Identification of a novel recessive ataxia gene

by

Randi M. Burns

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2014

Doctoral Committee:
Professor Margit Burmeister, Chair
Associate Professor William T. Dauer
Assistant Professor James J. Dowling, University of Toronto
Associate Professor Donna M. Martin
Associate Professor Miriam H. Meisler
Assistant Professor Vikram Shakkottai

To

Mario and Taylor

Randal and Kathlyn

Acknowledgments

I would like to acknowledge my mentor, Dr. Margit Burmeister, who provided endless support during my graduate training and always encouraged critical thinking and resourcefulness. I am grateful to the past and current members of the Burmeister lab for their wonderful advice, challenging questions, continuous encouragement and wonderful friendships, especially Dr. Viktoriya Strumba, Dr. Cynthia Schoen, Dr. Kristine Sikora, Dr. Erin Sandford, Dr. Sandra Villafuerte, Elzbieta Sliwerska, and Kendal Walker. I would also like to acknowledge the lab undergraduates, especially Mikhail Ognenovski, for his wonderful experiment preparation skills.

I have learned so much from interacting with my committee members Dr. Miriam Meisler, Dr. Vikram Shakkottai, Dr. Donna Martin, Dr. William Dauer and Dr. James Dowling for their invaluable advice during the development and execution of my research. I would also like to acknowledge the grants that supported this work from the National Institutes of Health, National Institute on Neurological Disorders grants (NS078560 and NS078560S1) to Margit Burmeister and National Institutes of Health Cellular and Molecular Biology Training Grant (T32-GM007315) to Randi Burns. I am also grateful to those who provided additional mentorship and support, especially Dr. Jessica Schwartz and Dr. Robert Fuller. I would like to acknowledge my support from PIBS, especially Tiffany Porties, the Rackham Graduate School, the Program in Cellular and Molecular Biology, and the Molecular and Behavioral Neuroscience Institute.

I thank my precious family members, especially my parents, Kathlyn and Randal, for encouraging my curiosity and instilling determination and perseverance in me. They are the best parents a young female scientist could ever have. I would like to thank my adopted sisters, Adrienne, Brandi, Candace, Sarah, Shanika, Shannon, and Shilpa for their love, faith, and for always being there. I also thank my fairy grandmother, Karen James, for welcoming me into her family, caring for my daughter when I needed to focus on my research, and for moving with my family to Texas in the near future. I acknowledge my in-laws, Angela James, Charles James Jr. and the twins, Deron and Devante McRoyal, for providing laughs throughout this process. Most of all, I thank my husband, Mario, and my daughter, Taylor, for inspiring me to work harder everyday, even when it wasn't in their best interest. Their infinite love and support made this work possible.

Table of Contents

Chapter V. Conclusion and Future Directions

List of Figures

List of Tables

Notes

Video recordings of phenotypes of morphant fish are available upon request to the author or the Burmeister laboratory.

Abstract:

While the genes involved in most forms of sporadic or recessive ataxia with mental retardation are still unknown, exome sequencing is a promising tool to identify novel genes in rare disorders. Previously, two siblings in a consanguineous Turkish family were reported, who presented with a non-progressive ataxia syndrome including congenital truncal and extremity ataxia, cerebellar hypoplasia, hypotonia, developmental delay, mental retardation and nystagmus. After exome sequencing and filtering by homozygosity, we identified a homozygous mutation at the invariant +1 position (c. 964+1 G>A) in intron 9 of the *CWF19L1* (complexed with cdc5 protein 19-like 1) gene. This mutation is absent in >6,500 European and African American individuals and 200 Turkish control DNAs. In lymphoblastoid cell lines from affected individuals, the mutation causes exon skipping, reduction in mRNA levels, and protein loss. Morpholino-mediated knockdown in a zebrafish model demonstrates that loss of the evolutionarily highly conserved *CWF19L1*, whose normal biological function is unknown, alters cerebellar morphology and causes movement abnormalities. Preliminary data suggests this protein is expressed in a tissue-specific manner and that this protein is localized in the nucleus. Our results suggest that we have identified a novel cause of recessive ataxia and developmental delay.

Chapter I.

Introduction

Ataxia is a disorder of the cerebellum characterized by lack of coordinated, voluntary movements. Acquired ataxias are caused by insult to the cerebellum, either by acute alcohol or toxic incidence, blunt force trauma to the head, or as a secondary symptom in other disorders including paraneoplastic diseases and multiple system atrophy [1,2]. While ataxia can be acquired, many ataxia disorders have an underlying genetic cause.

Hereditary ataxias are a clinically and genetically heterogeneous group of disorders. To date, greater than 50 genes have been implicated in ataxia, with these genes involved in many molecular mechanisms in the cell[3,4]. Many inherited ataxias are caused by gain- or loss-of-function mutations in genes, however the genetic etiology is unexplained in ~40% of cases[5,6]. Hereditary ataxias have several modes of inheritance including dominant, recessive, x-linked, and mitochondrial but the majority of the ataxia disorders are of dominant or recessive inheritance[2,5,7].

Dominant ataxias

Autosomal dominant cerebellar ataxias (ADCAs) consist of the spinocerebellar ataxias (SCA), which are generally caused by repeat expansions, and episodic ataxias (EAs) generally caused by defects in ion channels. ADCAs often have a later age of onset, over 40 years of age, and are often progressive, with worsening ataxia symptoms and cerebellar atrophy, significantly impacting the quality of life of the affected individuals[1,2,5,8]. Anticipation is a common phenomenon in the repeat expansion disorders where severity of the disease increases, due to expansion of the repeat, in succeeding generations[1,2,8–10]. Along with ataxic gait and cerebellar deficits, ophthalmological abnormalities and non-neurological signs and symptoms can accompany dominant ataxias. With the advent of advanced sequencing technologies, recent studies provide evidence for conventional (frameshift, missense, and nonsense) mutations in dominant ataxia[11]. Although the genetic etiology is defined for many of the dominant ataxias, it is estimated that many patients have rare mutations that are currently undefined[5].

Recessive ataxias

Autosomal recessive cerebellar ataxias (ARCAs) are recessively inherited neurological disorders that are characterized by incoordination, delayed motor development milestones, and early age of onset, before 25 years of age[1,8–10]. While most forms of ataxia are individually rare, recessive ataxias are cumulatively not uncommon, with an estimated frequency of 1/20,000 that varies

between countries[5,7]. ARCAs are genetically heterogeneous and are clinically classified in four different groups based on genetic etiology and common phenotypes: congenital ataxias, usually defined by abnormal development of the cerebellum or brainstem; metabolic ataxias, marked by metabolic disorder and where ataxia may be a minor symptom; DNA repair defect ataxias, marked by mutations in genes that are involved in DNA repair; and degenerative and progressive ataxias, due to mutations in mitochondrial genes and severe progression of the disorder[9,12].

Currently 21 ataxia genes are routinely tested in the clinic in cases of suspected recessive inheritance. In many of these cases, however, results are negative, suggesting that several recessive ataxia genes are still unknown[5]. Identifying additional recessive ataxia genes may lead to improved clinical testing, which would help in diagnosis. Establishment of a genetic cause can help provide prognostic information to the family. Additionally, gene identification may reveal a molecular mechanism or pathway that is currently unknown to play a role in ataxia or cerebellar function. Alternatively, the gene may play a role in an already known ataxia pathway, in which case knowledge of that ataxia pathway or network will be expanded[13,14]. As ataxias are genetically heterogeneous and some rare ataxias are caused by private mutations existing only in that particular family, development of ataxia pathways will be essential to provide larger scale targets for therapies.

Rare recessive ataxia in a consanguineous Turkish family

Two Turkish families were identified by Yapici and Eraksoy (families 1 and 3), family 3 is consanguineous[15]. Two pairs of siblings were affected with a non-progressive, congenital ataxia in an autosomal recessive inheritance pattern. Symptoms include hypotonia, developmental delay, mental retardation, nystagmus, truncal and extremity ataxia and cerebellar hypoplasia. Blood was collected from the affected children in the two families and mailed to the Burmeister lab for DNA extraction. Because the same physician referred both families from Turkey, we hypothesized that a founder mutation of common origin was the cause of the ataxia. Linkage analysis was performed on the 4 children of 2 families, and several regions of interest were identified. Several candidate genes were identified by linkage analysis, including *KCNK1* and *GLRX2*, however no variants were found. High density genotyping demonstrated no overlapping linkage and homozygosity regions, suggesting a founder mutation is unlikely. Therefore, we focused on identifying the mutation in the consanguineous family as consanguinity adds power to linkage analysis in small families[16]. The parents and a third sibling were clinically healthy. Negative testing for ataxia genes suggested a mutation in a novel gene or mutation in a gene that is not currently included in a clinical panel.

Gene identification methods

There are multiple methods to identify genetic etiologies for disorders[17–19]. Two traditional methods include functional cloning and positional cloning. In functional

cloning, the position of the gene is not necessary; genes are chosen based on the biological defect in the disorder[17] . This method has been used successfully to clone the genes for phenylketonuria (PAH) and sickle cell anemia (HBB)[17]. This method is most useful for disorders that have a biochemical defect and are homogeneous in their causes. Positional cloning is another method that is useful for identifying genetic candidates. Positional cloning does not assume any particular function of the candidate genes; it is based only on map location. A major step in positional cloning is linkage analysis. Linkage analysis uses a panel of genome-wide polymorphic markers to identify regions where the disease locus is located. It is based on the premise that markers nearest the mutation will segregate through generations with the mutation, as the probability of recombination between those regions is reduced, demonstrating proximity, or linkage, to the mutation. Linkage analysis maps regions where candidate genes may be located. These regions are then searched for potential candidate genes based on their function or involvement in similar disorders. These candidates can then be sequenced by the Sanger method to identify variants[17]. Positional cloning has proven to be effective for cloning the genes in multiple genetic disorders including Duchenne muscular dystrophy, Bardet-Biedl syndrome, and SCA1[20–22].

Homozygosity mapping is a special form of linkage analysis generally applied to families of consanguineous marriage, although it can be used for outbred individuals as well[18,16,23] . This method is used for identifying loci that remain

homozygous over several generations. Homozygosity mapping tests the assumption that the causative mutation is “identical by descent”, or that the affected individual (or individuals) received one copy of the mutated allele from each parent. Mutations of interest are likely to exist in large spans of homozygous loci due to decreased opportunity for recombination between these loci and the disease causing mutation. These homozygous regions can then be mapped and used to identify candidate genes in order to screen for mutations[16,24].

Next generation sequencing has recently become a more popular method for identifying novel variants. Next generation sequencing allows for rapid sequencing of either the whole genome or the coding genome (exome) of a patient. Exome sequencing has proven to be an effective method to identify novel variants because 85% of all known disease causing mutations are found within coding or nearby regulatory regions (5’UTR, 3’UTR, splice sites)[25]. Exome sequencing is also cost effective at less than $1000 per sample [5,25].

These methods have proven to be useful in identifying disease-causing variants; however, they have their limitations. The functional cloning approach is implausible in heterogeneous disorders, as our knowledge is too limited to identify genes based on function. This method would prohibit gene discovery for genes that encode proteins in which the function is unknown or has multiple functions. Rare genetic disorders are often studied in single, sporadic, individuals or individual families providing another barrier to gene identification. These

families are not large enough to perform informative linkage analysis or the mutations are private. In these cases, positional cloning and homozygosity mapping are expensive and laborious methods as they can provide a cumbersome number of candidates and difficulty increases with heterogeneity of the disorder. Exome sequencing also identifies a large number of potential causal variants; therefore, in rare or heterogeneous disorders it is often necessary to combine exome sequencing with linkage or homozygosity regions to reduce the candidate list and identify the variant[26–28]. Furthermore, exome sequencing may exclude variants by filtration methods such as non-coding variants, lack of functional association, or extent of allele frequency in non-ataxic populations.

The above methods are useful for identifying gene variants; however, damaging variants exist in personal genomes[29,30], therefore independent lines of evidence are necessary to show that a variant is causal. Ideally, rare, novel variants are found in additional individuals with the same disorder, demonstrating that the mutation causes the disorder. However, it is often the case that rare mutations are unique to a family, requiring verification by independent methods.

Animal models of ataxia

Development of animal models can demonstrate the impact of a mutation at the organismal level and can be used to discover the molecular mechanisms by which a gene causes a disorder. Animal models are advantageous research

tools in neuro-genetic disorders because they provide a source of tissue to study the effects of the mutation at the cellular and molecular levels, as human brain samples are usually unavailable to researchers. These animal models can then be used to test potential therapies[31–35].

Animal models have been particularly useful in identifying or validating genes involved in ataxia disorders[36]. Naturally occurring mouse models, such as *Nagoya* which has a mutation in *Cacna1*, have provided essential information about the function and physiology of their affected genes and mutations in many of these genes have been implicated in human ataxias as well[36,37]. Knockout and conditional mouse and Drosophila models are also generated to study the impact of ataxia mutations. Multiple models for the same or similar disorders prove advantageous when differences in genetic mutations, phenotypes, and physiology cannot be effectively studied in a particular model due to embryonic lethality or differential development[36]. For example, in SCA7, loss of the Drosophila ataxin-7 protein caused neural and retinal degeneration and early lethaiity[38]. Additionally, severe knockdown of zebrafish ataxin-7 led to deformed embryos and early lethality, however, moderate knockdown of ataxin-7 demonstrated the requirements of ataxin-7 in neuronal and retinal development allowing[39].

Another recent ataxia animal model is the zebrafish[40]. As zebrafish share basic vertebrate brain structure, they have been used to study human neuro-genetic

disorders, including ataxia, Parkinson's disease and bipolar disorder[40–43]. Zebrafish are advantageous research animals because of their rapid generation time, large numbers of offspring, quantifiable behavior, *ex-utero* development, and transparency of developing embryos[40,42,43]. Zebrafish have been utilized to validate mutations[44,45], demonstrated conserved function[46,47], and investigate molecular mechanisms of genes implicated in ataxia[48,49,39]. Expression of mutant dominant negative *RNF170* in zebrafish has been used to show disrupted embryonic development[44]. Knockdown of *rnf216* and *otud4,* discovered to cause ataxia with hypogonadism in a digenic manner, demonstrate abnormal development of the optic tectum and cerebellum of the fish, validating pathogenicity of the mutations[45]. These studies, suggest zebrafish are a feasible model to use to analyze the behavioral and developmental impact of ataxia mutations.

The zebrafish cerebellum has organizational similarities to the human cerebellum as it has a molecular layer, Purkinje cell layer, and granule cell layer[50,51]. The cells contained in these regions are similar to humans as the Purkinje cell layer contains inhibitory neurons that use gamma-aminobutyric acid (GABA) and/or glycine as their major neurotransmitter and the granule cell layer consists of excitatory neurons that use glutamate as a primary neurotransmitter[50]. The zebrafish cerebellum is, however, different in some ways from the human cerebellum. The zebrafish cerebellum is composed of three parts, two of which have the three-layer structure and one that only contains granule cells.

Additionally, the zebrafish cerebellum contains eurydendroid cells, which are thought to be similar to the deep cerebellar nuclei (DCN) that are the major output tract from the human cerebellum, but the molecular machinery of the eurydendroid cells is not well described[50]. Additionally, the zebrafish brain has cerebellum-like structures outside of the cerebellum that contain granule cells, but the function of these structures is unknown[50].

While the zebrafish brain has some differences from the human brain, models of genes implicated in ataxias have demonstrated changes in brain, particularly in the cerebellum, suggesting some conservation of the function of these genes and the importance of these genes in the cerebellum. Knockdown of Ataxin-7, the mutated gene in SCA7, showed defects in development of photoreceptors and Purkinje and granule cells of the cerebellum[39]. Knockdown of the zebrafish *CA8* gene demonstrated a decrease in cerebellar size, increase in neuronal apoptosis and increase in abnormal motor movements similar to the phenotype seen in patients[47].

Additionally, behavioral analysis has been used to demonstrate the impact of ataxia gene mutations in zebrafish models[47,48]. Early movement in fish is largely attributed to spinal cord neurons. However, lesions of the hindbrain, where the cerebellum is located, has been shown to affect swimming ability in larval zebrafish suggesting the hindbrain is important in larval movement[52]. Pathogenesis of the potassium (Kv3.3) channel mutation in SCA13 was

demonstrated by expression of mutant Kv3.3 in zebrafish. This model revealed changes in the excitability of motor neurons and changes in motor response to startle[48]. A zebrafish model of the epilepsy, ataxia, sensorineural deafness, and tubulopathy (EAST) syndrome shows that knockdown of *kcnj10* in fish demonstrates abnormal motor behavior and kidney dysfunction[41]. This model of EAST in the zebrafish is also advantageous to study the brain regions affected by *kcnj10* knockdown as knockout mice demonstrate abnormal brain morphology and are embryonic lethal[53]. Taken together, these studies indicate the zebrafish is a feasible model to analyze the impact of ataxia mutations.

Summary

Although greater than 50 genes have been implicated in ataxia, it is estimated that 40% still have an unexplained genetic etiology[5]. Many of the unexplained ataxias occur sporadically or in families that are not amenable to use of traditional methods, indicating the necessity for new methods to identify genetic causes. Recently, exome sequencing has been used to identify genes involved in rare neurological disorders, including ataxia[5,54–57]. Positive results are often obtained in families with consanguinity[5], as homozygosity further narrows down the linkage evidence[16], and homozygous mutations are easier to detect than two compound heterozygotes[58]. Discovery of these additional ataxia genes will help to increase our understanding of the pathophysiology of this disorder, establish methods to test gene function and may lead to targets for therapeutic treatment.

The goal of this dissertation was to identify the gene involved in a rare ataxia disorder using a combination of homozygosity mapping and exome sequencing in a small consanguineous family. As the mutation was in a gene not previously demonstrated to play a role in ataxia, we utilized molecular methods to demonstrate the impact of the mutation on expression of the gene, characterized normal expression of the protein in cells and tissues, and established a zebrafish animal model to demonstrate functional importance of the gene in an organism.

References

1. Manto, M., and Marmolino, D. (2009). Cerebellar ataxias. Curr. Opin. Neurol. *22*, 419–429.

2. Jayadev, S., and Bird, T.D. (2013). Hereditary ataxias: overview. Genet. Med. Off. J. Am. Coll. Med. Genet. *15*, 673–683.

3. Manto, M., and Marmolino, D. (2009). Cerebellar disorders--at the crossroad of molecular pathways and diagnosis. Cerebellum Lond. Engl. *8*, 417–422.

4. Sailer, A., and Houlden, H. (2012). Recent advances in the genetics of cerebellar ataxias. Curr. Neurol. Neurosci. Rep. *12*, 227–236.

5. Sailer, A., and Houlden, H. (2012). Recent advances in the genetics of cerebellar ataxias. Curr. Neurol. Neurosci. Rep. *12*, 227–236.

6. Németh, A.H., Kwasniewska, A.C., Lise, S., Parolin Schnekenberg, R., Becker, E.B.E., Bera, K.D., Shanks, M.E., Gregory, L., Buck, D., Zameel Cader, M., et al. (2013). Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. Brain J. Neurol. *136*, 3106–3118.

7. Finsterer, J. (2009). Ataxias with autosomal, X-chromosomal or maternal inheritance. Can. J. Neurol. Sci. J. Can. Sci. Neurol. *36*, 409–428.

8. Finsterer, J. (2009). Ataxias with autosomal, X-chromosomal or maternal inheritance. Can. J. Neurol. Sci. J. Can. Sci. Neurol. *36*, 409–428.

9. Palau, F., and Espinós, C. (2006). Autosomal recessive cerebellar ataxias. Orphanet J. Rare Dis. *1*, 47.

10. Embiruçu, E.K., Martyn, M.L., Schlesinger, D., and Kok, F. (2009). Autosomal recessive ataxias: 20 types, and counting. Arq. Neuropsiquiatr. *67*, 1143–1156.

11. Hersheson, J., Haworth, A., and Houlden, H. (2012). The inherited ataxias: genetic heterogeneity, mutation databases, and future directions in research and clinical diagnostics. Hum. Mutat. *33*, 1324–1332.

12. Fogel, B.L., and Perlman, S. (2007). Clinical features and molecular genetics of autosomal recessive cerebellar ataxias. Lancet Neurol. *6*, 245–257.

13. Watase, K., Gatchel, J.R., Sun, Y., Emamian, E., Atkinson, R., Richman, R., Mizusawa, H., Orr, H.T., Shaw, C., and Zoghbi, H.Y. (2007). Lithium therapy improves neurological function and hippocampal dendritic arborization in a spinocerebellar ataxia type 1 mouse model. PLoS Med. *4*, e182.

14. Rai, M., Soragni, E., Jenssen, K., Burnett, R., Herman, D., Coppola, G., Geschwind, D.H., Gottesfeld, J.M., and Pandolfo, M. (2008). HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model. PloS One *3*, e1958.

15. Yapici, Z., and Eraksoy, M. (2005). Non-progressive congenital ataxia with cerebellar hypoplasia in three families. Acta Paediatr. Oslo Nor. 1992 *94*, 248–253.

16. Lander, E.S., and Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science *236*, 1567–1570.

17. Collins, F.S. (1995). Positional cloning moves from perditional to traditional. Nat. Genet. *9*, 347–350.

18. Hildebrandt, F., Heeringa, S.F., Rüschendorf, F., Attanasio, M., Nürnberg, G., Becker, C., Seelow, D., Huebner, N., Chernin, G., Vlangos, C.N., et al. (2009). A systematic approach to mapping recessive disease genes in individuals from outbred populations. PLoS Genet. *5*, e1000353.

19. Wang, Z., Liu, X., Yang, B.-Z., and Gelernter, J. (2013). The role and challenges of exome sequencing in studies of human diseases. Front. Genet. *4*, 160.

20. Koenig, M., Hoffman, E.P., Bertelson, C.J., Monaco, A.P., Feener, C., and Kunkel, L.M. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell *50*, 509–517.

21. Kwiatkowski, T.J., Jr, Orr, H.T., Banfi, S., McCall, A.E., Jodice, C., Persichetti, F., Novelletto, A., LeBorgne-DeMarquoy, F., Duvick, L.A., and Frontali, M. (1993). The gene for autosomal dominant spinocerebellar ataxia (SCA1) maps centromeric to D6S89 and shows no recombination, in nine large kindreds, with a dinucleotide repeat at the AM10 locus. Am. J. Hum. Genet. *53*, 391–400.

22. Nishimura, D.Y., Searby, C.C., Carmi, R., Elbedour, K., Van Maldergem, L., Fulton, A.B., Lam, B.L., Powell, B.R., Swiderski, R.E., Bugge, K.E., et al. (2001). Positional cloning of a novel gene on chromosome 16q causing Bardet-Biedl syndrome (BBS2). Hum. Mol. Genet. *10*, 865–874.

23. Alkuraya, F.S. (2013). The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. Hum. Genet. *132*, 1197–1211.

24. Alkuraya, F.S. (2012). Discovery of rare homozygous mutations from studies of consanguineous pedigrees. Curr. Protoc. Hum. Genet. Editor. Board Jonathan Haines Al *Chapter 6*, Unit6.12.

25. Cooper, D.N., Chen, J.-M., Ball, E.V., Howells, K., Mort, M., Phillips, A.D., Chuzhanova, N., Krawczak, M., Kehrer-Sawatzki, H., and Stenson, P.D. (2010). Genes,

mutations, and human inherited disease at the dawn of the age of personalized genomics. Hum. Mutat. *31*, 631–655.

26. Ashraf, S., Gee, H.Y., Woerner, S., Xie, L.X., Vega-Warner, V., Lovric, S., Fang, H., Song, X., Cattran, D.C., Avila-Casado, C., et al. (2013). ADCK4 mutations promote steroid-resistant nephrotic syndrome through CoQ10 biosynthesis disruption. J. Clin. Invest. *123*, 5179–5189.

27. Gee, H.Y., Otto, E.A., Hurd, T.W., Ashraf, S., Chaki, M., Cluckey, A., Vega-Warner, V., Saisawat, P., Diaz, K.A., Fang, H., et al. (2013). Whole-exome resequencing distinguishes cystic kidney diseases from phenocopies in renal ciliopathies. Kidney Int.

28. He, M., Tang, B.-S., Li, N., Mao, X., Li, J., Zhang, J.-G., Xiao, J.-J., Wang, J., Jiang, H., Shen, L., et al. (2014). Using a combination of whole-exome sequencing and homozygosity mapping to identify a novel mutation of SCARB2. Clin. Genet.

29. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science *335*, 823–828.

30. Singleton, A.B. (2011). Exome sequencing: a transformative technology. Lancet Neurol. *10*, 942–946.

31. Rai, M., Soragni, E., Chou, C.J., Barnes, G., Jones, S., Rusche, J.R., Gottesfeld, J.M., and Pandolfo, M. (2010). Two new pimelic diphenylamide HDAC inhibitors induce sustained frataxin upregulation in cells from Friedreich's ataxia patients and in a mouse model. PloS One *5*, e8825.

32. García-Corzo, L., Luna-Sánchez, M., Doerrier, C., Ortiz, F., Escames, G., Acuña-Castroviejo, D., and López, L.C. (2014). Ubiquinol-10 ameliorates mitochondrial encephalopathy associated with CoQ deficiency. Biochim. Biophys. Acta.

33. Silva-Fernandes, A., Duarte-Silva, S., Neves-Carvalho, A., Amorim, M., Soares-Cunha, C., Oliveira, P., Thirstrup, K., Teixeira-Castro, A., and Maciel, P. (2014). Chronic Treatment with 17-DMAG Improves Balance and Coordination in A New Mouse Model of Machado-Joseph Disease. Neurother. J. Am. Soc. Exp. Neurother.

34. Le Corre, S., Eyre, D., and Drummond, I.A. (2014). Modulation of the Secretory Pathway Rescues Zebrafish Polycystic Kidney Disease Pathology. J. Am. Soc. Nephrol. JASN.

35. Nóbrega, C., Nascimento-Ferreira, I., Onofre, I., Albuquerque, D., Hirai, H., Déglon, N., and de Almeida, L.P. (2013). Silencing mutant ataxin-3 rescues motor deficits and neuropathology in Machado-Joseph disease transgenic mice. PloS One *8*, e52396.

36. Manto, M., and Marmolino, D. (2009). Animal models of human cerebellar ataxias: a cornerstone for the therapies of the twenty-first century. Cerebellum Lond. Engl. *8*, 137–154.

37. Perdomini, M., Hick, A., Puccio, H., and Pook, M.A. (2013). Animal and cellular models of Friedreich ataxia. J. Neurochem. *126 Suppl 1*, 65–79.

38. Mohan, R.D., Dialynas, G., Weake, V.M., Liu, J., Martin-Brown, S., Florens, L., Washburn, M.P., Workman, J.L., and Abmayr, S.M. (2014). Loss of Drosophila Ataxin-7, a SAGA subunit, reduces H2B ubiquitination and leads to neural and retinal degeneration. Genes Dev. *28*, 259–272.

39. Yanicostas, C., Barbieri, E., Hibi, M., Brice, A., Stevanin, G., and Soussi-Yanicostas, N. (2012). Requirement for zebrafish ataxin-7 in differentiation of photoreceptors and cerebellar neurons. PloS One *7*, e50705.

40. Kabashi, E., Brustein, E., Champagne, N., and Drapeau, P. (2011). Zebrafish models for the functional genomics of neurogenetic disorders. Biochim. Biophys. Acta *1812*, 335–345.

41. Mahmood, F., Mozere, M., Zdebik, A.A., Stanescu, H.C., Tobin, J., Beales, P.L., Kleta, R., Bockenhauer, D., and Russell, C. (2013). Generation and validation of a zebrafish model of EAST (epilepsy, ataxia, sensorineural deafness and tubulopathy) syndrome. Dis. Model. Mech. *6*, 652–660.

42. Amsterdam, A., and Hopkins, N. (2006). Mutagenesis strategies in zebrafish for identifying genes involved in development and disease. Trends Genet. TIG *22*, 473–478.

43. Becker, T.S., and Rinkwitz, S. (2012). Zebrafish as a genomics model for human neurological and polygenic disorders. Dev. Neurobiol. *72*, 415–428.

44. Valdmanis, P.N., Dupré, N., Lachance, M., Stochmanski, S.J., Belzil, V.V., Dion, P.A., Thiffault, I., Brais, B., Weston, L., Saint-Amant, L., et al. (2011). A mutation in the RNF170 gene causes autosomal dominant sensory ataxia. Brain J. Neurol. *134*, 602–607.

45. Margolin, D.H., Kousi, M., Chan, Y.-M., Lim, E.T., Schmahmann, J.D., Hadjivassiliou, M., Hall, J.E., Adam, I., Dwyer, A., Plummer, L., et al. (2013). Ataxia, dementia, and hypogonadotropism caused by disordered ubiquitination. N. Engl. J. Med. *368*, 1992–2003.

46. Imamura, S., and Kishi, S. (2005). Molecular cloning and functional characterization of zebrafish ATM. Int. J. Biochem. Cell Biol. *37*, 1105–1116.

47. Aspatwar, A., Tolvanen, M.E.E., Jokitalo, E., Parikka, M., Ortutay, C., Harjula, S.-K.E., Rämet, M., Vihinen, M., and Parkkila, S. (2013). Abnormal cerebellar

development and ataxia in CARP VIII morphant zebrafish. Hum. Mol. Genet. *22*, 417–432.

48. Issa, F.A., Mazzochi, C., Mock, A.F., and Papazian, D.M. (2011). Spinocerebellar ataxia type 13 mutant potassium channel alters neuronal excitability and causes locomotor deficits in zebrafish. J. Neurosci. Off. J. Soc. Neurosci. *31*, 6831–6841.

49. Miller, G.W., Ulatowski, L., Labut, E.M., Lebold, K.M., Manor, D., Atkinson, J., Barton, C.L., Tanguay, R.L., and Traber, M.G. (2012). The α-tocopherol transfer protein is essential for vertebrate embryogenesis. PloS One *7*, e47402.

50. Bae, Y.-K., Kani, S., Shimizu, T., Tanabe, K., Nojima, H., Kimura, Y., Higashijima, S., and Hibi, M. (2009). Anatomy of zebrafish cerebellum and screen for mutations affecting its development. Dev. Biol. *330*, 406–426.

51. Apps, R., and Garwicz, M. (2005). Anatomical and physiological foundations of cerebellar information processing. Nat. Rev. Neurosci. *6*, 297–311.

52. Saint-Amant, L., and Drapeau, P. (1998). Time course of the development of motor behaviors in the zebrafish embryo. J. Neurobiol. *37*, 622–632.

53. Bockenhauer, D., Feather, S., Stanescu, H.C., Bandulik, S., Zdebik, A.A., Reichold, M., Tobin, J., Lieberer, E., Sterner, C., Landoure, G., et al. (2009). Epilepsy, ataxia, sensorineural deafness, tubulopathy, and KCNJ10 mutations. N. Engl. J. Med. *360*, 1960–1970.

54. Foo, J.-N., Liu, J.-J., and Tan, E.-K. (2012). Whole-genome and whole-exome sequencing in neurological diseases. Nat. Rev. Neurol. *8*, 508–517.

55. Doi, H., Yoshida, K., Yasuda, T., Fukuda, M., Fukuda, Y., Morita, H., Ikeda, S., Kato, R., Tsurusaki, Y., Miyake, N., et al. (2011). Exome sequencing reveals a homozygous SYT14 mutation in adult-onset, autosomal-recessive spinocerebellar ataxia with psychomotor retardation. Am. J. Hum. Genet. *89*, 320–327.

56. Li, M., Pang, S.Y.Y., Song, Y., Kung, M.H.W., Ho, S.-L., and Sham, P.-C. (2013). Whole exome sequencing identifies a novel mutation in the transglutaminase 6 gene for spinocerebellar ataxia in a Chinese family. Clin. Genet. *83*, 269–273.

57. Lee, Y.-C., Durr, A., Majczenko, K., Huang, Y.-H., Liu, Y.-C., Lien, C.-C., Tsai, P.-C., Ichikawa, Y., Goto, J., Monin, M.-L., et al. (2012). Mutations in KCND3 cause spinocerebellar ataxia type 22. Ann. Neurol. *72*, 859–869.

58. Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. *12*, 745–755.

Chapter II.

Exome sequencing identifies a splice mutation in *CWF19L1* in a consanguineous recessive ataxia family

Introduction

Autosomal recessive cerebellar ataxias are a clinically and genetically heterogeneous group of neurological disorders characterized by deficiencies in the coordination of movements, most prominently the limbs, trunk and eyes. While most forms of ataxia are individually rare, recessive ataxias are cumulatively not uncommon, with an estimated frequency of 1/20,000 that varies between countries [5,7]. Most suspected recessive ataxia cases test negative for the 21 ataxia genes that are routinely included in clinical genetic testing [5], suggesting that many recessive ataxia genes are still unknown. Many families that present in the clinic with ataxia of unknown origin are too small to allow gene identification by traditional genetic methods, such as positional cloning. Additionally, the genetic heterogeneity of the ataxia disorders presents a time consuming and costly barrier in the laboratory for identifying gene candidates in small families.

In 2005, a small consanguineous (parents are first cousins) Turkish family was

described in which two siblings were affected with a non-progressive, congenital ataxia in an autosomal recessive inheritance pattern. Symptoms include hypotonia, developmental delay, mental retardation, and truncal and extremity ataxia. The patients were negative for clinical tests for other organ system involvement, including skeletal X-rays, optic atrophy, and metabolic tests commonly administered to individuals with unexplained ataxia. MRI demonstrated vermian hypoplasia in association with hypoplasia of the cerebellar hemispheres[15]. The parents and a third sibling were clinically healthy.

Recently, next generation sequencing has proven to be a time- and cost-effective method when identifying novel gene candidates especially when used in combination with other genetic approaches, such as gene expression analysis and homozygosity mapping[23,58]. Next generation sequencing has recently been used to identify genes involved in rare neurological disorders [5,59], including ataxia[6,60–63], which has been helpful in annotating gene pathways and creating new targets for therapeutic intervention. Thus we set out to identify a novel genetic cause for this rare form of recessive ataxia in this family by combining next generation sequencing, linkage analysis and expression analysis. We also investigated the impact of the novel variant on mRNA stability and protein expression.

Materials and Methods

Human Subjects

Informed consent was obtained from participants and the Institutional Review Board of the University of Michigan Medical School approved this study. Heparinized blood from the affected individuals was separated by density centrifugation and transformed with Epstein Barr virus (EBV)[64] to create patient-specific lymphoblastoid cell lines. After growth initiation, aliquots of the cells were frozen and grown as needed.

Homozygosity mapping and exome sequencing

Homozygosity mapping was performed by hybridizing DNA from both affected individuals to high-density Sentrix Human Hap 550 genotyping chips (Illumina). Linkage analysis was performed by hybridizing DNA from both affected individuals to Infinium HumanLinkage-12 genotyping chips (Illumina) and data analyzed using Merlyn. Note: these linkage chips are no longer being sold. Exome capture was performed with the NimbleGen SeqCap EZ Exome Library v1.0 kit (Roche, Indianapolis, IN). The exon-enriched DNA from both affected individuals was sequenced with an Illumina HiSeq2000 instrument at the University of Michigan DNA Sequencing Core to an average depth of coverage of 20x. The exome data was filtered to variants that were 1) in the homozygosity regions, 2) homozygous in both individuals, and 3) predicted to change the protein sequence or expression (missense, nonsense, splice variants).

PCR and Sequencing

DNA was extracted from EDTA (lavender) blood samples using the Puregene Blood Core Kit (Qiagen). RNA was extracted from lymphoblastoid cell lines using TRIzol reagent according to the manufacturer's (Life Technologies, Grand Island, NY) instructions. RNA was subjected to DNase I treatment (Ambion, Grand Island, NY) and reverse transcribed using the Invitrogen (now Life Technologies, Grand island, NY) SuperScript II reverse transcription kit using Oligo dTs and random hexamers.

Sequence data were compared with the *CWF19L1* reference sequence NC_000010.11 (DNA) or NM_018294 (mRNA) with Lasergene software (DNAStar, Madison, WI).

CWF19L1 splice variant was also screened in 203 DNA samples (406 chromosomes) from controls consisting of the parents and unaffected sibling and 200 unaffected Turkish individuals. Additionally, 64 DNA samples (128 chromosomes) of unrelated ataxia patients were screened for other *CWF19L1* variants by Sanger sequencing.

Primers for sequencing the novel *CWF19L1* variant spanned exon 9 and the flanking intronic regions (F: 5'-CAG GAA GAA TCA GCC TGT CAG TT-3'; R: 5'-GGC CAA GGC CAT GTT TAT ATT T-3'). RT-PCR included primers that spanned exons 8 to 10 (F: 5'-GCC TCC GGA TGT CAC TGA AAA CCC T-3'; R: 5'-GGG CTA GCA AGG CAA AAC CAG CA-3'), exon 9 (F: 5'-GCC CCT GTG

GAA GAA TCA GCC TGT C-3'; R: 5'-GCA GGG TCC TGG AGG CTG AGG AG-3'), exons 7 to11 (F: 5'-TGC ACA GCA TGC CAC CCG GTT T-3'; R: 5'-TCC TTT GGC CAG GGC AAG GTA GC-3') and exons 10 to14 (F: 5'- ACC CTG CTG GTT TTG CCT TGC T-3'; R: 5'- GGG CTC AAA GTC TTT CCG GAA GCG G-3').

Microarray and qRT-PCR

RNA was extracted from lymphoblastoid cell lines (LCLs) of the affected siblings and from 28 cell lines from controls or individuals affected by other conditions. cRNAs was prepared by standard methods and hybridized to Illumina human genome expression micro-array (RefSeq8). Data was analyzed using Illumina BeadStudio.

qRT-PCR was performed on iQ5 cycler (BioRad, Hercules, CA). Assays were performed in 20 µl reaction mixtures, using a SYBR Green Master Kit, following the manufacturer's protocol. All measurements were done in triplicate. The threshold cycle value for each product was determined and normalized to that of the internal controls, GAPDH and β-actin. qRT-PCR results were analyzed using MyQ (BioRad, Hercules, CA).

Primers amplified sequence that spanned exon 6 (F: 5'-TTG GGA ATT CTT CTG GAG AAG TGG A-3'; R: 5'-GGT CTT TTC CAA AGC AGC AAA ATG G-3'), exons 8-10 (F: 5'-GCC TCC GGA TGT CAC TGA AAA CCC T-3'; R: 5'GGG CTA GCA AGG CAA AAC CAG CA-3'), and exon 11(F: 5'-GCC CTG GCC AAA GGA

GGC TT-3'; R: 5'-CTG TAG CTG GAG GTG ATG GCT CTT-3'). The reference gene primers amplified *GAPDH* (F: 5'-GAG TCA ACG GAT TTG GTC GT-3'; R: 5'-AAT GAA GGG GTC ATT GAT GG-3') and *RNA POL II* (RPII) (F: 5'-CTC TTC CAG CCT TCC TTC CT-3'; R: 5'-AGC ACT GTG TTG GCG TAC AG-3').

Protein electrophoresis and immunoblotting

Three batches of LCLs, each consisting of two homozygotes and two controls were cultured for 2 to 3 weeks, and lysate was prepared in RIPA buffer containing HALT protease inhibitor mixture (Thermo Scientific). Equal amounts of protein (30 µg per experiment) were electrophoresed on SDS-PAGE gels. SDS-PAGE gels were prepared to 10% polyacrylamide and were large gels (separating gel volume of 25ml volume, ~10 inches in length, 1mm thick). Samples were mixed with 2X Laemmli buffer with β-mercaptoethanol and heated at 100^{o}C for 5 minutes. Gels were run in 1X Tris-Glycine-SDS running buffer at a constant current of 15-30 milliamps and electrophoretically transferred at a constant current of 100 milliamps overnight at room temperature in 0.5X Tris-Glycine buffer onto a nitrocellulose membrane (Pall Life Sciences, BioTraceTM NT, Cat. No. 66485). Ponceau S staining of nitrocellulose membranes was performed according to the manufacturer's protocol (USB, Cat. No. 32819, prepared as directed) to confirm transfer of proteins onto the membrane.

Membranes were incubated in blocking buffer (5% wt/vol non-fat dry milk powder in TBST) for at least 1 hour. Primary and secondary antibodies were diluted in blocking buffer and subsequently incubated with intermediate washes

for 1 hour at room temperature. For C19L1, we used primary antibody HPA036889 (Sigma, St. Louis, MO) at a dilution of 1:2000; the secondary antibody was HRP-conjugated antibody against rabbit IgG diluted 1:10,000 (BioRad, Hercules, CA, Cat. No. 170-6515). All washes were performed 3 times, 10 minutes each, in TBST. Reactions were visualized using BioRad Clarity Western ECL substrate (BioRad, Hercules, CA, Cat. No. 170-5061). Blots were stripped with Restore Western Blot Stripping Buffer (Thermo Scientific Cat No. 21059) and reprobed with primary antibody β-Tubulin (Abcam, Cambridge, MA) at a dilution of 1:1,000 using the same secondary antibody and ECL Western Blotting Substrate (BioRad, Hercules, CA).

In silico **analysis**

H. sapiens (C19L1): NP_060764.3

S. pombe (mug161): NP_593012.1

H. sapiens (C19L2): NP_689647.2

S. pombe (cwf19): NP_593208.1

BioGPS[65], PomBase[66] and UniProt[67] were used to identify all cwf genes, subcellular localization and function.

Results

Exome sequencing identifies a mutation in *CWF19L1* in affected siblings

The consanguineous family (parents are first cousins) has been previously described[15]. Two siblings, but not their parents or their unaffected sibling, are

affected with hypotonia, developmental delay, mental retardation, and non-progressive truncal and extremity ataxia. MRI demonstrated vermian hypoplasia in association with hypoplasia of the cerebellar hemispheres[15] (Figure 2.1). To identify candidate genes causing ataxia in this family, we used whole-exome sequencing filtered by homozygosity. Homozygosity mapping identified 13 regions > 500kb, spanning a total of 71 Mb, ~2% of the genome, as homozygous and shared between the siblings, reaching LOD scores of 1.0-1.8 (Figure 2.1). These regions contained 485 candidate genes, but no previously known ataxia genes.

Exome sequence variants were filtered for (a) being on target (in or near exons), (b) the homozygosity regions, and (c) predicted damaging function (nonsense, missense and splice variants, see Figure 2.2). 4,969 of the exome variants were shared between the siblings, and of these shared variants 713 were homozygous. 28 of the shared and homozygous variants mapped to homozygosity regions, with onl 17 of those within coding regions (Table 2.1). Three variants were potentially damaging: 2 missense mutations in conserved amino acids and one splice site mutation (Table 2.1). We tested all 17 variants using Sequenom's MassARRAY genotyping analysis in the patients and 202 unrelated individuals. None of the three most promising variants were present in 202 American controls, but several of the 14 others could be excluded because they were present in controls. The 3 promising variants were also absent from the 13,000 chromosomes available in the EVS server (National Heart, Lung, and

Blood Institute Exome Variant Server (NHLBI GO Exome Sequencing Project ESP), Seattle, WA (URL:http://evs.gs.washington.edu/EVS/) [October 2011, similar results February 2014]).

Of the three candidate gene mutations, two were unlikely to cause ataxia: a missense mutation in *CDC73*, since other mutations in this gene cause hyperparathyroidism, jaw syndromes and cancer[68,69], and a missense mutation in *GSTO1*, since mutations in *GSTO1* in humans and loss of *GSTO1* in mice and drosophila do not cause any discernible phenotype [70–72]. The third candidate mutation was a mutation in *CWF19L1* in an obligatory splice site (c. 964+1 G>A, i.e. the first base of intron 9, always G, is changed to an A, Figure 2.3). Use of bioinformatic sites such as HPRD[10], Ensembl[11], and FANS[12] identified the G>A variant in *CWF19L1* as a change in an obligatory splice site that is predicted to be highly damaging to the protein. While the function of the *CWF19L1* gene is unknown, it is highly conserved across species, suggesting the evolutionary importance of this gene. Taken together, *CWF19L1* was the most promising candidate gene variant (Figure 2.2).

Sanger sequencing confirmed the *CWF19L1* mutation in the affected siblings and demonstrated that it is absent in 200 Turkish individuals without neurological disorder. The parents are each heterozygous for this mutation and the unaffected sibling is homozygous for the reference allele (Figure 2.3). In 64 other subjects

with ataxia, we detected none who were homozygous or compound heterozygous for damaging *CWF19L1* mutations.

Splice mutation causes exon skipping, decreased mRNA levels, and loss of protein

The splice donor site mutation after exon 9 is predicted to eliminate a splice site immediately following exon 9. We tested this prediction by analyzing mRNA from lymphoblastoid cell lines (LCLs) of the two affected siblings and control cell lines. RT-PCR for the segment spanning exons 8-10 revealed a product of 234bp in control individuals. However, RT-PCR of RNA from both affected individuals demonstrated a smaller product of 119 bp. Sanger sequence analysis of the RT-PCR products showed a short cDNA splice product missing 119 bp that are equivalent to all of exon 9 (Figure 2.4). In order to determine the full impact of the mutation on mRNA processing, we performed RT-PCR using primers specific to exon 9 and other exons of *CWF19L1* and found no evidence of any exon 9 transcript in the affected individuals and no additional splicing aberrations (Figure 2.5).

As the cellular machinery reads the bases of mRNA transcripts in sets of 3 for protein translation, deletion of the 119 bases of exon 9, which is not divisible by three, causes a shift in the reading frame of this mRNA. This frameshift is predicted to cause a different translation from the canonical protein and is predicted to cause translation of an early “STOP” codon 60 amino acids

downstream of exon 9. The truncation of this aberrant protein is predicted to cause mRNA transcripts to undergo nonsense-mediated decay, which would cause decreased levels of *CWF19L1* mRNA in affected individuals. To test if the mutation affects gene expression, we extracted RNA from lymphoblastoid cell lines (LCLs) from the affected siblings and 28 other individuals, prepared cRNAs, and hybridized them to Illumina human genome expression micro-arrays (RefSeq8). The level of expression of *CWF19L1* in (LCLs) from the two affected children was 2 to 8-fold less when compared to 28 other individuals (ranked within the 10 largest expression changes for these two affected individuals) (Figure 2.6) while expression of *GSTO1* and *CDC73* were unchanged. However, because our sample size is only 2, a formal analysis of whether the difference is significant was not possible. Since the other samples are not from children, and not from Turkey, this result by itself is not convincing, but is an additional piece of evidence to consider in the overall assessment of a candidate gene. To further quantify the expression changes, we performed qRT-PCR using primers designed to *CWF19L1*. Normalized against *GAPDH* and *ACTIN-*β mRNA from the affecteds is ~6-fold reduced compared to that of control (n=3) individuals (Figure 2.7).

The *CWF19L1* gene is predicted to produce 3 different protein isoforms from different methionine start sites (in exon 1, in exon 5 and in exon 9)[67]. The *CWF19L1* variant was shown to cause skipping of exon 9 and decreased levels of *CWF19L1* mRNA, potentially causing a significant decrease or ablation of

C19L1, the protein encoded by *CWF19L1*. Since exon 9 is included in all isoforms, the *CWF19L1* variant would be predicted to impact all protein isoforms (Figure 2.8A). In order to test the hypothesis that protein is affected by the *CWF19L1* splice site mutation, we performed Western blots using proteins extracted from LCLs of the affected siblings and control individuals. Immunoblot was performed using commercial antisera for C19l1 (Sigma) and β-Tubulin. The C19L1 antisera recognizes an epitope 5' to the mutation, causing it to detect two of the three isoforms (Figure 2.8A). Western blot analysis demonstrated that control individuals (lanes 1 and 3, Figure 2.8B) have a two-band pattern for C19L1 while the patients (lanes 2 and 4, Figure 2.8B) show no bands for C19L1. The absence of protein bands in the patients demonstrates specificity of this antiserum. In affected individuals, there was no evidence for presence of normal or the predicted truncated (344 amino acid; mw 38kDa) protein, even when overloading or overexposing the blot suggesting true loss of the protein in the affected siblings.

Discussion

Using a combination of homozygosity mapping and next generation sequencing, we have identified a novel genetic candidate for a rare, recessive ataxia-mental retardation syndrome. Many recent studies have demonstrated the use of homozygosity mapping and exome sequencing in the identification of candidate genes in small families and even in single individuals. These studies include genetically heterogeneous disorders such as ataxia[63,73–75]. It is estimated that

40% of recessive ataxias still have unknown etiology [5]; therefore, discovery of this gene identifies a novel cause for ataxia which may lead to the identification of novel pathways in ataxia or may uncover a role for this gene in current ataxia pathways.

While *CWF19L1* is a good candidate gene, all other damaging variants in the exome should be evaluated for their pathogenicity as to not exclude *de novo* variations in this family. Analysis of other homozygous variants that were in genic regions and were predicted to change the encoded protein (missense, nonsense, splice changes; Table 2.2) identified no variants or genes that were previously associated with ataxia or neurological disorders. Also, none of these genes turned out to be potential causal genes based on their function or brain expression. Interestingly, most of the homozygous exome variants were intronic, non-damaging, or in noncoding regions (5' UTR and 3'UTR). Therefore, the genetic and molecular data that demonstrate linkage and homozygosity at the locus and damaging impact of the mutation on the expression of the gene provide convincing evidence that *CWF19L1* is the causal gene.

The identification of a mutation in *CWF19L1* in siblings with congenital ataxia suggests that this gene plays a role in neuronal development. We postulate that the C19L1 protein is important in the normal development of the brain, and especially the cerebellum, since the affected individuals are affected with cerebellar hypoplasia[15]. This suggests that severe deficiency or loss of the

protein causes abnormal development. *CWF19L1* encodes three protein isoforms. Our data suggest that the mutation leads to a null allele, as RT-PCR detected no evidence of mRNA that included exon 9 in affected individuals and no protein could be detected even by overexposure (not shown). We also did not detect the truncated protein product, which translates 60 aberrant amino acids after the end of exon 8 before the "STOP" codon, suggesting rapid degradation of this aberrant protein. Our data, however, does not differentiate between a severe hypomorph and a null allele, since we can't exclude the possibility that some normal splicing persists, and a small amount of normal protein is made below our level of detection.

The function of the *CWF19L1* encoded C19L1 protein is unknown. *CWF19L1* is the gene symbol for "complexed with cdc5 protein-like 1" after the cwf19 gene in *S. pombe*. There are 29 genes annotated in Pombase, an online resource for fission yeast, as "complexed with cdc5 protein" . Many of these cwf proteins have human orthologs (Table 2.3) [66,67,76]. These 29 cwf genes were identified in a large-scale proteomics assay that utilized tandem affinity purification and mass spectrometry to identify the components of the fission and budding yeast multi-protein splicing complexes[77,78]. While the cwf19 protein was found in the spliceosome in yeast and humans[77,79], it is unclear whether it has an essential role in the spliceosome or if it performs an accessory role[77–79]. Some of the other cwf proteins have been shown to play an accessory role in the spliceosome, but also perform other functions in the cell, suggesting multiple actions of those cwf

proteins (Table 2.3, column 8)[67,80]. *CWF19L1* may play additional roles in the cell, as *CWF19L1* has been shown in a large human yeast two-hybrid study to interact with *TOM1L1*, a Src-activating and signaling molecule, which is not present in yeast.

While human *CWF19L1* is named for *cwf19*, the orthologous *S. pombe* gene is *mug161* (meiosis up-regulated 161)[66,67]. Human *CWF19L1* is part of the CWF19 protein family, proteins that contain two C-terminal CwfJ domains (Figure 4.9). Interestingly, the orthologs to *mug161* also have an N-terminal metallophosphatase (MPP) domain (Figure 2.9) suggesting they have different or additional function when compared with the *cwf19* orthologs[66,67,76].

The CwfJ domain is particular to the cwf19 family of genes and is not described in the literature. Within the first CwfJ domain, C19L1 is predicted to have a Histidine Triad (HIT)-like domain[66,67,76]. Proteins with HIT- domains belong to a superfamily of nucleotide hydrolases and transferases that act on ribonucleotides[81]. These proteins are traditionally classified into three branches: Histidine Triad Nucleotide Binding (HINT), which are adenosine 5'-monophosphoramide hydrolases that have tumor suppressor properties; Fragile histidine triad (FHIT), which are diadenosine polyphosphate hydrolases and also have tumor suppressor properties, and finally the Galactose-1-phospahate uridylyltransferase (GalT) branch, which consist of specific nucleoside monophosphate transferases[81–83]. While the HIT domain containing proteins

have been linked with tumor suppression, mutations in these proteins have also been shown to cause axonal neuropathy[84,85]. Although these HIT domain-containing proteins have been studied and implicated in disorders, many of their particular substrates are still unknown and how these genes cause disorders are poorly understood. Interestingly, aprataxin, which causes ataxia with oculomotor apraxia 1 (AOA1), has been identified as a HIT family protein but is thought to form a new branch of HIT proteins because it binds DNA, in addition to RNA, and is implicated in DNA repair, in addition to its nucleotide hydrolase abilities[82,86–88]. The C19L1 domain is identified as "HIT-like" so it is unclear if *CWF19L1* has hydrolase or transferase activity similar to the HIT-domain containing proteins.

C19L1 also contains a metallophosphatase (MPP) domain in the N-terminus [66,67,76]. Metallophosphatases are a superfamily of enzymes that include a conserved domain that has an active site that bind two metal ions (manganese, iron, or zinc) and a cage of histidine, aspartate, and asparagine residues[89,90]. The MPP superfamily includes phosphoprotein phosphatases (PPPs), Mre11/SbcD-like exonucleases, Dbr1-like RNA lariat debranching enzymes, YfcE-like phosphodiesterases, purple acid phosphatases (PAPs), YbbF-like UDP-2,3-diacylglucosamine hydrolases, and acid sphingomyelinases (ASMases). These proteins are involved in a variety of cellular processes and encompass many different proteins. The MRE11 protein, which causes ataxia-telangiectasia-like disorder or progressive myoclonic ataxia, has an MPP domain and has been shown to be a part of the ATM DNA strand break repair pathway in ataxia[91].

Additionally, DRN1, debranching enzyme-associated ribonuclease 1, is the *S. cerevisiae* ortholog of *mug161* suggesting similarity to the debranching enzymes branch of the MPP superfamily. Further studies will be needed to determine the role of the C19L1 protein in brain development and function.

Acknowledgements

Linkage analysis data was run in MERLIN by Dr. Karen Macjenko. Next generation sequencing data was assessed for quality and alignment in the lab of Dr. Jun Li. Technical assistance for DNA preparation, cell culture and data assessment were provided by Elzbieta Sliwerska, Linda Gates and Dr. Erin Sandford.

We thank the family for their participation. We also acknowledge the University of Michigan DNA Sequencing Core (Dir. Robert Lyons) for exome sequencing, Dr. Miriam Meisler (University of Michigan, Dept of Human Genetics) and Dr. Mustafa Cengiz Yakicier (ACIBADEM University, Turkey Dept. of Medical Biology and Genetic Diagnostic Center) for ataxia and Turkish control DNA samples, respectively.

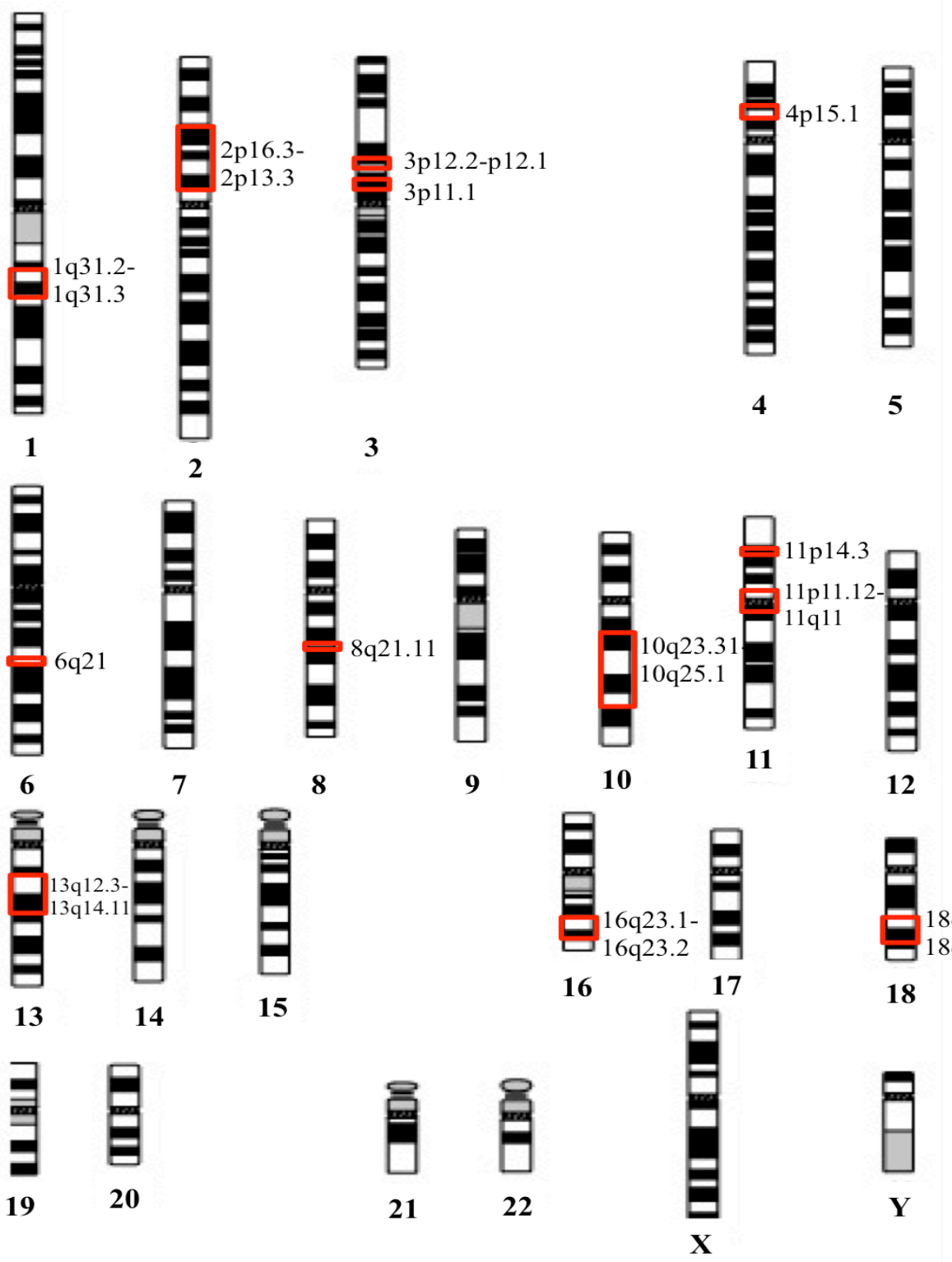


Figure 2.1. 13 genetic loci for ataxia identified by homozygosity mapping in Turkish consanguineous family

Schematic of each chromosome is shown. Red boxes and cytogenetic map locations indicate approximate positions of positive regions of homozygosity larger than 500kB.

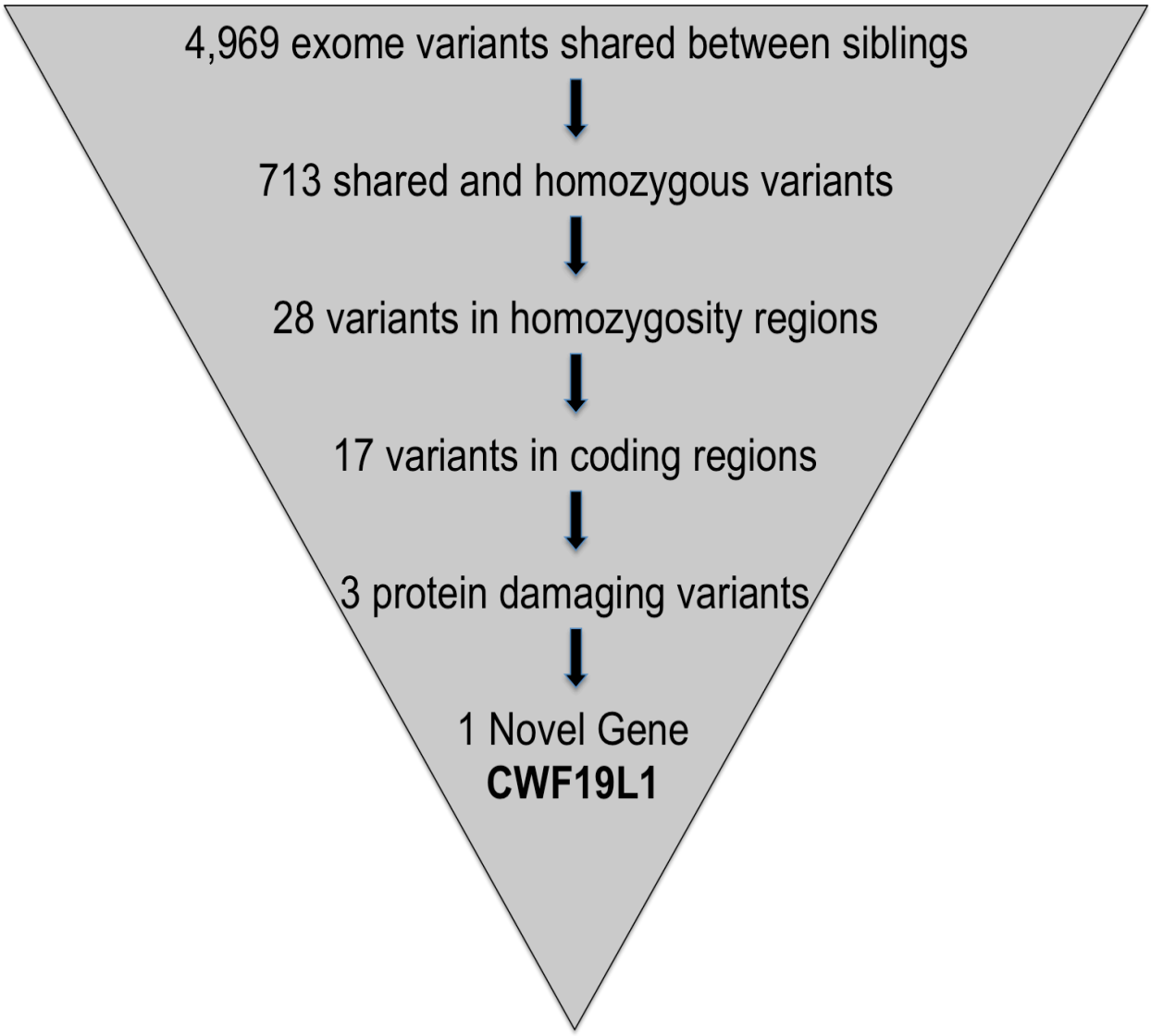


Figure 2.2. Exome sequencing data filtered by homozygosity identifies *CWF19L1* as promising candidate gene

Schematic demonstrates data filtration method used. Variants were narrowed by homozygosity mapping regions, variants that were contained in coding regions and that were predicted to be damaging (missense, nonsense, or splice site variants) to their encoded protein.

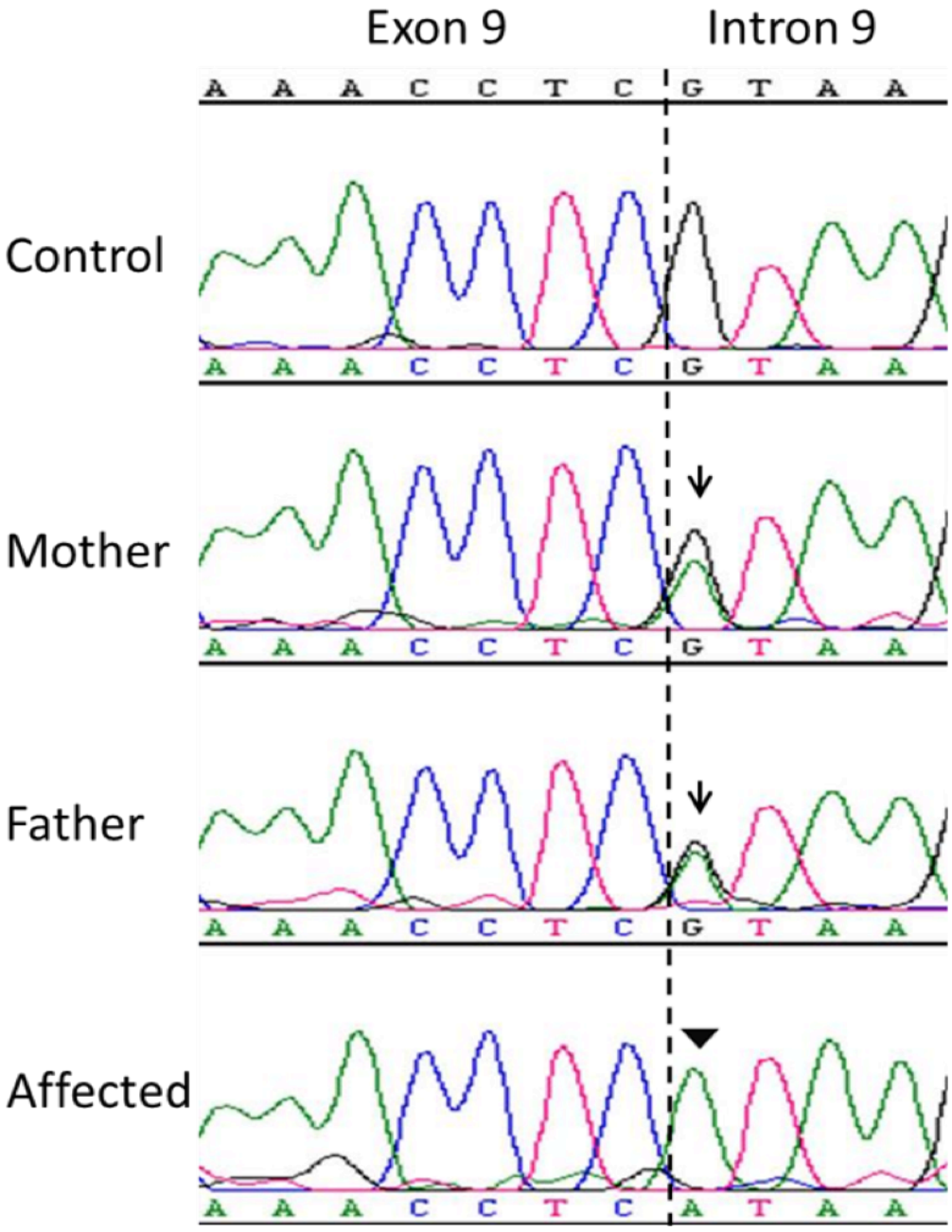


Figure 2.3. Genomic sequencing chromatogram validates exome sequence data

Sequencing chromatogram demonstrates homozygous reference allele (G) in control individual. Sequencing of parental DNA demonstrates heterozygous (G/A) alleles indicated by black arrows. Affected individual shows homozygous variant allele (A) at the position indicated by filled arrowhead.

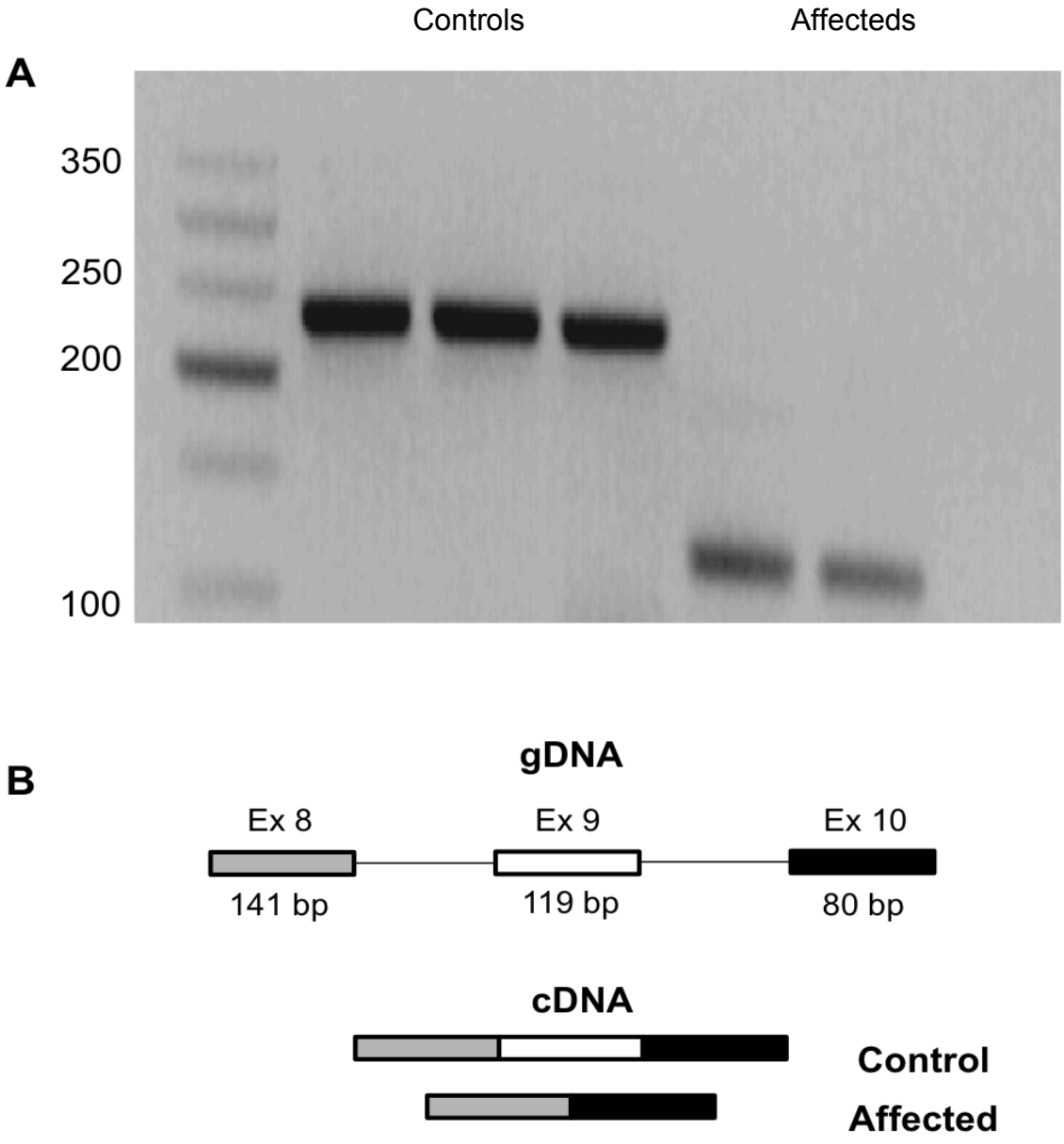


Figure 2.4. RT-PCR and sequencing demonstrates excision of exon 9

(**A**) RT-PCR using primers in exons 8 and 10 show normal product bands in control individuals (lanes 2-4) but a ~100bp smaller product band in the affected individuals (lanes 5-6). No cDNA control shown in lane 7. Sanger sequencing verifies splicing out of exon 9. (**B**) Schematic of RT-PCR Sanger sequencing results.

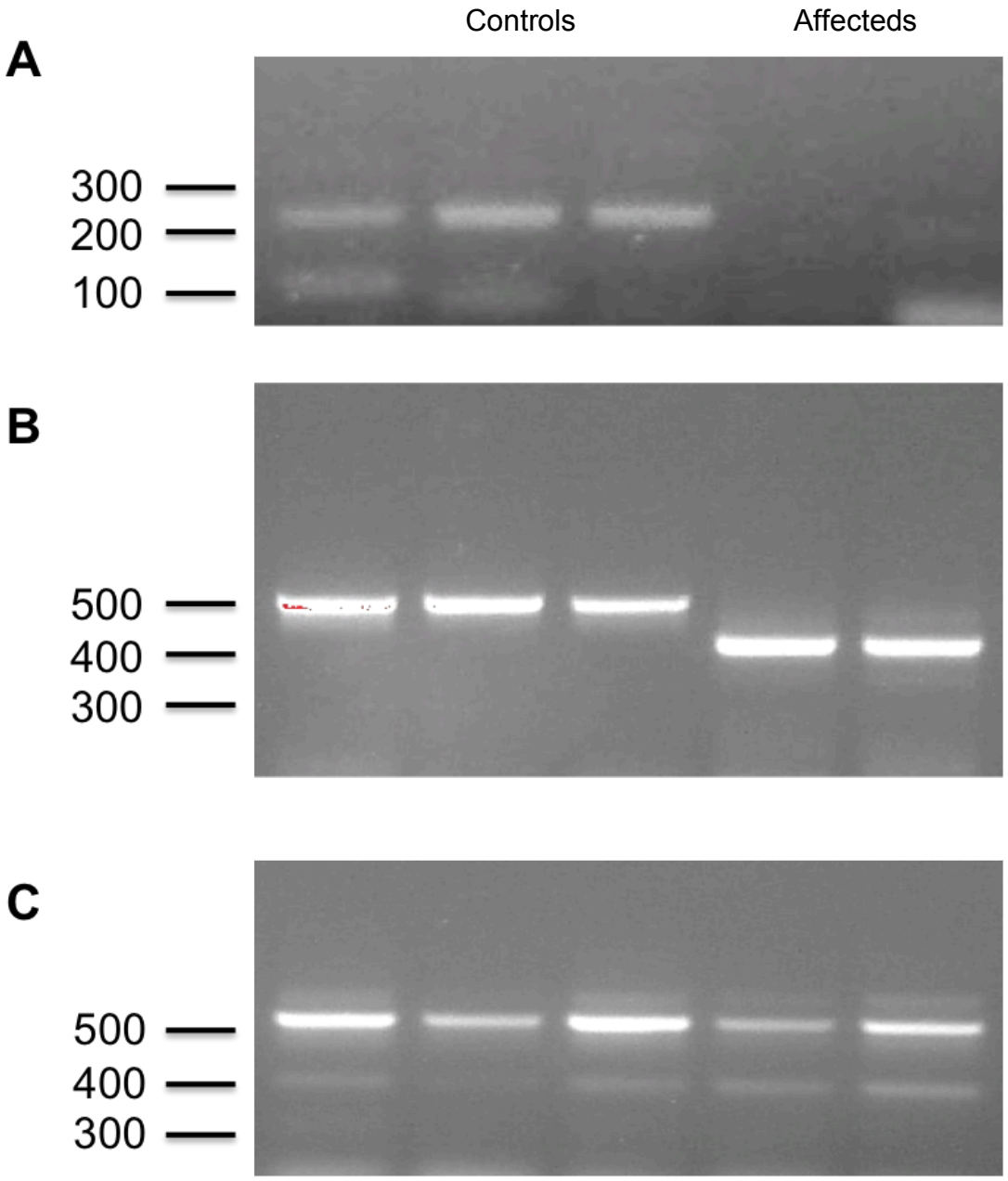


Figure 2.5. No additional splicing aberrations shown by RT-PCR

RT-PCR using primers in exon 9 (**A**), exons 7-11 (**B**), and exons 10-14 (**C**). Control samples in lanes 1-3. Affecteds in lanes 4-5.

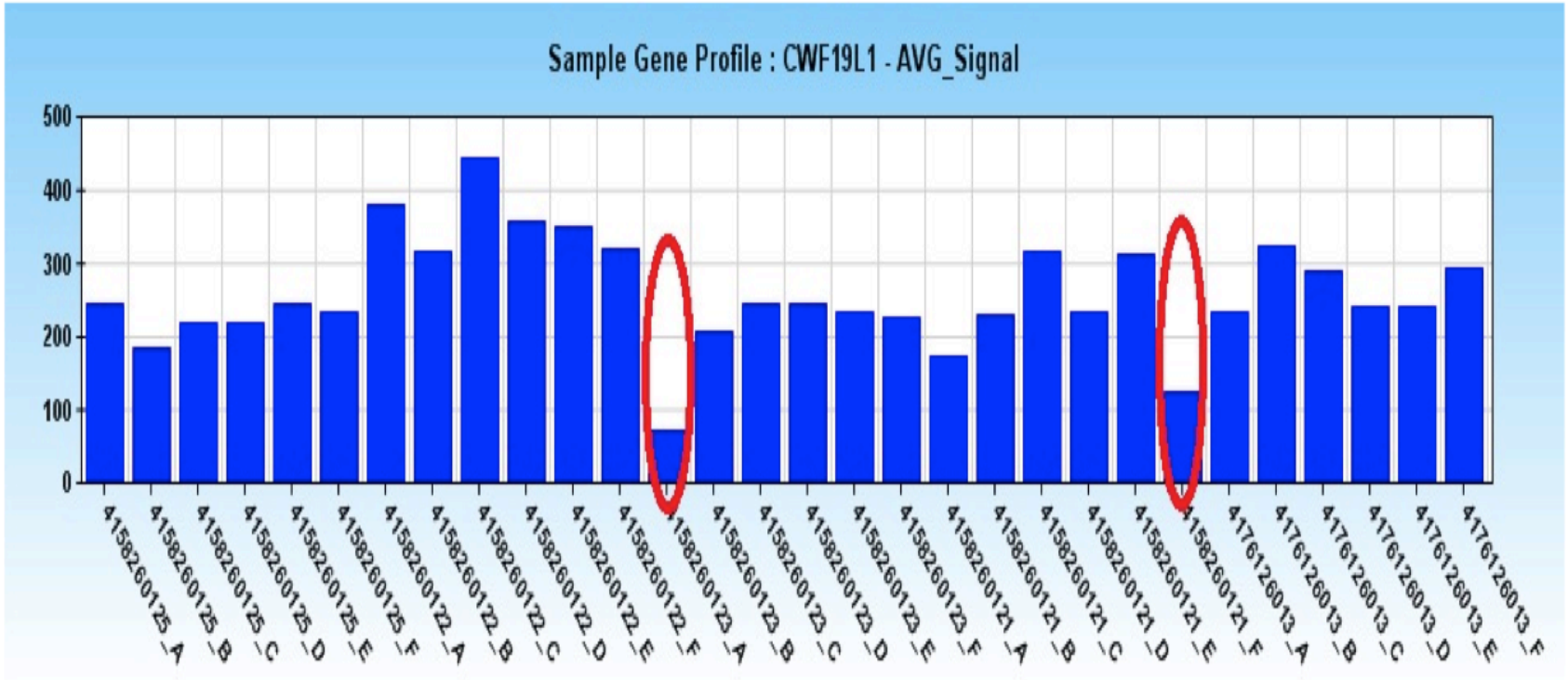


Figure 2.6. Microarray shows decrease in mRNA levels in affected individuals

Expression level of *CWF19L1* in LCLs from the two affected children (red circles) compared to 28 other samples run in the same batch on Illumina expression microarray.

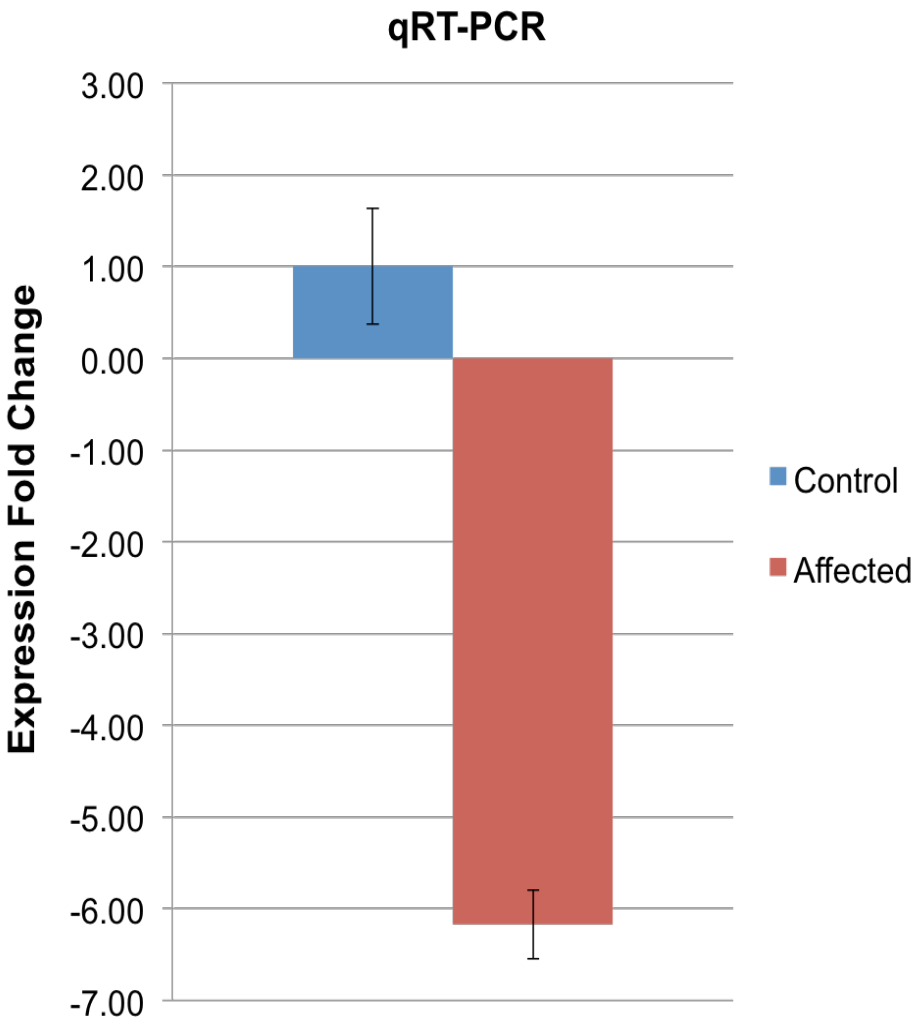


Figure 2.7. qRT-PCR shows decrease in *CWF19L1* mRNA in affected individuals

Primers were designed to exons 8-10 and normalized to *GAPDH.* Affected individuals, shown by red bar, demonstrate ~6 fold decrease in *CWF19L1* mRNA in affected individuals

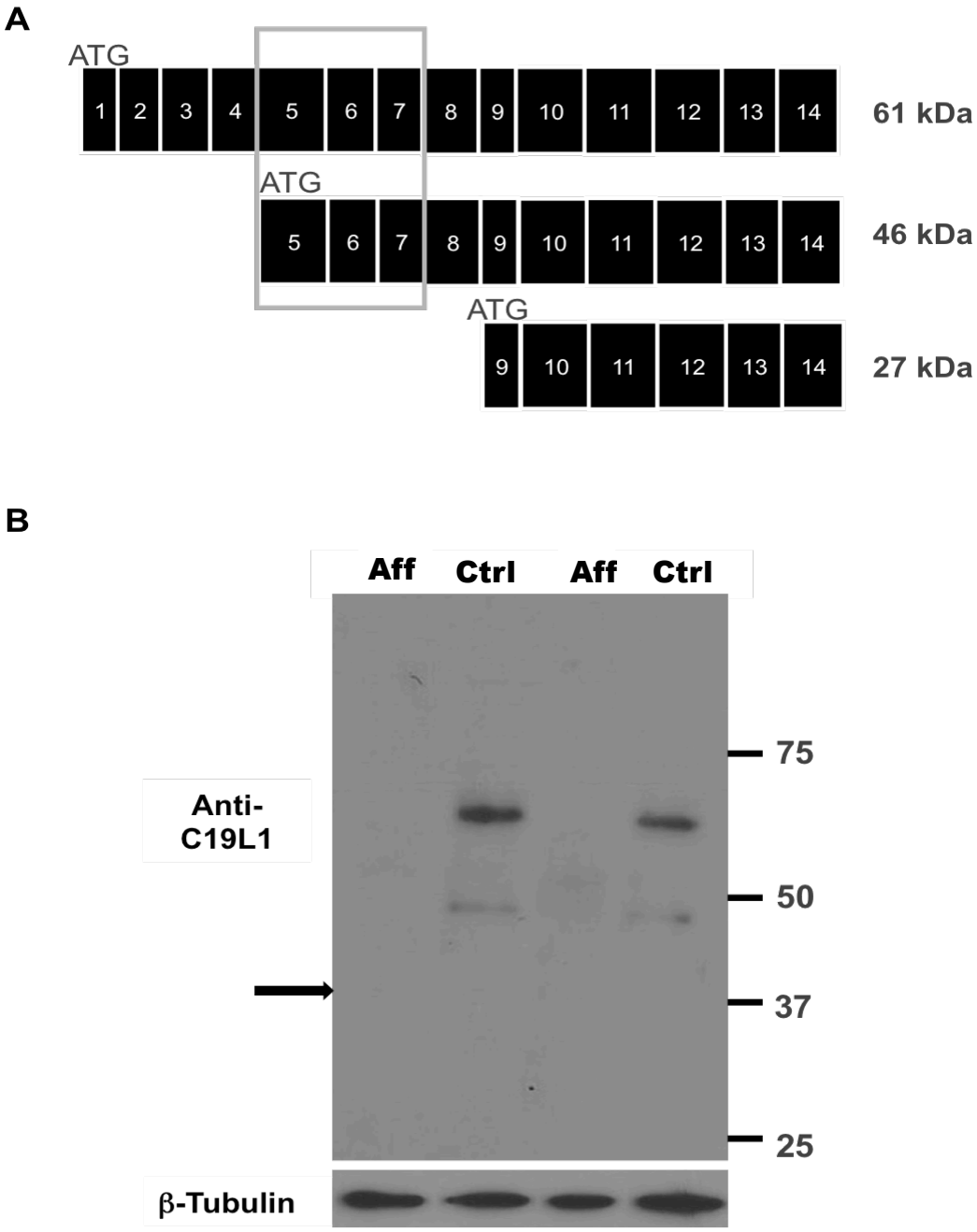


Figure 2.8. Immunoblot demonstrates loss of protein in LCLs of affected individuals

A) Schematic of C19L1 protein isoforms. Red box indicates C19L1 antibody epitope. (**B**) Western blot using C19L1 antiserum (Sigma) detects two C19L1 protein bands in control individuals. No protein bands are detected in **affecteds**. 30ug protein loaded. β-tubulin loading control. Predicted size of mutant truncated protein product is indicated by arrow.

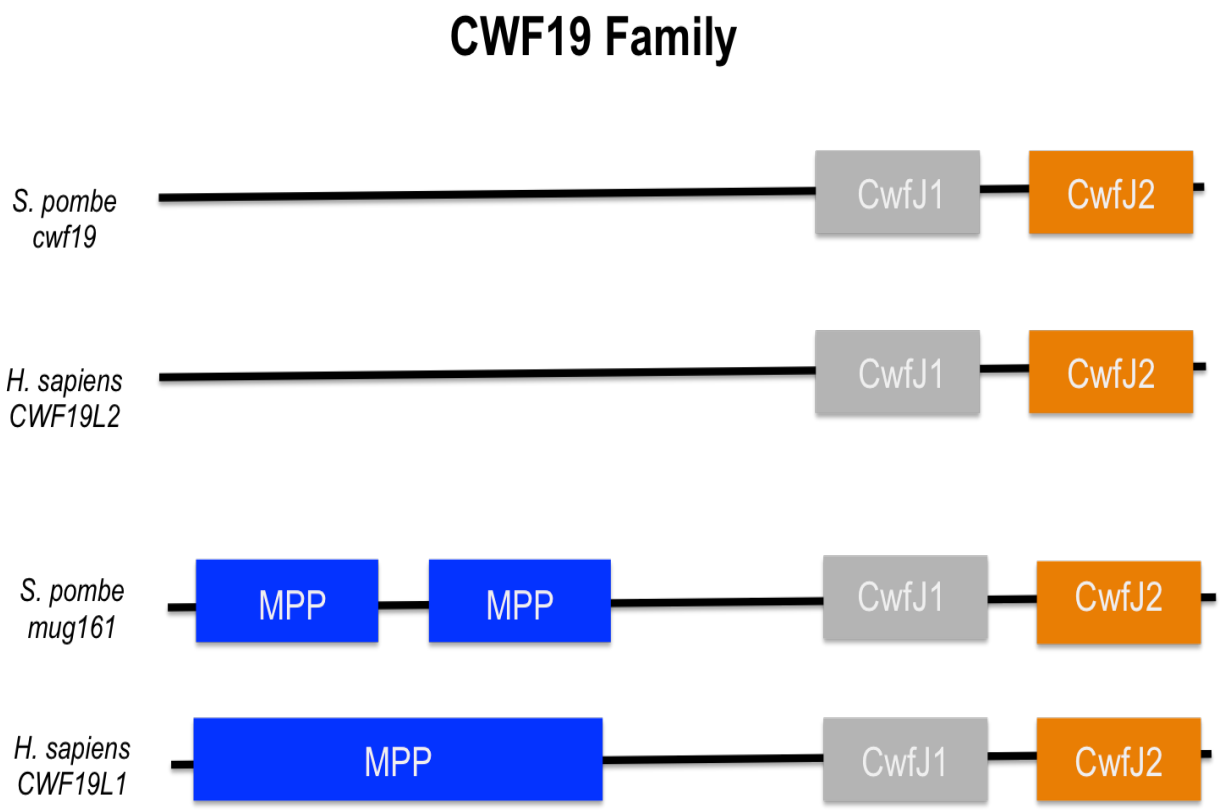


Figure 2.9. Functional domains of CWF19 family members

Schematic of CWF19 family of proteins in *S. pombe*, *H. sapiens* and *D. rerio*. Schematic showing C19L1 and C19L2 proteins demonstrating location of N-terminal metallophosphatase domain (MPP, blue box) and two C-terminal CwfJ domains (grey and orange box). The HIT-like domain is predicted to be buried within CwfJ domain 1 (grey box).

Table 2.1. Homozygous exome sequencing variants mapped in the homozygosity regions

Gene	Chr	Genomic position	Nucleotide Change	Protein	Protein Change	GERP Score	Present in EVS
CDC73	1	193,218,978	A->T	Parafibromin	Missense	0.305	N
ATP6V1G3	1	198,498,132	C->A	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G3	Intronic	-0.733	N
C2orf63	2	55,403,220	G->C	Clathrin heavy chain linker domain containing 1	Intronic	0.168	N
CCDC88A	2	55,582,926	A->G	Coiled-coil domain containing 88A	Intronic	1.75	N
XPO1	2	61,717,713	A->G	Exportin 1	Intronic	-0.268	N
C1D	2	68,269,974	G->T	C1D Nuclear Receptor Corepressor	3' UTR	1.9	N
CLEC4F	2	71,043,130	G->A	C-type lectin domain family 4, member F	Synonomous	-6.62	N
LIPA	10	91,122,063	T->C	Cholesteryl Esterase	Intronic	-0.691	N
RPP30	10	92,656,145	T->C	Rnase P Protein P30	Synonomous	-10.7	Y
BTAF1	10	93,746,700	C-.T	RNA Polymerase II, B-TFIID Transcription Factor Associated 170kD	Intronic	1.44	N
SORBS1	10	97,143,933	A->T	Sorbin and SH3 Domain Containing 1	Intronic	-2.91	Y
COX15	10	101,486,384	C->T	Cytochrome C Oxidase Assembly Homolog 15	Intronic	0.821	N
CWF19L1	10	102,005,555	C->T	CWF19-like protein 1	Splice Site	5.65	N
GSTO1	10	106,025,888	C->A	Glutathione S Transferase omega 1	Missense	4.2	N
C13orf23	13	39,585,497	A->G	Proline and Serine Rich 1	UTR-3	-5.73	N
DCLK1	13	36,402,011	G->A	Doublecortin-Like Kinase 1	Intronic	-6.26	N

SOHLH2	13	36,744,740	C->T	Spermatogenesis Associated 28	Synonomous	2.2	Y

Table 2.2. Complete list of shared homozygous variants that change encoded protein

Gene	Chr	Genomic position	Nucleotide Change	Protein	Protein Change	GERP Score	Damaging
CDC73	1	193,218,978	A->T	Parafibromin	Missense	0.305	Tolerated
CWF19L1	10	102,005,555	C->T	CWF19-like protein 1	Splice Site	5.65	-
GSTO1	10	106,025,888	C->A	Glutathione S-Transferase, Omega 1	Missense	4.2	Damaging
FCGR2C	1	161,560,943	T->G	Fc Fragment of IgG	Missense	1.52	Probably
OR2T34	1	248,737,259	C->G	Olfactory receptor 2T34	Missense	1.84	Benign
OR2T34	1	248,801,778	A->T	Olfactory receptor 2T34	Missense	1.6	Probably
RAB6C	2	130,738,163	G->A	Ras-related protein Rab-6C	Missense	-1.04	Tolerated
UGT2A3	4	69,795,626	C->T	UDP-Glucuronosyltransferase 2A3	Missense	2.12	Tolerated
AKAP9	7	91,646,406	G->A	A-Kinase Anchor Protein 9	Missense	-1.03	Tolerated
UCHL3	13	76,169,058	G->A	Ubiquitin Thioesterase	Missense	-0.712	Tolerated
POTEG	14	19,562,052	T->A	POTE-14	Missense	0.98	Damaging
OR3A3	17	3,324,004	C->T	Olfactory receptor 3A3	Missense	0.178	Tolerated

KRTAP9-8	17	39,394,755	C->A	Keratin Associated Protein 9-8	Missense	-4.31	Tolerated
KIAA1267	17	44,144,993	C->G	Centromere Protein	Missense	3.99	Tolerated
LRRC37A2	17	44,626,660	C->G	Leucine-Rich Repeat Containing 37A2	Missense	-4.47	Tolerated
POTEC	18	14,533,105	G->C	POTE-18	Missense	-2.08	Tolerated

Table 2.3. Listing of *S. pombe* "complexed with cdc5p" (cwf) genes and human orthologs

Yeast Gene	Human Ortholog	Protein	Deletion Phenotype	Subcellular Localization	Functional Domain(s)	Associated Disease/ Animal Model	Function
prp5 cwf1	PLRG1	**Pleiotropic regulator 1**	Inviable	Nucleus speckle	7 WD40 repeats	Embryonic lethal mouse and zebrafish	Interacts with cdc5L for pre-mRNA splicing
prp3 cwf2	PRPF3	**U4/U6 snRNP 90 kDa protein**	Inviable	Nucleus speckle	PWI domain	AD Retinitis Pigmentosa 18	Part of the U4/U5/U6 tri-snRNP complex, one of the building blocks of the spliceosome
syf1 cwf3	XAB2	**XPA-binding protein 2**	Inviable	Nucleus	14 HAT domains, TPR domain	None	Involved in transcription-coupled repair (TCR), transcription and pre-mRNA splicing
syf3 cwf4	CRNKL1	**Crooked neck-like protein 1**	Inviable	Nucleus/ Cytoplasm	17 HAT repeats	None	Involved in pre-mRNA splicing process

cwf5	RBM22	**RNA-binding motif protein 22**	Inviable	Nucleus/ Cytoplasm	RNA recognition motif; C3H1 type zinc finger	Mitotic arrest with shRNA knockdown in cells	Involved in the first step of pre-mRNA splicing
cwf6	PRPF8	**220 kDa U5 snRNP-specific protein**	Inviable	Nucleus Speckle	MPN (JAB/Mov) domain; RRM domain	AD Retinitis Pigmentosa 13	Functions as a scaffold that mediates the ordered assembly of spliceosomal proteins and snRNAs
spf27 cwf7	BCAS2	**Breast carcinoma-amplified sequence 2**	Inviable	Nucleus-Nucleolus	Coiled-coil	None	May have a scaffolding role in the spliceosome assembly as it contacts all other components of the core complex.
prp19 cwf8	PRP19	**Pre-mRNA-processing factor 19**	Inviable	Nucleus/ Cytoplasm	7WD40 repeats, U-box domain	None	DNA double-strand break (DSB) repair. Essential for spliceosome assembly and stability. May have E3 ubiquitin ligase activity.

smd2 cwf9	snrpd2	**snRNP core protein D2**	Inviable	Nucleus/ Cytoplasm	None	None	Core component of the spliceosomal U1, U2, U4 and U5 small nuclear ribonucleoproteins (snRNPs), Thereby, plays a role in the splicing of cellular pre-mRNAs.
cwf10	EFTUD2	**116 kDa U5 small nuclear ribonucleoprotein component**	Inviable	Nucleus	Elongation factor C, P-loop NTPase	Mandibulofacial dystosis with microcephaly (MFDM)	Component of the U5 snRNP and the U4/U6-U5 tri-snRNP complex required for pre-mRNA splicing. Binds GTP.
cwf11	AQR	**Intron-binding protein Aquarius**	Viable curved elongated	Nucleus	None	None	Intron-binding spliceosomal protein required to link pre-mRNA splicing and snoRNP (small nucleolar ribonucleoprotein) biogenesis.
isy1 cwf12	ISY1	**Pre-mRNA-splicing factor ISY1 homolog**	Viable elongated	Nucleus	None	None	regulates the fidelity of pre-mRNA splicing

prp45 cwf13	SNW1	**SNW domain-containing protein 1**	Inviable	Nucleus	SNW, Proline rich domain	None	Involved in transcriptional regulation. Is required in the specific splicing of CDKN1A pre-mRNA.
cwf14	BUD31	**Protein BUD31(G10) homolog**	Viable Elongated	Nucleus	G10 protein	None	Component of the SF3b subcomplex of the U2 snRNP
cwf15	CWC15	**Spliceosome-associated protein CWC15 homolog**	Inviable	Nucleus	Coiled coil	None	Component of the PRP19-CDC5L complex that forms an integral part of the spliceosome and is required for activating pre-mRNA splicing
cwf16	CCDC94	**Coiled-coil domain-containing protein 94**	Viable Normal	Nucleus	Coiled coil	None	may be involved in splicing
spf38 cwf17	SNRP40	**U5 small nuclear ribonucleoprotein 40 kDa protein**	Viable Normal	Nucleus	7 WD40 repeats	None	Component of the U5 small nuclear ribonucleoprotein (snRNP) complex.

cwf18	CCDC12	**Coiled-coil domain-containing protein 12**	Viable Normal	Nucleus/ Cytoplasm	Coiled coil	None	mRNA cis splicing, via spliceosome
cwf19	CWF19L2	**complexed with Cdc5 protein Cwf19**	Viable Normal	Unknown	HIT-like, CWF_J	None	mRNA cis splicing, via spliceosome
cwf20	PRCC	**Proline-rich protein PRCC**	Inviable	Nucleus	PRCC C-term	translocation with TFE3 expressed in papillary renal cell carcinoma	May regulate cell cycle progression through interaction with MAD2L2.
cwc21 cwf21	SRRM2	**Serine/arginine repetitive matrix protein 2**	Viable Normal	Nucleus Speckle	Coiled coil, RRM	None	Involved in pre-mRNA splicing
cwf22	CWC22	**Pre-mRNA-splicing factor CWC22 homolog**	Inviable	Nucleus Speckle	MI domain, MIF4G domain	None	Required for pre-mRNA splicing and for exon-junction complex (EJC) assembly. Role in nonsense mediated decay

cwf23	DNAJC17	**DnaJ homolog subfamily C member 17**	Inviable	Nucleus Nucleuolus	DNAJ domain, RRM domain	None	mRNA cis splicing, via spliceosome
cwf24	RNF113A	**RING finger protein 113A**	Inviable	Nucleus	RING type zinc finger, C3H1 type Zinc Finger	None	probably involved in mediating protein-protein interactions
cwf25	CWC25	**Pre-mRNA-splicing factor CWC25 homolog**	Inviable	Cytoplasm	Coiled coil	None	mRNA cis splicing, via spliceosome
bud13 cwf26	BUD13	**BUD13 homolog**	Inviable	Nucleus/ Cytoplasm	Coiled coil	None	mRNA cis splicing, via spliceosome
cyp7 cwf27	CWC27	**Peptidyl-prolyl cis-trans isomerase CWC27 homolog**	Viable Normal	Unknown	PPIase cyclophillin type	None	peptidyl-prolyl cis-trans isomerase activity
cwf28	GPKOW	**G patch domain and KOW motifs-containing protein**	Inviable	Nucleus	G Patch domain, KOW domain	None	The encoded protein interacts directly with protein kinase A and protein kinase X and is also found associated with the

							spliceosome
cwf29	RBMX2	**RBMX2 protein**	Viable Elongated	Nucleus	RRM domain	None	spliceosomal complex assembly
mug161	CWF19L1	**CWF19-like protein 1**	Viable Normal	Nucleus	HIT-like, CwfJ, MPP	None	mRNA cis splicing, via spliceosome

References

1. Finsterer, J. (2009). Ataxias with autosomal, X-chromosomal or maternal inheritance. Can. J. Neurol. Sci. J. Can. Sci. Neurol. *36*, 409–428.

2. Sailer, A., and Houlden, H. (2012). Recent advances in the genetics of cerebellar ataxias. Curr. Neurol. Neurosci. Rep. *12*, 227–236.

3. Yapici, Z., and Eraksoy, M. (2005). Non-progressive congenital ataxia with cerebellar hypoplasia in three families. Acta Paediatr. Oslo Nor. 1992 *94*, 248–253.

4. Lander, E.S., and Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science *236*, 1567–1570.

5. Alkuraya, F.S. (2013). The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. Hum. Genet. *132*, 1197–1211.

6. Foo, J.-N., Liu, J.-J., and Tan, E.-K. (2012). Whole-genome and whole-exome sequencing in neurological diseases. Nat. Rev. Neurol. *8*, 508–517.

7. Doi, H., Yoshida, K., Yasuda, T., Fukuda, M., Fukuda, Y., Morita, H., Ikeda, S., Kato, R., Tsurusaki, Y., Miyake, N., et al. (2011). Exome sequencing reveals a homozygous SYT14 mutation in adult-onset, autosomal-recessive spinocerebellar ataxia with psychomotor retardation. Am. J. Hum. Genet. *89*, 320–327.

8. Li, M., Pang, S.Y.Y., Song, Y., Kung, M.H.W., Ho, S.-L., and Sham, P.-C. (2013). Whole exome sequencing identifies a novel mutation in the transglutaminase 6 gene for spinocerebellar ataxia in a Chinese family. Clin. Genet. *83*, 269–273.

9. Lee, Y.-C., Durr, A., Majczenko, K., Huang, Y.-H., Liu, Y.-C., Lien, C.-C., Tsai, P.-C., Ichikawa, Y., Goto, J., Monin, M.-L., et al. (2012). Mutations in KCND3 cause spinocerebellar ataxia type 22. Ann. Neurol. *72*, 859–869.

10. Németh, A.H., Kwasniewska, A.C., Lise, S., Parolin Schnekenberg, R., Becker, E.B.E., Bera, K.D., Shanks, M.E., Gregory, L., Buck, D., Zameel Cader, M., et al. (2013). Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. Brain J. Neurol. *136*, 3106–3118.

11. Morino, H., Miyamoto, R., Ohnishi, S., Maruyama, H., and Kawakami, H. (2014). Exome sequencing reveals a novel TTC19 mutation in an autosomal recessive spinocerebellar ataxia patient. BMC Neurol. *14*, 5.

12. Doyle, A. (1990). Establishment of Lymphoblastoid Cell Lines. In Animal Cell Culture, J. Walker, J. Pollard, and J. Walker, eds. (Humana Press), pp. 43–47.

13. Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd, et al. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. *10*, R130.

14. Wood, V., Harris, M.A., McDowall, M.D., Rutherford, K., Vaughan, B.W., Staines, D.M., Aslett, M., Lock, A., Bähler, J., Kersey, P.J., et al. (2012). PomBase: a comprehensive online resource for fission yeast. Nucleic Acids Res. *40*, D695–699.

15. Magrane, M., and Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. Database J. Biol. Databases Curation *2011*, bar009.

16. Frank-Raue, K., Haag, C., Schulze, E., Keuser, R., Raue, F., Dralle, H., and Lorenz, K. (2011). CDC73-related hereditary hyperparathyroidism: five new mutations and the clinical spectrum. Eur. J. Endocrinol. Eur. Fed. Endocr. Soc. *165*, 477–483.

17. Kang, T.H., Guibinga, G.-H., Jinnah, H.A., and Friedmann, T. (2011). HPRT deficiency coordinately dysregulates canonical Wnt and presenilin-1 signaling: a neuro-developmental regulatory role for a housekeeping gene? PloS One *6*, e16572.

18. Chowdhury, U.K., Zakharyan, R.A., Hernandez, A., Avram, M.D., Kopplin, M.J., and Aposhian, H.V. (2006). Glutathione-S-transferase-omega [MMA(V) reductase] knockout mice: enzyme and arsenic species concentrations in tissues after arsenate administration. Toxicol. Appl. Pharmacol. *216*, 446–457.

19. Hsu, L.-I., Chen, W.-P., Yang, T.-Y., Chen, Y.-H., Lo, W.-C., Wang, Y.-H., Liao, Y.-T., Hsueh, Y.-M., Chiou, H.-Y., Wu, M.-M., et al. (2011). Genetic polymorphisms in glutathione S-transferase (GST) superfamily and risk of arsenic-induced urothelial carcinoma in residents of southwestern Taiwan. J. Biomed. Sci. *18*, 51.

20. Kim, K., Kim, S.-H., Kim, J., Kim, H., and Yim, J. (2012). Glutathione s-transferase omega 1 activity is sufficient to suppress neurodegeneration in a Drosophila model of Parkinson disease. J. Biol. Chem. *287*, 6628–6641.

21. Farhan, S.M.K., Wang, J., Robinson, J.F., Lahiry, P., Siu, V.M., Prasad, C., Kronick, J.B., Ramsay, D.A., Rupar, C.A., and Hegele, R.A. (2014). Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency. Mol. Genet. Genomic Med. *2*, 73–80.

22. Dor, T., Cinnamon, Y., Raymond, L., Shaag, A., Bouslam, N., Bouhouche, A., Gaussen, M., Meyer, V., Durr, A., Brice, A., et al. (2014). KIF1C mutations in two

families with hereditary spastic paraparesis and cerebellar dysfunction. J. Med. Genet. *51*, 137–142.

23. Miyamoto, R., Morino, H., Yoshizawa, A., Miyazaki, Y., Maruyama, H., Murakami, N., Fukada, K., Izumi, Y., Matsuura, S., Kaji, R., et al. (2014). Exome sequencing reveals a novel MRE11 mutation in a patient with progressive myoclonic ataxia. J. Neurol. Sci. *337*, 219–223.

24. NCBI Resource Coordinators (2014). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *42*, D7–17.

25. Ohi, M.D., Link, A.J., Ren, L., Jennings, J.L., McDonald, W.H., and Gould, K.L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol. Cell. Biol. *22*, 2011–2024.

26. Ren, L., McLean, J.R., Hazbun, T.R., Fields, S., Vander Kooi, C., Ohi, M.D., and Gould, K.L. (2011). Systematic two-hybrid and comparative proteomic analyses reveal novel yeast pre-mRNA splicing factors connected to Prp19. PloS One *6*, e16719.

27. Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res. *12*, 1231–1245.

28. Löscher, M., Fortschegger, K., Ritter, G., Wostry, M., Voglauer, R., Schmid, J.A., Watters, S., Rivett, A.J., Ajuh, P., Lamond, A.I., et al. (2005). Interaction of U-box E3 ligase SNEV with PSMB4, the beta7 subunit of the 20 S proteasome. Biochem. J. *388*, 593–603.

29. Brenner, C. (2002). Hint, Fhit, and GalT: function, structure, evolution, and mechanism of three branches of the histidine triad superfamily of nucleotide hydrolases and transferases. Biochemistry (Mosc.) *41*, 9003–9014.

30. Martin, J., St-Pierre, M.V., and Dufour, J.-F. (2011). Hit proteins, mitochondria and cancer. Biochim. Biophys. Acta *1807*, 626–632.

31. Hassan, M.I., Naiyer, A., and Ahmad, F. (2010). Fragile histidine triad protein: structure, function, and its association with tumorogenesis. J. Cancer Res. Clin. Oncol. *136*, 333–350.

32. Zhao, H., Race, V., Matthijs, G., De Jonghe, P., Robberecht, W., Lambrechts, D., and Van Damme, P. (2013). Exome sequencing reveals HINT1 mutations as a cause of distal hereditary motor neuropathy. Eur. J. Hum. Genet. EJHG.

33. Zimoń, M., Baets, J., Almeida-Souza, L., De Vriendt, E., Nikodinovic, J., Parman, Y., Battaloğlu, E., Matur, Z., Guergueltcheva, V., Tournev, I., et al. (2012). Loss-of-function mutations in HINT1 cause axonal neuropathy with neuromyotonia. Nat. Genet. *44*, 1080–1083.

34. Hirano, M., Yamamoto, A., Mori, T., Lan, L., Iwamoto, T., Aoki, M., Shimada, K., Furiya, Y., Kariya, S., Asai, H., et al. (2007). DNA single-strand break repair is impaired in aprataxin-related ataxia. Ann. Neurol. *61*, 162–174.

35. Kijas, A.W., Harris, J.L., Harris, J.M., and Lavin, M.F. (2006). Aprataxin forms a discrete branch in the HIT (histidine triad) superfamily of proteins with both DNA/RNA binding and nucleotide hydrolase activities. J. Biol. Chem. *281*, 13939–13948.

36. Moreira, M.C., Barbot, C., Tachi, N., Kozuka, N., Uchida, E., Gibson, T., Mendonça, P., Costa, M., Barros, J., Yanagisawa, T., et al. (2001). The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. Nat. Genet. *29*, 189–193.

37. Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., et al. (2013). CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. *41*, D348–352.

38. Marchler-Bauer, A., Panchenko, A.R., Shoemaker, B.A., Thiessen, P.A., Geer, L.Y., and Bryant, S.H. (2002). CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. *30*, 281–283.

39. Regal, J.A., Festerling, T.A., Buis, J.M., and Ferguson, D.O. (2013). Disease-associated MRE11 mutants impact ATM/ATR DNA damage signaling by distinct mechanisms. Hum. Mol. Genet. *22*, 5146–5159.

Chapter III.

Zebrafish animal model demonstrates decreased protein levels, and abnormal behavior and development

Introduction

A novel splice mutation in *CWF19L1* was found in affected Turkish siblings with congenital ataxia. The siblings also suffer from severe mental retardation and cerebellar hypoplasia (described in Chapter II)[92]. Exome sequencing in combination with homozygosity mapping identified a splice site mutation in the *CWF19L1* gene. RT-PCR and immunoblot demonstrated decreased mRNA level and loss of C19L1 protein in patient-specific LCLs suggesting loss of the protein causes the ataxia. These results clearly demonstrate absence of the protein in a linked and homozygous region. However, private null mutations exist in about 20% of humans[93], and a function of *CWF19L1* in ataxia is not immediately apparent, so we could not exclude the possibility that this could be a rare Turkish variant without pathological implications. One strategy to determine the function of a gene is to decrease or abolish expression of the gene to determine the effects of reduction or loss of the gene product.

One method to accomplish this is to utilize morpholino-mediated gene knockdown in zebrafish. Morpholinos, which are small antisense oligonucleotides, can inhibit zygotic transcripts and can target specific

sequences, creating a simple and effective method for studying mutations[94]. Morpholinos can be ubiquitously introduced into the developing zebrafish by injection into 1-2 cell embryos, and they result in gene knockdown that lasts a minimum of 4-7 days. Rapid generation time, large offspring number (100-300 per clutch), *ex utero* development, and easily quantifiable motor behaviors make zebrafish an advantageous animal model to investigate human neurological disorders[39,42,95].

Recently, zebrafish animal models have been utilized to test the effects of genes implicated in ataxia (cite some of the papers here). Additionally, zebrafish models can be used to identify the roles of newly discovered genes in development and other cellular processes so that we truly understand the role of a gene in normal brain function and in the ataxia disorders[96–100]. FANS, a mutation prediction program, predicts the *CWF19L1* variant we have identified to eliminate a splice site and to be highly damaging[101]. Splicing mutations have been studied in zebrafish using morpholino-mediated knockdown, in a variety of disorders, including primary microcephaly[100], collagen myopathy[14,15], and ataxia[96,97,102]. Many splice site morpholinos have been shown to cause a locomotor phenotype in zebrafish which can often be shown using multiple morpholinos to the same exon or rescued using the human RNA for the gene[96,97,99]. These manipulations establish a causal relationship between the splicing mutation and disease phenotypes in humans.

The zebrafish *cwf19l1* gene has 14 exons and shares 67% coding DNA identity with the human *CWF19L1* gene[67,103,104]. The *cwf19l1* gene is expressed ubiquitously in zebrafish, with higher expression in blood and head[104,105] but currently the function and effects of mutations in this gene are unknown. Here, we use morpholino-mediated knockdown of the zebrafish *cwf19l1* gene in combination with touch-evoked escape response and immunostaining to determine if severe deficiency or loss of *cwf19l1* affects behavior and development of the fish as a model of the human disorder.

Materials and Methods

Zebrafish experimentation

Wild type (AB strain) adult zebrafish were maintained in accordance with IACUC approved standards. Adult fish were mated to generate embryos for all subsequent analyses.

Morpholino experiments

A morpholino was designed to the exon 9 splice acceptor and donor sites of the zebrafish *cwf19l1* gene (Figure 3.2). All studies were done as a comparison between the *cwf19l1* morpholino and a standardized control morpholino (Gene Tools, Inc, Philomath, OR). The sequence of the *cwf19l1* exon 9 splice acceptor morpholino is TGC TGG TTC TTC CTG ATC AAA GAG A and the sequence of the *cwf19l1* exon 9 splice donor morpholino is AGA GTG CAT GTG AAT GGA CTC ACG T. The standard control morpholino sequence is CCT CTT ACC TCA

GTT ACA ATT TAT A. Morpholinos were injected into 1-2 cell stage embryos. Increasing concentrations were injected and screened by RT-PCR to determine efficacy in terms of interrupting splicing. The minimal dose for effect was 0.15 mM for the splice acceptor (I8E9) and 0.45 mM for the splice donor (E9I9) therefore these concentrations were used for all subsequent experimentation.

In silico **analysis**

All sequences were obtained from the NCBI database[76]:

mRNA: *H. sapiens* (*CWF19L1*): NM_018294.4,

D. rerio (cwf19l1):NM_001044758.1

C19L1 Protein: *H. sapiens* (C19L1): NP_060764.3,

D. rerio (c19l1): NP_001038223.1

Multiple Alignment using Fast Fourier Transform (MAFFT)[106] (http://mafft.cbrc.jp/alignment/software/) website was used to align C19L1 sequences of human and zebrafish. Sequences used for MAFFT program are listed above.

RNA extraction

RNA was extracted from a minimum of ten fish using TRIzol reagent (Life Technologies, Grand Island, NY) and Zymogen Directzol RNA MiniPrep kit

(Zymo Research Corp, Irvine, CA) according to the manufacturer's instructions. RNA was subjected to DNase I treatment (Ambion, Grand Island, NY) and reverse transcribed using the Invitrogen (now Life Technologies, Grand island, NY) SuperScript II reverse transcription kit using Oligo dTs and random hexamers. RT-PCR included primers that spanned exons 6 to 10 (F: 5'- TCG CAG ATC TCG CTG ACA AGC-3'; R: 5'- GCA AGG CAG AAC CAA CAG GGT-3') and primers to GAPDH (F: 5'- GTG TAG GCG TGG ACT GTG GT-3'; R: 5'- TGG GAG TCA ACC AGG ACA AAT A-3') for assessment of RNA.

Protein electrophoresis and immunoblotting

Described in Chapter II with the following changes: a minimum of 20 fish were collected from each condition at 72hpf (ctrl, I8E9 morphant, E9I9 morphant) and lysate was prepared in SDS sample buffer (1M Tris-HCL, pH 6.8, Glycerol, β-mercaptoethanol, 20% SDS, and H_2O) containing HALT protease inhibitor mixture (Thermo Scientific). Equal amounts of protein (15 µg per lane) were electrophoresed on SDS-PAGE gels. For c19l1, we used primary antibody HPA036890 (Sigma, St. Louis, MO) at a dilution of 1:2000; the secondary antibody was HRP-conjugated antibody against rabbit IgG diluted 1:10,000 (BioRad, Hercules, CA, Cat. No. 170-6515). Blots were stripped with Restore Western Blot Stripping Buffer (Thermo Scientific Cat No. 21059) and reprobed with primary antibody GAPDH (Abcam, Cambridge, MA) at a dilution of 1:5,000 using the same secondary antibody and ECL Western Blotting Substrate (BioRad, Hercules, CA).

Touch-evoked escape response

Escape response was measured in morphant zebrafish at 72 hpf. It was elicited by touching embryo tails with an eyelash filament. Response was graded from 0 to 3; 0= no response to repeated stimulation; 1= flickers of movement but no swimming; 2= abnormal swimming in response to touch; 3= normal escape response[107,108]. Time-lapse video microscopy was used to record results. Fish were scored by three blinded investigators and scores averaged. A minimum of 50 embryos per condition was utilized for behavioral assays. All experiments were performed 2-3 times.

Whole-mount immunostaining

Zebrafish larvae were fixed at 25°C in 4% PFA in PBST (PBS, 0.1% Triton X-100) for 3 h. The fixed larvae were washed with PBST, and incubated in acetone at − 20°C for 15 min. Larvae were washed once with PBST and twice with PBS-DT (PBS, 1% BSA, 1% DMSO, 1% Triton X-100), and incubated in 5% goat serum (Life Technologies, Grand Island, NY), PBS-DT at room temperature for 1 h. The samples were incubated with the primary antibody (Anti-zebrin II (1:200, hybridoma supernatant)[109,110], anti-Pvalb7 (1:1000, mouse ascites), anti-Vglut1 (1:1000, purified antibody)) solution at 4°C overnight. After four washes with PBST, the samples were incubated with secondary antibodies (1:1000 dilution, Alexa Fluor 488 goat anti-mouse or Alexa Fluor 555 goat anti-rabbit, Molecular Probes (Invitrogen, Grand Island, NY)) and stained with DAPI (Thermo Fisher,

Rockford, IL). Fish were mounted in 1% agarose and viewed on an Olympus Fluoview FV1000 confocal microscope. Confocal images were obtained using a 10X objective. Images were acquired with Olympus FV1000 software and analyzed using ImageJ.

For each condition, two investigators, who were blind to the injection cocktail, scored 25 embryos from 3 different experiments. Statistical significance was calculated using a Chi-squared test.

Results

Morpholino-mediated knockdown of *cwf19l1* produces abnormal motor behavior and cerebellar development in zebrafish

To test the effect of the *CWF19L1* splice site mutation *in vivo*, we used a morpholino knock-down strategy (Figure 3.1) to the homologous intron-exon junction in zebrafish. Zebrafish *cwf19l1* has 14 exons, shares 65% protein identity with the human *CWF19L1* gene (Figure 3.2)[67,104] and has the same predicted intron-exon structure.

We designed morpholinos to the exon 9 splice acceptor and donor site (*cwf19l1* MOs) to target excision of exon 9, as in the affected individuals (schematic Figure 3.1). MOs were injected into 1-2 cell stage embryos and RNA was extracted at 3 days post fertilization (dpf). RT-PCR, using primers to exon 6 and exon 10, revealed the expected band of approximately 450bp in fish injected with

a standard control morpholino. However, RT-PCR from morphant fish injected with the *cwf19l1* MOs demonstrated knockdown of the RT-PCR product with increasing morpholino concentration (Figure 3.3). Additionally, western blots performed with a C19L1 antiserum (HPA036890) demonstrated knockdown and loss of the c19l1 protein with increasing MO concentration (Figure 3.4), indicating deficiency, and loss at higher doses, of c19l1 in morphant fish.

The gross morphology of the *cwf19l1* morphant fish was normal overall. Images of the fish show no significant differences in size or morphology (Figure 3.5) at 0.15 mM I8E9 and 0.45 mM E9I9.

Behavioral analysis of *cwf19l1* morphant embryos (as compared to control morphants) revealed reduced swim speed and abnormal touch-evoked escape response, a stereotyped behavior that is prominent at 3 dpf (see Figure 3.6 and supplemental videos (available upon request). In contrast to their control morphant counterparts, *cwf19l1* morphant fish, when touched with the eyelash filament, moved slowly and made circular movements instead of rapid movement straight away from the stimulus. As this response was elicited with both *cwf19l1* morpholinos, this suggests knockdown of *cwf19l1* causes the abnormal motor behavior in these fish. Because the E9I9 MO required a higher dose to show morpholino effect, touch-evoked escape response was only quantified for comparison of 0.15 mM I8E9 and 0.45 mM E9I9 morphant fish. Doses above 0.5 mM are associated with an increase in mortality rate in control MO injected fish

and potential off target effects of the individual morpholinos[94,111], therefore doses above 0.5 mM were not used for the E9I9 morpholino.

Because the affected individuals showed cerebellar hypoplasia, we tested whether knockdown of *cwf19l1* in morphant fish altered cerebellar development. Immunostaining with the zebrin II monoclonal antibody, which labels aldolase C positive cells [109,110], demonstrated altered/diminished cerebellar staining in I8E9 morphants that worsened with increasing MO concentration (Figure 3.7). This suggests a defect in cerebellar structure. These results corroborate our earlier observation that c19l1 protein deficiency is associated with a neurological movement disorder and with cerebellar abnormalities.

Discussion

While the function of the cwf19l1 gene is unknown, its conservation from yeast to humans indicates an evolutionary importance. One strategy to determine the functional importance of a gene in an organism is to decrease or abolish expression of the gene to determine the effects of reduction or loss of the gene product. Zebrafish have recently been used to successfully study mutations in ataxia genes [96–99,102,112]. In our study, knockdown of *cwf19l1* in fish caused abnormal motor behavior and alteration of cerebellar structure. These data provide *in vivo* validation of our whole-exome sequencing results, demonstrating that loss of *cwf19l1* causes an abnormal motor phenotype.

Specificity of the knockdown of the *cwf19l1* gene was demonstrated by reproducible phenotypes obtained by two different morpholinos to the same exon. While the splice donor morpholino (E9I9) is predicted to cause direct splicing out of exon 9, the splice acceptor morphoino (I8E9) is predicted to cause inclusion of intron 8. However, since intron 8 is not divisible by 3 (as described in chapter II), inclusion of intron 8 is still expected to lead to nonsense-mediated decay of the RNA transcript, which is demonstrated by knockdown on RT-PCR. The specificity of knockdown is further confirmed by western blotting with a commercial antiserum (HPA036890) to the C-terminal end of c19l1. Decreased detection of c19l1 in the morphant fish demonstrates specificity of this antiserum.

Additionally, examination of brain, and specifically cerebellar development, revealed decreased staining with zebrin II antibody, suggesting that the knockdown of *cwf19l1* affected cerebellar development in the fish. Many forms of ataxia in human patients have a cerebellar structural defect as one of the clinical findings. Recent studies of ataxia genes that, when mutated, lead to cerebellar abnormalities in humans, have found evidence of abnormal cerebellar development in the zebrafish animal model [97,98,102]. These studies, in addition to our data, suggest conservation of genes that play a functional role in the development of the cerebellum between fish and humans, even if the exact morphology and primary cellular function are different.

While zebrin II staining has been used to demonstrate cerebellar abnormlities in zebrafish models of ataxia, we cannot differentiate between specific knockdown of cells that express aldolase C and abnormal cerebellum development. As such, I immunostained 5dpf fish with anti-Pvalb7 and anti-Vglut1 antibodies obtained from the Masahiko group in Japan[50]. Pvalb7 specifically labels the Purkinje cell layer and Vglut1 specifically labels the granule cell layer[50]. Unfortunately, staining with these antibodies according to previous published protocol revealed non-specific positive staining in all brain regions of control and morphant zebrafish. Future studies using the α1-T-GFP transgenic zebrafish of the Goldman lab group at the University of Michigan[113,114], which labels all α-tubulin positive neurons in the fish with GFP, could be performed to analyze general neuronal development, including cerebellar development, without requiring immunostaining.

There are currently no therapies for ataxia, therefore determining new pathways or targets for therapy are important. Recently, mouse models for spinocerebellar ataxia 1 and Friedreich's ataxia have led to potential drug targets and clinical trials [13,14]. Zebrafish are currently being used in studies for treatment of neurological disorders including bipolar disorder[115] and epilepsy[116,117]. Since development and behavior can be easily visualized and manipulated in zebrafish larvae they represent a promising model for studying the effects of loss of *cwf19l1* and for high throughput screening of therapeutic compounds to treat this recessive ataxia. Therefore in the future, this zebrafish model may also be used

to understand the function of C19L1, analyze and confirm cell experiments and ultimately test potential therapies for this form of ataxia.

Acknowledgements

Technical assistance for zebrafish injections, staining, mounting and imaging were provided by Angela Busta, Dr. Aaron Reifler, and Fairouz Elsaeidi.

We thank Richard Hawkes, University of Calgary, Canada for the anti-zebrin II antiserum, James Dowling, University of Toronto (previously University of Michigan) for the wild-type AB zebrafish. We also thank Masahiko Hibi, Nagoya University, Nagoya, Japan and RIKEN Center for Developmental Biology, Kobe, Hyogo, Japan for the Pvalb7 and Vglut1 antibodies.

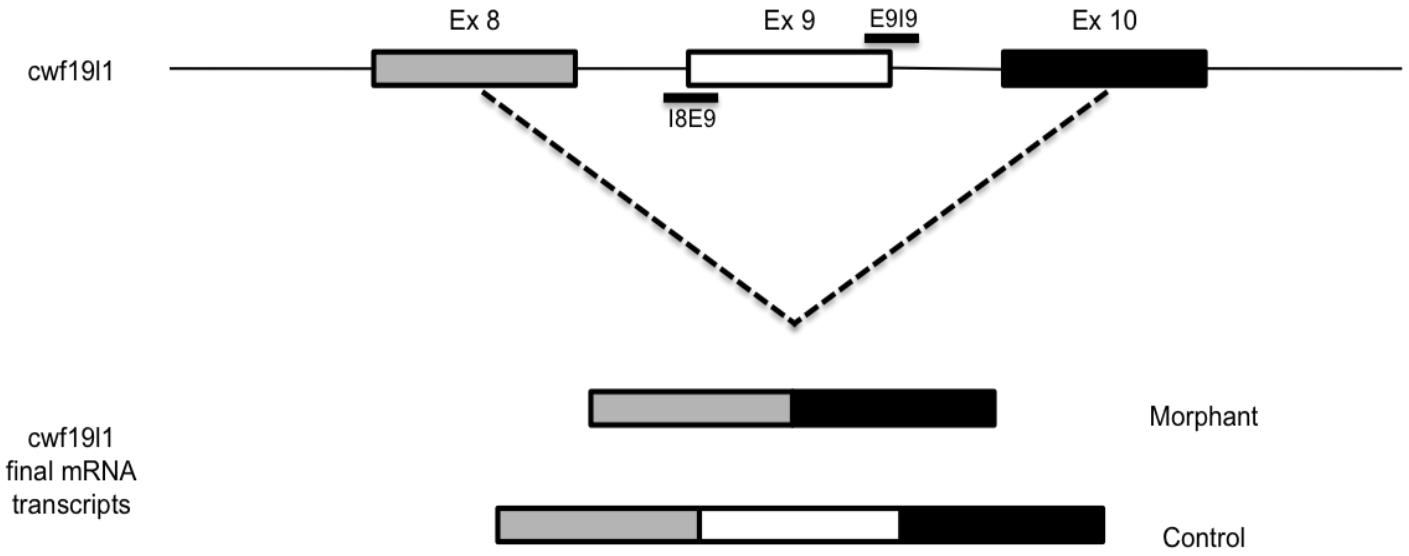


Figure 3.1. Morpholino-mediated knockdown design

Splice targeting morpholinos were designed to the intron 8/exon 9 (I8E9) splice acceptor site and to the exon 9 /intron 9 (E9I9) splice donor site.

Figure 3.2. Sequence similarity between human and zebrafish C19L1

Amino acid alignment of the coding region for human (*H. sapiens*) and zebrafish (*D. rerio*) C19L1. * (asterisk) indicates conserved amino acids, . (period) indicates conservation between groups of weakly similar properties, : (colon) indicates conservation between groups of strongly similar properties

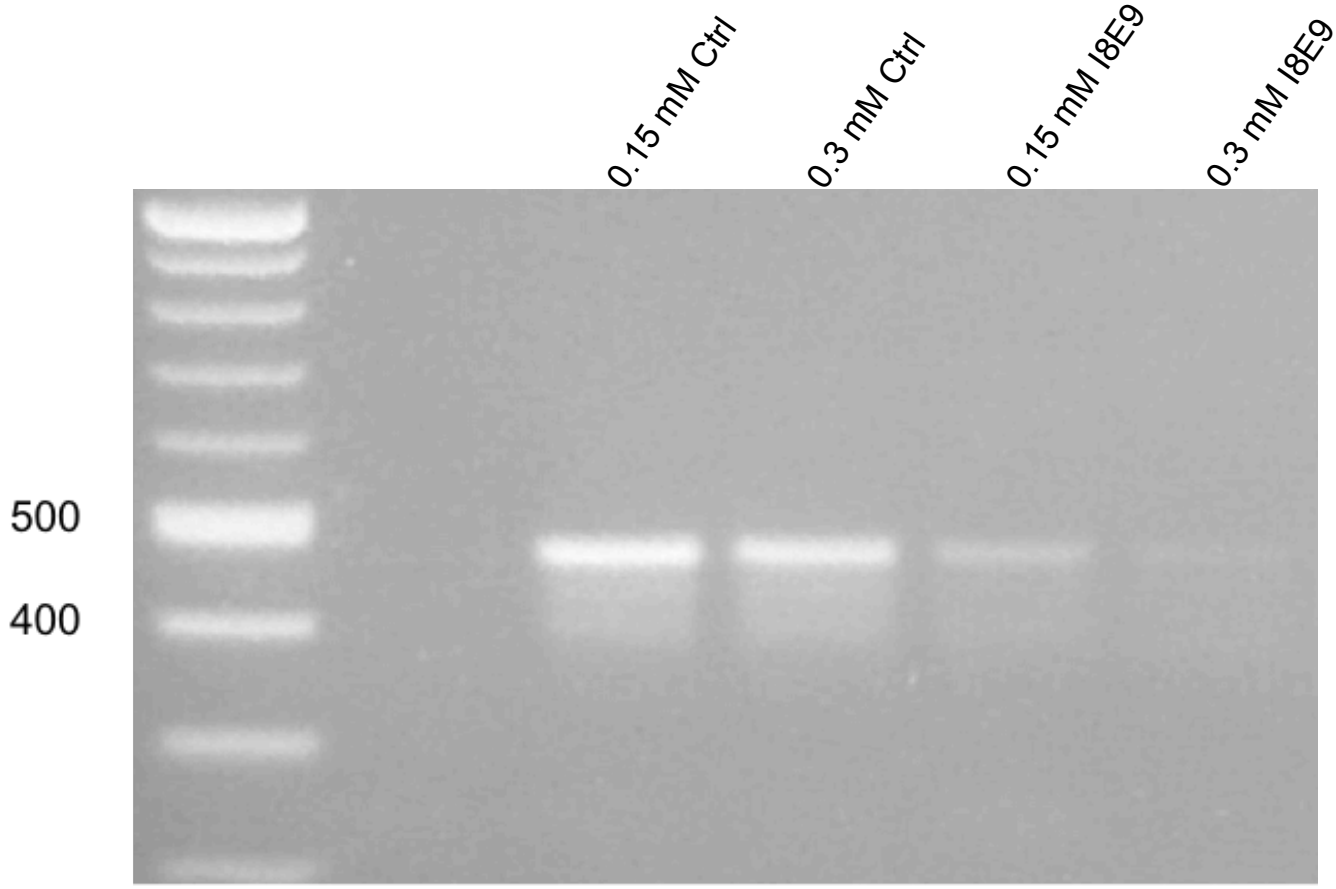


Figure 3.3. RT-PCR shows knockdown of cwf19l1 mRNA in morphant zebrafish

RT-PCR of zebrafish cwf19l1 exons 6-10 indicate a loss of *cwf19l1* RNA in morphants (0.15 mM I8E9 and 0.3 mM I8E9; lanes 4-5) with increasing morpholino doses as opposed to normal product in uninjected and control MO injected zebrafish (lanes 2-3). Lane 1: no DNA control

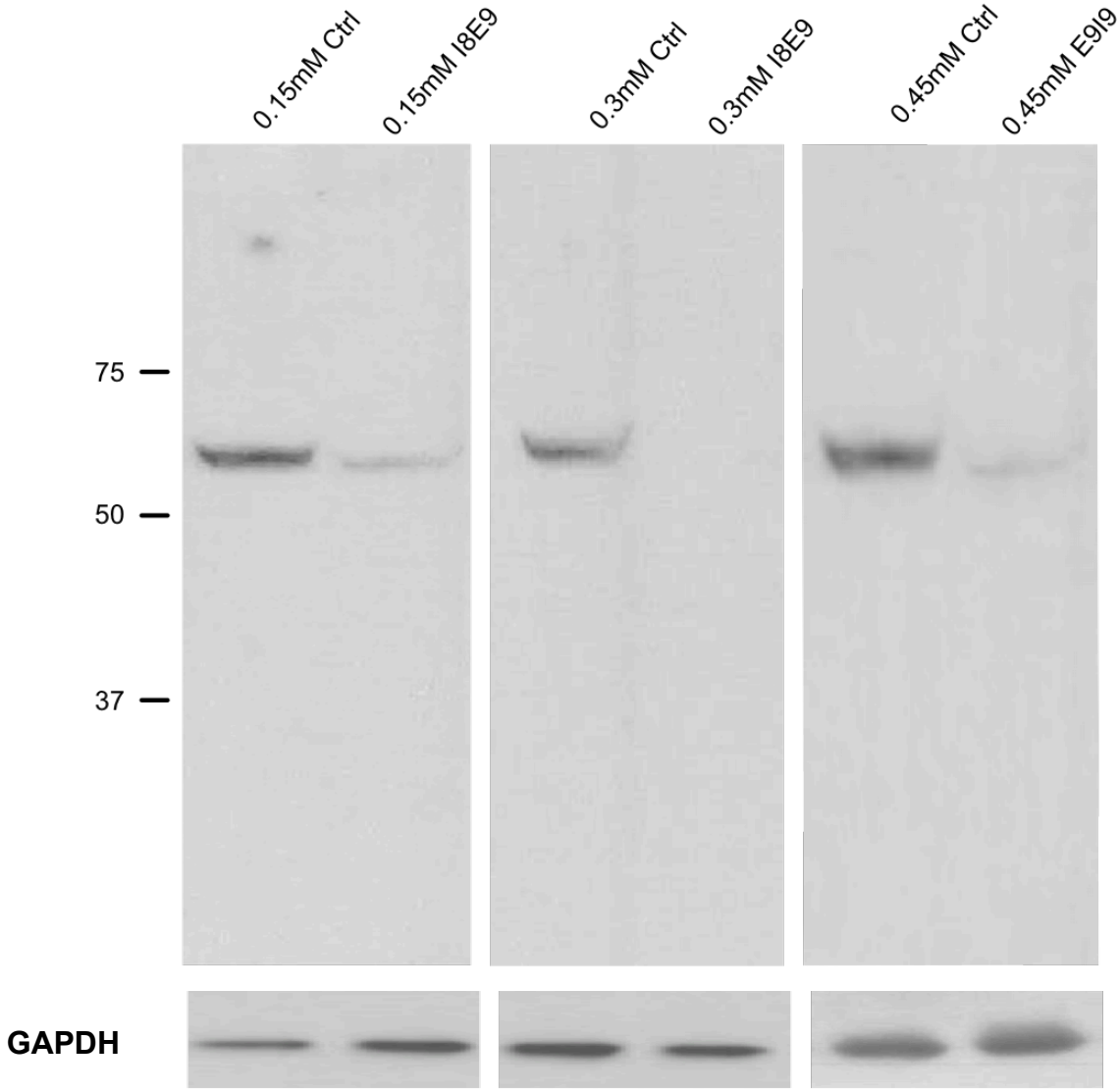


Figure 3.4. Western blot shows knockdown and loss of *cwf19l1* with increasing MO dose in cwf19l1 morphant fish

Western blot with antiserum that recognizes 3' end of c19l1 protein demonstrates loss of protein in morphant fish with increasing MO dose. GAPDH loading control

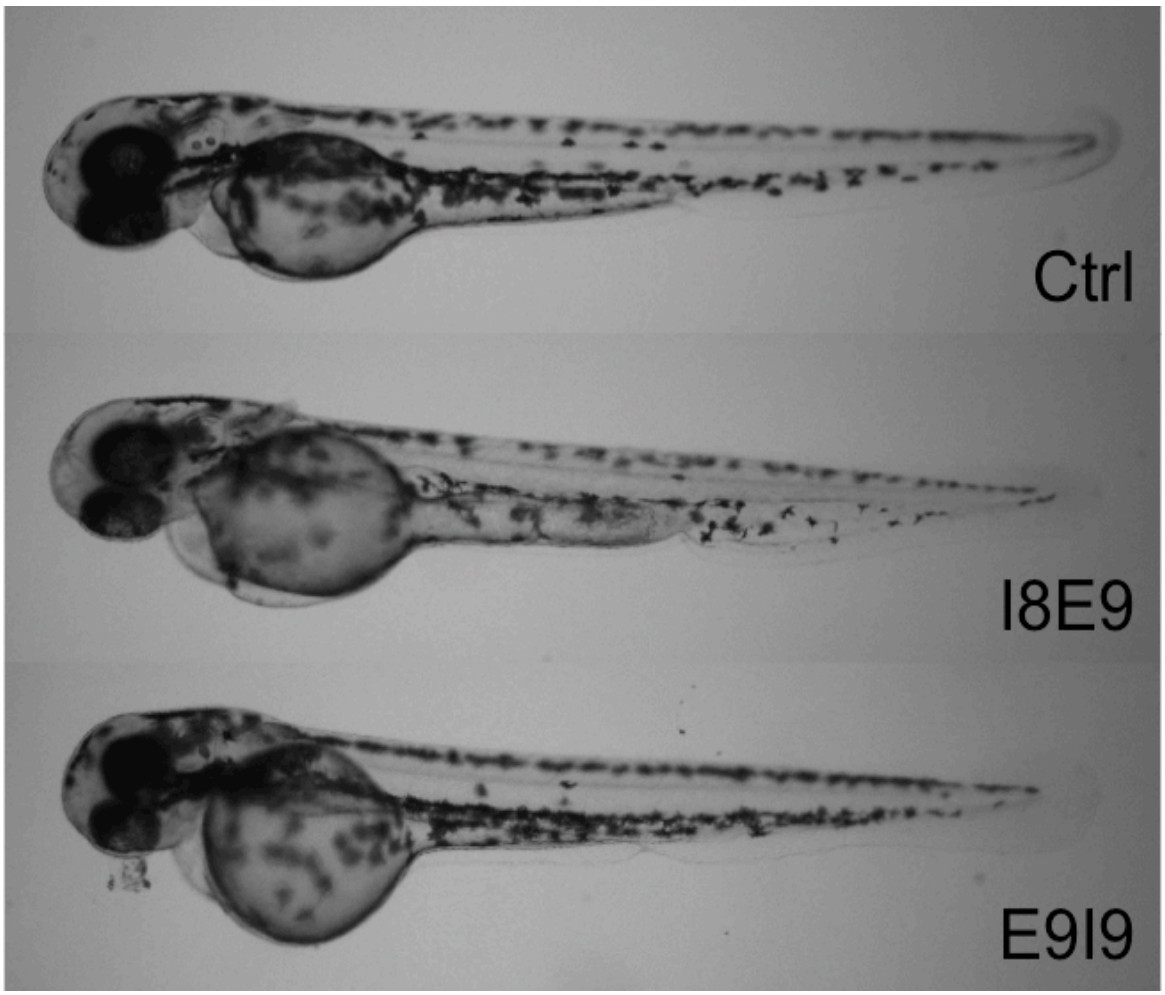


Figure 3.5. Gross morphology shows normal development in morphant fish

Size and shape of Ctrl, I8E9, and E9I9 morphant fish are comparable in size and morphology

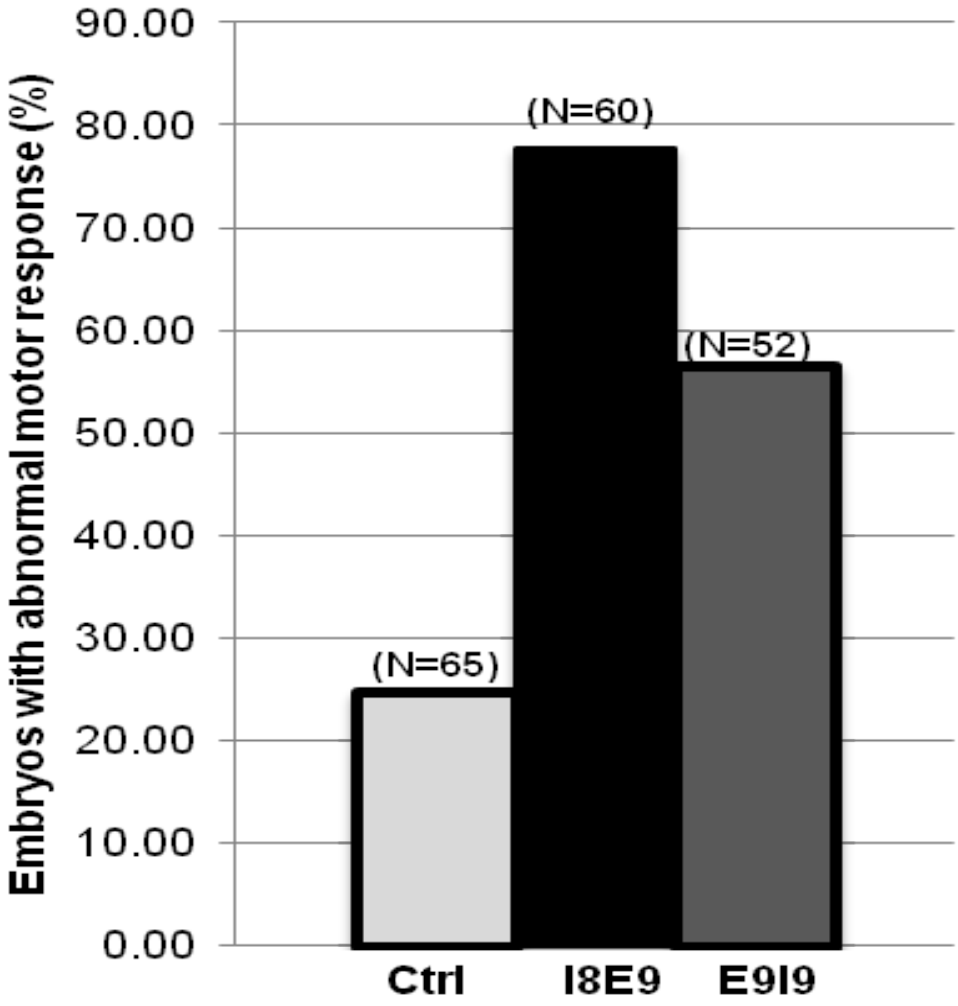
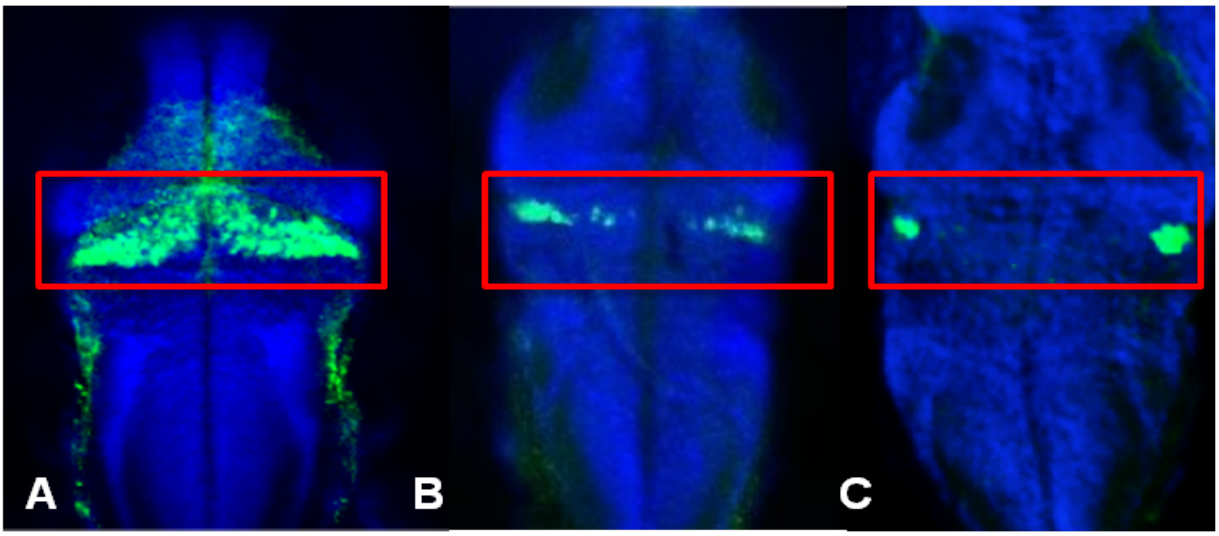


Fig 3.6. Touch-evoked escape response shows abnormal motor behavior in morphant fish

cwf19l1 morphants had abnormal motor behavior. Depicted is a graph of the touch-evoked escape response (see also Movies S1- S4) (n = 5 independently injected clutches, and a minimum of 10 embryos were assessed per clutch).



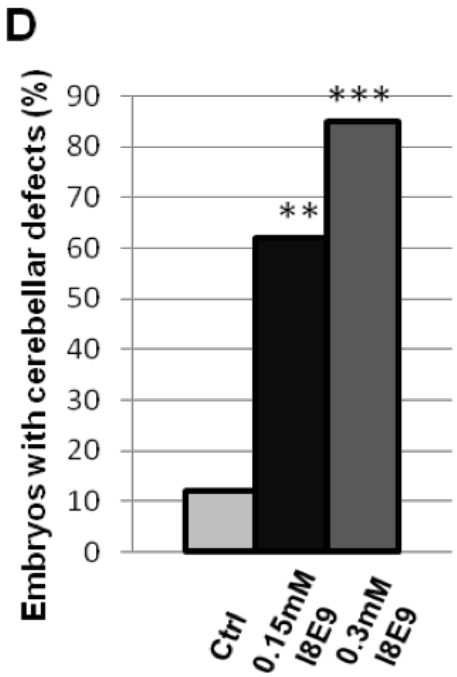


Figure 3.7. Zebrin II staining shows abnormal staining in cwf19l1 morphant fish

Zebrin II staining in 5dpf (**A**) control MO fish, (**B**) 0.15 mM I8E9 MO fish and (**C**) 0.3 mM I8E9 MO fish demonstrating gradually diminished staining in cerebellum with increasing MO dose (**D**) quantification (n=20 fish per condition, ** $p<0.01$, *** $p<0.001$).

References

1. Yapici, Z., and Eraksoy, M. (2005). Non-progressive congenital ataxia with cerebellar hypoplasia in three families. Acta Paediatr. Oslo Nor. 1992 *94*, 248–253.

2. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science *335*, 823–828.

3. Bill, B.R., Petzold, A.M., Clark, K.J., Schimmenti, L.A., and Ekker, S.C. (2009). A primer for morpholino use in zebrafish. Zebrafish *6*, 69–77.

4. Becker, T.S., and Rinkwitz, S. (2012). Zebrafish as a genomics model for human neurological and polygenic disorders. Dev. Neurobiol. *72*, 415–428.

5. Kabashi, E., Brustein, E., Champagne, N., and Drapeau, P. (2011). Zebrafish models for the functional genomics of neurogenetic disorders. Biochim. Biophys. Acta *1812*, 335–345.

6. Rinkwitz, S., Mourrain, P., and Becker, T.S. (2011). Zebrafish: an integrative system for neurogenomics and neurosciences. Prog. Neurobiol. *93*, 231–243.

7. Mahmood, F., Mozere, M., Zdebik, A.A., Stanescu, H.C., Tobin, J., Beales, P.L., Kleta, R., Bockenhauer, D., and Russell, C. (2013). Generation and validation of a zebrafish model of EAST (epilepsy, ataxia, sensorineural deafness and tubulopathy) syndrome. Dis. Model. Mech. *6*, 652–660.

8. Aspatwar, A., Tolvanen, M.E.E., Jokitalo, E., Parikka, M., Ortutay, C., Harjula, S.-K.E., Rämet, M., Vihinen, M., and Parkkila, S. (2013). Abnormal cerebellar development and ataxia in CARP VIII morphant zebrafish. Hum. Mol. Genet. *22*, 417–432.

9. Yanicostas, C., Barbieri, E., Hibi, M., Brice, A., Stevanin, G., and Soussi-Yanicostas, N. (2012). Requirement for zebrafish ataxin-7 in differentiation of photoreceptors and cerebellar neurons. PloS One *7*, e50705.

10. Issa, F.A., Mazzochi, C., Mock, A.F., and Papazian, D.M. (2011). Spinocerebellar ataxia type 13 mutant potassium channel alters neuronal excitability and causes locomotor deficits in zebrafish. J. Neurosci. Off. J. Soc. Neurosci. *31*, 6831–6841.

11. Kim, H.-T., Lee, M.-S., Choi, J.-H., Jung, J.-Y., Ahn, D.-G., Yeo, S.-Y., Choi, D.-K., and Kim, C.-H. (2011). The microcephaly gene aspm is involved in brain development in zebrafish. Biochem. Biophys. Res. Commun. *409*, 640–644.

12. Liu, C.-K., Chen, Y.-H., Tang, C.-Y., Chang, S.-C., Lin, Y.-J., Tsai, M.-F., Chen, Y.-T., and Yao, A. (2008). Functional analysis of novel SNPs and mutations in human and mouse genomes. BMC Bioinformatics *9 Suppl 12*, S10.

13. Margolin, D.H., Kousi, M., Chan, Y.-M., Lim, E.T., Schmahmann, J.D., Hadjivassiliou, M., Hall, J.E., Adam, I., Dwyer, A., Plummer, L., et al. (2013). Ataxia, dementia, and hypogonadotropism caused by disordered ubiquitination. N. Engl. J. Med. *368*, 1992–2003.

14. Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., et al. (2011). The UCSC Genome Browser database: update 2011. Nucleic Acids Res. *39*, D876–882.

15. Ramachandran, S., Ruef, B., Pich, C., and Sprague, J. (2010). Exploring zebrafish genomic, functional and phenotypic data using ZFIN. Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al *Chapter 1*, Unit 1.18.

16. Magrane, M., and Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. Database J. Biol. Databases Curation *2011*, bar009.

17. Thisse, B., Heyer, V., Lux, A., Alunni, V., Degrave, A., Seiliez, I., Kirchner, J., Parkhill, J.-P., and Thisse, C. (2004). Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. Methods Cell Biol. *77*, 505–519.

18. NCBI Resource Coordinators (2014). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *42*, D7–17.

19. Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. *9*, 286–298.

20. Telfer, W.R., Busta, A.S., Bonnemann, C.G., Feldman, E.L., and Dowling, J.J. (2010). Zebrafish models of collagen VI-related myopathies. Hum. Mol. Genet. *19*, 2433–2444.

21. Dowling, J.J., Low, S.E., Busta, A.S., and Feldman, E.L. (2010). Zebrafish MTMR14 is required for excitation-contraction coupling, developmental motor function and the regulation of autophagy. Hum. Mol. Genet. *19*, 2668–2681.

22. Lannoo, M.J., Brochu, G., Maler, L., and Hawkes, R. (1991). Zebrin II immunoreactivity in the rat and in the weakly electric teleost Eigenmannia (gymnotiformes) reveals three modes of Purkinje cell development. J. Comp. Neurol. *310*, 215–233.

23. Lannoo, M.J., Ross, L., Maler, L., and Hawkes, R. (1991). Development of the cerebellum and its extracerebellar Purkinje cell projection in teleost fishes as determined by zebrin II immunocytochemistry. Prog. Neurobiol. *37*, 329–363.

24. Eisen, J.S., and Smith, J.C. (2008). Controlling morpholino experiments: don't stop making antisense. Dev. Camb. Engl. *135*, 1735–1743.

25. Valdmanis, P.N., Dupré, N., Lachance, M., Stochmanski, S.J., Belzil, V.V., Dion, P.A., Thiffault, I., Brais, B., Weston, L., Saint-Amant, L., et al. (2011). A mutation in the RNF170 gene causes autosomal dominant sensory ataxia. Brain J. Neurol. *134*, 602–607.

26. Bae, Y.-K., Kani, S., Shimizu, T., Tanabe, K., Nojima, H., Kimura, Y., Higashijima, S., and Hibi, M. (2009). Anatomy of zebrafish cerebellum and screen for mutations affecting its development. Dev. Biol. *330*, 406–426.

27. Gulati-Leekha, A., and Goldman, D. (2006). A reporter-assisted mutagenesis screen using alpha 1-tubulin-GFP transgenic zebrafish uncovers missteps during neuronal development and axonogenesis. Dev. Biol. *296*, 29–47.

28. Ramachandran, R., Reifler, A., Parent, J.M., and Goldman, D. (2010). Conditional gene expression and lineage tracing of tuba1a expressing cells during zebrafish development and retina regeneration. J. Comp. Neurol. *518*, 4196–4212.

29. Watase, K., Gatchel, J.R., Sun, Y., Emamian, E., Atkinson, R., Richman, R., Mizusawa, H., Orr, H.T., Shaw, C., and Zoghbi, H.Y. (2007). Lithium therapy improves neurological function and hippocampal dendritic arborization in a spinocerebellar ataxia type 1 mouse model. PLoS Med. *4*, e182.

30. Rai, M., Soragni, E., Jenssen, K., Burnett, R., Herman, D., Coppola, G., Geschwind, D.H., Gottesfeld, J.M., and Pandolfo, M. (2008). HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model. PloS One *3*, e1958.

31. Ellis, L.D., and Soanes, K.H. (2012). A larval zebrafish model of bipolar disorder as a screening platform for neuro-therapeutics. Behav. Brain Res. *233*, 450–457.

32. Rahn, J.J., Bestman, J.E., Josey, B.J., Inks, E.S., Stackley, K.D., Rogers, C.E., Chou, C.J., and Chan, S.S.L. (2014). Novel Vitamin K analogs suppress seizures in zebrafish and mouse models of epilepsy. Neuroscience *259*, 142–154.

33. Baraban, S.C., Dinday, M.T., and Hortopan, G.A. (2013). Drug screening in Scn1a zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. Nat. Commun. *4*, 2410.

Chapter IV.

Tissue distribution and location of C19L1, the CWF19L1-encoded protein – preliminary data

Introduction

We have demonstrated that the ataxia/retardation syndrome in a Turkish sibship is associated with a splice site mutation that leads to splicing defects and loss of the c19l1 protein encoded by the *CWF19L1* gene (Chapter II). While our genetic and molecular sequencing data were consistent with this mutation causing the syndrome, there are some deleterious mutations in every personal genome so we acquired independent evidence to demonstrate that our mutation causes ataxia[1,2]. To link *CWF19L1* better to the ataxic phenotype, we generated a zebrafish model. We showed that reducing the *cwf19l1* mRNA and C19L1 protein in zebrafish causes abnormal motor behavior and cerebellar development (Chapter III). However, while these results convincingly link the loss of this protein to ataxia, our experiments have not given us any information on the function of this gene at the cellular level. Much of the current information on *CWF19L1* is based on predictions in databases. BioGPS, a database of gene expression studies, suggests this protein is ubiquitously expressed (Figure 4.1)[3]. While these databases give us a starting point for functional implications of this

gene, many of these predictions are based on high throughput experiments that have not been verified. One step to begin to characterize function is to determine tissue distribution and location in the cell[4–6]. Here we set out to determine the subcellular localization and tissue-specific expression pattern of *CWF19L1*.

Materials and Methods

Commercial lysates reagents

Commercial tissue lysates for liver, stomach, cerebellum (left and right), spleen, heart, whole brain, frontal lobe, kidney, skeletal muscle, and lung were obtained from ProSci Incorporated (Poway, CA). HeLa cell line cytoplasmic lysates were also obtained from ProSci Incorporated. HeLa cell line nuclear lysates were obtained from Santa Cruz Biotechnology (Dallas, TX) and Abcam (Cambridge, MA). Commercial normal tissue brain lysates blot (15mg protein/lane; Cat. No. 1526) and normal tissue lysates blot (15mg protein/lane, Cat. No. 1521) were obtained from ProSci Incorporated (Poway, CA).

Tissue lysate preparation

All protein extractions were performed on ice with pre-chilled reagents and materials. Mouse tissues (Cerebellum, brain without cerebellum, lungs, heart, liver, stomach, kidney, and spleen) were homogenized in a glass dounce homogenizer in 1-2 mL of homogenization buffer (50mM Tris, 500mM NaCl, 1mM EDTA, 0.5% NP40, 1x Protease Inhibitor Cocktail (Thermo Scientific), 1x PMSF, 0.1% SDS or 2% SDS). Homogenization was performed for 1-2 minutes

(approximately 50 strokes). Homogenate was centrifuged at 4oC in a pre-cooled microcentrifuge at 15,000 x g for 10-15 minutes to clear large debris. Supernatant was used immediately or flash-frozen with liquid nitrogen and stored at -80oC. Protein concentrations were measured using a commercial Bradford Assay (Bio-Rad, Cat. No. 500-0006).

Protein electrophoresis and immunoblotting

Described in Chapter II, Materials and Methods

Immunofluorescence

HeLa cells and human motor neuron (MN1) cells were grown and maintained in complete growth medium (1:1 mixture of DMEM and F12 medium, with 10% FBS) in 5% CO_2 at 37°C.

Cells were fixed in 4% PFA diluted in 1x PBS at room temperature for 1 hour. Cells were washed with TBS + TX100 for 10 minutes at room temperature, and then blocked with TBST-S for 30 minutes at room temperature. Primary antibodies were diluted in TBST-S at the following dilutions: anti-c19l1 antiserum (HPA036890, Sigma, St. Louis, MO) at 1:250 and anti- b-tubulin at 1:500, and incubated overnight at 4°C. Cells were washed 5x for 5 minutes each in TBST. Secondary antibodies and DAPI were diluted in TBST-S at the following dilutions: Alexa Fluor 555 (goat anti-rabbit, Molecular Probes, Invitrogen, Grand Island, NY) at 1:1000 and Alexa Fluor 488 (goat anti-mouse, Molecular probes, Grand Island, NY) at 1:1000 and incubated for 1.5 hours at room temperature. Five final

washes were performed: 4 times with TBST, and once with TBS. Cells were visualized using a fluorescence microscope.

Results

Tissue specific distribution of C19L1

Immunoblots using HPA036889 antiserum revealed detection of this protein in almost all brain tissues including cerebellum (left and right), occipital lobe, parietal lobe, spinal cord, temporal lobe, and thalamus (Figure 4.2). Interestingly, we did not detect this protein in the frontal lobe using brain lysates (Figure 4.2). Additionally, only 1 band was detected in brain tissues, though lymphoblastoid cell lines and tissue lysates had 2 bands (Figure 2.8 and Figure 4.3). Examination of a commercial tissue lysates blot also revealed high detection of C19L1 protein in the brain, colon, heart, and kidney (Figure 4.3).

Subcellular localization of C19L1

In order to explore the impact of loss of the protein at the cellular level, we sought to determine the localization of the protein at the cellular level. Immunostaining using the HPA036890 antiserum showed nuclear localization of this protein in HeLa cells and MN1 cells (a motor neuron cell line) (Figure 4.5). To validate this result, immunoblot was performed on commercial nuclear and cytoplasmic lysates using the HPA036889 antiserum. Immunoblot showed a single band in the HeLa nuclear extract and no bands in the cytoplasmic extract. Ponceau stain of this blot demonstrates equal protein loading as antibodies used for loading control cannot detect protein in both cellular compartments (Figure 4.6).

Discussion

The function of the C19L1 protein is unknown. We have demonstrated loss of this protein in individuals affected with ataxia and demonstrated abnormal behavior and development in a fish model; however our experiments do not provide functional information about this gene. The first step toward gaining insight into the function of a gene is to characterize the protein in tissues and at the cellular level. Immunoblots showed protein expression in the brain, heart, kidney, and colon. We detected no protein in the frontal lobe, although we detected protein in all other regions of the brain. Immunoblot revealed only the canonical protein band in the brain, suggesting this is the isoform present in the brain, though we cannot exclude undetectable levels of the other predicted isoforms of C19L1, as lower amounts of protein were loaded on the commercial blot (15mg vs 30mg for LCLs, Chapter II). As commercial lysates can vary significantly, I attempted to perform immunoblotting using the Sigma HPA036889 antiserum on mouse tissue lysates however, the results between experiments were inconsistent suggesting the antiserum is not effective for mouse tissue lysates (Figure 4.7).

It is unclear why loss of the c19l1 protein causes only a neurological phenotype. Importantly, there are examples of proteins that are expressed in multiple tissues but where a gain or loss of function mutation causes a phenotype only in a subset of organs in which the gene is expressed. Mutations in the SMN1 gene

cause spinal muscular atrophy[7] although the protein is ubiquitously expressed and mutations in the gene cause motor neuron loss and skeletal muscle atrophy specifically in motor neurons[7–12]. Differences in impact of genetic mutations in specific tissues can be due to increased sensitivity of particular organs to changes in protein expression or due to tissue-specific interacting proteins or regulators [7–10]. Further assessment of the function of C19L1 in neuronal cells is necessary in order to understand the requirements of this protein in neuronal tissue when compared with non-neuronal tissues.

My studies of subcellular localization through immunofluorescence demonstrated nuclear localization of *CWF19L1* in HeLa and MN1 cells (Figure 4.5). Immunostaining cells using the HPA036889 antiserum, even at very low dilutions, gave no signal suggesting this antibody cannot detect wild type protein. Immunostaining with the HPA036890 antiserum gave positive signal in the nucleus, however this antiserum (HPA036890) gave a non-specific minor band on LCL lysate western blots (Figure 4.7), suggesting it could be binding a non-specific protein in the cell lysates. As such, nuclear and cytoplasmic lysates were obtained to validate these results using the HPA036889 antiserum. Immunoblot of cell compartment specific lysates showed a single band in the HeLa nuclear lysates suggesting C19L1 is localized in the nucleus.

Yeast two-hybrid studies suggest that C19L1 interacts with proteins that are located in different compartments of the cell, providing inconclusive evidence[13,14].

Proteomics assay suggest it is involved with the spliceosome and therefore part of the nucleus[14–16], while another study suggests it interacts with TOM1L1, which is involved in a variety of cellular processes, but is generally localized in the cytoplasm[13,17]. Altogether, these results suggest nuclear localization and tissue-specific distribution of C19L1.

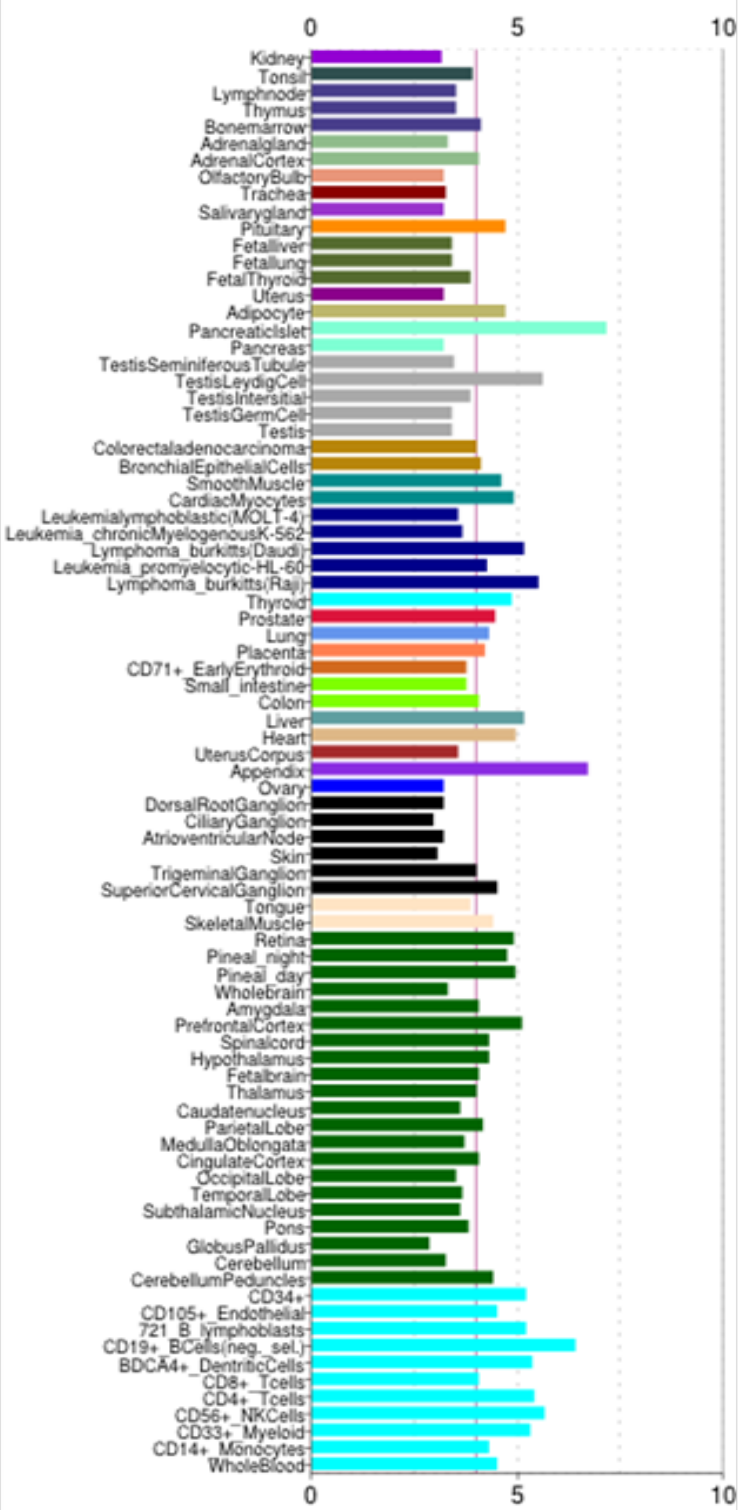


Figure 4.1. BioGPS indicates ubiquitous expression of *CWF19L1*
BioGPS expression data indicate human *CWF19L1* mRNA is expressed in all cell types. Dataset: GeneAtlas U133A, gcrma; probeset: 218787_x_at.

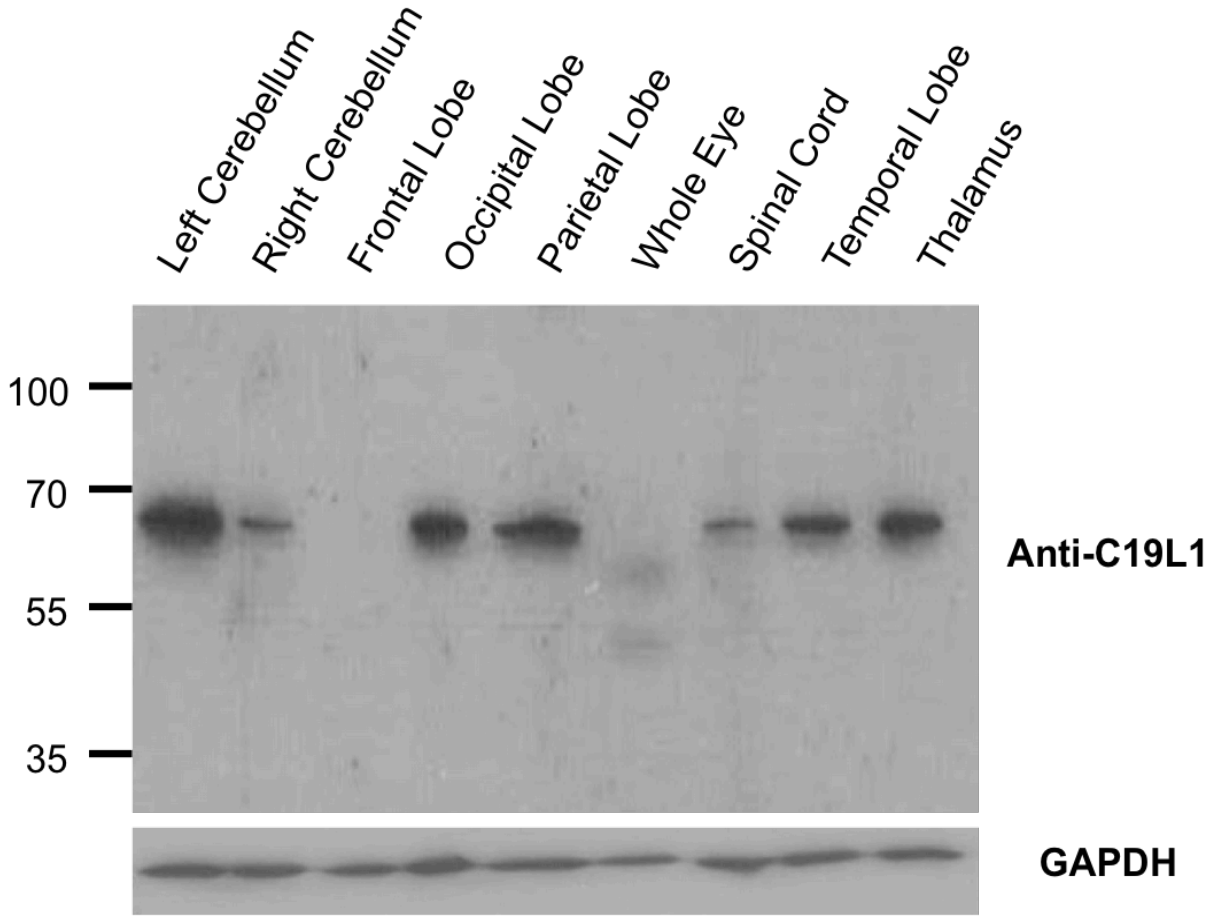


Figure 4.2. Expression of C19L1 in brain tissue

Western blot using C19L1 antiserum (Sigma) on brain lysates blot detects C19L1 in most brain tissues tested (61 kDa). 15ug protein loaded. GAPDH loading control.

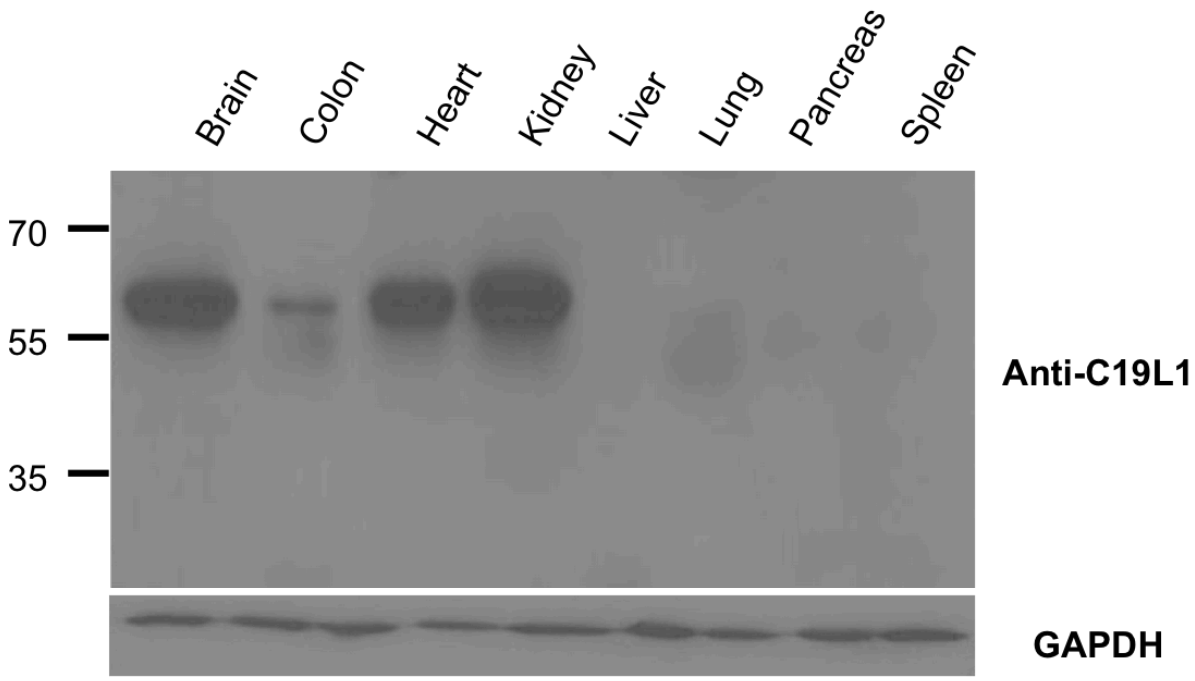


Figure 4.3. Distribution of C19L1 in tissue lysates

Western blot using C19L1 antiserum (Sigma) on commercial tissue lysates blot detects C19L1 in specific tissue lysates. 15ug protein loaded. GAPDH loading control.

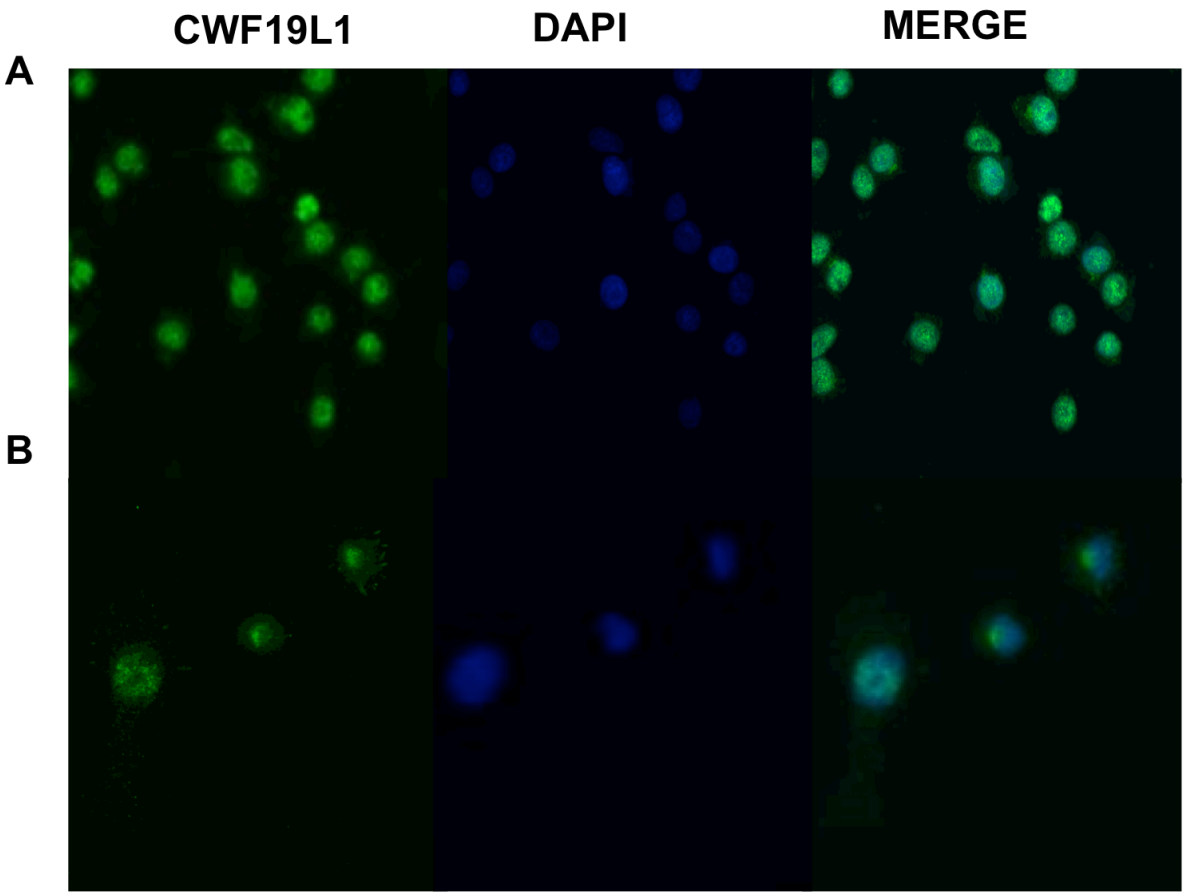


Figure 4.4. Immunofluorescence detects C19L1 in nucleus

Immunofluorescence for C19L1 (green, left panels) and nuclear marker DAPI (blue, middle panels) in (**A**) HeLa and (**B**) MN1 cell lines. Right panels show C19L1 and DAPI panels merged to indicate regions of co-localization (green-blue, right panels).

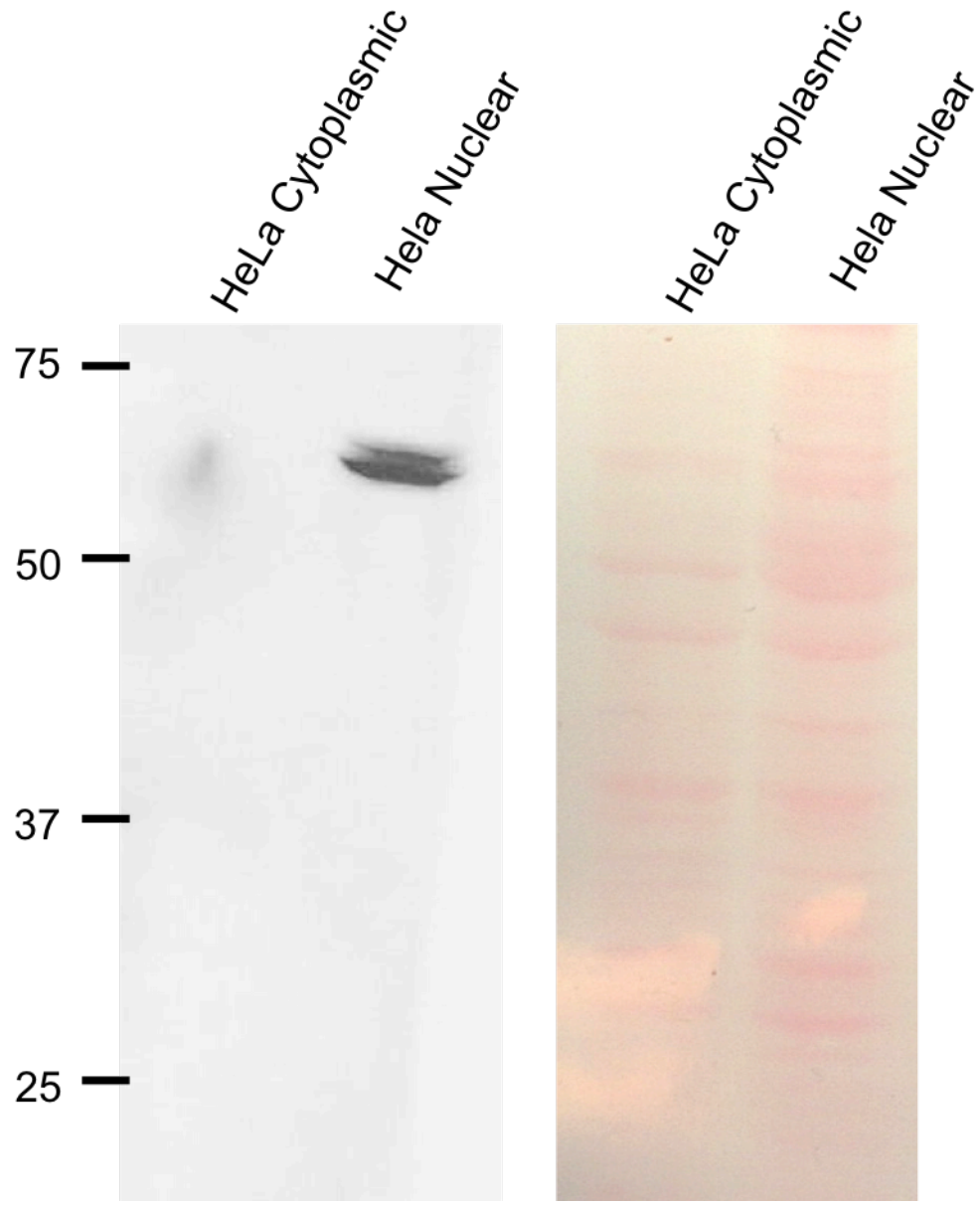


Figure 4.5. Immunoblot detects C19L1 in nucleus

(**A**) Western blot using C19L1 antiserum on commercial nuclear and cytoplasmic lysates detects C19L1 in nuclear lysate. 25ug protein loaded. (**B**) Ponceau stain shows loaded protein.

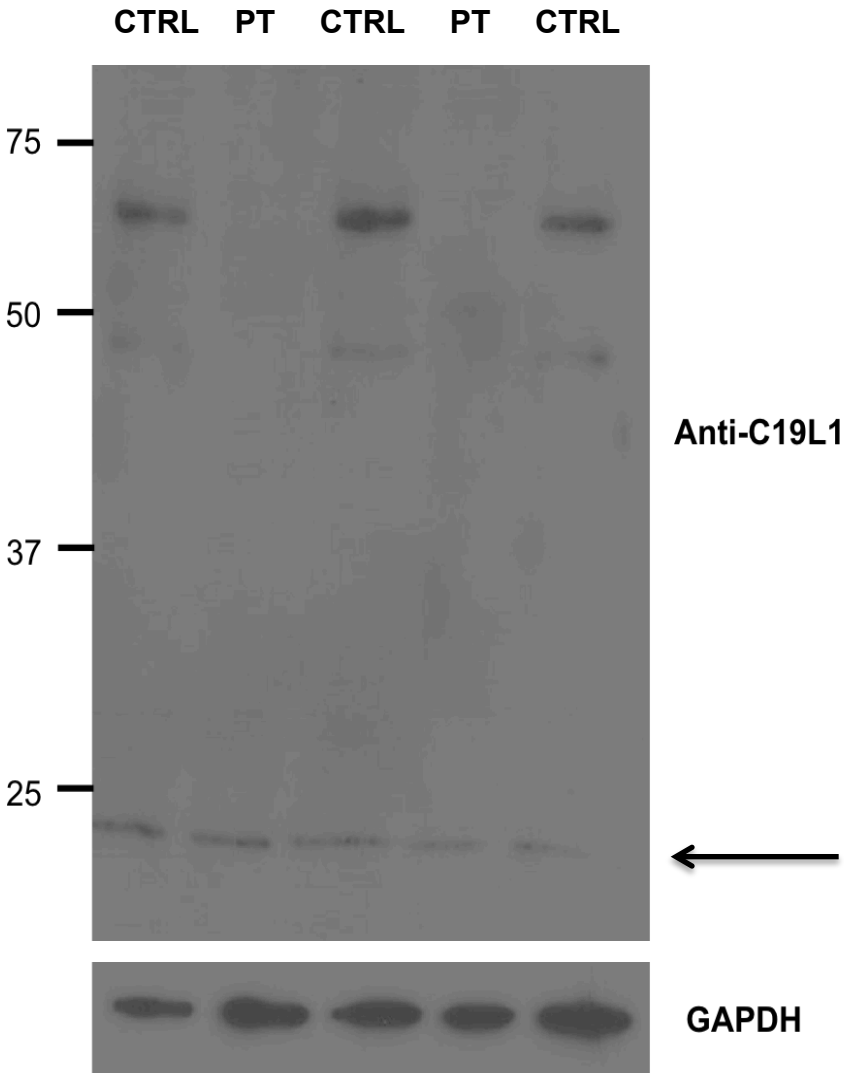


Figure 4.6 Immunoblot detects non-specific band in LCLs

Western blot using C19L1 antiserum (Sigma) on LCL lysates shows minor band (arrow) at 23 kDa in all individuals. Pt- affected, ctrl- control. 30ug protein loaded. GAPDH loading control.

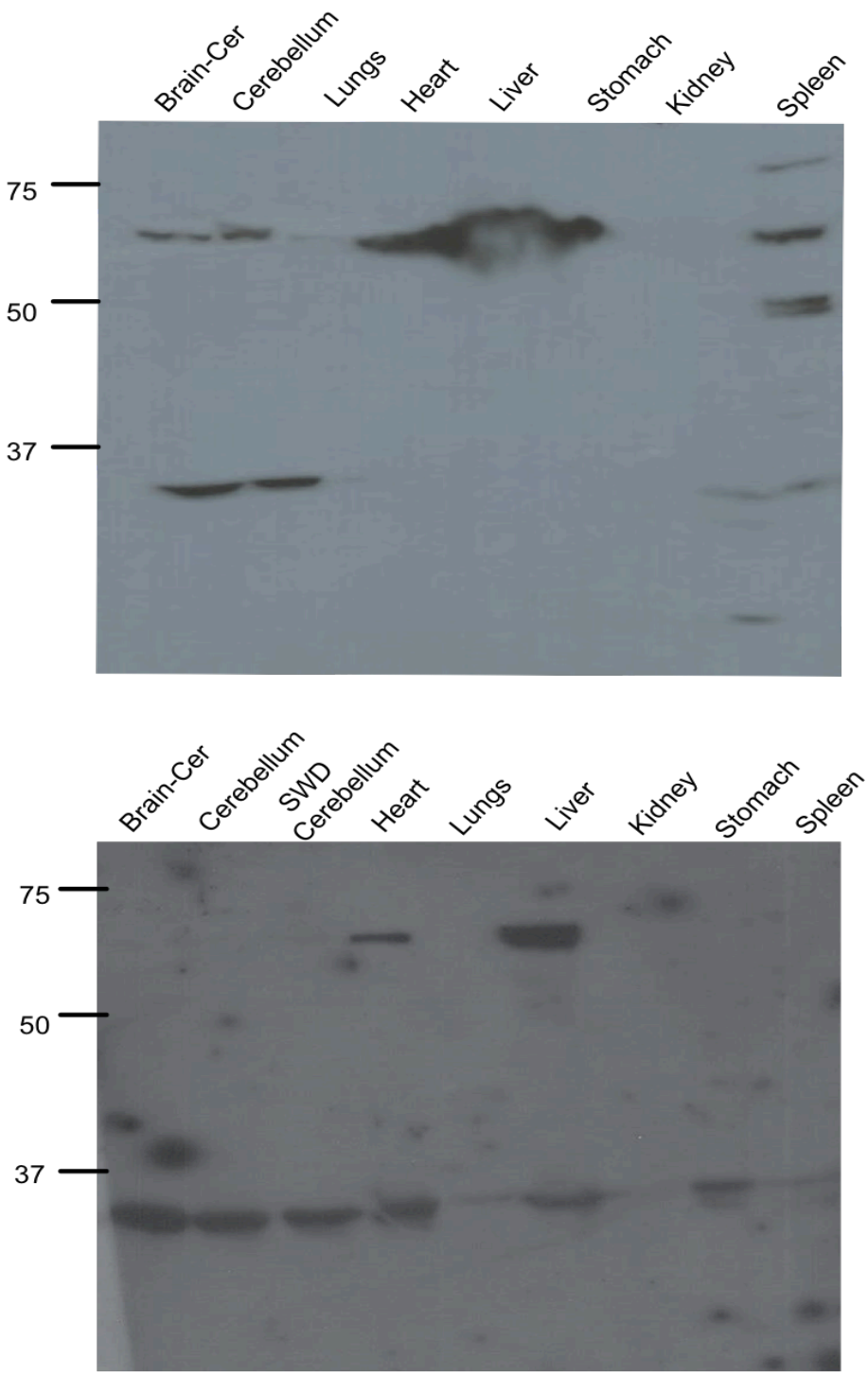


Figure 4.7. Immunoblot reveals inconsistent results between mouse tissue lysates

References

1. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science *335*, 823–828.

2. Singleton, A.B. (2011). Exome sequencing: a transformative technology. Lancet Neurol. *10*, 942–946.

3. Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd, et al. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. *10*, R130.

4. Tristan, C., Shahani, N., Sedlak, T.W., and Sawa, A. (2011). The diverse functions of GAPDH: views from different subcellular compartments. Cell. Signal. *23*, 317–323.

5. Bickmore, W.A., and Sutherland, H.G.E. (2002). Addressing protein localization within the nucleus. EMBO J. *21*, 1248–1254.

6. Hung, M.-C., and Link, W. (2011). Protein localization in disease and therapy. J. Cell Sci. *124*, 3381–3392.

7. Monani, U.R. (2005). Spinal muscular atrophy: a deficiency in a ubiquitous protein; a motor neuron-specific disease. Neuron *48*, 885–896.

8. Casad, M.E., Abraham, D., Kim, I.-M., Frangakis, S., Dong, B., Lin, N., Wolf, M.J., and Rockman, H.A. (2011). Cardiomyopathy is associated with ribosomal protein gene haplo-insufficiency in Drosophila melanogaster. Genetics *189*, 861–870.

9. Jia, Y., Mu, J.C., and Ackerman, S.L. (2012). Mutation of a U2 snRNA gene causes global disruption of alternative splicing and neurodegeneration. Cell *148*, 296–308.

10. Cañizares, J., Blanca, J.M., Navarro, J.A., Monrós, E., Palau, F., and Moltó, M.D. (2000). dfh is a Drosophila homolog of the Friedreich's ataxia disease gene. Gene *256*, 35–42.

11. Lee, A.J.-H., Awano, T., Park, G.-H., and Monani, U.R. (2012). Limited phenotypic effects of selectively augmenting the SMN protein in the neurons of a mouse model of severe spinal muscular atrophy. PloS One *7*, e46353.

12. Lee, Y.-C., Durr, A., Majczenko, K., Huang, Y.-H., Liu, Y.-C., Lien, C.-C., Tsai, P.-C., Ichikawa, Y., Goto, J., Monin, M.-L., et al. (2012). Mutations in KCND3 cause spinocerebellar ataxia type 22. Ann. Neurol. *72*, 859–869.

13. Wang, T., Liu, N.S., Seet, L.-F., and Hong, W. (2010). The emerging role of VHS domain-containing Tom1, Tom1L1 and Tom1L2 in membrane trafficking. Traffic Cph. Den. *11*, 1119–1128.

14. Ohi, M.D., Link, A.J., Ren, L., Jennings, J.L., McDonald, W.H., and Gould, K.L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol. Cell. Biol. *22*, 2011–2024.

15. Ren, L., McLean, J.R., Hazbun, T.R., Fields, S., Vander Kooi, C., Ohi, M.D., and Gould, K.L. (2011). Systematic two-hybrid and comparative proteomic analyses reveal novel yeast pre-mRNA splicing factors connected to Prp19. PloS One *6*, e16719.

16. Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res. *12*, 1231–1245.

17. Liu, N.S., Loo, L.S., Loh, E., Seet, L.-F., and Hong, W. (2009). Participation of Tom1L1 in EGF-stimulated endocytosis of EGF receptor. EMBO J. *28*, 3485–3499.

Chapter V.

Conclusions and Future Directions

Recent scientific advances have greatly increased our ability to study brain function. The lowered cost of exome sequencing and the improvements in cell lines and animal models have led to the identification of many disease genes and methods to study their pathogenesis[1–4]. Studies of rare disorders are not just important for the few families involved, but provide important new information about function and dysfunction of genes and associated proteins, especially when the gene/protein is novel and has not been studied. Discovery of these genes is necessary to begin to gain information on how these particular genes lead to disease[1,3].

In this dissertation, I have identified a mutation in *CWF19L1*, a novel ataxia gene, utilizing exome sequencing and homozygosity mapping. Sanger sequencing confirmed this mutation in affected individuals and determined that it was absent in American and Turkish controls, demonstrating that this

is not a common Turkish-specific polymorphism. RT-PCR and western blotting demonstrated instability of the mRNA transcript and deficiency of the protein. These results suggest that deficiency of the C19L1 protein, encoded by *CWF19L1*, causes the ataxia syndrome in the affected individuals. We also describe development of a zebrafish animal model in which to analyze the effect of mutation in this gene. Morpholino-mediated knockdown in zebrafish demonstrated increasing deficiency of c19l1 protein in a dose-dependent manner. Behavioral analysis of this model revealed abnormal motor behavior by touch-evoked escape response in a dose-dependent manner. Examination of the cerebellum in these fish showed decreased zebrin II immunostaining, also in a dose-dependent manner. These results suggest deficiency in the c19l1 protein causes abnormality in zebrafish motor behavior and development. Finally, initial characterization of this protein in human cells and tissues demonstrated nuclear localization of C19L1 and tissue specific distribution in the brain, heart, colon, and kidney but not in liver, lung, pancreas or spleen.

Identification of a novel ataxia gene can be useful in the future for developing a clinical test for ataxia cases with unexplained etiology. Since it has been estimated that 40% of ataxias are not currently associated with genetic mutations, *CWF19L1* is another potential gene candidate. Since this mutation was identified in a consanguineous family from a remote region in Turkey, this information can also be used for genetic testing.

Additionally, identification of the *CWF19L1* gene can be used to analyze its role in cerebellar development. The affected individuals showed non-progressive cerebellar hypoplasia, suggesting abnormal development of their cerebellum due to deficiency of this protein. Our results demonstrated that morpholino-mediated knockdown of c19l1 in fish caused decreased zebrin II staining in the cerebellum. Additional work in the zebrafish can be utilized to help explain the mechanism for abnormal cerebellar development with loss of this gene. Our results suggested decrease in zebrin II staining in the *cwf19l1* morphant zebrafish, however we cannot differentiate between loss of cerebellar cells and loss of the zebrin II marker. Future studies should analyze cell loss. Performing immunostaining using other antibodies to particular cerebellum cell types can be used to determine if the entire organ is affected or if particular cerebellum populations are affected. Pvalb7 and Vglut 1 antibodies have been used in studies to demonstrate specific loss of Purkinje and/or granular cell populations in zebrafish[5–7]. Conditional knockdown methods can also be used to determine the contribution of particular cell populations to the phenotype. Recent studies have demonstrated conditional knockdown of retina genes in zebrafish using electroporation[8]. PhotoMorphs are another novel method for conditional knockdown using morpholinos. PhotoMorphs are morpholinos with a photocleavable linkage that delays the activation of the morpholino until later stages of development. These PhotoMorphs are activated by UV light giving researchers spatiotemporal control of the morpholinos[9,10].

Additionally, zebrafish are advantageous as they can be used to perform high throughput drug screening for therapy. Zebrafish are also effective models because of high conservation of many genes from human to fish and because of easily observable traits in early larvae, including behavior, *ex utero* development and transparency at early stages[11]. Recent studies have used these advantages to test therapies for epilepsy, bipolar disorder, and Dravet syndrome[12–14].

Although the zebrafish model validated the effects of the *CWF19L1* mutation, it is a limited model. While zebrafish share vertebrate brain structure, it is clear that there are also differences. Mouse models might provide further information about cerebellar development as mice have greater conservation of brain regions, including cerebellum organization[15,16]. Mouse models would also allow researchers to explore the extent of the ataxic phenotypes through multiple behavioral tests, including rotarod, grip strength, and footprint pathway tests[17,18]. Mice are also useful for testing cognition, as there are reliable methods to examine memory and learning[17,18].

While our results are important in establishing *CWF19L1* as a disease gene, this work still has not described the function of this gene. We began to characterize C19L1 by immunoblot and found that it is distributed in brain, heart, kidney, and colon. Immunostaining revealed it is localized in the nucleus of HeLa and MN1 motor neuron cells. C19L1 is implicated in the spliceosome machinery suggesting it might function to process mRNAs. RNA-Seq is a promising tool that

can be used to analyze splicing globally in our affected individuals to determine the effect of loss of C19L1. This method is advantageous because it can be used to sequence all RNAs in our affected individuals for abnormalities in splicing and can survey transcript isoforms and quantitate expression levels[19]. Furthermore, C19L1 has an MPP domain and a HIT-like domain that are found in other ataxia genes[20–23]. Further studies that analyze the interactors of cwf19 in the spliceosome and that determine the effects of loss of cwf19 protein family members will be necessary to determine the protein's function at the cellular level. Additionally, studies that disrupt the functional domains will help to assess the importance of the functional domains contained within it. Deletion studies in yeast demonstrate that *cwf19* and *mug161* knockout yeast are viable, however, they do not reveal phenotypes of the viable yeast[24,25]. As there are easily detected phenotypes for multiple cellular processes in the yeast, yeast may be a good model organism to explore the function of cwf19 family proteins and their encoded domains[26,27].

It is unclear why this mutation causes only a neurological phenotype when our data suggest it is expressed in multiple tissues. This phenomenon is not uncommon as there are other gain of or loss of function mutations that cause specific phenotypes when they are expressed in multiple tissues or even ubiquitously. For example, overexpression of *DIAPH3* causes hearing loss though the gene is ubiquitously expressed[28]. Additionally, mutations in the *SMN1* and *HD* genes specifically cause spinal muscular atrophy and Huntington's

disease respectively, though the genes are also ubiquitously expressed[29,30]. Current studies suggest specific effects of disease mutations rely on the requirements of genes in particular cell types, for example, heterozygous mutations in the PRPF3 (cwf2) lead to Retinitis Pigmentosa because of increased elevated splicing activity in the retina[31]. Additionally, gain of function or loss of function of a protein may cause downstream effects in the regulation of other proteins that are specific to particular tissues[32,33]. Further studies that address the function of this protein and the requirement of this protein in normal brain development will be necessary to understand why this mutation causes a specific neurological phenotype.

In this work, we identified a mutation in *CWF19L1*, a novel ataxia gene, and demonstrated that the mutation causes deficiency of the protein and abnormal behavior and development in a zebrafish model. These findings revealed a novel gene that can be studied in future work to elucidate the role of this gene in brain development and function.

References

1. Sailer, A., and Houlden, H. (2012). Recent advances in the genetics of cerebellar ataxias. Curr. Neurol. Neurosci. Rep. *12*, 227–236.

2. Kabashi, E., Brustein, E., Champagne, N., and Drapeau, P. (2011). Zebrafish models for the functional genomics of neurogenetic disorders. Biochim. Biophys. Acta *1812*, 335–345.

3. Manto, M., and Marmolino, D. (2009). Cerebellar ataxias. Curr. Opin. Neurol. *22*, 419–429.

4. Perdomini, M., Hick, A., Puccio, H., and Pook, M.A. (2013). Animal and cellular models of Friedreich ataxia. J. Neurochem. *126 Suppl 1*, 65–79.

5. Bae, Y.-K., Kani, S., Shimizu, T., Tanabe, K., Nojima, H., Kimura, Y., Higashijima, S., and Hibi, M. (2009). Anatomy of zebrafish cerebellum and screen for mutations affecting its development. Dev. Biol. *330*, 406–426.

6. Hibi, M., and Shimizu, T. (2012). Development of the cerebellum and cerebellar neural circuits. Dev. Neurobiol. *72*, 282–301.

7. Yanicostas, C., Barbieri, E., Hibi, M., Brice, A., Stevanin, G., and Soussi-Yanicostas, N. (2012). Requirement for zebrafish ataxin-7 in differentiation of photoreceptors and cerebellar neurons. PloS One *7*, e50705.

8. Thummel, R., Bailey, T.J., and Hyde, D.R. (2011). In vivo electroporation of morpholinos into the adult zebrafish retina. J. Vis. Exp. JoVE e3603.

9. Tomasini, A.J., Schuler, A.D., Zebala, J.A., and Mayer, A.N. (2009). PhotoMorphs: a novel light-activated reagent for controlling gene expression in zebrafish. Genes. N. Y. N 2000 *47*, 736–743.

10. Tallafuss, A., Gibson, D., Morcos, P., Li, Y., Seredick, S., Eisen, J., and Washbourne, P. (2012). Turning gene function ON and OFF using sense and antisense photo-morpholinos in zebrafish. Dev. Camb. Engl. *139*, 1691–1699.

11. Kabashi, E., Champagne, N., Brustein, E., and Drapeau, P. (2010). In the swim of things: recent insights to neurogenetic disorders from zebrafish. Trends Genet. TIG *26*, 373–381.

12. Rahn, J.J., Bestman, J.E., Josey, B.J., Inks, E.S., Stackley, K.D., Rogers, C.E., Chou, C.J., and Chan, S.S.L. (2014). Novel Vitamin K analogs suppress seizures in zebrafish and mouse models of epilepsy. Neuroscience *259*, 142–154.

13. Baraban, S.C., Dinday, M.T., and Hortopan, G.A. (2013). Drug screening in Scn1a zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. Nat. Commun. *4*, 2410.

14. Ellis, L.D., and Soanes, K.H. (2012). A larval zebrafish model of bipolar disorder as a screening platform for neuro-therapeutics. Behav. Brain Res. *233*, 450–457.

15. Fritzsch, B. (1998). Of mice and genes: evolution of vertebrate brain development. Brain. Behav. Evol. *52*, 207–217.

16. Rakic, P. (2000). From spontaneous to induced neurological mutations: a personal witness of the ascent of the mouse model. Results Probl. Cell Differ. *30*, 1–19.

17. Crawley, J.N. (1999). Behavioral phenotyping of transgenic and knockout mice: experimental design and evaluation of general health, sensory functions, motor abilities, and specific behavioral tests. Brain Res. *835*, 18–26.

18. Karl, T., Pabst, R., and von Hörsten, S. (2003). Behavioral phenotyping of mice in pharmacological and toxicological research. Exp. Toxicol. Pathol. Off. J. Ges. Für Toxikol. Pathol. *55*, 69–83.

19. Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

20. Moreira, M.C., Barbot, C., Tachi, N., Kozuka, N., Uchida, E., Gibson, T., Mendonça, P., Costa, M., Barros, J., Yanagisawa, T., et al. (2001). The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. Nat. Genet. *29*, 189–193.

21. Kijas, A.W., Harris, J.L., Harris, J.M., and Lavin, M.F. (2006). Aprataxin forms a discrete branch in the HIT (histidine triad) superfamily of proteins with both DNA/RNA binding and nucleotide hydrolase activities. J. Biol. Chem. *281*, 13939–13948.

22. Hirano, M., Yamamoto, A., Mori, T., Lan, L., Iwamoto, T., Aoki, M., Shimada, K., Furiya, Y., Kariya, S., Asai, H., et al. (2007). DNA single-strand break repair is impaired in aprataxin-related ataxia. Ann. Neurol. *61*, 162–174.

23. Miyamoto, R., Morino, H., Yoshizawa, A., Miyazaki, Y., Maruyama, H., Murakami, N., Fukada, K., Izumi, Y., Matsuura, S., Kaji, R., et al. (2014). Exome sequencing reveals a novel MRE11 mutation in a patient with progressive myoclonic ataxia. J. Neurol. Sci. *337*, 219–223.

24. Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., et al. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. *28*, 617–623.

25. Hayles, J., Wood, V., Jeffery, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Salas-Pino, S., Heichinger, C., and Nurse, P. (2013). A genome-wide resource of cell cycle and cell shape genes of fission yeast. Open Biol. *3*, 130053.

26. Amsterdam, A., and Hopkins, N. (2006). Mutagenesis strategies in zebrafish for identifying genes involved in development and disease. Trends Genet. TIG *22*, 473–478.

27. Roux, A.E., Chartrand, P., Ferbeyre, G., and Rokeach, L.A. (2010). Fission yeast and other yeasts as emergent models to unravel cellular aging in eukaryotes. J. Gerontol. A. Biol. Sci. Med. Sci. *65*, 1–8.

28. Schoen, C.J., Emery, S.B., Thorne, M.C., Ammana, H.R., Sliwerska, E., Arnett, J., Hortsch, M., Hannan, F., Burmeister, M., and Lesperance, M.M. (2010). Increased activity of Diaphanous homolog 3 (DIAPH3)/diaphanous causes hearing defects in humans with auditory neuropathy and in Drosophila. Proc. Natl. Acad. Sci. U. S. A. *107*, 13396–13401.

29. Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S.W., et al. (1996). Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell *87*, 493–506.

30. Lee, A.J.-H., Awano, T., Park, G.-H., and Monani, U.R. (2012). Limited phenotypic effects of selectively augmenting the SMN protein in the neurons of a mouse model of severe spinal muscular atrophy. PloS One *7*, e46353.

31. Tanackovic, G., Ransijn, A., Thibault, P., Abou Elela, S., Klinck, R., Berson, E.L., Chabot, B., and Rivolta, C. (2011). PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. Hum. Mol. Genet. *20*, 2116–2130.

32. Casad, M.E., Abraham, D., Kim, I.-M., Frangakis, S., Dong, B., Lin, N., Wolf, M.J., and Rockman, H.A. (2011). Cardiomyopathy is associated with ribosomal protein gene haplo-insufficiency in Drosophila melanogaster. Genetics *189*, 861–870.

33. Jia, Y., Mu, J.C., and Ackerman, S.L. (2012). Mutation of a U2 snRNA gene causes global disruption of alternative splicing and neurodegeneration. Cell *148*, 296–308.