UNIVERSITY OF CALIFORNIA SAN DIEGO

Elucidating the Genetic and Molecular Mechanisms of Recessive Pediatric Brain Disease

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Shereen Georges Ghosh

Committee in charge:

Professor Joseph G. Gleeson, Chair
Professor Susan Ackerman
Professor Asa Gustafsson
Professor Gabriel Haddad
Professor Alysson Muotri

2020

The dissertation of Shereen Georges Ghosh is approved, and it is acceptable in quality and form for publication on microfilm electronically:

Chair

University of California San Diego

2020

DEDICATION

To my greatest supporter and husband, Jason; my beautiful warrior, Magdalene; and my parents for their unconditional support.

EPIGRAPH

"Entrust your works to the Lord,

and your plans will succeed."

-Proverbs 16:3

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge and warmly thank my thesis advisor, Professor Joseph Gleeson for the support and opportunities he provided throughout my graduate studies. I would also like to thank past and present members of the Gleeson lab for their technical support, friendships, and stimulating discussions, which have helped me in an immeasurable way. I would also like to thank my thesis committee members Susan Ackerman, Gabriel Haddad, Alysson Muotri, and Asa Gustaffson for their valuable insights, advice, and feedback they have provided me throughout my years as a graduate student.

I would also like to acknowledge and thank the Biomedical Sciences Graduate Program at UCSD for providing a stimulating environment and community. I would also like to thank the funding programs that have supported me throughout my tenure as a graduate student, in particular, the Genetics Training Program and its director Bruce Hamilton and the National Institute of Health.

Lastly, I would like to thank my family, friends, and church community for their endless support and encouragement over the years. My parents, George and Sara Georges deserve special notice for always pushing me to do my best, allowing me to explore my own potential, and nurturing my scholarly interests throughout my life. To my church community for their endless support and prayers throughout graduate school. I would not have made it this far without my faith at the center. Finally, I would like to thank my husband, best friend, and biggest supporter, Jason, for his endless love, patience and support throughout these last five years. I also cannot forget my beautiful daughter, Magdalene, who has been the greatest light and comfort since the first day she was born.

Chapter 2, in full, is a reprint of the material as it appears in American Journal of Human Genetics 2018. Ghosh, Shereen G.; Becker, Kerstin; Huang, He; Dixon-Salazar, Tracy; Chai, Guoliang; Salpietro, Vincenzo; Al-Gazali, Lihadh; Waisfisz, Quinten; Wang, Haicui; Vaux, Keith K.; Stanley, Valentina; Manole, Andreea; Akpulat, Ugur; Weiss, Marjan, M; Efthymiou, Stephanie; Hanna, Michael G.; Minetti, Carlo; Striano, Pasquale; Pisciotta, Livia; De Grandis, Elisa; Altmuller, Janine; Nurnberg, Peter; Thiele, Holger; Yis, Uluc; Okur, Tuncay D.; Polat, Ayse I.; Amiri, Nafise; Doosti, Mohammad; Karimani, Ehsan G.; Toosi, Mehran B.; Haddad, Gabriel; Karakaya, Mert; Wirth, Brunhilde; van Hagen, Johanna M.; Wolf, Nicole I.; Maroofian, Reza; Houlden, Henry; Cirak, Sebahattin; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Journal of Medical Genetics 2019. Ghosh, Shereen G.; Wang, Lu; Breuss, Martin W.; Green, Joshua D.; Stanley, Valentina; Yang, Xiaoxu; Ross, Danica; Traynor, Bryan J.; Alhashem, Amal M.; Azam, Matloob; Selim, Laila; Bastaki, Laila; Elbastawisy, Hanan I.; Temtamy, Samia; Zaki, Maha S.; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in European Journal of Human Genetics 2020. Ghosh, Shereen G.; Scala, Marcello; Beetz, Christian; Helman, Guy; Stanley, Valentina; Yang, Xiaoxu; Breuss, Martin W.; Mazaheri, Neda; Selim, Laila; Hadipour, Fatemeh; Pais, Lynn; Stutterd, Chloe; Karageorgou, Vasiliki; Begtrup, Amber; Crunk, Amy; Juusola, Jane; Willaert, Rebecca; Flore, Leigh A.; Kennelly, Kelly; Spencer, Christopher; Brown, Martha; Trapane, Pamela; Hurst, Anna C.E.; Rutledge, S. Lane; Goodloe, Dana H.; McDonald, Marie T.; Shashi, Vandana; Schoch, Kelly; Undiagnosed Diseases Network; Tomoum, Hoda; Zaitoun, Raghda; Hadipour, Zahra; Galehdari, Hamid; Pagnamenta, Alistair T.; Mojarrad, Majid;

Sedaghat, Alireza; Dias, Patricia; Quintas, Sofia; Eslahi, Atiyeh; Shariati, Gholamreza; Bauer, Peter; Simons, Cas; Houlden, Henry; Issa, Mahmoud I.; Zaki, Maha S.; Maroofian, Reza; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, has been submitted for publication of the material as it may appear in European Journal of Human Genetics 2020. Ghosh, Shereen G.; Breuss, Martin W.; Schlachetzki, Zinayida; Chai, Guoliang; Ross, Danica; Stanley, Valentina; Sonmez, F. Mujgan; Topaloglu, Haluk; Zaki, Maha S.; Hosny, Heba; Gad, Shaimaa; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this material.

Chapter 6, in full, has been submitted for publication of the material as it may appear in Genetics in Medicine 2020. Ghosh, Shereen G.; Lee, Sangmoon; Fabunan, Rudy; Chai, Guoliang; Zaki, Maha S.; Abdel-Salam, Ghada; Sultan, Tipu; Ben-Omran, Tawfeg; Alvi, Javeria R.; McEvoy-Venneri, Jennifer; Stanley, Valentina; Patel, Aakash; Ross, Danica; Ding, Jeffrey; Jain, Mohit; Pan, Daqiang; Lubbert, Philipp; Kammerer, Bernd; Wiedemann, Nils; Verhoeven-Duif, Nanda M.; Jans, Judith J.; Murphy, David; Toosi, Mehran B.; Ashrafzadeh, Farah; Imannezhad, Shima; Karimiani, Ehsan, G.; Ibrahim, Khalid; Waters, Elizabeth R.; Maroofian, Reza; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this material.

Chapter 7, in full, has been submitted for publication of the material as it may appear in American Journal of Medical Genetics 2020. Ghosh, Shereen G.; Hsiao, Katie K.; Gregor, Anne; Shimazu, Junko; Musaev, Damir; Stanley, Valentina; Roosing, Susanne; Silhavy, Jennifer L.; Issa, Mahmoud Y.; Elbendary, Hasnaa; Sahin, Yasin; Kariminejad, Ariana; Zaki, Maha S.; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this material.

VITA

2011-2013	Research Assistant, HIV Neurobehavioral Research Program
2011-2015	Research Assistant, Salk Institute for Biological Studies
2012-2015	Teaching Assistant, University of California San Diego
2013-2015	Student Worker, Pfizer Pharmaceuticals
2014	Bachelor of Science, University of California San Diego
2014	Bachelor of Science, University of California San Diego
2015	Master of Science, University of California San Diego
2015	Teaching Assistant, University of California San Diego
2020	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Ghosh, S.G., Hsaio, K.K., Gregor, A., Shimazu, J., Musaev, D., Stanley, V., Gleeson, J.G. "Mutations in *ARMC9*, Encoding Armadillo Repeat-Containing Protein 9, cause Joubert Syndrome with Frequency Corpus Callosal Defects." (*submitted*).

Ghosh, S.G.*, Breuss, M.W.*, Schlachetzi, Z., Chai, G., Ross, D., Stanley, V., Sonmez, F.M., Topaloglu, H., Hosny, H., Gad, S., Zaki, M.S., Gleeson, J.G. "Biallelic Hypomorphic Mutations in HEATR5B, Encoding HEAT repeat-containing protein 5B, Lead to Pontocerebellar Hypoplasia." *(under review).*

Ghosh, S.G.*, Lee, S.*, Fabunan, R., Chai, G., Zaki, M.S., Abdel-Salam, G., Sultan, T., Ben-Omran, T., Alvia, J.R., McEvoy-Venneri, J., Stanley, V., Patel, A., Ross, D., Ding, J., Jain, M., Pan, D., Lubbert, P., Kammerer, B., Wiedemann, N., Verhoeven-Duif, N.M., Jans, J.J., Murphyl, D., Toosi, M.B., Ashrafzadeh, F., Imannezhad, S., Karimiani, E.G., Ibrahim, K., Waters, E.R., Maroofian, R., Gleeson, J.G. "Biallelic Variants in *HPDL*, Encoding 4-Hydroxyphenylpyruvate Dioxygenase-Like Protein, lead to an Infantile Neurodegenerative Condition." *(in press*).

Ghosh, S.G., Beetz, C., Xu, X., Breuss, M.W., Stanley, V., Masaheri, N., Selim, L., Hadipour, F., Pais, L., Helman, G., Stutterd, C., Simons, C., Karageorgou, V., Tomoum, H., Zeitun, R., Hadipour, Z., Mahmoud, H., Galehdari, H., Mojarrad, M., Aktuna, S., Dias, P., Quintas, S., Eslahi, A., Shariati, G., Bauer, P., Houlden, H., Zaki, M.S., Maroofian, R., Gleeson, J.G. "A Recurrent Homozygous TRAPPC4 Loss-of-Function Variant in Multiple Families with an

Early-infantile Lethal Neurodegenerative Syndrome." *EJHG* (*2020).*

Kerman, B.E.*, Genoud, S.*, Vatandaslar, B.K., Denli, A.M., **Ghosh, S.G**., Xu, X., Yeo, G.W., Aimone, J.B., Gage, F.H. "Motoneuron expression profile analysis identifies a correlation between an axonal splice variant of HDGF-related protein 3 and myelination." *JBC (*2020*).*

Wang, L., Li, Z., Smith, D.E.C., Mendes, M.I., Stanely, V., **Ghosh, S.G**., Chen, D., Wang, Y., Salomons, G.S., Gleeson, J.G. "Biallelic loss of human NARS encoding Asparaginyl- tRNA synthetase blocks proliferation of Radial glial cells (RGs) in cortical organoids leading to Microcephaly." *Nat. Comm* (2020).

Hanzlikova, H., Prokhorova, E., Krejcikova, K., Cihlarova, Z., Kalasova, I., Kubovciak, J., Sachova, J., Hailstone, R., Brazina, J., **Ghosh, S.G.,** Cirak, S., Gleeson, J.G., Ahel, I., Caldecott, K.W. Pathogenic ARH3 Mutations Result in ADP-ribose Chromatin Scars during DNA Strand Break Repair." *Nat. Comm* (2020).

Wenderski, W., Wang, L., Krokhotin, A., Walsh, J., Li, H., Shoji, H., **Ghosh, S.G.**, George, R.D., Miller, E., Elias, L., Gillespie, M., Son, E., Staahl, B., Baek, S.T., Stanley, V., Moncada, C., Shipony, Z., Linker, S.B., Marchetto, M., Gage, F.H., Chen, D., Sultan, T., Zaki, M., Ranish, J.A., Miyakawa, T., Luo, L., Malenka, R., Crabtree, G.R., Gleeson, J.G. "Loss of the neural-specific BAF subunit ACTL6B relieves repression of early response genes and causes recessive autism." *PNAS* (2020).

Ghosh, S.G., Wang, L., Breuss, M.W., Green, J.D., Stanley, V., Yang, X., Ross, D., Traynor, B.J., Alhashem, A., Azam, M., Selim, L., Bastaki, L., Elbastawisy, H.I., Temtamy, S., Zaki, M.S., Gleeson, J.G. "Recurrent Homozygous Damaging Mutation in TMX2, Encoding a Protein Disulfide Isomerase, in Four Families with Microlissencephaly." *J Med Genet* (2019).

Ghosh, S., Becker, K., Huang, He., Salazar, T.D., Chai, G., Salpietro, V., Al-Gazali, Lihadh, Waisfisz, Q., Wang, Haicui, Vaux, K.A., Stanley, V., Manole, A., Akpulat, U., Weiss, M.M., Efthymiou, S., Hanna, M.G., Minetti, C., Striano, P., Pisciotta, L., De Grandis, E., Altmuller, J., Nurnberg, P., Thiele, H., Yis, U., Okur, T., Polat, A., Amiri, N., Doosti, M., Karimani, E., Toosi, M.B., Haddad, G., Karakaya, M., Wirth, B., van Hagen, J.M., Wolf, N.I., Maroofian, R., Houlden, H., Cirak, S., Gleeson, J.G. "Biallelic mutations in *ADPRHL2*, encoding ADP-ribosylhydrolase 3, lead to a degenerative pediatric stress- induced epileptic ataxia syndrome." *Am J Hum Genet* (2018).

Kerman, B.E., Kim, H.J., Padmanabhan, K., Mei, A., **Georges, S**., Joens, M.S., Fitzpatrick, J.A.J., Jappelli, R., Chandross, K.J., August, P., and Gage, F.H. "In Vitro Myelin Formation Using Embryonic Stem Cells." *Development* 142.12 (2015): 2213-225. Web.

Georges, Shereen. "A New Axonal Splice Variant of HDGF-related Protein 3 Increases Mature Oligodenrocyte Numbers." Thesis. University of California, San Diego, 2015. Print.

Montoya, J.L., **Georges, S**., Poquette, A., Depp, C.A., Atkinson, J.H., Moore, D.J., Translational Methamphetamine AIDS Research Center (TMARC) Group. "Refining a personalized mHealth intervention to promote medication adherence among HIV+ methamphetamine users." *AIDS Care* (2014): 2014;26(12):1477-81.

Moore, D. J., Montoya, J.L., Blackston, K., Rooney, A., Gouaux, B., **Georges, S.,** Depp, C.A., Atkinson, J.H., and The TMARC Group. "Preliminary Evidence for Feasibility, Use, and Acceptability of Individualized Texting for Adherence Building for Antiretroviral Adherence and Substance Use Assessment among HIV-Infected Methamphetamine Users." *AIDS Research and Treatment* (2013).

FIELDS OF STUDY

Major Field: Biomedical Science

Studies in Neurogenetics and Neurodevelopment
Professor Joseph G. Gleeson

Major Field: Genetics

Studies in Neurodevelopment
Professors Fred H. Gage and Bilal Kerman

ABSTRACT OF THE DISSERTATION

Elucidating the Genetic and Molecular Mechanisms of Recessive Pediatric Brain Disease

by

Shereen Georges Ghosh

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2020

Professor Joseph G. Gleeson, Chair

The structural organization and maturation of the brain are the result of a precisely orchestrated series of developmental processes with complex genetic regulation that occur during embryonic gestation and are critical for neuronal identity, wiring and connectivity. Mutations affecting any of the cellular processes involved in proper human brain development can lead to altered neurodevelopment and result in a number of different neurological diseases. Such diseases arising as a consequence of a biallelic mutation in a single gene are much more prevalent among offspring of consanguineous marriages, increasing the odds that a deleterious

mutation will be inherited on both chromosomes. The identification of a disease-causing gene in families with inherited brain disorders is one of the most useful methods for gaining insight into human brain development, function, and pathology. This study largely focuses on identifying novel mutations in both known and novel recessive pediatric brain diseases using next-generation sequencing approaches in combination with *in vitro* and *in vivo* modeling. By taking advantage of the large number of consanguineous families in our cohort, we are not only able to explore the cause of disease directly in humans, but also able to determine genetic disease risk and identify novel disease-causing variants with a high level of certainty. A total of 88 affected individuals were identified and studied from 46 families displaying a range of Structural Brain Defects (SBDs) including neurodegeneration (i.e. cortical and cerebellar atrophy), microlissencephaly, Pontocerebellar Hypoplasia, and Joubert syndrome, among other symptoms. A total of six disease-causing genes, *ADPRHL2, TMX2, TRAPPC4, HEATR5B, HPDL,* and *ARMC9* were identified, four of which had never been implicated in disease prior to this study. We were further able to model disease *in vivo* for three out of the six genes (*ADPRHL2, HEATR5B,* and *HPDL)* using either fly or mouse models. Taken together, this study not only provides further knowledge in identifying novel causes of both previously unknown diseases as well as known diseases with unknown cause, but also provides further promise to the application of NGS in revealing the full spectrum of mutations associated with these diseases and in further delineating the relevant molecular pathways involved.

CHAPTER 1

Introduction

1.1 Overview of Neurodevelopment and Neurodevelopmental Disease

The human cerebral cortex is made up of approximately 20 billion neurons, each of which makes an average of 7,000 connections.[1] This complex circuitry of the human brain is further exacerbated by the equally complex network of genes required to facilitate its self-assembly.[2] Thus, the structural organization and maturation of the brain are the result of a precisely orchestrated series of developmental processes with complex genetic regulation that occur during embryonic gestation and are critical for neuronal identity, wiring and connectivity. The process of neurodevelopment begins in the early prenatal stages with a complex process that starts with proliferation of distinct cell types, followed by differentiation into various fates, migration to their appropriate locations, and formation of integral circuitries.[3,4] The utmost precision is required to ensure all four processes are properly established. To understand how each of these steps occur, physicians and scientists have often focused on disorders of brain function. By studying how brain development and function go awry, we can better understand the critical components of normal development and function.

Mutations affecting any of the cellular processes involved in proper human brain development can lead to altered neurodevelopment and result in a number of different 'neurodevelopmental diseases.'[5,6] These diseases comprise a group of complex and heterogenous disorders characterized by impairments or delays in developmental milestones (i.e. cognition,

communication, adaptive behavior) and psychomotor skills due to abnormal brain development,[5,6] and are among the most difficult to treat in comparison to other organ systems. Only with a better understanding of the development and function of the nervous system will we be able to improve the quality of life of those who suffer from such disorders.

1.2 Structural Brain Defects (SBDs)

Structural Brain Defects (SBDs) represent a large subset of neurodevelopmental diseases leading to disruption of the cerebral, cerebellar or deep brain structures and are major causes of severe neurological disability including epilepsy, intellectual disability, and chronic motor disability, often leading to death.[7,8] Among SBDs, the three largest categories are structural disorders of: 1] forebrain (i.e. cerebral cortex) (Figure 1.1); 2] midbrain (i.e. upper brainstem) and 3] hindbrain (i.e. lower brainstem and cerebellum).[4,9]

1.2.1 Forebrain SBDs

As the cerebral cortex is by far the largest and most anatomically distinct, defects are further subdivided into those caused by defective neuronal proliferation, neuronal migration, and neuronal organization and connectivity. Diseases affecting each stage have been identified (Figure 1.1).[4]

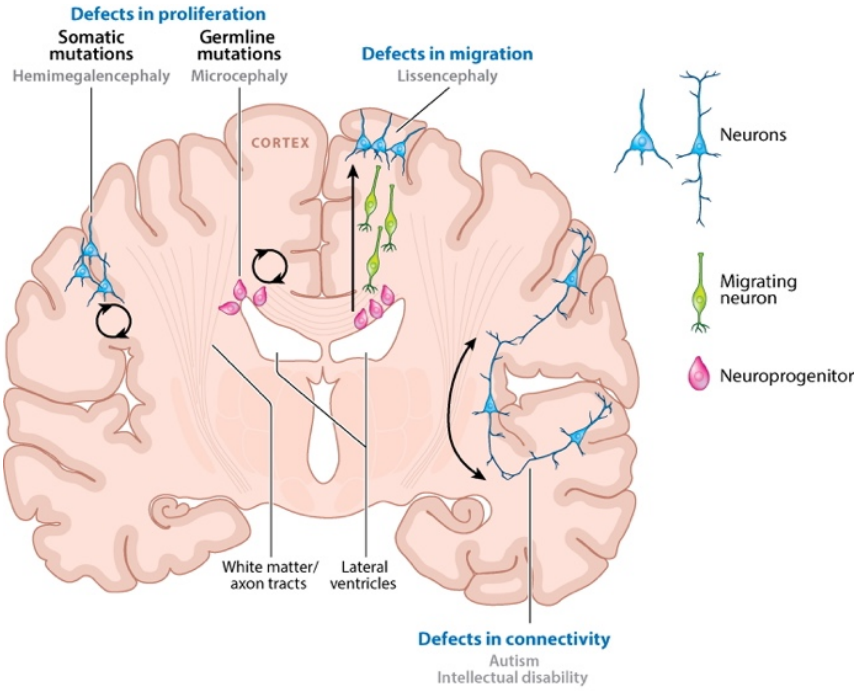

Figure 1.1: Schematic showing how SBD mutations affect genes that function at different stages of cortical development.
Defects in the proliferation of progenitors in the ventricular zone or the cortex can lead to microcephaly or hemimegalencephaly, respectively. Defects affecting the migration of neurons from the ventricular zone to the cortex give rise to lissencephaly. Connectivity defects are thought to underlie autism, intellectual disability, and neuropsychiatric disorders (adapted from W.F. Hu, et al. 2014).

For example, microcephaly ("small brain") is caused by defects in neuronal proliferation, resulting in a pathological state defined by a decreased number of neurons and a head circumference smaller than three standard deviations (SDs) below the mean (Figure 1.2).[10] The reduced brain size is caused by defects in neuronal progenitor cells, resulting either in a reduction of the output of progenitors or an increase in apoptosis during development.[11] Depending on the severity, it is associated with mental retardation, delayed motor functions, hyperactivity and epileptic seizures, with the majority of cases arising from genetic defects.[12] Recessive genetic mutations in a number of genes have been identified as causative of microcephaly, including *ASPM, MICROCEPHALIN, CDK5RAP2, CENPJ*, and *STIL*.[13-18]

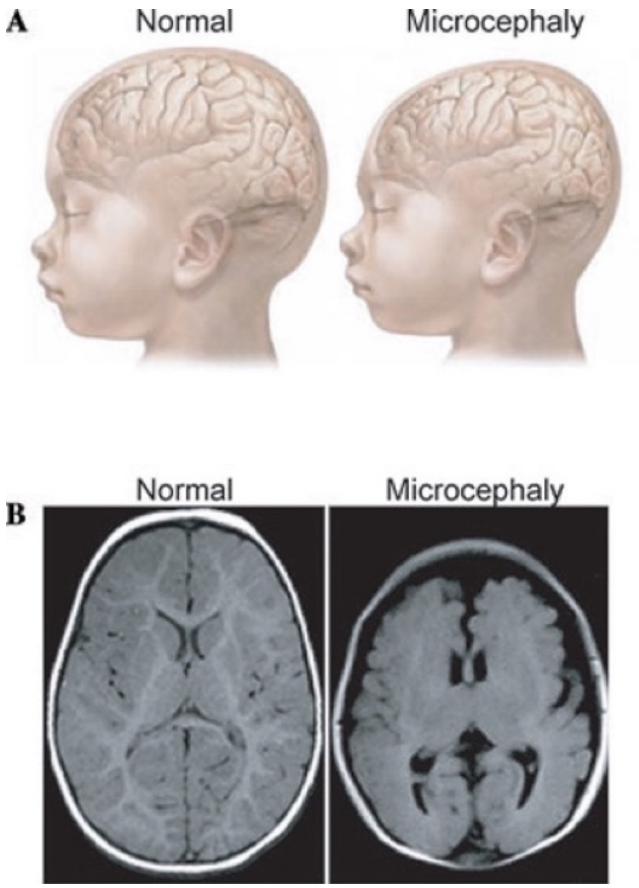

Figure 1.2: Anatomical features of primary microcephaly.
(A) Children with microcephaly have a reduced head circumference beginning at birth and a small, although architecturally normal, brain as compaired to a healthy control. (B) MRI scans of a healthy child and a child with microcephaly. Microcephalic brain shows reduced brain volume, with the largest volume loss seen in the cortical areas (Figure and caption adapted from T. Dixon-Salazar, et al. 2010).

Disorders of neuronal migration, such as Lissencephaly ("smooth brain") and other cortical gyration defects, show a wide variety of phenotypes and range in their severity.[19] Lissencephaly is defined as a loss in gyri and sulci patterning due to defective neuronal migration, resulting in a cortex that is thick, disorganized and contains only four layers instead of the usual six (Figure 1.3).[20] Children with lissencephaly present with significant developmental delays, mental disability, psychomotor impairment, and epileptic seizures; however, the severity of such phenotypes varies from case to case depending on the degree of brain malformation and

intractable epilepsy that develops.[20,21] Several genes have been identified as causing this condition, including *LIS1, DCX* (male-specific) and *TUBA1A*.[22-26]

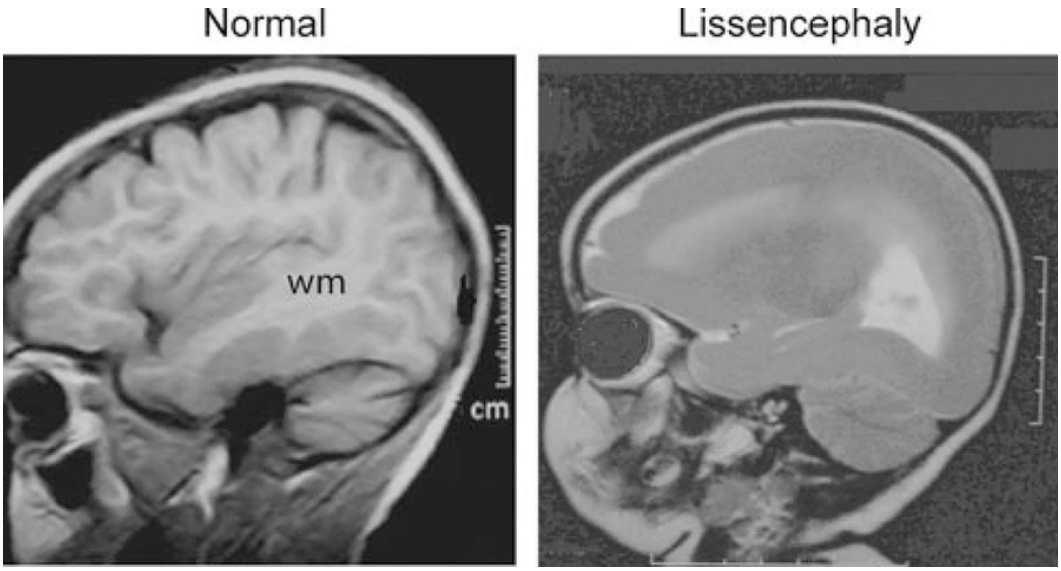

Figure 1.3: Canonical features of human lissencephaly.
MRI scans of a healthy child and a child with classical lissencephaly. Lissencephalic ('smooth') brains lack folds and risges in the outer cortical layers and typically have a smaller head size compared to healthy controls (wm, white matter) (Figure and caption adapted from T. Dixon-Salazar, et al. 2010).

Finally, defects in axon guidance and proper connectivity present with a series of symptoms such as intellectual disability, autism spectrum disorders, epilepsy, and movement disorders, in any combination.[27] One classic example of an axon guidance disorder is complete or partial agenesis of the corpus callosum, in which axons fail to cross the midline as a result of deficient neurogenesis or migration or directly by impaired axon guidance and growth.[28] This is often part of a more general neurological syndrome and is often seen in patients suffering from cortical malformations due to mutations in TUBA1A or FOXG1B.[29,30]

1.2.2 Midbrain and Hindbrain SBDs

Malformations of the midbrain and hindbrain, containing structures including the cerebellum and brain stem were nearly impossible to identify until MRI became a commonly used tool in clinical diagnosis.[31] However, with advances in neuroimaging, developmental

biology and molecular genetics, the number and complexity of recognized malformations of these structures has been steadily increasing.[9,31]

Disorders involving hypoplasia, or underdevelopment, of the cerebellum include Pontocerebellar Hypoplasia (PCH), where patients exhibit severe atrophy of the cerebellum and pons;[32] Dandy-Walker malformation, which consists of an enlarged posterior fossa that is filled with a dilated cystic-appearing ventricle;[31] diencephalon-mesencephalon (DM) junction dysplasia, which results in a poorly-defined junction between the midbrain and diencephalon leading to a 'butterfly' appearance in the upper midbrain;[31] and Joubert syndrome (JBTS), a group of disorders consisting of abnormalities with white matter tracts in the brainstem and dysplasia of the cerebellar vermis among additional accompanying abnormalities, ultimately giving rise to a radiological 'molar tooth sign' (Figure 1.4).[31]

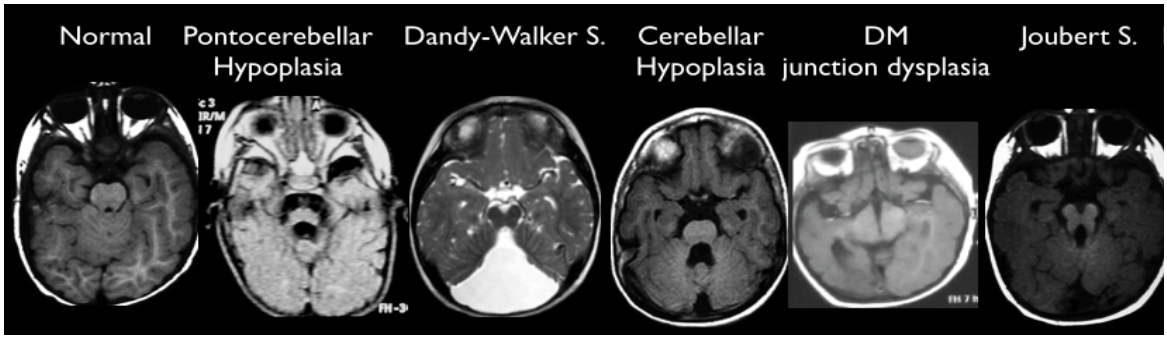

Figure 1.4: Classification of cerebellar and hindbrain SBDs according to brain MRI appearance.

Pontocerebellar hypoplasia has reduction of pons and cerebellar volume. Dandy-Walker syndrome has cystic dilation of the posterior fossa. Cerebellar hypoplasia has reduced cerebellar volume but intact brainstem volume. DM (Diencephalic-Mesencephalic) junction dysplasia has a 'butterfly' appearance of the upper midbrain. Joubert syndrome has a 'molar tooth' sign.

1.3 Pediatric-onset Neurodegenerative Disease

As opposed to neurodevelopmental disorders, which are characterized by defects involved in the organization and maturation of the brain during prenatal development, neurodegenerative disorders are characterized by progressive loss of selectively vulnerable

populations of neurons.[33] The most common types of neurodegenerative disorders are amyloidoses, like Alzheimer's disease; tauopathies, such as chronic traumatic encephalopathy; synucleinopathies, including Lewy body disorders; and TDP-43 Proteinopathies, such as Amyotrophic lateral sclerosis.[33] Although the majority of these neurodegenerative diseases do not manifest until late adulthood and arise sporadically, a wide range of neurodegenerative syndromes are being identified in early-infantile or childhood years, with Mendelian inheritance documented.[34] As opposed to neurodevelopmental SBDs, which are stable over time, patients harboring pediatric-onset neurodegenerative diseases display progressive atrophy across a wide range of brain structures including the cortex,[35] pons,[36] cerebellum,[36,37] spinal motor neurons,[37] and peripheral nerves,[37] indicating targeted cellular death of these neuronal populations.

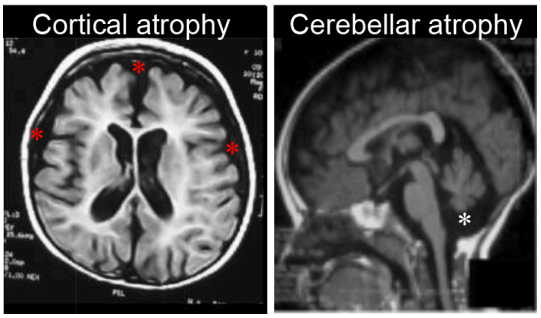

Figure 1.5: Examples of pediatric-onset neurodegenerative disorders according to brain MRI appearance.
Cortical atrophy has loss of cortical white matter, as indicated by the increased fluid surrounding the area of neuronal loss (red asterisks). Cerebellar atrophy, is a progressive loss of cerebellar volume over time due to loss of neurons in that region (white asterisk).

1.4 Mendelian Gene Inheritance and Consanguinity

Classical Mendelian patterns of inheritance are often used to describe diseases that can be linked to a single causative gene. Mendelian inheritance refers to the inheritance of phenotypes controlled by a single gene with two alleles.[38] Inheritance patterns depend on whether the genes are on autosomes or sex chromosomes, and whether the gene is dominant or recessive. If a gene resides on any one of the 22 pairs of human autosomes (all the chromosomes except the X or Y

chromosome), the inherited phenotype is considered to be 'autosomal' and is inherited in the same way regardless of the sex of the parent or offspring. Genes on autosomes can be dominant or recessive by nature. Autosomal dominant disorders only require one mutated copy of the gene for a person to be affected; however, autosomal recessive disorders require the gene to be mutated on both alleles (biallelic) for a person to be affected (Figure 1.6). If both parents are carriers of a recessive gene for a disorder, thus each having only one copy of the mutated gene, there is a 25% chance that their child will inherit the gene from each parent and have the disorder. There is a 50% chance that the child will be a carrier of the disorder. Carrier screening can be performed before or during pregnancy to determine whether one or both parents carries a gene for certain genetic disorders to determine risk of having a child with a genetic disorder.

Much of what is known about the relationship between gene function and human phenotypes is based on the study of rare variants underlying Mendelian phenotypes. The study of human phenotypes that result from inherited, mutated alleles is the most direct evidence for the requirement of a gene in normal human physiology. Thus, the study of Mendelian central nervous system (CNS) diseases can be an extremely powerful approach to not only elucidate genotype-phenotype correlations, but also to increase our understanding of how the human brain develops.

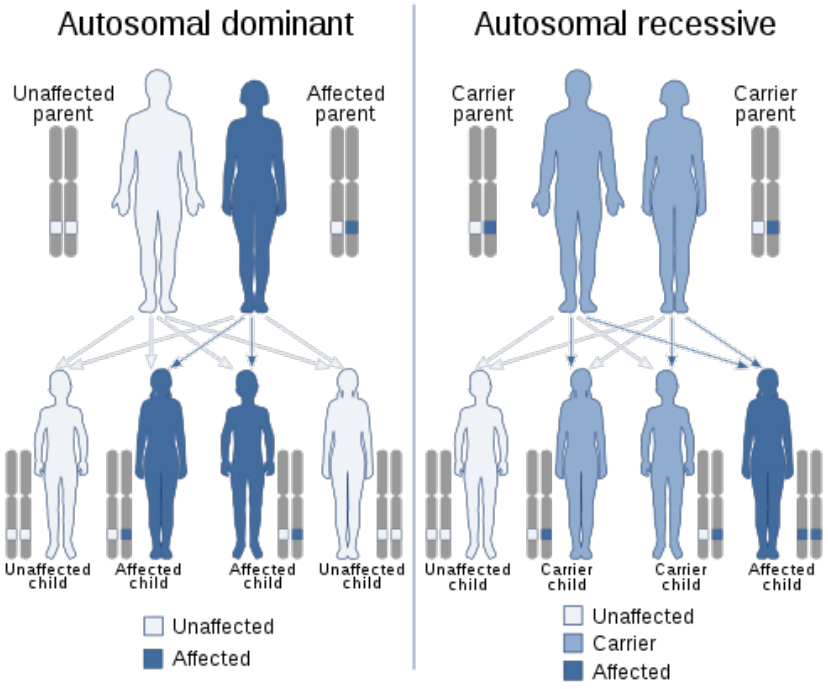

Figure 1.6: Mendelian inheritance patterns for autosomal dominant and recessive disorders. For autosomal dominant inheritance, one mutated copy of the gene is inherited from an affected parent and is sufficient to cause disease. In autosomal recessive inheritance, the parents of an individual with disease each carry one copy of the mutated gene, but typically do not show symptoms of the condition themselves (Figure adapted from https://en.wikipedia.org/wiki/Dominance_(genetics)).

Autosomal pathogenic alleles are introduced at a rate proportionate to the human mutation rate of 1.2 in 10^8 per nucleotide per generation,[39] suggesting that unaffected individuals carry a number of disease-causing recessive alleles in the heterozygous state. Monogenic diseases are conditions causally related to genome change(s), or variant(s), in a single gene.[40] These variants interfere with the efficient functioning of a gene product and may be inherited from a parent or occur *de novo* as a mutation in the germ cell of one of the parents.[40] While not uncommon in the US population,[41] diseases arising as a consequence of a biallelic mutation in a single gene are much more prevalent among offspring of consanguineous marriages, who are homozygous across 6% of the genome on average.[42] A consanguineous marriage is defined as a union between two individuals who are related as second cousins or closer.[43] Such consanguinity increases the

odds that a deleterious mutation will be inherited on both chromosomes, and results in a higher-than-expected recurrence risk for future pregnancies in these families, compared to diseases caused by *de novo* mutatons, which can arise spontaneously in the offspring during development.

The prevalence of consanguinity varies from one population to another depending on ethnicity, religion, culture and geography,[43] and confers a major public health concern on newborns.[44] In addition to the risk of acquiring a recessive genetic disease, the offspring of consanguineous parents are at an increased risk of preterm birth, congenital defects, and mortality.[44] In the Middle East, Central Asia, and North Africa, the occurrence rates are doubled where consanguinity rates approach 60%.[44]

1.5 Impact of Next-Generation Sequencing (NGS) on Recessive Disease

Advances in genetic tools and sequencing technology in the past few years have vastly expanded our understanding of the genetics of neurological disorders. Molecular diagnosis, carrier screening, prenatal diagnosis and developing new therapies are major goals for the health care system and society. Thus, the ability to unravel the variants underlying rare diseases potentiates new avenues for intervention, prevention, and treatment. Historically, successful approaches of identifying disease-causing Mendelian genes required mapping disease genes to their physical location in the human genome without any prior knowledge about their biological roles. One of the best strategies of positional mapping is linkage analysis, which analyzes the tendency for genes and other genetic markers to be inherited together.[45] However, these approaches present a major bottleneck limited by the number of affected individuals and the pedigree structure of a family.

The advent of next-generation sequencing (NGS), which enables sequencing at the scales of genomes, has largely bypassed this bottleneck and thus ushered in an era of unprecedented

pace of Mendelian disease gene discovery, replacing the need of positional mapping.[46] Monogenic diseases are most amenable to diagnosis through NGS because the disease-cuasing variants frequently involve one or a few DNA nucleotides in one or a handful of genes.[40] Arguably, NGS works best in consanguineous families with recessive disease, because of the ability to filter all but the rare homozygous variants. Thus, recessive disease-causing genes are a great target, since the expected pedigree stuctures within the families are known and recurrence rates can be measured. Using NGS, the underlying causative genes are directly distinguished via a systematic filtering, in which the identified gene variants are checked for novelty and functionality. Thus, the pace of discovery of new genes for SBDs was substantially accelerated with the advent of NGS in the research arena, especially in the area of recessive disease with parental consanguinity.

Whole-genome sequencing (WGS), one of the NGS technologies, allows for the detection of a full range of common and rare genetic variants of different types, covering almost the entire genome (Figure 1.7).[47] Whole-exome sequencing (WES), unlike WGS, focuses only on the exons (i.e. protein-coding regions) of the genome (Figure 1.7). Exons account for 1.5% of the DNA, comprising of approximately 22,000 genes.[48] In most cases, WES is preferred over WGS due to a number of reasons: 1) most identified genes implicated in Mendelian disease involve exons, 2) WES is more cost-effective than WGS, 3) the ability to interpret intronic regions of the genome is still limited.[49]

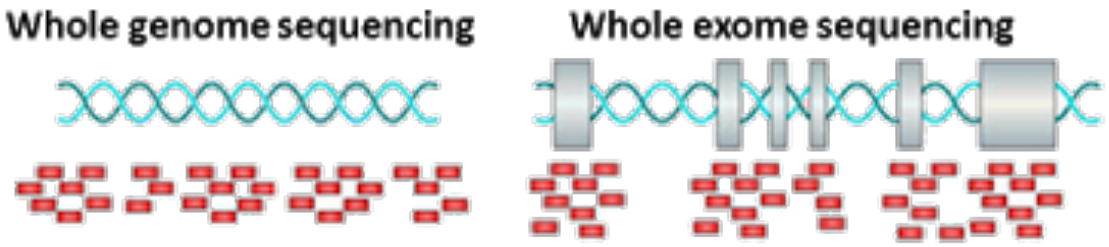

Figure 1.7: Schematic of whole genome and whole exome sequencing.
Whole genome sequencing allows for sequencing of the entire genome and can identify all kinds of variants. Whole exome sequencing allows for sequencing of just the coding regions (i.e. exons) of the genome and can identify all kinds of variants in the coding regions only (Figure adapted from http://www.genomesop.com/somatic-mutations/).

As costs for sequencing continue to drop, identification of inherited disease-causing genes is rapidly accelerating. In fact, numerous gene mutations underlying recessive SBDs have already been identified using these NGS techniques.[37, 50-53]

1.6 Rationale and Aims of this Study

The identification of a disease-causing gene in families with inherited brain disorders is one of the most useful methods for gaining insight into human brain development, function, and pathology. Through our worldwide recruitment effort, we have identified and sampled about 8000 families with neurodevelopmental or pediatric-onset neurodegenerative disorders, ~2/3 of which have SBDs. From these, we have identified the cause of disease in about 2500 (either a known gene or a high-confidence candidate gene), leaving 5500 families with an SBD of unknown cause.

This study largely focuses on identifying additional novel mutations in both known and novel recessive pediatric brain diseases using NGS approaches in combination with *in vitro* and *in vivo* modeling. By taking advantage of the large number of consanguineous families in our cohort, we are not only able to explore the cause of disease directly in humans, but also able to determine genetic disease risk and identify novel disease-causing variants with a high level of certainty. This approach also allows us to create appropriate, translationally relevant cell-based

and animal models for studying the mechanisms that underlie a given genetic disorder and the normal brain process it disrupts. A total of 88 affected individuals were identified and studied from 46 families displaying a range of SBDs including neurodegeneration (i.e. cortical and cerebellar atrophy), microlissencephaly, PCH, and JBTS, among other symptoms. A total of six disease-causing genes, *ADPRHL2, TMX2, TRAPPC4, HEATR5B, HPDL,* and *ARMC9* were identified, four of which had never been implicated in disease prior to this study. We were further able to model disease *in vivo* for three out of the five genes (*ADPRHL2, HEATR5B,* and *HPDL)* using either fly or mouse models.

Taken together, this study not only provides further knowledge in identifying novel causes of both previously unknown diseases as well as known diseases with unknown cause, but also provides further promise to the application of NGS in revealing the full spectrum of mutations associated with these diseases and in further delineating the relevant molecular pathways involved.

1.7 References

1. David A. Drachman. Do we have brain to spare? *Neurology,* 2005.

2. Byoung-il Bae, Divya Jawaraman, and Christopher A. Walsh. Genetic changes shaping the human brain. *Dev Cell,* 2016.

3. Jean-Marie Saudubray. An overview of inborn errors of metabolism affecting the brain: from neurodevelopment to neurodegenerative disorders. *Dialogues Clin Neurosci,* 2018.

4. Wen F. Hu, Maria H. Chahrour, and Christopher A. Walsh. The diverse genetic landscape of neurodevelopmental disorders. *Annu Rev Genomics Hum Genet,* 2014.

5. Ariana P. Mullin, Avanti Gokhale, Andres Moreno-De-Luca, Subhabrata Sanyal, John L. Waddington, and Victor Faundez. Neurodevelopmental disorders: mechanisms and boundary definitions from genomes, interactomes and proteomes. *Transl Psychiatry,* 2013.

6. Ana R. Cardoso, Monica Lopes-Marques, Raquel M. Silva, Catarina Serrano, Antonio Amorim, Maria J. Prata, and Luisa Azevedo. Essential genetic findings in neurodevelopmental disorders. *Hum Genomics,* 2019.

7. Matthew T. Whitehead, Stanley T. Fricke, and Andrea L. Gropman. Structural brain defects. *Clin Perinatol,* 2015.

8. Ute Hehr and Gerhardt Schuierer. Genetic assessment of cortical malformations. *Neuropediatrics,* 2011.

9. A. James Barkovich, Kathleen J. Millen, and William B. Dobyns. A developmental and genetic classification for midbrain-hindbrain malformations. *Brain,* 2009.

10. Aamod Nawathe, Jane Doherty, and Pranav Pandya. Fetal microcephaly. *BMJ,* 2018.

11. Gemma K. Thornton and C. Geoffrey Woods. Primary microcephaly: do all roads lead to Rome? *Trends Genet,* 2009.

12. C. Geoffrey Woods, Jacquelyn Bond, and Wolfgang Enard. Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *Am J Hum Genet,* 2005.

13. Jacquelyn Bond, Emma Roberts, Ganesh H. Mochida, Daniel J. Hampshire, Sheila Scott, Jonathan M. Askham, Kelly Springell, Meera Mahadevan, Yanick J. Crow, Alexander F. Markham, Christopher A. Walsh, C. Geoffrey Woods. ASPM is a major determinant of cerebral cortical size. *Nat Genet,* 2002.

14. Jacquelyn Bond, Emma Roberts, Kelly Springell, Sofia B. Lizarraga, Sheila Scott, Julie Higgins, Daniel J. Hampshire, Ewan E Morrison, Gabriella F. Leal, Elias O. Silva, Suzana M.R. Costa, Diana Baralle, Michela Raponi, Gulshan Karbani, Yasmin Rashid, Hussain Jafri, Christopher Bennett, Peter Corry, Christopher A. Walsh, and C.

Geoffrey Woods. A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat Genet,* 2005.

15. Andrew P. Jackson, Helen Eastwood, Sandra M. Bell, Jimi Adu, Carmel Toomes, Ian M. Carr, Emma Roberts, Daniel J. Hampshire, Yanick J. Crow, Alan J. Mighell, Gulshan Karbani, Hussain Jafri, Yasmin Rashid, Robert F. Mueller, Alexander F. Markham, and C. Geoffrey Woods. Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am J Hum Genet,* 2002.

16. Andrew P. Jackson, D.P. McHale, D.A. Campbell, Hussain Jafri, Yasmin Rashid, Jovaria Mannan, Gulshan Karbani, Peter Corry, M.I. Levene, Robert F. Mueller, Alexander F. Markham, Nicholas J. Lench, and C. Geoffrey Woods. Primary autosomal recessive microcephaly (MCPH1) maps to chromosome 8p22-pter. *Am J Hum Genet,* 1998.

17. Arun Kumar, Satish C. Girimaji, Mahesh R. Duvvari, and Susan H. Blanton. Mutations in *STIL*, Encoding a Pericentriolar and Centrosomal Protein, Cause Primary Microcephaly. *Am J Hum Genet,* 2009.

18. Asma Gul, Muhammad Jawad Hassan, Sabir Hussain, Syed Irfan Raza, Muhammad Salman Chishti, and Wasim Ahmad. A novel deletion mutation in *CENPJ* gene in a Pakistani family with autosomal recessive primary microcephaly. *J Hum Genet*, 2006.

19. Y. Feng and Christoper A. Walsh. Protein-protein interactions, cytoskeletal regulation and neuronal migration. *Nat Rev Neurosci,* 2001.

20. Mohammad F. Kattuoa and Joe M. Das. Lissencephaly. StatPearls Publishing, 2020.

21. Marina Bershteyn, Tomasz J. Nowakowski, Alex A. Pollen, Elizabeth Di Lullo, Aishwarya Nene, Anthony Wynshaw-Boris, and Arnold R. Kriegstein. Human iPSC-Derived Cerebral Organoids Model Celllar Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia. *Cell Stem Cell,* 2017.

22. Vincent des Portes, Jean-Marc Pinard, P. Billuart, Marie-Claude Vinet, Annette Koulakoff, Alain Carrie, Antoinette Gelot, Elisabeth Dupuis, Jacques Motte, Yoheved Berwald-Netter, Martin Catala, Axel Kahn, Cherif Beldjord, and Jamel Chelly. A Novel CNS Gene Required for Neuronal Migration and Involved in X-linked Subcortical Laminar Heterotopia and Lissencephaly Syndrome. *Cell,* 1998.

23. William B. Dobyns, Orly Reiner, Romeo Carrozo, and David H. Ledbetter. Lissencephaly: A Human Brain Malformation Associated with Deletion of the *LIS1* Gene Located at Chromosome 17p13. *JAMA,* 1993.

24. Joseph G. Gleeson, Kristina M. Allen, Jeremy W. Fox, Edward D. Lamperti, Samuel Berkovic, Ingrid Scheffer, Edward C. Cooper, William B. Dobyns, Sharon R. Minnerath, M. Elizabeth Ross, and Christopher A. Walsh. *doublecortin*, a Brain-specific Gene Mutated in Human X-linked Lissencephaly and Double Cortex Syndrome, Encodes a Putative Signaling Protein. *Cell,* 1998.

25. David A. Keays, Guoling Tian, Karine Poirier, Guo-Jen Huang, Christian Siebold, James Cleak, Peter L. Oliver, Martin Fray, Robert J. Harvey, Zoltán Molnár, Maria C. Piñon, Neil Dear, William Valdar, Steve D.M. Brown, Kay E Davies, J. Nicholas P. Rawlins, Nicholas J. Cowan, Patrick Nolan, Jamel Chelly, Jonathan Flint. Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans. *Cell,* 2007.

26. Orly Reiner, Romeo Carrozzo, Ying Shen, Manfred Wehnert, Fabrizia Faustinella, William B. Dobyns, C. Thomas Caskey, and David H. Ledbetter. Isolation of a Miller-Dicker lissencephaly gene containing G protein beta-subunit-like repeats. *Nature,* 1993.

27. Elisenda Cortes-Saladelafont, Noa Lipstein, and Angels Garcia-Cazorla. Presynaptic disorders: a clinical and pathophysiological approach focused on the synaptic vesicle. *J Inherit Metab Dis,* 2018.

28. Elizabeth C. Engle. Human Genetic Disorders of Axon Guidance. *Cold Spring Harbor Perspect Biol,* 2010.

29. Deborah J. Morris-Rosendahl, Juliane Najm, Augusta M.A. Lachmeijer, Laszlo Sztriha, Marcia Martins, Alma Kuechler, Verena Haug, Christine Zeschnigk, P. Martin, M. Santos, Carla Vasconcelos, Heymut Omran, Uwe Kraus, Marjo S. van der Knaap, Gerhardt Schuierer, Kerstin Kutsche, and Gokhan Uyanik. Refining the phenotype of alpha-1a Tubulin (TUBA1A) mutation in patients with classical lissencephaly. *Clin Genet,* 2008.

30. Chayim Can Schell-Apacik, Kristina Wagner, Moritz Bihler, Birgit Ertl-Wagner, Uwe Heinrich, Eba Klopocki, Vera M. Kalscheuer, Maximilian Muenker, and Hubertus von Voss. Agenesis and dysgenesisi of the corpus callosum: clinical, genetic and neuroimaging findings in a series of 41 patients. *Am J Med Genet A,* 2008.

31. A. James Barkovich. Developmental disorders of the midbrain and hindbrain. *Front Neuroanat,* 2012.

32. Tessa van Dijk, Frank Baas, Peter G. Barth and The Bwee Tien Poll. What's new in pontocerebellar hypoplasia? An update on genes and subtypes. *Orphanet J Rare Dis,* 2018.

33. Brittany N. Dugger and Dennis W. Dickson. Pathology of Neurodegenerative Diseases. *Cold Spring Harbor Perspect Biol,* 2017.

34. Antoni Matilla-Dueñas, Marc Corral-Juan, Agustí Rodríguez-Palmero Seuma, Dolores Vilas, Lourdes Ispierto, Sara Morais, Jorge Sequeiros, Isabel Alonso, Víctor Volpini, Carmen Serrano-Munuera, Guillem Pintos-Morell, Ramiro Álvarez, and Ivelisse Sánchez. Rare Neurodegenerative Diseases: Clinical and Genetic Update. *Adv Exp Med Biol,* 2017.

35. Michael E. Coulter, Damir Musaev, Ellen M. DeGennaro, Xiaochang Zhang, Katrin Henke, Kiely N. James, Richard S. Smith, R. Sean Hill, Jennifer N. Partlow, Muna Al-Saffar, A. Stacy Kamumbu, Nicole Hatem, A James Barkovich, Jacqueline Aziza, Nicolas Chassaing, Maha S. Zaki, Tipu Sultan, Lydie Burglen, Anna Rajab, Lihadh Al-Gazali, Ganeshwaran H. Mochida, Matthew P. Harris, Joseph G. Gleeson, and Christopher A. Walsh. Regulation of human cerebral cortical development by EXOC7 and EXOC8, components of the exocyst complex, and roles in neural progenitor cell proliferation and survival. *Genet Med,* 2020.

36. Ashleigh E. Schaffer, Veerle R.C. Eggens, Ahmet Okay Caglayan, Miriam S. Reuter, Eric Scott, Nicole G. Coufal, Jennifer L. Silhavy, Yuanchao Xue, Hulya Kayserili, Katsuhito Yasuno, Rasim Ozgur Rosti, Mostafa Abdellateef, Caner Caglar, Paul R. Kasher, J. Leonie Cazemier, Marian A. Weterman, Vincent Cantagrel, Na Cai, Christiane Zweier, Umut Altunoglu, N. Bilge Satkin, Fesih Aktar, Beyhan Tuysuz, Cengiz Yalcinkaya, Huseyin Caksen, Kaya Bilguvar, Xiang-Dong Fu, Christopher R. Trotta, Stacey Gabriel, André Reis, Murat Gunel, Frank Baas, and Joseph G. Gleeson. CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. *Cell,* 2014.

37. Vandana Shashi, Maria M. Magiera, Dennis Klein, Maha Zaki, Kelly Schoch, Sabine Rudnik-Schöneborn, Andrew Norman, Osorio Lopes Abath Neto, Marina Dusl, Xidi Yuan, Luca Bartesaghi, Patrizia De Marco, Ahmed A. Alfares, Ronit Marom, Stefan T. Arold, Francisco J. Guzmán-Vega, Loren Dm Pena, Edward C. Smith, Maja Steinlin, Mohamed Oe Babiker, Payam Mohassel, A. Reghan Foley, Sandra

Donkervoort, Rupleen Kaur, Partha S. Ghosh, Valentina Stanley, Damir Musaev, Caroline Nava, Cyril Mignot, Boris Keren, Marcello Scala, Elisa Tassano, Paolo Picco, Paola Doneda, Chiara Fiorillo, Mahmoud Y. Issa, Ali Alassiri, Ahmed Alahmad, Amanda Gerard, Pengfei Liu, Yaping Yang, Birgit Ertl-Wagner, Peter G. Kranz, Ingrid M. Wentzensen, Rolf Stucka, Nicholas Stong, Andrew S. Allen, David B. Goldstein, Undiagnosed Diseases Network, Benedikt Schoser, Kai M. Rösler, Majid Alfadhel, Valeria Capra, Roman Chrast, Tim M. Strom, Erik-Jan Kamsteeg, Carsten G. Bönnemann, Joseph G. Gleeson, Rudolf Martini, Carsten Janke, and Jan Senderek. Loss of tubulin deglutamylase CCP1 causes infantile-onset neurodegeneration. *EMBO J,* 2018.

38. Jean Gayon. From Mendel to epigenetics: History of genetics. *C R Biol,* 2016.

39. Catarina D. Campbell and Evan E. Eichler. Properties and rates of germline mutations in humans. *Trends Genet,* 2013.

40. Joshua E. Petrikin, Laurel K. Willig, Laurie D. Smith, and Stephen F. Kingsmore. Rapid Whole Genome Sequencing and Precision Neonatology. *Semin Perinatol,* 2016.

41. Daniel G. MacArthur, Suganthi Balasubramanian, Adam Frankish, Ni Huang, James Morris, Klaudia Walter, Luke Jostins, Lukas Habegger, Joseph K. Pickrell, Stephen B. Montgomery, Cornelis A. Albers, Zhengdong D. Zhang, Donald F. Conrad, Gerton Lunter, Hancheng Zheng, Qasim Ayub, Mark A. DePristo, Eric Banks, Min Hu, Robert E. Handsaker, Jeffrey A. Rosenfeld, Menachem Fromer, Mike Jin, Xinmeng Jasmine Mu, Ekta Khurana, Kai Ye, Mike Kay, Gary Ian Saunders, Marie-Marthe Suner, Toby Hunt, If H.A. Barnes, Clara Amid, Denise R. Carvalho-Silva, Alexandra H. Bignell, Catherine Snow, Bryndis Yngvadottir, Suzannah Bumpstead, David N. Cooper, Yali Xue, Irene Gallego Romero, 1000 Genomes Project Consortium, Jun Wang, Yingrui Li, Richard A. Gibbs, Steven A. McCarroll, Emmanouil T. Dermitzakis, Jonathan K. Pritchard, Jeffrey C. Barrett, Jennifer Harrow, Matthew E. Hurles, Mark B. Gerstein, and Chris Tyler-Smith. A systematic survey of loss-of-function variants in human protein-coding genes. *Science,* 2012.

42. C. Geoffrey Woods, James Cox, Kelly Springell, Daniel J. Hampshire, Moin D. Mohamed, Martin McKibbin, Rowena Stern, F. Lucy Raymond, Richard Sandford, Saghira Malik Sharif, Gulshan Karbani, Mustaq Ahmed, Jacquelyn Bond, David Clayton, and Chris F. Inglehearn. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet,* 2006.

43. Hanan Hamamy. Consanguineous marriages. *J Community Genet,* 2012.

44. Mohd Fareed and Mohammad Afzal. Genetics of consanguinity and inbreeding in health and disease. *Ann Hum Biol,* 2017.

45. Jessica X. Chong, Kati J. Buckingham, Shalini N. Jhangiani, Corinne Boehm, Nara Sobreira, Joshua D. Smith, Tanya M. Harrell, Margaret J. McMillin, Wojciech Wiszniewski, Tomasz Gambin, Zeynep H. Coban Akdemir, Kimberly Doheny, Alan F. Scott, Dimitri Avramopoulos, Aravinda Chakravarti, Julie Hoover-Fong, Debra Mathews, P. Dane Witmer, Hua Ling, Kurt Hetrick, Lee Watkins, Karynne E. Patterson, Frederic Reinier, Elizabeth Blue, Donna Muzny, Martin Kircher, Kaya Bilguvar, Francesc López-Giráldez, V. Reid Sutton, Holly K. Tabor, Suzanne M. Leal, Murat Gunel, Shrikant Mane, Richard A. Gibbs, Eric Boerwinkle, Ada Hamosh, Jay Shendure, James R. Lupski, Richard P. Lifton, David Valle, Deborah A. Nickerson, Centers for Mendelian Genomics, and Michael J. Bamshad. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges and Opportunities. *Am J Hum Genet,* 2015.

46. Fowzan S. Alkuraya. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases.

47. Aziz Belkadi, Alexandre Bolze, Yuval Itan, Aurelie CObat, Quentin B. Vincent, Alexander Antipenko, Lei Shang, Bertrand Boisson, Jean-Laurent Casanova, and Laurent Abel. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci USA,* 2015.

48. Sarah B. Ng, Emily H. Turner, Peggy D. Robertson, Steven D. Flygare, Abigail W. Bigham, Choli Lee, Tristan Shaffer, Michelle Wong, Arindam Bhattacharjee, Evan E. Eichler, Michael Bamshad, Deborah A. Nickerson, and Jay Shendure. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature,* 2009.

49. Jacek Majewski, Jeremy Schwartzentruber, Emilie Lalonde, Alexandre Montpetit, and Nada Jabado. What can exome sequencing do for you? *J Med Genet,* 2011.

50. Pamela Magini, Daphne J. Smits, Laura Vandervore, Rachel Schot, Marta Columbaro, Esmee Kasteleijn, Mees van der Ent, Flavia Palombo, Maarten H. Lequin, Marjolein Dremmen, Marie Claire Y. de Wit, Mariasavina Severino, Maria Teresa Divizia, Pasquale Striano, Natalia Ordonez-Herrera, Amal Alhashem, Ahmed Al Fares, Malak Al Ghamdi, Arndt Rolfs, Peter Bauer, Jeroen Demmers, Frans W. Verheijen, Martina Wilke, Marjon van Slegtenhorst, Peter J. van der Spek, Marco Seri, Anna C. Jansen, Rolf W. Stottmann, Robert B. Hufnagel, Robert J. Hopkin, Deema Aljeaid, Wojciech Wiszniewski, Pawel Gawlinski, Milena Laure-Kamionowska, Fowzan

S. Alkuraya, Hanah Akleh, Valentina Stanley, Damir Musaev, Joseph G. Gleeson, Maha S. Zaki, Nicola Brunetti-Pierri, Gerarda Cappuccio, Bella Davidov, Lina Basel-Salmon, Lily Bazak, Noa Ruhrman Shahar, Aida Bertoli-Avella, Ghayda M. Mirzaa, William B. Dobyns, Tommaso Pippucci, Maarten Fornerod, and Grazia M.S. Mancini. Loss of SMPD4 Causes a Developmental Disorder Characterized by Microcephaly and Congenital Arthrogryposis. *Am J Hum Genet,* 2019.

51. Caroline M. Dias, Jaya Punetha, Céline Zheng, Neda Mazaheri, Abolfazl Rad [5], Stephanie Efthymiou, Andrea Petersen, Mohammadreza Dehghani, Davut Pehlivan, Jennifer N. Partlow, Jennifer E. Posey, Vincenzo Salpietro, Alper Gezdirici, Reza Azizi Malamiri, Nihal M Al Menabawy, Laila A. Selim, Mohammad Yahya Vahidi Mehrjardi, Selina Banu, Daniel L. Polla, Edward Yang, Jamileh Rezazadeh Varaghchi, Tadahiro Mitani, Ellen van Beusekom, Maryam Najafi, Alireza Sedaghat, Jennifer Keller-Ramey, Leslie Durham, Zeynep Coban-Akdemir, Ender Karaca, Valeria Orlova, Lieke L M Schaeken, Amir Sherafat, Shalini N. Jhangiani, Valentina Stanley, Gholamreza Shariati, Hamid Galehdari, Joseph G. Gleeson, Christopher A. Walsh, James R. Lupski, Elena Seiradake, Henry Houlden, Hans van Bokhoven, and Reza Maroofian. Homozygous Missense Variants in NTNG2, Encoding Presynaptic Netrin-G2 Adhesion Protein, Lead to a Distinct Neurodevelopmental Disorder. *Am J Hum Genet,* 2019.

52. Rea M. Lardelli, Ashleigh E. Schaffer, Veerle R.C. Eggens, Maha S. Zaki, Stephanie Grainger, Shashank Sathe, Eric L. Van Nostrand, Zinayida Schlachetzki, Basak Rosti, Naiara Akizu, Eric Scott, Jennifer L Silhavy, Laura Dean Heckman, Rasim Ozgur Rosti, Esra Dikoglu, Anne Gregor, Alicia Guemez-Gamboa, Damir Musaev, Rohit Mande, Ari Widjaja, Tim L. Shaw, Sebastian Markmiller, Isaac Marin-Valencia, Justin H. Davies, Linda de Meirleir, Hulya Kayserili, Umut Altunoglu, Mary Louise Freckmann, Linda Warwick, David Chitayat, Susan Blaser, Ahmet Okay Çağlayan, Kaya Bilguvar, Huseyin Per, Christina Fagerberg, Henrik T. Christesen, Maria Kibaek, Kimberly A. Aldinger, David Manchester, Naomichi Matsumoto, Kazuhiro Muramatsu, Hirotomo Saitsu, Masaaki Shiina, Kazuhiro Ogata, Nicola Foulds, William B. Dobyns, Neil C. Chi, David Traver, Luigina Spaccini, Stefania Maria Bova, Stacey B Gabriel, Murat Gunel, Enza Maria Valente, Marie-Cecile Nassogne, Eric J Bennett, Gene W. Yeo, Frank Baas, Jens Lykke-Andersen, and Joseph G. Gleeson. Biallelic mutations in the 3'. Exonuclease TOE1 cause pontocerebellar hypoplasia and uncover a role in snRNA processing. *Nat Genet,* 2017.

53. Alicia Guemez-Gamboa, Long N. Nguyen, Hongbo Yang, Maha S. Zaki, Majdi Kara, Tawfeg Ben-Omran, Naiara Akizu, Rasim Ozgur Rosti, Basak Rosti, Eric Scott, Jana Schroth, Brett Copeland, Keith K. Vaux, Amaury Cazenave-Gassiot, Debra Q.Y. Quek, Bernice H. Wong, Bryan C. Tan, Markus R. Wenk, Murat Gunel, Stacey Gabriel, Neil C. Chi, David L. Silver, Joseph G. Gleeson. Inactivating mutations in

MFSD2A, required for omega-3 fatty acid transport in brain, cause a lethal microcephaly syndrome. *Nat Genet,* 2015.

CHAPTER 2

Biallelic Mutations in *ADPRHL2*, Encoding ADP-Ribosylhydrolase Like-2, Lead to a Lethal Pediatric Neurodegenerative Epilepsy Syndrome

2.1 Abstract

ADP-ribosylation, the addition of poly-ADP ribose (PAR) onto proteins, is a response signal to cellular challenges, such as excitotoxicity or oxidative stress. This process is catalyzed by a group of enzymes referred to as poly(ADP-ribose) polymerases (PARPs). Because the accumulation of proteins with this modification results in cell death, its negative regulation restores cellular homeostasis: a process mediated by poly-ADP ribose glycohydrolases (PARGs) and ADP-ribosylhydrolase proteins (ARHs). Using linkage analysis and exome or genome sequencing, we identified recessive inactivating mutations in *ADPRHL2* in six families. Affected individuals exhibited a pediatric-onset neurodegenerative disorder with progressive brain atrophy, developmental regression, and seizures in association with periods of stress, such as infections. Loss of the *Drosophila* paralog *Parg* showed lethality in response to oxidative challenge that was rescued by human *ADPRHL2*, suggesting functional conservation. Pharmacological inhibition of PARP also rescued the phenotype, suggesting the possibility of postnatal treatment for this genetic condition.

2.2 Introduction

ADP-ribosylation is a tightly regulated posttranslational modification of proteins involved in various essential physiological and pathological processes, including DNA repair, transcription, telomere function, and apoptosis.[1-3] The addition of poly-ADP-ribose (PAR) is mediated by a group of enzymes, referred to as poly(ADP-ribose) polymerases (PARPs), in response to cellular stressors, such as excitotoxicity or reactive oxygen species. PARylated proteins can subsequently initiate cellular stress response pathways. After resolution of the original insult, ADP-ribose polymers are rapidly removed.[4,5] Although PAR modification can protect the cell from death in the setting of cellular stress, excessive PAR accumulation or failure to reverse PAR modification can trigger a cell-death response cascade.[6,7]

Humans have two genes encoding specific PAR-degrading enzymes: *ADPRHL2* (MIM: 610624) and *PARG* (MIM: 603501). Both are capable of hydrolyzing the glycosidic bond between ADP-ribose moieties and are ubiquitously expressed.[8,9] *ADPRH* (MIM: 603081) and putatively *ADPRHL1* (MIM: 610620) encode proteins that can cleave mono-ADP-ribosylated residues and thus are not functionally redundant with *ADPRHL2* and *PARG*.[8] Studies of *in situ* hybridization have shown high *Adprhl2* expression in the developing mouse forebrain and that its expression remains high in the cerebellum, cortex, hippocampus, and olfactory bulb in early postnatal ages and persists into adulthood.[10] *Parg*$^{-/-}$ mice die embryonically as a result of PAR accumulation and cellular apoptosis.[11] There are no reports of *Adprhl2*$^{-/-}$ animals, but *Adprhl2*$^{-/-}$ mouse embryonic fibroblasts (MEFs) engineered to express the catalytic domain of nuclear PARP1 in mitochondria show PAR accumulation, as well as longer mitochondrial PAR polymers.[12,13]

Drosophila melanogaster has a single *Parg*-like gene, and null flies are lethal in the larval stage; however, when grown at a permissive temperature, a few can survive. The surviving

flies display PAR accumulation, neurodegeneration, reduced locomotion, and premature death,[14] suggesting increased neuronal vulnerability to PAR accumulation. Although mutations in enzymes PARG and PARP have not been reported in human disease, other members of this pathway have been implicated in human phenotypes.[15] For example, mutations in *XRCC1* (MIM: 194360), encoding a molecular scaffold protein involved in complex assembly during the repair of DNA-strand breaks, lead to PARP-1 overactivation and are associated with cerebellar ataxia, ocular motor apraxia, and axonal neuropathy.[16]

2.3 Materials and Methods

2.3.1 Linkage Analysis and Exome Sequencing

This study was approved by the Institutional Review Board at the respective host institutions. All study participants signed informed consent documents, and the study was performed in accordance with Health Insurance Portability and Accountability Act (HIPAA) Privacy Rules. DNA was extracted from peripheral blood leukocytes with salt extraction and genotyped with the Illumina Linkage IVb mapping panel.[17] LOD scores were calculated using easyLinkage-Plus software using genotype results of all consenting family members in the parental and affected generations.[18] Fine mapping of selected individuals was performed using the Affymetrix 250K Nsp1 SNP array and results analyzed using identity-by-descent mapping. Exome capture was performed using the Agilent SureSelect Human All Exome 50Mb Kit and paired-end sequenced on an Illumina instrument.[19] Depth of coverage was 30±16 (s.d) per exome and exome data was analyzed as previously described.[20] Whole exome sequencing for case II-1 family 3 was performed on an Illumina Hiseq 4000, using the SureSelect Human All Exon V6 enrichment kit and a paired-end 75bp sequencing protocol and data analysis and

filtering for autosomal recessive inheritance was performed as previously described.[21]

2.3.2 Whole Genome Sequencing (WGS)

Both affected siblings from family 4 underwent whole genome sequencing. In brief, sample prep (TruSeq DNA, input 250 ng) and the Illumina's HiSeqX sequencing platform were used according to manufacturer's instructions to perform WGS with 2x150bp read length. Illumina data were processed with the inhouse developed pipeline (https://github.com/hartwigmedical/pipeline), with settings validated for clinical genetics version 1.12 (https://github.com/hartwigmedical/pipeline/blob/master/settings/KG.ini). The 20x coverage of the exome target sequence was >95% for all samples. Variants present in the protein coding region or those that potentially affect splicing, +/- 6 base pairs into the intron were prioritized using Cartagenia Bench Lab NGS (Agilent Technologies, Santa Clara, USA) based on population allele frequency and anticipated inheritance patterns. Further interpretation was performed based on predicted functional impact of the variants and literature.

For case II-3 from family 6, the DNA libraries for Trio-Whole Genome Sequencing were prepared for WGS using Illumina´s TruSeq PCR-Free sample preparation kit according to manufacturer instructions. In short, 1µg of gDNA was fragmented to a mean target size of 350-450 bp using a Covaris E220 instrument, followed by end-repair, 3'-adenylation and ligation of indexed sequencing adaptors. The quality and concentration of all sequencing libraries was assessed using a LabChip GX instrument (Perkin Elmer), followed by further quality control using a MiSeq sequencer (Illumina) to obtain optimal cluster densities, insert size and library diversities. Sequencing libraries (one sample per lane) were hybridized to the surface of HiSeqX flowcells (v2 or v2.5) using the Illumina cBot™. Paired-end sequencing-by-synthesis (SBS) was performed on Illumina HiSeqX sequencers, using 2x150 cycles of incorporation and imaging.

Real-time analysis involved conversion of image data to base-calling in real-time. All steps in the sample preparation and sequencing workflow were monitored using an in-house laboratory information management system (LIMS) with barcode tracking of all samples and reagents.

The GRCh37 reference sequence was used to map reads. The raw sequences were aligned to reference sequence using BWA version 0.7.10. The sequences in the BAM files were realigned around indels with GATKLite version 2.3.9 using a public set of known indels. PCR duplicates marked with Picard tool version 1.117. We filtered out the variants with low map quality, low read depth (<6x), low Phred-scaled SNP quality score (<20) and strong strand bias. Variants were analyzed using the Clinical Sequence Miner application (DeCODE Genetics, Iceland). Low-quality variants, variants with call ratios of less than 20%, and variants with global allele frequencies higher than 0.01 were excluded from analysis according to Exome Aggregation Consortium (ExAC), Genome Aggregation Database (GnomAD) and Greater Middle East Variome project (GME Variome). Variants were analyzed under the assumption of being present or homozygous in the affected individual and absent or heterozygous in the unaffected parents. Variants were prioritized according to in silico pathogenicity score using Phred-scaled score provided by Combined Annotation Dependent Depletion (CADD).

2.3.3 Reverse Transcription (RT) PCR

Fibroblasts from affected individuals were harvested from punch biopsy and grown as previously described.[22] For the relative quantification of mRNA expression, total RNA from quantified subject and control fibroblasts was reverse-transcribed using the Superscript III First-Strand cDNA Kit (Invitrogen). PCR analysis of cDNA was performed using dHPLC-purified primers designed against exons 4-6 of ADPRHL2 excluding introns and PCR products were visualized using standard techniques.

2.3.4 Antibodies

ADPRHL2 (Sigma HPA027104); Tubulin (Sigma T6074).

2.3.5 Western Blot

Patient and healthy control fibroblasts were grown to 90% confluence, pelleted, and lysed in the treatment buffer (20mM Tris-HCl pH 8, 137mM NaCl, 1% Triton X-100, 2mM EDTA) with a cocktail of protease inhibitors (Complete Protease Inhibitor CocktailTablets, Roche). Protein quantification was performed with the BCA Protein Assay Kit (Pierce), 25 µg of the protein samples were run on standard 10% SDS-PAGE gels, and transferred to a PVDF membrane. Membranes were then blocked in either 5% non-fat milk powder or 4% BSA in Tris-buffered saline Tween (TBST), incubated overnight in the primary antibody solutions, washed, and incubated in the HRP-conjugated secondary antibody solutions for 1 hour. Bands were visualized using an appropriate chemiluminescence detection reagent.

2.3.6 Protein Purification

Wildtype *ADPRHL2* cDNA was cloned into pET15b vector and mutations were introduced via mutagenesis with NEB Q5 mutagenesis kit. Recombinant proteins were purified with *Escherichia coli BL21-CodonPlus (DE3)-RIPL* competent cells (Agilent) according to previous study.[23]

2.3.7 Thermofluor Assay

6 µg of purified protein was re-buffered in Hepes buffer (10 mM Hepes; 1 mM MgCl2; 5% Glycerol; pH 7.0) and mixed with 10 x SYPRO™ Orange Protein Gel Stain in DMSO (Invitrogen™) to a total volume of 25 µl. 100 µM ligands (adenosine diphosphate ribose (ADPr); adenosine triphosphate (ATP); ribose-1-phosphat (r-1-P); ribose-5-phosphat (r-5-P)) were added to test their influence on the thermal stability. Triplicates were pipette in a 96 well plate, heated

up to different temperatures in a rage of 15-70 °C and measured the fluorescence with Applied Biosystems 7500 Real-Time PCR System, using the ROX filter. Data were analyzed with Protein Thermal Shift™ Software (Thermo Scientific™).

2.3.8 Circular Dichroism (CD) Spectroscopy

40-50 µg purified protein were re-buffered in NaPO4 (10 mM; pH 7.0) and diluted in a total volume of 200 µl. Proteins were heated from 10°C to 70°C in 5-10°C steps in the range of 180-260 nm or with variable temperature from 10-70 °C (1°C/minute) while measuring the θ mdeg with the Jasco J-715 CD-spectrometer at 221 nm. Data were collected in cuvette with 1 mm path length and analyzed with DichroWeb using the CONTINLL algorithm and the reference set 7 (25;26). Data were normalized for molarity to obtain the mean residual elipticity.

2.3.9 Drosophila Stocks and Crosses

UAS-*Dicer 2* (#24650), daughterless (*da*)-Gal4 (#5460), and *elav*-Gal4 (#458) were obtained from the Bloomington Stock Center (Indiana). Parg$^{27.1}$ was kindly provided by A. Tulin as described.[14] UAS-*parg* RNAi (#23964GD) was obtained from VDRC (Vienna). All Gal4 and UAS strains were out-crossed into w1118 at least 6 times. Flies were maintained at room temperature. Experiments were performed at 25°C. Dicer-2 (Dcr) was included in RNAi experiments together with *da*>Gal4 or *elav*-Gal4 to enhance RNAi effects. Homozygous *elav*-Gal4; UAS-*Dcr* and *da*-Gal4;UAS-*Dcr* stocks were established by standard genetic crosses, then crossed to either to UAS-*parg* RNAi for RNA knock-down or to w1118 for control in respective experiments.[24]

2.3.10 Quantitative PCR (qPCR)

Total RNA was isolated from whole flies, heads, or bodies using Trizol (Invitrogen) and the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. cDNA synthesis was

performed with oligo-dT and random primers using SuperScript III first-strand synthesis system (Invitrogen). qPCR was performed in duplicate using SYBR Green on an ABI 7900HT qPCR system (Applied Biosystems) according to the manufacturer's protocol. All samples were analyzed from at least 3 independent experiments. Data was normalized to the level of rp49 mRNA prior to quantifying the relative levels of mRNA between controls and experimentally treated samples.

2.3.11 Oxidative Treatment

Four- to five-day old males, grouped with 20 flies per vial, were fed on a 3mm Whatman paper soaked with 1% hydrogen peroxide (Sigma) in 5% sucrose/PBS. Control flies were fed 5% sucrose/PBS. Under this condition, flies live up for 10 days without consequence. Scores were generated every 12 hours for the number of dead flies. Fresh hydrogen peroxide was added daily. For Minocycline experiments, flies were fed first with 5% sucrose/PBS or 5% sucrose/PBS/Minocycline for 24 hours prior to adding hydrogen peroxide (n=160 flies per genotype for all experiments).

2.3.12 Hypoxic Treatment

Male flies were collected within 24 hours of eclosion, placed at a density of 20 flies per vial, and aged for 2-3 days before subject to hypoxia. Flies were fed with standard food, maintained in a hypoxic chamber (COY Laboratory Products INC) filled with 2% oxygen at 22°C, and transferred every 2 day to new vials within the chamber without exposure to air, therefore avoiding re-oxygenation. The number of dead flies was scored daily (n=200 flies per genotype).

2.3.13 Lifespan

Parg$^{27.1}$ and w1118 flies were allowed to produce eggs and develop until second instar larvae at 25°C. The larvae were then transferred to an incubator at 29°C until eclosion. Adult flies were collected within 24 hours of eclosion, grouped 10 males per vial on standard fly food, and placed back at 25°C. Flies were passed every day and dead flies were scored until all Parg$^{27.1}$ flies were dead (n=60 Parg$^{27.1}$ and n=120 w1118 flies).

2.3.14 Statistical Analyses

Survival was analyzed by Kaplan–Meier Log-rank Test, and gene expression was analyzed by the Student's t-test or One-way ANOVA with Dunnett's posttest using Graph Pad Prism4 software. $p < 0.05$ was considered statistically significant.

2.4 Results

2.4.1 Clinical Presentation of Families Harboring ADPRHL2 Mutations

In this study, we show that mutations in *ADPRHL2* underlie an age-dependent recessive epilepsy-ataxia syndrome initiating with sudden severe seizures in otherwise healthy individuals followed by progressive loss of milestones, brain atrophy, and death in childhood. We describe six independent families carrying *ADPRHL2* mutations leading to a nearly identical epilepsy-ataxia syndrome (Figure 2.1A). One of the six families (family 2) lacked documentation of parental consanguinity, and the parents from this family were from the same small village. The clinical details of subjects from all included families are shown in Supplemental File 1. The emerging clinical picture is one of a stress-induced neurodegenerative disease of variable progression with developmental delay, intellectual disability, mild cerebellar atrophy (Figure 2.1B), and recurring seizures. Individuals with mutations in this gene are asymptomatic early

after birth but gradually develop a cyclic pattern of illness-related spontaneous epileptic seizures or present with a neurodegenerative course including weakness, ataxia, and loss of milestones followed by clinical deterioration that ultimately leads to premature death. Most of the subjects succumbed to sudden

unexpected death in epilepsy (SUDEP) or an apnoic-attack-like clinical presentation, suggesting a hyperacute presentation prior to the family's recognition of a predisposition. We could not establish an obvious genotype-phenotype correlation given that we show below that the missense mutation also leads to a severe loss of function. Thus, the clinical variability in the age of onset might occur because the genetic background or environmental challenges lead to variable susceptibility to illness-related cellular stress.

The differential diagnosis for this condition was based upon the presentation of a recessive condition with recurrent exacerbations and predominant features of global developmental delay, intellectual disability, seizures, neurogenic changes on electromyography, hearing impairment, regression, and mild cerebellar atrophy but not microcephaly or cataracts. The differential diagnosis in our families included mitochondrial disorders, spastic ataxia, and peripheral neuropathy.

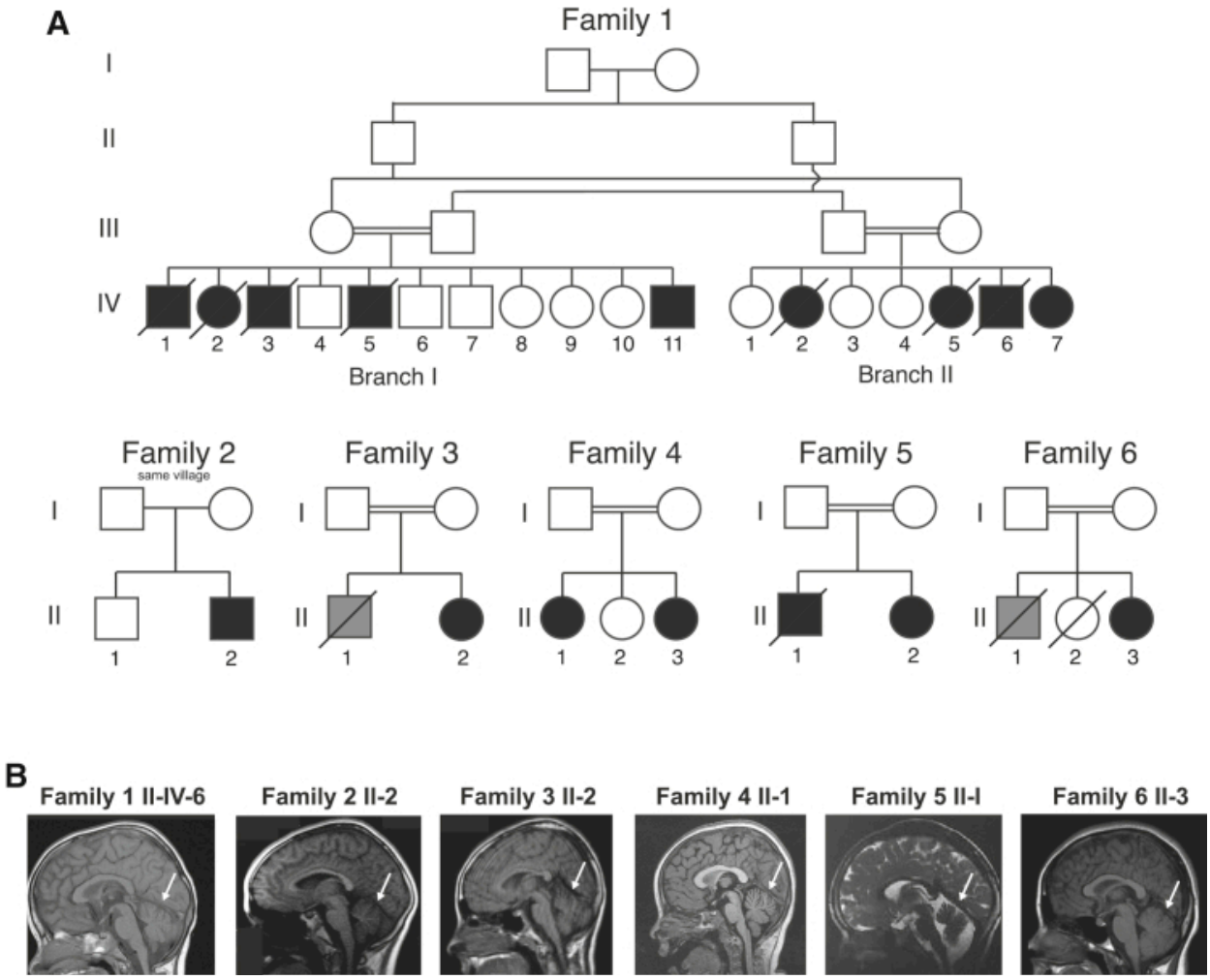

Figure 2.1: Pedigrees of families with mutations in *ADPRHL2* and their clinical presentation. (A) Pedigrees of families 1-6 show consanguineous unions (double bar) and a total of 16 affected individuals. Slashes represent deceased individuals. Black shading indicates affected individuals. Gray shading indicates individuals who passed away from SUDEP; however, no DNA is available. (B) Panels show midline sagittal MRI for one affected individual from each of the six families. White arrows indicate cerebellar atrophy, evidenced by widely spaced cerebellar foci.

2.4.2 Identification of Truncating and Missense Mutations in ADPRHL2

Genome-wide linkage analysis of 14 members of family 1 mapped the disease locus to an 11 Mb locus in chromosomal region 1p36 with a genome-wide-significant multipoint LOD score of 3.4 (Figure 2.2). Exome sequencing of individual II-IV-6 at >30× read depth for 96.9% of the exome revealed a single rare (allele frequency < 1:1,000) potentially deleterious variant within

the linkage interval: a frameshift *ADPRHL2* mutation that segregated with the phenotype according to a recessive mode of inheritance.

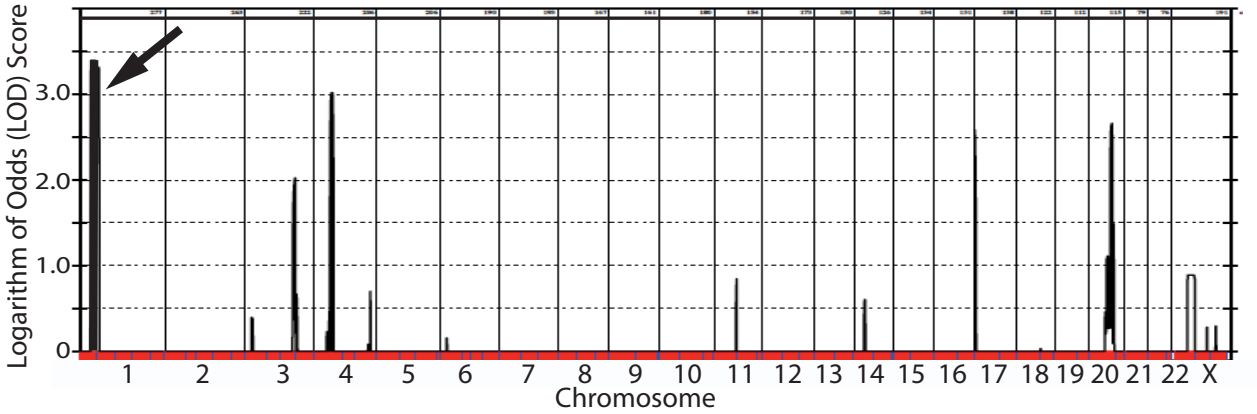

Figure 2.2: Linkage analysis for family 1.
Multipoint linkage plot for family 1 shows a peak on chromosome 1 with a maximum LOD score of 3.4 (black arrow) and lower peaks on chromosomes 4 and 20.

Using GeneMatcher, this international collaborative group of authors identified further pathogenic alleles in *ADPRHL2*.[25] After obtaining informed consent from all participating individuals in accordance with the ethical standards of the responsible committee on human experimentation at the University of California, San Diego, we identified a total of six distinct mutations in ADP-ribosylhydrolase-like 2 (*ADPRHL2* [NCBI Gene ID: 54936]) in the six families by whole-exome or genome sequencing. All variants were prioritized by allele frequency, conservation, blocks of homozygosity, and predicted effect on protein function, and in all families the homozygous variant in *ADPRHL2* was the top candidate. Variants were confirmed by Sanger sequencing and segregated with the phenotype according to a recessive mode of inheritance. All variants were predicted to be disease causing by MutationTaster.[26] These variants were not encountered in dbGaP, the ExAC Browser, 1000 Genomes, genomAD, or the Greater Middle East Variome.

ADPRHL2 contains six coding exons, yielding a single protein-coding transcript, ADP-ribosylhydrolase 3 (ARH3) (Figure 2.3A). The encoded 363 amino acid ARH3 is predicted to have a mitochondrial localization sequence and single enzymatic ADP-ribosyl-glycohydrolase domain (Figure 2.3B). Family 1 carried the homozygous exon 6 mutation c.1000C>T (GenBank: NM_017825), which introduces a premature stop codon (p.Gln334Ter) predicted to truncate the highly conserved last 30 amino acids of the protein, including part of the ADP-ribosylhydrolase domain. Family 2 harbored the homozygous exon 3 mutation c.316C>T (GenBank: NM_017825), which also introduces a premature stop codon (p.Gln106Ter) in the ADP-ribosylhydrolase domain. Family 3 revealed the homozygous exon 2 missense mutation c.235A>C (GenBank: NM_017825), which leads to an amino acid change (p.Thr97Pro) in a residue that is highly conserved among vertebrates (Figure 2.4A). Using a previously published crystal structure of ARH3, we localized this residue to an α-helical loop within the ADP-ribosylhydrolase domain and the substrate binding site, which is defined by the position of two Mg^{2+} ions located in adjacent binding sites; thus, the residue is predicted to affect protein structure and enzymatic activity (Figure 2.4B).[27] Family 4 carried the homozygous 5 bp, exon 3 deletion c.414_418TGCCC (GenBank: NM_017825), which results in a frameshift (p.Ala139GlyfsTer5) in the ADP-ribosylhydrolase domain. Family 5 carried the homozygous exon 4 missense mutation c.530C>T (GenBank: NM_017825), which leads to an amino acid change (p.Ser177Leu) that is also highly conserved among vertebrates. It is localized in a critical α-helical loop within the ADP-ribosylhydrolase domain, also suggesting an effect on protein structure and activity. Family 6 carried the homozygous exon 1 missense mutation c.100G>A (GenBank: NM_017825), which leads to an amino acid change (p.Asp34Asn) that is highly

conserved among vertebrates. This change is also localized in a critical α-helical loop within the ADP-ribosylhydrolase domain, suggesting a potential impact on protein structure and activity.

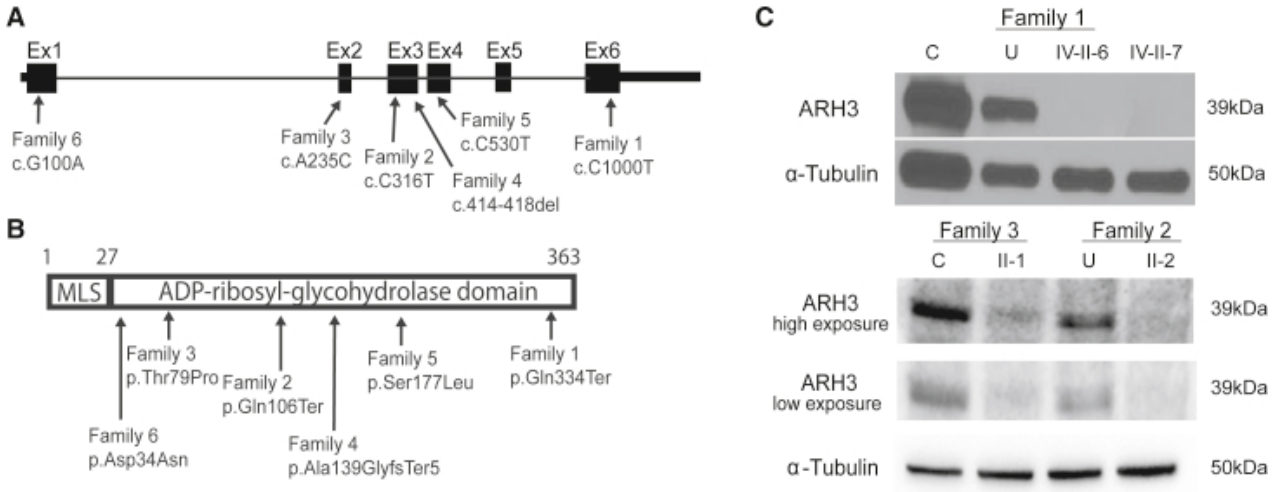

Figure 2.3: Truncating and missense *ADPRHL2* mutations in six independent families are predicted to be inactivating.

(A) Schematic of ADPRHL2 depicts the coding sequence spanning six exons and the 5' and 3' UTRs. Black arrows indicate the positions of the six identified mutations and their coordinates within the cDNA (Gene ID: 54936). (B) Schematic of ARH3 depicts the mitochondrial localization sequence (MLS) and the ADP-ribosyl-glycohydrolase domain. Black arrows indicate the position and coordinates of the impact of the described mutations. (C) Western blot of fibroblasts from an unrelated control individual (C), the unaffected carrier father (U), and affected individuals IV-II-6 and IV-II-7 from family 1 shows the absence of ARH3 in affected fibroblasts. α-tubulin was used as the loading control. Western blot of fibroblasts from an unrelated control individual (C) and affected individual II-1 from family 3 and the unaffected carrier mother (U) and affected individual II-3 from family 2 shows significantly reduced amounts of ARH3. α-tubulin was used as the loading control.

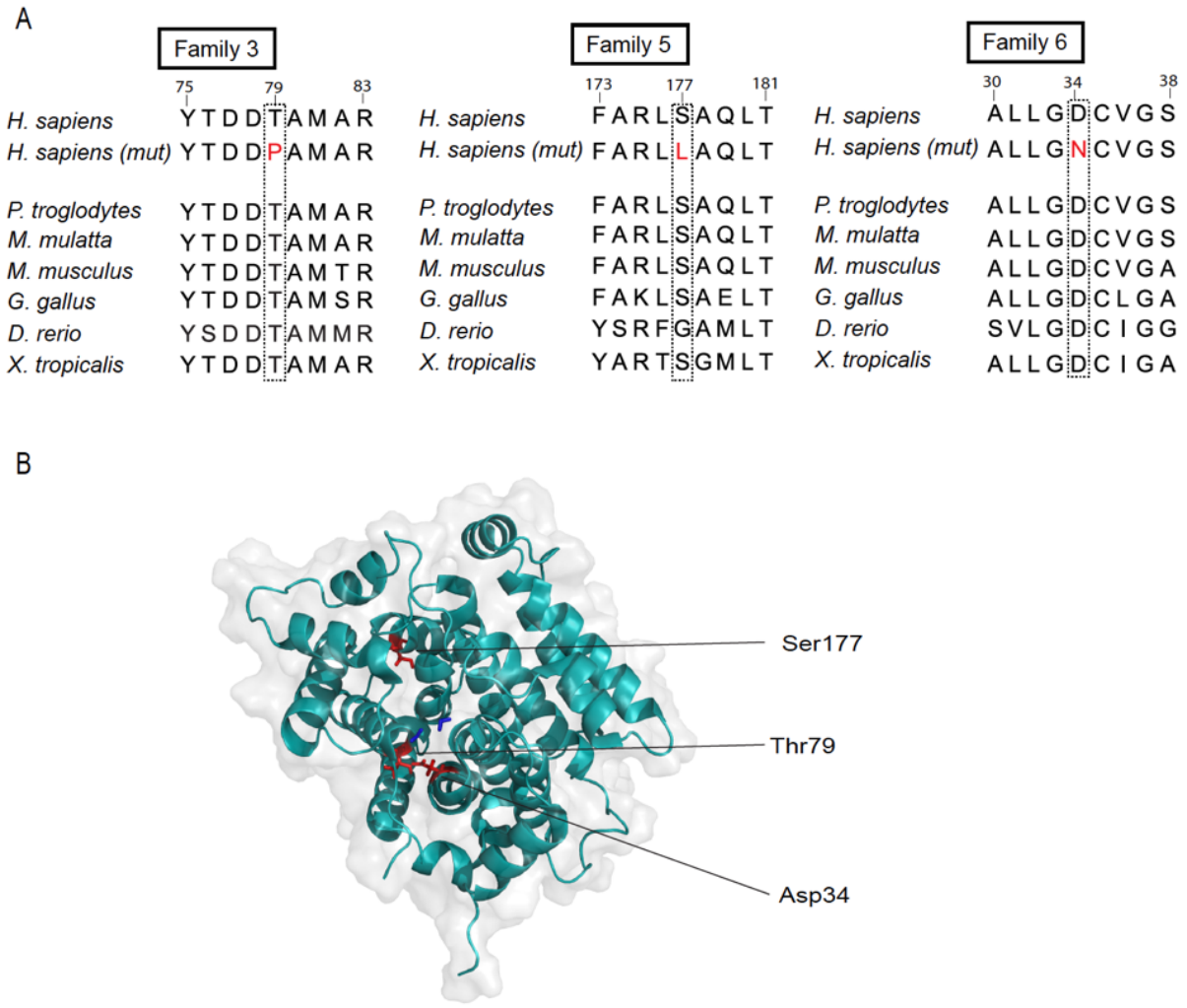

Figure 2.4: Families 3, 5 and 6 carry *ADPRHL2* missense mutations that affect conserved amino acid residues.

(A) The amino acid missense mutations in Families 3, 5, and 6 are conserved in almost all vertebrates. (B) Cartoon model of ADPRHL2 as solved at 1.6 Angstrom resolution.[23] Depicted are the mutated residues, Ser177, Thr79, and Asp34 in red within alpha-helical domains. Mg^{2+} ions are shown in blue.

2.4.3 ARH3 Mutations Display Improper Folding and Impaired Binding Activity

To determine the impact of these mutations on protein folding and binding activity, we generated recombinant proteins in *E.coli* and purified them by His-tag affinity chromatography. Our results showed that the p.Gln334Ter protein was not evident in the soluble fraction, whereas the wild-type (WT) was recovered with good purity (Figure 2.5A). The p.Thr79Pro protein was expressed and soluble, although possibly recovered with slightly less purity than WT ARH3. We

studied the deleterious impact of p.Thr79Pro by using circular dichroism spectroscopy (Figure 2.5B and C). Compared with the WT, this mutant exhibited reduced α-helical content and an altered secondary structure, in agreement with the fact that p.Thr79Pro occurred within an α-helical domain. Further, the melting temperature (T_m) of p.Thr79Pro was reduced by more than 10°C, confirming destabilization of the mutant (Figure 2.5D-F). We also found that in contrast to WT ARH3, the p.Thr79Pro protein was not stabilized by ligands such as adenosine diphosphate ribose (ADPr) (Figure 2.5G-I). We confirmed the specificity of this assay by using adenosine triphosphate (ATP) and ribose-5-phosphate as negative controls, which were not predicted to bind or stabilize ARH3. Together, these data suggest that both disease-causing, truncating mutants and amino acid substitutions should be destabilized when expressed in cells.

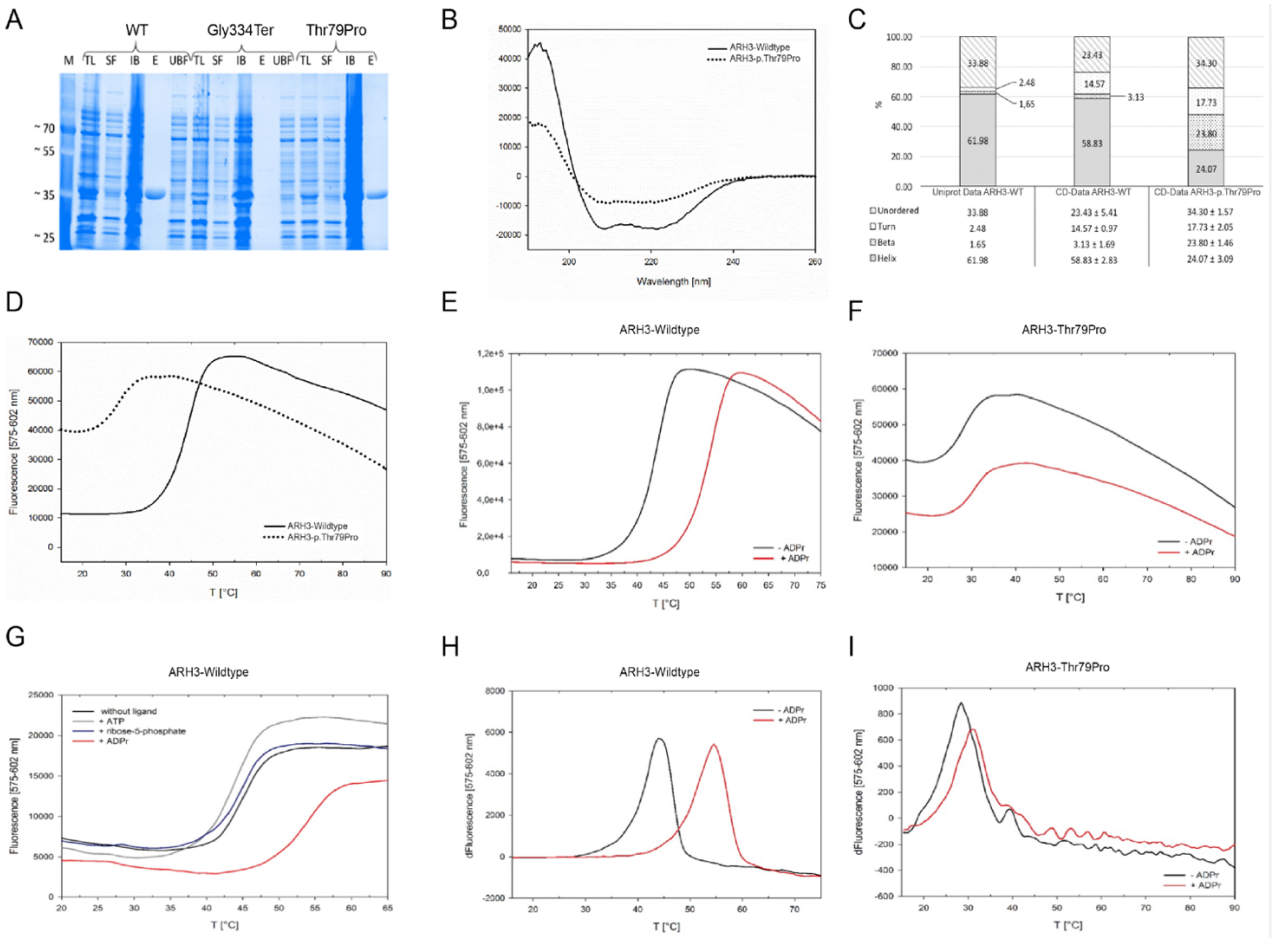

Figure 2.5: Biochemical characterization of wildtype ARH3 and mutants on protein stability and substrate binding.

Recombinant protein purification of ARH3 Wildtype (WT) and mutants with His-tag affinity chromatography. TL: total lysate, SF: soluble fraction, IB: inclusion body, E: elution of pure proteins, UBF: unbound flow through. (B-C) CD-spectroscopy measurement of secondary structure of ARH3-Wildtype and mutants. (B) Normalized exemplary presentation of the CD-spectra of recombinant ARH3-Wildtype and ARH3-p.Thr79Pro. Molar ellipticity (θ) measured at 10°C in a range of 190-260nm. (C) Experimentally determined composition of the secondary structural elements of ARH3-Wildtype and ARH3-p.Thr79Pro. Comparison of Uniprot data and CD-derived data from ARH3-Wildtype and ARH3-p.Thr79Pro. The data analysis was carried out with DichroWeb and is based on the algorithm CONTINLL.[28,29] The mean values and standard deviation of the CD-data analysis are tabulated in percent. ARH3-Wildtype n=3 NRMSD >0.05; ARH3-p.Thr79Pro n=3 NRMSD >0.08 (NRMSD=Normalized root-mean-square deviation). (D-I) Thermal stabilities of ARH3-Wildtype and ARH3-Thr79Pro measured via the thermofluor assay. Measured using 6 µg protein, ROX filter (575-602nm) and SYPRO orange dye. (D) Thermofluor melt curves of recombinant ARH3-Wildtype and ARH3-p.Thr79Pro. Average melting temperatures: ARH3-Wildtype= 44.79 °C; ARH3-p.Thr79Pro= 28.37 °C. (E) Thermofluor melt curves of recombinant ARH3-Wildtype with different ligands (100 µM). Average melting temperatures: ARH3-Wildtype= 45.74 °C; ARH3-Wildtype+ATP=44.57°C; ARH3-Wildtype+ribose-5-phosphate=45.75 °C; ARH3-Wildtype+ADPr= 54.52°C. (F) Normalized and (G) derivative ARH3-Wildtype melt curve in the presence of adenosine diphosphate ribose (ADPr). Average melting temperatures: ARH3-Wildtype= 43.87 °C; ARH3-Wildtype+ADPr= 54.59 °C. (H) Normalized and (I) derivative ARH3-p.Thr79Pro melt curve. Average melting temperatures: ARH3-p.Thr79Pro= 28.37 °C; ARH3-p.Thr79Pro + ADPr= 31.05 °C.

Because the c.1000C>T (p.Gln334ter) mutation of family 1 was in the last exon, we first excluded nonsense-mediated decay (NMD) of the mutant mRNA. We collected skin biopsies from the father (III-II) and two affected individuals (II-IV-6 and II-IV-7), generated primary fibroblasts, and then performed RT-PCR by using primers designed to amplify the last three exons of *ADPRHL2* (Figure 2.6). The father's and affected individuals' cells revealed a band of the expected size and of similar intensity to that of a healthy control individual, arguing against NMD, as expected. However, when we used an antibody recognizing amino acids 231–245, lysates derived from the affected individuals showed no detectable ARH3 (Figure 2.3C), consistent with a null effect of the truncating mutation. Further, western blot analysis of individual II-2 from family 2 showed an absence of the protein, as predicted by the early stop codon. Fibroblasts from individual II-1 from family 3 showed a severely reduced amount of ARH3 (Figure 2.3C), consistent with the thermal instability of this mutant protein (Figure 2.5D-F) and the severe alteration of its secondary structure (Figure 2.5B and C).

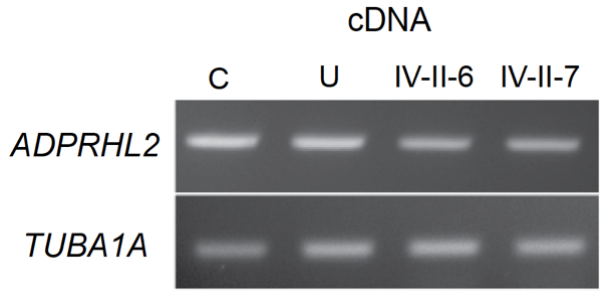

Figure 2.6: RT-PCR analysis for family 1.
RT-PCR results from fibroblasts from unrelated control (C), unaffected carrier Father (U), and affected individuals (IV-II-6 and IV-II-7) from Family 1 shows normal *ADPRHL2* mRNA levels.

2.4.4 Premature Death and Locomotor Defects in Drosophila Parg Mutants

Whereas humans have two known genes with specific PARG activity (*PARG* and *ADPRHL2*; Figure 2.7A), *Drosophila* have a single gene that regulates this process: *Parg*. Using the Gal4-UAS system to drive RNAi expression, we found

that *Parg* knockdown led to a 60% decrease in total *Parg* mRNA for flies with the ubiquitous *da* promoter and a 50% decrease with the neuron-specific promoter, *elav* (*embryonic lethal abnormal visual system*) (Figure 2.8A). Whereas the *da*-Gal4 and $Parg^{RNAi}$ lines showed normal survival, crossing the two together led to *daughterless* (*da*)-mediated expression of $Parg^{RNAi}$, which reduced survival substantially (Figure 2.8B). Ubiquitous knockdown of *Parg* also led to decreased survival when animals were exposed to stress with either hydrogen peroxide (H_2O_2) in their water or environmental hypoxia (2% O_2) (Figure 2.8C and D). Furthermore, knockdown of *Parg* specifically in neurons largely recapitulated this phenotype by using the same two environmental stressors (Figure 2.8E and F). These data provide evidence that stress leads to premature death in the absence of *Parg* and that neurons play an important role in this phenotype.

However, lethality of these flies was not as severe as in the $Parg^{27.1}$ line, which carries a p-element insertion that deletes two-thirds of the open reading frame (nucleotides 34,622–36,079 of GenBank: Z98254)[14], suggesting that $Parg^{RNAi}$ is partially inactivating. These mutant flies with *Parg* loss of function lack the protein Parg and show elevated amounts of PAR, especially in nervous tissue.[14] Mutant flies die in larval stages, but 25% of the animals survive when grown at the permissive 29°C temperature. These adult flies display progressive neurodegeneration, reduced locomotion, and reduced lifespan,[14] consistent with the individuals' phenotypes in our families. We confirmed lethality of the $Parg^{27.1}$ line and found that forced expression of *Drosophila Parg* under the ubiquitous *da* promoter in the mutant background increased both survival and motor activity as measured by an established "climbing index" (Figure 2.7B and C).[30] Likewise, expression of the human *ADPRHL2* under the same *da* promoter showed a nearly

identical degree of rescue of both survival and locomotor activity (Figure 2.7B and C). These results suggest that human *ADPRHL2* is a functional paralog of *Drosophila Parg*.

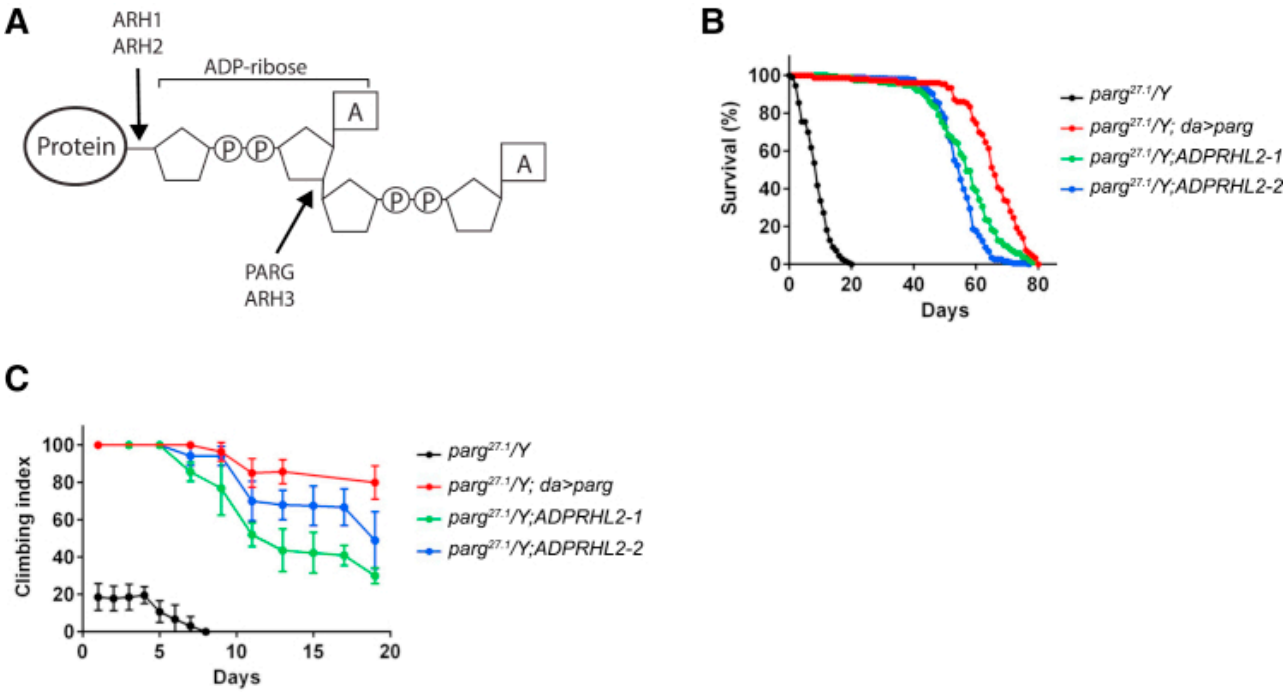

Figure 2.7: Premature death and locomotor defects in *Drosophila parg* mutants are rescued by human *ADPRHL2*.
Schematic of a poly-ADP-ribosylated protein and the location of cleavage. *PARG* and *ADPRHL2* both remove poly-ADP-ribose (PAR) from proteins and cleave the same site. *Drosophila melanogaster* has one PAR-removing enzyme, Parg. (B) *Parg*$^{27.1}$ mutant flies (black) show a severe climbing defect, which was rescued by ubiquitous forced expression of *Parg* (red) or mis-expression of human *ADPRHL2* in two different transgenic lines (green and blue). (C) *Parg*$^{27.1}$ mutant flies (black) displayed decreased survival, which was rescued with ubiquitous forced expression of *Parg* (red) and two different transgenic lines expressing human *ADPRHL2* (green and blue). Data represent the mean ± SEM of eight experiments.

We next tested whether this phenotype might be ameliorated by inhibition of protein PARylation. We reasoned that the requirement for dePARylation should be reduced by the blockage of stress-induced PARylation. Minocycline displays PARP inhibitory activity with an IC_{50} of 42 nM in humans[31] and is well tolerated in flies.[32] We fed flies with a range of concentrations from 0 to 1 mg/mL minocycline for 24 hr before stress and measured survival

rates 96 hr after stress induction. Drug treatment of flies with ubiquitous knockdown of *Parg* revealed a dose-dependent partial rescue of the lethality (Figure 2.7G). This rescue was also seen when the drug was given to flies with neuron-specific knockdown of *Parg* (Figure 2.7H), providing evidence that PARP inhibition can rescue lethality *in vivo*. Although we expect that the effect of minocycline on survival in this assay was due to its effect on PARP, we cannot exclude off-target or non-specific effects.[32]

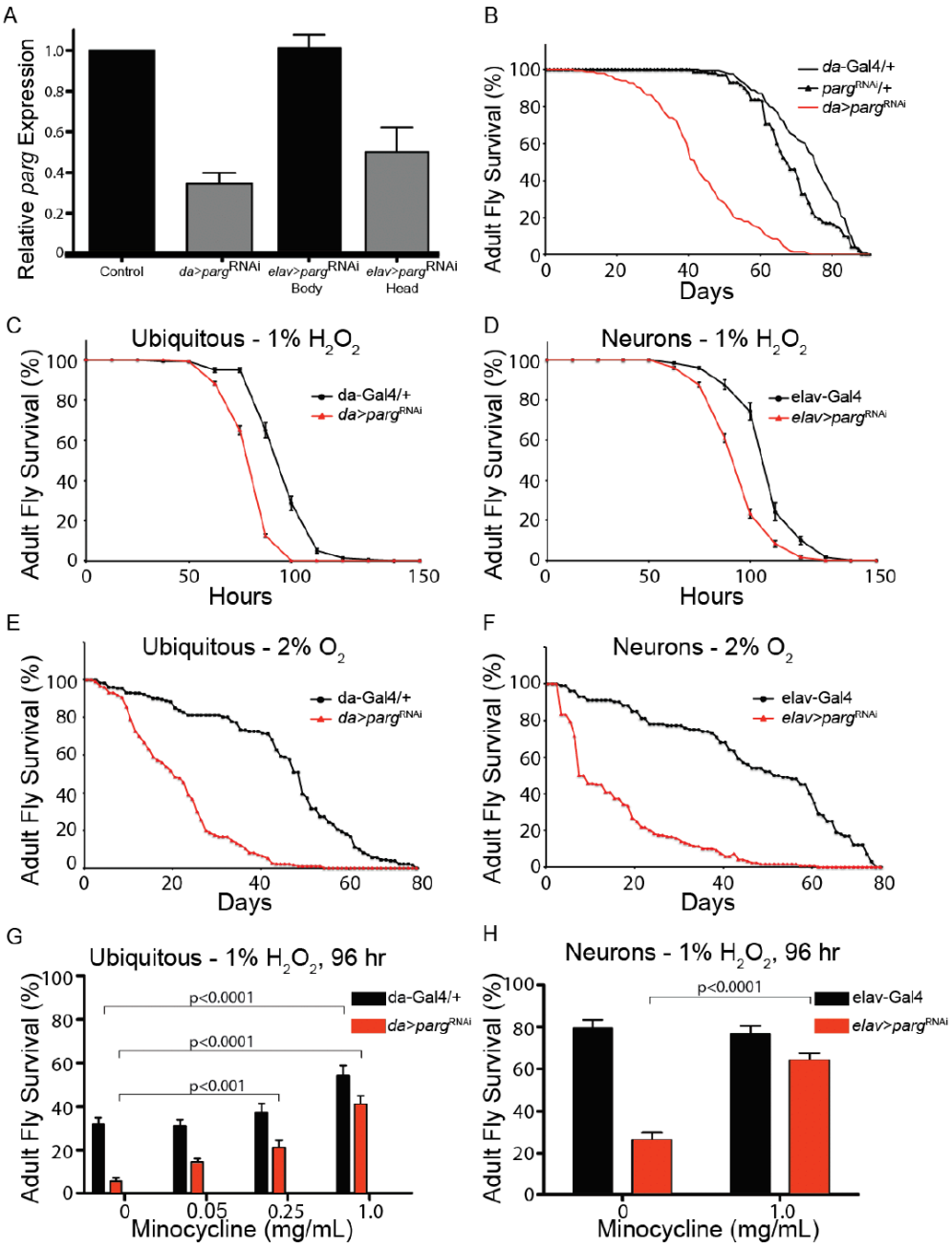

Figure 2.8: *Drosophila* mutants display increased survival with PARP inhibition.
Relative *parg* expression in ubiquitous knockdown animals under the *daughterless* (*da*) promoter shows that RNAi leads to an ~60% in *parg* mRNA levels. Relative *parg* expression in neuron-specific knockdown animals under the *elav* promoter shows that RNAi leads to an ~50% in *parg* mRNA levels in the head of the animal only. (B) Mutant flies with a ubiquitous decrease in *parg* by RNAi die prematurely (red) compared with parental stocks (black). (C,E) Flies with ubiquitous knockdown of *parg* under the *daughterless* (*da*) promoter display decreased survival when exposed to oxidative stress or low oxygen. (D,F) Flies with knockdown of *parg* under the neural *elav* promoter display decreased survival when exposed to oxidative stress or low oxygen. (G,H) Flies with ubiquitous and neuronal knockdown of *parg* can show increased survival in the presence of hydrogen peroxide when pre-treated with Minocycline. Kaplan–Meier Log-rank Test. Data is mean ± s.e.m. of n=8 experiments.

2.5 Discussion

Given that PARP inhibitors are currently in trials for various types of cancer, it is possible that these drugs could be tested for clinical effectiveness in this orphan disease, where they could have a positive effect. Potentially clinically relevant PARP inhibitors include (1) minocycline, an FDA-approved tetracycline derivative that displays PARP inhibitory activity; (2) dihydroisoquinoline (DPQ), a non-FDA-approved potent PARP-1 inhibitor used in experimental research; and (3) veliparib (ABT: 888), a potent PARP-1 and PARP-2 inhibitor currently in clinical trials for the treatment of various type of cancers (IC_{50} = 42, 37, and 4.4 nM, respectively).[31,33]

The extent to which *ADPRHL2* and *PARG* functionally diverge or converge is not well understood, partly because of a lack of detailed comparative expression analysis and biochemical function. PARG demonstrates greater specific activity than ARH3 for removing PAR from proteins,[8] and loss of *Parg* in mice is embryonically lethal.[13] Together, these data suggest that *PARG* is likely to be the major contributor to PAR removal in cells that express both genes under basal conditions. One possibility is that *ADPRHL2* acts as a backup for *PARG* to remove excessive PAR moieties under stress conditions. This would be consistent with the clinical pesentation of individuals with loss of *ADPRHL2*, where phenotypes appear to be induced by environmental stress. Recent studies have shown that ARH3 acts on a recently discovered form of Ser-ADP ribosylation.[34] For example, studies have illustrated an excessive accumulation of Ser-poly-ADP-ribosylated enzymes in *ADPRHL2*$^{-/-}$ cell lines and that ARH3 acts mainly on Ser-ADPr removal.[35] This would be consistent with the phenotype we see in subjects with loss of ARH3, where phenotypes do not emerge until environmental stress insults are encountered. Finally, ARH3 contains a mitochondrial localization signal, and thus another possibility is that

ARH3 functions as a mitochondrial-specific glycohydrolase that is required after the induction of oxidative stress.[13]

PAR signaling has been shown to play a role in a number of cellular processes—including the regulation of transcription, telomere function, mitotic spindle formation, intracellular trafficking, and energy metabolism—in addition to apoptosis-inducing-factor (AIF)-mediated apoptosis.[2,3] Although we hypothesize that the disease mechanism is through cell death, it is possible that PAR accumulation could affect other cellular processes before this. Further work is needed to characterize these effects in the context of this disease.

Chapter 2, in full, is a reprint of the material as it appears in American Journal of Human Genetics 2018. Ghosh, Shereen G.; Becker, Kerstin; Huang, He; Dixon-Salazar, Tracy; Chai, Guoliang; Salpietro, Vincenzo; Al-Gazali, Lihadh; Waisfisz, Quinten; Wang, Haicui; Vaux, Keith K.; Stanley, Valentina; Manole, Andreea; Akpulat, Ugur; Weiss, Marjan, M; Efthymiou, Stephanie; Hanna, Michael G.; Minetti, Carlo; Striano, Pasquale; Pisciotta, Livia; De Grandis, Elisa; Altmuller, Janine; Nurnberg, Peter; Thiele, Holger; Yis, Uluc; Okur, Tuncay D.; Polat, Ayse I.; Amiri, Nafise; Doosti, Mohammad; Karimani, Ehsan G.; Toosi, Mehran B.; Haddad, Gabriel; Karakaya, Mert; Wirth, Brunhilde; van Hagen, Johanna M.; Wolf, Nicole I.; Maroofian, Reza; Houlden, Henry; Cirak, Sebahattin; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this paper.

2.6 References

1. Paul O. Hassa, Sandra S. Haenni, Michael Elser, and Michael O. Hottiger. Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? *Microbiol Mol Biol Rev*, 2006.

2. Xin Luo and W. Lee Kraus. On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. *Genes Dev* 2012.

3. Valérie Schreiber, Françoise Dantzer, Jean-Christophe Ame, and Gilbert de Murcia. Poly(ADP-ribose): novel functions for an old molecule. *Nat Rev Mol Cell Biol* 2006.

4. Mike De Vos, Valérie Schreiber, and Françoise Dantzer. The diverse roles and clinical relevance of PARPs in DNA damage repair: current state of the art. *Biochem Pharmacol*, 2012.

5. Zhifeng Wang, Fengli Wang, Tieshan Tang, and Caixia Guo. The role of PARP1 in the DNA damage response and its application in tumor therapy. *Front Med*, 2012.

6. Shaida A. Andrabi, No Soo Kim, Seong-Woon Yu, Hongmin Wang, David W. Koh, Masayuki Sasaki, Judith A. Klaus, Takashi Otsuka, Zhizheng Zhang, Raymond C. Koehler, Patricia D. Hurn, Guy G. Poirier, Valina L. Dawson, and Ted M. Dawson. Poly(ADP-ribose) (PAR) polymer is a death signal. *Proc Natl Acad Sci USA*, 2006.

7. Yingfei Wang, Valina L. Dawson, and Ted M. Dawson. Poly(ADP-ribose) signals to mitochondrial AIF: a key event in parthanatos. *Exp Neurol*, 2009.

8. Shunya Oka, Jiro Kato, and Joel Moss. Identification and characterization of a mammalian 39-kDa poly(ADP-ribose) glycohydrolase. *J Biol Chem*, 2006.

9. Marc F. Poitras, David W. Koh, S.W. Wu, Shaida A. Andrabi, Allen S. Mandir, Guy G. Poirier, Valina L. Dawson, and Ted M. Dawson . Spatial and functional relationship between poly(ADP-ribose) polymerase-1 and poly(ADP-ribose) glycohydrolase in the brain. *Neuroscience*, 2007.

10. Susan Magdaleno, Patricia Jensen, Craig L. Brumwell, Anna Seal, Karen Lehman, Andrew Asbury, Tony Cheung, Tommie Cornelius, Diana M. Batten, Christopher Eden, Shannon M. Norland, Dennis S. Rice, Nilesh Dosooye, Sundeep Shakya, Perdeep Mehta, and Tom Curran. BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol*, 2006.

11. Marc Niere, Stefan Kernstock, Friedrich Koch-Nolte, and Mathias Ziegler. Functional localization of two poly(ADP-ribose)-degrading enzymes to the mitochondrial matrix. *Mol Cell Biol*, 2008.

12. Marc Niere, Masato Mashimo, Line Agledal, Christian Dolle, Atsushi Kasamatsu, Jiro Kato, Joel Moss, and Mathias Ziegler. ADP-ribosylhydrolase 3 (ARH3), not poly(ADP-ribose) glycohydrolase (PARG) isoforms, is responsible for degradation of mitochondrial matrix-associated poly(ADP-ribose). *J Biol Chem*, 2012.

13. David W. Koh, Ann M. Lawler, Marc F. Poitras, Masayuki Sasaki, Sigrid Wattler, Michael C. Nehls, Tobias Stoger, Guy G. Poirier, Valina L. Dawson, and Ted M. Dawson. Failure to degrade poly(ADP-ribose) causes increased sensitivity to cytotoxicity and early embryonic lethality. *Proc Natl Acad Sci USA*, 2004.

14. Shuji Hanai, Masayuki Kanai, Sayaka Ohashi, Keiji Okamoto, Mitsunori Yamada, Hitoshi Takahashi, and Masanao Miwa. Loss of poly(ADP-ribose) glycohydrolase causes progressive neurodegeneration in Drosophila melanogaster. *Proc Natl Acad Sci USA*, 2004.

15. Mareie Butepage, Laura Eckei, Patricia Verheugd, and Bernhard Luscher. Intracellular mono-ADP-ribosylation in signaling and disease. *Cells,* 2015.

16. Nicolas C Hoch, Hana Hanzlikova, Stuart L Rulten, Martine Tétreault, Emilia Komulainen, Limei Ju, Peter Hornyak, Zhihong Zeng, William Gittens, Stephanie A Rey, Kevin Staras, Grazia M S Mancini, Peter J McKinnon, Zhao-Qi Wang, Justin D Wagner, Care4Rare Canada Consortium, Grace Yoon, and Keith W Caldecott. XRCC1 mutation is associated with PARP-1 hyperactivation and cerebellar ataxia. *Nature,* 2017.

17. Sarah Shaw Murray, Arnold Oliphant, Richard Shen, Celeste McBride, Rhoberta J Steeke, Stuart G Shannon, Todd Rubano, Bahram G Kermani, Jian-Bing Fan, Mark S Chee, and Mark S T Hansen. A highly informative SNP linkage panel for human genetic studies. *Nat Methods*, 2004.

18. Katrin Hoffman and Tom H. Lindner. easyLINKAGE-Plus--automated linkage analyses using large-scale SNP data. *Bioinformatics*, 2005.

19. Andreas Gnirke, Alexandre Melnikov, Jared Maguire, Peter Rogov, Emily M. LeProust, William Brockman, Timothy Fennell, Georgia Giannoukos, Sheila Fisher, Carsten Russ,

Stacey Gabriel, David B. Jafe, Eric S. Lander, and Chad Nusbaum. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol*, 2009.

20. Tracy J. Dixon-Salazar, Jennifer L. Silhavy, Nitin Udpa, Jana Schroth, Stephanie Bielas, Ashleigh E. Schaffer, Jesus Olvera, Vineet Bafna, Maha S. Zaki, Ghada H. Abdel-Salam, Lobna A. Mansour, Laila Selim, Sawsan Abdel-Hadi, Naima Marzouki, Tawfeg Ben-Omran, Nouriya A. Al-Saana, F. Müjgan Sonmez, Figen Celep, Matloob Azam, Kiley J. Hill, Adrienne Collazo, Ali G. Fenstermaker, Gaia Novarino, Naiara Akizu, Kiran V. Garimella, Carrie Sougnez, Carsten Russ, Stacey B. Gabriel, and Joseph G. Gleeson. Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med*, 2012.

21. Salem Alawbathani, Amit Kawalia, Mert Karakaya, Janine Altmüller, Peter Nurnberg, and Sebahattin Cirak. Late diagnosis of a truncating WISP3 mutation entails a severe phenotype of progressive pseudorheumatoid dysplasia. *Cold Spring Harb Mol Case Stud*, 2018.

22. Jacob Villegas and Michael McPhaul. Establishment and culture of human skin fibroblasts. *Curr Protoc Mol Biol Chapter 28*, 2005.

23. Christoph Mueller-Dieckmann, Stefan Kernstock, Michael Lisurek, Jens Peter von Kries, Friedrich Haag, Manfred S. Weiss, and Friedrich Koch-Nolte. The structure of human ADP-ribosylhydrolase 3 (ARH3) provides insights into the reversibility of protein ADP-ribosylation. *Proc Natl Acad Sci USA*, 2006.

24. Georg Gietzl, Doris Chen, Frank Schnorrer, Kuan-Chung Su, Yulia Barinova, Michaela Fellner, Beate Gasser, Kaolin Kinsey, Silvia Oppel, Susanne Scheiblauer, Africa Couto, Vincent Marra, Krystyna Keleman, and Barry J. Dickson. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature*, 2007.

25. Lina Nashef, J.F. Annegers, and S.W. Brown. Introduction and overview. Sudden unexpected death in epilepsy. *Epilepsia*, 1997.

26. Jan E. Leestma, John F. Annegers, Martin J. Brodie, Stephen Brown, Paul Schraeder, David Siscovick, Braxton B. Wannamaker, Patricia S. Tennis, Mark A. Cierpial, and Nancy L. Earl. Sudden unexplained death in epilepsy: observations from a large clinical development program. *Epilepsia*, 1997.

27. Jana Marie Schwarz, Christian Rödelsperger, Markus Schuelke, and Dominik Seelow. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods*, 2010.

28. Lee Whitmore and B.A. Wallace. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*, 2004.

29. Narasimha Sreerama and Robert W. Woody. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem*, 2000.

30. Surya T. Madabattula, Joel C. Strautman, Andrew M. Bysice, Julia A. O'Sullivan, Alaura Androschuk, Cory Rosenfelt, Kacy Doucet, Guy Rouleau, and Francois Bolduc. Quantitative analysis of climbing defects in a Drosophila model of neurodegenerative disorders. *J Vis Exp*, 2015.

31. Conrad C. Alano, Tiina M. Kauppinen, Andreu Viader Valls, and Raymond A. Swanson. Minocycline inhibits poly(ADP-ribose) polymerase-1 at nanomolar concentrations. *Proc Natl Acad Sci USA,* 2006.

32. Gang Jun Lee, Jin Ju Lim, and Seogana Hyun. Minocycline treatment increases resistance to oxidative stress and extends lifespan in *Drosophila* via FOXO. *Oncotarget*, 2017.

33.. Cherrie K. Donawho, Yan Luo, Yanping Luo, Thomas D. Penning, Joy L. Bauch, Jennifer J. Bouska, Velitchka D. Bontcheva-Diaz, Bryan F. Cox, Theodore L. DeWeese, Larry E. Dillehay, Debra C. Ferguson, Nayereh S. Ghoreishi-Haack, David R. Grimm, Ran Guan, Edward K. Han, Rhonda R. Holley-Shanks, Boris Hristov, Kenneth B. Idler, Ken Jarvis, Eric F. Johnson, Lawrence R Kleinberg, Vered Klinghofer, Loren M. Lasko, Xuesong Liu, Kennan C. Marsh, Thomas P. McGonigal, Jonathan A. Meulbroek, Amanda M. Olson, Joann P. Palma, Luis E. Rodriguez, Yan Shi, Jason A. Stavropoulos, Alan C. Tsurutani, Gui-Dong Zhu, Saul H. Rosenberg, Vincent L. Giranda, and David J. Frost. ABT-888, an orally active poly(ADP-ribose) polymerase inhibitor that potentiates DNA-damaging agents in preclinical tumor models. *Clin Cancer Res*, 2007.

34. Xiaoxing Feng, Yiran Zhou, Alicia M. Porter, Mandi M. Hopkins, Mengwei Liu, and David W. Koh. Silencing of Apoptosis-Inducing factor and poly(ADP-ribose) glycohydrolase reveals novel roles in breast cancer cell death after chemotherapy. *Mol Cancer*, 2012.

34. Pietro Fontana, Juan Jose Bonfiglio, Luca Palazzo, Edward Bartlett, Ivan Matic, and Ivan Ahel. Serine ADP-ribosylation reversal by the hydrolase ARH3. *eLife*, 2017.

35. Luca Palazzo, Orsolya Leidecker, Evgeniia Prokhorova, Helen Dauben, Ivan Matic, and Ivan Ahel. Serine is the major residue for ADP-ribosylation upon DNA damage. *eLife*, 2018.

CHAPTER 3

Recurrent homozygous damaging mutation in *TMX2*, encoding a protein disulfide isomerase, in four families with microlissencephaly

3.1 Abstract

Protein disulfide isomerase (PDI) proteins are part of the thioredoxin protein superfamily. PDIs are involved in the formation and rearrangement of disulfide bonds between cysteine residues during protein folding in the endoplasmic reticulum and are implicated in stress response pathways. In this study, we recruited eight children from four consanguineous families residing in distinct geographies within the Middle East and Central Asia for study. All probands showed structurally similar microcephaly with lissencephaly (microlissencephaly) brain malformations. DNA samples from each family underwent whole exome sequencing, assessment for repeat expansions and confirmatory segregation analysis. An identical homozygous variant in *TMX2* (c.500G>A), encoding thioredoxin-related transmembrane protein 2, segregated with disease in all four families. This variant changed the last coding base of exon 6, and impacted mRNA stability. All patients presented with microlissencephaly, global developmental delay, intellectual disability and epilepsy. While *TMX2* is an activator of cellular *C9ORF72* repeat expansion toxicity, patients showed no evidence of *C9ORF72* repeat expansions.

The *TMX2* c.500G>A allele associates with recessive microlissencephaly, and patients show no evidence of *C9ORF72* expansions. *TMX2* is the first PDI implicated in a recessive disease, suggesting a protein isomerisation defect in microlissencephaly.

3.2 Introduction

Lissencephaly (LIS, lissos means smooth in Greek) refers to a smooth surface of the cerebral cortex.[1] Previously termed 'type I' and 'type II' to refer to the absence or presence of coexistent microcephaly,[1] respectively, recent lissencephaly molecular classification has identified eight recurrently mutated genes in recessive forms of disease: *RELN* (MIM:600514), *NDE1* (MIM:609449), *LAMB1* (MIM:150240), *KATNB1* (MIM:602703), *CDK5* (MIM:123831), *TMTC3* (MIM:617218).[2-7] Primary microcephaly (MPCH) refers to reduced head size with reduced cerebral volume; and at the more severe end of the spectrum, there is often a reduction in the complexity of the folding of the cerebral cortex, termed 'simplified gyral pattern.'[8] Therefore, most cases with microcephaly show only mild disruption of the gyral folding on brain MRI, but at the severe end of microcephaly spectrum the cerebral cortex can show gyri paucity, and can be difficult to distinguish from forms of lissencephaly. A distinct entity termed 'microlissencephaly' describes the combination of lissencephaly with microcephaly.[9] Of the recessive causes for lissencephaly only a few show notable additional microcephaly: *RELN, NDE1, KATNB1,* and a few chromosomal deletion syndromes.[2,3,5,10] This suggests the existence of genetic conditions in which the degree of lissencephaly is greater than would be expected from the degree of microcephaly alone, supporting the use of the term microlissencephaly for appropriate conditions.

Some genes linked to microcephaly with lissencephaly, such as *TMTC3*, are implicated in the endoplasmic reticulum (ER) stress response, which is activated upon misfolding of secretory proteins or calcium balance perturbation.[11,12] Overload of ER protein folding capacity can lead to an accumulation of misfolded proteins, ER stress and subsequent cell death. Protein disulfide isomerases (PDIs) are resident transmembrane ER proteins that catalyze thiol-disulfide

interchanges, which are critical for proper protein folding.[13] There are 21 known genes encoding PDIs, including three members of the Thioredoxin-related transmembrane (TMX) family, TMX-1 (MIM:610527), −2 (MIM:616715), and −3 (MIM:616102). Each TMX family member encodes an N-terminal signal peptide, single-pass transmembrane domain, C-terminal thioredoxin (Trx)-like domain, and a Di-lysine ER retention signal domain, but how they maintain protein homeostasis in the ER remains unknown.[14]

3.3 Materials and Methods

3.3.1 Patient Recruitment

The procedures followed for recruitment and data collection were in accordance with the ethical standards of the responsible committee on human experimentation at the University of California, San Diego.

3.3.2 DNA Extraction and Whole Exome Sequencing

This study was approved by the Institutional Review Board at the respective host institution. All study participants and/or their guardians signed informed consents, and the study was performed in accordance with Health Insurance Portability and Accountability Act (HIPAA) Privacy Rules. DNA was extracted from peripheral blood leukocytes with salt extraction. DNA from the probands was subjected to Agilent Sure-Select Human All Exon v2.0 (44Mb target) and Illumina Rapid Capture Enrichment (37Mb target) library preparation and sequenced on Illumina HiSeq 2000 or 4000 instruments.[15]

3.3.3 Computational Analysis

Variant calling and filtering were performed according to a previously described whole exome sequencing pipeline.[15] Variants were filtered if not present in all affected individuals in

the family, with zygosity based upon the presumed mode of inheritance. For all families, variants were filtered for minor allele frequency (MAF) >1:1000, PolyPhen-2 scores of <0.9 or GERP score <4.5. Variants were assessed with MutationTaster, and runs of homozygosity were defined with HomozygosityMapper.[16-18]

3.3.4 Sanger sequencing

Sanger sequencing of PCR products covering exon 6 of *TMX2* was performed using standard procedures. Sequencing tracks were visualized with SnapGene Viewer (SnapGene).

3.3.5 GTEx and Brainspan

Data was obtained from the GTEx and BrainSpan (© 2010 BrainSpan Atlas of the Developing Human Brain. Available from brainspan.org) portals and processed using Python (3.64) with Pandas (0.22.0).

3.3.6 Repeat-primed PCR

Repeat-primed PCR was performed using 100ng of genomic DNA from the following individuals: 1673-III-1, 2525-III-1, and 3501-III-7 as previously described.[19]

3.3.7 TMX2 mRNA Assessment

Total RNA was extracted from fresh frozen blood on available parents and probands using TRIzol™, quantified by spectrophotometry, and reverse-transcribed using the Superscript III First-Strand cDNA Kit (Invitrogen). PCR analysis of cDNA was performed using dHPLC-purified primers, visualized using the BioRad GelDoc™ and quantified by Azure™ software.

3.4 Results

3.4.1 Identification of Four Consanguineous Families Segregating with Microlissencephaly

We studied eight affected individuals with neurological phenotypes consistent with microlissencephaly from four unrelated consanguineous families. Family 1673 presented with one affected child (1673-III-1) from a first-cousin marriage. The individual was born at full term with a significantly reduced head circumference (−2SD) and exhibited generalized tonic-clonic (GTC) seizures from birth (Figure 3.1A, Supplemental File 2). At her last evaluation, she exhibited severe intellectual disability, hypertonia, absence of motor skills, and increased deep tendon reflexes. Brain MRI revealed a thickened cerebral cortical mantle, with diminished gyral folds, with an appearance of pachygyria/lissencephaly, cortical atrophy, corpus callosum hypogenesis, and ventriculomegaly (Figure 3.2A and B).

Family 2525 had two affected children (2525-III-2, and 2525-III-3) from a consanguineous marriage (Figure 3.1B). Both children were born full term with significantly reduced head circumference (−3SD and −2SD, respectively). They also exhibited GTC seizures in the neonatal period with an onset of 2 weeks and 3 weeks, respectively. At their last evaluation, they were diagnosed with severe intellectual disability, increased deep tendon reflexes, delayed gross motor development, and absence of fine motor and language development (Supplemental File 2). Brain MRI assessment of the older sibling revealed lissencephaly, reduced white matter volume, cortical atrophy, corpus callosum hypogenesis, and ventriculomegaly. (Figure 3.2C and D).

Family 4984 had three affected children (4984-III-1, 4984-III-3, and 4984-III-4) and one healthy sibling from a consanguineous marriage (Figure 3.1C). The oldest affected child was born at full term with a reduced head circumference (−1SD). He was also diagnosed with severe intellectual disability, delayed gross motor and language development, absent fine motor development, and epilepsy. He died at 5 years of age due to pneumonia. His younger two

affected siblings (4984-III-3, and 4984-III-4) were both born at 36 weeks and showed reduced head circumference at last evaluation (−4SD and −3SD, respectively). They exhibited similar symptoms as their older sibling. Brain MRIs were obtained from one child (4984-III-3), and showed lissencephaly, corpus callosum hypoplasia, reduced white matter volume, ventriculomegaly, and cerebellar atrophy (Figure 3.2E and F).

Family 3501 had two affected children (3501-III-6 and 3501-III-7) and five healthy siblings from a consanguineous marriage (Figure 3.1D). Both affected individuals were born at full term with significantly reduced head circumferences (−2SD). They also exhibited neonatal epilepsy, which predated their diagnoses of severe intellectual disability, delayed/absent psychomotor development, and increased tendon reflexes at last evaluation. Brain MRIs for both children showed lissencephaly, cerebral mantle thickening, ventriculomegaly, reduced white matter volume, cerebellar atrophy, and corpus callosum hypoplasia (Figure 3.2G and H).

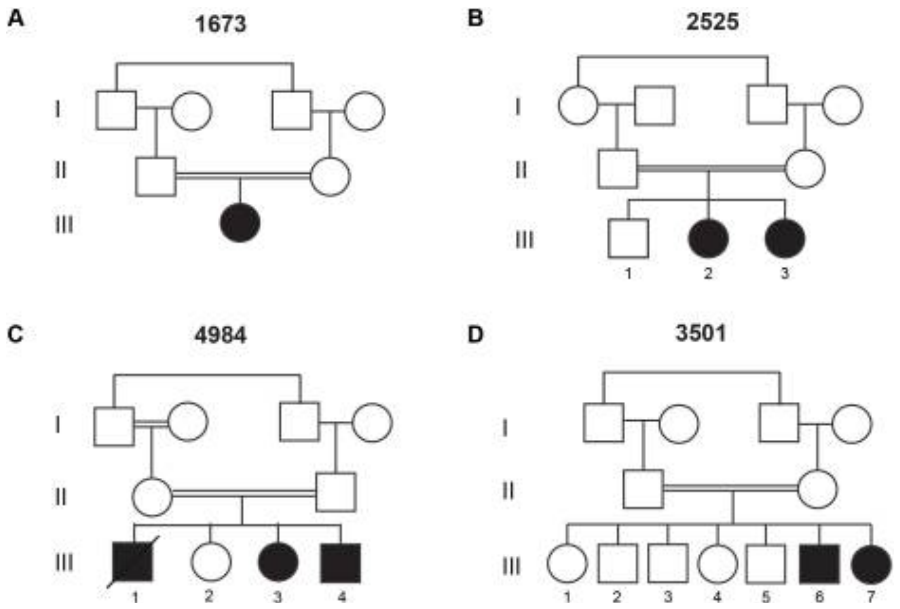

Figure 3.1: Consanguineous families with a homozygous recessive mutation in *TMX2*.
(A-D) Pedigrees of all four families showing first-cousin consanguineous marriages (double bar) with a total of eight affected children. All unfilled members are without neurological disease.

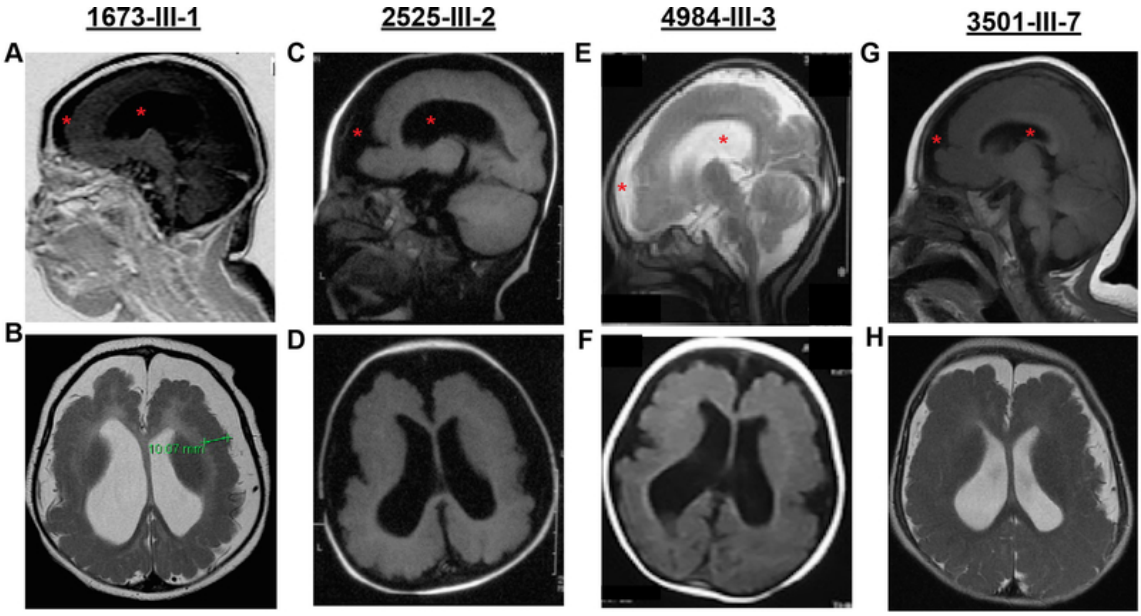

Figure 3.2: MRIs from affected children display lissencephaly, microcephaly, and brain atrophy.
Top row: Midline saggital MRI. Bottom row: Axial MRI. Subject ID provided at top. All images are T1-weighted ecept (B, E, H), which are T2 weighted, selected to best demonstrate the main imaging defects. The lack of cortical folding is apparent in both frontal and occipital regions, consistent with lissencephaly. Red asterisk in top row highlights ventriculomegaly and increased extra-axial fluid accumulation, consistent with brain atrophy and microcephaly. Bottom row shows paucity of cortical gyri and thickened gray matter, which was measured at 10.07mm (green bracket) in (B), which is roughly twice the width of normal.

3.4.2 Identification of a Homozygous TMX2 c.500G>A Variant in Families with Microlissencephaly

Whole exome sequencing (WES) was performed on blood-derived DNA from the affected child in Family 1673. For variant identification Genome Analysis Toolkit (GATK) workflow identified variants that were intersected with identity-by-descent blocks from HomozygosityMapper.[18,20] Rare potentially deleterious variants were prioritized against an in-house exome database consisting of over 4,000 ethnically matched individuals, in addition to publicly available exome datasets, cumulatively numbering over 10,000 individuals. From this analysis, we identified three variants that passed these filter parameters in the genes *RAB3GAP2* (MIM:609275), *TMX2,* and *MOCOS* (MIM:613274). Patient phenotypes did not match the phenotypes reported for *RAB3GAP2* and *MOCOS,* leaving a single, homozygous hg38:Chr11:g.57739039G>A (GenBank: NM_001347898.1) variant in *TMX2*, located at the last base of exon six of eight (c.500G>A), as the sole candidate. We then performed WES on one affected child each from Families 2525 and 4984, and both affected children from Family 3501, based on matching phenotypic features between the families. Interestingly, we identified the same *TMX2* variant recurring in all three families. Further, each of the four consanguineous families showed a run of homozygosity that contained *TMX2*, consistent with the parental consanguinity (Figure 3.3), although the haplotype blocks showed no commonality of shared SNPs of any size (Figure 3.4). This suggests that the variant arose independently in each family on separate haplotypes, or was so distantly derived that a common haplotype could not be identified.

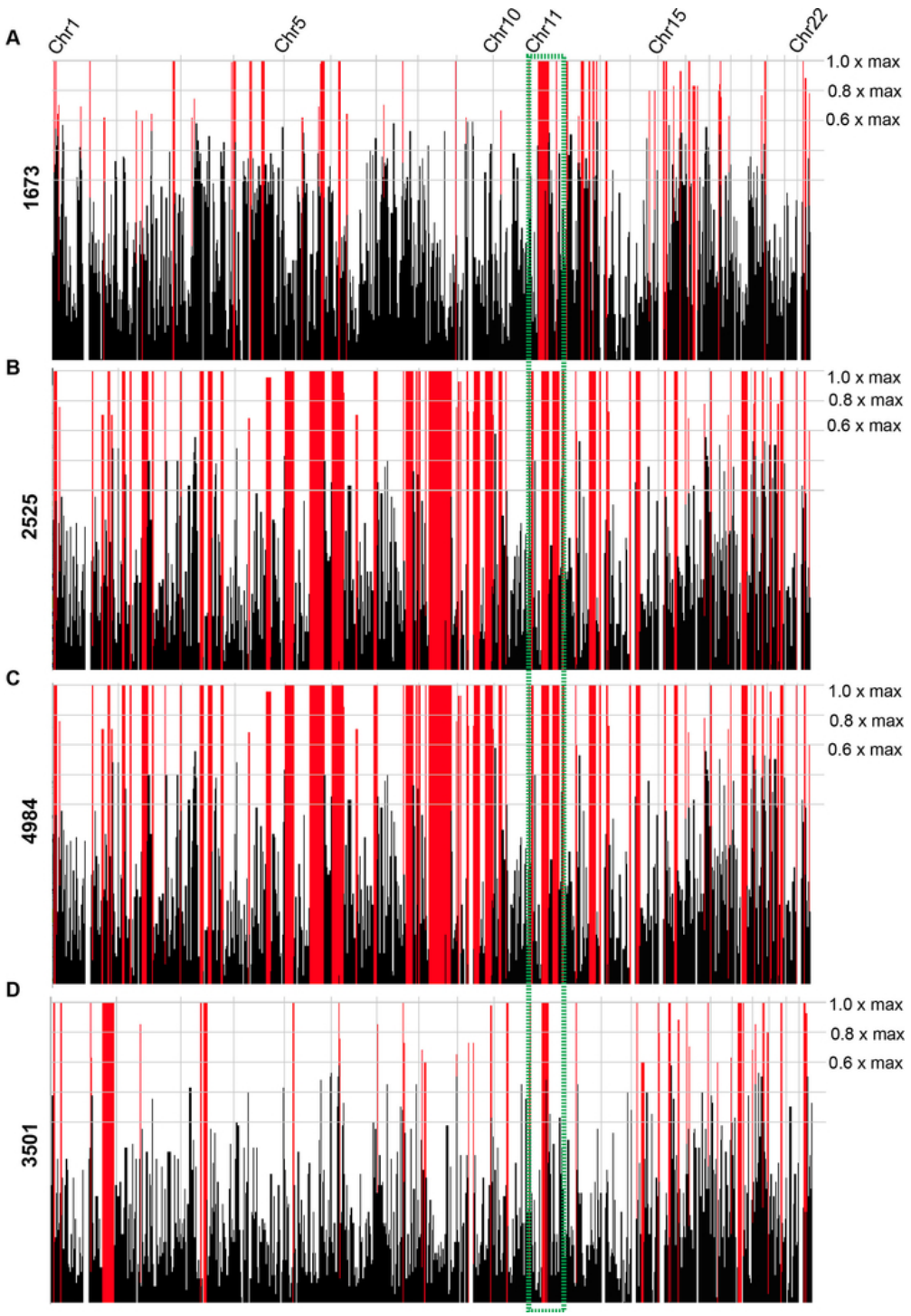

Figure 3.3: Homozygosity plots for all four families display homozygosity surrounding the *TMX2* gene.

(A-D) Graphical representation of regions of homozygosity, generated by HomozygosityMapper for all 22 autosomes for the affected individual in Family 1673 (III-1, A), one affected individual from Family 2525 (III-2, B), one affected individual from Family 4984 (III-3, C), and both affected individuals from Family 3501 (III-6 and III-7, D). Red peaks indicate homozygosity scores above the given cutoff at 0.8x of the maximum. All plots show a region of homozygosity containing the *TMX2* gene (green box).

					Patient ID				
CHR	POS	RS	REF	ALT	1673-III-1	2525-III-2	4984-III-3	3501-IV-6	3501-IV-7
11	57369951	rs2276038	C	T	1/1	0/0	0/0	0/0	0/0
11	57379543	rs540687	A	G	1/1	1/1	1/1	1/1	1/1
11	57380702	rs669661	A	G	1/1	1/1	1/1	1/1	1/1
11	57380725	rs116078737	C	T	0/0	0/0	1/1	0/0	0/0
11	57387528	rs536455	G	A	1/1	1/1	0/0	1/1	1/1
11	57389932	rs490358	A	G	1/1	1/1	1/1	1/1	1/1
11	57423979	rs113953954	C	A	0/0	0/0	1/1	0/0	0/0
11	57467940	rs11229040	G	A	0/0	0/0	1/1	0/0	0/0
11	57542834	rs80123614	A	G	1/1	0/0	0/0	0/0	0/0
11	57545996	rs12223229	C	G	0/0	1/1	0/0	0/0	0/0
11	57611697	rs2511988	A	G	0/0	0/0	1/1	0/0	0/0
11	57696028	rs1783978	T	C	0/0	1/1	1/1	1/1	1/1
11	**57739039**	**rs370455806**	**G**	**A**	**1/1**	**1/1**	**1/1**	**1/1**	**1/1**
11	57796519	rs10896644	C	T	1/1	0/0	0/0	0/0	0/0
11	57803760	rs11229137	C	T	1/1	1/1	1/1	1/1	1/1
11	58031251	rs921135	A	G	0/0	1/1	0/0	1/1	1/1
11	58031470	rs2513726	A	G	0/0	1/1	0/0	1/1	1/1
11	58031711	rs7123727	C	T	0/0	0/0	0/0	0/0	0/0
11	58031899	rs1374570	G	C	1/1	0/0	0/0	0/0	0/0
11	58118518	rs78218902	C	G	0/0	1/1	0/0	0/0	0/0
11	58179691	rs111629598	G	A	1/1	0/0	0/0	0/0	0/0
11	58179792	rs11229273	C	G	0/0	1/1	0/0	0/0	0/0
11	58191025	rs34337292	T	C	0/0	0/0	1/1	0/0	0/0
11	58191303	rs7120468	C	T	1/1	1/1	0/0	0/0	0/0
11	58191324	rs73470064	C	T	0/0	0/0	1/1	0/0	0/0

Figure 3.4: Haplotype blocks of the four families containing the *TMX2* c.500G>A mutation. Haplotypes from all individuals resulting from WES. Single nucleotide polymorphisms (SNPs) are highlighted based on their family (1673-III-1: red, 2525-III-2: orange, 4984-III-3: blue, 3501-IV-6: pink, 3501-IV-7: green). The shared region is boxed. CHR: chromosome. POS: coordinate position on chromosome. RS: reference SNP. REF: reference allele. ALT: alternate allele. 0/0: homozygous for the REF allele. 0/1: heterozygous for the ALT allele. 1/1: homozygous for the ALT allele.

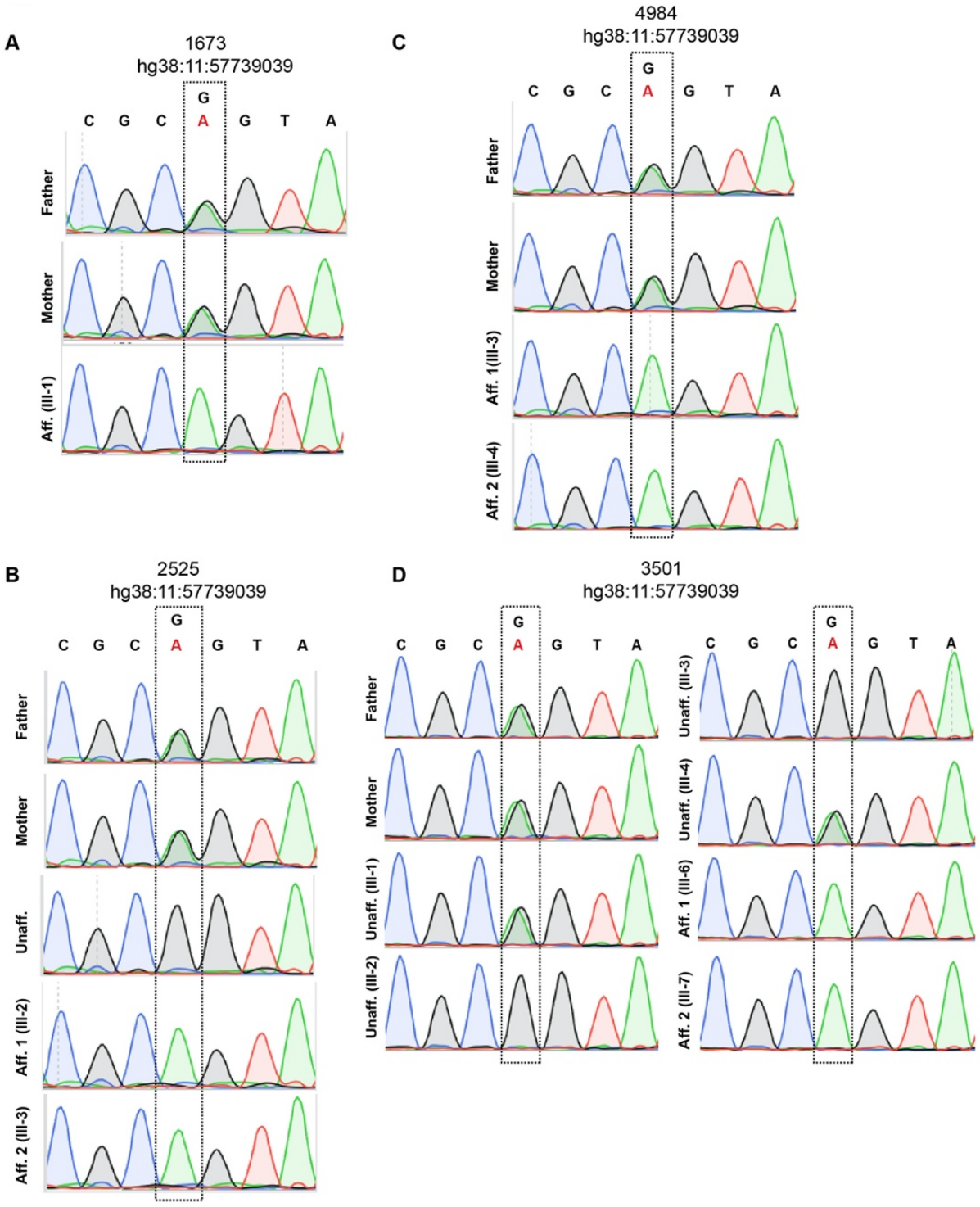

Figure 3.5: Sanger sequencing traces for all four families.
(A-D) Sanger sequencing traces for Families 1673 (A), 2525 (B), 4984 (C), and 3501 (D) depicting the G>A conversion resulting in the missense variant. Shown are traces for father, mother, unaffected siblings (Unaff.) and affecteds (Aff.) depicting the G>A conversion. All parents are heterozygous and the affected child is homozygous for the variant. Two unaffected siblings are heterozygous (III-1 and III-4), while the other two are homozygous wildtype (III-2 and III-3) in Family 3501.

This variant was confirmed by direct Sanger sequencing analysis, and testing of all available family members demonstrated segregation with the phenotype according to a recessive mode of inheritance (Figure 3.5). We found no additional families in our own cohort of over 5000 individuals, or in the GeneMatcher website, of patients displaying likely causative variants in this gene. The variant was also not identified in the Greater Middle Eastern (GME) Variome (consisting of 2,497 individuals) or any other public database.[21] These analyses made *TMX2* the top candidate for pathogenesis.

TMX2 encodes for Thioredoxin-related transmembrane protein 2, and the variant introduced an amino acid substitution in the Trx-like domain of the protein (p.Arg205Gln, Figure 3.6A and B). The p.Arg205Gln substitution was also predicted to be disease-causing by MutationTaster.[17] TMX2 has two protein-coding isoforms, isoform 1 and 2 (NP_001334827.1 and NP_057043.1), which differ in the length of the Extracellular Topological Domain (TAD) and absence of the transmembrane domain in isoform 2 (Figure 3.6B), but functional differences between these isoforms are not reported.

As this homozygous variant occurred in the last nucleotide of exon 6 (Figure 3.6D), we considered a potential impact on *TMX2* splicing or mRNA stability. In addition to the two canonical protein-coding isoforms, *TMX2* is also predicted to undergo alternative splicing at the exon 6–7 junction, which is a particularly small intron (Figure 3.6C). Approximately 95% of observed transcripts from NCBI conform to a 91 bp intron, with the amino acid sequence reading …**DVSTRYKVS**…. One minor isoform, representing approximately 2% of transcripts, contains a 69 bp intron, with the sequence reading …**DVSTR**YELSGPCR**YKVS**…, due to the use of an alternative 3' splice acceptor site. The other minor isoform, representing approximately 3% of transcripts, conforms to a 102 bp intron, with the sequence reading …**DV**QSE… (Figure 3.6E),

yet contains a premature stop codon approximately 26 amino acids downstream, suggesting a non-functional isoform. Whether or not these isoforms are found in nature or encode for protein has yet to be elucidated.

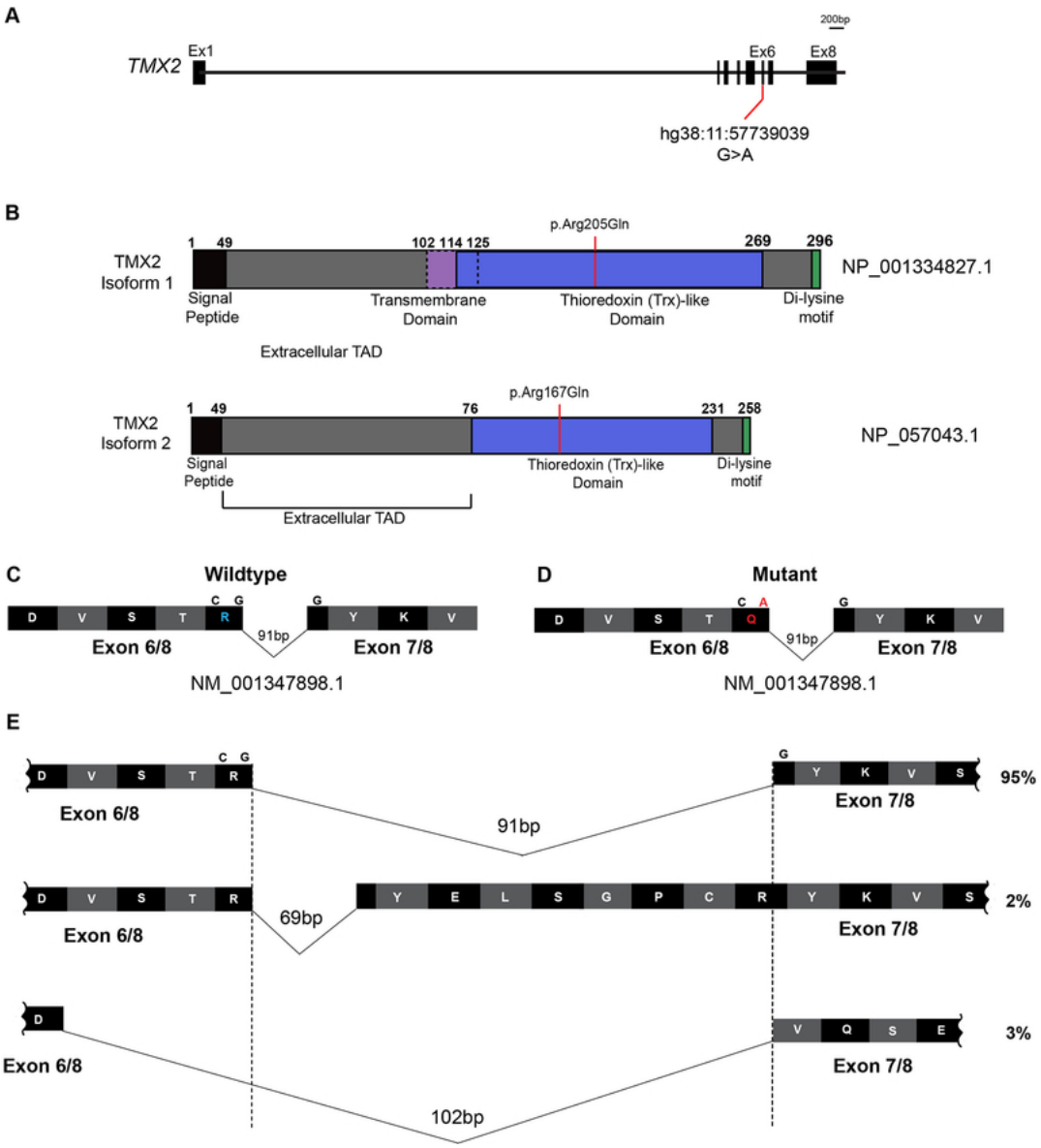

Figure 3.6: *TMX2* gene and predicted protein isoforms.
(A) Schematic of *TMX2* depicting the genomic map spanning eight exons. Red line indicates the position of the identified variant in exon six in all four families and coordinates within the cDNA map (RefSeq NM_001347898.1). (B) Schematic of the two protein-coding isoforms of Thioredoxin-related transmembrane protein 2 (TMX2) depicting the location of the 'signal peptide' (black), extracellular topological associated domain (TAD), transmembrane domain (purple), Thioredoxin (Trx)-like domain (blue), and Di-lysine motif (green). Isoform 2 is 38 amino acids shorter than isoform 1 due to the absence of the transmembrane domain and part of the extracellular TAD. Red line indicates the position and coordinates of the amino acid change in affected individuals from all four families within the Trx-like domain. (C-D) Illustration of partial wildtype (C) and mutant (D) residues from exons six and seven surrounding the variant shown in blue and red in wildtype and mutant, respectively. (E) Schematic of the affitional splice isoforms of the exon6-7 junction of *TMX2*, based on the UCSC Genome Browser with predicted percent existence of each isoform in human tissues from GenBank. The percentage of each isoform was calculated based on the average number of Expressed Sequence Tags (ESTs) available, displaying that particular isoform (e.g. BI764082, BG170274, and BI561678, respectively).

3.4.3 The TMX2 c.500G>A Mutation Results in Reduced mRNA Levels in Affected Individuals

TMX2 is ubiquitously expressed in adult human tissues (GTEx data, Figure 3.7). Additionally, *TMX2* is expressed in human brain during development at all assessed stages (Figure 3.8), consistent with a role in this phenotype.

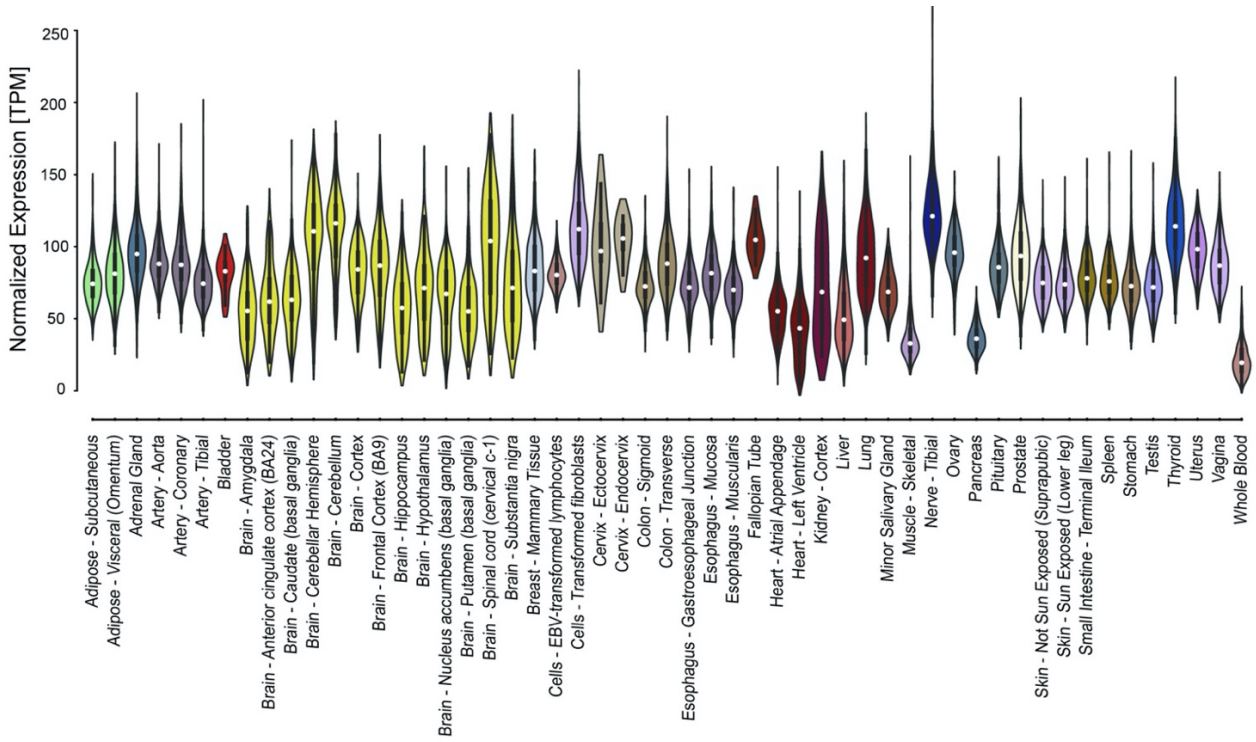

Figure 3.7: *TMX2* expression across various tissues in adult humans (GTEx Expression Data). Expression analysis of *TMX2* transcripts across various adult tissues, using data derived from the GTEx portal. *TMX2* shows ubiquitous expression across all tissues. Expression is shown as Transcripts per Kilobase Million (TPM). Shown are standard violin plots with an inner box plot.

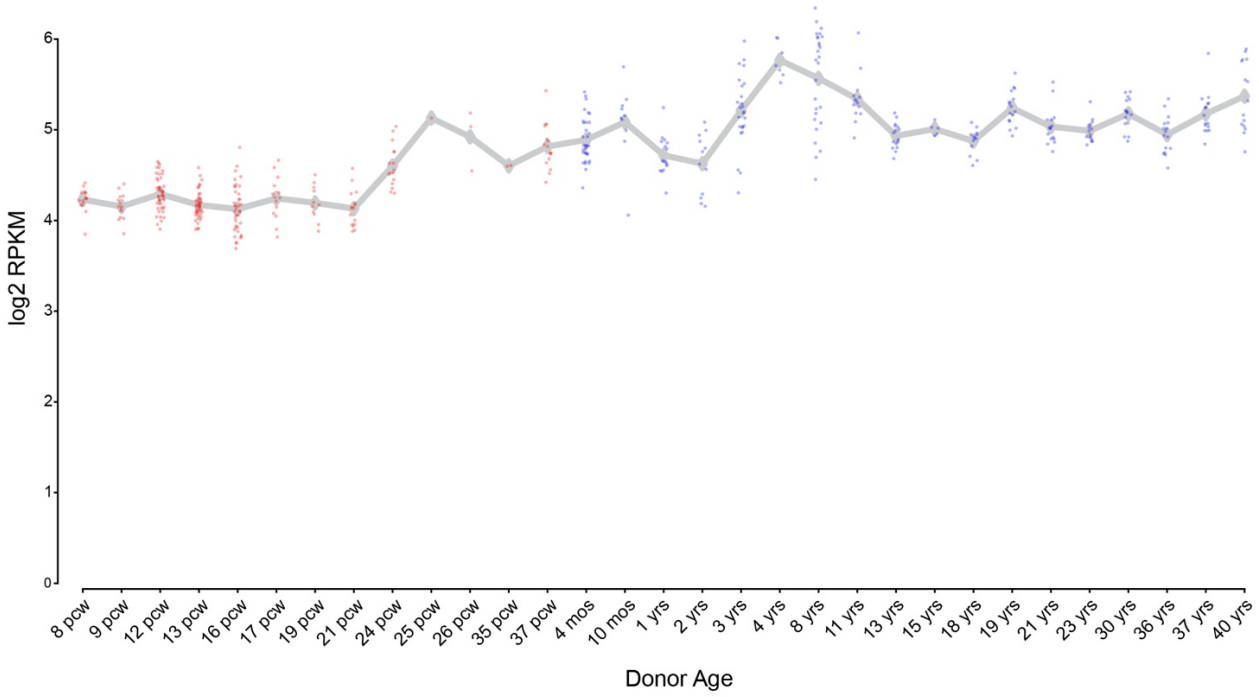

Figure 3.8: *TMX2* expression in human brain across development (BrainSpan data). Expression analysis of *TMX2* transcripts in developing and adult brain, using data derived from RNA-Sequencing data from BrainSpan. *TMX2* shows ubiquitous expression across all time points. Expression is shown as Reads Per Kilobase of transcript, per Million mapped reads (RPKM). The mean ± standard deviation (SD) and individual data points are shown for each time point. Pcw: post-conception weeks; mos: months; yrs: years.

Because the variant alters the last base of an exon, we considered whether splicing or RNA stability could be impacted. Thus, RNA was extracted from frozen whole blood obtained from the father and one affected individual from Family 2525 as well as both parents and one affected from Family 4984. mRNA transcript levels were assessed using RT-PCR across various regions flanking the variant (Figure 3.9A) and compared to an unrelated control sample. Three primer pairs were designed: 1] the region between exons five and seven, 2] the region between exon six and seven, 3] the region between exons six and eight (Figure 3.9B). We did not observe evidence of any novel splice isoforms, such as might result from exon skipping or intron inclusion; however, *TMX2* mRNA levels were reduced to 15–50% of control in both samples from affecteds compared with the healthy control, whereas unaffected obligate carriers showed

levels between 50–90% of normal (Figure 3.9C and D). The reduced mRNA levels in patients and carriers is consistent with an effect of *TMX2* c.500G>A on splicing or RNA stability, and the absence of novel bands in the affected and carrier individuals suggests nonsense mediated decay of aberrantly spliced mRNA.

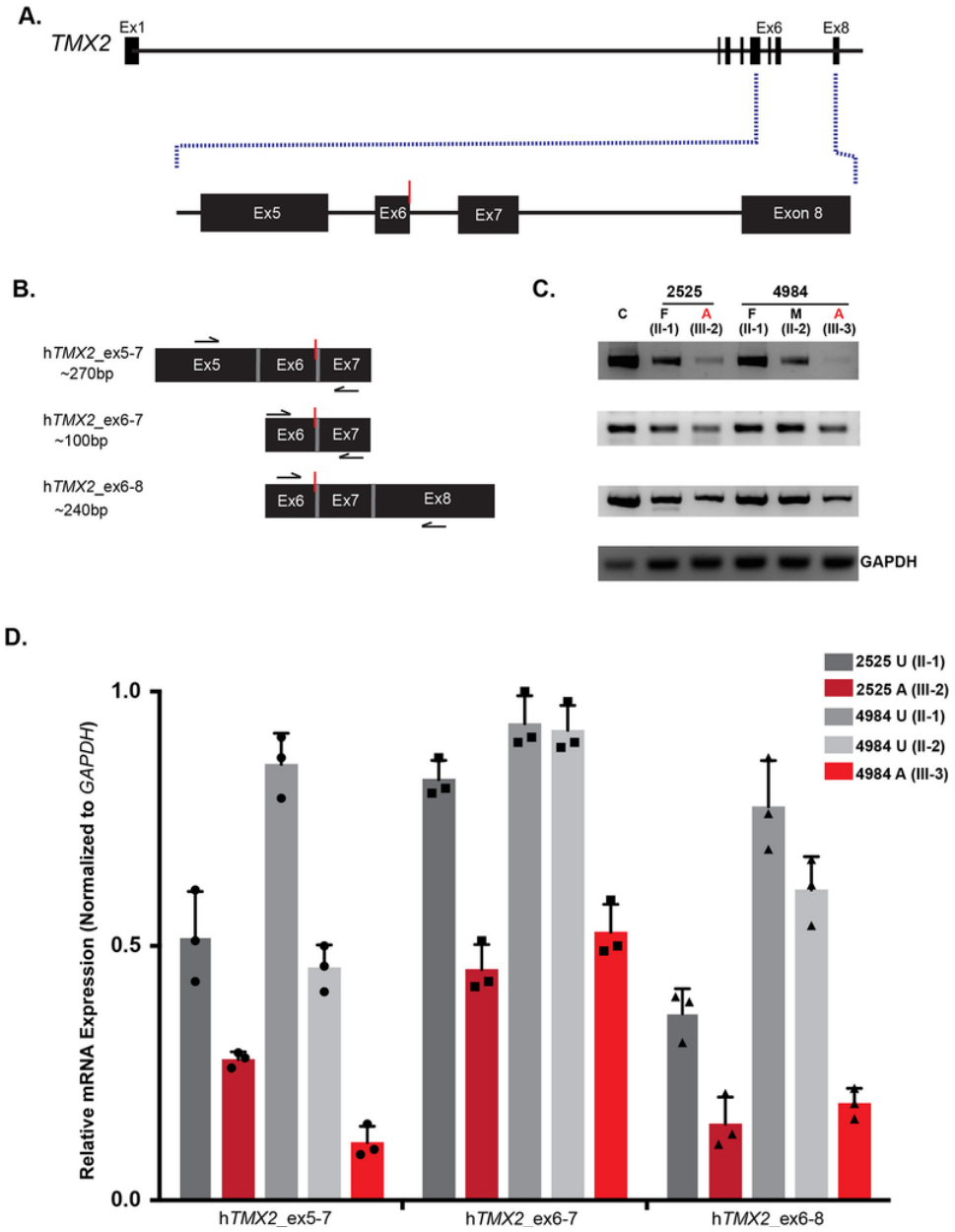

Figure 3.9: *TMX2* c.500G>A allele associates with reduced mRNA levels in carriers and affected individuals.

Illustration of *TMX2* with an expanded view of exons five through eight. Red line indicates the location of the variant. (B) Schematic of primer pairs designed for RT-PCR analysis and expected base pair sizes for the corresponding products. Black arrows indicate the approximate location of the primers. Red line indicates the location of the variant. (C) RT-PCR analysis of RNA extracted from whole blood derived from an unrelated control (C), the unaffected father (II-1) and one affected individual (III-2) from Family 2525, and both parents (II-1 and II-2) and one affected individual (III-3) from Family 4984. There was reduction of *TMX2* mRNA levels in affected individuals (red) compared to the normalized expression of 1.0 in the unrelated control (n=3). *TMX2* levels of parent carriers (grey) were intermediate compared with unrelated control. *GAPDH* was used as a loading control. (D) Quantification of RT-PCR results displaying relative mRNA expression normalized to *GAPDH*. n=3 independent experiments. Bars show mean ± standard deviation (SD) and individual data points.

3.4.4 Lack of C9ORF72 Repeat Expansion in TMX2 Mutated Samples

C9ORF72 gene mutations are the most common cause of Amyotrophic Lateral Sclerosis (ALS, MIM:105400) and Frontotemporal Dementia (FTD, MIM:600274).[22,23] The damaging mutation in *C9ORF72* associated with these diseases is an expanded noncoding hexanucleotide repeat (GGGGCC), leading to repeat-associated non-ATG (RAN) translation of a dipeptide repeat protein (DPR).[22,23] Previous transcriptomic analysis using *C9ORF72*-mutant ALS brain tissue revealed upregulation of genes involved in ER stress,[24] suggesting failed ER stress response in *C9ORF72* pathogenesis. Following a CRISPR-Cas9 KO screen for genetic modifiers of DPR toxicity, *TMX2* emerged as one of the two strongest protective modifier genes identified.[19] In these experiments, loss of TMX2 suppressed the toxic effect of the *C9ORF72* DPR, suggesting that TMX2 is an activator of DPR toxicity. Therefore, we wanted to test the possibility that the *TMX2* mutation occurred in the genetic background of *C9ORF72* expansion.

To test this, we performed repeat-primed PCR using a previously validated detection method to screen all available DNA samples from three families for the presence of the GGGGCC hexanucleotide repeat expansion (Family 3501 failed analysis due to insufficient amounts of DNA available).[23] All remaining samples were absent for the expanded *C9ORF72* repeat, as compared to the negative control and an unrelated positive control sample (Figure 3.10)

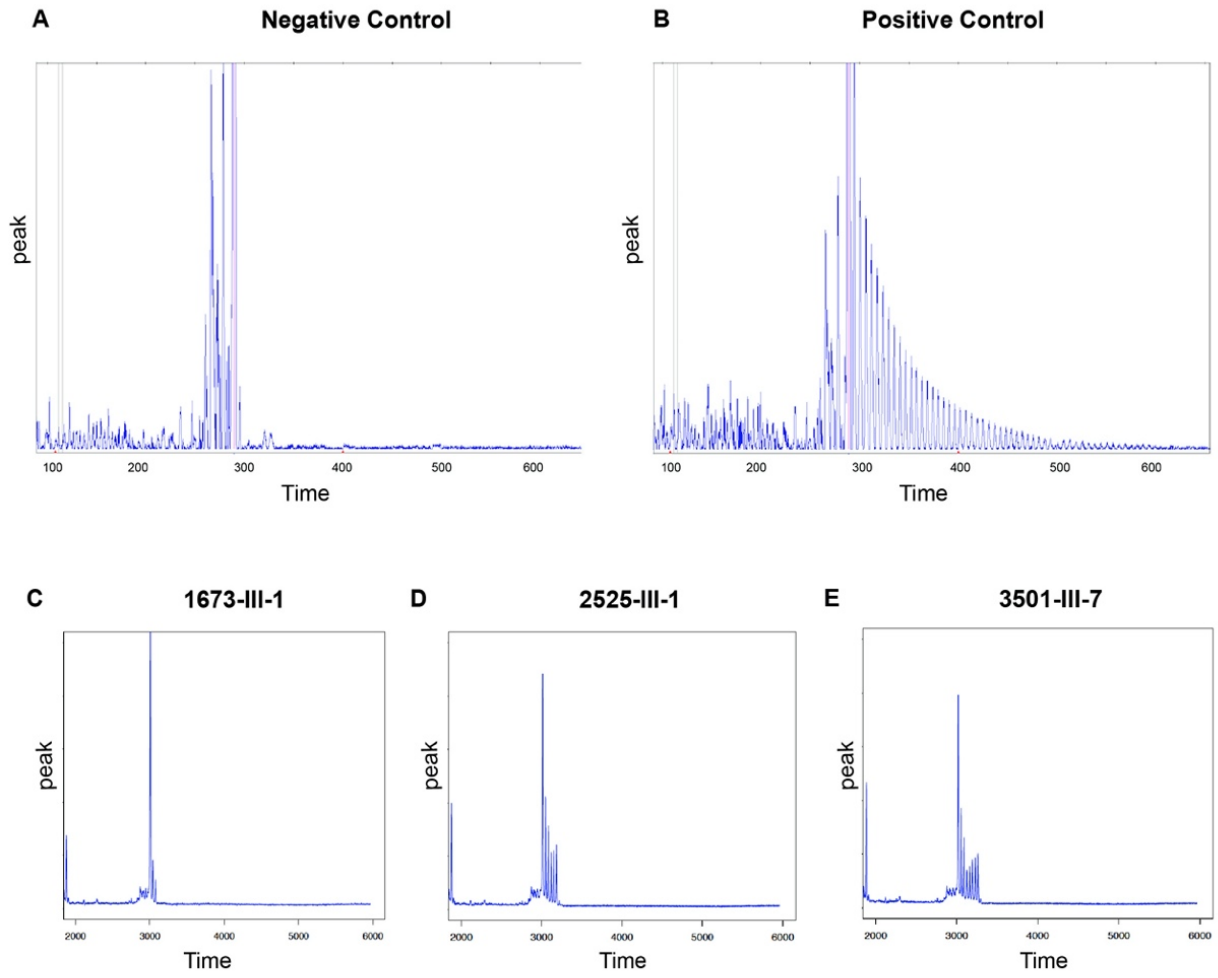

Figure 3.10: Repeat-primed PCR for families 1673, 2525, and 3501 depicts absence of *C9ORF72* repeat expansion.

(A) Negative control (B) Positive control (C-E) Repeat-primed PCR analysis on genomic DNA from three affected individuals from Families 1673 (III-1, C), 2525 (III-1, D), and 3501 (III-7, E). Family 4984 was not included in the analysis due to insufficient amounts of DNA available.

3.5 Discussion

Here we demonstrate a recurrent variant in *TMX2* associated with a severe, recessive neurological disease characterized by microcephaly with lissencephaly. All affected individuals in four families carried an identical homozygous c.500G>A variant in *TMX2* and displayed microlissencephaly, global developmental delay, intellectual disability, and epilepsy. The variant led to decreased mRNA steady-state levels, and was not associated with expansion in

the *C9ORF72* hexanucleotide repeat. Our results implicate a relatively common single allele as a cause for microlissencephaly in humans.

TMX2 is arguably the least studied member of the PDI family. *TMX2* was cloned in 2003, and encodes a multidomain transmembrane protein that is enriched on the mitochondria-associated membrane (MAM) of the ER via palmitoylation of two of its cystolically-exposed cysteines.[13,25] The MAM is a domain of the ER that mediates the exchange of ions, lipids and metabolites between the ER and mitochondria, suggesting a potential function in either mitochondria, ER, or both. TMX2 shows ubiquitous expression, with highest levels detected in heart, brain, liver, kidney and pancreas.[12,13,26] Whether TMX2 displays catalytic activity remains unclear, given that the canonical CXXC motif required for oxioreductase activity is replaced by an SXXC motif.[27]

There was one prior mention of *TMX2* in human disease, where a cohort of patients with malformations of cortical development was sequenced to identify causes.[28] One family was identified with two children affected with lissencephaly where two candidate genes emerged, one being *TMX2*.[28] In this family, *TMX2* showed compound heterozygous mutations [(c.326A>G; p.Asp109Gly) and (c.691C>T; p.Arg231Trp)], but brain MRI was not available to allow comparisons with our cases. Our results, linking *TMX2* to lissencephaly, suggest that, in the previous family, *TMX2* was likely the relevant gene.

Mutations in genes that encode ER stress response proteins have been implicated in a range of neuroprogressive disorders among other human diseases, including Parkinson's disease, Huntington's disease, and ALS, the latter of which has been studied in connection with ER stress response pathways and apoptosis.[29-32] *TMX2* knockdown or CRISPR targeting was able to rescue the cell death and ER stress response toxicity induced by application of the DPR encoded by

the *C9ORF72* expansion. CRISPR targeting of *TMX2* in DPR-expressing human cells showed upregulation of prosurvival unfolded-protein-response-pathway genes including *Atf3*, and downregulation of calcium-binding and apoptotic proteins.[19] We found no evidence of expansion of the *C9ORF72* repeat in patient cells, suggesting that loss of *TMX2* was not secondary to effects of *C9ORF72,* but it is possible that the mutated *TMX2* allele is under positive selection, as it appears to have arisen independently on four separate haplotypes. It would be interesting as a next step to understand mechanism by generating an animal model of disease; however, initial attempts indicate that homozygous *Tmx2* null mutations in mouse are lethal prior to weaning, with complete penetrance.[33]

The presence of a Trx-like domain in TMX2 suggests a role regulating redox pathways, specifically in relation to the ER stress response. Partial or complete loss of *TMX2* may thus predispose to cell death. While we have not linked the phenotypes from our cases to a molecular function in this study, increased ER stress during neurodevelopment has been linked to aberrant neuronal maturation and the subsequent development of neurodevelopmental disorders caused by abnormal neuronal differentiation and maturation.[34,35] We hypothesize that loss of TMX2 in humans may lead to cell-death mediated depletion of neural progenitor cells in the developing cerebral cortex during embryogenesis, producing microlissencephaly. Whether correct, or whether there are secondary effects of *TMX2* in neurogenesis or neuronal migration, remain to be explored.

Chapter 3, in full, is a reprint of the material as it appears in Journal of Medical Genetics 2019. Ghosh, Shereen G.; Wang, Lu; Breuss, Martin W.; Green, Joshua D.; Stanley, Valentina; Yang, Xiaoxu; Ross, Danica; Traynor, Bryan J.; Alhashem, Amal M.; Azam, Matloob; Selim,

Laila; Bastaki, Laila; Elbastawisy, Hanan I.; Temtamy, Samia; Zaki, Maha S.; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this paper.

3.6 References

1. Andrew E. Fry, Thomas D. Cushion, and Daniela T. Pilz. The genetics of lissencephaly. *Am J Med Genet C Semin Med Genet*, 2014.

2. Maha Zaki, Marwa Shehab, Alice Abd El-Aleem, Ghada Abdel-Salam, Hajira B. Koeller, Yesim Ilkin, M. Elizabeth Ross, William B. Dobyns, and Joseph G. Gleeson. Identification of a novel recessive RELN mutation using a homozygous balanced reciprocal translocation. *Am J Med Genet*, 2007.

3. Fowzan S. Alkuraya, Xuyu Cai, Carina Emery, Ganeshwaran H. Mochida, Mohammed S. Al-Dosari, Jillian M Felie, R. Sean Hill, Brenda J. Barry, Jennifer N Partlow, Generoso G. Gascon, Amal Kentab, Mohammad Jan, Ranad Shaheen, Yuanyi Feng, and Christopher A. Walsh. Human mutations in NDE1 cause extreme microcephaly with lissencephaly. *Am J Hum Genet*, 2011.

4. Farid Radmanesh, Ahmet Okay Caglayan, Jennifer L. Silhavy, Cahide Yilmaz, Vincent Cantagrel, Tarek Omar, Başak Rosti, Hande Kaymakcalan, Stacey Gabriel, Mingfeng Li, Nenad Sestan, Kaya Bilguvar, William B. Dobyns, Maha S. Zaki, Murat Gunel, and Joseph G. Gleeson. Mutations in LAMB 1 cause cobblestone brain malformation without muscular or ocular abnormalities. *Am J Hum Genet*, 2013.

5. Ketu Mishra-Gorur,[*] Ahmet Okay Çağlayan,[*] Ashleigh E. Schaffer, Chiswili Chabu, Octavian Henegariu, Fernando Vonhoff,[7] Gözde Tuğce Akgümüş, Sayoko Nishimura, Wenqi Han, Shu Tu, Burcin Baran, Hakan Gumus, Cengiz Dilber, Maha S. Zaki, Heba AA Hossni, Jean-Baptiste Rivière, Hülya Kayserili, Emily G. Spencer, Rasim O. Rosti, Jana Schroth, Hüseyin Per, Caner Cağlar, Cagri Cağlar, Duygu Dölen, Jacob F. Baranoski, Sefer Kumandaş, Frank J. Minja, E. Zeynep Erson-Omay, Shrikant M. Mane, Richard P. Lifton, Tian Xu, Haig Keshishian, William B Dobyns, Neil C. Chi, Nenad Sestan, Angeliki Louvi, Kaya Bilgüvar, Katsuhito Yasuno, Joseph G. Gleeson, and Murat Günel. Mutations in KATNB1 cause complex cerebral malformations by disrupting asymmetrically dividing neural progenitors. *Neuron*, 2014.

6. Daniella Magen, Ayala Ofir, Liron Berger, Dorit Goldsher, Ayelet Eran, Nasser Katib, Yousif Nijem, Euvgeni Vlodavsky, Shay Tzur, Doron M. Behar, Yakov Fellig, and Hanna

Mandel. Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with a loss-of-function mutation in CDK5. *Hum Genet*, 2015.

7. Julie Jerber, Maha S. Zaki, Jumana Y. Al-Aama, Rasim Ozgur Rosti, Tawfeg Ben-Omran, Esra Dikoglu, Jennifer L. Silhavy, Caner Caglar, Damir Musaev, Beate Albrecht, Kevin P. Campbell, Tobias Willer, Mariam Almuriekhi, Ahmet Okay Çağlayan, Jiri Vajsar, Kaya Bilgüvar, Gonul Ogur, Rami Abou Jamra, Murat Günel, and Joseph G. Gleeson. Biallelic mutations in TMTC3, encoding a transmembrane and TPR-containing protein, lead to cobblestone lissencephaly. *Am J Hum Genet*, 2016.

8. Muhammad Faheem, Muhammad Imran Naseer, Mahmood Rasool, Adeel G. Chaudhary, Taha A. Kumosani, Asad Muhammad Ilyas, Peter Natesan Pushparaj, Farid Ahmed, Hussain A. Algahtani, Mohammad H. Al-Qahtani, and Hasan Saleh Jamal. Molecular genetics of human primary microcephaly: an overview. *BMC Med Genomics*, 2015.

9. A. James Barkovich, Donna M. Ferriero, Rebeca M. Barr, Pierre Gressens, W.B. Dobyns, Charles L. Truwit, and Philipe P. Evrard. Microlissencephaly: a heterogeneous malformation of cortical development. *Neuropediatrics*, 1998.

10. Elena Parrini, Valerio Conti, William B. Dobyns, and Renzo Guerrini. Genetic Basis of Brain Malformations. *Mol Syndromol*, 2016.

11. Maud Racapé, Jean-Paul Duong Van Huyen, Richard Danger, Magali Giral, Françoise Bleicher, Yohann Foucher, Annaïck Pallier, Paul Pilet, Petra Tafelmeyer, Joanna Ashton-Chess, Emilie Dugast, Ségolène Pettré, Béatrice Charreau, Jean-Paul Soulillou, and Sophie Brouard. The involvement of SMILE/TMTC3 in endoplasmic reticulum stress response. *PLoS One*, 2011.

12. Toshiyuki Nakagawa, Hong Zhu, Nobuhiro Morishima, En Li, Jin Xu, Bruce A. Yankner, and Junying Yuan. Caspase-12 mediates endoplasmic-reticulum-specific apoptosis and cytotoxicity by amyloid-beta. *Nature*, 2000.

13. Xianfang Meng, Chun Zhang, Jinzhong Chen, Shuying Peng, Yaoqiong Cao, Kang Ying, Yi Xie, and Yumin Mao. Cloning and identification of a novel cDNA coding thioredoxin-related transmembrane protein 2. *Biochem Genet*, 2003.

14. Motonori Matsusakim Shingo Kanemura, Misaki Kinoshita, Yuong-Ho Lee, Kenji Inaba, and Masaki Okumura. The Protein Disulfide Isomerase Family: from proteostasis to pathogenesis. *Biochim Biophys Acta Gen Subj*, 2019.

15. Tracy J Dixon-Salazar, Jennifer L. Silhavy, Nitin Udpa, Jana Schroth, Stephanie Bielas, Ashleigh E. Schaffer, Jesus Olvera, Vineet Bafna, Maha S. Zaki, Ghada H. Abdel-Salam, Lobna A. Mansour, Laila Selim, Sawsan Abdel-Hadi, Naima Marzouki, Tawfeg Ben-Omran, Nouriya A. Al-Saana, F. Müjgan Sonmez, Figen Celep, Matloob Azam, Kiley J. Hill, Adrienne Collazo, Ali G. Fenstermaker, Gaia Novarino, Naiara Akizu, Kiran V. Garimella, Carrie Sougnez, Carsten Russ, Stacey B. Gabriel, and Joseph G. Gleeson. Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med*, 2012.

16. Dominik Seelow and Markus Schuelke. HomozygosityMapper 2012--bridging the gap between homozygosity mapping and deep sequencing. *Nucleic Acids Res*, 2012.

17. Jana Marie Schwarz, Christian Rödelsperger, Markus Schuelke, and Dominik Seelow. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*, 2010.

18. Dominik Seelow, Markus Schuelke, Friedhelm Hildebrandt, and Peter Nurnberg. HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic Acids Res*, 2009.

19. Nicholas J. Kramer, Michael S. Haney, David W. Morgens, Ana Jovičić, Julien Couthouis, Amy Li, James Ousey, Rosanna Ma, Gregor Bieri, C. Kimberly Tsui, Yingxiao Shi, Nicholas T. Hertz, Marc Tessier-Lavigne, Justin K. Ichida, Michael C. Bassik, and Aaron D. Gitler . CRISPR–Cas9 screens in human cells and primary neurons identify modifiers of *C9ORF72* dipeptide-repeat-protein toxicity. *Nat Genet*, 2018.

20. Mark A. DePristo, Eric Banks, Ryan Poplin, Kiran V. Garimella, Jared R. Maguire, Christopher Hartl, Anthony A. Philippakis, Guillermo del Angel, Manuel A. Rivas, Matt Hanna, Aaron McKenna, Tim J. Fennell, Andrew M. Kernytsky, Andrey Y. Sivachenko, Kristian Cibulskis, Stacey B. Gabriel, David Altshuler, and Mark J. Daly. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, 2011.

21. Eric M. Scott, Anason Halees, Yuval Itan, Emily G. Spencer, Yupeng He, Mostafa Abdellateef Azab, Stacey B. Gabriel, Aziz Belkadi, Bertrand Boisson, Laurent Abel, Andrew G. Clark, Greater Middle East Variome Consortium, Fowzan S. Alkuraya, Jean-Laurent Casanova, and Joseph G. Gleeson. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet*, 2016.

22. Mariely DeJesus-Hernandez, Ian R. Mackenzie, Bradley F. Boeve, Adam L. Boxer, Matt Baker, Nicola J. Rutherford, Alexandra M. Nicholson, Nicole A. Finch, Heather Flynn, Jennifer Adamson, Naomi Kouri, Aleksandra Wojtas, Pheth Sengdy, Ging-Yuek R. Hsiung, Anna Karydas, William W. Seeley, Keith A Josephs, Giovanni Coppola, Daniel H. Geschwind, Zbigniew K. Wszolek, Howard Feldman, David S. Knopman, Ronald C. Petersen, Bruce L. Miller, Dennis W. Dickson, Kevin B. Boylan, Neill R. Graff-Radford, and Rosa Rademakers. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, 2011.

23. Alan E. Renton, Elisa Majounie, Adrian Waite, Javier Simón-Sánchez, Sara Rollinson, J. Raphael Gibbs, Jennifer C. Schymick, Hannu Laaksovirta, John C. van Swieten, Liisa Myllykangas, Hannu Kalimo, Anders Paetau, Yevgeniya Abramzon, Anne M. Remes, Alice Kaganovich, Sonja W. Scholz, Jamie Duckworth, Jinhui Ding, Daniel W. Harmer, Dena G. Hernandez, Janel O. Johnson, Kin Mok, Mina Ryten, Danyah Trabzuni, Rita J. Guerreiro, Richard W. Orrell, James Neal, Alex Murray, Justin Pearson, Iris E. Jansen, David Sondervan, Harro Seelaar, Derek Blake, Kate Young, Nicola Halliwell, Janis Bennion Callister, Greg Toulson, Anna Richardson, Alex Gerhard, Julie Snowden, David Mann, David Neary, Michael A. Nalls, Terhi Peuralinna, Lilja Jansson, Veli-Matti Isoviita, Anna-Lotta Kaivorinne, Maarit Hölttä-Vuori, Elina Ikonen, Raimo Sulkava, Michael Benatar, Joanne Wuu, Adriano Chiò, Gabriella Restagno, Giuseppe Borghero, Mario Sabatelli, ITALSGEN Consortium. David Heckerman, Ekaterina Rogaeva, Lorne Zinman, Jeffrey D. Rothstein, Michael Sendtner, Carsten Drepper, Evan E. Eichler, Can Alkan, Ziedulla Abdullaev, Svetlana D. Pack, Amalia Dutra, Evgenia Pak, John Hardy, Andrew Singleton, Nigel M. Williams, Peter Heutink, Stuart Pickering-Brown, Huw R. Morris, Pentti J. Tienari, and Bryan J. Traynor. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, 2011.

24. Mercedes Prudencio, Veronique V. Belzil, Ranjan Batra, Christian A. Ross, Tania F. Gendron, Luc J. Pregent, Melissa E. Murray, Karen K. Overstreet, Amelia E. Piazza-Johnston, Pamela Desaro, Kevin F. Bieniek, Michael DeTure, Wing C. Lee, Sherri M. Biendarra, Mary D. Davis, Matthew C. Baker, Ralph B. Perkerson, Marka van Blitterswijk, Caroline T. Stetler, Rosa Rademakers, Christopher D. Link, Dennis W. Dickson, Kevin B. Boylan, Hu Li, and Leonard Petrucelli. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci*, 2015.

25. Emily M. Lynes, Michael Bui, Megan C. Yap, Matthew D. Benson, Bobbie Schneider, Lars Ellgaard, Luc G. Berthiaume, and Thomas Simmen. Palmitoylated TMX and calnexin target to the mitochondria-associated membrane. *EMBO J*, 2012.

26. James J. Galligan and Dennis R. Petersen. The human protein disulfide isomerase gene family. *Hum Genomics*, 2012.

27. Lars Ellgaard and Lloyd W. Ruddock. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep*, 2005.

28. Julia Lauer Zillhardt, Karine Poirier, Loïc Broix, Nicolas Lebrun, Adrienne Elmorjani, Jelena Martinovic, Yoann Saillour, Giuseppe Muraca, Juliette Nectoux, Bettina Bessieres, Catherine Fallet-Bianco, Stanislas Lyonnet, Olivier Dulac, Sylvie Odent, Imen Rejeb, Lamia Ben Jemaa, Francois Rivier, Lucile Pinson, David Geneviève, Yuri Musizzano, Nicole Bigi, Nicolas Leboucq, Fabienne Giuliano, Nicole Philip, Catheline Vilain, Patrick Van Bogaert, Hélène Maurey, Cherif Beldjord, François Artiguenave, Anne Boland, Robert Olaso, Cecile Masson, Patrick Nitschké, Jean-François Deleuze, Nadia Bahi-Buisson, and Jamel Chelly. Mosaic parental germline mutations causing recurrent forms of malformations of cortical development. *Eur J Hum Genet*, 2016.

29. Brett K. Kaiser, Daesong Yim, I-Ting Chow, Segundo Gonzalez, Zhenpeng Dai, Henning H. Mann, Roland K. Strong, Veronika Groh, and Thomas Spies. Disulphide-isomerase-enabled shedding of tumour-associated NKG2D ligands. *Nature*, 2007.

30. Benjamin G. Hoffstrom, Anna Kaplan, Reka Letso, Ralf S. Schmid, Gregory J. Turmel, Donald C. Lo, and Brent R. Stockwell. Inhibitors of protein disulfide isomerase suppress apoptosis induced by misfolded proteins. *Nat Chem Biol*, 2010.

31. Julie D. Atkin, Manal A. Farg, Bradley J. Turner, Doris Tomas, Judith A. Lysaght, Janelle Nunan, Alan Rembach, Phillip Nagley, Philip M. Beart, Surindar S. Cheema, and Malcolm K. Horne. Induction of the unfolded protein response in familial amyotrophic lateral sclerosis and association of protein-disulfide isomerase with superoxide dismutase 1. *J Biol Chem*, 2006.

32. Adam M. Benham. The Protein Disulfide Isomerase Family: Key Players in Health and Disease. *Antioxid Redox Signal*, 2012.

33. https://www.mousephenotype.org/data/genes/MGI:1914208, month accessed: June 2019.

34. Koichi Kawada, Takaaki Iekumo, Masayuki Kaneko, Yasuyuki Nomura, and Yasunobu Okuma. ER stress-induced aberrant neuronal maturation and neurodevelopmental disorders. *Yakugaku Zasshi*, 2016.

35. Martin W. Breuss, An Nguyen, Qiong Song, Thai Nguyen, Valentina Stanley, Kiely N. James, Damir Musaev, Guoliang Chai, Sara A. Wirth, Paula Anzenberg, Renee D. George, Anide Johansen, Shaila Ali, Muhammad Zia-Ur-Rehman, Tipu Sultan, Maha S. Zaki, and Joseph G. Gleeson. Mutations in LNPK, Encoding the Endoplasmic Reticulum Junction Stabilizer Lunapark, Cause a Recessive Neurodevelopmental Syndrome. *Am J Hum Genet*, 2018.

CHAPTER 4

A Relatively Common Homozygous *TRAPPC4* Splicing Variant is Associated with an Early-Infantile Neurodegenerative Syndrome

4.1 Abstract

Trafficking protein particle (TRAPP) complexes, which include the TRAPPC4 protein, regulate membrane trafficking between lipid organelles in a process termed vesicular tethering. *TRAPPC4* was recently implicated in a recessive neurodevelopmental condition in four unrelated families due to a shared c.454+3A>G splice variant. Here we report 23 patients from 17 independent families with an early infantile-onset neurodegenerative presentation, where we also identified the homozygous variant hg38:11:119020256 A>G (NM_016146.5:c.454+3A>G) in *TRAPPC4* through exome or genome sequencing. No other clinically relevant *TRAPPC4* variants were identified among any of over 10,000 patients with neurodevelopmental conditions. We found the carrier frequency of *TRAPPC4* c.454+3A>G was 2.4-5.4 per 10,000 healthy individuals. Affected individuals with the homozygous *TRAPPC4* c.454+3A>G variant showed profound psychomotor delay, developmental regression, early-onset epilepsy, microcephaly and progressive spastic tetraplegia. Based upon RNA sequencing, the variant resulted in partial exon 3 skipping and generation of an aberrant transcript due to use of a downstream cryptic splice donor site, predicting a premature stop codon and nonsense mediated decay. These data confirm the pathogenicity of the *TRAPPC4* c.454+3A>G variant, and refine the clinical presentation of *TRAPPC4*-related encephalopathy.

4.2 Introduction

Trafficking protein particle (TRAPP) complexes play a vital role in regulating membrane trafficking from the ER to the Golgi and plasma membrane. There are two TRAPP complexes identified in mammalian cells, TRAPPII and TRAPPIII, sharing core subunits TRAPPC1, TRAPPC2, TRAPPC3, TRAPPC4, TRAPPC5, TRAPPC6, and TRAPPC2L, with additional unique subunits specific to each of the complexes.[1] Biallelic variants in genes encoding human TRAPP subunits cause a range of disorders known as TRAPPopathies, including encephalopathy (*TRAPPC2L*, MIM:610970; *TRAPPC6B*, MIM:610397; *TRAPPC12*, MIM:614139), neurodevelopmental disorders (*TRAPP6A*, MIM:610396; *TRAPPC9*, MIM:611966), encephalomyopathy (*TRAPPC11*, MIM: 614138) and spondyloepiphyseal dysplasia tarda (*TRAPPC2*, MIM:300202).[2-6] However, the phenotypic heterogeneity remains poorly understood.

Very recently, a homozygous, variant in a well-conserved splice donor site (hg38:11:119020256 A>G (NM_016146.5:c.454+3A>G) within *TRAPPC4* (trafficking protein particle complex, subunit C4, (MIM:610971)) was identified in 8 affected individuals from four independent families presenting a neurodevelopmental disorder characterized by severe-to-profound developmental delay/intellectual disability, microcephaly, early-onset seizures, and spastic quadriparesis.[7,8] The TRAPPC4-related phenotype acronym is known as NEDESBA (MIM:618741), which stands for NEurodevelopmental Disorder with Epilepsy, Spasticity, and Brain Atrophy. Here we report 23 additional cases from 17 independent families with an early infantile-onset neurodegeneration harboring the same homozygous *TRAPPC4* splice site variant, recurring in all the families from different ethnicities, and resulting in decreased *TRAPPC4* expression due to aberrant splicing.

4.3 Materials and Methods

4.3.1 Participant Recruitment and Sequencing

The institutional review boards of the University of California, San Diego, as well as the respective host institution approved this study. All study participants and/or their parents or guardians signed informed consent forms allowing for participation and to publishing of photographs and other identifying information. Genomic DNA was extracted from peripheral blood leukocytes using a QIAamp DNA blood Midi kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. For families 1-6, 16 and 17, exome sequencing and data analysis were performed as previously described.[9] For families 7-10, exome sequencing and analysis were conducted at CENTOGENE AG and for families 12-15, exome sequencing and analysis were conducted at GeneDx.[10,11] For family 15, genome sequencing was performed through the Undiagnosed Diseases Network. In families 11 and 12, trio genome sequencing and proband-only RNA sequencing were performed. Sequencing for the remaining families was performed at the Broad Institute (Cambridge, Massachusetts, United States of America). The candidate variant was confirmed, and segregation analysis was performed by Sanger sequencing. Further details are available in the Supplementary Material.

4.3.2 RNA Preparation and Short Read RNA Sequencing

RNA was isolated from cultured fibroblasts obtained from a skin biopsy from patient 15, in addition to 24 unrelated controls. RNA sequencing was performed at the Broad Institute (Cambridge, Massachusetts, United States of America). Libraries were prepared using a TruSeq Stranded mRNA Library Prep Kit (Cat. No. RS-122-2101, RS-122-2102, and RS-122-2103). Paired-end sequencing was performed using a read length of up to 100bp on an Illumina HiSeq4000 instrument to achieve a minimum sequencing depth of 50 million paired-end reads.

RNA sequencing reads were processed using a Bpipe (Version 0.9.9.6, release 21/07/2018) pipeline for quality control checks, trimming, and alignment.[12] FastQC and Trimmomatic were used for sequencing quality checks and trimming of poor-quality reads.[13] Alignment was performed using STAR (version 2.7.3a, release 08/10/2019), in two-pass mode for read alignment to the Human Reference Genome Build 38 (excluding "ALT" contigs).[14] Duplicate reads were marked with Picard MarkDuplicates and quantification was performed using FeatureCounts from the R Subread package Version 1.34.7, release 03/01/2019).[15] Differential expression analysis was performed using the DESeq2 package (version 1.25.9, release 31/07/2019) comparing expression of the affected individual to 24 unrelated control samples.[16] Sashimi plots were prepared using ggsashimi.[17]

4.4 Results

We report a cohort of 23 patients from 17 families with a *TRAPPC4*-related neurodevelopmental disorder (Fig.1A). The families are of different ancestries (Iranian, Egyptian, Portuguese, English, mixed European-American) and were identified through collaboration among the involved study centers.

4.4.1 Identification of a Homozygous TRAPPC4 c.454+3A>G Variant

Among a collective cohort of over 10,000 patients with mostly recessive neurodevelopmental disease from the University of California, San Diego and University College London cohorts, approximately 50% remain unsolved after whole exome sequencing. Among these, there were 23 patients from 17 families where both university groups independently identified the homozygous c.454+3A>G variant (hg38:11:119020256 A>G (NM_016146.5, rs375776811) in intron 3 of *TRAPPC4*, among a search for potential variants affecting splicing

(Figure 1B-D). *TRAPPC4* encodes for Trafficking Protein Particle Complex Subunit 4 (TRAPPC4). No other clinically relevant variants in *TRAPPC4* were identified in any of the remaining unsolved families. This variant had been initially filtered out from exome analysis because the allele frequency (AF) rose above our 0.02% AF-cutoff. It was encountered as homozygous only in affected members, and not in any unaffected members, and segregated with the phenotype consistent with a recessive mode of inheritance with full penetrance in each of the 17 families. Comparison of the exome calls among the probands of these families did not identify any shared haplotype, suggesting that either it is recurrently mutated or is an ancient genetic variant.[7]

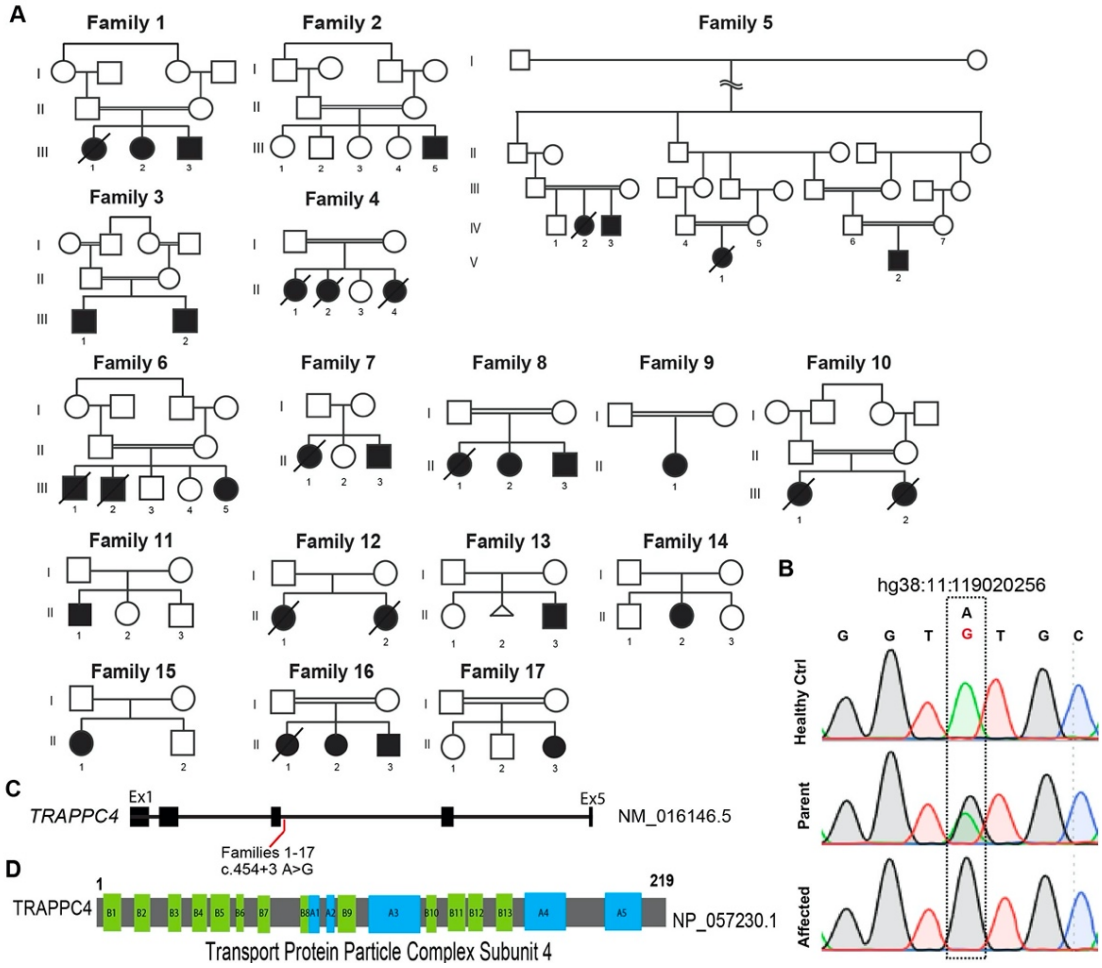

Figure 4.1: Pedigrees, chromatogram, and TRAPPC4 gene and protein.
(a) Pedigrees of the 17 families, including an extended Iranian family composed of three smaller distantly related subfamilies (family 5). Double bar: consanguinity. Slash: deceased. (**b**) Sanger sequence chromatograms with the c.454+3A>G variant (NM_016146.5) in a healthy control (Healthy Ctrl), heterozygous carrier (Father) and patient from family 5 (Affected). (**c**) Schematic of NM_016146.5 *TRAPPC4* transcript with 5 predicted exons and the c.454+3A>G variant (red) identified in the 17 families. Exons were numbered using the UC Santa Cruz (UCSC) Genome Browser as a reference. (**d**) TRAPPC4 219 amino acid length protein (NP_057230.1) with predicted beta strands (B1-B13) and alpha helices (A1-A5).

4.4.2 Search for the TRAPPC4 c.454+3A>G Variant in clinical exome and genome databases

The gnomAD (v2.1.1) reports the minor allele frequency (MAF) of the *TRAPPC4* c.454+3A>G variant to be 0.024% with 68 heterozygotes reported with representation in the South Asian, European, and African populations. In the GeneDx® database of patients referred for

molecular diagnostics, there were four individuals that were homozygous for the *TRAPPC4* c.454+3A>G variant (Supplemental File 3, families 12, 13, 14, and 15), all of whom consented to publication. In the Centogene® database, comprising clinically ascertained exomes from 47,811 individuals, there were eight individuals homozygous for the *TRAPPC4* c.454+3A>G variant, and amongst these, consent to publish was obtained for four (Supplemental File 3, families 7, 8, 9, 10). All known individuals homozygous for the variant were found to be clinically affected, suggesting full penetrance, and bringing the total number of patients reported here to 27.

In addition to the patients carrying a homozygous *TRAPPC4* c.454+3A>G variant, the same variant was found heterozygous in two individuals in our Middle Eastern cohort of over 5000 individuals (MAF = 0.040%) who underwent whole exome sequencing, and in whom a causative gene other than *TRAPPC4* was identified or families remain unsolved.[18,19] In the GeneDx® unaffected adult cohort, *TRAPPC4* c.454+3A>G was heterozygous in 64 out of 189,846 alleles (MAF = 0.033%). In the 100K Genomes Project, the variant was seen in 64 of 59,464 participants (all heterozygous, MAF = 0.054%). In the NHLBI Exome Sequencing Project database the MAF = 0.046%, and in the UK10K database the MAF = 0.03%. The Centogene® database comprising clinically ascertained exomes of 47,811 individuals, we identified 14 carriers from unrelated families, with MAF = 0.026%. In the Baylor Genetics Laboratories database, of 14,775 sequenced individuals, 13 heterozygous individuals were identified (MAF = 0.088%). Six further heterozygous individuals were found in the dataset of ~15,500 exomes of the Queen Square Genomic Center database (MAF = 0.039%). Accordingly, the carrier frequency of the *TRAPPC4* c.454+3A>G variant ranges from 2.4-5.4 per 10,000 individuals in our collective cohort consisting mostly of individuals with neurological disease. There was no reported health consequence of carrying the heterozygous variant.

4.4.3 TRAPPC4 mRNA Analysis Reveals Defective Splicing

The c.454+3A>G variant occurred immediately following exon 3, suggesting defective splicing may result. RNA sequencing of dermal fibroblasts from patient 15 revealed *TRAPPC4* as the only genome-wide significant differentially expressed gene, compared to 24 similarly sequenced controls (Adjusted p-value = 0.016) (Fig. 4.2A), a finding that helped identify the *TRAPPC4* c.454+3A>G variant as the cause in these families. Inspection of reads mapping to *TRAPPC4* revealed exon 3 skipping in approximately 46% (7 of 15) of transcripts utilizing the acceptor splice site of exon 4 (Fig. 4.2B). This splice junction was also observed at low levels in the sequenced control samples where it was utilized by an average 3% of transcripts. A further small percentage (13%) of reads in the patient contain the NM_016146.5; c.454+3A>G variant and use a novel donor splice site at position c.454+41, resulting in a 40-nucleotide extension of exon 3. In summary, more than half (59%) of expressed *TRAPPC4* transcripts from patient 15 fibroblasts display aberrant splicing while wild-type splicing of exon 3 was supported by 41% of reads. Both of the aberrant transcripts detected predict a frameshift and nonsense mediated decay, consistent with the observed decrease in *TRAPPC4* expression. RT-PCR and subsequent Sanger sequencing analysis of the RT-PCR fragments have been described by Van Bergen et al.[7]

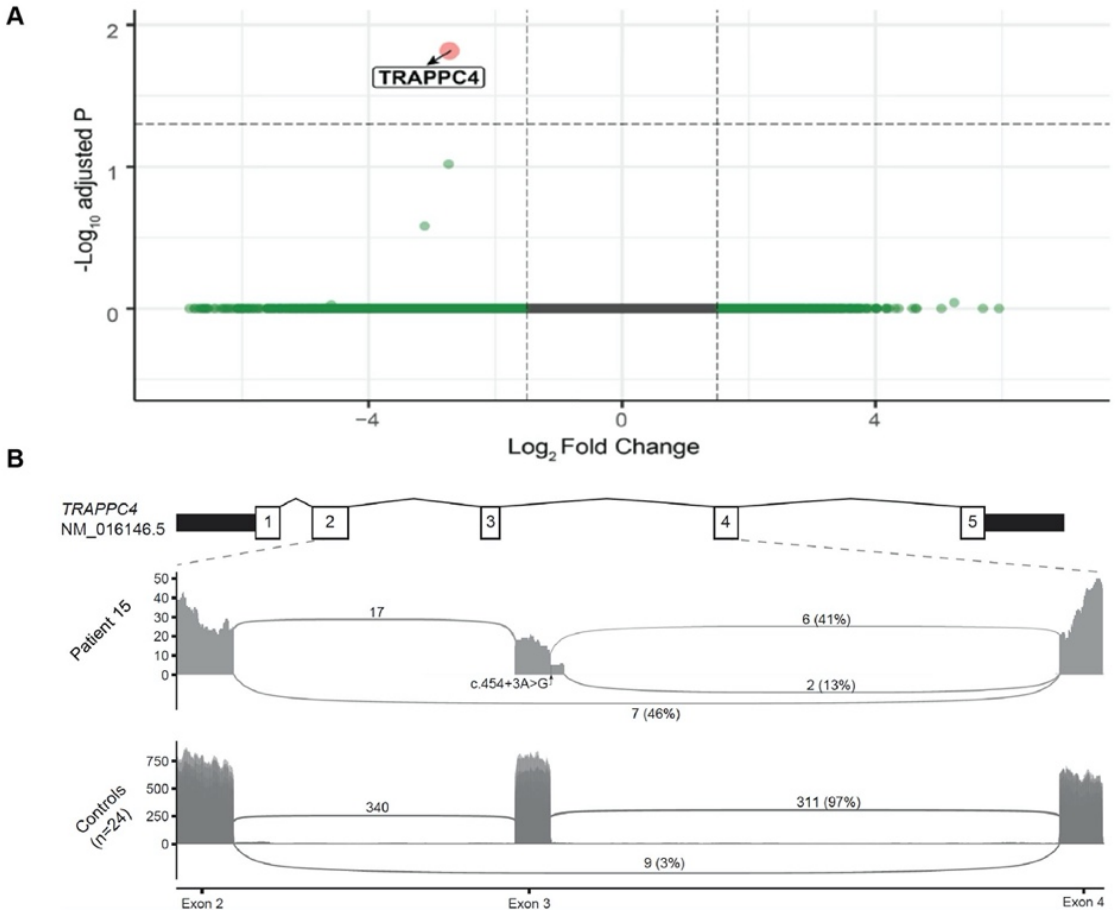

Figure 4.2: *TRAPPC4* mRNA analysis.
(A) Volcano plot illustrating total RNA transcript level of all transcripts mapped to the reference genome for fold change vs. adjusted p-value of RNA sequencing from fibroblasts derived from Patient 15 compared to 24 control fibroblasts lines. *TRAPPC4* (red dot) was the only gene significantly differentially expressed with a $\log_2$-fold change between 2-4 and a minus-$\log_{10}$ adjusted P value above genome wide significance (gray dashed line) between 1-2. (B) Sashimi plot depicting individual reads from patient 15 (top) compared with sum of 24 controls (bottom) mapped to the reference genome from exon 2 to exon 4 of *TRAPPC4*. Site of location is shown with arrow. In control, splicing from exon 2 to 4 occurs in only 3% of reads in control but 46% of reads in patient 15. A novel splice donor site 40 bp downstream of the canonical site in exon 3 was seen in 13% of reads in patient 15, and not seen in control. Splicing from the canonical exon 3 donor site was seen in 97% of transcripts in controls but 41% in patient 15.

4.4.4 Clinical Evaluation of the Affected Individuals

Following a generally uncomplicated pregnancy and neonatal course, in the first few months of life all affected individuals developed a severe, progressive encephalopathy characterized by motor, language, and social developmental stagnation. Psychomotor regression

was further observed in some cases (Table 4.1) With disease progression, patients experienced a critical cognitive decline and demonstrated acquired microcephaly. The degree of microcephaly was generally severe, with a mean occipito-frontal circumference of -5.77 standard deviations (SDs) at around 5 years of age as a mean. Subtle and non-specific facial dysmorphic features, including bitemporal narrowing, thick eyebrows, full cheeks, long philtrum, wide mouth, thin and tented upper lip, and pointed chin (Fig. 4.3A) were observed in most cases, in addition to all previously reported cases over two years of age.[7,8] In all patients, seizure onset was in the first 6 months of life with highly variable seizure types (e.g., infantile spasms and focal, tonic-clonic, atonic, and tonic seizures) and frequencies. In most cases, there was a partial response to antiepileptic drugs (AEDs), including levetiracetam and clobazam, but seizures were refractory in patients 9 and 19. Electroencephalographic features were not specific, showing generalized disorganization and epileptiform discharges. Neurologic examination revealed spastic tetraplegia and hyperreflexia in all subjects, similar to previously reported subjects where data was available.[7,8] Abnormal movements mainly consisting of ataxia and dystonia were observed in patients 5, 6, 16, and 18. Patients 14, 15, and 18 showed bilateral cataracts. Despite the presence of visual fixation in most individuals, visual pursuits were almost invariably absent, suggesting impaired visual function probably independent of cataracts. Hearing loss was observed in patients 11 and 12. Early death due to infections occurred in five out of the 23 reported individuals at a mean of 8.8 years.

Brain MRIs showed a variable degree of cerebral atrophy in all cases (Fig. 4.3B). This was also evident in five out of the eight previously reported cases where MRIs were available.[7,8] In many subjects, loss of white matter, ventriculomegaly, and cerebellar atrophy were also dramatic, and appeared more severe in older children, suggestive of progressive atrophy. The degree of

ventriculomegaly varied considerably, but all subjects showed enlarged subarachnoid spaces and widely spaced cortical gyri. Hypoplasia of the corpus callosum (CCH) and brainstem were also observed (Supplemental File 3).

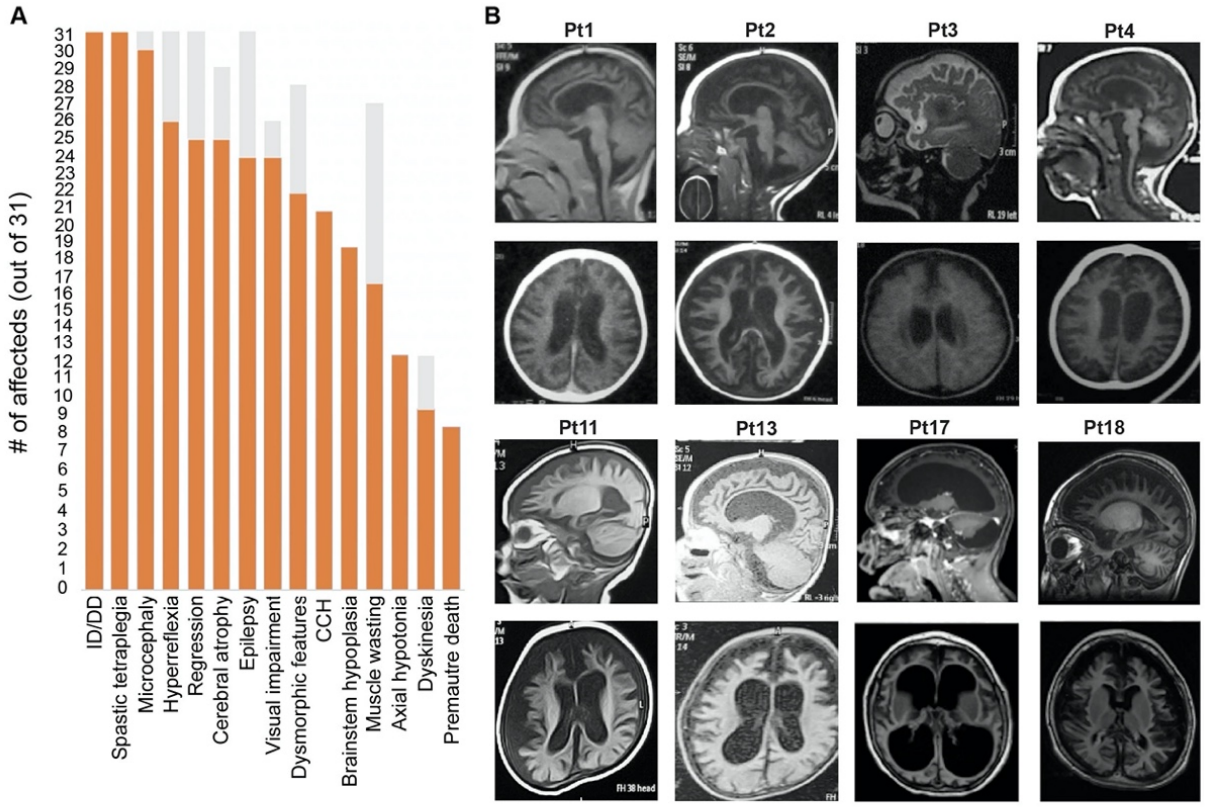

Figure 4.3: Clinical and radiological characterization of *TRAPPC4* patients.
(**A**) Bar graph showing the distribution of the most relevant clinical features from the 17 ascertained families and the eight patients identified from clinical testing facilities. Orange: number of patients out of 27 showing each feature. Gray: data not available. (**B**) Brain MRI scans showing global cerebral atrophy in all patients, including cortex, corpus callosum (arrow in Patient 1), cerebellum (double arrows in patient 4) and ventricular enlargement (asterisk in patient 17).

4.5 Discussion

This study supports the pathogenicity of the recurrent c.454+3A>G variant in *TRAPPC4* as the cause of a severe developmental encephalopathy characterized by a profound global developmental delay, and frequent psychomotor regression and lethality, and delineates the carrier frequencies throughout sampled populations. Affected individuals developed debilitating motor

impairments, with spastic tetraplegia, cognitive decline, and movement disorders. Common facial dysmorphism includes bitemporal narrowing, thick eyebrows, full cheeks, long philtrum, wide mouth, thin and tented upper lip, and pointed chin. Our report of additional affected individuals helps clarify the expressivity in NEDESBA.

Early-onset epilepsy with variable seizure phenotypes and electroencephalographic findings is a feature of NEDESBA. In most subjects, there is a partial response to AEDs, however refractory seizures were observed in some patients. Interestingly, two affected siblings with gelastic seizures and episodes of temperature dysregulation have been previously reported, one of whom also had premature adrenarche and osteopenia.[7] These findings, suggestive of hypothalamic dysregulation, have not been observed in the other cases or in our cohort. Similarly, dystonia or ataxia were only observed in some affected individuals. Accordingly, NEDESBA might have a wide phenotypic spectrum, and variable expressivity. The report of our further cases helps to clarify the expressivity of these rare features.

An additional critical aspect of NEDESBA involves psychomotor regression. In previous reports, this condition was mainly described as a progressive encephalopathy. Most subjects from our cohort showed absent development of motor, verbal, and social skills. Although, some children attained developmental milestones in the first few months of life, these were invariably lost with disease progression. Psychomotor regression was observed in all phenotyped individuals, representing a cardinal feature of this disorder (Supplemental File 3). Together with the progressive course of the neuroimaging abnormalities, this observation suggests that *TRAPPC4* encephalopathy displays early onset neurodegenerative features.

An individual with *TRAPPC4* variant displaying severe muscle involvement associated with biochemical alterations suggestive of episodic rhabdomyolysis (increased plasma lactate,

creatine phosphokinase, and transaminases) has been very recently reported, resembling the progressive early-onset encephalopathy with episodic rhabdomyolysis (PEERB, MIM:618331) caused by recessive *TRAPPC2L* variants (MIM:618331).[8] Muscle wasting was observed in our cohort, most likely secondary to the profound limitation of voluntary movements. In one case (patient 19), blood tests revealed elevated lactate on one occasion (level 424 U/L, reference range 20-200 U/L), but this was not observed on repeat testing. Although primary muscle involvement is still possible, many *TRAPPC4* subjects also present clinical manifestations which might lead to a transient increase in serum muscle enzymes (e.g., seizures, spasticity, and dystonia). Accordingly, the assessment of biochemical alterations for creatine kinase and other muscle enzymes on a case-by-case basis is essential.

Although not pathognomonic, the neuroimaging features of *TRAPPC4* subjects demonstrate consistent cerebral atrophy with predominant involvement of subcortical white matter and ventricular enlargement, often in association with basal ganglia, cerebellar and brainstem atrophy. The remarkable white matter involvement is also supported by observation of thin corpus callosum evident in most patients. However, in contrast to cases from a previous report,[7] a significant sparing of the basal ganglia was not noted in our cohort. Accordingly, the exact progression of the neurodegenerative process and the temporal relationship between cerebral basal ganglia, brainstem and cerebellar atrophy remain to be elucidated.

The c.454+3A>G variant is relatively common across populations worldwide, as shown by the relatively high carrier frequency in public and private databases. This is also supported by the absence of consanguinity in some of our families (9, 11, 12, 13, and 15) and in one previously reported family.[19] This variant might have derived from a possible shared common ancestor, however, neither our study nor previous studies were able to detect a shared haplotype. We showed

that this variant leads to a failed splicing based on the presence of *TRAPPC4* aberrant transcripts. These findings support the pathogenicity of the *TRAPPC4* variant as the cause of the disorder in the studied subjects. Furthermore, the identification of the c.454+3A>G variant in RNAseq data in patient 15 stresses the power of RNAseq as an ancillary diagnostic tool in individuals with undiagnosed rare diseases, especially in cases where exome sequencing yields negative results.[20]

Table 4.1: TRAPP subunit subcomplex components and clinical syndrome so far associated with disease.

Most TRAPP subunits are part of both TRAPPII and III complexes except for C10, which is exclusively part of TRAPPCII, and TRAPPC11 ,-12, and -13, which are exclusively part of TRAPPIII. All conditions are autosomal recessive except the *TRAPPC2* gene which is X-linked.

Subunit	Complex	Clinical Condition	OMIM#
C1	II and III		
C2	II and III	Spondyloepiphyseal dysplasia tarda, X-linked	313400
C2L	II and III	Encephalopathy, progressive, early-onset, with episodic rhabdomyolysis	618331
C3	II and III		
C4	II and III	Neurodevelopmental disorder with epilepsy, spasticity, and brain atrophy	618741
C5	II and III		
C6A	II and III		
C6B	II and III	Neurodevelopmental disorder with microcephaly, epilepsy, and brain atrophy	617862
C9	II and III	Mental retardation, autosomal recessive 13	613192
C10	II		
C11	III	Muscular dystrophy, limb-girdle, autosomal recessive 18	615356
C12	III	Encephalopathy, progressive, early-onset, with brain atrophy and spasticity	617669
C13	III		

A range of conditions are now described in the TRAPPopathies primarily involving the developing brain and muscle (Table 4.1). Our study highlights the involvement of this relatively common single allele as the cause for a severe recessive neurological disease characterized by progressive microcephaly, profound cognitive decline, infantile-onset epilepsy, and progressive spastic tetraplegia. Our observations expand the spectrum of TRAPPopathies, supporting the existence of phenotypic overlap in this complex group of disorders. To date, around 23 families of European, Mediterranean, Middle-Eastern, and Indian ancestries are known to be homozygous for the c.454+3A>G variant in *TRAPPC4*. The highest frequency of this variant is found in individuals with European and Mediterranean ancestries, whereas it appears to be rare in African and East Asian populations. The commonality of the c.454+3A>G variant in the general human population warrants the search for *TRAPPC4* variants in children with microcephaly, especially when associated with neurodegenerative features. The frequency of this variant in European and Mediterranean basin populations suggests that a genetic screening-based approach might prove useful in early carrier detection.

Chapter 4, in full, is a reprint of the material as it appears in European Journal of Human Genetics 2020. Ghosh, Shereen G.; Scala, Marcello; Beetz, Christian; Helman, Guy; Stanley, Valentina; Yang, Xiaoxu; Breuss, Martin W.; Mazaheri, Neda; Selim, Laila; Hadipour, Fatemeh; Pais, Lynn; Stutterd, Chloe; Karageorgou, Vasiliki; Begtrup, Amber; Crunk, Amy; Juusola, Jane; Willaert, Rebecca; Flore, Leigh A.; Kennelly, Kelly; Spencer, Christopher; Brown, Martha; Trapane, Pamela; Hurst, Anna C.E.; Rutledge, S. Lane; Goodloe, Dana H.; McDonald, Marie T.; Shashi, Vandana; Schoch, Kelly; Undiagnosed Diseases Network; Tomoum, Hoda; Zeitoun, Raghda; Hadipour, Zahra; Galehdari, Hamid; Pagnamenta, Alistair T.; Mojarrad, Majid; Sedaghat, Alireza; Dias, Patricia; Quintas, Sofia; Eslahi, Atiyeh; Shariati, Gholamreza; Bauer,

Peter; Simons, Cas; Houlden, Henry; Issa, Mahmoud I.; Zaki, Maha S.; Maroofian, Reza;

Gleeson, Joseph G. The dissertation author was the primary investigator and author of this paper.

4.6 References

1. Stephanie Brunet and Michael Sacher. In Sickness and in Health: The Role of TRAPP and Associated Proteins in Disease. *Traffic*, 2014.

2. Miroslav P. Milev, Claudio Graziano, Daniela Karall, Willemijn F.E. Kuper, Noraldin Al-Deri, Duccio Maria Cordelli, Tobias B. Haack, Katharina Danhauser, Arcangela Iuso, Flavia Palombo, Tommaso Pippucci, Holger Prokisch, Djenann Saint-Dic, Marco Seri, Daniela Stanga, Giovanna Cenacchi, Koen L.I. van Gassen, Johannes Zschocke, Christine Fauth, Johannes A. Mayr, Michael Sacher, and Peter M. van Hasselt. Bi-allelic mutations in TRAPPC2L result in a neurodevelopmental disorder and have an impact on RAB11 in fibroblasts. *J Med Genet,* 2018.

3. Agi K. Gedeon, Alison Colley, Robyn Jamieson, Elizabeth M. Thompson, John Rogers, David Sillence, George E. Tiller, John C. Mulley and Jozef Gecz. Identification of the gene (SEDL) causing X-linked spondyloepiphyseal dysplasia tarda. *Nature Genet,* 1999.

4. Nina Bögershausen, Nassim Shahrzad, Jessica X. Chong, Jürgen-Christoph von Kleist-Retzow, Daniela Stanga, Yun Li, Francois P. Bernier, Catrina M. Loucks, Radu Wirth, Eric G. Puffenberger, Robert A. Hegele, Julia Schreml, Gabriel Lapointe, Katharina Keupp, Christopher L. Brett, Rebecca Anderson, Andreas Hahn, A. Micheil Innes, Oksana Suchowersky, Marilyn B. Mets, Gudrun Nürnberg, D. Ross McLeod, Holger Thiele, Darrel Waggoner, Janine Altmüller, Kym M. Boycott, Benedikt Schoser, Peter Nürnberg, Carole Ober, Raoul Heller, Jillian S. Parboosingh, Bernd Wollnik, Michael Sacher, and Ryan E. Lamont. Recessive TRAPPC11 mutations cause a disease spectrum of limb girdle muscular dystrophy and myopathy with movement disorder and intellectual disability. *Am J Hum Genet,* 2013.

5. Isaac Marin-Valencia, Gaia Novarino, Anide Johansen, Basak Rosti, Mahmoud Y. Issa, Damir Musaev, Gifty Bhat, Eric Scott, Jennifer L. Silhavy, Valentina Stanley, Rasim O. Rosti, Jeremy W. Gleeson, Farhad B. Imam, Maha S. Zaki, and Joseph G. Gleeson. A homozygous founder mutation in TRAPPC6B associates with a neurodevelopmental disorder characterised by microcephaly, epilepsy and autistic features. *J Med Genet,* 2018.

6. Michael Sacher, Nassim Shahrzad, Hiba Kamel, and Miroslav P. Milev. TRAPPopathies: An emerging set of disorders linked to variations in the genes encoding transport protein particle (TRAPP)-associated proteins. *Traffic*, 2019.

7. Nicole J. Van Bergen, Yiran Guo, Noraldin Al-Deri, Zhanna Lipatova, Daniela Stanga, Sarah Zhao, Rakhilya Murtazina, Valeriya Gyurkovska, Davut Pehlivan, Tadahiro Mitani, Alper Gezdirici, Jayne Antony, Felicity Collins, Mary J.H. Willis, Zeynep H. Coban Akdemir, Pengfei Liu, Jaya Punetha, Jill V. Hunter, Shalini N. Jhangiani, Jawid M. Fatih, Jill A. Rosenfeld, Jennifer E. Posey, Richard A. Gibbs, Ender Karaca, Sean Massey, Thisara G. Ranasinghe, Patrick Sleiman, Chris Troedson, James R. Lupski, Michael Sacher, Nava Segev, Hakon Hakonarson, and John Christodoulou. Deficiencies in vesicular transport mediated by *TRAPPC4* are associated with severe syndromic intellectual disability. *Brain,* 2019.

8. Parneet Kaur, Rajagopal Kadavigere, Katta Mohan Girisha, and Anju Shukla. Recurrent bi-allelic splicing variant c.454+3A>G in TRAPPC4 is associated with progressive encephalopathy and muscle involvement. *Brain,* 2020.

9. Yaping Yang, Donna M. Muzny, Jeffrey G. Reid, Matthew N Bainbridge, Alecia Willis, Patricia A. Ward, Alicia Braxton, Joke Beuten, Fan Xia, Zhiyv Niu, Matthew Hardison, Richard Person, Mir Reza Bekheirnia, Magalie S. Leduc, Amelia Kirby, Peter Pham, Jennifer Scull, Min Wang, Yan Ding, Sharon E. Plon, James R. Lupski, Arthur L. Beaudet, Richard A. Gibbs, and Christine M. Eng. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *N Engl J Med,* 2013.

10. Peter Bauer, Krishna Kumar Kandaswamy, Maximilian E.R. Weiss, Omid Paknia, Martin Werber, Aida M. Bertoli-Avella, Zafer Yüksel, Malgorzata Bochinska, Gabriela E. Oprea, Shivendra Kishore, Volkmar Weckesser, Ellen Karges, and Arndt Rolfs. Development of an evidence-based algorithm that optimizes sensitivity and specificity in ES-based diagnostics of a clinically heterogeneous patient population. *Genet Med,* 2019.

11. Kyle Retterer, Jane Juusola, Megan T. Cho, Patrik Vitazka, Francisca Millan, Federica Gibellini, Annette Vertino-Bell, Nizar Smaoui, Julie Neidich, Kristin G Monaghan, Dianalee McKnight, Renkui Bai, Sharon Suchy, Bethany Friedman, Jackie Tahiliani, Daniel Pineda-Alvarez, Gabriele Richard, Tracy Brandt, Eden Haverfield, Wendy K. Chung, and Sherri Bale. Clinical application of whole-exome sequencing across clinical indications. *Genet Med,* 2016.

12. Simon P. Sadedin, Bernard Pope, and Alicia Oshlack. Bpipe: a tool for running and managing bioinformatics pipelines. *Bioinformatics,* 2012.

13. Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 2014.

14. Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 2013.

15. Yang Liao, Gordon K. Smyth, and Wei Shi. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res,* 2019.

16. Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 2014.

17. Diego Garrido-Martin, Emilio Palumbo, Roderic Guigo, and Alessandra Breschi. ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput Biol,* 2018.

18. Tracy J. Dixon-Salazar, Jennifer L. Silhavy, Nitin Udpa, Jana Schroth, Stephanie Bielas, Ashleigh E. Schaffer, Jesus Olvera, Vineet Bafna, Maha S. Zaki, Ghada H. Abdel-Salam, Lobna A. Mansour, Laila Selim, Sawsan Abdel-Hadi, Naima Marzouki, Tawfeg Ben-Omran, Nouriya A. Al-Saana, F. Müjgan Sonmez, Figen Celep, Matloob Azam, Kiley J. Hill, Adrienne Collazo, Ali G. Fenstermaker, Gaia Novarino, Naiara Akizu, Kiran V. Garimella, Carrie Sougnez, Carsten Russ, Stacey B. Gabriel, and Joseph G. Gleeson. Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med,* 2012.

19. Eric M. Scott, Anason Halees, Yuval Itan, Emily G. Spencer, Yupeng He, Mostafa Abdellateef Azab, Stacey B. Gabriel, Aziz Belkadi, Bertrand Boisson, Laurent Abel, Andrew G. Clark, Greater Middle East Variome Consortium, Fowzan S. Alkuraya, Jean-Laurent Casanova, and Joseph G. Gleeson. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet,* 2016.

20. Hernan D. Gonorazky, Sergey Naumenko, Arun K. Ramani, Viswateja Nelakuditi, Pouria Mashouri, Peiqui Wang, Dennis Kao, Krish Ohri, Senthuri Viththiyapaskaran, Mark A. Tarnopolsky, Katherine D. Mathews, Steven A. Moore, Andres N. Osorio, David Villanova, Dwi U Kemaladewi, Ronald D. Cohn, Michael Brudno, and James J. Dowling. Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *Am J Hum Genet,* 2019.

CHAPTER 5

Biallelic Hypomorphic Mutations in *HEATR5B*, Encoding HEAT repeat-containing protein 5B, Lead to Pontocerebellar Hypoplasia

5.1 Abstract

HEAT repeats are 37-47 amino acid flexible tandem repeat structural motifs occurring in a wide variety of eukaryotic proteins with diverse functions. Due to their ability to undergo elastic conformational changes, they often serve as scaffolds at sites of protein interactions. Here, we describe four affected children from two families presenting with pontocerebellar hypoplasia manifest clinically with neonatal seizures, severe intellectual disability and motor delay. Whole exome sequencing identified biallelic variants at predicted splice sites in intron 31 of *HEATR5B*, encoding the HEAT repeat-containing protein 5B segregating in a recessive fashion. Aberrant splicing was found in patient fibroblasts, which correlated with reduced levels of HEATR5B protein. *HEATR5B* is expressed during brain development in human, and we failed to recover live-born homozygous *Heatr5b* knockout mice. Taken together, our results implicate loss of *HEATR5B* in pontocerebellar hypoplasia.

5.2 Introduction

Pontocerebellar hypoplasia (PCH) refers to a group of recessive neurodevelopmental disorders characterized by loss of parenchymal volume of the pons and cerebellum. PCH frequently follows a degenerative course, with clinical hallmarks including defects in brainstem and cerebellar function, evident by difficulty with swallowing, oculomotor defects, spastic

quadriparesis, epilepsy, frequent pulmonary infections, and early death.[1] Currently, there are eleven overlapping clinical subtypes and seventeen PCH-related genes that have been described, implicating genes involved in RNA processing or protein translation.[2]

HEAT (Huntingtin, Elongation factor 3, the PR65/A subunit of protein phosphatase 2A, and TOR) repeats (HEATRs) are 37-47 amino-acid motifs consisting of duplicated anti-parallel α-helices linked by flexible inter-unit loops.[3] Structurally related to armadillo repeats, HEATRs occur in a variety of eukaryotic proteins mediating protein interactions involved in processes such as cytoplasmic and nuclear transport, microtubule dynamics, or chromosome segregation.[4] Upon interaction with binding partners HEATRs are capable of undergoing flexible and elastic changes.[5] Although HEAT repeats occur in dozens of proteins, a total of ten different genes are designated with the name HEATR in the human genome (HEATR 1, 2, 3, 4, 5A, 5B, 6, 7A, 7B1 and 8), and none yet have been linked to human disease, except in a single family with primary ciliary dyskinesia with a homozygous *HEATR2* missense allele (CILD18; MIM: 614874).[6]

5.3 Materials and Methods

5.3.1 Patient Recruitment

The Institutional Review Board at the University of California, San Diego, approved this study. All study participants signed informed consent documents, and the study was performed in accordance with Health Insurance Portability and Accountability Act (HIPAA) Privacy Rules. The procedures followed for recruitment and data collection were in accordance with the ethical standards of the responsible committee on human experimentation at the respective, participating institute and proper informed consent was obtained.

5.3.2 DNA Extraction and Whole Exome Sequencing

DNA was extracted on an Autopure LS instrument (Qiagen, Valencia, CA) according to the manufacturer's instructions. Samples of the three affected individuals and their unaffected parents were subjected to Agilent Sure-Select Human All Exon v2.0 (44Mb baited target) library preparation sequencing on an Illumina HiSeq 2000 with v2 chemistry (Read Length: 151). The accession number for these data is dbGAP: phs000288.v1.p1. The Genome Analysis Toolkit (GATK) workflow was used to identify variants that were homozygous in the children.[7]

5.3.3 Computational Analysis

Variant calling and filtering were performed according to an established whole exome sequencing pipeline.[7] Runs of homozygosity were assessed using HomozygosityMapper.[8-10] Variants were filtered if not consistent with recessive monogenic inheritance and if the minor allele frequency (MAF) was >1:10,000 in gnomAD or >1:1,000 in our in-house ethnically matched exome database consisting of 7,000 sequenced individuals. Missense variants were prioritized by their conservation scores in GERP++ and PhyloP and deleterious or pathogenic scores in at least two of the following *in silico* pathogenicity prediction tools: Combined Annotation Dependent Depletion (CADD), MutationTaster, Polyphen-2, and SIFT.

5.3.4 Reverse Transcription (RT) PCR

Patient fibroblasts for Family 2610 were harvested from a skin dermal punch biopsy as described.[11] Total RNA from quantified patient and control fibroblasts was reverse-transcribed using the Superscript III First-Strand cDNA Kit (Invitrogen, Cat: 18080051). PCR analysis of cDNA was designed against exons 22-23, 30-31, and 33-34 of *HEATR5B,* excluding introns using GoTaq Master Mix (Promega, Cat: M7833). Quantification of band intensity was performed using Image J software and normalized to loading control.

5.3.5 Gel Extraction

DNA was purified from agarose gels following RT-PCR and recovered using the Zymoclean Gel DNA Recovery Kit as per manufacturer's instructions (Zymo Research, Cat: D4001).

5.3.6 Western Blot

Cells were lysed with ice-cold RIPA buffer (150mM sodium chloride, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 50mM Tris pH 8.0) and assessed by Western blot using standard protocols. Primary antibodies used were anti-beta-Actin (Santa Cruz, sc-47778, 1:2000), and anti-HEATR5B (Sigma-Aldrich, HPA055639, 1:200). Western blot signal intensity (area of peaks) was analyzed using ImageJ and normalized to the loading control for quantification. N=3 independent extractions used for quantification.

5.3.7 Mouse

Animal use followed NIH guidelines and was approved by IACUC at the University of California, San Diego. The guide 5'- GGATTGTTCATTGCTCTCTT-3' was co-injected with recombinant CRISPR/Cas9 into C57Bl/6 zygotes according to standard protocols at the UCSD Transgenic Mouse Core. We recovered several damaging alleles and focused on a *Heatr5b* frameshift allele in constitutively spliced exon 28 of 36 (Refseq NM_001081179.1; c.4384_4385insA; p.Asp1462GlufsTer15). Mice were genotyped by PCR followed by Sanger sequencing (mHeatr5b_F/R; Table S1). Heterozygous mice were backcrossed to wildtype C57Bl/6 for at least five generations prior to intercross of heterozygotes to generate homozygous animal. Timed pregnant animals were obtained by plug checks, where the day of the observed vaginal plug was determined to be E0.5.

5.3.8 Statistical Analysis

Statistical analysis was done as indicated in figure legends using GraphPad Prism 7. Visualization of data was performed using GraphPad Prism 7.

5.4 Results

5.4.1 Clinical Evaluation of Four Affected Individuals from Two Families with Similar Neurological Phenotypes.

We identified four affected members from two unrelated families with similar neurological phenotypes. Family 2610 had documented parental consanguinity and presented with two affected children (II-1 and II-2) with severe global motor and cognitive delays and frequent neonatal-onset seizures (Figure 1, Table 5.1). The older sibling was born at 36 weeks of gestation and showed intractable convulsions and hypocalcemia in the neonatal period. Absence of meconium led to a diagnosis of aganglionic megacolon that required surgery at 11 days of age. Physical examination showed growth retardation, severe intellectual disability, and gross motor delay with absent head control. There was also microphthalmia, low set ears, talipes equinovarus, increased deep tendon reflexes, and limb spasticity. The younger sibling showed similar features, without megacolon, but instead showed transient polycythemia. Brain MRI for both children revealed pontocerebellar and corpus callosum hypoplasia with a greater degree of ventriculomegaly in the older child, suggesting a progressive course. The older sibling passed away at 3 years of age, while the younger sibling is still living.

Family 3195 had two affected children (III-3 and III-4) and two healthy siblings from a first-cousin marriage. Both children had severe global motor and cognitive delay with seizure onset at 3 and 7 months of age, respectively. Brain MRI for the younger sibling revealed

pontocerebellar hypoplasia, defective myelination, mild cortical atrophy, corpus callosum hypoplasia, and mild ventriculomegaly. Both children passed away at 4 years of age. DNA or MRIs for one of the affected children (III-4) could not be obtained, however phenotypic similarity between the two siblings suggested a single genetic etiology.

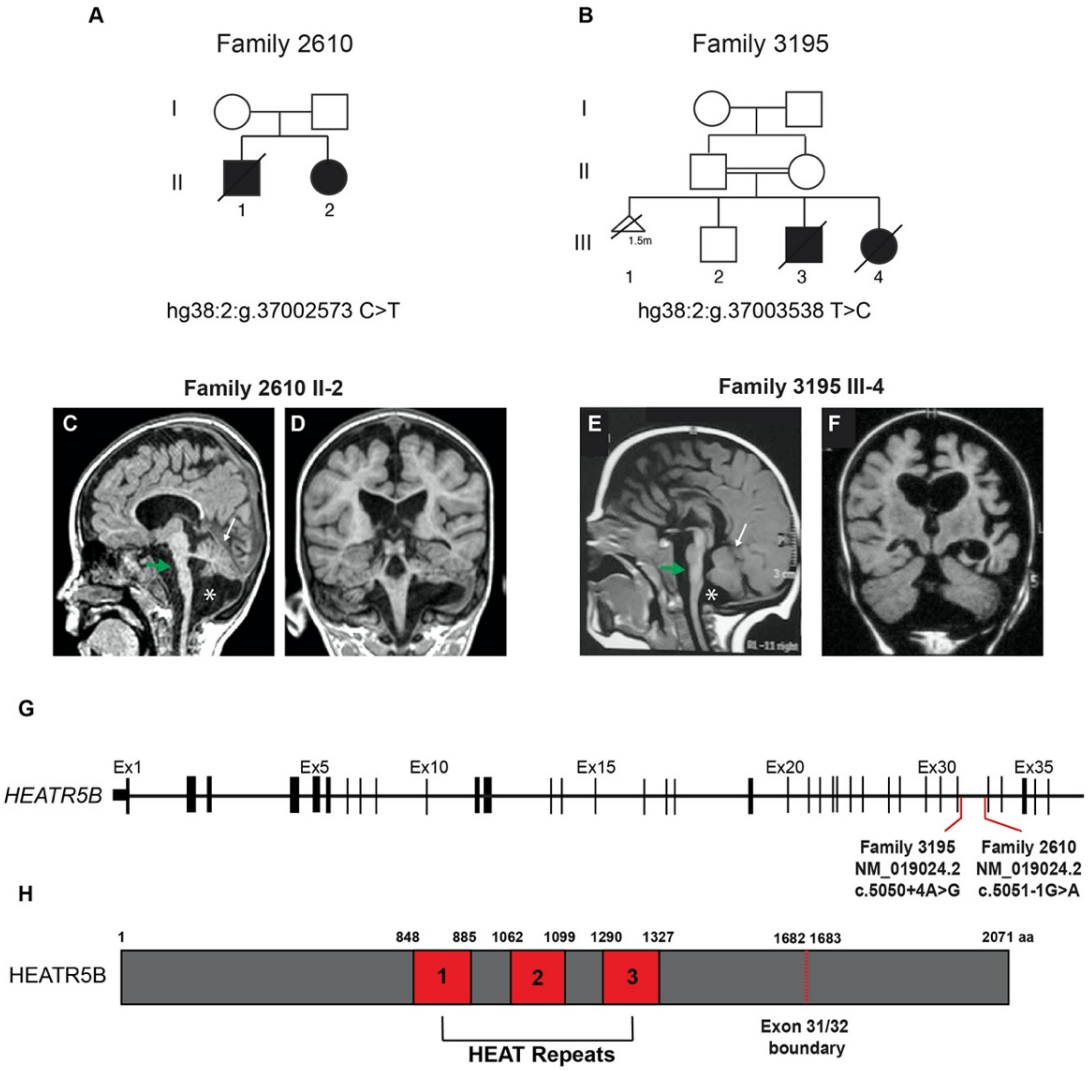

Figure 5.1: Clinical and genetic information for families 2610 and 3195.
(A-B) Family pedigrees for families 2610 and 3195. Hash line: deceased. Triangle: deceased pregnancy at 1.5 months gestation. Double line: reported consanguinity. (C-F) MRIs for individuals 2610-II-2 (C-D) and 3195-III-4 (E-F). Shown are midline sagittal (C, E) and coronal (D, F) images. Asterisk: mega cisterna magna. White arrow: cerebellar hypoplasia. Green arrow: pontine hypoplasia. (G) *HEATR5B* 36 exons with genomic position. Red lines: position of intron 31 pathogenic variants (RefSeq NM_019024.2). (H) Schematic of the HEATR5B protein with location of the three HEAT repeat domains (RefSeq NP_061897.1). Red dashed line highlights the exon 31/32 boundary at amino acids 1682-1683, between which the two splicing variants were present. aa: amino acid.

Table 5.1: Clinical information for families harboring *HEATR5B* mutations.

	PCH-2610-3-1	PCH-2610-3-2	DBD-3195-4-3	DBD-3195-4-4
Mutation gDNA (hg19)	hg38:2:g.37002573C>T	hg38:2:g.37002573C>T	N/A	hg38:2:g.37003538T>C
Mutation cDNA	c.5051-1G>A	c.5051-1G>A	N/A	c.5050+4A>G
Mutation Protein	p.?	p.?	N/A	p.?
Zygosity	homozygous	homozygous	N/A	homozygous
Gender	Female	Male	Male	Female
Ethnic origin	Turkey	Turkey	Egypt	Egypt
Consanguinity	none reported	none reported	1st degree cousins	1st degree cousins
Pregnancy duration (weeks)	36	39	39	38
Weight at birth (g)	1900	3000	2800	3000
OFC at birth (cm)	32	33	34.5	34
OFC at last examination (cm)	43.5 (-2SD)	39.1 (-2SD)	35 (-2SD)	32.5 (-2SD)
Deceased at age (years)	3	3	4	4
Psychomotor development				
Gross motor	Absent	Absent	Delayed	Absent
Fine motor	Absent	Absent	Delayed	Absent
Language	Absent	Absent	Absent	Absent
Social	Absent	Absent	Delayed	Absent
Regression	Y	Y	Y	Y
Seizures				
Onset	Birth	Birth	3 months	7 months
Type	Focal and Generalized	Focal	Myoclonic	Focal and Myoclonic
Frequency	Daily	Daily	Daily	Daily
Neurological findings				
Cognitive functions	Severe ID	Severe ID	Severe ID	Severe ID
Brainstem findings	Dysphagia	Dysphagia	Apenic spells	Dysphagia and apnea
Cerebellar deficits	Nystagmus	Nystagmus	Nystagmus	Nystagmus
Vision	Impaired	Impaired	Impaired	Impaired
Muscle tone	Axial hypotonia	Axial hypotonia	Axial hypotonia	Axial hypotonia
Reflexes	Increased	Increased	Increased	Increased
Neuroimaging			**CT**	**MRI**
Cerebellum	Hypoplasia	Hypoplasia	N/A	Hypoplasia
Pons	Hypoplasia	Hypoplasia	N/A	Hypoplasia
Cerebral cortex	Atrophy	Atrophy	N/A	Atrophy
Ventriculomegaly	Y	Y	N/A	Y
White matter	Volume loss	Volume loss	N/A	Volume loss
Corpus callosum	Hypoplasia	Hypoplasia	N/A	Hypoplasia
Other				
Eye	Micropthalmia	-	-	-
Immunodeficiency	Recurrent infections	Recurrent infections	Recurrent infections	Recurrent infections
Respiratory	Recurrent aspirations	Recurrent aspirations	Recurrent aspirations	Recurrent aspirations
Gastrointestinal	Aganglionic megacolon	Gallbladder stone	-	-
Other Investigations				
EEG	Epileptiform	Epileptiform	Epileptiform	Epileptiform
Metabolic (UOA, PAA)	Negative	Polycyhtemia	Negative	Negative

5.4.2 Homozygous HEATR5B Damaging Variants in Pontocerebellar Hypoplasia

Following informed consent from all participating individuals (or their guardians), whole exome sequencing (WES) was performed on blood-derived DNA from the father, mother, and both affected children of Family 2610, as well as the father, mother and the affected girl from Family 3195. Rare, potentially deleterious variants were prioritized using an in-house exome database consisting of over 5,000 ethnically matched individuals, in addition to publicly available exome datasets, cumulatively numbering over 10,000 individuals.[12,13] In family 2610, only one candidate passed filters and segregated with the phenotype, which we identified as a homozygous variant predicted to affected splicing in *HEATR5B* (Table S2). In Family 3195, the variant in *HEATR5B* was the only splicing variant we identified, among additional missense variants in other genes (Table S2). Thus, both families showed homozygous variants adjacent to canonical splice sites (c.5051-1 G>A for Family 2610 and c.5050+4 A>G for Family 3195) in intron 31 of *HEATR5B* (RefSeq NM_019024.2), predicted to impair splicing.

Neither *HEATR5B* variant was present in the gnomAD database, or in our in-house exome database. The predicted genotypes from next generation sequencing were confirmed through Sanger sequencing for all individuals where DNA was available (Figure 5.2).

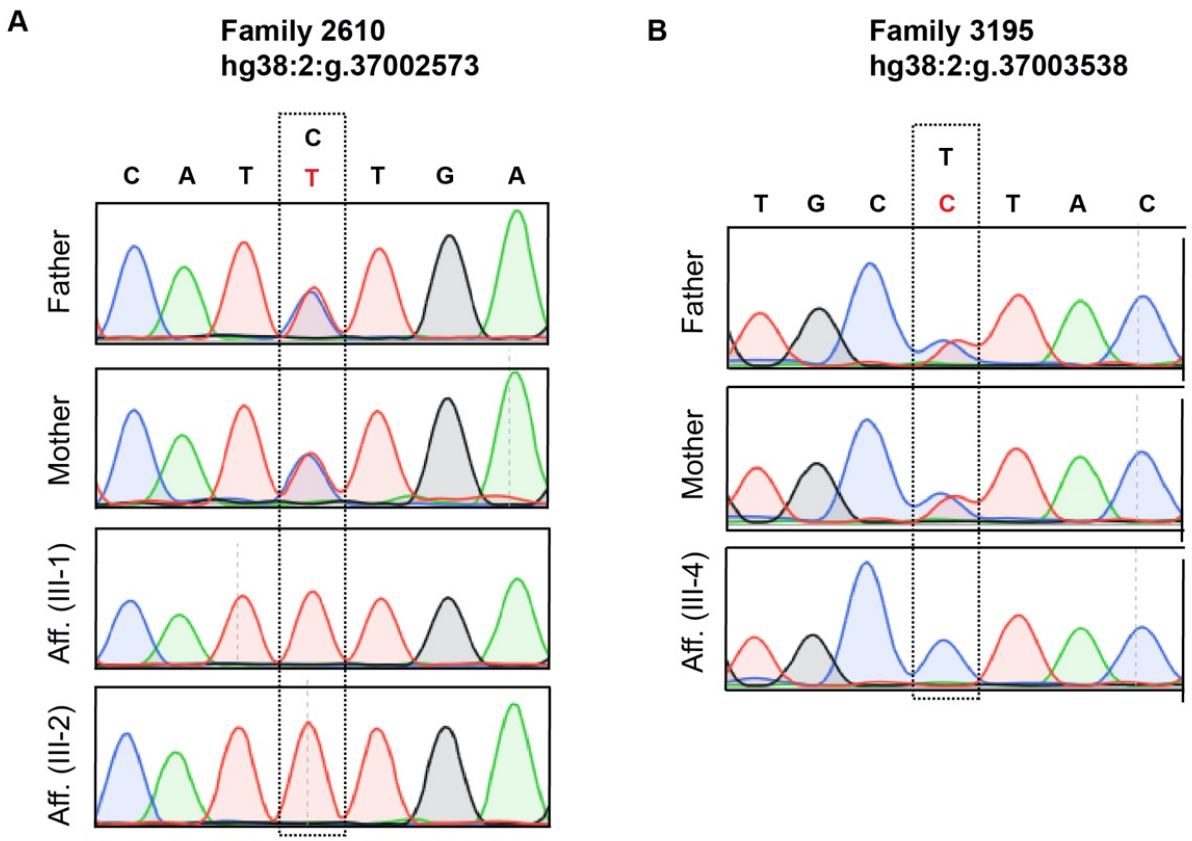

Figure 5.2: Sanger sequencing traces for families 2610 and 3195.
(A) Sanger sequencing traces depicting the C>T nucleotide change present in Family 2610. Shown are heterozygous traces for father, mother, and homozygous traces from both affected members (Aff.1 and 2). (B) Sanger sequencing traces depicting the T>C conversion present in Family 3195. Shown are heterozygous traces for father, mother, and homozygous traces from the affected individual (III-4).

Furthermore, both variants were adjacent to or within a stretch of extended homozygosity (HomozygosityMapper, Figure 5.3), and were predicted as 'disease-causing' by MutationTaster.[8-10] *HEATR5B* is ubiquitously expressed in adult human tissues (GTEx data, Figure 5.4) and during human brain development (BrainSpan data, Figure 5.5) at all assessed stages. The genetic support, the similar location of the variants, as well as the matching phenotypes rendered the *HEATR5B* variants as the most likely cause of disease.

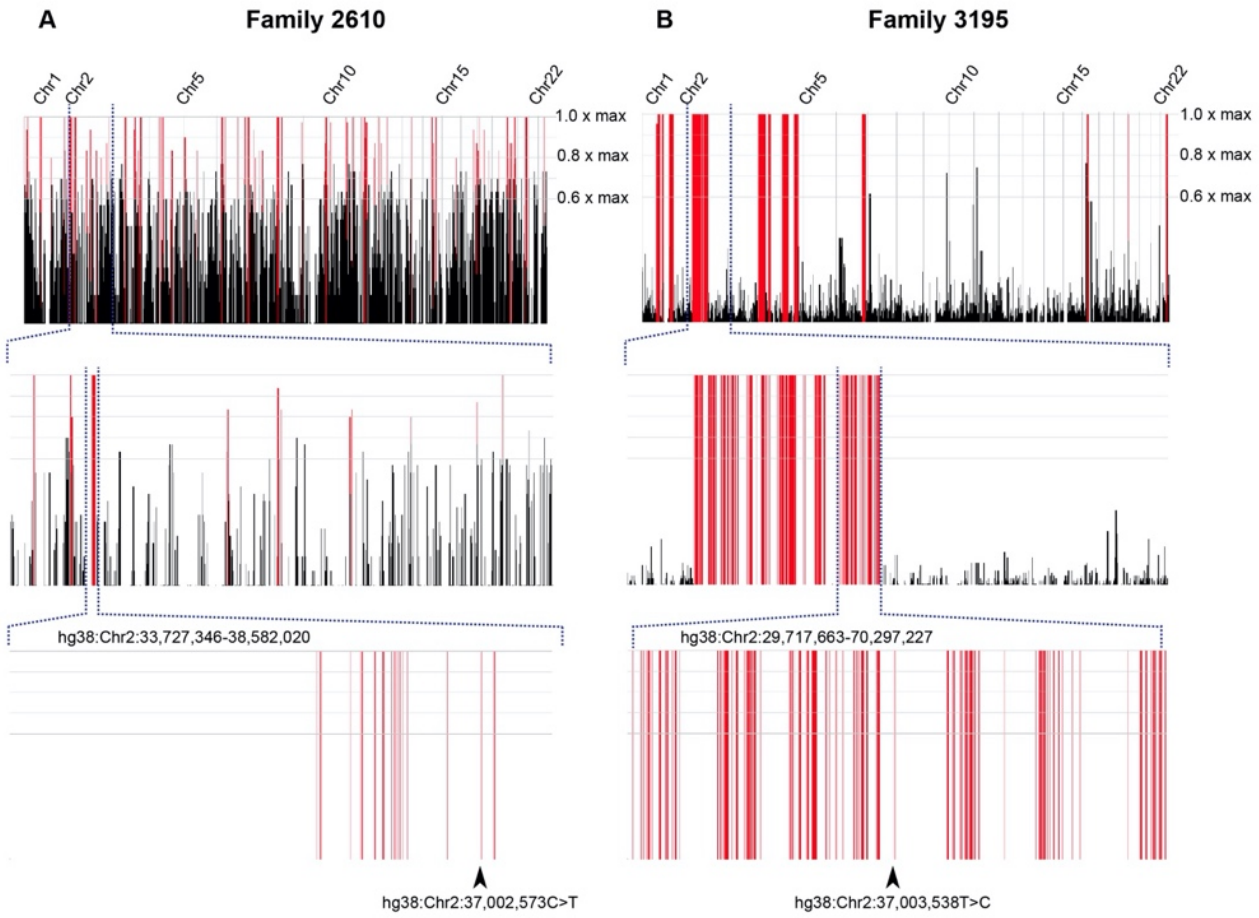

Figure 5.3: Homozygosity plots for families 2610 and 3195.
(A-B) Graphical representation of the homzygosity plots generated by HomozygosityMapper of all 22 autosomes for the affected individuals in Families 2610 (A) and 3195 (B). Red peaks indicate homozygosity scores above the cutoff at 0.8x of the maximum. There is an extensive stretch of homozygosity on chromosome 2. Arrowheads in the lowest panel indicate the positions of the mutations.

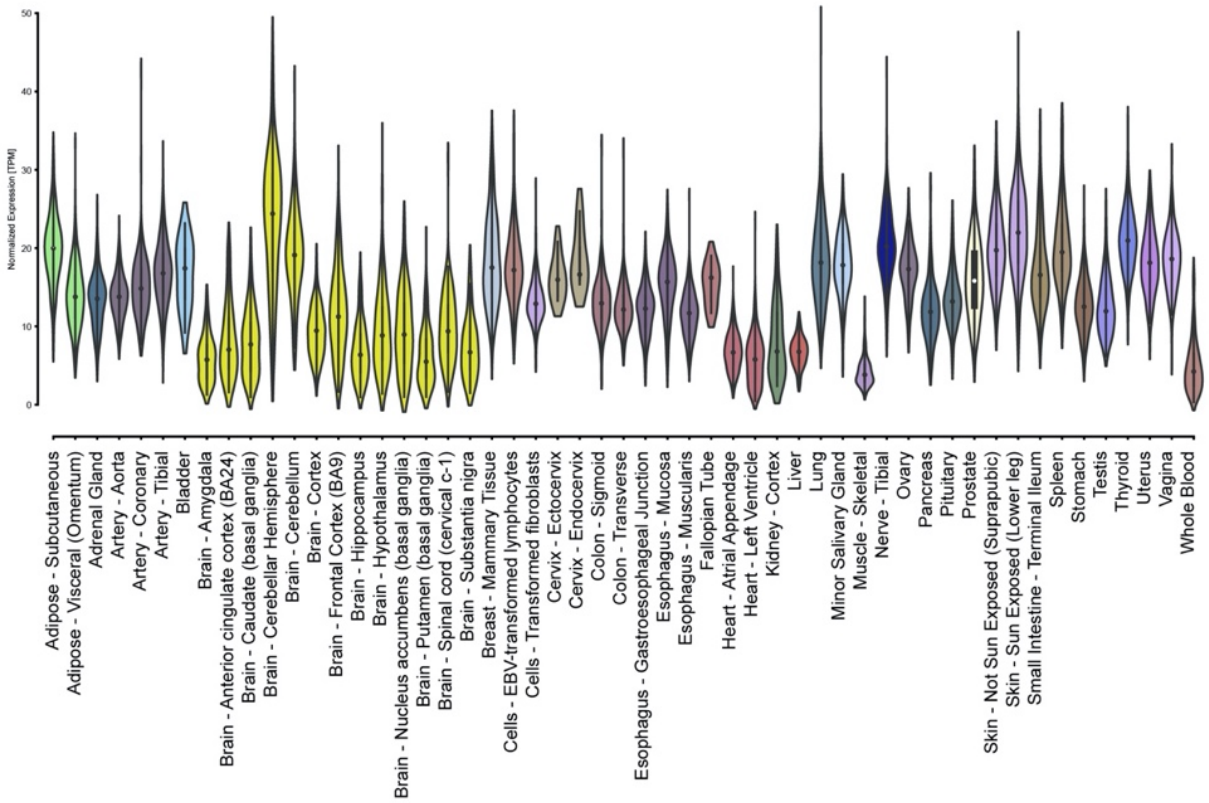

Figure 5.4: *HEATR5B* expression across various tissues in adult humans (GTEx data). Expression analysis of *HEATR5B* transcripts across various adult tissues. *HEATR5B* shows ubiquitous expression across all tissues. Expression is shown as Transcripts per Kilobase Million (TPM). Data used was derived from the GTEx portal.[14,15] Shown are standard violin plots with inner boxplots.

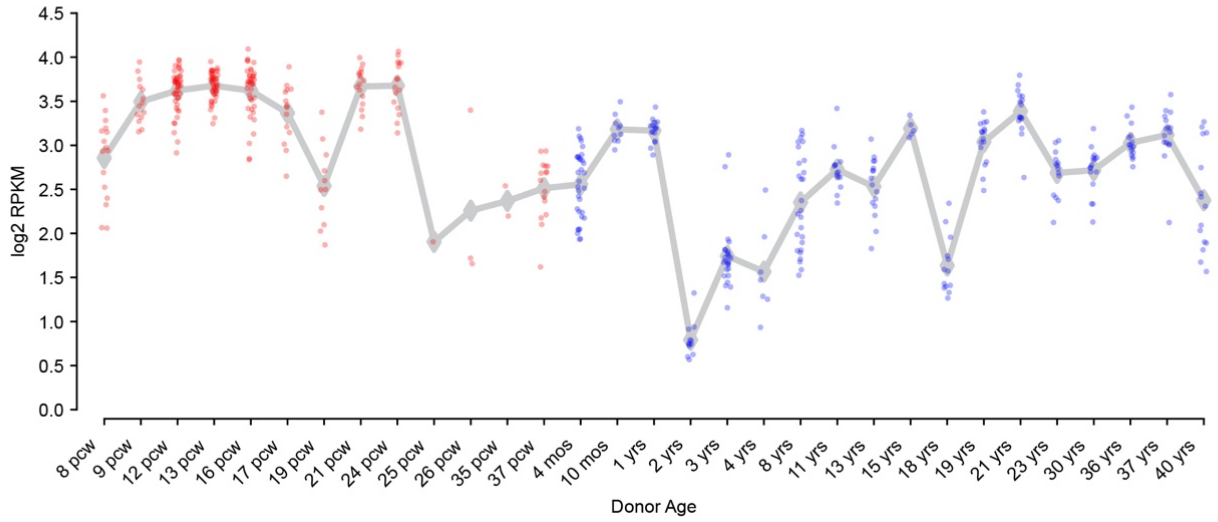

Figure 5.5: *HEATR5B* expression in the human cortical region along the developmental trajectory (BrainSpan data).
Expression analysis of *HEATR5B* transcripts in the developing (red) and postnatal (blue) brain. HEATR5B shows ubiquitous expression across all time points. Expression is shown as the log2 of Reads Per Kilobase of transcript, per Million mapped reads (RPKM). Data used was derived from RNA-Sequencing data from BrainSpan.[16] Shown are mean ± SD and individual data points for each time point. pcw: post-conception weeks; mos: months; yrs: years.

5.4.3 Pathogenic variants in HEATR5B result in aberrant splicing and subsequent reduced protein levels

To functionally assess splicing defects that may result from these variants, we generated dermal fibroblast cultures from the affected girl and both parents in Family 2610 following informed consent. We assessed the presence of *HEATR5B* mRNA using RT-PCR from exon 22 to 23, which are upstream of the variant in Family 2610, and observed no differences in expression level compared with control (Figure 5.6A,D). Primers located in exon 31 and 32, which flank the variant, demonstrated a lower band corresponding to spliced mRNA, which was of decreased intensity in the affected compared with control or parents after normalization of PCR to a GAPDH loading control ($P<0.0001$, Fig. 5.6B,E). Additionally, in the affected there was a new upper band, corresponding to a product in which intron 31 failed to splice out of the

transcript ($P<0.0001$, Fig. 5.6B,F). Primers located in exon 33 and 34, which are downstream of the variant, showed no difference in expression level. The data suggest that the variant in intron 31 can lead to aberrant splicing of the transcript.

We extracted DNA from both the upper and lower bands of the affected individual following RT-PCR. Sequencing of the extracted DNA confirmed intron retention in the upper band (Figure 5.6H-I), and showed a loss of six nucleotides (Figure 5.6H, J) in the lower band. This suggests that loss of the original splice site resulted in two new splicing events: 1] usage of a weaker splice site six base pairs downstream of the start of exon 32/36. 2] failure to splice the intron completely. These result in the loss of two amino acids in the final protein, or higher levels of intron retention, respectively.

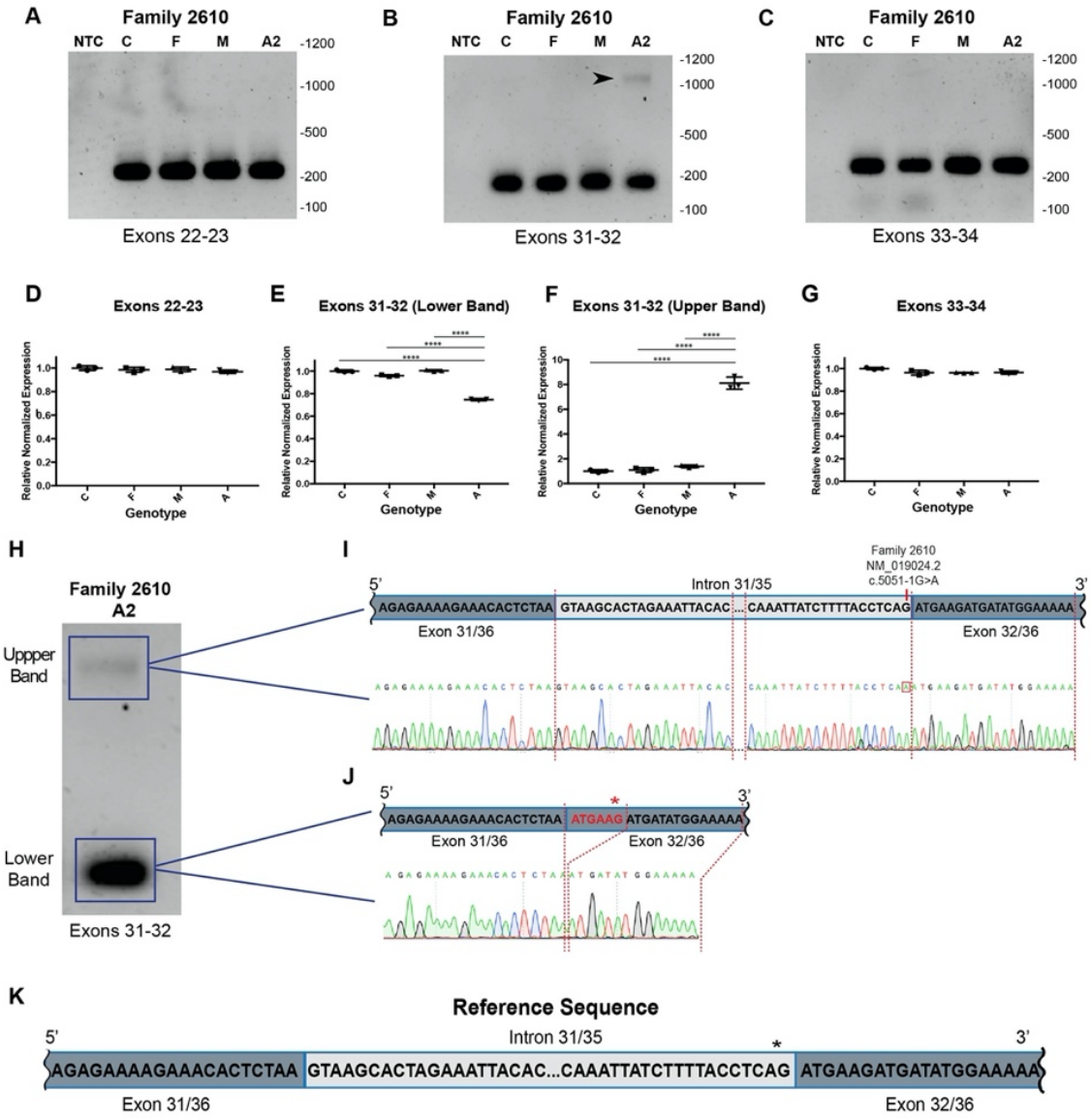

Figure 5.6: Pathogenic variants in *HEATR5B* result in mRNA intron retention.
(A and C) RT-PCR from fibroblasts from unaffected Father (F), Mother (M), and affected child (A2) from Family 2610, for primers located in exons 22 and 23 or exons 33 and 34. Results show indistinguishable band intensities and molecular weight in all. (B) RT-PCR for primers located in exons 31 and 32, spanning the variant, showing reduced intensity of wildtype band (lower band) and presence of a novel upper band, suggesting intron retention in A2 (arrowhead). NTC: Negative Template Control. (D-G) Band quantification of RT-PCR analyses. (H) RT-PCR of sample A2 showing upper and lower bands used for Sanger sequencing. (I) Sequence trace from the upper band revealed retention of intron 31. (J) Sequence trace from the lower band revealed a loss of six nucleotides from the start of exon 32, resulting in the predicted loss of two amino acids in the encoded protein, indicating use of the alternative splice site within exon 32. Red asterisk indicates the new splice site location resulting from the variant. (K) Schematic of *HEATR5B* pre-mRNA sequence of intron 31 of 35. Black asterisk indicates the location of the canonical splice site affected by the variant in Family 2610. Statistical analysis in D-G was done using a one-way ANOVA and Tukey's multiple comparisons test. Unless indicated, differences were not significant between the categories. ****P<0.0001. N=3 independent experiments and graphs show individual data points as well as mean ± SEM.

We next performed Western analysis for HEATR5B protein in whole cell lysates from Family 2610 and observed significantly reduced protein in the sample from the affected compared to both parents and an unrelated control (P<0.0001, Fig. 5.7A-B).

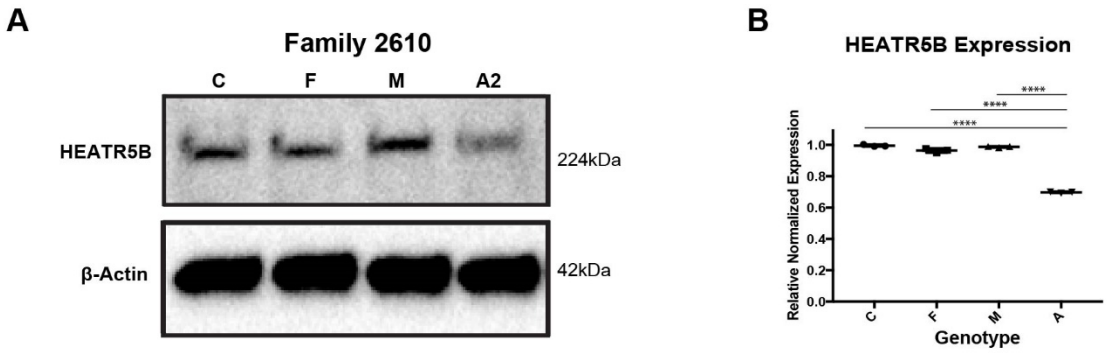

Figure 5.7: Primary dermal fibroblasts from affected member of family 2610 show reduced HEATR5B protein level.
(A) Western blot for HEATR5B full-length protein and β-actin using the same sample input. Full-length HEATR5B signal is reduced in the proband. (B) Band quantification of Western blot analyses. Statistical analysis done using a one-way ANOVA and Tukey's multiple comparisons test. Unless indicated, differences were not significant between the categories. ****P<0.0001. N=3 independent experiments and graph shows individual data points as well as mean ± SEM.

5.4.4 Embryonic Lethality of Heatr5b Mouse Model of Disease

To investigate the importance of HEATR5B in development, we generated a frameshift allele (Refseq NM_001081179.1; c.4384_4385insA; p.Asp1462GlufsTer15) in the mouse homolog *Heatr5b* using CRISPR/Cas9 pronuclear injection (Figure 5.8). While heterozygous animals were healthy and fertile, we failed to recover any homozygous mutant mice at P0. We thus established timed-pregnant matings but also failed to recover any healthy embryos genotyped as homozygous mutants at embryonic day (E) 12.5 or 14.5 (0/21 embryos total; P=0.0041). These results are consistent with an essential role in embryonic development and supported pathogenicity of HEATR5B function perturbation.

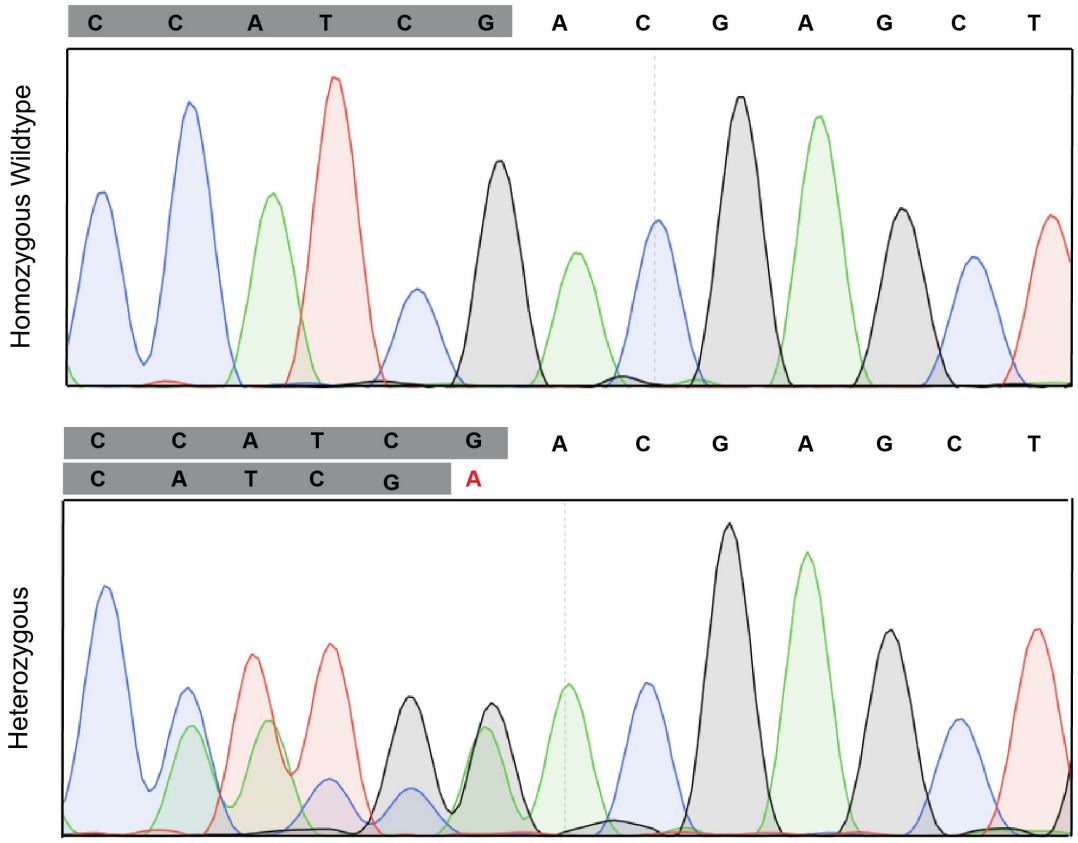

Figure 5.8: Sanger sequencing traces for wildtype and heterozygous mice.
Sanger sequencing traces depicting the frameshift allele generated in the heterozygous *Heatr5b* embryonic mice following CRISPR/Cas9 pronuclear injection. The A nucleotide insertion is shown in red and the shared wildtype sequence is shaded in grey.

5.5 Discussion

Here, we describe two unrelated families with four affected individuals that exhibited a similar phenotypic spectrum comprising severe global motor and developmental delay, pontocerebellar hypoplasia, and epilepsy. Both families harbored unique splice site variants in *HEATR5B* that segregated with the phenotype in affected individuals, likely impaired splicing, and resulted in reduced steady-state levels of the encoded HEATR5B protein in a fibroblast culture. Embryonic lethality of *Heatr5b* knockout mice further supports the significance of this protein in development. This severe phenotype in the knockout mouse suggests that these patient

splice site variants are hypomorphic, which is supported by the residual expression of protein in fibroblasts that are homozygous for one of these variants.

While the cellular role of the HEATR5B protein in mammals remains unclear, recent studies on the yeast homolog of this protein, Laa1, have provided evidence for its importance in endocytosis; this is based on its interaction with the adaptor protein complex-1 (AP-1),[17] a highly conserved clathrin adaptor involved in membrane trafficking.[18,19] This association of HEATR5 proteins with endocytosis is evolutionarily conserved from yeast to mammals.[20-23] Although disruption of the HEATR5 family causes strong defects in AP-1 recruitment in all systems tested so far, it remains unclear whether HEATR5B plays a direct role in AP-1 function.[20-22] At the minimum, the identification of deleterious splice site variants in patients with pontocerebellar hypoplasia and epilepsy, suggests that its function is important for mammalian neurodevelopment.

Eleven subtypes of PCH have been described thus far, with two subtypes (PCH 3 and PCH 11) previously implicated in intracellular vesicle transport or vesicle formation.[24-27] Due to the proposed role of HEATR5B in endocytosis, it is possible that this newly-associated HEATR5B functions in endocytosis although more work is required to assess function, and there is no obvious convergence on a subtype-specific set of phenotypes, due to the limited number of cases reported in the literature.

Chapter 5, in full, has been submitted for publication of the material as it may appear in European Journal of Human Genetics 2020. Ghosh, Shereen G.; Breuss, Martin W.; Schlachetzki, Zinayida; Chai, Guoliang; Ross, Danica; Stanley, Valentina; Sonmez, F.Mujgan; Topaloglu, Haluk; Zaki, Maha S.; Hosny, Heba; Gad, Shaimaa; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this material.

5.6 References

1. Sabine Rudnik-Schöneborn, Peter G Barth, and Klaus Zerres. Pontocerebellar hypoplasia. *Am J Med Genet C Semin Med Genet*, 2014.

2. Tessa van Dijk, Frank Baas, Peter G Barth, and Bwee Tien Poll-The. What's new in pontocerebellar hypoplasia? An update on genes and subtypes. *Orphanet J Rare Dis*, 2018.

3. Matthew R. Groves, Neil Hanlon, Patric Turowski, Brian A. Hemmings, and David Barford. The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell*, 1999.

4. Andrew Neuwald and Tatsuya Hirano. HEAT repeats associated with condensins, cohesins and other complexes involved in chromosome-related functions. *Genome Res*, 2000.

5. Alison Grinthal, Ivana Adamovic, Beth Weiner, Martin Karplus, and Nancy Kleckner. PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. *Proc Natl Acad Sci USA*, 2010.

6. Amjad Horani, Todd E. Druley, Maimoona A. Zariwala, Anand C. Patel, Benjamin T. Levinson, Laura G. Van Arendonk, Katherine C. Thornton, Joe C. Giacalone, Alison J. Albee, Kate S. Wilson, Emily H. Turner, Deborah A. Nickerson, Jay Shendure, Philip V. Bayly, Margaret W. Leigh, Michael R. Knowles, Steven L. Brody, Susan K. Dutcher, and Thomas W. Ferkol. Whole-exome capture and sequencing identifies HEATR2 mutation as a cause of primary ciliary dyskinesia. *Am J Hum Genet*, 2012.

7. Tracy J. Dixon-Salazar, Jennifer L. Silhavy, Nitin Udpa, Jana Schroth, Stephanie Bielas, Ashleigh E. Schaffer, Jesus Olvera, Vineet Bafna, Maha S. Zaki, Ghada H. Abdel-Salam, Lobna A. Mansour, Laila Selim, Sawsan Abdel-Hadi, Naima Marzouki, Tawfeg Ben-Omran, Nouriya A. Al-Saana, F. Müjgan Sonmez, Figen Celep, Matloob Azam, Kiley J. Hill, Adrienne Collazo, Ali G. Fenstermaker, Gaia Novarino, Naiara Akizu, Kiran V. Garimella, Carrie Sougnez, Carsten Russ, Stacey B. Gabriel, and Joseph G. Gleeson. Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med*, 2012.

8. Dominik Seelow, Markus Schuelke, Friedhelm Hildebrandt, and Peter Nurnberg. HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic Acids Res*, 2009.

9. Dominik Seelow and Markus Schuelke. HomozygosityMapper2012--bridging the gap between homozygosity mapping and deep sequencing. *Nucleic Acids Res*, 2012.

10. Jana Marie Schwarz, Christian Rödelsperger, Markus Schuelke, and Dominik Seelow. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*, 2010.

11. Jacob Villegas and Michael McPhaul. Establishment and culture of human skin fibroblasts. *Curr Protoc Mol Biol*, 2005.

12. Monkol Lek, Konrad J. Karczewski, Eric V. Minikel, Kaitlin E. Samocha, Eric Banks, Timothy Fennell, Anne H. O'Donnell-Luria, James S. Ware, Andrew J. Hill, Beryl B. Cummings, Taru Tukiainen, Daniel P. Birnbaum, Jack A. Kosmicki, Laramie E. Duncan, Karol Estrada, Fengmei Zhao, James Zou, Emma Pierce-Hoffman, Joanne Berghout, David N. Cooper, Nicole Deflaux, Mark DePristo, Ron Do, Jason Flannick, Menachem Fromer, Laura Gauthier, Jackie Goldstein, Namrata Gupta, Daniel Howrigan, Adam Kiezun, Mitja I. Kurki, Ami Levy Moonshine, Pradeep Natarajan, Lorena Orozco, Gina M. Peloso, Ryan Poplin, Manuel A. Rivas, Valentin Ruano-Rubio, Samuel A. Rose, Douglas M. Ruderfer, Khalid Shakir, Peter D. Stenson, Christine Stevens, Brett P. Thomas, Grace Tiao, Maria T. Tusie-Luna, Ben Weisburd, Hong-Hee Won, Dongmei Yu, David M. Altshuler, Diego Ardissino, Michael Boehnke, John Danesh, Stacey Donnelly, Roberto Elosua, Jose C. Florez, Stacey B. Gabriel, Gad Getz, Stephen J. Glatt, Christina M. Hultman, Sekar Kathiresan, Markku Laakso, Steven McCarroll, Mark I. McCarthy, Dermot McGovern, Ruth McPherson, Benjamin M. Neale, Aarno Palotie, Shaun M. Purcell, Danish Saleheen, Jeremiah M. Scharf, Pamela Sklar, Patrick F. Sullivan, Jaakko Tuomilehto, Ming T. Tsuang, Hugh C. Watkins, James G. Wilson, Mark J. Daly, Daniel G. MacArthur, and Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 2016.

13. Eric M. Scott, Anason Halees, Yuval Itan, Emily G. Spencer, Yupeng He, Mostafa Abdellateef Azab, Stacey B. Gabriel, Aziz Belkadi, Bertrand Boisson, Laurent Abel, Andrew G. Clark, Greater Middle East Variome Consortium, Fowzan S. Alkuraya, Jean-Laurent Casanova, and Joseph G. Gleeson. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet*, 2016.

14. Latarsha J. Carithers, Kristin Ardlie, Mary Barcus, Philip A. Branton, Angela Britton, Stephen A. Buia, Carolyn C. Compton, David S. DeLuca, Joanne Peter-Demchok, Ellen T. Gelfand, Ping Guan, Greg E. Korzeniewski, Nicole C. Lockhart, Chana A. Rabiner, Abhi K. Rao, Karna L. Robinson, Nancy V. Roche, Sherilyn J. Sawyer, Ayellet V. Segrè, Charles E. Shive, Anna M. Smith, Leslie H.

Sobin, Anita H. Undale, Kimberly M. Valentino, Jim Vaught, Taylor R. Young, Helen M. Moore, and GTEx Consortium. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank*, 2015.

15. GTEx-Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*, 2013.

16. Jeremy A. Miller, Song-Lin Ding, Susan M. Sunkin, Kimberly A. Smith, Lydia Ng, Aaron Szafer, Amanda Ebbert, Zackery L. Riley, Joshua J. Royall, Kaylynn Aiona, James M. Arnold, Crissa Bennet, Darren Bertagnolli, Krissy Brouner, Stephanie Butler, Shiella Caldejon, Anita Carey, Christine Cuhaciyan, Rachel A. Dalley, Nick Dee, Tim A. Dolbeare, Benjamin A.C. Facer, David Feng, Tim P. Fliss, Garrett Gee, Jeff Goldy, Lindsey Gourley, Benjamin W. Gregor, Guangyu Gu, Robert E. Howard, Jayson M. Jochim, Chihchau L. Kuan, Christopher Lau, Chang-Kyu Lee, Felix Lee, Tracy A. Lemon, Phil Lesnar, Bergen McMurray, Naveed Mastan, Nerick Mosqueda, Theresa Naluai-Cecchini, Nhan-Kiet Ngo, Julie Nyhus, Aaron Oldre, Eric Olson, Jody Parente, Patrick D. Parker, Sheana E. Parry, Allison Stevens, Mihovil Pletikos, Melissa Reding, Kate Roll, David Sandman, Melaine Sarreal, Sheila Shapouri, Nadiya V. Shapovalova, Elaine H. Shen, Nathan Sjoquist, Clifford R. Slaughterbeck, Michael Smith, Andy J. Sodt, Derric Williams, Lilla Zöllei, Bruce Fischl, Mark B. Gerstein, Daniel H. Geschwind, Ian A. Glass, Michael J. Hawrylycz, Robert F. Hevner, Hao Huang, Allan R. Jones, James A. Knowles, Pat Levitt, John W. Phillips, Nenad Sestan, Paul Wohnoutka, Chinh Dang, Amy Bernard, John G. Hohmann, and Ed S. Lein. Transcriptional landscape of the prenatal human brain. *Nature*, 2014.

17. Christopher J. Zysnarski, Sagar Lahiri, Fatima T. Javed, Jorge Y. Martínez-Márquez, Justin W. Trowbridge, and Mara C. Duncan. Adaptor protein complex-1 (AP-1) is recruited by the HEATR5 protein Laa1 and its co-factor Laa2 in yeast. *J Biol Chem*, 2019.

18. Bertram J. Canagarajah, Xuefeng Ren, Juan S. Bonifacino, and James H. Hurley. The clathrin adaptor complexes as a paradigm for membrane-associated allostery. *Protein Sci*, 2013.

19. Linton M. Traub. Common principles in clathrin-mediated sorting at the Golgi and the plasma membrane. *Biochim Biophys Acta*, 2005.

20. G. Esteban Fernandez and Gregory S. Payne. Laa1p, a conserved AP-1 accessory protein important for AP-1 localization in yeast. *Mol Biol Cell*, 2006.

21. Ghislain Gillard, Massiullah Shafaq-Zadah, Ophélie Nicolle, Raghida Damaj, Jacques Pécréaux, and Grégoire Michaux. Control of E-cadherin apical localisation and morphogenesis by a SOAP-1/AP-1/clathrin pathway in C. elegans epidermal cells. *Development*, 2015.

22. Yang Yu, Ayako Kita, Masako Udo, Yuta Katayama, Mami Shintani, Kwihwa Park, Kanako Hagihara, Nanae Umeda, and Reiko Sugiura. Sip1, a conserved AP-1 accessory protein, is important for Golgi/endosome trafficking in fission yeast. *PLoS One*, 2012.

23. Stéphanie Le Bras, Christine Rondanino, Géraldine Kriegel-Taki, Aurore Dussert, and Rolande Le Borgne. Genetic identification of intracellular trafficking regulators involved in Notch-dependent binary cell fate acquisition following asymmetric cell division. *J Cell Sci*, 2012.

24. Mustafa Y. Ahmed, Barry A. Chioza, Anna Rajab, Klaus Schmitz-Abe, Aisha Al-Khayat, Saeed Al-Turki, Emma L. Baple, Michael A. Patton, Ali Y. Al-Memar, Matthew E. Hurles, Jennifer N. Partlow, R. Sean Hill, Gilad D. Evrony, Sarah Servattalab, Kyriacos Markianos, Christopher A. Walsh, Andrew H. Crosby, and Ganeshwaran H. Mochida. Loss of PCLO function underlies pontocerebellar hypoplasia type III. *Neurology*, 2015.

25. Burak Durmaz, Bernd Wollnik, Ozgur Cogulu, Yun Li, Hasan Tekgul, Filiz Hazan, and Ferda Ozkinay. Pontocerebellar hypoplasia type III (CLAM): extended phenotype and novel molecular findings. *J Neurol*, 2009.

26. Ekaterina L. Ivanova, Frédéric Tran Mau-Them, Saima Riazuddin, Kimia Kahrizi, Vincent Laugel, Elise Schaefer, Anne de Saint Martin, Karen Runge, Zafar Iqbal, Marie-Aude Spitz, Mary Laura, Nathalie Drouot, Bénédicte Gérard, Jean-François Deleuze, Arjan P.M. de Brouwer, Attia Razzaq, Hélène Dollfus, Muhammad Zaman Assir, Patrick Nitchké, Maria-Victoria Hinckelmann, Hilger Ropers, Sheikh Riazuddin, Hossein Najmabadi, Hans van Bokhoven, and Jamel Chelly. Homozygous truncating variants in TBC1D23 cause pontocerebellar hypoplasia and Alter cortical development. *Am J Hum Genet*, 2017.

27. Isaac Marin-Valencia, Andreas Gerondopoulos, Maha S. Zaki, Tawfeg Ben-Omran, Mariam Almureikhi, Ercan Demir, Alicia Guemez-Gamboa, Anne Gregor, Mahmoud Y. Issa, Bart Appelhof, Susanne Roosing, Damir Musaev, Basak Rosti, Sara Wirth, Valentina Stanley, Frank Baas, Francis A. Barr, and Joseph G. Gleeson. Homozygous mutations in TBC1D23 lead to a non-degenerative form of pontocerebellar hypoplasia. *Am J Hum Genet*, 2017.

CHAPTER 6

Biallelic Variants in *HPDL*, Encoding 4-Hydroxyphenylpyruvate Dioxygenase-Like Protein, lead to an Infantile Neurodegenerative Condition

6.1 Abstract

Dioxygenases are oxidoreductase enzymes with roles in metabolic pathways necessary for aerobic life. 4-hydroxyphenylpyruvate dioxygenase-like protein (HPDL), encoded by *HPDL,* is an orphan paralogue of 4-hydroxyphenylpyruvate dioxygenase (HPD), an iron-dependent dioxygenase involved in tyrosine catabolism. The function and association of HPDL with human diseases remains unknown. We applied exome sequencing in a cohort of over 10,000 individuals with neurodevelopmental diseases. We identified biallelic variants in *HPDL* in 8 families displaying recessive inheritance. Effects of HPDL loss were investigated *in vitro* and *in vivo*, and through mass spectrometry analysis. Evolutionary analysis was performed to investigate the potential functional separation of HPDL from HPD. Knockout mice closely phenocopied humans and showed evidence of apoptosis in multiple cellular lineages within the cerebral cortex. *HPDL* is a single exonic gene that likely arose from a retrotransposition event at the base of the tetrapod lineage, and unlike HPD, HPDL is mitochondria-localized. Metabolic profiling of *HPDL* mutant cells and mice showed no evidence of altered tyrosine metabolites, but rather notable accumulations in other metabolic pathways. The mitochondrial localization, along with its disrupted metabolic profile, suggests *HPDL* loss in humans links to a unique neurometabolic mitochondrial infantile neurodegenerative condition.

6.2 Introduction

Disorders involving neurometabolism can lead to both structural and functional disturbances of the nervous system through multiple mechanisms that include abnormal accumulation of toxic substrates, depletion of key energy or metabolic intermediates, or cell death.[1] Although pediatric-onset brain diseases are often associated with genetic abnormalities, the link between metabolic impairments and brain disease has been associated with relatively common disorders of the nervous system.[2] Especially in highly consanguineous populations, congenital malformations of the nervous system and neurometabolic disorders are often linked.[3] Genes involved in inborn errors of metabolism often encode enzymes that catalyze specific biochemical reactions or required cofactors.[4] Absent or abnormal functioning of such proteins leads to an accumulation of upstream substrates or deficiency of downstream products, with detrimental disease-causing effects.

Dioxygenases are oxidoreductase enzymes that utilize dioxygen (i.e., O_2), most often complexed with iron, in oxidation reactions involved in metabolic pathways or as oxygen sensors.[5] To date, 15 dioxygenases have been linked to diseases that are mostly autosomal recessive including encephalopathy, alkaptonuria (MIM: 203500), and intellectual disability.[6-8] Dioxygenases can oxidize either small molecules, as in tryptophan dioxygenase,[9] or residues on macromolecules, as in EGLN1, a HIF proline dioxygenase acting as an oxygen sensor.[10]

4-hydroxyphenylpyruvate dioxygenase (HPD, MIM: 609695) is a non-heme Fe(II)-dependent dioxygenase that catalyzes the second of four steps in L-tyrosine catabolism in nearly all aerobic cells: the conversion of 4-hydroxyphenylpyruvate into homogentisate.[11] Biallelic loss associates with recessive tyrosinemia type III (MIM:276710), the least frequent form of three genetic tyrosinemias.[12] Clinical features are milder than those of other genetic forms, and can

include seizures, intellectual disability, and intermittent ataxia.[13,14] Serum tyrosine level is variably elevated by 5- to 20-fold (148-1,769 μmol/L, reference 29-86 μmol/L).[14-17]

In addition to biallelic variants, a particular gain-of-function *HPD* variant (p.Ala33Thr) associates with the dominant disease hawkinsinuria (MIM: 140350),[15] as mutant HPD can convert the normal substrate into a reactive epoxide intermediate but cannot rearrange the intermediate to homogentisate. As a result, the reactive epoxide is detoxified to either hawkinsin or hydroxycyclohexyl acetate by glutathione transferase and epoxide hydrolase, respectively. Hawkinsinuria manifests clinically with metabolic acidosis in early childhood, but most patients have normal long-term outcome.

HPDL encodes 4-hydroxyphenylpyruvate dioxygenase-like protein (HPDL), the only mammalian paralogue of HPD, and shares 44% sequence similarity. The origin and function of *HPDL* is not known, and the gene has not been linked to human disease previously. Here we describe that *HPDL* biallelic loss leads to a unique neurometabolic, infantile neurodegenerative condition.

6.3 Materials and Methods

6.3.1 Patient Recruitment

The Institutional Review Board at the University of California, San Diego, approved this study. All study participants signed informed consent documents, and the study was performed in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rules. The procedures followed for recruitment and data collection were in accordance with the ethical standards of the responsible committee on human experimentation at the respective participating institute, and proper informed consent was obtained.

6.3.2 DNA Extraction Sequencing

Samples of the probands were subjected to exome sequencing. DNA was extracted on an Autopure LS instrument (Qiagen, Valencia, CA) with Autopure chemistry according to the manufacturer's instructions. Samples of the probands were subjected to Agilent Sure-Select Human All Exon v2.0 (44Mb_baited target) library preparation sequencing on an Illumina HiSeq 2000 with v2 chemistry (Read Length: 151). The accession number for these data is dbGAP: phs000288.v1.p1.

6.3.3 Computational Analysis

Variant calling and filtering were performed according to an established exome sequencing pipeline.[18] Variants were filtered if not consistent with recessive monogenic inheritance, if the minor allele frequency (MAF) of gnomAD was >1:10,000, if MAF of local cohort was > 1:1,000, if not moderate or high impact, if CADD PHRED score ≤20, or if not predicted as damaging by either SIFT, PolyPhen, or MutationTaster.

6.3.4 Sanger Sequencing

Sanger sequencing of PCR products was performed using standard procedures. Individual allele sequences were obtained by PCR according to standard protocols. Sequencing tracks were visualized with ApE software.

6.3.5 Cell Culture

HEK293T cells (ATCC) were transfected with expression vectors using Lipofectamine 2000 according to manufacturer's protocol (100 ng of DNA and 0.5 µl of Lipofectamine 2000 per well) and immunostained 48h after transfection.

6.3.6 Cloning Mammalian Expression Vectors for HPD and HPDL

Open reading frame (ORF) of human *HPD* was extracted by reverse transcription PCR of total RNA extracted from donated human liver tissue using Trizol (primer: human_HPD_cloning, Table S1). The ORF of human *HPDL* was obtained by direct amplification of *HPDL* from human genomic DNA (primer: human_HPDL_cloning, Table S1). A stop codon was removed from reverse primers for C-terminal tags for both *HPD* and *HPDL*. ORFs of HPD or HPDL were shuttled into pENTR vector using pENTR™/D-TOPO™ Cloning Kit (Thermo Fisher Scientific, Cat #: K240020). ORFs in pENTR vectors were transferred to pEZYmyc-His (Addgene, Cat #: 18701) or pDEST-CMV-C-EGFP (Addgene, Cat #: 122844) using LR clonase II (Thermo Fisher Scientific, Cat #: 11791100) to produce C-terminal Myc-His or EGFP-tagged HPD or HPDL expression vectors in mammalian cells.

6.3.7 HEK293T Cell Culture and Plasmid Transfection

HEK293T cells (ATCC) were seeded in 8-well chamber slides (Thermo Fisher Scientific, Cat #: 154534PK) at a density of 5 x 10^5 cells per well and maintained in MEM (Gibco Life Technologies, Cat #: 11095-080) supplemented with 10% Fetal Bovine Serum (FBS, Gemini Bio-Products, Cat #: 100-106) and 1% penicillin and streptomycin (Thermo Fisher Scientific, Cat #: 15140-122). 24h after plating, the cells were transfected with the expression vectors using Lipofectamine 2000 according to manufacturer's protocol (100 ng of DNA and 0.5 µl of Lipofectamine 2000 per well). 48h after transfection, cells were washed with PBS, fixed with 4% PFA for 15 min, washed with PBS, and incubated for 1h in blocking solution. Cells were incubated overnight with TOMM20 (Santa Cruz Biotechnology, sc-17765, 1:500) and c-Myc (Santa Cruz Biotechnology, sc-788, 1:100) antibodies in blocking solution. Cells were then washed in PBS, incubated with secondary antibodies (Alexa Fluor™ 647 donkey anti-mouse

IgG (H+L), Thermo Fisher Scientific, Cat# A-31571, 1:1000; AlexaFluor™ 568 donkey anti-rabbit IgG (H+L), Thermo Fisher Scientific, Cat# A10042, 1:1000) for 1h at room temperature, stained with DAPI (ThermoFisher Scientific, D1306, 1:5000) for 15 min, washed in PBS, and mounted in Fluoromount Aqueous Mounting Medium (Millipore Sigma, Cat#: F4680). Single plane images were taken on a Zeiss 880 Airyscan Confocal microscope equipped with 405, 488, 568, 647 nm laser lines and 20x objective. Image processing was in ImageJ.[19]

6.3.8 CACO-2 Cell Culture, Generation of HPDL Knockout Line using CRISPR/Cas9 Technology, Immunofluorescence Stainings, and Imaging

CACO-2 cells were maintained in MEM (Gibco life technologies, Cat#: 11095-080) supplemented with 20% FBS, 1% MEM NEAA (Gibco life technologies, Cat#: 11140-050), and 1% penicillin and streptomycin (Thermo Fisher Scientific, Cat #: 15140-122). To obtain *HPDL* KO CACO-2 cell line, ribonucleoprotein complex of Cas9 and single-guide RNA (sgRNA, sequence: CGU CCG CCA CGU CGA AGC AC) was introduced into the cell by electroporation using Lonza Cell Line NucleofectorTM Kit V (Lonza, Cat#: VVCA-1003) and Nucleofector® 2b Device according to manufacturer's instruction. Transfected cells were plated on a cell culture plate and incubated for 24h. Cells were trypsinized, diluted, and seeded on 96-well plates with a concentration of 0.5 cells/well. Cells with a single colony were expanded and genotyped using Sanger sequencing.

For immunofluorescent analysis, WT and KO cells were seeded onto a 24-well plate (VWR, Cat#: 353047) with Poly-D-Lysine-treated coverslips (Fisher Scientific, Cat #: 08-774-384) at a density of 1 x 10^5 cells/well. 24h after, cells were washed PBS, fixed with 4% PFA for 15 min, washed with PBS, and incubated for 1h in blocking solution. Cells were incubated in primary antibodies in blocking solution at the following dilutions: HPDL (Proteintech, Cat#:

207771-AP, 1:100), HPD (Proteintech, Cat#: 17004-1-AP, 1:100), TOMM20. Cells were washed in PBS and incubated with fluorophore-coupled secondary antibodies (AlexaFluor™ 488 donkey anti-mouse IgG (H+L), Cat#: A-21202,1:1000; AlexaFluor™ donkey anti-rabbit IgG (H+L), Cat#: A-21206, 1:1000) for 1h in blocking solution, stained with DAPI for 15 min, washed in PBS, and mounted in Fluoromount Aqueous Mounting Medium for imaging. Single plane images were taken on a Zeiss 880 Airyscan Confocal microscope equipped with 405, 488, and 568 nm laser lines and 20x objective. Image processing was in Image J.[19]

6.3.9 Tissue Preparation, Immunofluorescence Staining, and Imaging of Mice

For histological analysis, postnatal day 0, 5, and 9-day old mice were anesthetized with isoflurane before cranial dislocation, and transcardially perfused with PBS followed by 4% PFA in PBS. Brain tissue was post-fixed with 4% PFA for 4h and dehydrated in 30% sucrose solution before sectioning. Sections used for immunofluorescence stainings were washed in PBS before Tris-Based Antigen Unmasking solution (Vector Laboratories, Cat#: H-3301) to 90°C. The sections were incubated at room temperature for 30 min, washed in PBS for 5 min and once in PBS-T for 10 min, and incubated in blocking solution (0.3% PBS-T and 5S NDS) for 1h at room temperature. Sections were incubated with primary antibodies at the following dilutions: CC3 (Cell Signaling Technology, Cat#: 9664, 1:200; CTIP2 (Abcam, Cat#: ab28448, 1:400); OLIG2 (Millipore Sigma, Cat#: MABN50, 1:200); GFAP (Abcam, Cat#: ab4674, 1:1000) overnight at 4°C, then washed with PBS-T for 10 min, incubated with secondary antibodies (Alexa Fluor™ 488 donkey anti-rabbit IgG (H+L), Cat# A-21206, 1:1000; AlexaFluor™ 568 donkey anti-mouse IgG (H+L), Cat#: A-10037, 1:1000; AlexaFluor™ 568 goat anti-chicken IgY (H+L), Cat#: A-11041, 1:1000) in blocking solution together with DAPI (ThermoFisher Scientific, Cat#: D1306, 1:5000) for 1h at room temperature, washed with PBS-T for 5 min, and mounted in Fluoromount

Aqueous Mounting Medium (Millipore Sigma, Cat#: F4680). Confocal single-plane images were generated at a Zeiss 880 Airyscan Confocal microscope. Image processing was in Image J.[19] Three mice were analyzed and 2 sections per mouse were counted for statistical analysis.

6.3.10 Animal Experiments

Animal use followed NIH guidelines and was approved by IACUC at the University of California, San Diego. The frameshift variant in *Hpdl* (NM_146256.3:c.59_65del) was introduced using CRISPR/Cas9 (sgRNA sequence: CUU CCA GCC CCU GGC GGU GA) according to standard protocols at the UCSD Transgenic Mouse Core. Mice were genotyped by PCR followed by Sanger sequencing, and those with correct genotypes were backcrossed for at least five generations prior to experimental crosses of heterozygous animals. Timed pregnant animals were obtained by plug checks, where the day of the observed vaginal plug was determined to be E0.5. All mice were group-housed under a 12 hr light/dark cycle with access to chow and water. Methods used for histological analysis, immunofluorescence staining, and imaging are described in Supplementary Methods.

6.3.11 Oxygen Consumption Rate Measurements

The oxygen consumption rate was measured using the Agilent Seahorse XFe96 Analyzer. Forty thousand cells were seeded in poly-D-lysine coated XF96 cell culture plates per well and incubated over night at 37°C with 5% CO_2. Before starting the assay, all wells were washed once with 200 µL XF base media (unbuffered DMEM supplemented with 2 mM L-glutamine, 11 mM glucose, 1 mM sodium pyruvate, pH 7.4) and equilibrated for 1 h at 37°C. During OCR measurement, following inhibitors were used: 2 µM oligomycin, 0.9 µM carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazon (FCCP) and 1 µM rotenone with 1 µM antimycin A. For each, control and knockout cells, ten technical replicates were measured (n = 10) and for each

condition cycles were performed in triplicates with 3 min mixing followed by 3 min measurement. After the assay, normalization was performed by determination of the protein content for each well by employing a Bradford assay (Roti-Quant). The absorption was measured at 595 nm using a CLARIOstar plate reader (BMG Labtech).

6.3.12 Collection of Cell Lysates and Mouse Brain for Metabolomics

HPDL WT and KO CACO-2 cells were plated in 6-well plates and cultured until confluent. The medium was refreshed 24h after plating and 24h before cell collection. Cell collection was done by washing cells with cold PBS (4ºC), followed by cell scraping in 1 ml ice-cold methanol. Mouse brain tissue was homogenized in methanol using a bullet blender, transferred into 1.5 ml Eppendorf tubes, and centrifuged (16,200 g for 10 min at 4ºC), and then supernatants were transferred to new 1.5 ml Eppendorf tubes. The samples were evaporated at 40ºC under nitrogen until dry, and reconstituted with 500 µl of UPLC-grade methanol (room temperature). The reconstituted samples were stored at -80ºC until analysis was performed. All the experiments were performed with three replicates for each condition.

6.3.13 Direct-Infusion Based Metabolomics

A non-quantitative Direct Infusion-High-Resolution Mass Spectrometry metabolomics method was used as previously described with minor adaptations for cell and tissue extracts.[20] In brief, 70 µl of cell lysate in methanol was diluted with 60 µl 0.3% formic acid (Emsure, Darmstadt, Germany) and filtered using a methanol preconditioned 96 well filter plate (Acro prep, 0.2 um GHP, NTRL, 1 ml well; Pall Corporation, Ann Arbor, MI, USA) and a vacuum manifold. The sample filtrate was collected in a 96 well plate (Advion, Ithaca, NY, USA). Samples were analyzed using a TriVersa NanoMate system (Advion, Ithaca, NY, USA) controlled by Chipsoft software (version 8.3.3, Advion). Data acquisition was performed using

Xcalibur software (version 3.0, Thermo Scientific, Waltham, MA, USA). A peak calling pipeline, developed in R programming language, annotated the raw mass spectrometry data according to the Human Metabolome DataBase (HMDB).

6.3.14 Evolutionary Model Construction

Three gene evolution models in codeml (PAML) were used to test for differences in evolutionary rate. The 1 model indicates one omega value for the full tree. The 2 model indicates one omega value for HPD and one omega value for HPDL. The 4 model indicates one omega value for HPD, one for HPD in tetrapods, one for HPDL, and one for HPDL in tetrapods. Parameters for the test are indicated below each phylogeny and the lnl value for the model. 1 model vs 2 model (p-value < .001), 2 model vs 4 model (p-value <<< .001). (Top) dN/dS ratios for three models tested using codeml. Branches scaled by dN/dS ratio values. (Bottom) dN estimates of the three models. Branches scaled by dN estimates.

6.3.15 Gene Order and Synteny

Syntenic regions were visualized using GEve on the comparative genomics online platform using the LastZ alignment algorithm. The HPDL gene was used as the center for synteny in *Homo sapiens*, *Xenopus tropicalis*, and *Danio rerio*. For HPDL, syntenic regions were extended 110000 bps to the left and right of the HPDL gene in *Danio rerio* (NP_001103178.1), *Homo sapiens* (NP_116145.1), and *Xenopus tropicalis* (XP_004914037.1).

6.3.16 Phylogenetic Tree Reconstruction

Amino acid sequences were aligned using PROMALs3d.[21] Three alignments were computed (1) all sequences, (2) HPD sequences, and (3) HPDL sequences. We used a maximum likelihood method using RAxML.[22] on XSEDE via CIPRES to obtain the gene tree topology

using the JTT model and 1000 bootstrap replicates. The gene tree was visualized using iTol,[23] and the final rooted topology was obtained rerooting the tree with a midpoint root.

6.3.17 Estimates of Evolutionary Rates

Branch lengths were obtained directly from the phylogenetic trees. To calculate dN/dS ratios over the tree and subtrees (HPD and HPDL), we used CODEML in the PAML package (version 4.8).[24] To test multiple rates on the tree on CODEML, we used the model = 2 parameter to estimate an HPD, HPDL, HPD tetrapod, and an HPDL tetrapod rate. Pairwise dN estimates were used using the runmode = -2 parameter. To perform the CODEML analysis, the CDS alignment was performed using TranslatorX.[25] The 'dist.alignment' function was called in seqinr to compute the pairwise distance matrix using the amino acid alignment using both identity and similarity.[26]

6.4 Results

6.4.1 Identification of an Infantile-onset Neurodegenerative Condition in 8 Families with Biallelic Mutations in HPDL

We identified eight independent consanguineous families carrying *HPDL* biallelic variants leading to a range of neurological phenotypes, which included spastic tetraplegia, microcephaly, brain atrophy, epilepsy, and severe intellectual and motor disability (Figure 6.1A and Supplemental File 4). After obtaining informed consent from all participating individuals in accordance with the ethical standards set by the University of California San Diego IRB, we identified a total of eight distinct variants in *HPDL* in the eight families by exome sequencing. We recruited Family 1 with three affected siblings showing limb spasticity and mild motor and intellectual disability. Although parental consanguinity suggested the presence of a homozygous

variant as the cause, no homozygous variants in any gene passed filter criteria. Instead, we identified compound heterozygous p.Gly50Asp; p.Tyr118* variants in *HPDL,* which was the only candidate identified by exome sequencing and segregated in the family. Family 2 was recruited with documented parental consanguinity with one affected girl showing a nearly identical clinical pattern as Family 1. We identified a homozygous p.Gly50Asp variant in the affected individual as the most likely cause of disease, one of the same variants identified in Family 1.

Through collaboration with colleagues, we identified four additional families (3, 4, 5, and 8) with homozygous missense variants and two additional families (6 and 7) with homozygous truncating variants in *HPDL* independently identified as likely most pathogenic based upon inheritance pattern, computational prediction, population allele frequency, and segregation (Tables 6.1). These variants included homozygous p.Ala78Thr (Family 3), p.Gly126Ser (Family 4), p.Leu164Pro (Family 5), p.Gly319Argfs*15 (Family 6), p.Gln32* (Family 7), and p.Gly301Val (Family 8). Thus, we identified a total of 8 families comprising 17 individuals with biallelic *HPDL* damaging variants with variable phenotypes. Because HPDL is encoded by a single exon, and thus not subject to alternative splicing, the presence of homozygous early truncating variants suggests *HPDL* loss of function as the mechanism of this infantile-onset neurodegenerative disease.

Affected children were born full-term without complications during pregnancy or delivery (Supplemental File 4). Where data was available, birth weight and head circumference were not remarkable. However, most affected individuals showed a trend toward smaller head circumference by 5-years of age; and 7 out of 14 affected individuals, where data was available, met criteria for microcephaly, defined as a head circumference <3 standard deviation (SD) below

mean (Supplemental File 4). 6 out of 14 affected individuals also displayed nonspecific facial dysmorphisms. Most individuals presented with minimal to absent psychomotor development, including gross motor skills, language, and social skills. Most displayed spontaneous epileptic seizures starting by 1 year. Seizures were typically myoclonic, focal, or tonic and occurred daily to weekly. Neurological findings included hypertonia, hyperreflexia, spastic gait, and positive Babinski sign, features that became more pronounced over time, suggesting a progressive course. Many individuals were unable to walk and displayed absent language skills. In six severely affected individuals, brain MRIs were available for review, demonstrating cortical atrophy, white matter hyperintensity, corpus callosum thinning, reduced size of the cerebellum, ventriculomegaly and increased extra-axial fluid (Figure 6.1B), suggesting loss of brain parenchyma. Brain atrophy and the progressive course were consistent with neurodegeneration.

HPDL encodes 371 amino acids containing a mitochondrial targeting sequence (MTS), two vicinal oxygen chelate (VOC) domains, and three iron-binding sites. The locations of all the variants occurred within the two VOC domains (Figure 6.1C), thus predicted to impact enzymatic function. Of note, the phenotypes between homozygous truncating variants (p.Gln32* and p.Gly319Argfs*15), destroying at least one iron-binding site, and three of the missense variants (p.Ala78Thr, p.Gly126Ser, and p.Leu164Pro), were indistinguishable, suggesting that the missense variants cause loss of function of HPDL. The p.Gly50Asp variant was found only in Families 1 and 2, and was associated with milder phenotypes compared to the more severely affected individuals in other families with brain atrophy and epilepsy, suggesting that this variant has a milder impact on HPDL function. The variants were unique in our dataset of >5,000 exomes from individuals with neurodevelopmental phenotypes, were predicted to be disease-causing, and were not represented (p.Ala78Thr, p.Leu164Pro, p.Gly301Val, and

p.Gly319Argfs*) or very rare {p.Gln32* [allele frequency (AF) = 8.46e-6], p.Gly50Asp [AF = 1.02e-5], p.Tyr118* [AF = 3.19e-5], and p.Gly126Ser [AF = 8.67e-6]}, in the gnomAD database.[27] No instances of biallelic damaging variants in *HPDL* among over 100,000 sequenced individuals occur in gnomAD (Fisher exact test P< 0.00001). Additionally, all missense variants are highly conserved among vertebrates (Figure 6.1D). All variants were confirmed by Sanger sequencing and segregated according to a recessive mode of inheritance, with carriers showing no apparent clinical features (Figure 6.2).

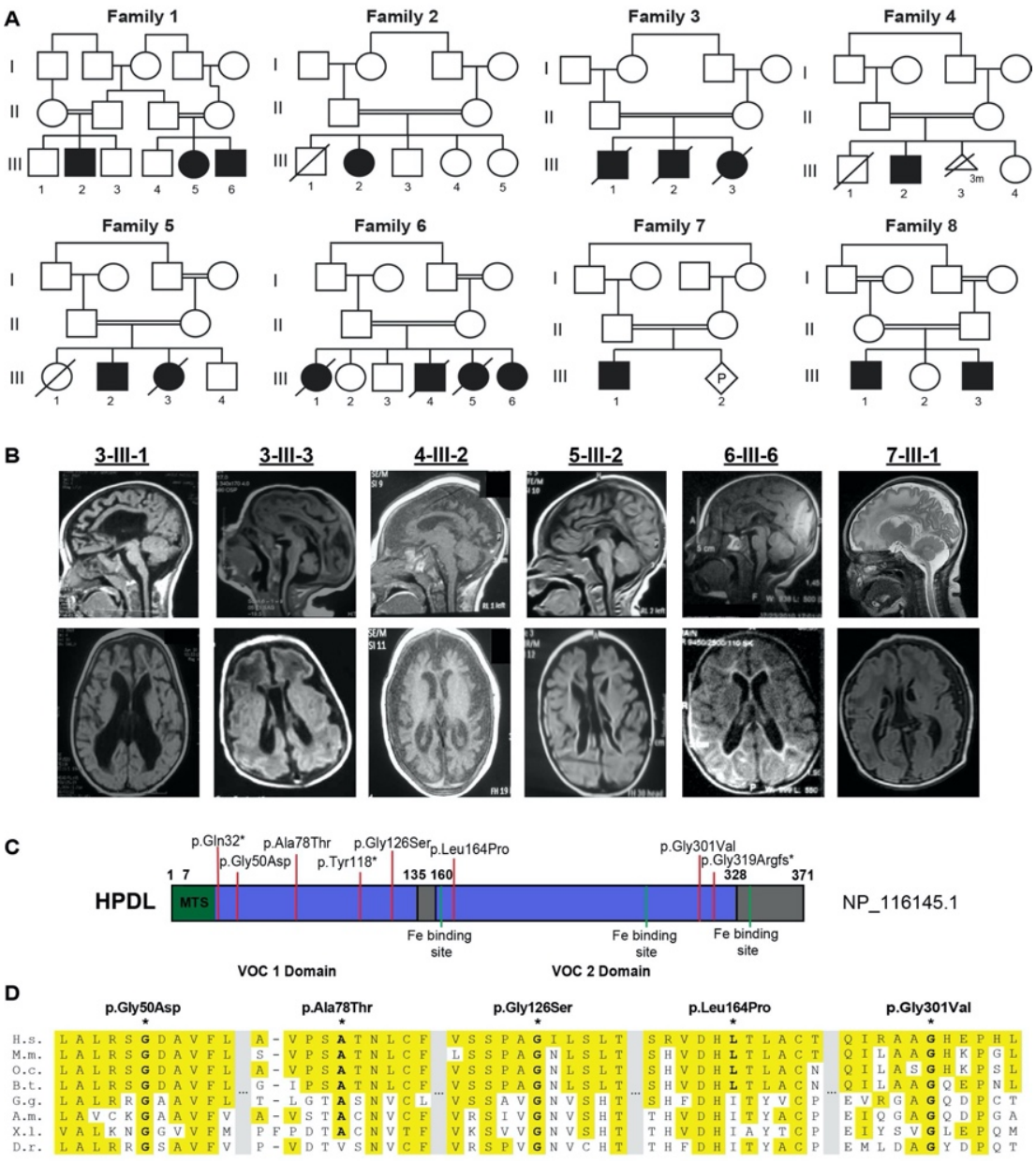

Figure 6.1: Variants in *HPDL* in eight independent consanguineous families lead to microcephaly and brain atrophy.

(A) Pedigrees of families 1-8 showing consanguineous marriages (double line) with a total of 17 affected children. All unfilled members are without neurological disease, epilepsy, or neurodegeneration. Dashed line: deceased. (B) Panels show MRI scans for six affected individuals from five different families. Shown are sagittal and axial images demonstrating enlarged ventricles, thin corpus callosum, and severe cortical and cerebellar atrophy. (C) Schematic of the HPDL protein structure depicting a predicted Mitochondrial Targeting Sequence (MTS, green), two vicinal oxygen chelate (VOC) domains from amino acids 7-135 and 160-328 (shown in blue) and three iron (Fe) binding sites (shown in green). Red lines indicate the locations of the variants. (D) Amino acid alignments for all identified missense variants across different vertebrate species. Amino acids highlighted in yellow indicate conserved residues. H.s.: *Homo sapiens;* M.m: *Mus musculus*; O.c.: *Oryctolagus cuniculus*; B.t.: *Bos taurus*; G.g.: *Gallus gallus*; A.m.: *Alligator mississippiensis*; X.l.: *Xenopus laevis*; D.r.: *Danio rerio*.

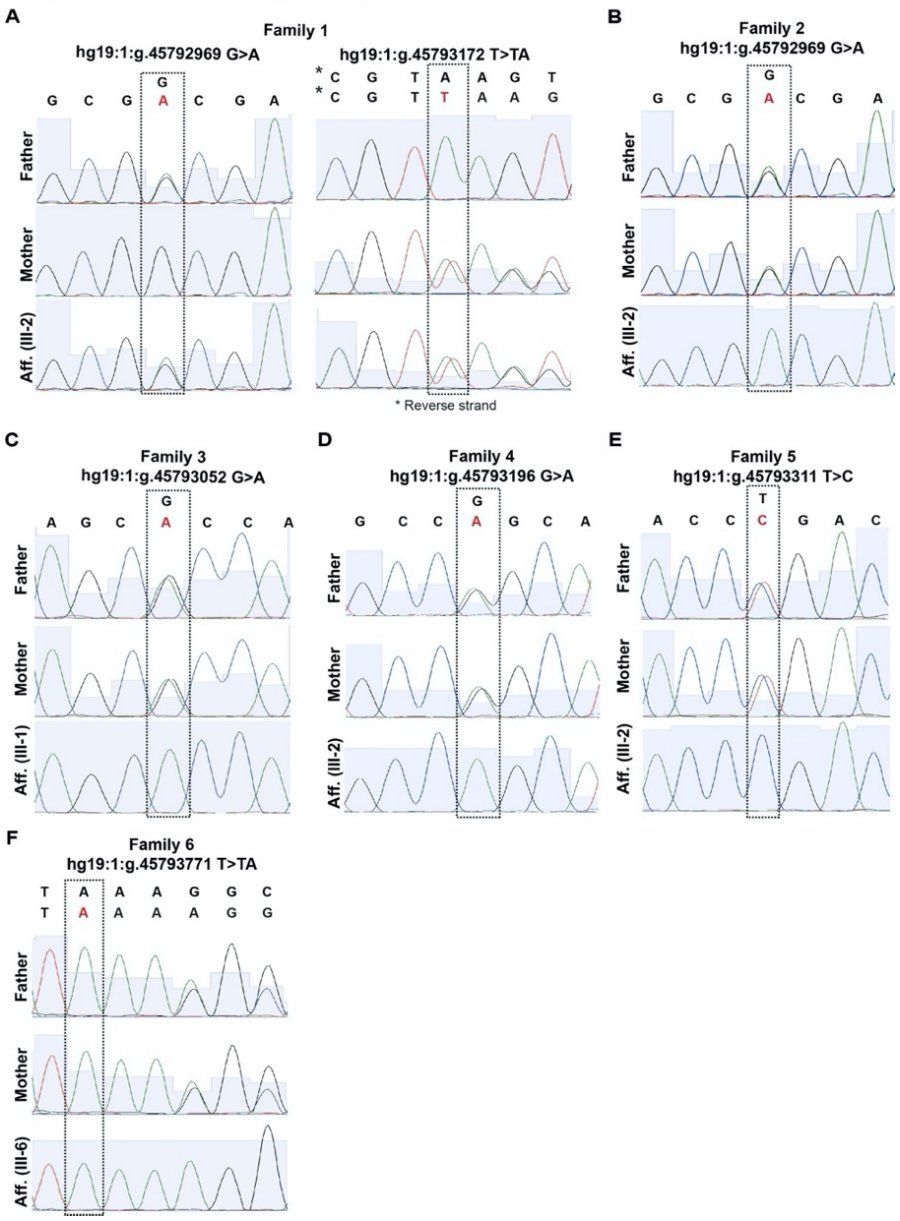

Figure 6.2: Sanger sequencing traces for families 1-6.

(A) Sanger sequencing traces in Father, Mother, and Affected (Aff., III-2) in Family 1 showing heterozygosity in the Affected for both the G>A nucleotide change and the T>TA insertion. Note the G>A nucleotide change was heterozygous in the father, but not the mother, while the T>TA insertion was heterozygous in the mother, but not the father. Both alleles are shown to highlight the additional nucleotide insertion in the affected individual. *: reverse strands are shown. All other traces are forward strand sequences. (B) Sanger sequencing traces for Father, Mother and Affected (Aff., III-2) in Family 2 depicting the G>A nucleotide change. (C) Sanger sequencing traces for Father, Mother, and Affected (Aff., III-1) in Family 3 depicting the G>A nucleotide change. (D) Sanger sequencing traces for Father, Mother, and Affected (Aff., III-2) in Family 4 depicting the G>A nucleotide change. (E) Sanger sequencing traces for Father, Mother, and Affected (Aff., III-2) in Family 5 depicting the T>C nucleotide change. (F) Sanger sequencing traces from Father, Mother, and Affected (Aff., III-6) in Family 6 depicting the T>TA insertion. Both alleles are shown to highlight shift in the Affected caused by the nucleotide insertion.

6.4.2 HPD and HPDL Have Differing Tissue-wide Expression and Subcellular Localization

To functionally distinguish between *HPDL* and the paralogous *HPD*, we sought differences in tissue expression and subcellular localization patterns. The GTEx database indicates that *HPD* is almost exclusively expressed in the liver and kidney, whereas *HPDL* is widely expressed in most organs (Figure 6.3).[28] Importantly, the expression level of *HPDL* is much higher than *HPD,* specifically in the brain (Figure 6.4).[28] Publicly available human and mouse brain single-cell RNA-seq data suggests that *HPDL* transcripts are encountered ~5x more commonly than *HPD* in the brain, with brain astrocytes as the major source of *HPDL* transcripts.[29-31]

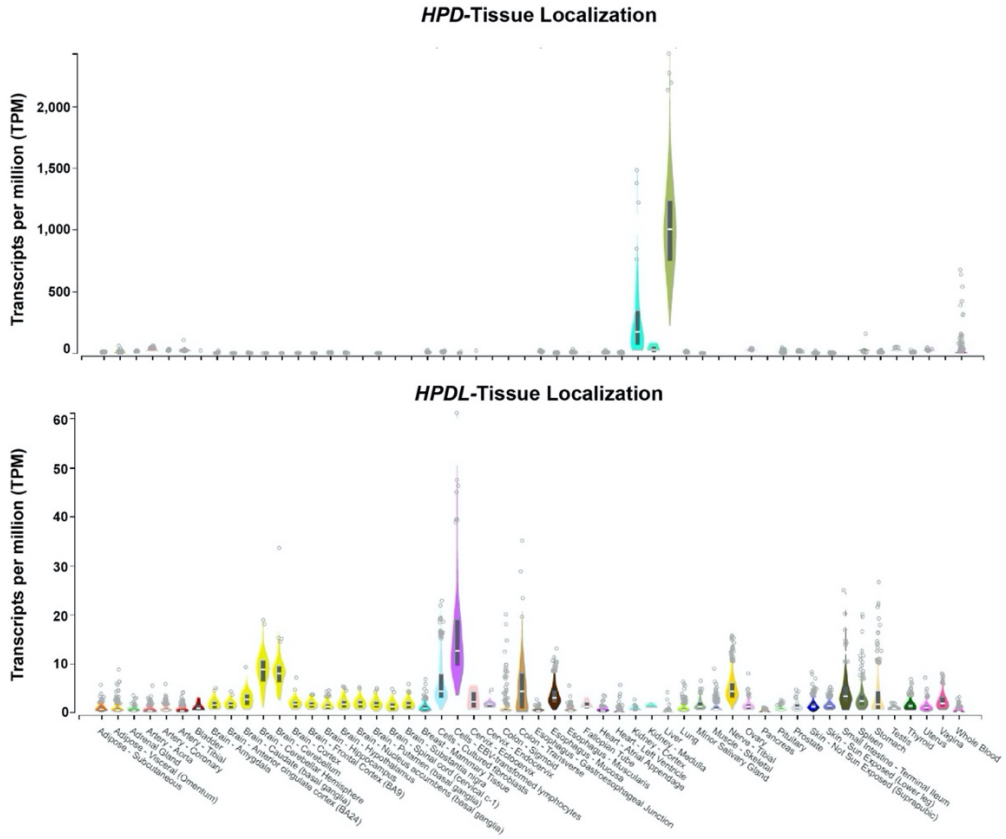

Figure 6.3: *HPD* and *HPDL* expression across various tissues in adult humans (GTEx data). Expression analysis of *HPD and HPDL* transcripts across various adult tissues. *HPDL* shows ubiquitous expression across all tissues, while HPD expression is restricted to the kidney and liver. Expression is shown as Transcripts per Kilobase Million (TPM). Data used was derived from the GTEx portal.[3] Shown are standard violin plots with inner boxplots.

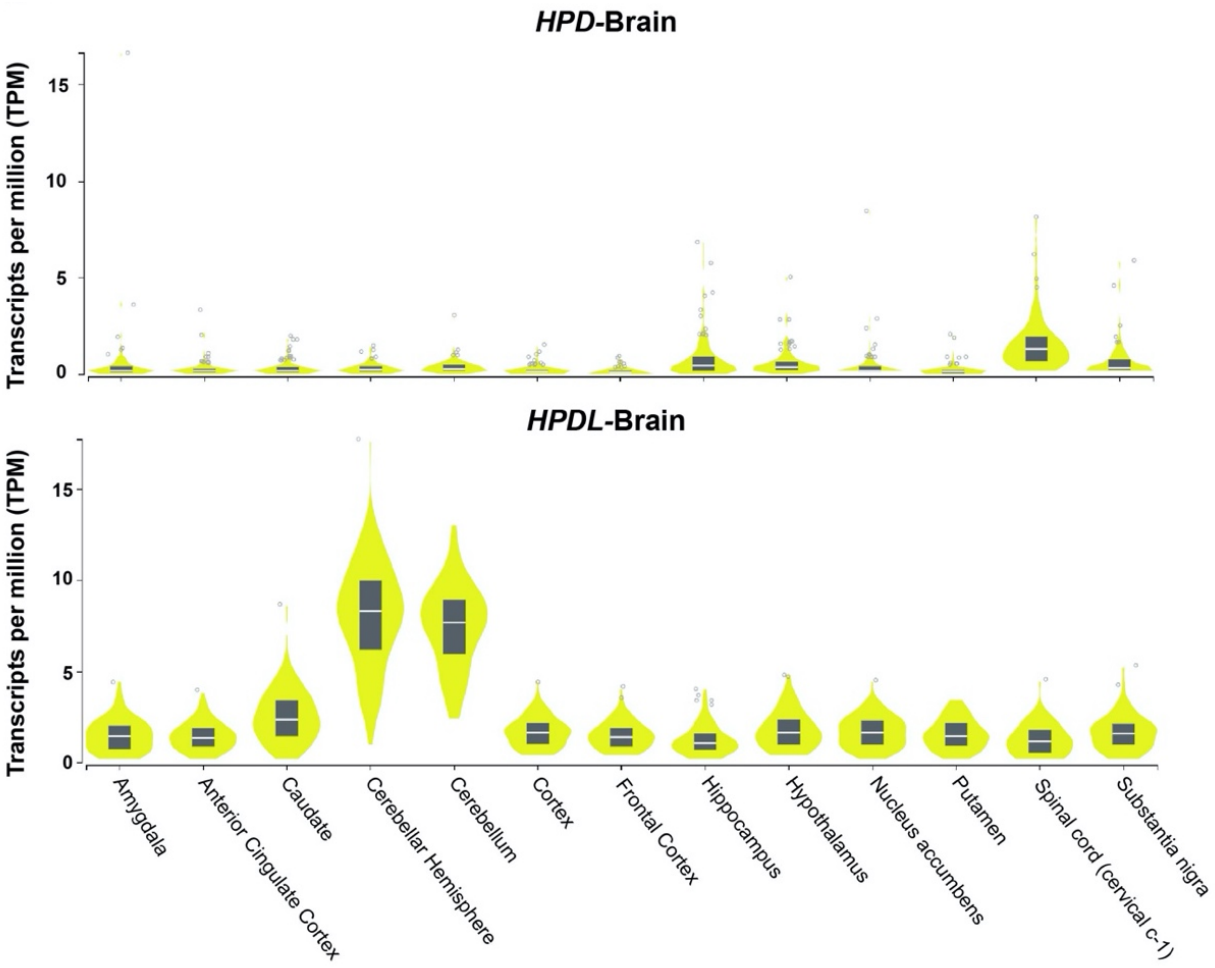

Figure 6.4: *HPD* and *HPDL* expression across various brain regions in adult humans (GTEx data). Expression analysis of *HPD and HPDL* transcripts across various adult brain regions. Note that *HPDL* expression is higher than *HPD* in all brain regions. The expression is shown as Transcripts per Kilobase Million (TPM). Data used was derived from the GTEx portal.[28] Shown are standard violin plots with inner boxplots.

Protein databases suggest HPD localizes to the cytosol while HPDL localizes to mitochondria.[32] We used Mitofates, an online mitochondrial targeting sequence (MTS) prediction tool, to assess for an N-terminal mitochondrial presequence or cleavable localization signal.[33] While HPD had no predicted MTS, HPDL contained three different recognition motifs for TOMM20, a mitochondrial import receptor subunit responsible for the recognition and translocation of cytosolically synthesized mitochondrial preproteins. In addition, there was a region of maximum positively charged amphiphilicity, another essential feature for

mitochondrial presequence function (Figure 6.5A).[34] To confirm this experimentally, we utilized CACO-2 cells, a human colon adenocarcinoma cell line expressing levels of *HPDL* higher than most other cell lines.[35] To validate the specificity of the anti-HPDL antibody, we generated *HPDL* knockout (KO) CACO-2 cells using CRISPR/Cas9, and confirmed correct biallelic targeting.

Immunofluorescence in WT cells confirmed the co-localization of TOMM20 and HPDL within mitochondria (Figure 6.5B), while staining in the KO cells confirmed absent HPDL signal (Figure 6.5C). HPD showed no co-localization with TOMM20, but instead was localized diffusely to the cytoplasm (Figure 6.5D). To assess localization of tagged versions of the proteins, we co-transfected HEK293T cells with C-terminally Myc-tagged HPDL and C-terminally EGFP-tagged HPD and co-stained with anti-TOMM20. We used C-terminally tagged vectors to ensure that any N-terminal MTS would not be disturbed. Again, TOMM20 overlapped in localization with HPDL, but not HPD. We also flipped the vectors and repeated the experiment (C-terminally Myc-tagged HPD and C-terminally EGFP-tagged HPDL) to ensure that that the tag did not affect localization, which provided further confirmation (Figure 6.5E). We utilized a Seahorse assay to compare the oxygen consumption rate in these cells, which illustrated slightly lower OCR activity in the HPDL KO cells compared to WT (Figure 6.6). Together, these results demonstrate that HPDL, but not HPD, localizes to mitochondria, suggesting that these enzymes act in different locations on either the same or on different substrates. The slightly reduced OCR activity in the KO cells further suggests that the absence of HPDL may have an effect on mitochondrial metabolic function; however, this difference was not significant and further characterization is required.

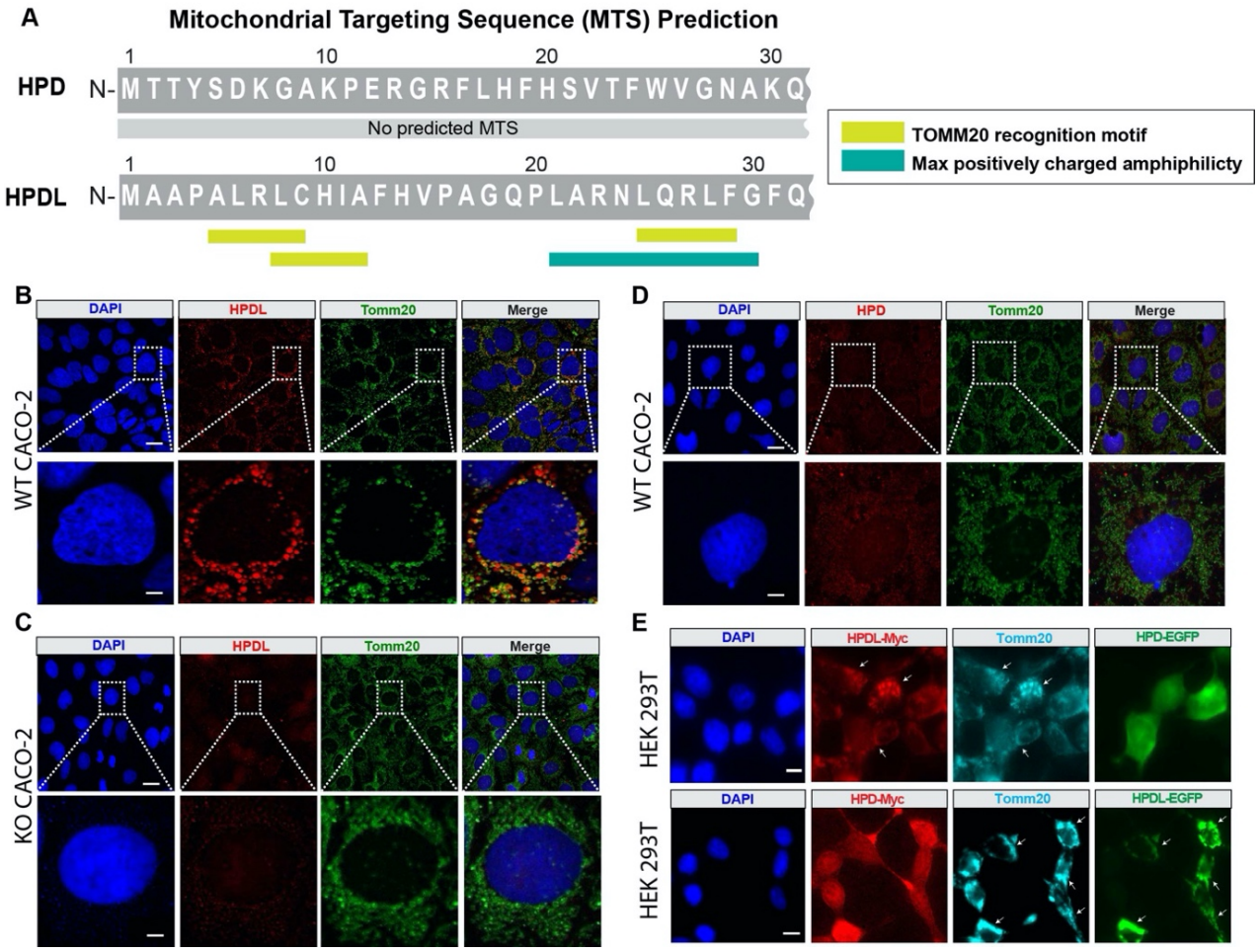

Figure 6.5: HPD and HPDL display different subcellular localization patterns.
(A) Schematic of HPD and HPDL N-terminal amino acid sequences highlighting the presence or absence of a predicted Mitochondrial Targeting Sequence (MTS) as predicted by Mitofates. HPD has no predicted MTS, while HPDL has three different TOMM20 recognition motifs (yellow) and a region of max positively charged amphiphilicity (green). (B) Immunofluorescent staining of HPDL WT CACO-2 cells with DAPI (blue), HPDL (red), and TOMM20 (green). Scale bars: 10 µm (upper panels) and 2µm (lower panels). (C) Immunofluorescent staining of HPDL KO CACO-2 cells with DAPI (blue), HPDL (red), and TOMM20 (green). Scale bars: 10 µm (upper panels) and 2µm (lower panels). (D) Immunofluorescent staining of HPDL WT CACO-2 cells with DAPI (blue), HPD (red), and TOMM20 (green). Scale bars: 10 µm (upper panels) and 2µm (lower panels). (E) Immunostaining for TOMM20 (cyan) and DAPI (blue) of HEK293T cells transfected with both HPDL-Myc (red) and HPD-EGFP (green) or HPD-Myc (red) and HPDL-EGFP (green) to determine the subcellular localization of HPD and HPDL. White arrows: cells with mitochondrial localization of HPDL. Scale bar: 10 µm.

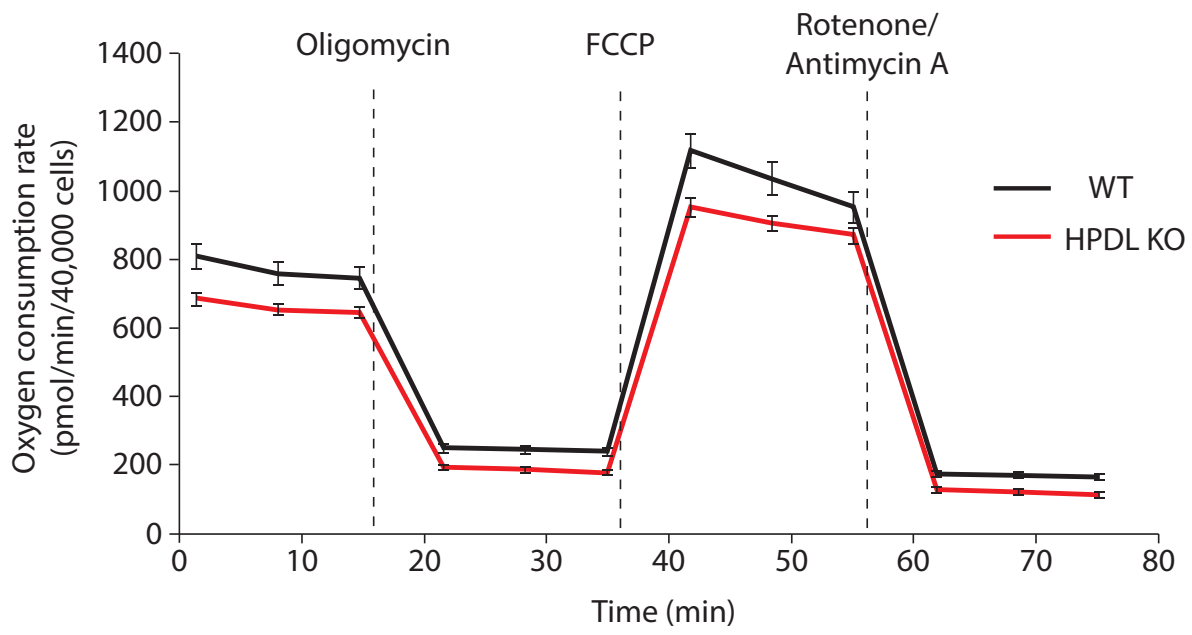

Figure 6.6: Seahorse assay of CACO-2 WT and *HPDL* KO cells.
Oxygen consumption rate of CACO-2 WT and HPDL KO cells after indicated treatments. Data are represented as ± standard error of the mean (n = 10; FCCP, carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone).

6.4.3 Hpdl KO Mice Display Epilepsy, Early Lethality, Smaller Brain Sizes, and Cellular Apoptosis in the Brain

To investigate the physiological role of HPDL, we created a KO mouse model with a homozygous frameshift deletion of *Hpdl* generated using CRISPR/Cas9 (Figure 6.7).

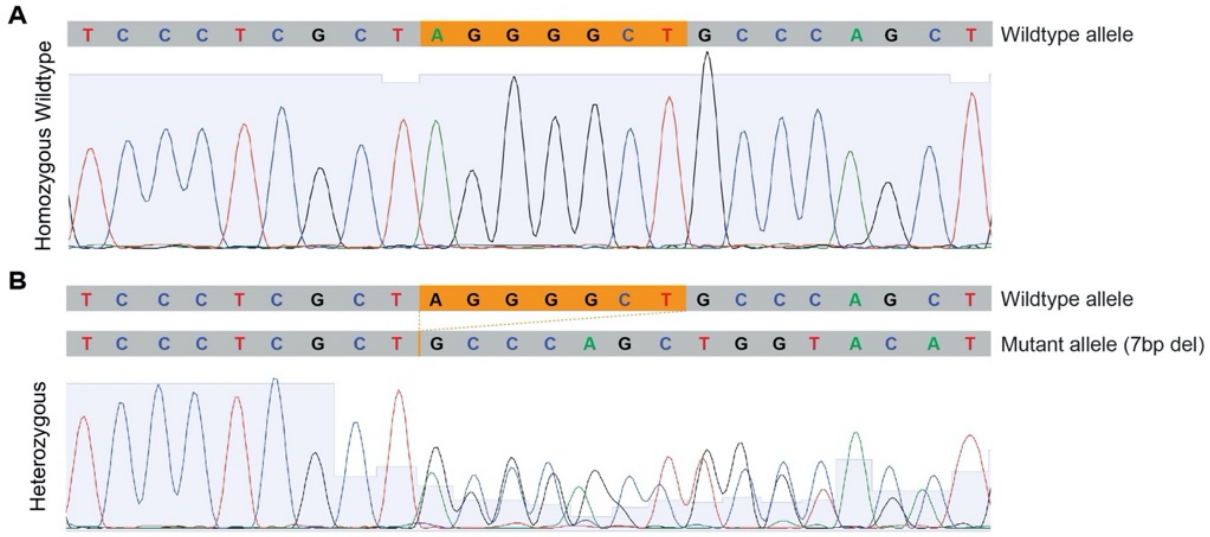

Figure 6.7: Sanger sequencing traces for WT and *Hpdl* heterozygous mice.
(A) Sanger sequencing trace from a WT mouse using the following primers: Hpdl_KO_F/R (Supplementary Table S3). (B) Sanger sequencing trace depicting a heterozygous 7bp deletion generated with the following sgRNA guide sequence: CUU CCA GCC CCU GGC GGU GA, ultimately giving rise to a frameshift variant in *Hpdl* (NM_146256.3:c.59_65del). Primers used for sequencing: Hpdl_KO_F/R (Supplementary Table S3).

KOs were born healthy and at Mendelian ratios (WT:heterozygote:homozygote = 30:68:26, test for Hardy-Weinberg equilibrium P=0.368). Although KOs were indistinguishable at birth, by postnatal day 5 (P5), KOs exhibit lethargy and spontaneous movements suggesting of epileptic seizures. Almost all KOs died by P10, and none survived to P15 (Figure 6.8A). Although body weights and brain size were similar at birth, by P9-10 KO mice weighed approximately half that of littermates (Figure 6.8B and C). Immunostaining of P9 brain tissue sections for DAPI and Cleaved Caspase 3 (CC3) demonstrated significant cellular apoptosis (CC3+ cells) in the KO brains compared to WT pups (23 vs. <1 CC3+ cell per 500µm column), which correlated with smaller brain size (Figure 6.8C-E). To determine the cellular identity of CC3+ cortical cells, we performed co-immunostaining with CTIP2 (upper layer cortical neurons), OLIG2 (oligodendrocytes), and GFAP (astrocytes). CTIP2/CC3 staining revealed the most co-

localization compared with OLIG2/CC3 and GFAP/CC3 (5.9% vs. 1.9% and 3.8%, respectively, Figure 6.8F-G). Taken together, this data suggests cellular apoptosis in the brain occurs in the absence of HPDL function.

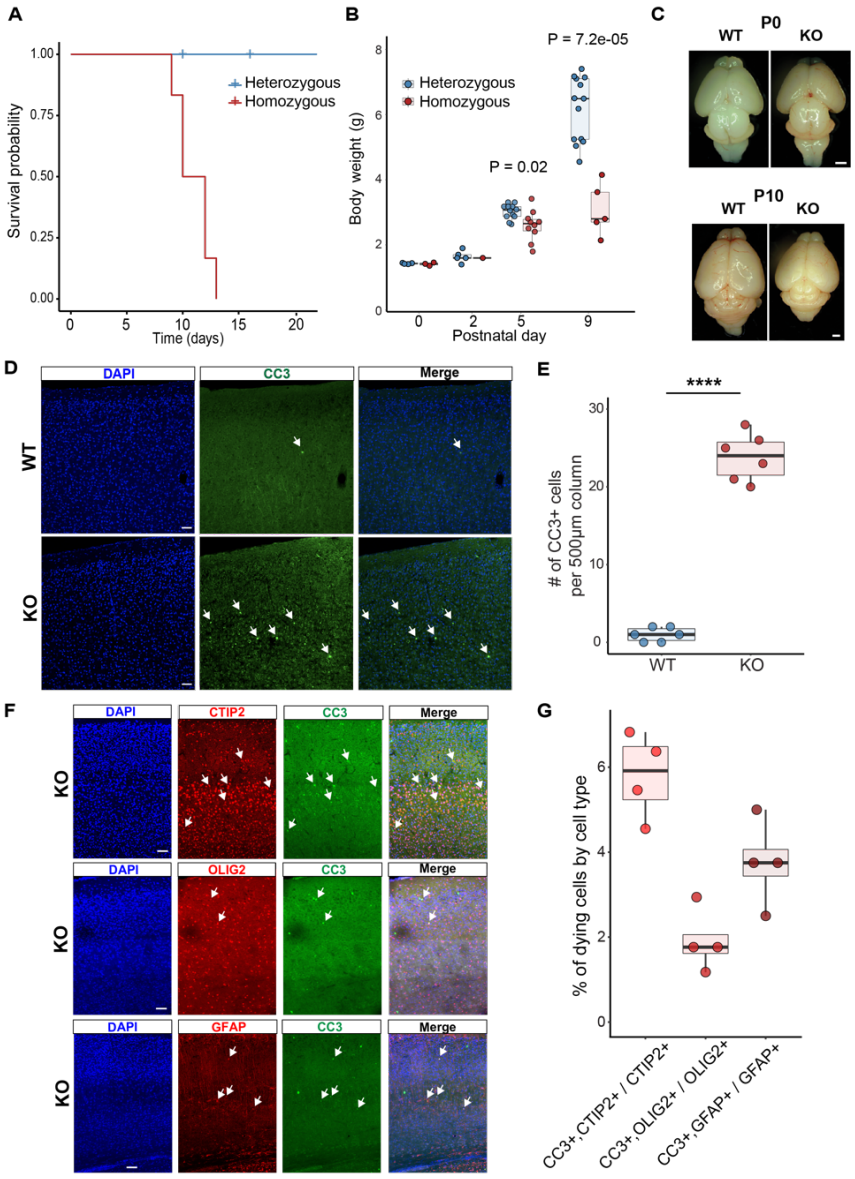

Figure 6.8: *Hpdl* KO mice display early lethality, smaller brain sizes, and cellular apoptosis. (A) Kaplan-Meyer curve displaying the survival probability of *Hpdl* heterozygous mice (n=10, blue) versus *Hpdl* KO mice (n=6, red). (B) Graph depicting body weight measurements of *Hpdl* heterozygous mice versus *Hpdl* KO mice at postnatal days 0, 2, 5, and 9. P-values were obtained by an independent t-test. (C) Whole-brain images of WT versus *Hpdl* KO mice at P0 and P10. Scale bar: 1 mm (D) Coronal sections of P9 WT and KO mouse brain were stained for DAPI (blue) and Cleaved Caspase 3 (CC3, green). The P9 KO shows strong CC3+ immunoreactivity. Note the absence of CC3 immunoreactivity in the P9 WT. Scale bar: 50µm (E) Quantification of CC3+ cells in *Hpdl* WT vs KO mouse brain per 500µm column. Statistical analysis (two-way ANOVA) was performed between the categories. N=3 independent experiments with two sections per mouse quantified (****P<0.0001). (F) Coronal sections of P9 KO mouse brain were stained for DAPI (blue) and CC3 (green) in combination with either CTIP2 (red), OLIG2 (red), or GFAP (red). White arrows depict co-localization of the cell identity markers with CC3. Scale bar: 50 µm. (G) Quantification of the number of CTIP2+, OLIG2+, or GFAP+ cells which were also CC3+. Analysis shows CTIP2+ cells to be most affected, followed by GFAP+ cells, followed by OLIG2+ cells. N=2 independent experiments. Two sections per mouse were used for quantification.

6.4.4 Metabolomic Analysis Reveals Disrupted Metabolic Signature but Normal Tyrosine Catabolism

Deficiencies of specific enzymes of the tyrosine catabolism pathway are known to give rise to a number of severe metabolic disorders in humans. In the oxidative degradation of tyrosine by vertebrates, the first conversion step is to 4-hydroxyphenylpyruvate in a reaction catalyzed by tyrosine transaminase (Figure 6.9A).[36] HPD then catalyzes the oxidation of 4-hydroxyphenylpyruvate to homogentisate.[12] Defects in these two steps in tyrosine catabolism lead to tyrosinemia type II and type III, respectively. Normally, homogentisate is further catabolized by the next enzyme in the pathway, homogentisate 1,2-dioxygenase.[8] In alkaptonuria, however, the enzyme is deficient, leading to this rare hereditary metabolic disorder. In tyrosinemia type I, fumarylacetoacetate hydrolase is mutated, causing the subsequent buildup of byproducts.[37] Because both HPD and HPDL have highly similar amino acid homology, we speculated that HPD and HPDL may have similar substrates. To test this, we performed mass spectrometry analysis on *HPDL* WT and KO CACO-2 cell extracts and conditioned cell culture media. The analysis revealed several significantly downregulated and upregulated metabolites in both the cell extracts and culture media of KO cells compared to WT; however, we did not detect an abnormal accumulation of any metabolites involved in tyrosine catabolism (Figure 6.9B and C, Table 6.1). We then performed a similar analysis in both *Hpdl* WT and KO mouse brain extracts and serum at P5, which again showed no abnormal accumulation of tyrosine catabolites (Figure 6.9D and E, Table 6.1). Taken together, this data suggests that HPDL likely acts on a different substrate than HPD, and perhaps on a different cellular pathway than tyrosine catabolism.

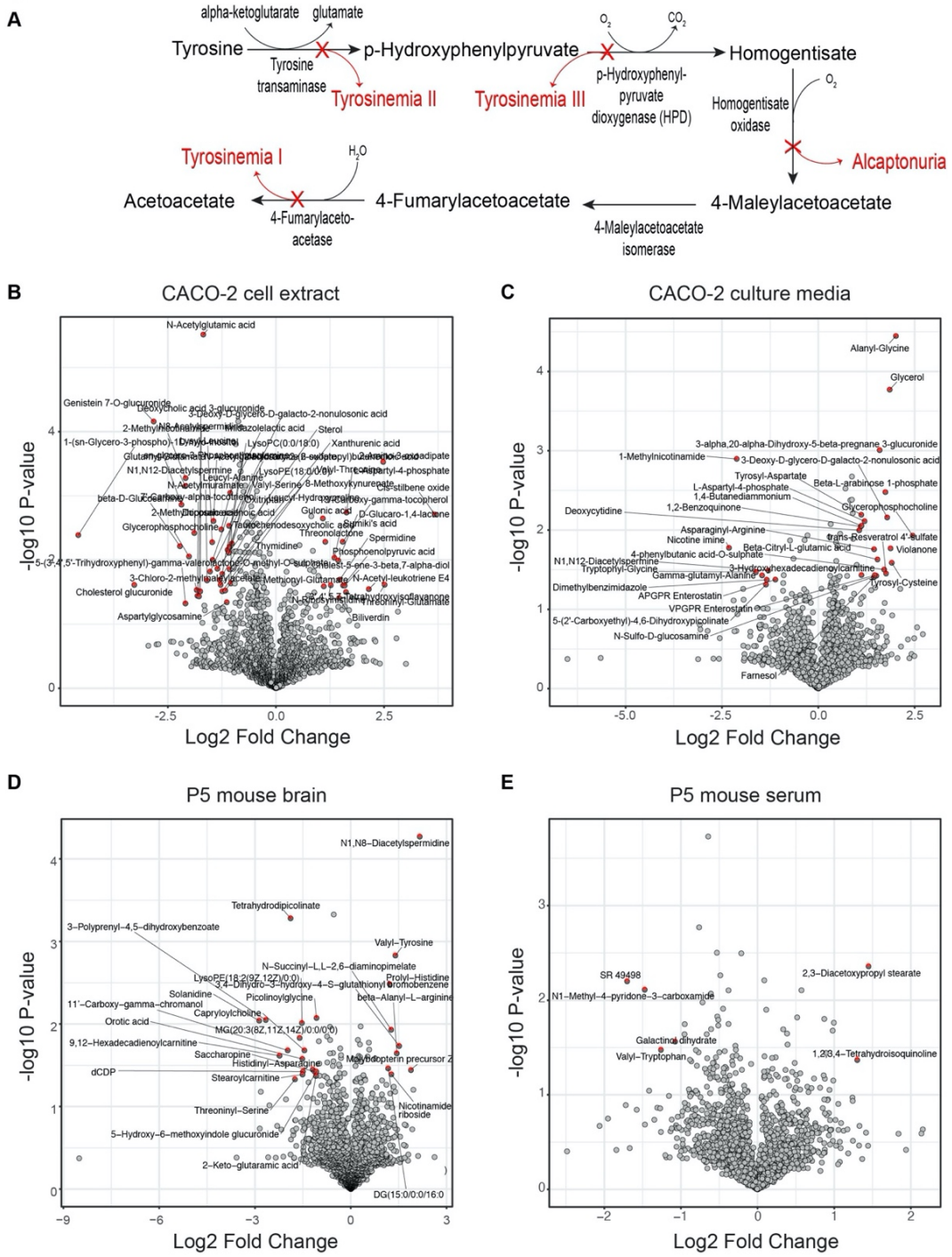

Figure 6.9: *In vitro* and *in vivo* metabolomic analysis reveals significant metabolic differences, yet no differences in tyrosine catabolism markers.

(A) Schematic of tyrosine catabolism pathway. Diseases caused by deficiencies at each step are shown in red. (B-C) MS analysis from *HPDL* KO CACO-2 shows significant metabolic disruption, but no abnormal accumulation of tyrosine metabolites. (D-E) MS analysis from P5 WT and KO mouse brain extracts and blood serum shows significant metabolic differences, but no abnormal accumulation of tyrosine metabolites. Red circles indicate >2-fold change in KO compared to WT and statistical significance ($P<0.05$).

Table 6.1: Mean 'counts' of tyrosine metabolites in *HPDL* WT vs KO CACO-2 cells and *Hpdl* WT vs KO mice.

Tyrosine metabolites	CACO-2				P5 mouse				Remarks
	Cell extract		Conditioned media		Brain		Plasma		
	WT	KO	WT	KO	WT	KO	WT	KO	
FAA/MAA	15567.4	20433.4	19695.0	14308.4	6403.5	6097.1	9514.5	6521.5	Indistinguishable due to same M.W.
Acetoacetic acid	407491.4	494147.4	1977801.0	1619625.0	495181.7	511316.1	1264628.8	1809750.0	
cis/trans-4-Hydroxycyclohexylacetic acid	607928.2	580355.2	530035.9	602645.4	988736.2	1110786.0	1387685.7	1425584.1	Indistinguishable due to same M.W.
Hydroxyphenyllactic acid	190195.7	158948.5	383484.9	274351.1	146728.6	172931.5	260558.4	255488.6	
4-HPPA/ $C_9H_8O_4$	75601.0	54318.5	110473.7	104272.5	198197.8	409628.2	77549.5	61367.7	Indistinguishable due to same M.W.
L-Tyrosine	6844628.5	4884269.3	10956105.8	12427637.2	2846352.0	2880003.3	6094334.1	4219806.8	
4-Hydroxy phenylacetylglutamine	15896.7	10874.1	13008.5	19637.9	75640.7	56615.5	25710.9	25689.5	

6.4.5 Independent HPD Duplication Events Gave Rise to HPDL in Vertebrates

HPD, conserved in all eukaryotes, is encoded on human chromosome 12, with its major isoform consisting of 14 coding exons (Figure 6.10A). One alternative splice isoform is described, however, the major isoform encodes a protein 393 amino acids long. *HPDL*, on the other hand, is a single-exon gene located on chromosome 1, with no alternative splice isoforms reported, encoding 371 amino acids, apparent in *Xenopus tropicalis* (frog) but not in lower species like *Danio rerio* (zebrafish) (Figure 6.10B). There are no other *HPD* paralogues in mammals, suggesting that *HPDL* arose as a result of a retrotransposition event at the base of the tetrapod lineage (Figure 6.10C). However, there is an *HPDL* paralogue in other lower deuterostomes, including *Ciona intestinalis* (sea vase) and zebrafish, but this *HPDL* contains introns. Thus, other models are possible by which *HPDL* arose from *HPD* during evolution.

Several lines of evidence suggest that mammalian *HPDL* arose as a separate gene duplication event from the event that yielded *HPDL* in lower deuterostomes. First, only mammalian HPDL contains a predicted MTS. Second, mammalian *HPDL* shows evidence of stronger positive selective pressure than fish *hpdl*, as evidenced by faster accumulation of nonsynonymous to synonymous base changes (Figure 6.10D and 6.11A). Third, mammalian *HPDL* is in a region of conserved synteny in frog but not in zebrafish (Figure 6.11B). Lastly, and most convincingly, it is more likely that a retrotransposition event yielded an intronless gene than removal of introns from an intron-containing gene. Human HPDL is more similar to zebrafish hpdl than to human HPD (40% vs 25% amino acid sequence identity). The data suggest that lower deuterostome *HPDL* was lost contemporaneously with its retrotransposition to intronless *HPDL*, which subsequently acquired an MTS and amino acid changes that provided a new essential function.

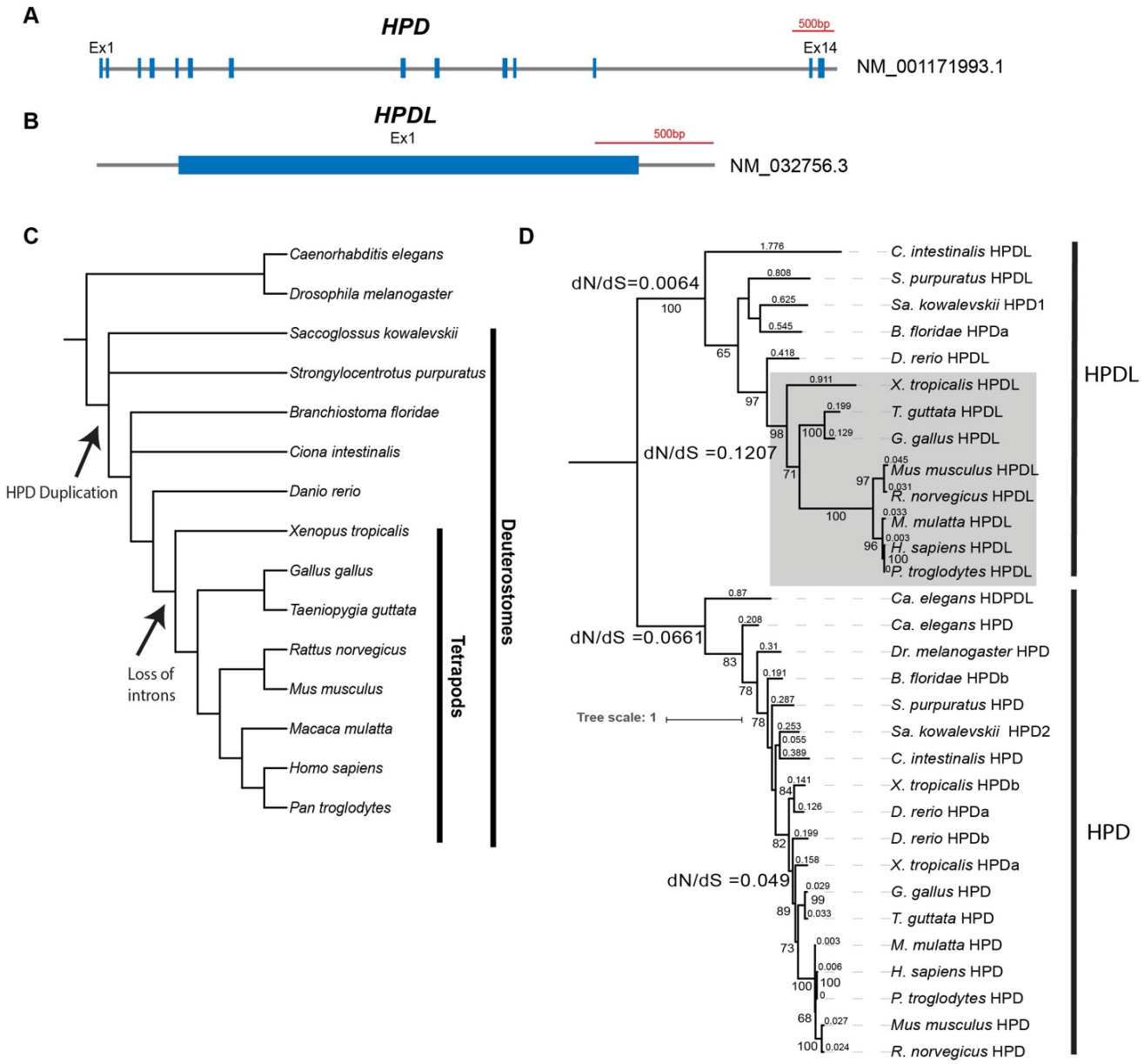

Figure 6.10: HPD and HPDL evolutionary analysis reveals two separate, yet significant, evolutionary events leading to the emergence of vertebrate HPDL with novel function.
(A) Schematic of the major human *HPD* isoform depicting the coding sequence spanning 14 exons using the UCSC Genome Browser as a reference. (B) Schematic of human *HPDL* depicting the coding sequence spanning a single exon and the 5' and 3' UTRs using the UCSC Genome Browser as a reference. Scale bars are shown in red. (C) Species tree of genomes used in this analysis. Arrows indicate *HPD* gene duplication and a possible retrotransposition event leading to the loss of exons in *HPDL*. (D) Gene tree for the *HPD* and *HPDL* genes constructed with RAxML with the JTT model. Branch lengths are on top of each branch, bootstrap support of 65 and above are shown below branches. Gray box indicates a loss of *HPDL* exons in tetrapods. dN/dS ratios are estimated using codeml and the model showing four different dN/dS ratios is shown (Figure 6.11).

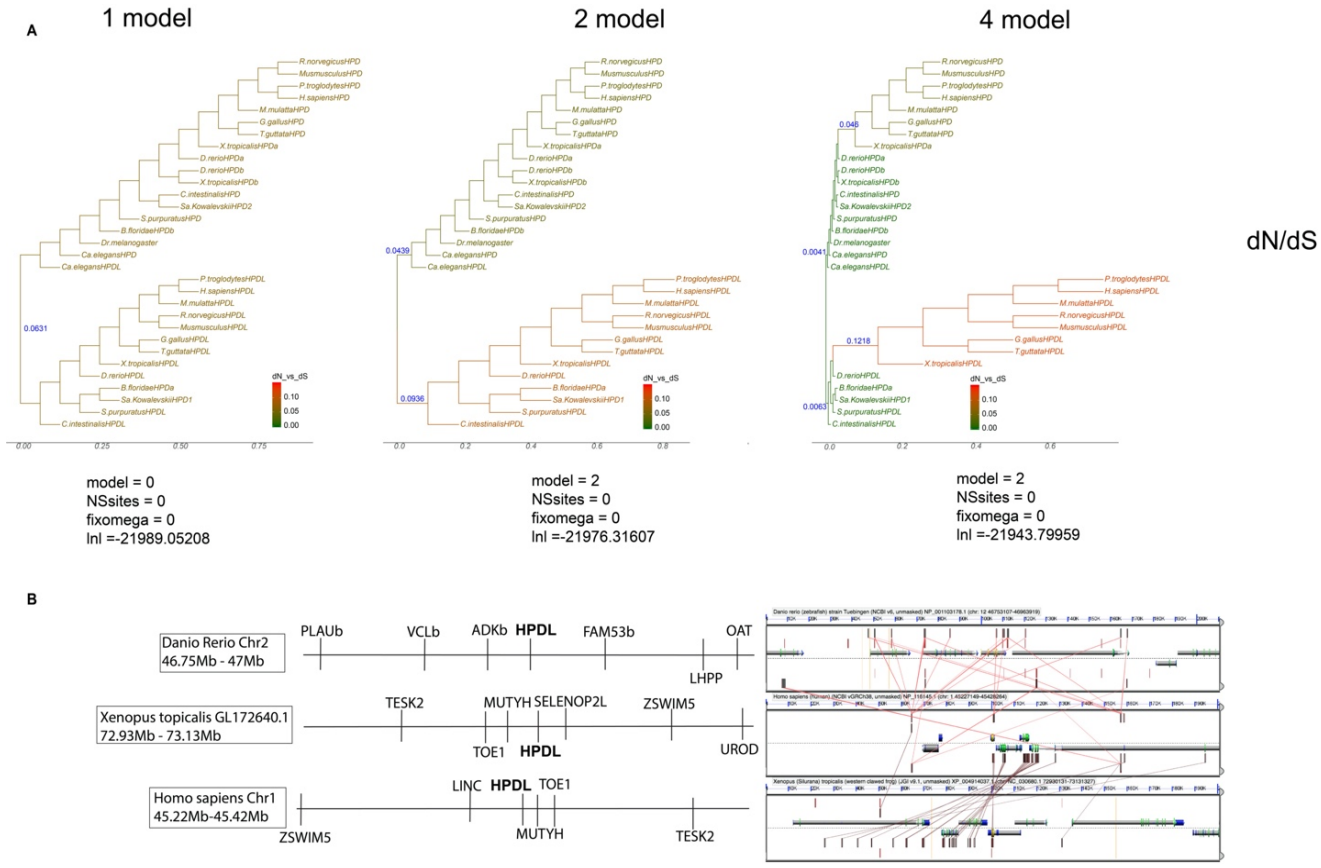

Figure 6.11: dN/dS ratios for three models tested using codeml and HPDL synteny analysis. (A) The 1 model indicates one omega value for the full tree, speculating that all *HPD* and *HPDL* genes have one rate of evolution. The 2 model indicates one omega value for *HPD* and one omega value for *HPDL*, suggesting two different rates of evolution: one for *HPD* and another for the resulting gene following the gene duplication event of *HPD*. The 4 model indicates one omega value for *HPD*, one for *HPD* in tetrapods, one for *HPDL*, and one for *HPDL* in tetrapods. These different omega values suggests that only genes resulting after the loss of *HPD* introns have their own rate of evolution, and all other *HPD* and *HPDL* genes evolved under a different evolutionary rate. Parameters for the test are indicated below each phylogeny and the lnl value for the model. (B) HPDL Synteny (right) and gene order (left) for *Danio rerio*, *Xenopus tropicalis*, and *Homo sapiens*.

6.5 Discussion

Here, we implicate *HPDL* in a recessive infantile-onset neurodegenerative disease, characterized by intellectual and motor disability, epilepsy, spastic tetraplegia, and brain atrophy. Radiographic phenotypes include severe cortical and cerebellar atrophy, thinning of corpus callosum, and ventriculomegaly. There are likely a range of developmental brain phenotypes resulting from loss of *HPDL*, although our subjects are likely to be at the most severe end of the

spectrum given the nature of the alleles. HPDL is not the first mitochondrial dioxygenase to be implicated in a neurometabolic disorder. ETHE1, another mitochondrial dioxygenase, is linked to ethylmalonic encephalopathy, an inherited disorder defined by elevated excretion of ethylmalonic acid, leading to developmental delay, seizures and hypertonia.[38] Future studies will be required to elucidate the full range of phenotypes, genotype-phenotype correlations, and mechanisms of pathogenicity.

We demonstrate that HPD and HPDL not only have different tissue-wide expression patterns, but also show different subcellular localizations. In addition to showing much higher expression in the brain, *HPDL* also displays notably high expression in lymphocytes, suggesting a potential role of this dioxygenase in activating or inhibiting immune responses. It is possible that HPDL's activity in cells of the immune system can critically alter their responses to immunological insults, as shown with other dioxygenase pathways.[39] Localization of HPDL to mitochondria, together with the disrupted metabolomes in *HPDL* knockout cells and mice, implicates this disease not only as mitochondrial neurodegenerative condition, but also as a likely neurometabolic condition. In a direct comparison between *Hpdl* WT and KO mice, we highlight the pathophysiological importance of the HPDL enzyme for neurodevelopment and survival. Interestingly, the fact that the mice are born healthy and begin regressing soon after birth, suggests that the mice lacking *Hpdl* may depend on an environmental or some other factor provided by their mother to stay well. Further, although astrocytes have been shown to express HPDL at the highest levels in the brain at the cellular level,[29-31] in *HPDL* KO brain we found more neurons than glia with evidence of apoptosis, suggesting that neurons are ultimately more sensitive to *HPDL* loss.

There are several possible reasons why we were unable to identify a specific substrate for HPDL despite comprehensive metabolic profiling. First, the metabolite may be unidentifiable by the current approach, where we focused on small molecules that have been pre-annotated in public databases. Second, the amount of the potential substrate may have been too small to detect due to large sample volume. Therefore, future work should aim towards more targeted approaches (i.e., analyzing specific brain regions rather than the whole brain, or analyzing mitochondrial-specific extracts rather than whole-cell extracts). Lastly, it is still possible that HPDL is not an enzyme, since we did not confirm its enzymatic activity. However, given its close sequence similarity to the enzyme HPD, presence of iron-binding motifs, disease caused by biallelic null variants, and metabolic dysregulation seen in the cell and mouse models, this hypothesis is less likely.

This finding was further supported by our evolutionary analysis which supported that HPDL and HPD are likely functionally different. We have shown that the *HPDL* found in humans, likely emerged through a series of *HPD* gene duplication events, the last of which probably involved a retrotransposition event at the transition from deuterostomes to tetrapods. This suggests that at some point in time, *HPDL* existed as a multi-exonic gene before losing its introns and gaining an MTS.

Taken together, we describe a novel, early-onset neurodegenerative disorder caused by variants in *HPDL*, and also illustrate the importance of this protein for proper neurodevelopment and survival. Through evolutionary modeling and analysis, we illustrate the route by which human *HPDL* emerged as a paralogue of *HPD*, with proposed novel biochemical functions. Although further work is needed to identify the specific function of mitochondrial HPDL, we

believe our work is a significant cornerstone to elucidate the downstream mechanisms of HPDL, and to potentially lead to a treatment for this disease.

Chapter 6, in full, has been submitted for publication of the material as it may appear in Genetics in Medicine 2020. Ghosh, Shereen G.; Lee, Sangmoon; Fabunan, Rudy; Chai, Guoliang; Zaki, Maha S.; Abdel-Salam, Ghada; Sultan, Tipu; Ben-Omran, Tawfeg; Alvi, Javeria R.; McEvoy-Venneri, Jennifer; Stanley, Valentina; Patel, Aakash; Ross, Danica; Ding, Jeffrey; Jain, Mohit; Pan, Daqiang; Lubbert, Philipp; Kammerer, Bernd; Wiedemann, Nils; Verhoeven-Duif, Nanda M.; Jans, Judith J.; Murphy, David; Toosi, Mehran B.; Ashrafzadeh, Farah; Imannezhad, Shima; Karimiani, Ehsan, G.; Ibrahim, Khalid; Waters, Elizabeth R.; Maroofian, Reza; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this material.

6.6 References

1. Anaiis Brassier, Chris Ottolenghi, Nathalie Boddaert, P. Sonigo, Tania Attie-Bitach, Anne-Elodie Millischer-Bellaiche, G. Baujat, V. Cormier-Daire, Vassili Valayannopoulos, Nathalie Seta, Monique Piraud, Bernadette Chadefaux-Vekemans, Christine Vianey-Saban, Roseline Froissart, and P. de Lonlay. Prenatal symptoms and diagnosis of inherited metabolic diseases. *Arch Pediatr,* 2012.

2. Maja Tarailo-Graovac, Casper Shyr, Colin J. Ross, Gabriella A. Horvath, Ramona Salvarinova, Xin C. Ye, Lin-Hua Zhang, Amit P. Bhavsar, Jessica J.Y. Lee, Britt I. Drogemoller, Mena Abdelsayed, and Majid Alfadhel. Exome Sequencing and the Management of Neurometabolic Disorders. *N Engl J Med,* 2016.

3.Ahmed Y. BoAli, Majid Alfadhel, and Brahim Tabarki. Neurometabolic disorders and congenital malformations of the central nervous system. *Neurosciences (Riyadh),* 2018.

4. Michèl A. Willemsen, Inga Harting, and Ron A Wevers. Neurometabolic disorders: Five new things. *Neurol Clin Pract,* 2016.

5. Mahdi M. Abu-Omar, Aristobulo Loaiza, and Nikos Hontzeas. Reaction mechanisms of mononuclear non-heme iron oxygenases. *Chem Rev,* 2005.

6. Alberto B. Burlina, Franco Zacchello, Carlo Dionisi-Visi, Enrico Bertini, Gaetano Sabetta, Michael J. Bennet, Daniel Hale, Eberhard Schmidt-Sommerfeld, and Piero Rinaldo. New clinical phenotype of branched-chain acyl-CoA oxidation defect. *Lancet,* 1991.

7. Sarah Boissel, Orit Reish, Karine Proulx, Hiroko Kawagoe-Takaki, Barbara Sedgwick, Giles S.H. Yeo, David Meyre, Christelle Golzio, Florence Molinari, Noman Kadhom, Heather C. Etchevers, Vladimir Saudek, I. Sadaf Farooqi, Philippe Froguel, Tomas Lindahl, Stephen O'Rahilly, Arnold Munnich, and Laurence Colleaux. Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. *Am J Hum Genet,* 2009.

8. Jose M. Fernández-Cañón, Begona Granadino, Daniel Beltran-Valero De Bernabe, Monica Renedo, Elena Fernandez-Ruiz, Miguel A. Penalva, and Santiago Rodriguez De Cordoba. The molecular basis of alkaptonuria. *Nat Genet,* 1996.

9. Patrick Ferreira, Inchul Shin, Iveta Sosova, Kednerlin Dornevil, Shailly Jain, Deborah Dewey, Fange Liu, and Aimin Liu. Hypertryptophanemia due to tryptophan 2,3-dioxygenase deficiency. *Mol Genet Metab,* 2017.

10. Richard K. Bruick and Steven Mcknight. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science,* 2001.

11. Graham R. Moran. 4-Hydroxyphenylpyruvate dioxygenase. *Arch Biochem Biophys,* 2005.

12. Ulla Ruetschi, Roberto Cerone, Celia Pérez-Cerda, Maria Christina Schiaffino, Sue Standing, Magdalene Ugarte and Elisabeth Holme. Mutations in the 4-hydroxyphenylpyruvate dioxygenase gene (HPD) in patients with tyrosinemia type III. *Hum Genet,* 2000.

13. Shuji Kubo, Kohji Kiwaki, Kistaka Awata, Hideki Katoh, Yumi Kanagae, Izumu Saito, Tetsuro Yamamoto, Jun-Ichi Miyazaki, Ichiro Matsuda and Fumio Endo. In vivo correction with recombinant adenovirus of 4-hydroxyphenylpyruvic acid dioxygenase deficiencies in strain III mice. *Hum Gene Ther,* 1997.

14. Edyta Szymanska, Malgorzata Sredzinska, Elzbieta Ciara, Dorota Piekutowska-Abramczuk, Rafal Ploski, Dariusz Rokicki, and Anna Tylki-Szymanska. Tyrosinemia type III in an asymptomatic girl. *Mol Genet Metab Rep,* 2015.

15. Kaede Tomoeda, Hisataka Awata, Toshinobu Matsuura, Ichiro Matsuda, Engelbert Ploechl, Tom Milovac, Avihu Boneh, Cynthia Scott, David M. Danks, and Fumio Endo. Mutations in the 4-hydroxyphenylpyruvic acid dioxygenase gene are responsible for tyrosinemia type III and hawkinsinuria. *Mol Genet Metab,* 2000.

16. Chike Bellarmine Item, Ivana Mihalek, Oliver Lichtarge, Anil Jalan, Julia Vodopiutz, Adolf Muhl, and Olaf A. Bodamer. Manifestation of hawkinsinuria in a patient compound heterozygous for hawkinsinuria and tyrosinemia III. *Mol Genet Metab,* 2007.

17. Fábio Barroso, Joana Correia, Anabela Bandeira, Carla Carmona, Laura Vilarinho, Manuela Almeida, Júlio César Rocha, and Esmeralda Martins. Tyrosinemia Type III: A Case Report of Siblings and Literature Review. *Rev Paul Pediatr,* 2020.

18. Tracy J. Dixon-Salazar, Jennifer L. Silhavy, Nitin Udpa, Jana Schroth, Stephanie Bielas, Ashleigh E. Schaffer, Jesus Olvera, Vineet Bafna, Maha S. Zaki, Ghada H. Abdel-Salam, Lobna A. Mansour, Laila Selim, Sawsan Abdel-Hadi, Naima Marzouki, Tawfeg Ben-Omran, Nouriya A. Al-Saana, F. Müjgan Sonmez, Figen Celep, Matloob Azam, Kiley J. Hill, Adrienne Collazo, Ali G. Fenstermaker, Gaia Novarino, Naiara Akizu, Kiran V. Garimella, Carrie Sougnez, Carsten Russ, Stacey B. Gabriel, and Joseph G. Gleeson. Exome sequencing can improve diagnosis and alter patient management. *Sci Transl Med*, 2012.

19. Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* 2012.

20. Hanneke A. Haijes, Marcel Willemsen, Maria van der Ham, Johan Gerrits, Mia L. Pras-Raves, Hubertus C. M. T. Prinsen, Peter M. van Hasselt, Monique G. M. de Sain-van der Velden, Nanda M. Verhoeven-Duif, and Judith J. M. Jans. Direct Infusion Based Metabolomics Identifies Metabolic Disease in Patients' Dried Blood Spots and Plasma. *Metabolites,* 2019.

21. Jimin Pei and Nick V. Grishin. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol Biol,* 2014.

22. Alexandros Stamatakis. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 2014.

23. Ivica Letunic and Peer Bork. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res,* 2016.

24. Ziheng Yang. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol,* 2007.

25. Federico Abascal, Rafael Zardoya, and Maximilian J. Telford. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res,* 2010.

26. D. Charif, Jean Thioulouse, JEan Lobry and Guy Perriere. Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics,* 2005.

27. Konrad J. Karczewski, Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Laura D Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A. Watts, Daniel Rhodes, Moriel Singer-Berk, Eleina M. England, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X. Chong, Kaitlin E. Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S. Ware, Christopher Vittal, Irina M. Armean, Louis Bergelson, Kristian Cibulskis, Kristen M. Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto, Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade, Michael E. Talkowski, Genome Aggregation Database Consortium, Benjamin M. Neale, Mark J. Daly, and Daniel G. MacArthur. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature,* 2020.

28. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science,* 2015.

29. Ye Zhang, Steven A. Sloan, Laura E. Clarke, Christine Caneda, Colton A. Plaza, Paul D. Blumenthal, Hannes Vogel, Gary K. Steinberg, Michael S.B. Edwards, Gordon Li, John A. Duncan 3rd, Samuel H. Cheshier, Lawrence M. Shuer, Edward F. Chang, Gerald A. Grant, Melanie G. Hayden Gephart, and Ben A. Barres. Purification and Characterization

of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron,* 2016.

30. Arpiar Saunders, Evan Z. Macosko, Alec Wysoker, Melissa Goldman, Fenna M. Krienen, Heather de Rivera, Elizabeth Bien, Matthew Baum, Laura Bortolin, Shuyu Wang, Aleksandrina Goeva, James Nemesh, Nolan Kamitaki, Sara Brumbaugh, David Kulp, and Steven A. McCarroll. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell,* 2018.

31. Michael Vanlandewijck, Liqun He, Maarja Andaloussi Mae, Johanna Andrae, Koji Ando, Francesca Del Gaudio, Khayrun Nahar, Thibaud Lebouvier, Barbara Lavina, Leonor Gouveia, Ying Sun, Elisabeth Raschperger, Markus Rasanen, Yvette Zarb, Naoki Mochizuki, Annika Keller, Urban Lendahl and Christer Betsholtz. A molecular atlas of cell types and zonation in the brain vasculature. *Nature,* 2018.

32. Peter J. Thul, Lovisa Åkesson, Mikaela Wiking, Diana Mahdessian, Aikaterini Geladaki, Hammou Ait Blal, Tove Alm, Anna Asplund, Lars Björk, Lisa M. Breckels, Anna Bäckström, Frida Danielsson, Linn Fagerberg, Jenny Fall, Laurent Gatto, Christian Gnann, Sophia Hober, Martin Hjelmare, Fredric Johansson, Sunjae Lee, Cecilia Lindskog, Jan Mulder, Claire M. Mulvey, Peter Nilsson, Per Oksvold, Johan Rockberg, Rutger Schutten, Jochen M. Schwenk, Åsa Sivertsson, Evelina Sjöstedt, Marie Skogs, Charlotte Stadler, Devin P. Sullivan, Hanna Tegel, Casper Winsnes, Cheng Zhang, Martin Zwahlen, Adil Mardinoglu, Fredrik Pontén, Kalle von Feilitzen, Kathryn S. Lilley, Mathias Uhlén, and Emma Lundberg. A subcellular map of the human proteome. *Science,* 2017.

33. Yoshinori Fukasawa, Junko Tsuji, Szu-Chin Fu, Kentaro Tomii, Paul Horton, and Kenichiro Imai. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol Cell Proteomics,* 2015.

34. David Roise, Franziska Theiler, Suzanna J. Horvath, John M. Tomich, John H. Richards, Daniel S. Allison, and Gottfried Schatz. Amphiphilicity is essential for mitochondrial presequence function. *Embo j,* 1988.

35. Mathias Uhlén, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szigyarto, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per-Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M. Schwenk, Marica Hamsten, Kalle von

Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar von Heijne, Jens Nielsen, and Fredrik Pontén. Proteomics. Tissue-based map of the human proteome. *Science,* 2015.

36. E. Natt, Kazuhiro Kida, Michael Odievre, Maja Di Rocco, and Gerd Scherer. Point mutations in the tyrosine aminotransferase gene in tyrosinemia type II. *Proc Natl Acad Sci USA,* 1992.

37. D. Phaneuf, M. Lambert, Rachel Laframboise, G. Mitchell, Francine Lettre, and Robert M. Tanguay. Type 1 hereditary tyrosinemia. Evidence for molecular heterogeneity and identification of a causal mutation in a French Canadian patient. *J Clin Invest,* 1992.

38. Valeria Tiranti, Carlo Viscomi, Tatjana Hildebrandt, Ivano Di Meo, Rossana Mineri, Cecilia Tiveron, Michael D. Levitt, Alessandro Prelle, Gigliola Fagiolari, Marco Rimoldi, and Massimo Zeviani. Loss of ETHE1, a mitochondrial dioxygenase, causes fatal sulfide toxicity in ethylmalonic encephalopathy. *Nat Med,* 2009.

39. Jacques C. Mbongue, Dequina A. Nicholas, Timothy W. Torrez, Nan-Sun Kim, Anthony F. Firek, William H.R. Langridge. The Role of Indoleamine 2, 3-Dioxygenase in Immune Suppression and Autoimmunity. *Vaccines*, 2015.

CHAPTER 7

Mutations in *ARMC9,* Encoding Armadillo Repeat-Containing Protein 9, cause Joubert Syndrome with Frequent Corpus Callosal Defects

7.1 Abstract

Joubert syndrome (JS), is characterized by cerebellar hypoplasia and neurological features, including ataxia, psychomotor delay, and oculomotor apraxia, with frequent association with retinal dystrophy and nephronophthisis. Corpus callosal defects were reported in a minority of patients with JS, but the genetic basis of the association has not been determined. Through review of a JS cohort of 300 families, we identified 10 patients from 9 families with biallelic mutations in *ARMC9,* a gene recently reported to associate with JS. From these 10 patients, 8 demonstrated thinning or misshapen corpus callosum. All variants segregated with the disease and affected individuals exhibited classical JS including the molar tooth sign on axial MRI, hypotonia, ataxia, and psychomotor delay with oculomotor apraxia and nystagmus. These data suggest *ARMC9* as a contributor to cases of JS with coexistent callosal defects, and joins a growing list of genotype-phenotype correlations in JS.

7.2 Introduction

Cilia are microtubule-based structures found on almost all vertebrate cells, responsible for mechanoreception and locomotion, either of the cell itself or of fluids on the cell surface.[1] Ciliopathies are disorders caused by defects of the nonmotile (i.e. primary) cilium, which acts as a signaling hub and a focus of cell polarity. A range of phenotypes are associated with ciliopathies,

reflecting their organ-specific and cell-type-specific roles throughout the body. Over 80 different genes are mutated collectively in the ciliopathies, nearly all of which are inherited in a recessive fashion. Despite shared phenotypes involving the most commonly involved organs of kidney, eye and brain, important genotype-phenotype correlations can suggest specific syndromes, or even hint at which of the potential genes is most likely to be mutated for a given presentation.[2]

Joubert syndrome (JS) is one of the ciliopathies, characterized by ataxia, hypotonia, and mental retardation with a pathognomonic 'molar tooth sign' on axial brain magnetic resonance imaging.[3] Associated phenotypes include oculomotor apraxia, ptosis, polydactytly, retinal blindness, nephronophthisis, and breathing dysregulation, with partial penetrance. Occasional mention of corpus callosal defects appear in the literature, but the genetic basis remains mostly unexplored. Biallelic or X-linked mutations in more than 35 genes are associated with the condition, and for nearly all of them, the protein product of the gene is located at or near the primary cilium.[4] Loss of these genes typically impacts ciliary growth and stability, seemingly acting on the 'transition zone,' a domain of the cilium between the basal body and the axoneme, which separates the ciliary membrane from the plasma membrane.[5]

Recently, biallelic variants (stop-gain, missense, splice-site, and single-exon deletion) within *ARMC9* (MIM:610971) were identified in 14 affected individuals from nine independent JS families. Most showed JS features limited to brain, but a few patients showed retinal dystrophy, pituitary insufficiency, Dandy-Walker malformation and/or synpolydactly.[6,7] Here, we report 10 additional cases from nine independent families with classical JS harboring deleterious biallelic variants in *ARMC9*, and demonstrate frequent association with defects of the corpus callosum, thus further refining the clinical presentation of *ARMC9*-related JS.

7.3 Materials and Methods

7.3.1 Participant Recruitment and Whole Exome Sequencing

The institutional review boards of the University of California, San Diego, as well as the respective host institution approved this study. Whole exome sequencing was performed on one affected member, or sometimes both parents and affected member per JS family. Genomic DNA was subject to Agilent Human All Exon 50Mb kit library preparation, then paired-end sequencing (2x150bp) on Illumina HiSeq 2000 instrument. For each patient sample, >96% of the exome was covered at >12x. GATK was used for variant identification. We tested for segregating rare structural variants using XHMM. We then prioritized homozygous variants, removed alleles with >0.1% MAF in the sequenced population and ExAC, and prioritized highly conserved variants identified by high scores to likely damage the protein function. Variants with CADD score of >20 and SIFT score of <0.1 were considered as damaging and tested for segregation with a disease.

7.3.2 Reverse Transcription (RT) PCR and Gel Extraction

Patient fibroblasts for Family 2612 and an unrelated control were harvested from a skin dermal punch biopsy as described.[8] Cells were cultured in serum free conditions for 36h to promote ciliogenesis prior to harvest. Total RNA from quantified patient and control fibroblasts was reverse-transcribed using the Superscript III First-Strand cDNA Kit (Invitrogen, Cat: 18080051). PCR analysis of cDNA was designed against exons 8-11 of *ARMC9*, using GoTaq Master Mix (Promega, Cat: M7833). DNA was purified from agarose gels following RT-PCR and recovered using the Zymoclean Gel DNA Recovery Kit as per manufacturer's instructions (Zymo Research, Cat: D4001). Sanger sequencing was performed on purified DNA for analysis.

7.4 Results

7.4.1 Clinical Evaluation of the Affected Individuals

We report a cohort of 10 patients from nine families of different ancestries (Turkish, Egyptian, Iranian) with an *ARMC9*-related neurodevelopmental disorder characterized by classical JS (molar tooth sign in cranial MRI, hypotonia, ataxia, and psychomotor delay) and additional features of oculomotor apraxia and nystagmus (Fig.7.1A-B, Supplemental File 5). Similar to the previously reported patients, all affected individuals displayed classical features of JS including hypotonia, psychomotor delay and intellectual disability, with none of the affected individuals noting kidney or liver involvement. Additional isolated phenotypes were also present in most individuals including oculomotor apraxia and nystagmus in eight out of 10 patients, and neonatal periodic breathing abnormalities in six of the 10 patients (Supplemental File 5). None noted a history of seizure activity.

Brain MRIs showed thickened and mal-oriented superior cerebellar peduncle, deepened interpeduncular fossa, and vermis hypoplasia, consistent with the 'molar tooth sign' in all patients (Fig. 7.1B). The 'molar tooth sign' was also evident in all previously reported cases where MRIs were available.[6,7] In eight of the 10 patients of our patients, a thin corpus callosum was also evident (Fig. 7.1B and Supplemental File 5). In these patients, the corpus callosum was either dramatically thinned, as in the affected member from family 1693, both children from families 2007, 2611, 3083 3202 and 3961 or was foreshortened or excessively rounded in appearance in family 2594. There were no notable clinical differences in cognition in patients with and without corpus callosal defects.

In our cohort of ~300 patients with JS, many with molecular diagnosis, there are a total of 20 with notable callosal defects, including the 10 reported here. Three of these 20 were previously

diagnosed with acrocallosal syndrome, and displayed *KIF7* mutations, but the remainder had been without a molecular diagnosis. Thus, among our cohort, *ARMC9* accounts for ~ 3% of all JS cases, and accounts for ~50% of all JS cases with callosal defects.

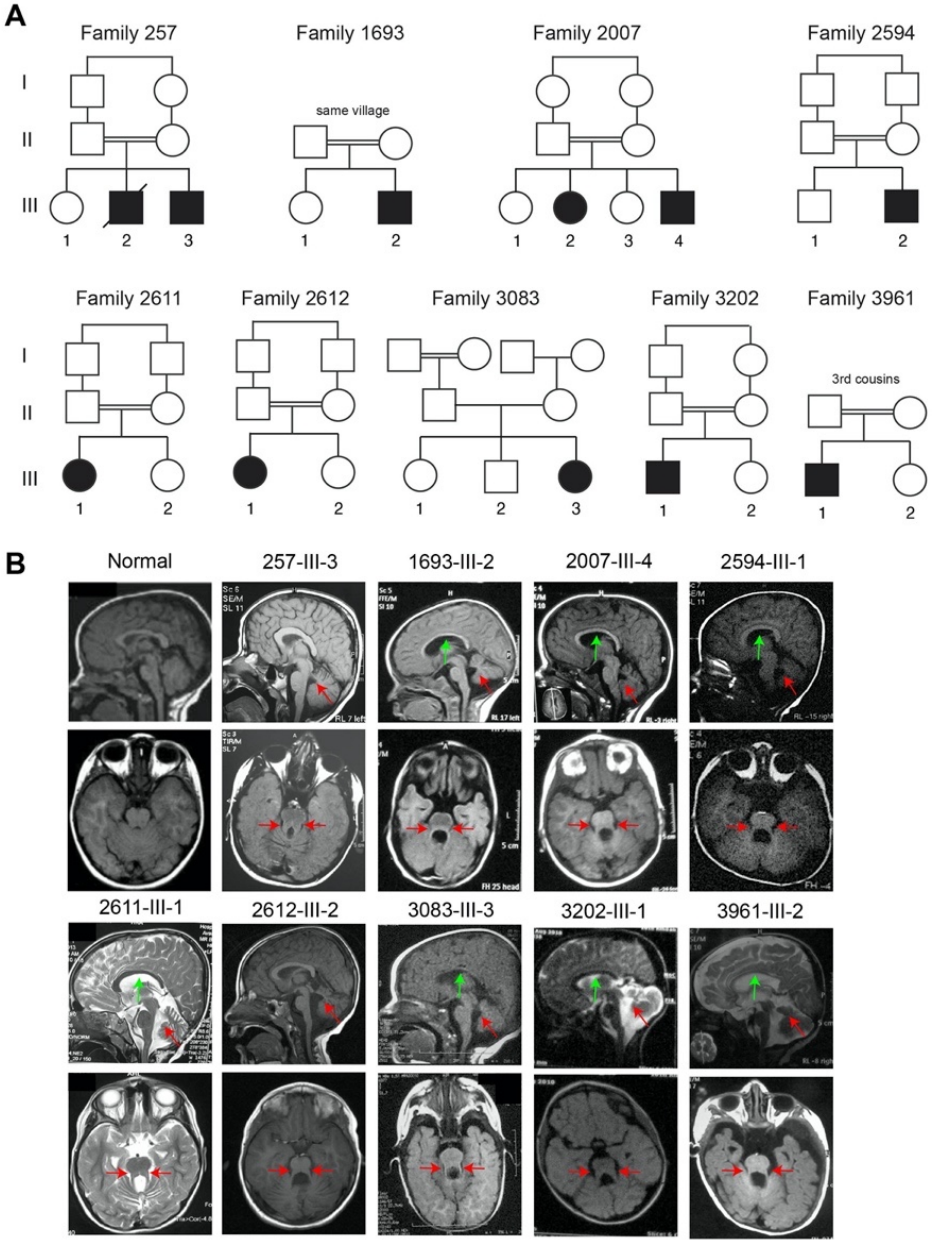

Figure 7.1: Pedigrees and radiological characterization of *ARMC9* patients.
(A) Double bar: consanguinity. Slash: deceased. (B) Brain MRI scans from a healthy individual and patients with *ARMC9* biallelic mutations, showing thickened and mal-oriented superior cerebellar peduncle (upper, red 'arrows'), deepened interpeduncular fossa constituting the 'molar tooth sign' on axial images (lower, red arrowheads). Corpus callosal defects (green arrows) were noted in all but case 257 and 2612, demonstrating thinning and excessive foreshortening and rounding.

7.4.2 Identification of Deleterious ARMC9 Variants

After obtaining informed consent from all participating individuals in accordance with the ethical standards set by the University of California, San Diego IRB, we identified a total of five distinct variants in *ARMC9* in nine independent families by exome sequencing, two of which have not been identified previously. Families 257 (c.1027C>T, p.Arg343Cys), 1693 (c.51+5G>T, p.?), 2007 (c.51+5G>T, p.?), 3961 (c.51+5G>T, p.?), 2611 (c.879G>A, p.Ile261_Thr293del), 2612 (c.879G>A, p.Ile261_Thr293del), and 3202 (c.1027C>T, p.Arg343Cys) carried mutations previously reported.[6,7] We then recruited Family 3083 with one affected female showing classical JS phenotypes. No homozygous variants in any gene passed filter criteria. Instead, we identified compound heterozygous c.51+5G>T, c.1307T>C variants in *ARMC9*, which was the only candidate identified by exome sequencing and segregated in the family. Family 2594 was recruited with documented parental consanguinity with one affected male showing a nearly identical clinical pattern as the other eight families. We identified a novel homozygous variant in *ARMC9* in the affected individual as the most likely cause of disease (c.1528C>G, p.Leu510Val and c.1307T>C, p. Leu436Pro) (Fig. 7.2A).

ARMC9 protein is predicted to have a Lissencephaly type-1-like homology (LisH) motif, and armadillo repeats (Figure 7.2A). Two of the missense variants occur within the armadillo repeat domain, while the remaining variants are not located in known domains. All missense variants identified alter evolutionarily conserved residues among vertebrates (Figure 7.2B).

7.4.3 ARMC9 RT-PCR Analysis Reveals Defective Splicing

The c.879G>A variant occurs at the last base of exon 9, suggesting defective splicing may result. We therefore examined the effect of the splicing mutation (c.879G>A, p.Ile261_Thr293del) by designing primers flanking the exon 9 region harboring the mutation (Figure 7.2C). RT-PCR

analysis from patient-derived fibroblasts confirmed skipping of exon 9 in *ARMC9* transcripts (Figure 7.2D-E).

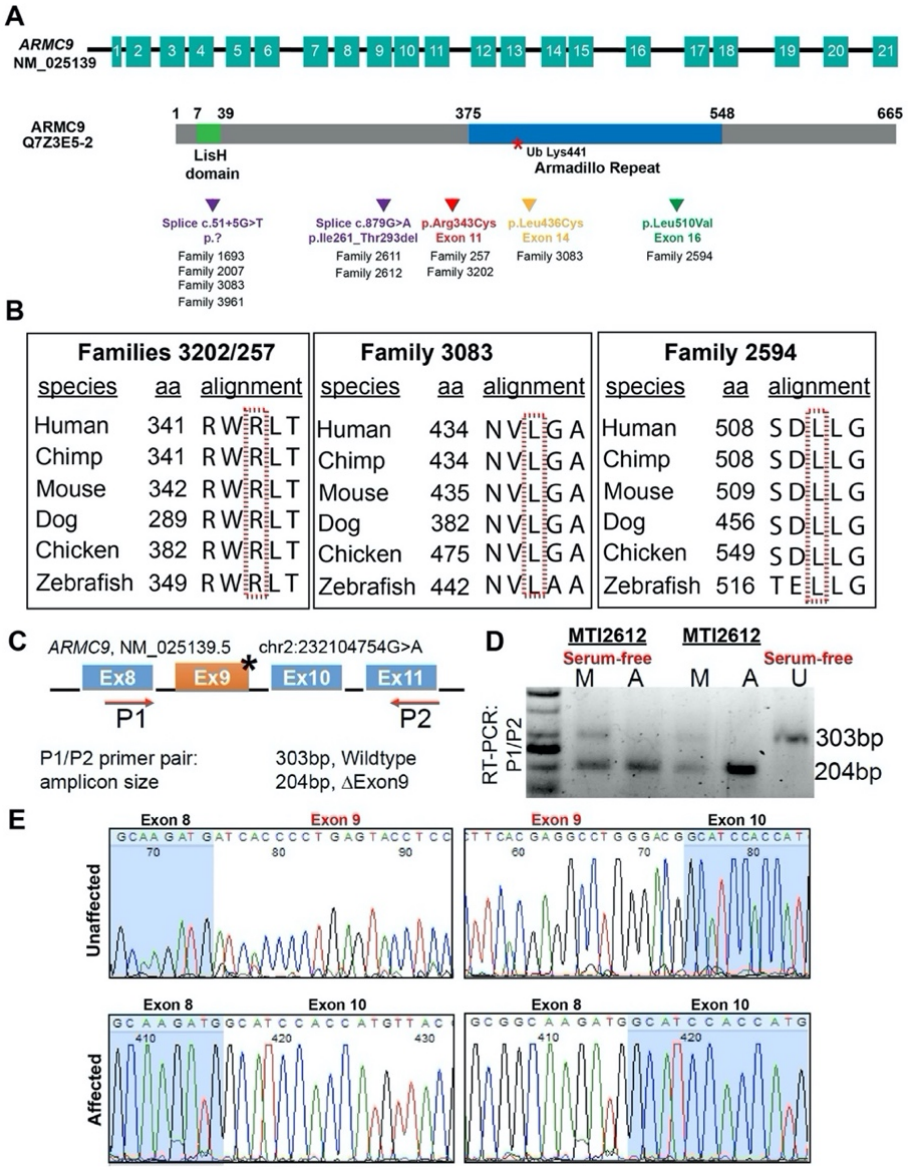

Figure 7.2: Locations and conservation of *ARMC9* mutations and splicing analysis.
(A) Genomic structure (NM_025139, top panel) and protein structure (Q7Z3E5-2, lower panel) of ARMC9 full-length isoform with 24 exons and 665 amino acids. The locations of the mutations are indicated below. (B) Missense mutations (boxed in red) occur at evolutionally conserved residues in all species shown. (C) RT-PCR primers relative to *ARMC9* genomic structure and the predicted amplicon size (bp). (D) RT-PCR results performed on cDNA from patient-derived fibroblasts. 50bp ladder (Lane 1). Fibroblasts from carrier (M, mother) and affected (A) patients from family MTI-2612, cultured in serum-free condition for 36 hours (Lane 2&3), or non-serum-free medium (Lane 4&5). Fibroblasts from unaffected individual, cultured in serum-free condition for 36 hours (Lane 6). (E) Chromatograms depict skipped exon 9 in cDNA of *ARMC9* transcript from affected MTI2612 patient fibroblasts.

7.5 Discussion

This study supports the pathogenicity of deleterious variants in *ARMC9* as the cause of a severe neurodevelopmental disorder characterized by classical JS (molar tooth sign in cranial MRI, hypotonia, ataxia, and psychomotor delay) and additional features including oculomotor apraxia and nystagmus. Through whole exome sequencing, we identified additional families harboring similar variants as previously reported, in addition to identifying two novel deleterious variants in *ARMC9.* We also confirmed the c.879G>A variant as leading to skipping of exon 9, as previously shown.[7]

Together with those previously reported, the *ARMC9* phenotype can be described a classic JS with variably observed postaxial polydactyly, seizures, and neonatal periodic breathing abnormalities. Neuroimaging features demonstrate consistent molar tooth sign; however, in contrast to cases from previous reports,[6,7] we also note remarkable thin or dysmorphic corpus callosum, observed in most of our patients. We postulate that these callosal defects represent a developmental rather than a degenerative origin, although serial MRIs were not available to address this question.

Defects in the corpus callosum have been occasionally observed in association with JS and other ciliopathy conditions, but in an inconsistent pattern, and without a clear genetic association, but more often in patients with other midline defects such as Dandy-Walker, encephalocele, midline lipomas or cysts.[9-11] In a study of 71 patients with JS, there was evidence of dysgenic CC in 6 patients (8.45%), but specific genes were not correlated with these findings.[12] The gene *C5orf42* is the major gene responsible for Oro-Facial-Digital syndrome type VI (i.e. Varadi-Papp syndrome), which includes midline features such as tongue hamartomas, frenulum and/or upper lip notch, mesaxial polydactyly and hypothalamic hamartomas, but

corpus callosal defects were observed in a single case.[13] The gene *MKS1* (MIM: 609883) was reported in a single JS case with callosal agenesis.[14] Although acrocallosal syndrome can show both corpus callosal defects and JS-like features (in fact, it is considered JS type 12), it can usually be clinically differentiated based upon other associated malformations.[15,16] Therefore, of the genes linked to JS to date, we suggest mutations in *ARMC9* to be most consistently associated with corpus callosal defects in relation to JS, with this phenotype presenting in 25% (8/24) of all previously and newly identified *ARMC9* patients.

Future studies will be required to characterize the full range of phenotypes, genotype-phenotype correlations, and mechanisms of pathogenicity; however, two previous studies have not only revealed mutations in *ARMC9* leading to JS in humans, but also illustrated ciliopathy features using Zebrafish modeling and yeast 2-hybrid screening, providing functional evidence supporting *ARMC9* as a ciliopathy-associated gene.[6,17] These studies confirmed localization of ARMC9 to the ciliary basal body (similar to other JS-associated proteins), illustrated ciliopathy phenotypes in CRISPR-engineered zebrafish harboring *armc9* mutations, and used ARMC9 in tandem-affinity purification and yeast 2-hybrid screens to identify a ciliary module whose dysfunction underlies JS,[6,17] thus providing further evidence linking ARMC9 with the biological mechanisms underlying JS.

ARMC9 is not the first ARM-containing protein to be implicated in a ciliopathy. ARMC4 (OMIM: 615408), a protein required for a late step in anchoring dynein outer arms, is linked to autosomal recessive primary ciliary dyskinesia (OMIM: 615451), a disorder resulting from defective ciliary motility leading to respiratory distress and recurrent airway infections.[18,19] Although the exact role of ARMC9 in causing JS remains to be elucidated, revelation of these

deleterious variants further extends the genetic and phenotypic heterogeneity of JS, while also expanding on the types of genes implicated in various ciliopathies.

Chapter 7, in full, has been submitted for publication of the material as it may appear in American Journal of Medical Genetics 2020. Ghosh, Shereen G.; Hsiao, Katie K.; Gregor, Anne; Shimazu, Junko; Musaev, Damir; Stanley, Valentina; Roosing, Susanne; Silhavy, Jennifer L.; Issa, Mahmoud Y.; Elbendary, Hasnaa; Sahin, Yasin; Kariminejad, Ariana; Zaki, Maha S.; Gleeson, Joseph G. The dissertation author was the primary investigator and author of this material.

7.6 References

1. Friedhelm Hildebrandt, Thomas Benzing, and Nicholas Katsanis. Ciliopathies. *N Eng J Med,* 2011.

2. Melissa A. Parisi. The molecular genetics of Joubert syndrome and related ciliopathies: The challenges of genetic and phenotypic heterogeneity. *Transl Sci Rare Dis*, 2019.

3. Shifteh Sattar, and Joseph G. Gleeson. The ciliopathies in neuronal development: a clinical approach to investigation of Joubert syndrome and Joubert syndrome-related disorders. *Dev Med Child Neurol*, 2011.

4. T. Tony Yang, Jimmy Su, Won-Jing Wang, Branch Craige, George B. Witman, Meng-Fu Bryan Tsou, and Jung-Chi Liao. Superresolution Pattern Recognition Reveals the Architectural Map of the Ciliary Transition Zone. *Sci Rep*, 2015.

5. Xiaoyu Shi, Galo Garcia III, Julie C. Van De Weghe, Ryan McGorty, Gregory J. Pazour, Dan Doherty, Bo Huang, Jeremy F. Reiter. Super-resolution microscopy reveals that disruption of ciliary transition-zone architecture causes Joubert syndrome. *Nat Cell Biol,* 2017.

6. Julie C. Van De Weghe, Tamara D.S. Rusterholz, Brooke Latour, Megan E. Grout, Kimberly A. Aldinger, Ranad Shaheen, Jennifer C. Dempsey, Sateesh Maddirevula, Yong-Han H.

Cheng, Ian G. Phelps, Matthias Gesemann, Himanshu Goel, Ohad S. Birk, Talal Alanzi, Rifaat Rawashdeh, Arif O. Khan, University of Washington Center for Mendelian Genomics, Michael J. Bamshad, Deborah A. Nickerson, Stephan C.F. Neuhauss, William B. Dobyns, Fowzan S. Alkuraya, Ronald Roepman, Ruxandra Bachmann-Gagescu, and Dan Doherty. Mutations in ARMC9, which Encodes a Basal Body Protein, Cause Joubert Syndrome in Humans and Ciliopathy Phenotypes in Zebrafish. *Am J Hum Genet*, 2017.

7. Anjana Kar, Shubha R. Phadke, Aneek Das Bhowmik, and Ashwin Dalal. Whole exome sequencing reveals a mutation in ARMC9 as a cause of mental retardation, ptosis, and polydactyly. *Am J Med Genet A,* 2018.

8. Jacob Villegas and Michael McPhaul. Establishment and Culture of Human Skin Fibroblasts. *Curr Protoc Mol Biol,* 2005.

9. Efsun U. Senocak, Kader Karli Oguz, Goknur Haliloglu, Meral Topcu, and Aysenur Cila. Structural abnormalities of the brain other than molar tooth sign in Joubert syndrome-related disorders. *Diagn Interv Radiol,* 2010.

10. Nelia Zamponi, B. Rossi, Gabriele Polonara, Luana Regnicolo, and C. Cardinali. Joubert syndrome with associated corpus callosum agenesis. *Eur J Paediatr Neurol,* 2002.

11. Ekrem Karakas, Nesat Cullu, Omer Karakas, Mustafa Calik, Fatima Nurefsan Boyaci, Sema Yildiz, Hasan Cece, and Ali Akal. Joubert syndrome: the clinical and radiological findings. *J Pak Med Assoc,* 2014.

12. Andrea Poretti, T.A.G.M Huisman, Ianina Scheer, and Eugen Boltshauser. Joubert syndrome and related disorders: spectrum of neuroimaging findings in 75 patients. *AJNR Am J Neuroradiol,* 2011.

13. Estelle Lopez, Christel Thauvin-Robinet, Bruno Reversade, Nadia El Khartoufi, Louise Devisme, Muriel Holder, Hélène Ansart-Franquet, Magali Avila, Didier Lacombe, Pascale Kleinfinger, Irahara Kaori, Jun-Ichi Takanashi, Martine Le Merrer, Jelena Martinovic, Catherine Noël, Mohammad Shboul, Lena Ho, Yeliz Güven, Ferechté Razavi, Lydie Burglen, Nadège Gigot, Véronique Darmency-Stamboul, Julien Thevenon, Bernard Aral, Hülya Kayserili, Frédéric Huet, Stanislas Lyonnet, Cédric Le Caignec, Brunella Franco, Jean-Baptiste Rivière, Laurence Faivre, and Tania Attié-Bitach. C5orf42 is the major gene responsible for OFD syndrome type VI. *Hum Genet,* 2014.

14. Ingrid Bader, Eva Decker, Johannes A. Mayr, Verena Lunzer, Johannes Koch, Eugen Boltshauser, Wolfgang Sperl, Peter Pietsch, B. Ertl-Wagner, H. Bolz, Carsten Bergmann, and Olaf Rittinger. MKS1 mutations cause Joubert syndrome with agenesis of the corpus callosum. *Eur J Med Genet,* 2016.

15. Audrey Putoux, Sophie Thomas, Karlien L M Coene, Erica E Davis, Yasemin Alanay, Gönül Ogur, Elif Uz, Daniela Buzas, Céline Gomes, Sophie Patrier, Christopher L Bennett, Nadia Elkhartoufi, Marie-Hélène Saint Frison, Luc Rigonnot, Nicole Joyé, Solenn Pruvost, Gulen Eda Utine, Koray Boduroglu, Patrick Nitschke, Laura Fertitta, Christel Thauvin-Robinet, Arnold Munnich, Valérie Cormier-Daire, Raoul Hennekam, Estelle Colin, Nurten Ayse Akarsu, Christine Bole-Feysot, Nicolas Cagnard, Alain Schmitt, Nicolas Goudin, Stanislas Lyonnet, Férechté Encha-Razavi, Jean-Pierre Siffroi, Mark Winey, Nicholas Katsanis, Marie Gonzales, Michel Vekemans, Philip L Beales, Tania Attié-Bitach. KIF7 mutations cause fetal hydrolethalus and acrocallosal syndromes. *Nat Genet,* 2011.

16. Bassam R Ali, Jennifer L Silhavy, Nadia A Akawi, Joseph G Gleeson, and Lihadh Al-Gazali. A mutation in KIF7 is responsible for the autosomal recessive syndrome of macrocephaly, multiple epiphyseal dysplasia and distinctive facial appearance. *Orphanet J Rare Dis,* 2012.

17. Brooke L. Latour, Julie C. Van De Weghe, Tamara Ds Rusterholz, Stef Jf Letteboer, Arianna Gomez, Ranad Shaheen, Matthias Gesemann, Arezou Karamzade, Mostafa Asadollahi, Miguel Barroso-Gil, Manali Chitre, Megan E. Grout, Jeroen van Reeuwijk, Sylvia Ec van Beersum, Caitlin V. Miller, Jennifer C. Dempsey, Heba Morsy, University of Washington Center for Mendelian Genomics, Michael J. Bamshad, Genomics England Research Consortium, Deborah A Nickerson, Stephan Cf Neuhauss, Karsten Boldt, Marius Ueffing, Mohammad Keramatipour, John A. Sayer, Fowzan S. Alkuraya, Ruxandra Bachmann-Gagescu, Ronald Roepman, Dan Doherty. Dysfunction of the ciliary ARMC9/TOGARAM1 protein module causes Joubert syndrome. *J Clin Invest,* 2020.

18. Rim Hjeij, Anna Lindstrand, Richard Francis, Maimoona A. Zariwala, Xiaoqin Liu, You Li, Rama Damerla, Gerard W. Dougherty, Marouan Abouhamed, Heike Olbrich, Niki T. Loges, Petra Pennekamp, Erica E. Davis, Claudia M.B. Carvalho, Davut Pehlivan, Claudius Werner, Johanna Raidt, Gabriele Köhler, Karsten Häffner, Miguel Reyes-Mugica, James R. Lupski, Margaret W Leigh, Margaret Rosenfeld, Lucy C. Morgan, Michael R. Knowles, Cecilia W. Lo, Nicholas Katsanis, Heymut Omran. ARMC4 mutations cause primary ciliary dyskinesia with randomization of left/right body asymmetry. *Am J Hum Genet,* 2013.

19. Alexandros Onoufriadis, Amelia Shoemark, Mustafa M. Munye, Chela T. James, Miriam Schmidts, Mitali Patel, Elisabeth M. Rosser, Chiara Bacchelli, Philip L. Beales, Peter J. Scambler, Stephen L. Hart, Jeannette E. Danke-Roelse, John J. Sloper, Sarah Hull, Claire Hogg, Richard D. Emes, Gerard Pals, Anthony T. Moore, Eddie M.K. Chung, UK10K, Hannah M. Mitchison. Combined exome and whole-genome sequencing identifies mutations in ARMC4 as a cause of primary ciliary dyskinesia with defects in the outer dynein arm. *J Med Genet,* 2014.

CHAPTER 8

Discussion

The structural organization and maturation of the brain are the result of a precisely orchestrated series of developmental processes with complex genetic regulation that occur during embryonic gestation.[1] Thus, the utmost precision is required to ensure all functional connections are properly established. Mutations affecting any of the cellular processes involved in proper human brain development can lead to altered neurodevelopment and result in a number of different 'neurodevelopmental diseases.'[2,3] Structural Brain Defects (SBDs) represent a large subset of neurodevelopmental diseases leading to disruption of the cerebral, cerebellar or deep brain structures and are major causes of severe neurological disability including epilepsy, intellectual disability, and chronic motor disability, often leading to death.[4,5]

As opposed to SBDs, which are stable over time, patients harboring pediatric-onset neurodegenerative diseases display progressive atrophy across a wide range of brain structures.[6-8] SBDs and pediatric-onset neurodegenerative diseases which arise as a consequence of a biallelic mutation in a single gene are much more prevalent among offspring of consanguineous marriages, who are homozygous across 6% of the genome on average.[9] Such consanguinity increases the odds that a deleterious mutation will be inherited on both chromosomes, and results in a higher-than-expected recurrence risk for future pregnancies in these families, compared to diseases caused by *de novo* mutations, which can arise spontaneously in the offspring during development.[10] The advent of next-generation sequencing (NGS), enabling sequencing at the scales of genomes, has ushered in an era of unprecedented pace of Mendelian disease gene

discovery.[11] Therefore, by taking advantage of the large number of consanguineous families in our cohort, in combination with NGS and experimental techniques, we were not only able to explore the cause of disease directly in humans, but also to determine genetic disease risk and to identify novel disease-causing variants with a high level of certainty.

8.1 Summary of Findings

Through this work, it is evident that the genes that cause these recessive brain defects when mutated fall on a broad spectrum of various cellular and molecular pathways, illustrating the high sensitivity of the developing CNS compared with other organ systems, making it more susceptible to disease. Of the six genes examined in this study, we illustrate *ADPRHL2* and *HPDL* as causing novel early-onset neurodegenerative syndromes with previously unknown cause, while *TMX2* and *HEATR5B* were found to cause neurodevelopmental diseases, microlissencephaly and PCH, respectively. Although *TRAPPC4* was linked to disease prior to our study,[12,13] our identification of 23 additional cases from 17 independent families with similar phenotypes, harboring the same homozygous *TRAPPC4* splice site variant, highlights the involvement of this relatively common single allele as the cause for a severe recessive neurological disease. Additionally, *ARMC9* was also linked to Joubert syndrome prior to our study;[14,15] however, we extended the phenotype by the occurrence of a thin or dysmorphic corpus callosum, observed in 8/10 assessed patients. Thus, of the genes linked to JBTS to date, we suggest mutations in *ARMC9* to be most consistently associated with corpus callosal defects. Prior to our study, defects in the corpus callosum had only been occasionally observed in this syndrome and other ciliopathy conditions, but in an inconsistent pattern, and without a clear genetic association.

Although individuals harbor mutations in the same gene, our observations show that their phenotypes can still differ drastically, even among siblings from the same family, which cannot always be explained genetically. The sources of such phenotypic variability may include both additional genetic factors in conjunction with non-genetic or environmental factors.[16] Thus, translating this work into the clinic may require personalized regimens catered to each patient individually based on epigenetic factors. Using genomics is essential to understand both the normal and disease-related functions of the genetic and epigenetic contributors to disease and the cellular pathways and biological processes in which they are involved.[17] Such an understanding is critical to the development of improved strategies for diagnosis, prevention and therapeutic intervention.

Our work also highlights WES to be among the most useful approaches in the identification of novel causal mutations.[18] WGS further elucidates the role of non-coding mutations in the development of SBD phenotypes, adding an extra dimension to the already complex etiology of these diseases. Through application of WGS in some of our families in this study, we were able to identify causative mutations that would have otherwise remained unknown with the sole use of WES. Therefore, WGS should be considered whenever exome analyses do not yield candidate mutations in SBD phenotypes. The vast information obtained from NGS has changed the lives of many individuals who suffer from these disorders without knowing the exact cause, and empowered families to make more informed decisions, a trend that is only likely to accelerate in the future.

8.2 Detection of Congenital Disorders

Congenital malformations occur in 2-4% of all births, and are the leading cause of infant mortality in the United States,[19-21] with many of these being monogenic.[22] Currently, newborns

and infants with congenital malformations, syndromes, and inherited disorders typically undergo an extensive diagnostic process, with relatively low rates of diagnosis.[23] As a result, decisions regarding management of care for such patients are often made in the absence of a definitive diagnosis, either leading to delays in effective treatment regimens, or to the use of ineffective treatments, which can often times have adverse effects and/or exacerbate symptoms.[24] Thus, prenatal or early perinatal diagnosis of such malformations provides families with information for decisions during pregnancy and appropriate treatment perinatally, improving both perinatal and long-term outcomes.

Over the years, prenatal diagnosis has greatly benefited from advances in ultrasound and other screening technologies and in our ability to detect single gene disorders, leading to significant improvements in detecting such congenital anomalies.[25] In a European multicenter study involving 3,686 malformed fetuses, congenital anomalies of the CNS represented 16% of all anomalies studied and ranked highest in detection sensitivity,[26] specifically highlighting the importance of such screening in providing families like the ones studied here with accurate information regarding etiology, prognosis, and recurrence risk for future pregnancies.

As disease progression is rapid in patients with SBDs, establishing an etiologic diagnosis must be equally fast to inform possible interventions that can lessen suffering and mortality. Traditionally, chromosomal microarray analysis is the recommended first-tier genomic test for children with genetic diseases,[27,28] with additional testing including newborn screening panels, metabolic testing, and single-gene sequencing, among others. Over the last few years, WGS and WES have started to gain broad use for diagnosis of infants and children with suspected genetic diseases, yet can still take weeks to return results. Most recent technological advances through the use of rapid whole-genome sequencing (rWGS) have led to the ability to sequence and

interpret the entire genome of a neonate in less than 50 hours.[24] Thus, in cases where diagnosis leads to palliation of symptoms or effective treatments, compared to WES or WGS, rWGS has the potential to shorten hospital stays and minimize ineffective treatment regimens, improving the overall outcome for these patients.

8.3 Therapeutic Strategies and Future Implications

As we have shown in these studies, taking advantage of unique family structures in conjunction with mutation modeling in a dish or in animals can begin to provide useful insights into pathogenic mechanisms. Few other approaches allow for such direct insight into the human brain; and identifying genes in families with inherited CNS disorders has led to a critical understanding of the genetic control of human brain development and function. Although most of these diseases are still far from therapeutic targeting, our approach and findings offer many benefits which can ultimately translate to clinical practice, including 1) the diagnosis of brain disorders with previously unknown cause 2) the ability to determine carrier status and provide genetic counseling 3) the potential to design new, gene-specific therapeutic targets, providing further accessibility to personalized genomic medicine in the future.

Although our list of genetic causes of SBDs is far from being exhaustive, significant advances have been made in parallel with elucidating the specific biological pathways that are underlying the molecular mechanisms of these illnesses. Discerning these underlying pathological mechanisms and the identification of therapeutic strategies to treat individuals affected with rare neurological diseases has proven challenging due to a host of factors. One key challenge is that most identified genes are not readily targeted with current small-molecule therapeutics, with the blood-brain barrier posing a major challenge from a drug-discovery perspective.[29] In hopes of overcoming this challenge, the recent discovery of antisense

therapeutics using synthetic RNA-binding oligonucleotides is providing promise in the ability to bind to targeted RNA and modulate its function by several different mechanisms.[29-31] One successful example of the use of this therapy, in conjunction with genome sequencing, is milasen: a splice-modulating antisense oligonucleotide drug specifically tailored to a particular patient with a novel mutation leading to neuronal ceroid lipfuscinosis 7, a rare and fatal neurodegenerative disease.[32] Identification of the mutation was followed by the development and administration of the personalized drug within 1 year, significantly extending the patient's lifespan and quality of life.[32] However, thus far, this approach has only shown potential in diseases that are neurodegenerative in nature, as opposed to pre-existing SBDs present at birth due to a neurodevelopmental defect.

In summary, our work provides a profound positive impact for the families that were studied and future families harboring similar mutations. Knowledge of the genetic etiology in such cases will help the health care system provide meaningful distinctions in genetic risk of great importance to individuals and their families, while providing clinicians with a roadmap to treating such patients whose diagnoses were previously unknown. The ability to offer a diagnosis frequently holds power for a family, regardless of the impact on clinical management. Currently, for many genetic diseases like the ones studied here, there is not an evidence-based therapeutic literature available. These diseases are very rare and often times progress rapidly, leaving a very narrow window for implementing investigational new treatments. The ultimate hope is to use genetic discoveries like ours to learn enough about the underlying neurobiology to generate novel and patient-specific therapeutic approaches for these common and devastating disorders, and our work provides a significant platform in this direction.

8.4 References

1. Byoung-il Bae, Divya Jawaraman, and Christopher A. Walsh. Genetic changes shaping the human brain. *Dev Cell,* 2016.

2. Ariana P. Mullin, Avanti Gokhale, Andres Moreno-De-Luca, Subhabrata Sanyal, John L. Waddington, and Victor Faundez. Neurodevelopmental disorders: mechanisms and boundary definitions from genomes, interactomes and proteomes. *Transl Psychiatry,* 2013.

3. Ana R. Cardoso, Monica Lopes-Marques, Raquel M. Silva, Catarina Serrano, Antonio Amorim, Maria J. Prata, and Luisa Azevedo. Essential genetic findings in neurodevelopmental disorders. *Hum Genomics,* 2019.

4. Matthew T. Whitehead, Stanley T. Fricke, and Andrea L. Gropman. Structural brain defects. *Clin Perinatol,* 2015.

5. Ute Hehr and Gerhardt Schuierer. Genetic assessment of cortical malformations. *Neuropediatrics,* 2011.

6. Michael E. Coulter, Damir Musaev, Ellen M. DeGennaro, Xiaochang Zhang, Katrin Henke, Kiely N. James, Richard S. Smith, R. Sean Hill, Jennifer N. Partlow, Muna Al-Saffar, A. Stacy Kamumbu, Nicole Hatem, A James Barkovich, Jacqueline Aziza, Nicolas Chassaing, Maha S. Zaki, Tipu Sultan, Lydie Burglen, Anna Rajab, Lihadh Al-Gazali, Ganeshwaran H. Mochida, Matthew P. Harris, Joseph G. Gleeson, and Christopher A. Walsh. Regulation of human cerebral cortical development by EXOC7 and EXOC8, components of the exocyst complex, and roles in neural progenitor cell proliferation and survival. *Genet Med,* 2020.

7. Ashleigh E. Schaffer, Veerle R.C. Eggens, Ahmet Okay Caglayan, Miriam S. Reuter, Eric Scott, Nicole G. Coufal, Jennifer L. Silhavy, Yuanchao Xue, Hulya Kayserili, Katsuhito Yasuno, Rasim Ozgur Rosti, Mostafa Abdellateef, Caner Caglar, Paul R. Kasher, J. Leonie Cazemier, Marian A. Weterman, Vincent Cantagrel, Na Cai, Christiane Zweier, Umut Altunoglu, N. Bilge Satkin, Fesih Aktar, Beyhan Tuysuz, Cengiz Yalcinkaya, Huseyin Caksen, Kaya Bilguvar, Xiang-Dong Fu, Christopher R. Trotta, Stacey Gabriel, André Reis, Murat Gunel, Frank Baas, and Joseph G. Gleeson. CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. *Cell,* 2014.

8. Vandana Shashi, Maria M. Magiera, Dennis Klein, Maha Zaki, Kelly Schoch, Sabine Rudnik-Schöneborn, Andrew Norman, Osorio Lopes Abath Neto, Marina Dusl, Xidi Yuan, Luca Bartesaghi, Patrizia De Marco, Ahmed A. Alfares, Ronit Marom, Stefan T. Arold, Francisco J. Guzmán-Vega, Loren Dm Pena, Edward C. Smith, Maja Steinlin, Mohamed Oe Babiker, Payam Mohassel, A. Reghan Foley, Sandra Donkervoort, Rupleen Kaur, Partha S. Ghosh, Valentina Stanley, Damir Musaev, Caroline Nava, Cyril Mignot, Boris Keren, Marcello Scala, Elisa Tassano, Paolo Picco, Paola Doneda, Chiara Fiorillo, Mahmoud Y. Issa, Ali Alassiri, Ahmed Alahmad, Amanda Gerard, Pengfei Liu, Yaping Yang, Birgit Ertl-Wagner, Peter G. Kranz, Ingrid M. Wentzensen, Rolf Stucka, Nicholas Stong, Andrew S. Allen, David B. Goldstein, Undiagnosed Diseases Network, Benedikt Schoser, Kai M. Rösler, Majid Alfadhel, Valeria Capra, Roman Chrast, Tim M. Strom, Erik-Jan Kamsteeg, Carsten G. Bönnemann, Joseph G. Gleeson, Rudolf Martini, Carsten Janke, and Jan Senderek. Loss of tubulin deglutamylase CCP1 causes infantile-onset neurodegeneration. *EMBO J,* 2018.

9. C. Geoffrey Woods, James Cox, Kelly Springell, Daniel J. Hampshire, Moin D. Mohamed, Martin McKibbin, Rowena Stern, F. Lucy Raymond, Richard Sandford, Saghira Malik Sharif, Gulshan Karbani, Mustaq Ahmed, Jacquelyn Bond, David Clayton, and Chris F. Inglehearn. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet,* 2006.

10. Mohd Fareed and Mohammad Afzal. Genetics of consanguinity and inbreeding in health and disease. *Ann Hum Biol,* 2017.

11. Fowzan S. Alkuraya. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases.

12. Nicole J. Van Bergen, Yiran Guo, Noraldin Al-Deri, Zhanna Lipatova, Daniela Stanga, Sarah Zhao, Rakhilya Murtazina, Valeriya Gyurkovska, Davut Pehlivan, Tadahiro Mitani, Alper Gezdirici, Jayne Antony, Felicity Collins, Mary J.H. Willis, Zeynep H. Coban Akdemir, Pengfei Liu, Jaya Punetha, Jill V. Hunter, Shalini N. Jhangiani, Jawid M. Fatih, Jill A. Rosenfeld, Jennifer E. Posey, Richard A. Gibbs, Ender Karaca, Sean Massey, Thisara G. Ranasinghe, Patrick Sleiman, Chris Troedson, James R. Lupski, Michael Sacher, Nava Segev, Hakon Hakonarson, and John Christodoulou. Deficiencies in vesicular transport mediated by *TRAPPC4* are associated with severe syndromic intellectual disability. *Brain,* 2019.

13. Parneet Kaur, Rajagopal Kadavigere, Katta Mohan Girisha, and Anju Shukla. Recurrent bi-allelic splicing variant c.454+3A>G in TRAPPC4 is associated with progressive encephalopathy and muscle involvement. *Brain,* 2020.

14. Julie C. Van De Weghe, Tamara D.S. Rusterholz, Brooke Latour, Megan E. Grout, Kimberly A. Aldinger, Ranad Shaheen, Jennifer C. Dempsey, Sateesh Maddirevula, Yong-Han H. Cheng, Ian G. Phelps, Matthias Gesemann, Himanshu Goel, Ohad S. Birk, Talal Alanzi, Rifaat Rawashdeh, Arif O. Khan, University of Washington Center for Mendelian Genomics, Michael J. Bamshad, Deborah A. Nickerson, Stephan C.F. Neuhauss, William B. Dobyns, Fowzan S. Alkuraya, Ronald Roepman, Ruxandra Bachmann-Gagescu, and Dan Doherty. Mutations in ARMC9, which Encodes a Basal Body Protein, Cause Joubert Syndrome in Humans and Ciliopathy Phenotypes in Zebrafish. *Am J Hum Genet*, 2017.

15. Anjana Kar, Shubha R. Phadke, Aneek Das Bhowmik, and Ashwin Dalal. Whole exome sequencing reveals a mutation in ARMC9 as a cause of mental retardation, ptosis, and polydactyly. *Am J Med Genet A,* 2018.

16. Miklos Toth. Mechanisms of Non-Genetic Inhertiance and Psychiatric Disorders. *Neuropsychopharmacology,* 2015.

17. Azam Moosavi and Ali Motevalizadeh Ardekani. Role of Epigenetics in Biology and Human Disease. *Iran Biomed J,* 2016.

18. Miriam S. Reuter, Hasan Tawamie, Rebecca Buchert, Ola Hosny Gebril, Tawfiq Froukh, Christian Thiel, Steffen Uebe, Arif B. Ekici, Mandy Krumbiegel, Christiane Zweier, Juliane Hoyer, Karolin Eberlein, Judith Bauer, Ute Scheller, Tim M. Strom, Sabine Hoffjan, Ehab R. Abdelraouf, Nagwa A. Meguid, Ahmad Abboud, Mohammad Ayman Al Khateeb, Mahmoud Fakher, Saber Hamdan, Amina Ismael, Safia Muhammad, Ebtessam Abdallah, Heinrich Sticht, Dagmar Wieczorek, Andre Reis, and Rami Abou Jamra. Diagnostic Yield and Novel Candidate Genes by Exome Sequencing in 152 Consanguineous Families with Neurodevelopmental Disorders. *JAMA Psychiat,* 2017.

19. Philip M. Marden, David W. Smith, and Michael J. McDonald. Congenital anomalies in the newborn infant, including minor variations. *J Pediatr,* 1964.

20. Paula W. Yoon, Richard S. Olney, Muin J. Khoury, William M. Sappenfield, Gilberto F. Chavez, and Don Taylor. Contribution of Birth Defects and Genetic Defects to Pediatric Hospitalizations. *Arch Pediatr Adolesc Med,* 1997.

21. Sherry L. Murphy, Jiaquan Xu, Kenneth D. Kochanek, Elizabeth Arias. Mortality in the United States, 2017. *NCHS Data Brief,* 2018.

22. Joyce A. Martin, Brady E. Hamilton, Stephanie J. Ventura, Michelle J.K. Osterman, T.J. Mathews. Births: final data for 2011. *Natl Vital Stat Rep,* 2013.

23. Laurel K. Willig, Josh E. Petrikin, Laurie D. Smith, Carol J. Saunders, Isabelle Thiffault, Neil A. Miller, Sarah E. Soden, Julie A. Cakici, Suzanne M. Herd, Greyson Twist, Aaron Noll, Mitchell Creed, Patria M. Alba, Shannon L. Carpenter, Mark A. Clements, Ryan T. Fischer, J. Allyson Hays, Howard Kilbride, Ryan J. McDonough, Jamie L. Rosterman, Sarah L. Tsai, Lee Zellmer, Emily G. Farrow, Stephen F. Kingsmore. Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *Lancet Respir Med,* 2015.

24. Joshua E. Petrikin, Laurel K. Willig, Laurie D. Smith, and Stephen F. Kingsmore. Rapid whole genome sequencing and precion neonatology. *Semin Perinatol,* 2015.

25. David Chitayat and Riyana Babul-Hirji. Genetic counselling in prenatally diagnosed non-chromosomal fetal abnormalities. *Curr Opin Obstet Gynecol,* 2000.

26. Helene Grandjean, Daniele Larroque, and Salvator Levi. The performance of routine ultrasonographic screening of pregnancies in the Eurofetus Study. *Am J Obstet Gynecol,* 1999.

27. David T. Miller, Margaret P. Adam, Swaroop Aradhya, Leslie G. Biesecker, Arthur R. Brothman, Nigel P. Carter, Deanna M. Church, John A. Crolla, Evan E. Eichler, Charles J. Epstein, W Andrew Faucett, Lars Feuk, Jan M. Friedman, Ada Hamosh, Laird Jackson, Erin B Kaminsky, Klaas Kok, Ian D. Krantz, Robert M. Kuhn, Charles Lee, James M. Ostell, Carla Rosenberg, Stephen W. Scherer, Nancy B. Spinner, Dimitri J. Stavropoulos, James H. Tepperberg, Erik C. Thorland, Joris R. Vermeesch, Darrel J. Waggoner, Michael S. Watson, Christa Lese Martin, David H. Ledbetter. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet,* 2010.

28. Sarah T. South, Charles Lee, Allen N. Lamb, Anne W. Higgins, Hutton M. Kearney, Working Group for the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med,* 2013.

29. C. Frank Bennette, Adrian R. Krainer and Don W. Cleveland. Antisense Oligonucleotide Therapies for Neurodegenerative Diseases. *Ann Rev Neurosci,* 2019.

30. C. Frank Bennett, Eric E. Swayze. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu Rev Pharmacol Toxicol,* 2010.

31. Stanley T. Crooke, Joseph L. Witztum, C. Frank Bennett, and Brenda F. Baker. RNA-Targeted Therapeutics. *Cell Metab,* 2018.

32. Jinkuk Kim, Chunguang Hu, Christelle Moufawad El Achkar, Lauren E. Black, Julie Douville, Austin Larson, Mary K. Pendergast, Sara F. Goldkind, Eunjung A. Lee, Ashley Kuniholm, Aubrie Soucy, Jai Vaze, Nandkishore R. Belur, Kristina Fredriksen, Iva Stojkovska, Alla Tsytsykova, Myriam Armant, Renata L. DiDonato, Jaejoon Choi, Laura Cornelissen, Luis M. Pereira, Erika F. Augustine, Casie A. Genetti, Kira Dies, Brenda Barton, Lucinda Williams, Benjamin D. Goodlett, Bobbie L. Riley, Amy Pasternak, Emily R. Berry, Kelly A. Pflock, Stephen Chu, Chantal Reed, Kimberly Tyndall, Pankaj B. Agrawal, Alan H. Beggs, P. Ellen Grant, David K. Urion, Richard O. Snyder, Susan E. Waisbren, Annapurna Poduri, Peter J. Park, Al Patterson, Alessandra Biffi, Joseph R. Mazzulli, Olaf Bodamer, Charles B. Berde, Timothy W. Yu. Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *N Engl J Med,* 2019.