEL	TVPFA**R**YLAMN	TICSS**V**DKLDK	-PKGR**K**VPCVG	RRTGQ**P**LCIC-
chimp	LPRRAWASLLS	TVPFA**R**YLAMN	AICSSIDKLDK	-PKGH**K**VPCVG	KRLSES-----
mouse	MADRA**A**LEELV	TVPFA**R**YLAMN	TICSS**V**DKLDK	-PKGR**K**VPCVG	RRTNQ**P**LSTC-
chicken	-MAD----EAA	TVPFA**R**YLAMN	TICSS**V**DKLDK	-PKGR**K**VPCVG	RRTNQL-----
frog	GADRAELEKAL	TVPFA**R**YLAMN	TICSS**V**DKLDK	-PKGR**K**VPCVG	KRTKIL-----
zebrafish	VSMRLCAGLMG	TVPFA**R**YLAMN	TICST**V**DKLDK	-PKAG**K**CPVWG	----------
fly	SDTREQILEQI	TVPLA**R**YLAMN	TICSA**V**DKLDK	-PRGKAVPCVG	RRQQASA----
worm	AERKAILMQEA	TVPFA**R**YLAMN	TICSS**V**DKLDK	-SKAN-VPCCG	DTLAAL-----
plant	VTTRLSSSFRP	TPSLA**R**LVIQK	RVCIIIDKIEK	---GDDFPACG	----------
baker's yeast	----------	TVPFA**R**YVAMN	KISSA**V**DKLDK	GKKSTQIPCVG	EKLSQIHEDGL
bacteria	----------	TAGCV**R**AGIEH	FLEQHKEKLD-	---GRATPAVG	TLLG-------

Figure 3: Localization of HARS protein in vitro

Variant HARS does not affect gross protein localization in cultured neurons. Cells were imaged using fluorescence microscopy, with staining for HARS-GFP expression (in green) and staining for nuclei with DAPI (in blue).

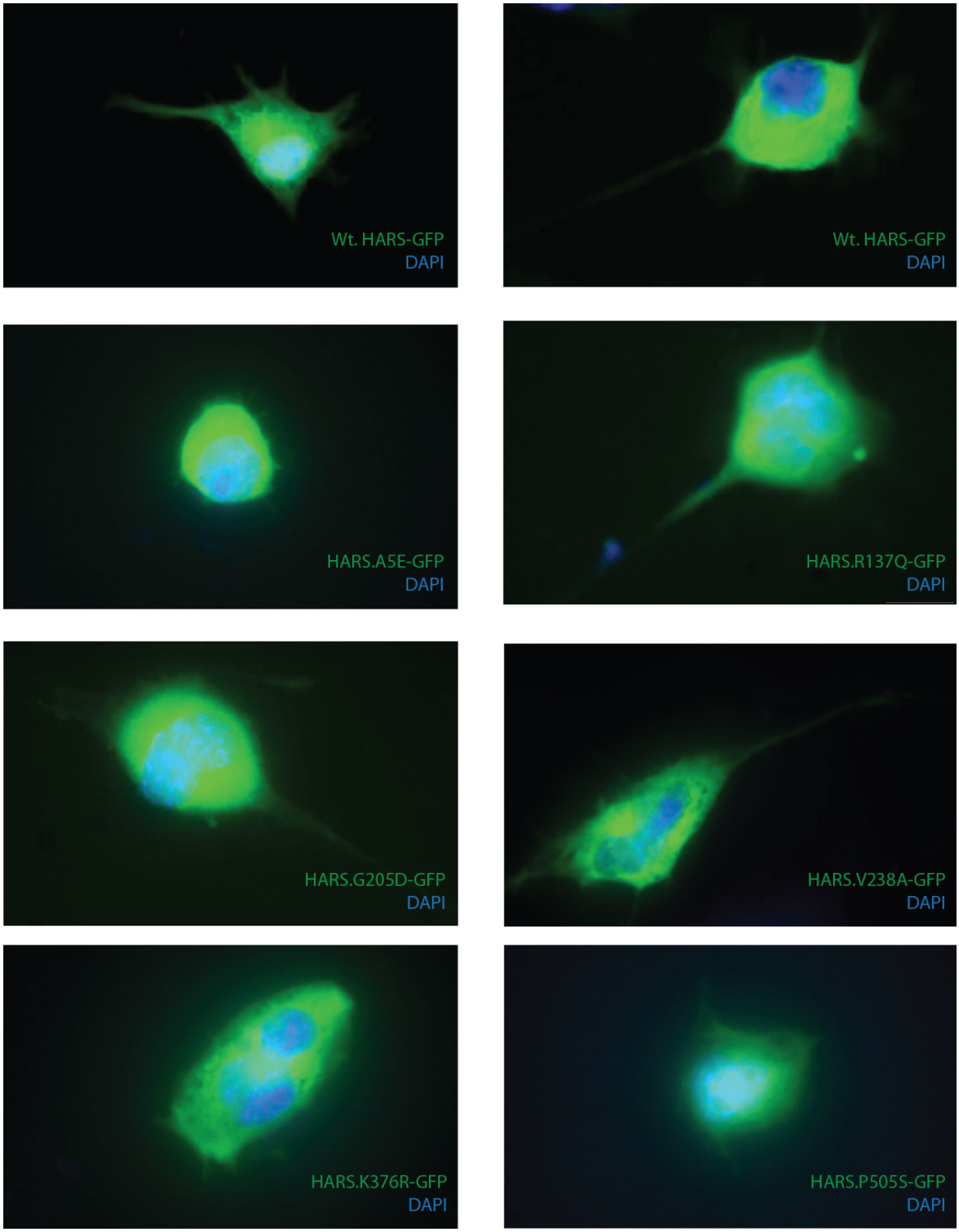


Figure 4: HARS yeast complementation assays

a.) ***HARS* variants expressed in a yeast model.** Human *HARS* variants were modeled in yeast. *HTS1* is the orthologous yeast protein sequence to *HARS*. One *HARS* mutation, A5E, was not modeled because the mutation could not be mapped in the yeast protein sequence (see Figure 2).

HARS variant	*HTS1* variant
R137Q	R116Q
G205D	K184D
V238A	V217A
K376R	K371R
P505S	V519S

b.) **R137Q *HARS* variant causes a loss of function in yeast.** Cultures of each yeast strain were grown on 5-fluoroorotic acid medium, at dilutions of 1:1, 1:10 and 1:100. Strains were transfected with the experimental vector. The experimental vector: 1) did not contain wild-type *HTS1* (empty pRS315); 2) contained wild-type *HTS1;* or 3) contained the human *HARS* variants expressed in *HTS1*, as labeled.

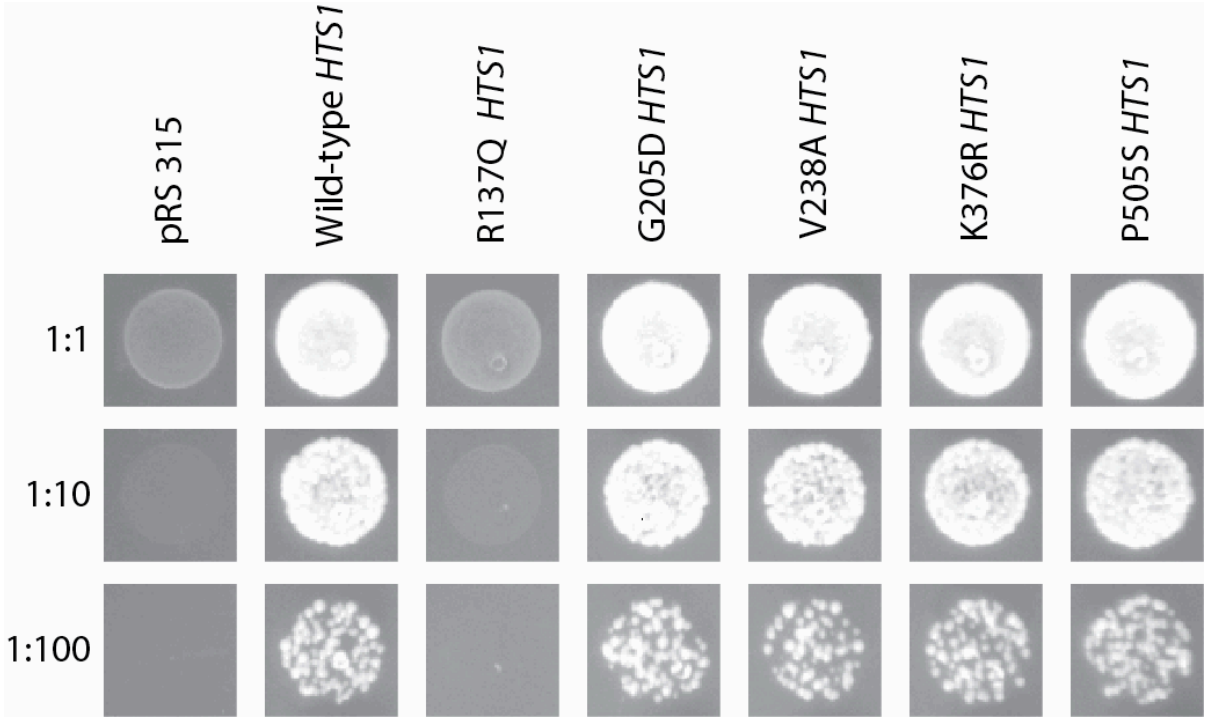


Figure 4: HARS yeast complementation assays (continued)

c.) **V238A and P505S *HARS* variants do not alter yeast growth.** The average OD_{600} measurement for three different clones of the wild-type strain, as well as for each variant, is shown. The ratio of variant to wild-type yeast growth is graphed. The error bars for each point indicate standard deviation. (Figure by Anthony Antonellis)

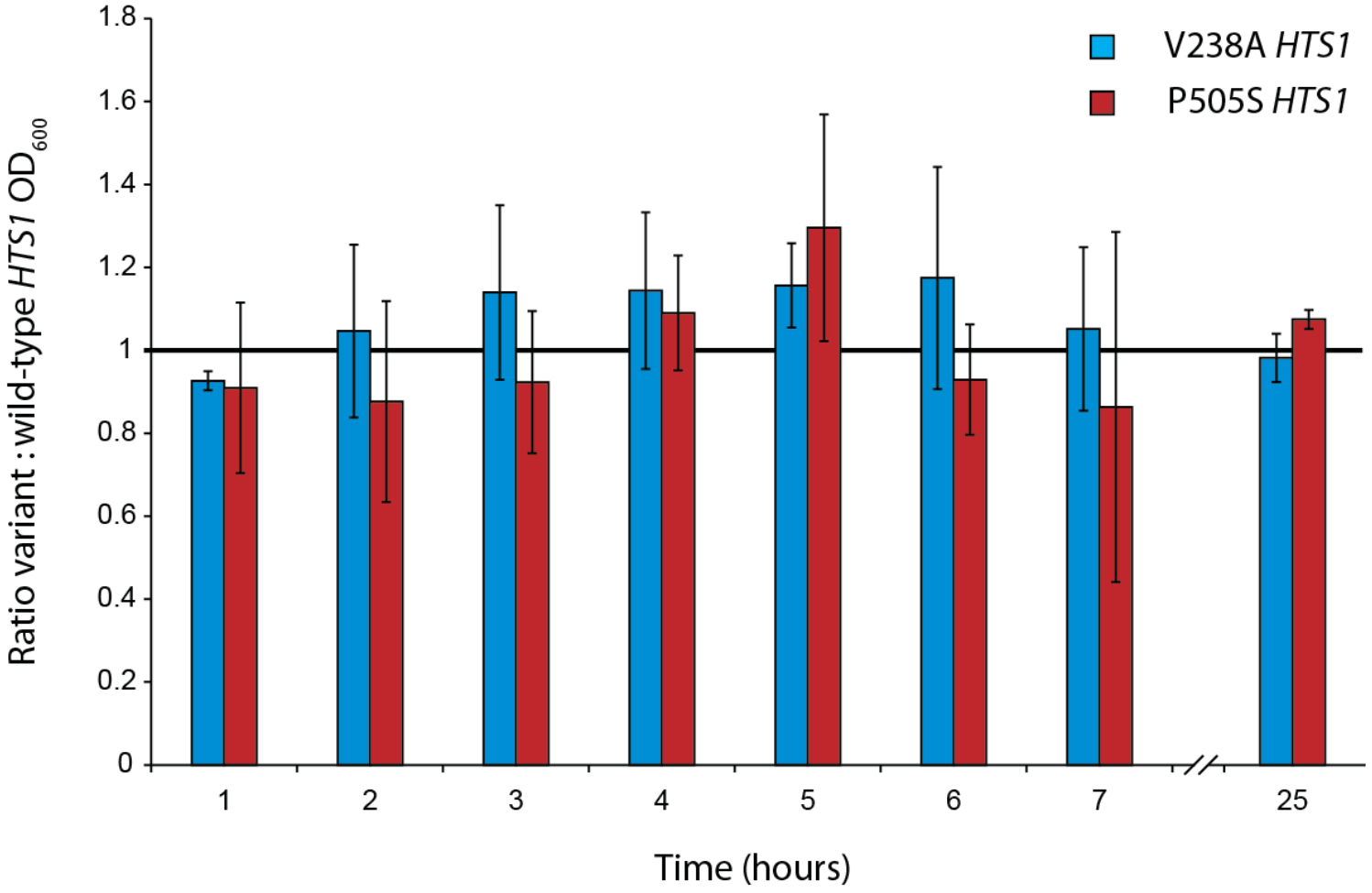


Figure 5: HARS dimerization assays

a.) **HARS dimerization studies via two-hybrid assays do not show expected wild-type HARS dimerization.** The ratio of firefly to *Renilla* luciferase activity was assessed in lysed MN-1 cells. Cells were transfected with the constructs labeled on the x-axis. Independent transfections were performed independently in 8 wells of a 96-well plate, and the average luciferase intensity value was used when graphing. The error bars indicate standard deviation.

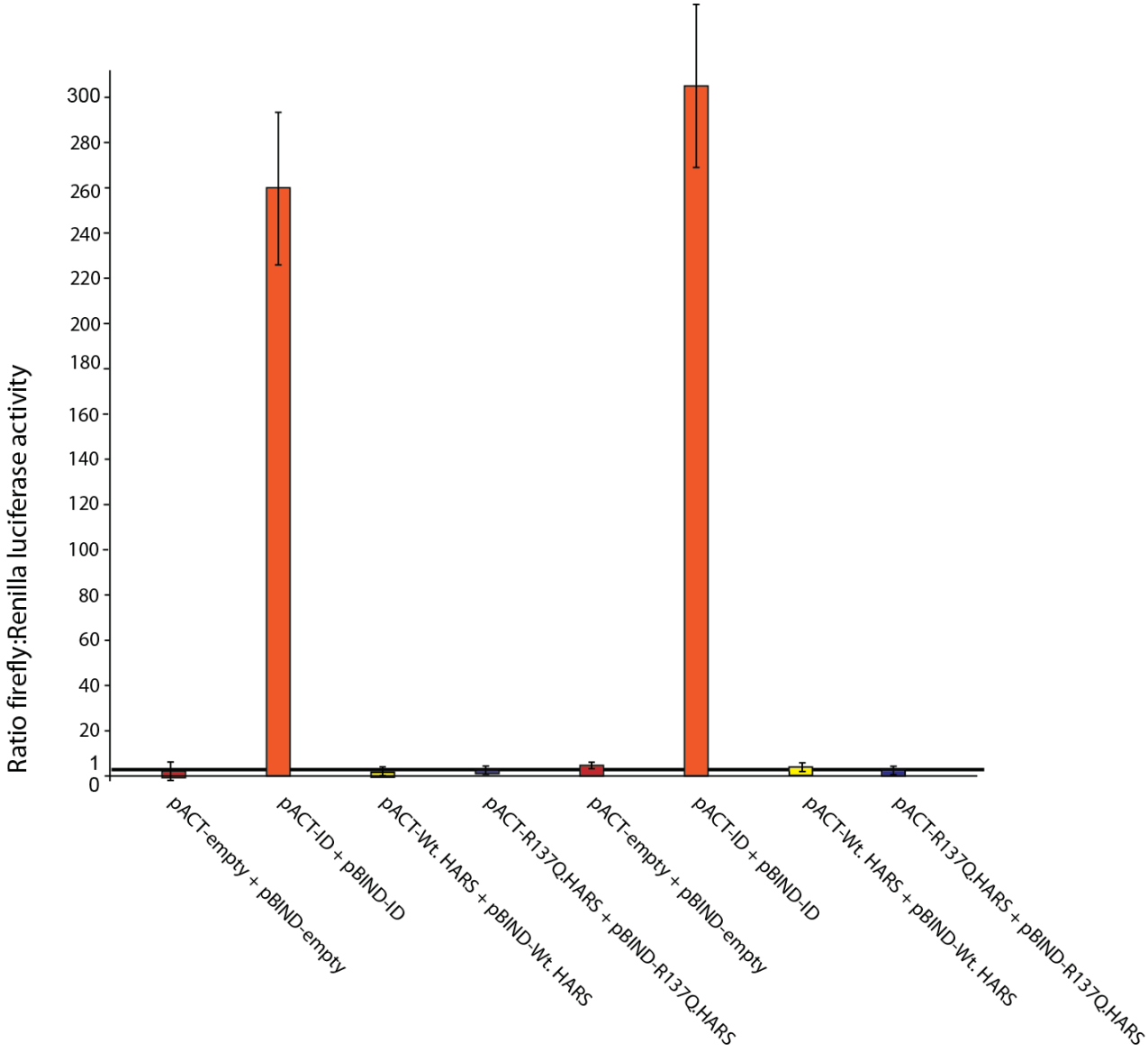


Figure 5: HARS dimerization assays (continued)

b.) **DSS-treated dimerization of wild-type and mutant R137Q HARS occurs *in vitro.*** Mutant and wild-type (Wt) HARS samples were treated with DSS ("treated") or were left untreated as a control ("untreated"). BenchMark Protein Ladder (Invitrogen) was used for visualization. Both wild-type and mutant HARS appear as a higher band when treated with DSS, indicating that dimerization occurred in MN-1 cells.

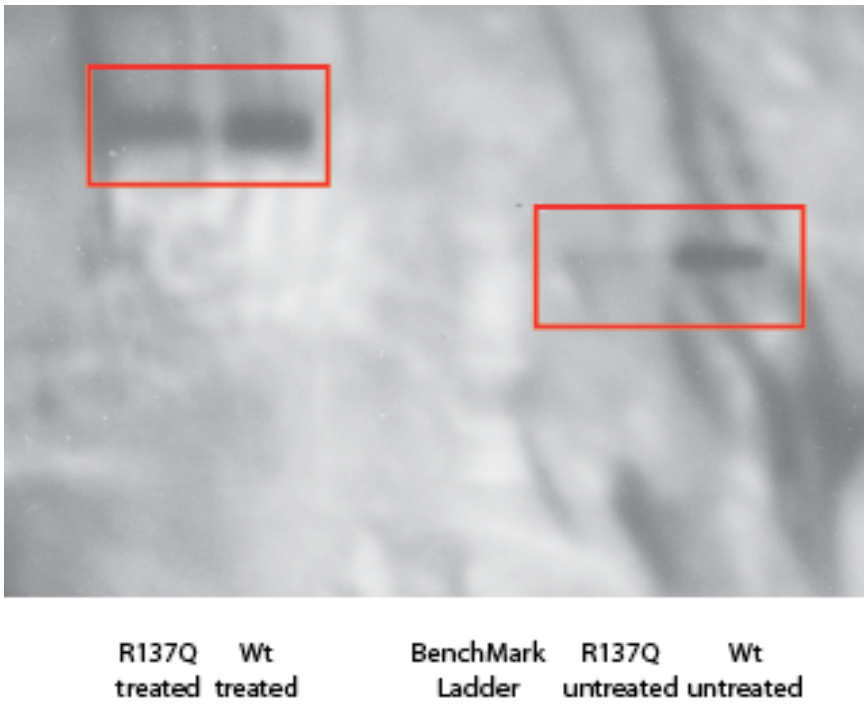
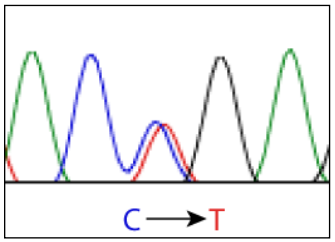


Figure 6: HARS bisulfite sequencing

a.) **Affected CpG island of R137Q *HARS* variant.** The chromatogram trace displays the R137Q *HARS* mutation in the patient with CMT. The mutation is present in the heterozygous state; thus the original cytosine (C, in blue) and mutated thymidine (T, in red) are present. The chromatogram trace was generated with Sequencher software (Gene Codes).



b.) **CpG methylation in *HARS* exon encompassing R137Q variant.** An illustration of the bisulfite sequencing products of eight CpGs in exon 5 of *HARS* is shown. Each circle represents a CpG island, and filled circles indicate CpG islands that are methylated. Three clones were analyzed for the first sample (Sample 1) of genomic DNA, and four clones were analyzed for the second sample (Sample 2) of genomic DNA. Genomic DNA samples were from individuals not affected by CMT. (Figure by Anthony Antonellis)

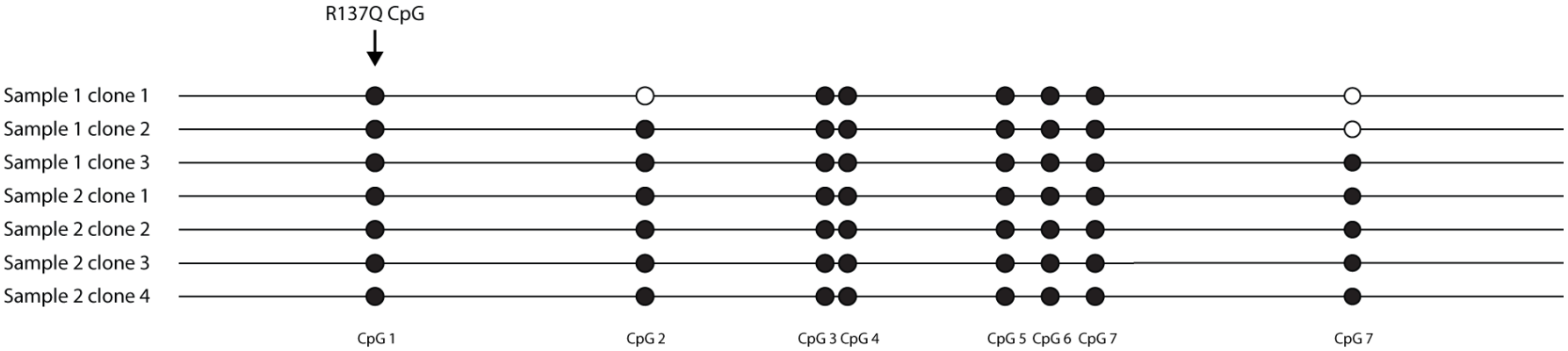


Figure 7: GARS protein localization in Drosophila motor neurons

Wild-type GARS localizes diffusely in *Drosophila* motor neurons. Motor neurons from wandering third instar larvae are stained for axonal membrane (anti-HRP) in red and wild-type *gars* (anti-myc) in green. HRP is a plant glycoprotein used to stain neural tissue (Paschinger et al. 2009).

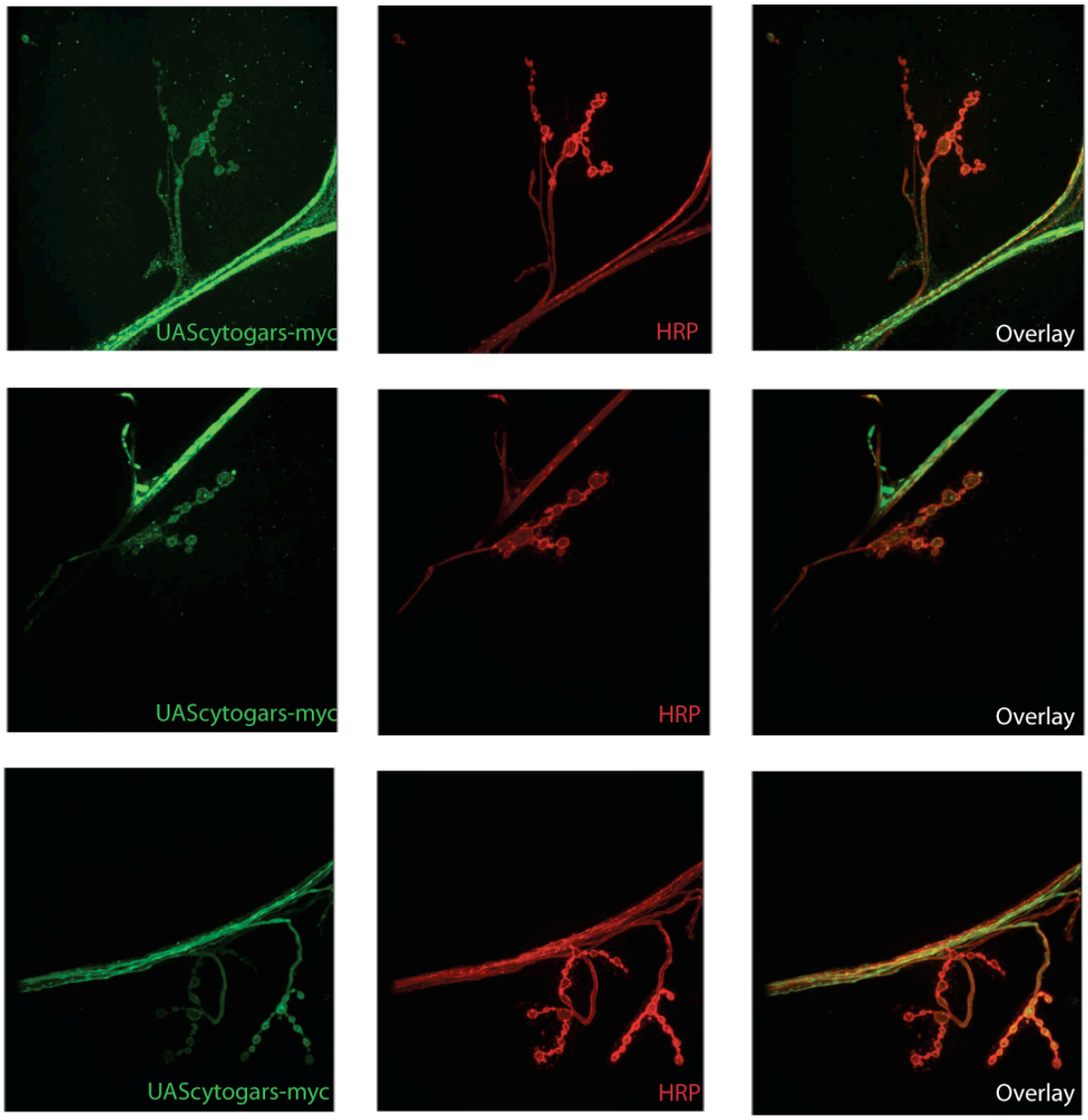


Figure 8: Localization of mitochondria and GARS in sensory neurons of Drosophila

a.) **Localization of mitochondria in *Drosophila* sensory neurons**. Sensory neurons from wandering third instar larvae are stained for axonal membrane (anti-RFP) in red and mitoGFP (anti-GFP) in green. CD8 is expressed at the membrane, and RFP was used for visualization. Likewise, GFP was expressed by mitochondria and was used to visualize them.

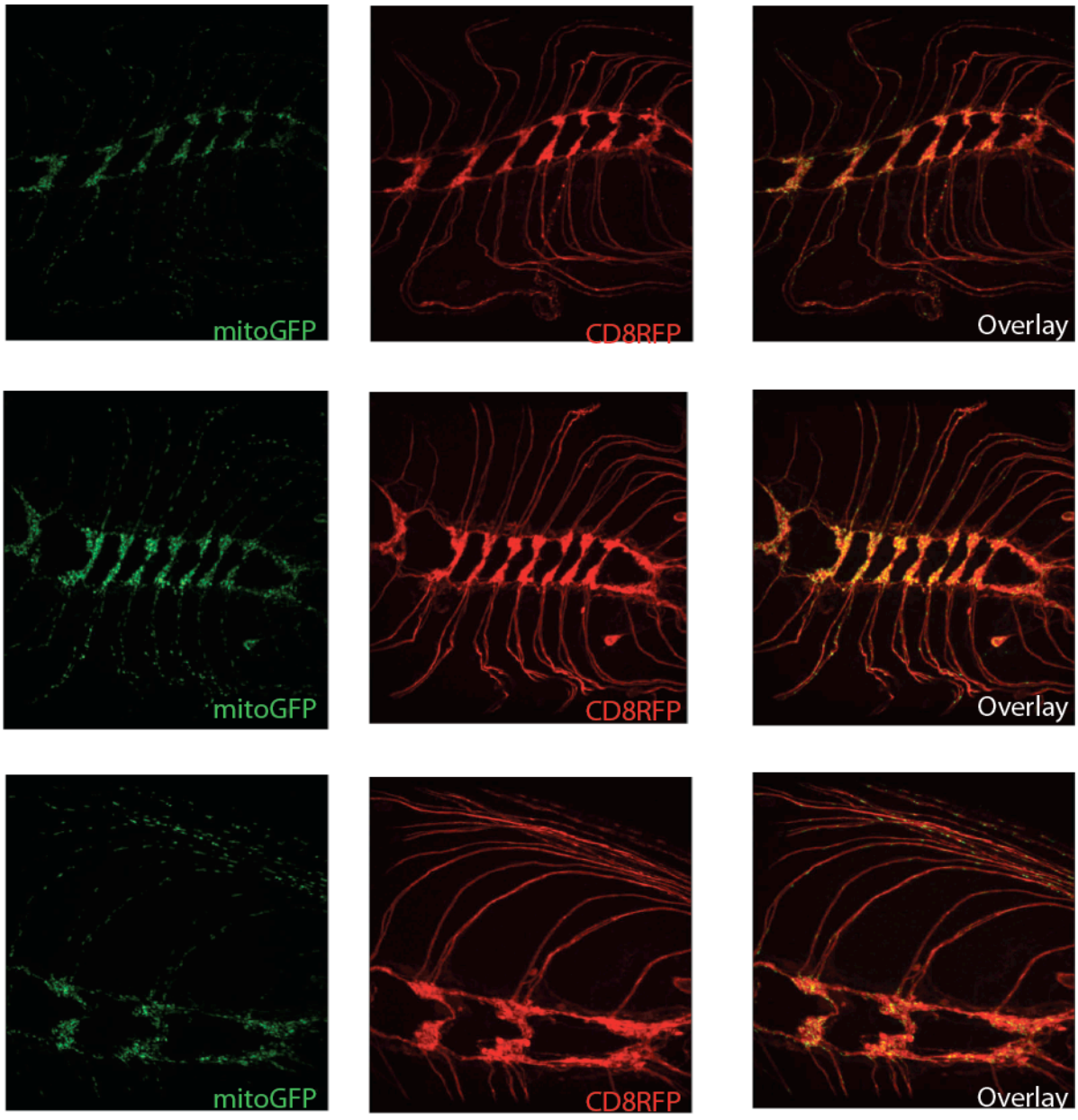


Figure 8: Localization of mitochondria and GARS in Drosophila sensory neurons

b.) **Co-staining of mitochondria and wild-type GARS in *Drosophila* sensory neurons.** Sensory neurons from wandering third instar larvae are stained for mitochondria (anti-GFP) in green, and wild-type *gars* (anti-myc) in purple.

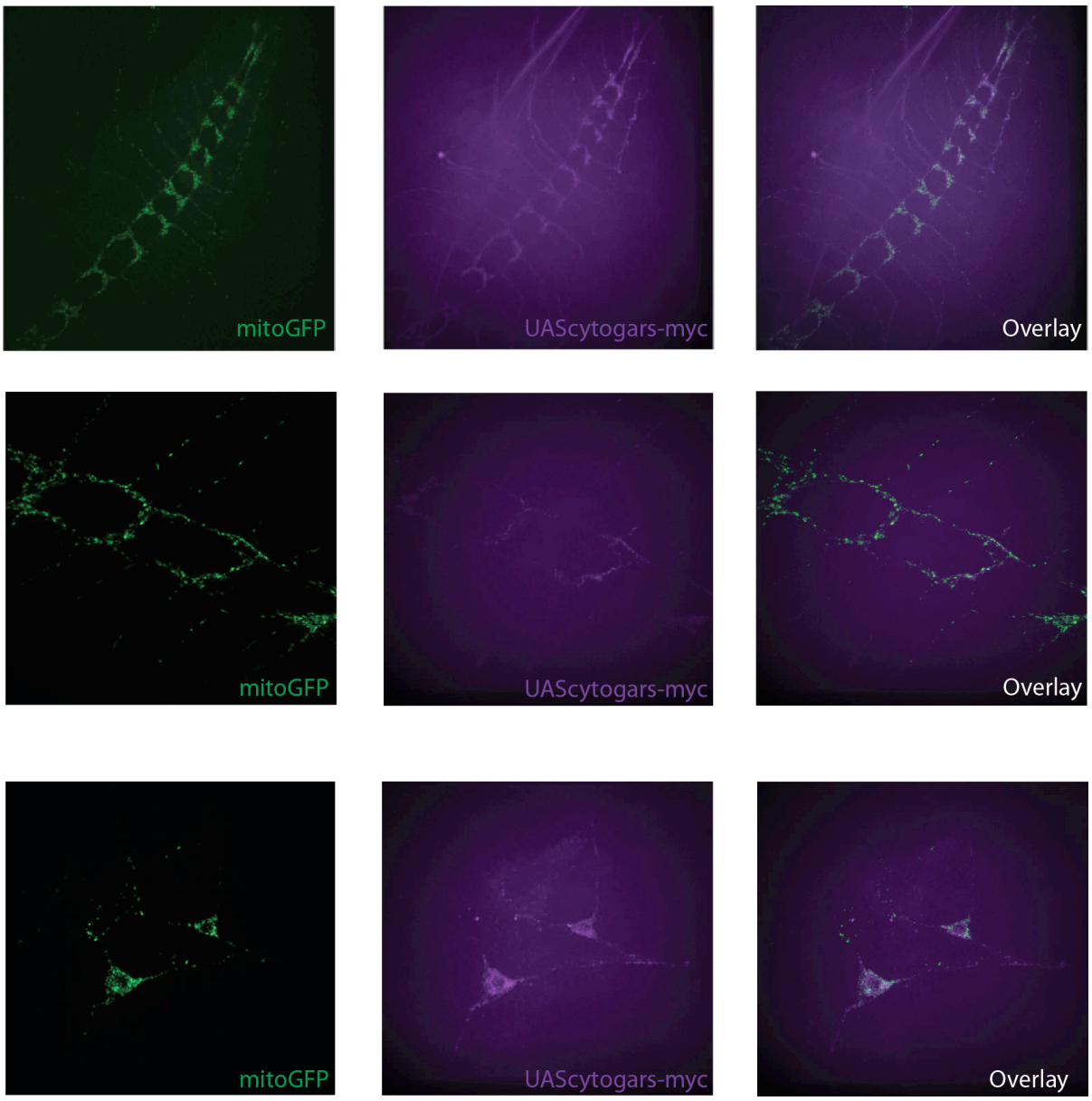


Figure 9: Wild-type and mutant GARS protein localization in Drosophila sensory neurons

a.) **Wild-type GARS localization in *Drosophila* sensory neurons.** Sensory neurons from wandering third instar larvae are stained for axonal membrane (anti-RFP) in red and cyto*gars*-myc (anti-myc) in green. CD8 is expressed at the membrane, and RFP was used for visualization.

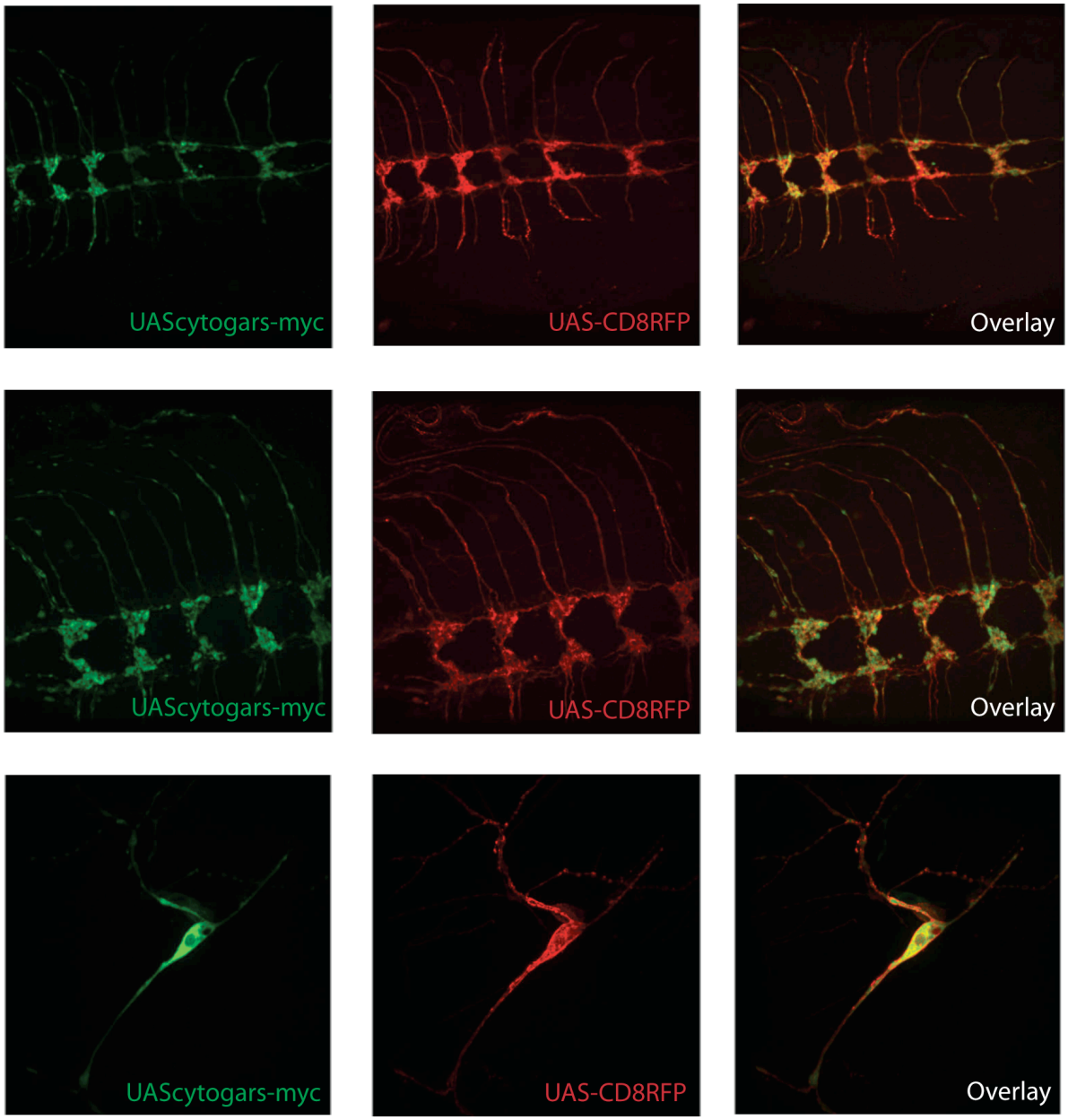
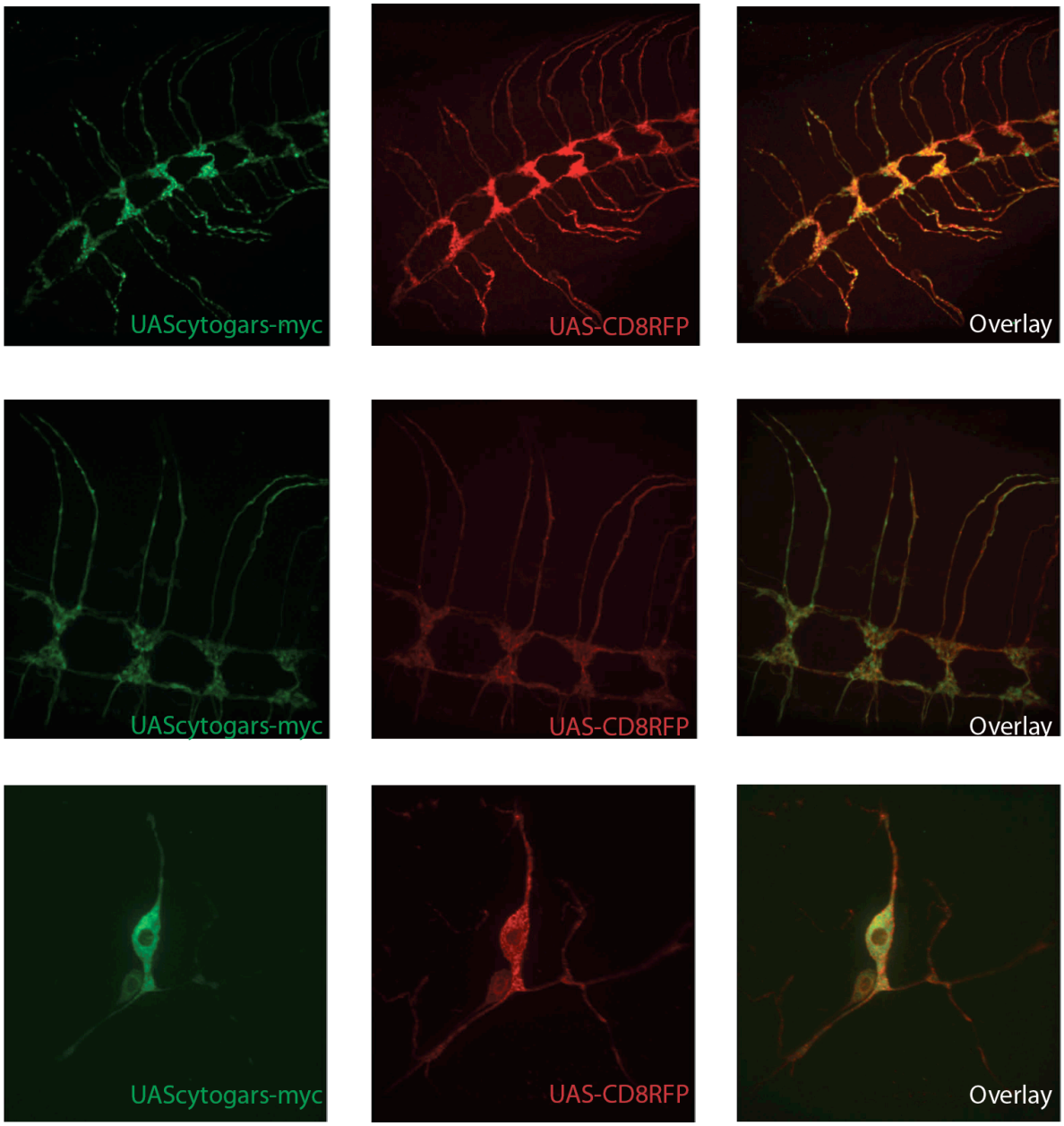


Figure 9: Wild-type and mutant GARS protein localization in Drosophila sensory neurons

b.) **Mutant G240R *GARS* does not significantly change GARS localization in *Drosophila* sensory neurons.** Sensory neurons from wandering third instar larvae are stained for axonal membrane (anti-RFP) in red and cyto*gars*-myc (anti-myc) in green. CD8 is expressed at the membrane, and RFP was used for visualization.



REFERENCES

1. Antonellis, A., Lee-Lin, S. Q., Wasterlain, A., Leo, P., Quezado, M., Goldfarb, L. G., . . . Green, E. D. (2006). Functional analyses of glycyl-tRNA synthetase mutations suggest a key role for tRNA-charging enzymes in peripheral axons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 26*(41), 10397-10406.

2. Antonellis, A., Ellsworth, R. E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S., . . . Green, E. D. (2003). Glycyl tRNA synthetase mutations in charcot-marie-tooth disease type 2D and distal spinal muscular atrophy type V. *The American Journal of Human Genetics, 72*(5), 1293-1299.

3. Antonellis, A., & Green, E. D. (2008). The role of aminoacyl-tRNA synthetases in genetic diseases. *Annual Review of Genomics and Human Genetics, 9*(1), 87-107.

4. Biesecker, L. G., Mullikin, J. C., Facio, F. M., Turner, C., Cherukuri, P. F., Blakesley, R. W., . . . Green, E. D. (2009). The ClinSeq project: Piloting large-scale genome sequencing for research in genomic medicine. *Genome Research, 19*(9), 1665-1674.

5. Budnik, V., Koh, Y., Guan, B., Hartmann, B., Hough, C., Woods, D., & Gorczyca, M. (1996). Regulation of synapse structure and function by the drosophila tumor suppressor gene dlg. *Neuron, 17*(4), 627-640.

6. Casasnovas, C., Cano, L. M., Albertí, A., Céspedes, M., & Rigo, G. (2008). Charcot-marie-tooth disease. *Foot & Ankle Specialist, 1*(6), 350-354.

7. Chihara, T., Luginbuhl, D., & Luo, L. (2007). Cytoplasmic and mitochondrial protein translation in axonal and dendritic terminal arborization. *Nature Neuroscience, 10*(7), 828-837.

8. Dyck, P. J., & Lambert, E. H. (1968). Lower motor and primary sensory neuron diseases with peroneal muscular atrophy. II. neurologic, genetic, and electrophysiologic findings in various neuronal degenerations. *Archives of Neurology, 18*(6), 619-625.

9. Freist, W., Verhey, J. F., Rühlmann, A., Gauss, D. H., & Arnez, J. G. (1999). Histidyl-tRNA synthetase. *Biological Chemistry, 380*(6), 623.

10. Hodonsky, C. J., Kleinbrink, E. L., Charney, K. N., Prasad, M., Bessling, S. L., Jones, E. A., . . . Antonellis, A. (2012). SOX10 regulates expression of the SH3-domain kinase binding protein 1 (Sh3kbp1) locus in schwann cells via an alternative promoter. *Molecular and Cellular Neuroscience, 49*(2), 85-96.

11. Ionasescu, V., Searby, C., Sheffield, V. C., Roklina, T., Nishimura, D., & Ionasescu, R. (1996). Autosomal dominant charcot-marie-tooth axonal neuropathy mapped on chromosome 7p (CMT2D). *Human Molecular Genetics, 5*(9), 1373-1375.

12. Jones, W. D. (2009). The expanding reach of the GAL4/UAS system into the behavioral neurobiology of drosophila. *BMB Reports, 42*(11), 705-712.

13. Jordanova, A., Irobi, J., Thomas, F. P., Van Dijck, P., Meerschaert, K., Dewil, M., . . . Timmerman, V. (2006). Disrupted function and axonal distribution of mutant tyrosyl-tRNA synthetase in dominant intermediate charcot-marie-tooth neuropathy. *Nature Genetics, 38*(2), 197-202.

14. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). Clustal W and clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948.

15. Latour, P., Thauvin-Robinet, C., Baudelet-Méry, C., Soichot, P., Cusin, V., Faivre, L., . . . Rousson, R. (2010). A major determinant for binding and aminoacylation of tRNAAla in cytoplasmic alanyl-tRNA synthetase is mutated in dominant axonal charcot-marie-tooth disease. *The American Journal of Human Genetics, 86*(1), 77-82.

16. Lee, M. K., & Cleveland, D. W. (1994). Neurofilament function and dysfunction: Involvement in axonal growth and neuronal disease. *Current Opinion in Cell Biology, 6*(1), 34-40.

17. Lupski, J. R., de Oca-Luna, R. M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B. J., . . . Patel, P. I. (1991). DNA duplication associated with charcot-marie-tooth disease type 1A. *Cell, 66*(2), 219-232.

18. McLaughlin, H. M., Sakaguchi, R., Giblin, W., NIH Intramural Sequencing Center, Wilson, T. E., Biesecker, L., . . . Antonellis, A. (2012). A recurrent loss-of-function alanyl-tRNA synthetase (AARS?) mutation in patients with charcot-marie-tooth disease type 2N (CMT2N). *Human Mutation, 33*(1), 244-253.

19. McLaughlin, H. M., Sakaguchi, R., Liu, C., Igarashi, T., Pehlivan, D., Chu, K., . . . Antonellis, A. (2010). Compound heterozygosity for loss-of-function lysyl-tRNA synthetase mutations in a patient with peripheral neuropathy. *The American Journal of Human Genetics, 87*(4), 560-566.

20. Motley, W. W., Talbot, K., & Fischbeck, K. H. (2010). GARS axonopathy: Not every neuron's cup of tRNA. *Trends in Neurosciences, 33*(2), 59-66.

21. Nangle, L. A., Zhang, W., Xie, W., Yang, X., & Schimmel, P. (2007). Charcot-marie-tooth disease-associated mutant tRNA synthetases linked to altered dimer interface and neurite distribution defect. *Proceedings of the National Academy of Sciences of the United States of America, 104*(27), pp. 11239-11244.

22. Patzkó, Á., & Shy, M. E. (2011). Update on charcot-marie-tooth disease. *Current Neurology and Neuroscience Reports, 11*(1), 78.

23. Paschinger, K., Rendic, D., & Wilson, I. B. (2009). Revealing the anti-HRP epitope in drosophila and caenorhabditis. Glycoconjugate Journal, 26(3), 385-395.

24. Reilly, M. M., Murphy, S. M., & Laurá, M. (2011). Charcot-marie-tooth disease. *Journal of the Peripheral Nervous System, 16*(1), 1-14.

25. Seburn, K. L., Nangle, L. A., Cox, G. A., Schimmel, P., & Burgess, R. W. (2006). An active dominant mutation of glycyl-tRNA synthetase causes neuropathy in a charcot-marie-tooth 2D mouse model. *Neuron, 51*(6), 715-726.

26. Sivakumar, K., Kyriakides, T., Puls, I., Nicholson, G. A., Funalot, B., Antonellis, A., . . . Goldfarb, L. G. (October 2005). Phenotypic spectrum of disorders associated with glycyl-tRNA synthetase mutations. *Brain, 128*(10), 2304-2314.

27. Soejima, M., Kang, E. H., Gu, X., Katsumata, Y., Clemens, P. R., & Ascherman, D. P. (2011). Role of innate immunity in a murine model of histidyl?transfer RNA synthetase (jo-1)?mediated myositis. *Arthritis & Rheumatism, 63*(2), 479-487.

28. Storkebaum, E., Leitao-Goncalves, R., Godenschwege, T., Nangle, L., Mejia, M., Bosmans, I., . . . Jordanova, A. (2009). Dominant mutations in the tyrosyl-tRNA synthetase gene recapitulate in drosophila features of human charcot-marie-tooth neuropathy. *Proceedings of the National Academy of Sciences, 106*(28), 11782-11787.

29. Stum, M., McLaughlin, H. M., Kleinbrink, E. L., Miers, K. E., Ackerman, S. L., Seburn, K. L., . . . Burgess, R. W. (2011). An assessment of mechanisms underlying peripheral axonal degeneration caused by aminoacyl-tRNA synthetase mutations. *Molecular and Cellular Neuroscience, 46*(2), 432-443.

30. Verhamme, C., King, R. H. M., ten Asbroek, A. L. M. A., Muddle, J. R., Nourallah, M., Wolterman, R., . . . van Schaik, I. N. (2011). Myelin and axon pathology in a long-term study of PMP22-overexpressing mice.. *Journal of Neuropathology & Experimental Neurology, 70*(5), 386-13.

31. Vester, A., Velez-Ruiz, G., McLaughlin, H.M., NIH Intramural Sequencing Center, Lupski, J.R., Talbot, K.,. . . Antonellis, A. (2012). A loss-of-function variant in the human histidyl-tRNA synthetase (*HARS*) gene is neurotoxic in vivo. *In preparation*.

32. Xiong, X., Wang, X., Ewanek, R., Bhat, P., Aaron DiAntonio, & Collins, C. A. (2010). Protein turnover of the Wallenda/DLK kinase regulates a retrograde response to axonal injury. *The Journal of Cell Biology, 191*(1), pp. 211-223.

33. Zhai, J., Lin, H., Julien, J., & Schlaepfer, W. W. (2007). Disruption of neurofilament network with aggregation of light neurofilament protein: A common pathway leading to motor neuron degeneration due to Charcot–Marie–Tooth disease-linked mutations in NFL and HSPB1. *Human Molecular Genetics, 16*(24), 3103-3116.