# Diagnosing Rare Disease: Next Generation Sequencing and Functional Studies of an Ultra Rare Neurodevelopmental and Neurodegenerative Disorder

Alexa Derksen
Integrated Program in Neuroscience
McGill University, Montreal
August 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of a Master of Science

# Contents

**Abstract**

Genetically determined leukoencephalopathies (gLEs) are a group of rare heritable white matter disorders primarily affecting children, for which there remains many unsolved genetic cases. Identifying the molecular cause of these neurodegenerative diseases is essential as it allows the patients and their families to know the disease course, gain access to therapeutic options if available and obtain genetic counselling to make reproductive decisions.

The first section of this thesis focuses on identifying the genetic cause of a cohort of patients with unsolved gLEs using whole exome sequencing (WES). The analysis was performed on 18 patients, 12 of which were second or third analyses, in accordance with the American College of Medical Genetics and Genomics (ACMG) Guidelines. We found a *de novo* pathogenic variant in *DYNC1H1* in two twin siblings, a strong candidate gene *ABHD16A* in 4 cases from two families, and variants of unknown significance in other genes in another 5 cases. We thereby report a solved rate of approximately 27% (5/18), wherein three of the second analyses, and both third analyses have strong candidate genes. Our solved rate greatly exceeds the anticipated 5-11% rate published in literature for other re-analyses cohorts.

In the second part of this thesis, we explore the functional defects caused by biallelic variants identified in the *LSM7* (NM_016199) gene, which encodes an RNA-binding protein that is part of two complexes playing a role in either RNA splicing or mRNA decay. Following a clinical WES, a homozygous variant in *LSM7* at position c.121G>A; p.Asp41Asn was identified in a patient (GB114.0) with a leukodystrophy. Through GeneMatcher, another individual with a different homozygous variant in *LSM7* at position c.206G>C; p.Arg69Pro was found. Structural data predicted that the mutant residues would result in the loss of key intra- and inter-molecular interactions. Affinity purification of wildtype and disease-associated *LSM7* variants from human

cells confirmed that both variants lead to a defect in assembly of both LSM complexes. Molecular investigations of the p.Asp41Asn variant showed decreased levels of *LSM7* mRNA and protein compared to healthy controls. We also found that knockout of *Lsm7* in zebrafish embryos led to a defect in nervous system development, supporting a role for *LSM7* mutations in neurodegeneration.

This thesis demonstrated that WES is an invaluable tool to identify the molecular causes of rare genetic diseases and that re-analysis of existing WES data should be considered before WGS and RNA sequencing. We describe a novel ultra rare neurodevelopmental and neurodegenerative disorder caused by biallelic pathogenic variants in *LSM7* and shed light on the disease pathogenesis. We are hopeful that our work will provide the foundation for future investigations into this rare human disease and lead to the development of potential therapies.

**Resumé**

Les leucoencéphalopathies génétiquement déterminées (gLEs) sont un groupe de maladies héréditaires rares affectant principalement les enfants et pour lesquelles plusieurs cas demeurent génétiquement non résolus. Identifier la cause moléculaire de ces maladies neurodégénératives est essentiel puisque cela permet aux patients ainsi qu'à leur famille de connaître l'évolution de la maladie, d'avoir accès à des options thérapeutiques si elles sont disponibles.

La première section de cette thèse porte sur l'identification de la cause génétique d'une cohorte de patients avec des gLEs non résolues en utilisant le séquençage d'exome entier (WES). L'analyse a été effectuée sur 18 patients, parmi lesquels 12 étaient des secondes ou des troisièmes analyses, en accord avec les lignes directrices du American College of Medical Genetics and Genomics (ACMG). Nous avons trouvé une mutation pathogénique *de novo* à l'intérieur du gène *DYNC1H1* chez deux jumelles, un gène candidat fort, *ABHD16A*, dans quatre cas chez deux familles, ainsi que des variantes génétiques de signification inconnue à l'intérieur d'autres gènes pour cinq autres cas. Nous rapportons donc un taux de résolution d'environ 27% (5/18), parmi lequel trois des secondes analyses et les deux troisièmes analyses ont des gènes candidats forts. Notre taux de résolution excède grandement le 5-11% anticipé par les publications dans la littérature pour d'autres cohortes de réanalyses.

Dans la seconde partie de cette thèse, nous explorons les défauts fonctionnels causés par des variantes génétiques biallélіques identifiées dans le gène *LSM7* (NM_016199), qui code pour une protéine de liaison à l'ARN faisant partie de deux complexes qui jouent un rôle soit dans l'épissage de l'ARN, soit dans la dégradation de l'ARN messager. À la suite d'un WES clinique, une mutation homozygote dans *LSM7* à la position c.121G>A; p.Asp41 Asn a été identifiée chez un patient (GB114.0) avec une leucodystrophie. À l'aide de GeneMatcher, un autre individu avec

une mutation homozygote différente dans *LSM7* à la position c.206G>C; p.Arg69Pro a été trouvé. Les données structurelles ont prédit que les résidus mutants engendreraient la perte d'interactions intra et intermoléculaires clés. Les purifications par affinité des LSM7 du type sauvage ainsi que de la forme mutée causant la maladie provenant de cellules humaines ont confirmé que les deux mutations mènent à un défaut d'assemblage des deux complexes LSM. Les investigations moléculaires des mutations p.Asp41Asn ont montré des niveaux diminués d'ARN messager ainsi que de protéine LSM7 en comparaison avec des contrôles en santé. Nous avons également trouvé que le knock-out de *Lsm7* chez l'embryon du poisson-zèbre menait à un défaut dans le développement du système nerveux central, appuyant un rôle pour des mutations du gène *LSM7* dans la neurodégénérescence.

Cette thèse a démontré que WES est un outil inestimable pour identifier les causes moléculaires des maladies génétiques rares et que la réanalyse des WES existants devrait être considérée avant WGS et le séquençage de l'ARN. Nous décrivons une nouvelle maladie neurodéveloppementale et neurodégénérative ultra rare causée par des mutations pathogéniques bialléliques dans le gène *LSM7* et mettons en lumière la pathogénèse de la maladie. Nous espérons que notre travail fournira la fondation nécessaire pour de futures investigations par rapport à cette maladie humaine rare et mènera au développement de thérapies potentielles.

**Acknowledgements**

First and foremost, I'd like to thank Dr. Geneviève Bernard for being an outstanding supervisor, mentor and role model. Thank you for believing in me, pushing me to accomplish my goals, and supporting my research endeavors. I would also like to thank all the members of the MyeliNeuroGene Lab for their help and continued support throughout my master's degree. To myfellow students, Stefanie Perrier, Travis Moore, Alexandra Chapleau and Mackenzie Michell-Robinson: thank you for putting up with my shenanigans and keeping me clearheaded throughout the past two years. A special shout out to Aaron Spahr who began this crazy journey with me 2 years ago, the classes we took together would not have been the same without you. I would also like to thank Lama Darbelli for helping me design and troubleshoot many of my experiments while also lending an attentive ear for my ideas and stories. To Luan Tran for commuting with me to the lab everyday, being the best neighbour and for raising my goddogger, Copper. Finally, I would like to thank Lei Cao for attentively reviewing and editing my thesis and Marie-Lou St-Jean for translating my abstract.

I would also like to thank Dr. Benoit Coulombe for welcoming me into his lab and providing me with guidance throughout my Master's. Thank you to the members of the Coulombe lab: Marie-Soleil Gauthier, Maxine Pinard, Phillipe Cloutier, and Esen Sokullu. Thank you particularly to Diane Forget for patiently training me, answering my many questions, and for helping me troubleshoot my experiments. I would also like to recognize Christian Poitras for his assistance with data analysis and figure design.

Finally, I would like to thank my parents, Beverly and Robert, and my sister Robyn for believing in me. Thank you for being my support system and always being there for me.

**Contributions of Authors**

**Chapter 1: Literature Review**

A.D. wrote this section in collaboration with Dr. Geneviève Bernard.

**Chapter 2: Rationale and Hypothesis**

A.D. wrote this section in collaboration with Dr. Geneviève Bernard.

**Chapter 3: Project 1 – Determining the genetic causes of molecularly unsolved cases of genetically determined leukoencephalopathies**

A.D. wrote this section in collaboration with Dr. Geneviève Bernard. WES analysis, experiments, and interpretations were performed by A.D. Dr. Geneviève Bernard contributed to interpretations of WES analysis.

**Chapter 4: Manuscript - A novel *de novo* variant in *DYNC1H1* causes spinal muscular atrophy lower extremity predominant in identical twins with discordant phenotypes**

A.D. wrote this section in collaboration with Dr. Amytice Mirchi and Dr. Geneviève Bernard.

A.D. designed and analysed experiments and A.M. conducted a review of the clinical data.

**Chapter 5: Project 2 – Bi-allelic pathogenic variants in *LSM7* cause a novel ultra-rare neurodevelopmental and neurodegenerative disorder**

A.D. wrote this section in collaboration with Dr. Geneviève Bernard. RT-qPCR experiments and analysis were performed by A.D. with guidance from Lama Darbelli and Dr. Geneviève Bernard.

**Chapter 6: Manuscript - Biallelic mutations in *LSM7* impair assembly of LSM complexes and lead to a rare neurodevelopmental and neurodegenerative disorder**

A.D. wrote this section in collaboration with Dr. Geneviève Bernard and Dr. Benoit Coulombe.

A.D. designed and analysed experiments with guidance from Lama Darbelli, Diane Forget and

Christian Poitras. The zebrafish animal model was created in the lab of Dr. Joshua Bonkowsky by Post doctoral fellow Hung-Yu Shih at the University of Utah.

**Chapter 7: Discussion and Conclusion**

A.D. wrote this section in collaboration with Dr. Geneviève Bernard.

**List of Figures**

Figure 6.6 Zebrafish *lsm7* is conserved to human and expressed in developing nervous system, and is necessary for embryonic morphology, eye size, motor behaviour, and survival of oligodendrocytes

**List of Tables**

**List of Abbreviations**

| | |
|---|---|
| 3'UTR | 3 prime untranslated region |
| 4H | Hypomyelination with hypogonadotropic hypogonadism and hypodontia |
| 5S RNA | 5S ribosomal ribonucleic acid |
| 7SK RNA | 7SK small nuclear ribonucleic acid |
| *AARS1* | Alanyl transfer ribonucleic acid synthetase 1 |
| *ABCD1* | ATP binding cassette subfamily D member 1 |
| ACMG | American college of medical genetics |
| AD | Autosomal dominant |
| *ADAR* | Adenosine deaminase, RNA-specific gene |
| ADLD | Adult onset autosomal dominant leukodystrophy |
| AGS | Aicardi-Goutières syndrome |
| ANNOVAR | Annotate variation |
| AR | Autosomal recessive |
| *ARSA* | Arylsulfatase A gene |
| ATP | Adenosine triphosphate |
| *ATXN2* | Ataxin 2 gene |
| *Brr2* | Pre-mRNA splicing helicase gene |
| *CNOT6* | CCR4-NOT transcription complex subunit 6 gene |
| CNS | Central nervous system |
| *DARS1* | Aspartyl transfer ribonucleic acid synthetase 1 gene |
| *DARS2* | Aspartyl transfer ribonucleic acid synthetase 2 gene |
| *Dcp1* | Messenger ribonucleic acid decapping enzyme subunit 1 gene |
| *DCP1A* | Decapping enzyme 1*, Saccharomyces cerevisiae*, homolog of A gene |
| *DCP1B* | Decapping enzyme 1, *Saccharomyces cerevisiae*, homolog of B gene |
| *Dcp2* | mRNA-decapping enzyme subunit 2 gene |
| *DDX6* | DEAD (Asp-Glu-Ala-Asp) box helicase 6 gene |
| *Dhh1* | Adenosine triphosphate-dependent ribonucleic acid helicase gene |
| *DHX16* | DEAH (Asp-Glu-Ala-His) box helicase 16 gene |
| DNA | Deoxyribonucleic acid |
| *EARS2* | Glutamyl transfer ribonucleic acid synthetase 2 gene |
| *Edc4* | Enhancer of messenger ribonucleic acid decapping 4 gene |
| *EFTUD2* | Elongation factor Tu guanosine triphosphate-binding domain containing 2 gene |
| EIF2B | Eukaryotic translation initiation factor 2B |
| *EIF2B1* | Eukaryotic translation initiation factor 2B subunit 1 gene |
| *EIF2B2* | Eukaryotic translation initiation factor 2B subunit 2 gene |
| *EIF2B3* | Eukaryotic translation initiation factor 2B subunit 3 gene |
| *EIF2B4* | Eukaryotic translation initiation factor 2B subunit 4 gene |
| *EIF2B5* | Eukaryotic translation initiation factor 2B subunit 5 gene |
| *EPRS1* | Glutamyl-prolyl transfer ribonucleic acid synthetase 1 gene |
| *GALC* | Galactosylceramidase gene |
| GATK | Genome analysis toolkit |
| *GJC2* | Gap junction protein gamma 2 gene |
| gLE | Genetically determined leukoencephalopathy |

| | |
|---|---|
| HBSL | Hypomyelination with brainstem and spinal cord involvement and leg spasticity |
| HLD | Hypomyelinating leukodystrophy |
| LBSL | Leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation |
| LD | Leukodystrophy |
| *LMNB1* | Lamin B1 gene |
| *LSM* | Like-Sm gene |
| *LSM7* | Like-Sm 7 gene |
| LTBL | Leukoencephalopathy with thalamus and brain stem involvement and high lactacte |
| MAF | Minor allele frequency |
| *MFA2pG* | Mating Factor A with poly G sequence |
| MLC | Megaencephalic leukoencephalopathy with subcortical cysts |
| *MLC1* | Modulator of volume-regulated anion channel current 1 gene |
| MLD | Metachromatic leukodystrophy |
| MRI | Magnetic resonance imaging |
| mRNA | Messenger ribonucleic acid |
| NCBI | National centre for biotechnology information |
| NGS | Next generation sequencing |
| Non-HLD | Non-hypomyelinating leukodystrophy |
| OMIM | Online mendelian inheritance in man |
| *PABP* | Polyadenylate binding protein gene |
| *PAN2* | Polyadenylate binding protein-dependent poly(A) nuclease 2 gene |
| *PAN3* | Polyadenylate binding protein-dependent poly(A) nuclease 3 gene |
| *Pat1b* | Protein associated with topoisomerase II gene |
| *PATL1* | Protein associated with topoisomerase 1 homology 1, processing body mRNA decay factor gene |
| PCR | Polymerase chain reaction |
| *PLP1* | Proteolipid protein 1 gene |
| PMD | Pelizaeus-Merzbacher disease |
| PMLD1 | Pelizaeus-Merzbache-like disease |
| PNS | Peripheral nervous system |
| *POLD* | Pigmentary orthochromatic leukodystrophy gene |
| *POLR1C* | Polymerase I RNA subunit C gene |
| POLR3 | RNA polymerase III |
| *POLR3A* | Polymerase III RNA subunit A gene |
| *POLR3B* | Polymerase III RNA subunit B gene |
| *POLR3K* | Polymerase III RNA subunit K gene |
| *Prp2* | Precursor messenger ribonucleic acid splicing factor adenosine triphosphate-dependent ribonucleic acid helicase-like protein gene |
| *Prp24* | U4/U6 small nuclear ribonucleic acid associated splicing factor gene |
| *Prp3* | U4/U6 small nuclear ribonucleoprotein 3 gene |
| *Prp31* | Precursor messenger ribonucleic acid processing factor 31 gene |
| *Prp4* | U4/U6 small nuclear ribonucleoprotein 4 gene |
| *Prp6* | Precursor messenger ribonucleic acid splicing factor 6 gene |

| | |
|---|---|
| *Prp8* | Precursor messenger ribonucleic acid splicing factor 8 gene |
| *PRPF3* | Precursor messenger ribonucleic acid processing factor 3 gene |
| *PRPF31* | Precursor messenger ribonucleic acid processing factor 31 gene |
| *PRPF4* | Precursor messenger ribonucleic acid processing factor 4 gene |
| *PRPF6* | Precursor messenger ribonucleic acid processing factor 6 gene |
| *PRPF8* | Precursor messenger ribonucleic acid processing factor 8 gene |
| *PSAP* | Prosaposin gene |
| PVL | Periventricular leukomalacia |
| *RARS1* | Arginyl transfer ribonucleic acid synthetase 1 gene |
| RNA | Ribonucleic acid |
| *RNASEH2A* | Ribonuclease H2 subunit A gene |
| *RNASEH2B* | Ribonuclease H2 subunit B gene |
| *RNASEH2C* | Ribonuclease H2 subunit C gene |
| *SAMHD1* | Sterile alpha motif domain and HD (His-Asp) domain-containing protein 1 gene |
| *SART3* | Squamous cell carcinoma antigen recognized by T cells 3 gene |
| SCA-2 | Spinocerebellar ataxia type 2 |
| *Sen1* | Splicing endonuclease 1 gene |
| sgRNA | Single guide ribonucleic acid |
| SIFT | Scale-invariant feature transform |
| *SLBP* | Stem-loop binding protein gene |
| *SLC17A5* | Solute carrier family 17 (acidic sugar transporter) member 5 gene |
| SMA | Spinal muscular atrophy |
| *SMN* | Survival motor neuron gene |
| *SMN1* | Survival motor neuron 1 gene |
| snRNA | Small nuclear ribonucleic acid |
| snRNP | Small ribonucleoprotein |
| *SNRNP200* | Small nuclear ribonucleoprotein U5 subunit 200 gene |
| *Tpt1* | Transfer ribonucleic acid 2'-phosphotransferase gene |
| *TREX1* | 3-prime repair exonuclease 1 gene |
| tri-snRNP | Triple small nuclear ribonucleoprotein complex |
| tRNA | Transfer ribonucleic acid |
| UCSC | University of California Santa Cruz |
| WES | Whole exome sequencing |
| WT | Wild type |
| X-ALD | X-linked adrenoleukodystrophy |
| *XRN1* | 5'-3' exoribonuclease 1 gene |
| *Xrn1* | 5'-3' exoribonuclease 1 gene |

## Introduction

Leukodystrophies (LD) are a heterogeneous group of genetically determined rare neurodegenerative disorders characterized by abnormal white matter on brain imaging[3]. They can be broadly divided into hypomyelinating (HLD) or non-hypomyelinating (non-HLD) based on magnetic resonance imaging (MRI) characteristics and whether the pathophysiological problem is abnormal myelin deposition during development or abnormal myelin homeostasis, respectively[4,5]. These diseases collectively affect as many as 1 in 7,663 individuals with a broad range of onset spanning from prenatal life to senescence[6-8].

Leukodystrophy patients are typically previously healthy children presenting with developmental regression[8]. As the disease progresses, their disabilities increase, and can include, reduced mobility, impaired communication, swallowing difficulties and cognitive and behavioural difficulties, amongst others[8]. Patients can also suffer from other systemic health-related problems including: musculoskeletal and skin issues, nutritional and gastrointestinal complications, respiratory difficulties, sleep disturbances, endocrine complications, etc[8]. Unfortunately, due to the rarity and heterogeneity of these disorders, many patients and their families may not receive a definitive diagnosis[9]. Devastatingly, one third of children with a leukodystrophy will succumb to their underlying disease and its complications by the age of eight[6]. Currently, there are no known curative treatments for most leukodystrophies, and as such treatments tend to focus on alleviating the symptoms of the patients[8,10,11].

This thesis is divided into two parts, the first focuses on uncovering the molecular causes in a cohort of unsolved genetically determined leukoencephalopathies patients. In chapter 3, using whole exome sequencing (WES), and with the help of our international collaborators, and technological genetic platforms such as GeneMatcher, we aimed to identify the genetic cause

of their disease and, ideally to administer more personalized care and treatments for these patients and their families[12,13]. This section is followed by chapter 4, containing the manuscript titled: "A novel *de novo* variant in *DYNC1H1* causes spinal muscular atrophy lower extremity predominant in identical twins with discordant phenotypes", which details one of our solved cases.

The second part of this thesis centres on the study of two novel biallelic variants in *LSM7* (OMIM #607287) which were uncovered in two unrelated individuals with an ultra-rare neurodevelopmental disorder. *LSM7* has not previously been implicated in any human disease. However, its function and role make it an extremely interesting candidate for study. As such, functional testing was conducted to better understand the role of the native and mutant LSM7 proteins. Most of these results are presented and discussed (chapter 5) in the manuscript titled: "Biallelic mutations in *LSM7* impair assembly of LSM complexes and lead to a rare neurodevelopmental and neurodegenerative disorder". This thesis concludes with a chapter containing a general discussion and presents directions for future experiments to study *LSM7*.

# Chapter 1: Literature Review

## 1.0 Genetically Determined Leukoencephalopathies

### 1.0.1 Preface

The first section of this chapter provides a literature review of genetically determined leukoencephalopathies (gLEs) and leukodystrophies (LDs).

### 1.0.2 Genetically Determined Leukoencephalopathies Classification

Genetically determined leukoencephalopathies (gLEs) are genetic disorders with either primary glial, neuronal, vascular or systemic involvement which result in primary or secondary white matter abnormalities[3,14,15]. Leukodystrophies (LDs) are a type of genetically determined leukoencephalopathy and considered primarily disorders of the white matter of the central nervous system (CNS) with or without peripheral nervous system (PNS) involvement[3,14-16]. LDs and gLEs do not include acquired CNS myelin disorders, such as multiple sclerosis or any related acquired demyelinating processes[3,6,14-16].

The term leukodystrophy was first used in 1928 to describe metachromatic leukodystrophy (MLD), a hereditary progressive disease with white matter degeneration[7,9,15]. However, it wasn't until the 1980s that the first genes associated with leukodystrophies were identified, including that of *PLP1* (OMIM #300401) for Pelizaeus-Merzbacher Disease (PMD)[9]. PLP1, or proteolipid protein is the primary constituent of myelin in the CNS, making PMD a perfect fit for the original leukodystrophy definition which was extremely myelin focused[5,7,9]. The more recent discovery of disease-causing genes involved in the housekeeping processes of mRNA transcription and translation have required adaptations to the original LD definition to be made[17-19]. The refined definition for LDs includes disorders of the white matter of the CNS that have primarily glial cell or myelin sheath abnormalities, the neuropathology of which can be characterized not only by the

involvement of oligodendrocytes (the myelin producing cells of the CNS), but also of astrocytes, microglia and other non-neural cell types[7]. LDs are thus an extremely heterogeneous group of disorders, and include neonatal, childhood, adolescent and even adult presentations[7,14,15].

So far, over 30 conditions have be categorized as primary LDs while numerous others are considered heritable conditions and fall in the broader category of gLEs[3,14,15]. The combined incidence of these heterogenous LDs ranges broadly. A study published following national health reporting systems in Germany indicated an incidence of approximately 1 in 50,000 live births[20]. More recently, statewide data in Utah, USA reported an overall incidence of 1 in 7,663. The large disparity in prevalence reported in these studies indicates that additional research is required to establish incidences outside of these two populations. Moreover, these heterogeneous genetic diseases can have possible inheritance patterns of autosomal recessive, autosomal dominant or *de novo* dominant, X-linked and mitochondrial further complicating their genetic diagnoses[6,14,15]. All LDs share white matter abnormalities on imaging or upon pathology of the CNS with the majority having motor deficits which dominate the clinical presentation[4,14,16]. Some motor phenotypes include spastic paraparesis or quadriparesis, dystonia, and ataxia[5-7,19]. There are also a variety of non-neurological findings that can be associated with specific LDs and gLEs which may be helpful in the determination of a molecular diagnosis. Examples of such manifestations include adrenal insufficiency or other endocrine disturbances, ophthalmologic abnormalities, hypodontia or oligodontia, dysmorphic facial features, hearing impairment, skeletal abnormalities, hepatosplenomegaly and gastrointestinal symptoms[14,19].

While symptoms may be useful in the classification of LDs and gLEs, LDs are most commonly first diagnosed based on specific patterns on brain magnetic resonance imaging (MRI)[4,16]. As myelination of the brain takes place during the first 2 years of life, a single MRI

within the first year of life is not sufficient to distinguish between delayed myelination, hypomyelination, or the early stages of a LD[4]. Furthermore, since delayed myelination is a nonspecific feature observed in children with delayed development, two MRIs must be taken with a significant time interval between for a diagnosis of a LD[4,14]. There are 3 major MRI characteristics that are used to help discriminate between the different types of LDs and gLEs[14]. The first is hypomyelination, which is characterized by an unchanged pattern of deficient myelination on 2 MRI scans at least 6 months apart in a child 1 year or older[14]. Second, white matter abnormalities can generally be classified as confluent which normally implies a gLE or LD or as multifocal which implies an acquired disorder[4,14-16]. Lastly, several LDs and gLEs can be distinguished by the specific location of the abnormalities on brain MRI[14]. For these reasons, MRIs are used as the primary diagnostic tool for LDs and may expose features characteristic of a more specific diagnosis[4,14,21].

### 1.0.3 Leukodystrophy Classification

Leukodystrophies are a subgroup of genetically determined leukoencephalopathies where the primary abnormality involves myelin producing cells[4,5]. LDs can be further classified as either hypomyelinating (HLD) or non-hypomyelinating (non-HLD) based on MRI characteristics[4,5]. The pathophysiological problem in HLDs is that of abnormal myelin deposition during development while in non-HLDs the issue is abnormal myelin homeostasis[4,5]. In HLDs, T2-weighted MRI images show mild hyperintensity of white matter compared to gray matter while T1-weighted images can be hyperintense, isointense or mildly hypointense compared to grey matter structures[3,4,15,16,19]. In contrast, non-HLDs have prominent T2 hyperintensity and prominent T1 hypointensity relative to gray matter structures on MRI[3,4,15,16].

### 1.0.4 Hypomyelinating Leukodystrophies

A wide spectrum of genes ranging from those which encode myelin specific proteins to those that are involved in housekeeping processes such as transcription and translation have been found to cause HLDs and non-HLDs (Table 1.1)[5,7,9,17-19,21-23]. One of the most common HLDs is caused by mutations in genes encoding subunits of the RNA polymerase III (POLR3) and is named 4H (Hypomyelination, Hypodontia, and Hypogonadotropic Hypogonadism) or POLR3-related leukodystrophy[22]. POLR3 is responsible for transcribing noncoding RNAs including ribosomal 5S RNA, tRNAs, U6 snRNA, and 7SK RNA[17]. Thus far, mutations in POLR3 components *POLR3A, POLR3B, POLR1C*, and *POLR3K* have all been found to cause HLD[17,19,21]. More recently, it has also been shown that mutations in genes encoding tRNA synthetases such as *AARS1* and *EPRS1*, encoding alanyl-tRNA synthetase and glutamyl-prolyl-tRNA synthetase, respectively also cause HLDs[18,23]. Although these genes are widely accepted to cause leukodystrophies, the mechanisms through which they cause this CNS specific disease remains poorly understood. Currently, the most widely accepted hypothesis is that mutations in the POLR3 subunits result in a decrease in the levels of tRNAs which are critical for the synthesis of essential CNS proteins[18,24,25]. Furthermore, alterations in the tRNA-synthetase pathway could lead to abnormal translation and result in insufficient production of specific proteins at critical developmental stages[18,26].

### 1.0.5 Non-Hypomyelinating Leukodystrophies

Non-HLDs are a group of leukodystrophies with heterogeneous clinical and MRI presentation, and are generally defined as having abnormal myelin homeostasis[3,4]. This group consists of demyelinating (*i.e.;* loss of previously deposited myelin), dysmyelinating (i.e. deposition of structurally or biochemically abnormal myelin) and myelinolytic diseases (i.e.

myelin vacuolization)[7]. Some of the more common non-HLDs include metachromatic leukodystrophy (MLD), Krabbe disease, and X-linked adrenoleukodystrophy (X-ALD)[7,14-16](Table 1.1). MLD is an autosomal recessive disorder caused by biallelic pathogenic variants in *ARSA* which encodes the lysosomal enzyme arylsulfatase A. Reduced activity of this enzyme results in the accumulation of sulfatides in both the central and peripheral nervous systems and result in subsequent demyelination[7,10,27-29]. Krabbe disease is a progressive lysosomal storage disorder that results from biallelic pathogenic variants in *GALC* which encodes the enzyme involved in the catabolism of galactosylceramide, a major lipid constituent of myelin[14,16,30]. Finally, X-ALD caused by hemizygous loss of function mutations in *ABCD1*, results in defects in peroxisomal beta oxidation and subsequently the accumulation of saturated very long chain fatty acids[31,32]. Additionally, like HLDs, several non-HLDs are also caused by defects in genes involved in housekeeping processes. One example is eIF2B-related disorder (Vanishing white matter disease OMIM #603896) which is caused by biallelic pathogenic variants in genes encoding subunits of the eukaryotic translation initiation factor eIF2B[33-35]. Furthermore, a subgroup of non-HLDs called mitochondrial leukoencephalopathies are caused by mutations in genes encoding mitochondrial proteins. Examples include leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL) as well as leukoencephalopathy with thalamus and brainstem involvement and high lactate (LTBL) which are caused by biallelic pathogenic variants in the genes encoding the mitochondrial aminoacyl-tRNA synthetases for aspartic acid (*DARS2*) and glutamic acid (*EARS2*), respectively[36-38]. In summary, there are a wide variety of genetic causes for non-HLDs, HLDs and gLEs, many of which have yet to be discovered.

## Table 1.1 List of leukodystrophies discussed in this thesis

| Disease Name | OMIM | Inheritance | Gene(s) | Classification |
|---|---|---|---|---|
| **Hypomyelinating Leukodystrophies** | | | | |
| *EPRS1*-related leukodystrophy[18] | 617951 | AR | *EPRS1* | HLD |
| Hypomyelination with brainstem and spinal cord involvement and leg spasticity (HBSL)[15] | 615281 | AR | *DARS1* | HLD |
| Pelizaeus-Merzbacher disease (PMD)[14] | 312080 | X-linked | *PLP1* | HLD |
| Pelizaeus-Merzbacher-like disease (PMLD1)[14,39] | 608804 | AR | *GJC2* | HLD |
| POLR3-related leukodystrophy[21,24,40] | 607694, 614381, 616494. | AR | *POLR3A, POLR3B, POLR1C, POLR3K* | HLD |
| *RARS1*-related leukodystrophy[15] | 616140 | AR | *RARS1* | HLD |
| **Non-Hypomyelinating Leukodystrophies** | | | | |
| Aicardi-Goutières Syndrome (AGS)[14] | 225750, 610333, 610181, 610329, 612952. 615010 | Usually AR; may be AD | *TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR* | Non-HLD |
| eIF2B related disorder (Vanishing White Matter Disease)[33] | 169500 | AR | *EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5* | Non-HLD |
| Leukoencephalopathy with thalamus and brainstem involvement and high lactate (LTBL)[36] | 614924 | AR | *EARS2* | Non-HLD |
| Krabbe Disease (Globoid cell LD)[11,30] | 245200. 611722 | AR | *GALC, PSAP* | Non-HLD |
| Leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL)[38] | 611105 | AR | *DARS2* | Non-HLD |
| Metachromatic Leukodystrophy (MLD)[14] | 607574, 249900 | AR | *ARSA, PSAP* | Non-HLD |
| Megaencephalic leukoencephalopathy with subcortical cysts (MLC)[14] | 604004, 613925, 613926 | AR and AD | *MLC1, HEPACAM* | Non-HLD |
| X-linked Adrenoleukodystrophy (X-ALD)[14] | 300100 | X-linked | *ABCD1* | Non-HLD |

## 1.1 Next-Generation Sequencing

### 1.1.1 Preface

This section gives a brief history of the use of next generation sequencing (NGS) techniques in rare disease research while describing the strengths and limitations of whole exome sequencing analysis (WES).

### 1.1.2 Background

The total number of rare diseases is thought to be between 6,000-7,000, with approximately 72% being genetic in origin and 70% having pediatric onset[12,13,41-43]. Additional analyses have shown that the prevalence of rare diseases in the population ranges from 3.5-5.9% which is equivalent to 263-446 million persons worldwide[41]. Unfortunately, approximately half of the patients with a rare genetic diseases will never receive a genetic diagnosis, resulting in slow, costly, and invasive diagnostic odysseys[13]. Recently, it has been demonstrated that the time to genetic diagnosis in 25% of patients ranged from 5 to 30 years[13]. During that time, 40% of patients received an incorrect diagnosis and an estimated 30% received inappropriate care or underwent unnecessary procedures[13]. This highlights the need for a shortened diagnostic odyssey and calls for earlier confirmation of genetic diagnoses for patients with these rare genetic diseases[13].

Definitive genetic diagnoses allow for improvements in disease management, such as targeted treatment options, and personalized health care as well as allow for genetic counselling, knowledgeable risk assessments and prenatal diagnosis options for families[12,13]. Historically, if an individual was suspected of having a rare genetic disease, genetic testing would be performed on a gene-to-gene basis or through analysis of gene panels containing only a handful of candidate genes[12]. Although this process has proven to be successful in providing many individuals with a genetic diagnoses it can be extremely labor intensive and expansive to sequence genes one at a

time[13]. Regrettably, following the use of these traditional methods there still remained a large portion of patients whose cases remain unsolved[12,13].

### 1.1.3 Molecular Identification using Whole Exome Sequencing

The emergence of new and easily accessible technologies such as next generation sequencing (NGS) has resulted in improved rates of discovery of novel disease-causing genes as well as timely genetic diagnosis for genes already with disease associations[12]. Whole exome sequencing (WES), a type of NGS is a powerful approach in which about 1% of the human genome, also known as the coding portion or exome, consisting of approximately 20,000 genes and their exon-intron boundaries is sequenced and subsequently analyzed[13]. This method allows for the detection of genetic variations within the coding region of an individual and is particularly beneficial for those that, as a result of potential genetic heterogeneity and disease rarity, almost certainly would not have received a genetic diagnosis using traditional methods, such as gene panels[12,13]. Studies have shown that significant proportions of patients will have their treatments and therapies adjusted or initiated based on the genetic results from WES[13]. This demonstrates just how crucial it is to provide patients and their families with an appropriate molecular diagnosis to ensure that treatable conditions and windows of opportunity for treatments are not missed[12,13].

### 1.1.4 Whole Exome Sequencing Limitations

Unfortunately, approximately 20-30% of the patients who have undergone WES analysis remain without a molecular diagnosis[9,44]. Like any method, WES has its limitations, including its inability to detect certain types of mutations with high confidence[45]. WES cannot identify large rearrangements, copy number variations, mutations in repetitive regions and mutations in high GC content regions[45]. Furthermore, WES is unable to detect epigenetic factors or mosaic mutations which may lead to or cause disease[45]. The choice of tools and algorithms used to analyze WES is

another source of error and potential variation[46]. Different software packages differ in their abilities to call bases, and various algorithms exist for aligning reads, some which are more accurate for certain types of variations than others[46]. The coverage depth of a particular base can also be affected by the quantity and quality of DNA that was used for the sequencing[46]. Moreover, the computational abilities of the software used to perform the sequencing also has limitations in terms of its ability to read the DNA[46]. As such, current capture methods are not perfect but still cover upwards of 99% of the regions of interest[47,48]. For the cases of WES that remain unsolved it has been shown that through collaboration and the sharing of information that re-analyses performed 6-12 months later may uncover disease-causing variants[12]. Indeed, it has been found that the success rate of these WES re-analyses ranges from 5-11% in the literature[49-52]. However, for the cases that remain unsolved even following re-analysis, whole genome sequencing (WGS) combined with RNA sequencing is the proposed next step[53]. WGS not only has the advantage of potentially revealing variants in the exome that were poorly captured in WES, but also allows for the improved detection of structural variants and variants in noncoding regions which could have impacts on gene expression or alternative splicing[54,55]. Performing RNA sequencing in combination with WGS allows for the quantification of transcript diversity thereby providing a measurement of the consequences of coding and noncoding variants on gene expression levels and alternative splicing[54,55].

## 1.2 LSM Proteins

### 1.2.1 Preface

This section gives a brief background on Sm proteins and their relation to LSM proteins. The roles of LSM proteins and the functions of the two principal LSM complexes LSM1-7 and LSM2-8 are then described in more detail. Key interactors associated with the LSM complexes are also considered, as well as some of less well characterized functions of these complexes. Finally, this section ends with a short discussion of diseases linked indirectly to LSM proteins.

### 1.2.2 Sm Proteins

Sm proteins were first discovered as antigens targeted by Anti-Sm antibodies in a patient with systemic lupus erythematosus, commonly referred to as lupus[56]. There are 7 distinct Sm genes in eukaryotes SmB/B' (B' is a result of alternative splicing), SmD1, SmD2, SmD3, SmE, SmF and SmG which range in size from 8.5kDa to 24kDa[57-59]. These proteins come together to form a doughnut shaped heteroheptameric complex in the cytoplasm upon binding to the U-rich track of the U1, U2, U4 and U5 snRNAs[57,59]. These snRNAs are key components of the primary spliceosome that are transcribed by RNA polymerase II and subsequently exported into the cytoplasm for further modifications[57,59,60]. The interaction of the Sm complex with these snRNAs promotes their stability and the hypermethylation of their 5' cap[57]. This mediates their nuclear re-import and allows for the subsequent assembly of the major spliceosome in the nucleus[57,59]. Thereby, the Sm proteins play a critical role in the stabilization of these snRNAs which make up several of the key components of the major spliceosome[59].

### 1.2.3 LSM Proteins

LSM proteins were first uncovered due to their similarities with the family of Sm proteins and were named Sm-like (LSM)[56]. Both Sm and LSM proteins share important

characteristics including their primary structures, which consist of a conserved Sm-domain containing a Sm motif-I and Sm motif-II, separated by a non-conserved linker region of variable length[59,61-64] (Figure 1.1A). The secondary structures are also highly conserved and consist of an N-terminal alpha helix next to a 5 strongly bent antiparallel beta sheets, referred to collectively as the Sm-fold[59,61,64] (Figure 1.1B). The Sm motif-I includes beta sheets 1 through 3 while the Sm motif-II consists of beta sheets 4 and 5[57]. Furthermore, the tertiary structural beta barrel shape of the Sm and LSM proteins is preserved across the three kingdoms of life[57]. LSM proteins, similar to Sm proteins form heteroheptameric doughnut shaped complexes which function in RNA binding. The LSM proteins, however, participate in a wider array of RNA related functions including mRNA decay, pre-mRNA splicing, RNA processing, and antisense RNA-mediated regulation of gene expression[61]. Interestingly, it is thought that the *LSM* genes arose before Sm genes, as they have RNA processing functions which evolutionarily precede splicing[57].

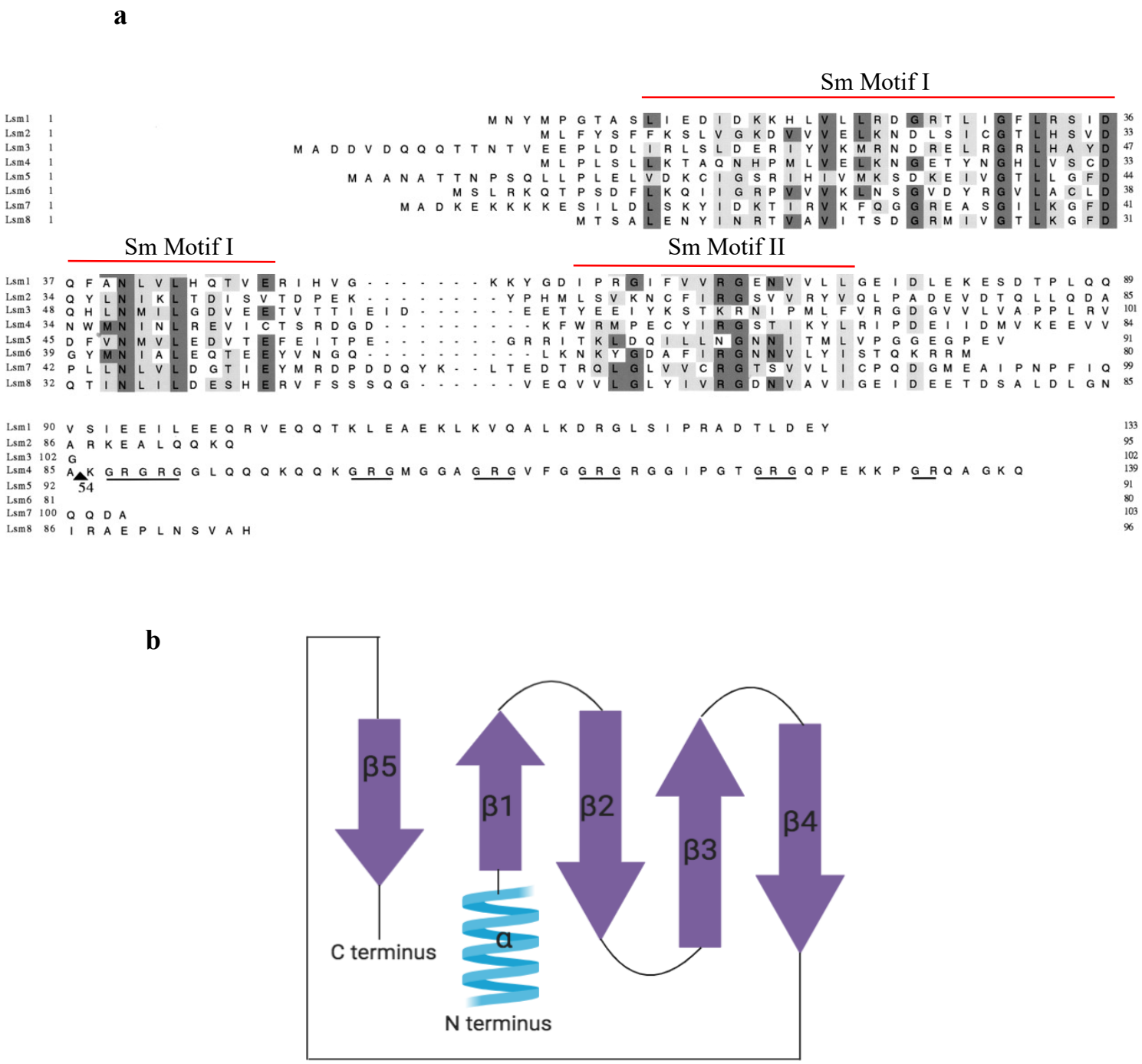



**Figure 1.1 Structure of LSM proteins. a,** Amino acid sequence of human LSM proteins (LSM1, 2, 3, 4, 5, 6, 7, and 8) with structural Sm-motifs indicated. **b,** Secondary structural motifs of LSM proteins. This figure (1.1A) was originally published in the Journal of Biological Chemistry Friesen, WJ and Dreyfuss, G. Specific Sequences of the Sm and Sm-like (Lsm) Proteins Mediate Their Interaction with Spinal Muscular Atrophy Disease Gene Product (SMN). J Biol Chem. 2000; 34:26370-5. © the American Society for Biochemistry and Molecular Biology

There are 8 main LSM proteins in eukaryotes, LSM1 through LSM8 which exist principally in two heteroheptameric ring shaped complexes *in vivo*, namely LSM1-7 and LSM2-8[61] (Figure 1.2). These quaternary structures are stabilized by the interactions between beta sheets 4 and 5 of adjacent monomers[61]. It is the loops connecting beta sheets 2 to 3 and 4 to 5 which face the inner side of the heteroheptameric ring and interact directly with RNA to carry out their principal functions[57]. Unlike the Sm complex, both of the LSM complexes, LSM1-7 and LSM2-8, are able to form spontaneously in the absence their associated RNAs[65]. The LSM1-7 complex, like the Sm complex is located in the cytoplasm. However, here it plays a role in 5' to 3' mRNA decay rather than in the stabilization of spliceosomal precursors[58,61,65]. The LSM2-8 complex, on the other hand, is found in the nucleus and plays a role in splicing, specifically in chaperoning and stabilizing U6 snRNA, the fifth and final U snRNA of the major spliceosome[59,61,66].

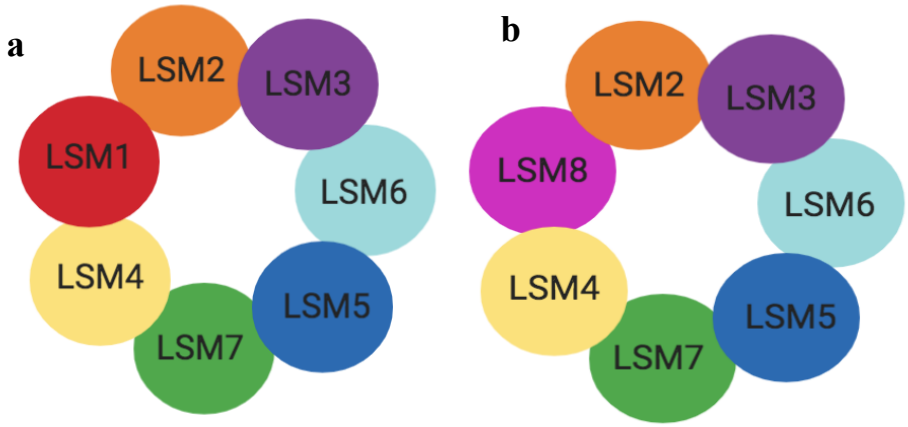



**Figure 1.2 Quaternary structure of LSM complexes. a,** LSM1-7 complex and **b,** LSM2-8 complex

### 1.2.4 LSM1-7 Complex

Cytoplasmic mRNA decay is a highly regulated process which is critical for eukaryotic cells[67]. The two main pathways through which mRNA is decayed in the cytoplasm are highly

conserved in eukaryotes and have been well studied in yeast[58]. The initial step in both pathways involves the shortening of the poly(A) tail of the mRNA[58]. Following this step, the mRNA is degraded in either a 3' to 5' direction, by the exosome complex, or is decapped and subsequently decayed in the 5' to 3' direction by the exoribonuclease XRN1 (Figure 1.3)[58,65,66]. Both pathways are highly selective and ensure that polyadenylated mRNAs are not degraded but rather available for translation[1,67,68]. Interestingly, mammalian mRNA is made with a much longer Poly(A) tail consisting of about 200 residues compared to the only 70-80 residues in yeast and deadenylation appears to be biphasic[58]. The first phase is mediated by the PABP-dependent poly(A) nucleases 2 and 3 (PAN2 and PAN3) and involves the shortening of the poly(A) tail to an intermediate length. In the second phase the poly(A) tail is further deadenylated to an oligoadenylated form by the CCR4-NOT transcription complex, subunit 6 (CNOT6)[58]. This oligoadenylated mRNA can then be degraded through either the 5' to 3' or 3' to 5' decay pathway[58].

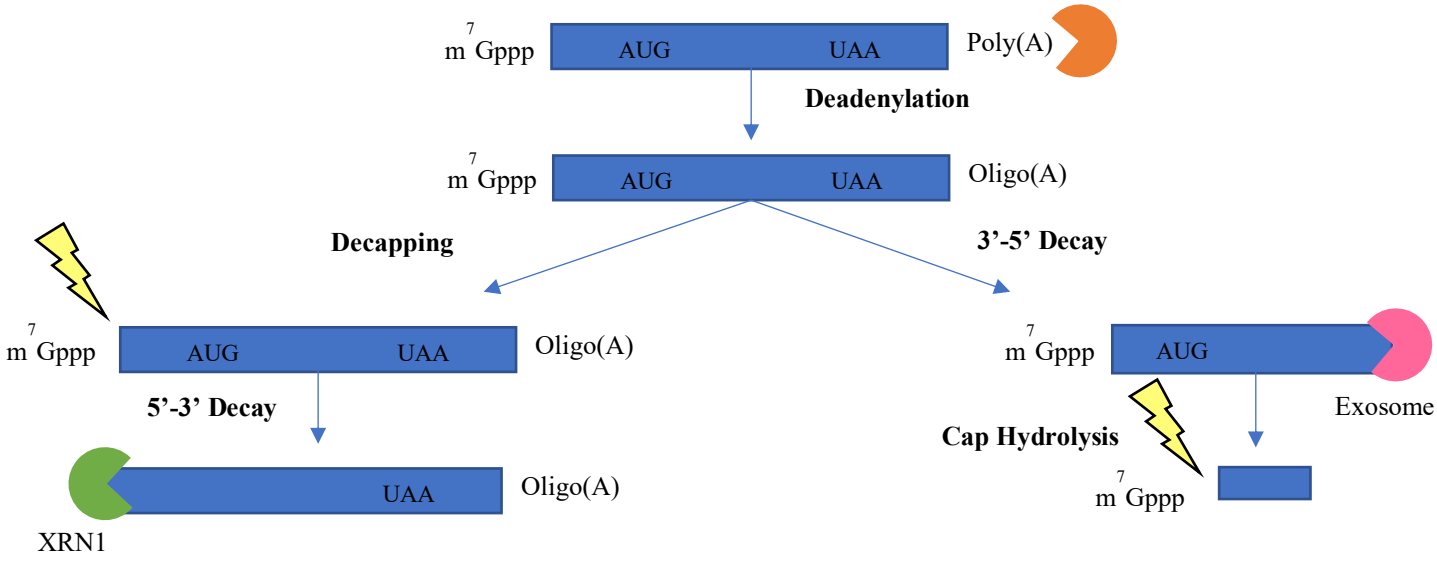



**Figure 1.3 Cytoplasmic mRNA decay in the 5' to 3' and 3' to 5' directions in yeast.** Permission to reprint is granted by Copyright clearance Center's RightLink® service.

The LSM1-7 complex plays a role in the 5' to 3' decay of mRNA, and has a strong binding preference for oligoadenylated mRNA or unadenylated RNAs containing U tracts at their 3' ends[58]. These LSM1-7 associated oligoadenylated or unadenylated mRNAs are then decapped and subsequently degraded in a 5' to 3' exonucleolytic manner[58] (Figure 1.4). While translation initiation machinery and the poly(A) binding protein (PABP) are known to be inhibitory to decapping, several other factors including the Processing Body mRNA Decay Factor (PATL1) and LSM1-7 complex have been found to facilitate it[58,69]. Previous studies in yeast have concluded that mRNA decay in cells with mutations in Lsm1p through Lsm7p are blocked at decapping stage[70,71]. This leads to the accumulation of a capped oligoadenylated form of mRNA whereas the wild type cells have a more uniform distribution of the A-tail length at steady state[58,71,72]. In general, mRNA decay in the cytoplasm is a highly coordinated process as it is a critical step in the regulation of gene expression and determining the transcriptome of eukaryotic cells[73]. It is well known that loss of mRNA stability regulation can result in a variety of diseases depending on the specific mRNAs affected, supporting the role of LSM proteins as good candidates for disease[58].

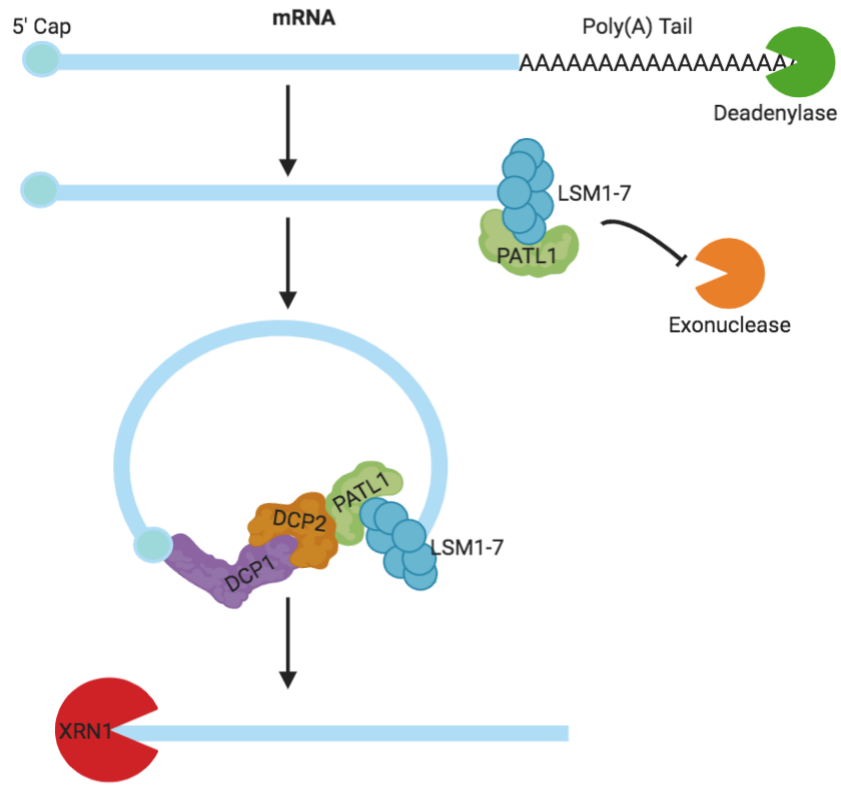



**Figure 1.4 LSM deadenylation dependent mRNA decay pathway in eukaryotes.**

### 1.2.5 LSM2-8 Complex

The second LSM complex, LSM2-8, consists of LSM proteins 2 through 8 and has been localized to the nucleus[73]. LSM2-8 is found to interact with U6 snRNA, a component of the primary spliceosome[1,74]. U6 snRNA is produced by RNA Polymerase III, acquires a 5' $\gamma$-monomethyl phosphate cap and is retained in the nucleus[59,75]. This differs from U1, U2, U4 and U5 snRNAs which are transcribed by RNA Polymerase II and exported into the cytoplasm for further modification and maturation before their eventual return to the nucleus[57,59]. The major spliceosome is composed of all 5 of these snRNAs (U1, U2, U4, U5 and U6 snRNA) as well as many other interacting proteins[75-77] (Figure 1.5). The splicing reaction is complex and consists first of the U1 snRNP being recruited to the 5' splice site with other non-snRNP factors[66,67]. U2 snRNP then stably associates with one of the non-snRNP factors forming the A complex, or the pre-spliceosome[66,67]. Next, the U4/U6.U5 tri-snRNP which is assembled by way of the U5 and U4/U6 snRNP interaction is recruited and generates the pre-catalytic B complex[66,67]. Following the association of all U snRNPs, major RNA-protein and RNA-RNA rearrangements lead to the destabilization and subsequent release of U1 and U4 snRNPs, creating the activated spliceosome[66,67]. This activated spliceosome becomes catalytically active by way of the DEAH-Box RNA helicase Prp2 (DHX16 in humans) which sets off the remaining two steps of the spliceosome reaction[66,67]. Importantly, following a round of splicing the spliceosome is dissociated and the snRNPs are released becoming free to take part in additional rounds of splicing[66,67].

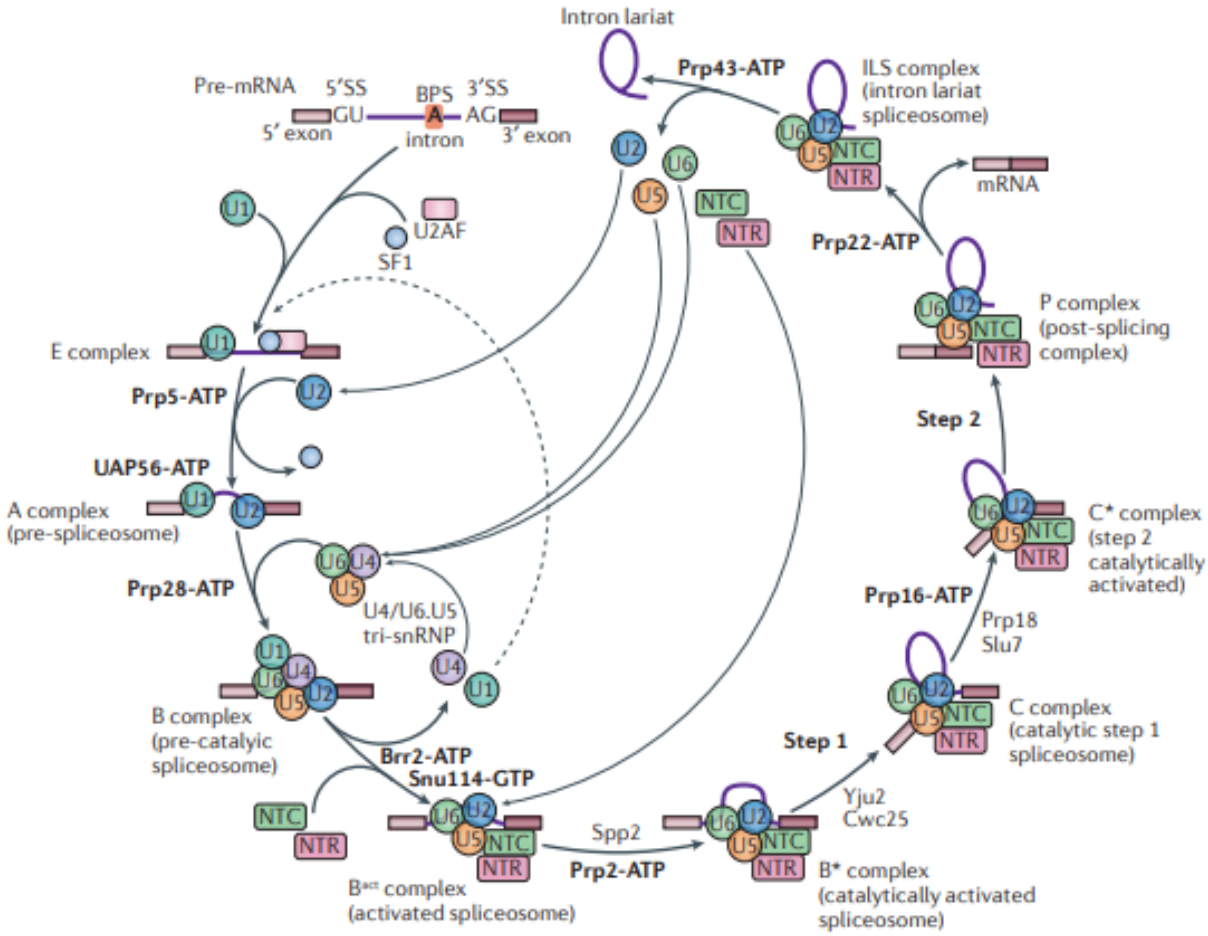



**Figure 1.5 The pre-mRNA splicing reactions and the splicing cycle in yeast[2].**
Permission to reprint is granted by Spring Nature and Copyright Clearance Center.

The role of the Lsm2-8 complex in splicing has been well established in yeast where Lsm2p through Lsm8p mutants show reduced levels of U6 snRNA[78]. It is known that Lsm2-8 interacts directly with U6 snRNA and is required in order to maintain its stability[59,74,75]. Importantly, U4/U6 duplexed RNAs are also found to directly associate with the Lsm2-8 complex[59,74]. It has thereby been suggested that Lsm2-8 is a chaperone which supports the many rearrangements of U6 snRNA containing complexes including its recycling following the spliceosome reaction[74,79]. As splicing is an extremely ordered process, mis-splicing events and/or the de-regulation of the entire spliceosome, could have grave consequences and possibly lead to disease[76,77,80].

### 1.2.6 Key Interactors

One of the most important interactors of the LSM1-7 complex which has been shown to play a role in facilitating decapping in the 5' to 3' decay of mRNA, is that of PATL1, or processing body mRNA decay factor[1]. The yeast homolog, Pat1b interacts stably and directly with Lsm2 and Lsm3 of the Lsm1-7 complex[65]. Pat1b is a multifunctional protein, binding to the decapping complex Dcp1-Dcp2 (DCP1A and DCP1B in humans) as well as another decapping activator Dhh1 (DDX6 in humans). It has also been shown to interact with the Xrn1 exoribonuclease and the Ccr4-Not deadenylase complex in yeast[1,65,81]. Interestingly, immunoprecipitation experiments found that FLAG-tagged Pat1b not only immunoprecipitated its predicted cytoplasmic partners including Lsm1-7, Dhh1, CNOT1, Xrn1, and decapping activator Edc4 but also the entire set of tri-snRNP factors. This included the splicing factors Prp24 (SART3 in humans), Prp31 (PRPF31 in humans), Brr2 (SNRNP200 in humans), Prp4 (PRPF4 in humans), Prp3 (PRPF3 in humans) and Lsm2-8[1] (Figure 1.6). Prp24 promotes U4 and U6 annealing while the association of U4/U6 with U5 is mediated by Prp31 and Prp6 (PRPF6 in humans)[1,82,83]. Prp3 and Prp4 interact with each other in the U4/U6 di-snRNP while snRNP200, Prp8 (PRPF8 in humans) and EFTUD2 are U5 snRNP proteins[1,83]. This study demonstrated that not only is Pat1b a key interactor of the Lsm1-7 complex playing a role in mRNA decay but it also interacts with Lsm2-8 and plays a role in the formation, many rearrangements, and recycling of the U4/U6.U5 tri-snRNP[1].

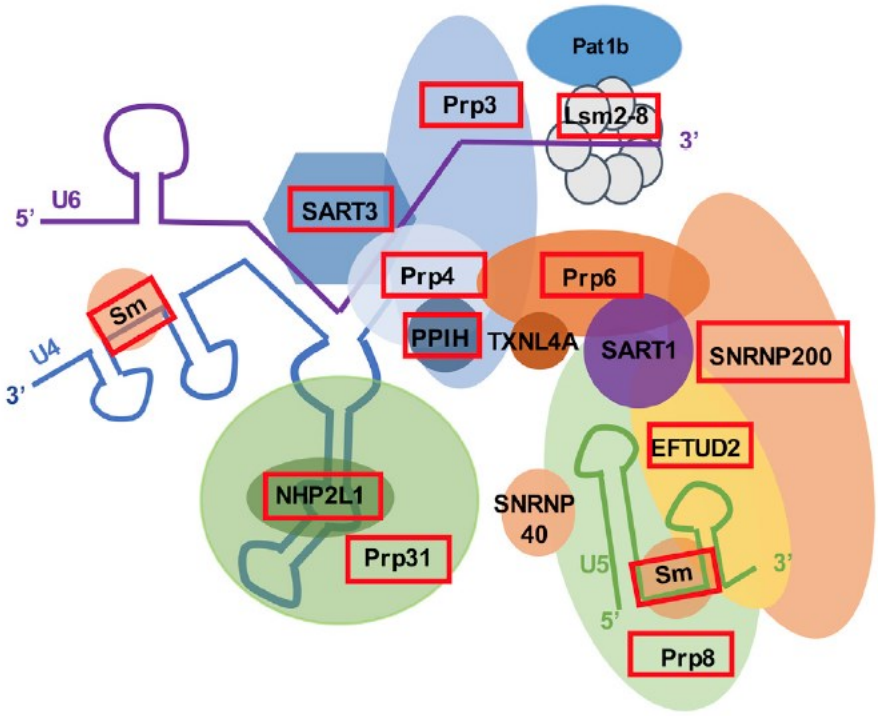



**Figure 1.6 Tri-snRNP proteins and RNAs.** Peptides in red boxes indicate proteins that co-immunoprecipitate with Pat1b. Permission to reprint is granted from Elsevier open access under the terms of the Creative Commons CC-BY license[1].

### 1.2.7 Other Functions of the LSM Complexes

Besides the two main LSM complexes, LSM1-7 which functions in 5' to 3' mRNA decay and LSM2-8 which functions in the maturation and stabilization of U6 snRNA, LSM proteins have also been found to play roles in other cellular processes[58,74]. In yeast, Lsm1p has been co-immunoprecipitated in an RNA dependent manner with the stem loop binding protein (SLBP), a key regulator in histone mRNA processing, translation, and stability[58]. Specifically, *Lsm1Δ* yeast have defective recovery from replication fork stalling and show DNA damage sensitivity[67]. One study found that the deletion of *Lsm1* increased the stability of histone mRNAs by 4 to 6-fold while another showed that abnormally high amounts of histones accumulate in *Lsm1Δ* yeast[67]. As such, it has been suggested that the Lsm1-7-Patb1 complex recruits decay factors to avoid the toxic effect of excessive histones in human cells[67]. This is supported by the finding that in yeast the degradation of mRNA histones occurs mainly by the 5' to 3' degradation pathway but this is severely impaired in the *Lsm1Δ* yeast[67]. Lsm proteins in yeast also play a role in the splicing of pre-tRNAs[59]. One study showed that the depletion of any of the 5 essential Lsm proteins (Lsm2, Lsm3, Lsm4, Lsm5 and Lsm8) in yeast resulted in the accumulation of unspliced pre-tRNA species in addition to the accumulation of 5' to 3' unprocessed species[59]. This relationship was first uncovered through yeast 2-hybrid screening which found that Lsm8 interacts with Sen1, an ATP dependent RNA helicase that acts as a positive effector of tRNA splicing[59]. The 2-hybrid screen also found an interaction between Lsm2 and Tpt1, a 2' phosphotransferase that functions in tRNA splicing[59]. Overall, these studies demonstrate that the roles that Lsm proteins play within the cell are quite extensive and should be studied in greater detail.

### 1.2.8 Disease Associations

There are currently no human diseases associated with LSM proteins, however, proteins within the LSM protein family, as well as proteins that interact with LSM proteins have been found to cause disease. One such connection is the interaction of Sm and LSM proteins with the SMN1 (Survival of Motor Neurons 1) protein[75,84]. Biallelic variants in the *SMN1* gene have a well-established link to spinal muscular atrophy (SMA), a severe neuromuscular disease resulting from motor neuron degeneration[75]. It is thought that the SMA phenotype is caused by a motor neuron specific response to defects in the cytoplasmic assembly and/or the nuclear cycling of the splicing machinery[75]. SMN is crucial for snRNP biogenesis in the cytoplasm and SMN oligomers have been found to interact with snRNP core proteins SmB, SmD1, and SmD3. Upon further investigation it was revealed that SMN actually exists in a complex with Sm proteins, Gemin2, Gemin3 and Gemin4 which form around and stabilize the U1, U2, U4 and U5 snRNAs[75,84]. An interaction between SMN and LSM4 and LSM6 has also been found, suggesting that the SMN complex also has a role in the assembly and/or regeneration of the LSM containing complexes including U6 snRNP[75].

Another disease of interest is that of spinocerebellar ataxia type 2 (SCA-2) which is an autosomal dominant neurodegenerative disease caused by trinucleotide (CAG) expansion in the gene *ATXN2*[85,86]. Interestingly, ataxin-2 is a member of the LSM protein family, carrying an LSM-domain and participating in functions related to RNA processing and metabolism[85,86]. ATXN2 is evolutionarily conserved from yeast and is widely expressed in the mammalian nervous system with heightened expression in cerebellar Purkinje cells[85,86]. ATXN2 specifically regulates global mRNA stability and translation by binding to uridine rich elements in the 3'UTR of its targets using its LSM domain[86]. The implication that LSM protein interactors and/or LSM domain

containing proteins cause disease supports the role of *LSM* genes as candidates for neurodegenerative diseases such as leukodystrophies.

## Chapter 2: Rationale and Hypothesis

### Project 1: Determining the genetic causes of molecularly unsolved cases of genetically determined leukoencephalopathies

In the first part of this project I used whole exome sequencing (WES) analysis to uncover the genetic causes of previously unsolved genetically determined leukoencephalopathies (gLEs). Many of the cases examined had previously been analyzed and no strong candidates found. As such, I completed a 2nd or even 3rd analysis 6-12 months following their initial negative analyses.

**Hypothesis:** We hypothesize we will obtain solve rates of 20-30% for our first analyses and of 5-11% for re-analyses as predicted by the literature[44,49-52].

**Aim:** To use WES to analyze or re-analyze molecularly unsolved cases of genetically determined leukoencephalopathies.

### Project 2: Biallelic mutations in *LSM7* impair assembly of LSM complexes and lead to a rare neurodevelopmental and neurodegenerative disorder

The second component of this study aimed at gaining a better understanding of the functions of native LSM7, as well as the two variant LSM7 proteins, and their respective roles in both the LSM1-7 and LSM2-8 complexes. By doing so we gained a greater understanding of cellular mechanisms through which these variants lead to the disease phenotype (Figure 2.1).

**Hypothesis**: The biallelic *LSM7* variants result in deregulated cellular functions which can be predicted by the type and location of the variants.

**Aim 1:** To determine the effect of the biallelic variants on *LSM7* mRNA and protein levels. We hypothesized that there will be a decrease in the levels of LSM7 protein in the two mutant proteins compared to wildtype.

**Aim 2:** To determine if the observed *LSM7* variants lead to defects in complex formation and interactions. As a result of the homology of the LSM7 protein with its yeast counterpart, it is anticipated that the p.Arg69Pro variant which corresponds to Arg87 in yeast, a residue predicted to contact the neighbouring Lsm5 residue[57,65,87], will lead to defects in complex assembly. It is also expected that there will be defects in both mRNA decay and splicing, the functions of the LSM1-7 and LSM2-8 complexes respectively, in addition to a loss of important protein interactors with these complexes. The second variant, p.Asp41Asn is predicted to correspond to an RNA binding residue in yeast[57,87] and therefore we hypothesized that it will also cause a loss in important protein interactors with the complexes, however not to the same severity as the p.Arg69Pro variant.

**Aim 3:** To characterize the role of Lsm7 in the nervous system development of zebrafish. **Part 3.1**: Knockdown of Lsm7 in zebrafish embryos using single guide RNAs. We predict that this will lead to nervous system developmental abnormalities similar to those seen in our living patient. **Part 3.2:** Rescue of Lsm7 knockout using human wildtype and variant constructs. We predict that the wildtype Lsm7 construct will be able to rescue the deficits in the knockdown zebrafish, and that the mutant constructs will lead to varying degrees of phenotypes.

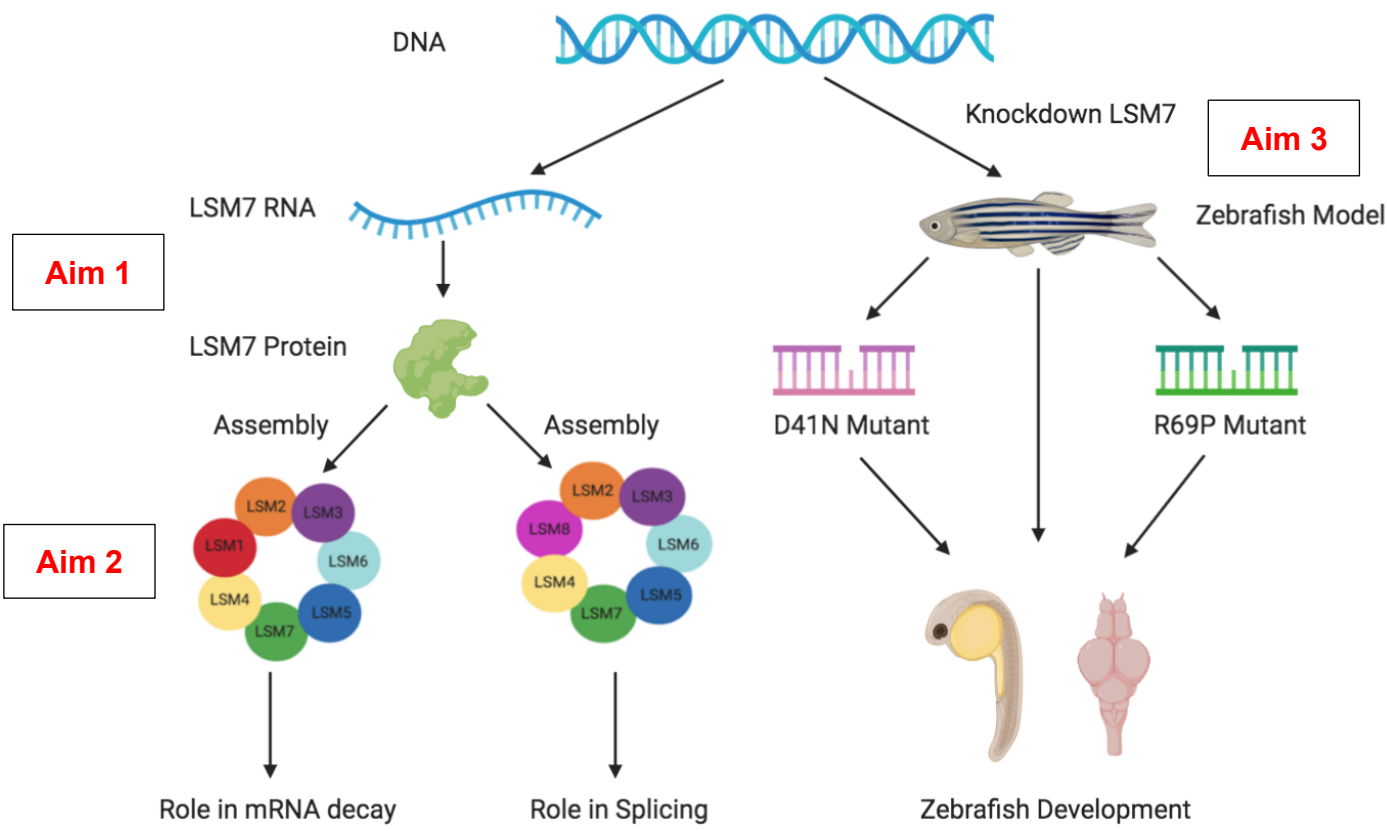



**Figure 2.1 Workflow of Project 2.**

# Chapter 3: Project 1 – Determining the genetic causes of molecularly unsolved cases of genetically determined leukoencephalopathies

## 3.0 Preface

Whole exome sequencing (WES) is frequently used in rare disease research when the cause of disease is believed to be of genetic origin and previous genetic testing has been negative[12,13]. The first part of this chapter describes the methods used to complete project 1 of this thesis including the collection of patient information, extraction of DNA, WES and subsequent WES analysis. The remainder of the chapter details the results uncovered by the WES analysis in a cohort of patients with molecularly unsolved genetically determined leukoencephalopathies. Of note, the majority of these WES analyses were 2nd or even 3rd analyses conducted 6-12 months following an initial negative analysis.

## 3.1 Materials and Methods

### 3.1.1 Patient Information

Informed consent was obtained from all participants. Approval for human subjects' research was obtained at the McGill University Health Centre Research Institute Research Ethics Board (projects 11-105-PED and 2019-4972). The medical charts and brain MRIs of all patients were reviewed.

### 3.1.2 DNA extraction

Genomic DNA was extracted from whole blood samples or fibroblasts using the Puregene Blood Core Kit C according to manufacturer's instructions (Qiagen Sciences, Hilden, Germany). The DNA quality and quantity were determined using a Nanodrop (Thermo Fisher Scientific).

### 3.1.3 WES Methods and Analysis

WES was performed on genomic DNA as previously described by Care4Rare[88] or at the Genomic Medical Center of Children's Mercy-Kansas City[24]. Illumina TruSeq Polymerase Chain Reaction (PCR) Free sample preparation or Nextera Rapid Capture Exome kits were used according to the manufacturer's protocols for exome enrichment, with validation performed by real-time PCR. The Illumina HiSeq 2000 or 2500 instruments with 2 × 100 nucleotide sequences, providing an average of 80-fold coverage were used for the sequencing of library samples. The subsequent sequences were aligned to the human reference genome (UCSC GRCh37/hg19) using the Burrows-Wheeler Aligner algorithm. Single nucleotide polymorphisms and small insertions and deletions were identified using SAM tools while the Genome Analysis Toolkit (GATK) was used for variant calling. Variant annotation was done using ANNOVAR.

WES analysis on patient singletons (Figure 3.1) was conducted in accordance the with the guidelines set out by the American College of Medical Genetics (ACMG) in association with the Association for Molecular Pathology[89]. As we are studying disorders that are rare and with an early onset variants present in healthy population databases with a minor allele frequency (MAF) of greater than 1% were filtered out as the carrier rates should be well below 1%[90,91]. Furthermore, the aggregation consortium and database GnomAD was used to filter and remove variants if they were found to be homozygous in the healthy control population as there should be no healthy adults reported with these homozygous variants[90,91]. Specific gene panels either developed in-house or acquired from the National Centre for Biotechnology Information (NCBI) Genetic Testing Registry were selected. Such panels were used to screen potential genetic candidates based on the clinical phenotypes of the patients. If no clear candidates were found the remainder of our in-house gene panels were used, additionally homozygous, compound heterozygous, X-linked and

novel not previously reported variants were screen for. Final screening of the variants was conducted using online tools including Online Mendelien Inheritance in Man (OMIM) as well as *in silico* prediction tools such as Mutation Taster, SIFT, Provean, and CADD to predict pathogenicity.

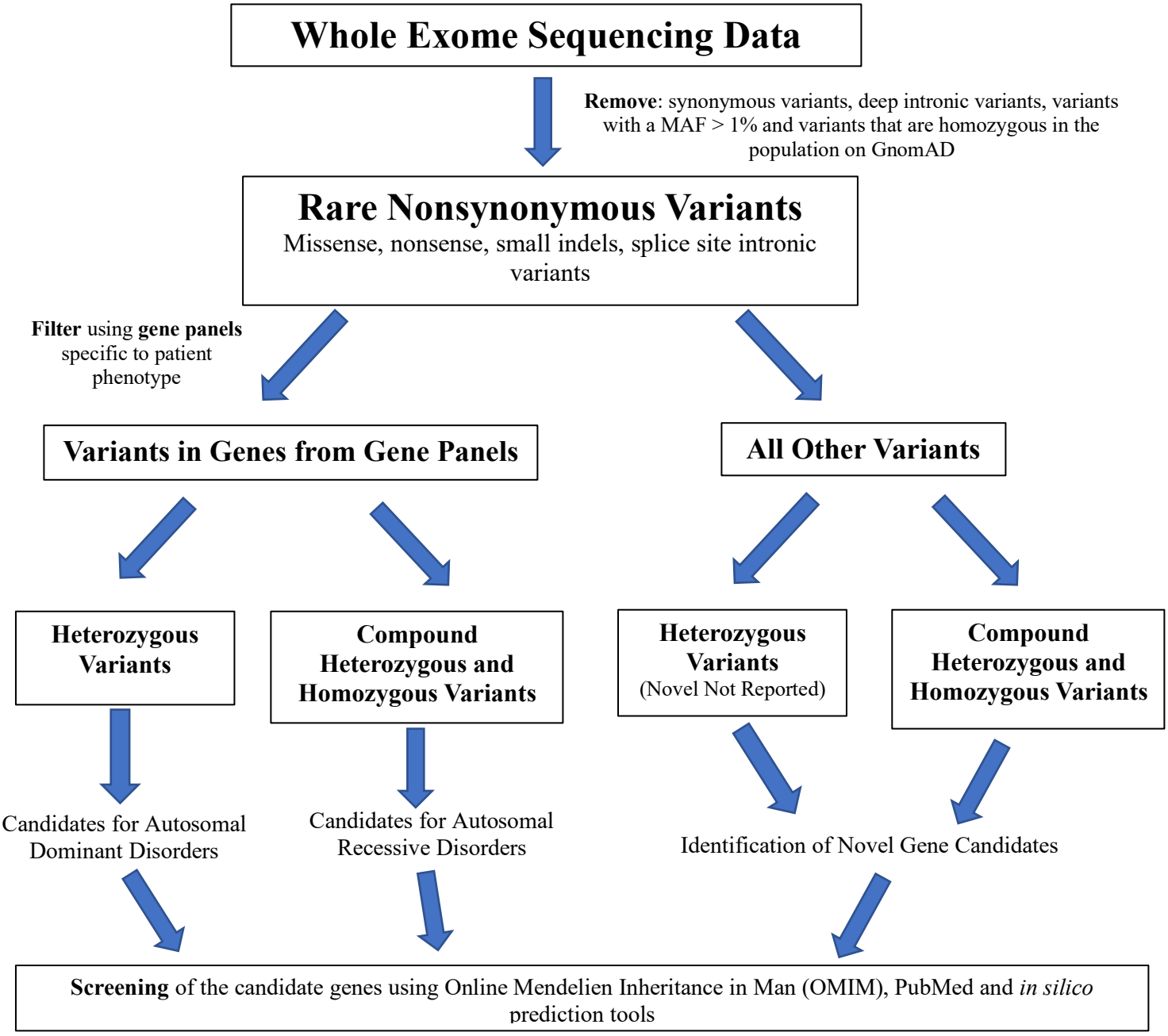



**Figure 3.1 Whole Exome Sequencing Analysis Workflow.**

### 3.1.4 Validation of Candidate Genes

#### Reporting

Following WES analysis, the pathogenicity status of candidate gene variants was determined in accordance with the ACMG guidelines[89]. If a variant/variants previously reported in the literature to cause a disease then Sanger sequencing would be performed on the patient and parents, and if segregation was confirmed, the case was reported as solved. If a novel variant in a gene associated with a disease was found, *in silico* prediction tools, conservation scores, as well as Sanger sequencing validation were used to predict if the novel variant was indeed pathogenic. If a genetic variant was found in a gene which had not been previously associated with a disease, a more intensive review of the literature would be conducted, including explorations into the function of the gene, potential cellular or animal models that have been created, in addition to *in silico* prediction tools and conservation scores to evaluate its pathogenicity. Furthermore, our cohort and those of our collaborators were searched to determine if other unsolved cases also had variants in this novel gene. Finally, functional studies could be performed on these genes and their specific variants in order to more clearly elucidate their disease association. These novel variants and/or genes are also commonly placed on matching tools such as GeneMatcher. GeneMatcher can be extremely valuable when looking to expand a clinical phenotype of a previous disease associated gene, or when characterizing the phenotype associated with a novel gene. Such sites allow for clinicians and researchers to be connected with others who report cases with overlapping phenotypes and variants in a gene of interest[12,13,92]. Finding another individual with variants in the same gene with an overlapping phenotype reinforce a novel gene as a possible candidate for disease and provide support to initiate functional testing[12]. These world-wide sharing efforts are

critical in the advancement of our knowledge of genes implicated in rare diseases as well as in the expansion of the phenotypic spectra associated with them[12,13].

**Sanger Sequencing**

Once potential candidate genes variants were uncovered in the singletons they were validated using Sanger sequencing to not only confirm the presence of the variant but also to determine, if parental DNA samples were available, if the variant(s) segregated. Primers were designed based on the reference sequences from UCSC GRCh37/hg19 and using Primer3 software to produce amplicon products of approximately 500-700 base pairs. Touchdown PCR was used to amplify the genomic DNA in the regions of interest of the candidate genes. Primer mixtures were prepared at 40uM and used at a final concentration of 0.8uM per sample. A total of 20ng of genomic DNA was used for each amplification. The cycling conditions used were as follows: 95°C for 10 minutes followed by 3 cycling conditions. The first condition were 5 cycles at 95°C for 15 seconds, then 60°C for 20 seconds, and finally, 72°C for 45 seconds. Next were 5 cycles with the annealing reduced from 60°C to 58°C. Third was 40 cycles with the annealing temperature further reduced to 54°C. Finally, the touchdown PCR ended with a single cycle of 72°C for 10 minutes for a final extension before cooling. The touchdown PCR was performed using the 2720 Thermo Cycler (LifeTechnologies). The product amplification was confirmed by running it on a 2% agarose gel with a 100 bp DNA ladder (FroggaBio). Lastly, the PCR products were forward and reverse sequenced at the McGill University and Genome Quebec Innovation Centre using an Applied Biosystems 3730xl DNA Analyzer and the results analyzed using SeqMan (DNAstar), following alignment to their respective genomic sequences.

## 3.2 Results

### 3.2.1 Description of Patient Cohort

Patients included in this cohort (Table 3.1) were referred to Dr. Geneviève Bernard or seen directly at the Montreal Children's Hospital Leukodystrophy and Neurometabolic Disorders Clinic. Patients were suspected to have a leukodystrophy, a genetically determined leukoencephalopathy or other genetically determined rare disease based on MRI imaging, clinical information, and biochemical abnormalities.

## Table 3.1 WES analysis for unsolved cases

| Patient Number | Patient ID | Analysis Number | Sex | Diagnosis |
|---|---|---|---|---|
| 1 | B5032 | 2nd | Male | Developmental delay<br>Non-specific white matter abnormalities |
| 2 | GB11.0 | 2nd | Female | Alternating hemiplegia of childhood |
| 3 | GB22.0 | 3rd | Male | Myelination delay<br>Hereditary spastic paraplegia<br>Global developmental delay<br>Myoclonic seizures |
| 4 | GB24.0 | 3rd | Female | Myelination delay<br>Hereditary spastic paraplegia<br>Global developmental delay |
| 5 | GB47.0A | 2nd | Male | Spastic quadraparesis<br>Ataxic gait<br>Periventricular leukomalacia-like MRI pattern<br>Global developmental delay |
| 6 | GB47.0B | 2nd | Male | Periventricular leukomalacia-like MRI pattern<br>Refractory epilepsy<br>Global developmental delay |
| 7 | GB52.0 | 2nd | Male | Cerebellar features, nystagmus, multifocal white matter abnormalities |
| 8 | GB53.0A | 2nd | Female | Spinal muscular atrophy, lower extremity predominant |
| 9 | GB58.0 | 2nd | Male | Spastic paraparesis<br>Cerebellar features |
| 10 | GB60.0A | 2nd | Female | Hereditary spastic paraplegia |
| 11 | GB60.0B | 2nd | Female | Hereditary spastic paraplegia |
| 12 | GB111.0 | 2nd | Male | Generalized dystonia |
| 13 | ET/D247.0 | 2nd | Male | Hypotonia<br>Gait ataxia<br>Global developmental delay |
| 14 | U106.0 | 2nd | Male | Hypomyelination disorder |
| 15 | WM195.0 | 1st | Female | Neurodegeneration with brain iron accumulation |
| 16 | WM279.0B | 1st | Male | Mitochondrial disorder |
| 17 | WM280.0 | 1st | Male | Aicardi-Goutières Syndrome-like disorder |
| 18 | WM283.0 | 1st | Male | Hypomyelinating leukodystrophy<br>Spastic paraparesis |

### 3.2.2 List of Candidate Genes for Patients with Molecularly Undetermined Diseases

WES analysis was performed on the 18 patients with unsolved genetically determined leukoencephalopathies or other undetermined rare genetic disorders as previously described[89]. Fourteen of the eighteen cases had previously been analyzed with no strong candidate gene found. As such, these were the second (12/18) or even the third (2/18) analyses performed. An overview of the patients' candidate genes, the variants identified, and a summary of the current status of their analysis is presented in Table 3.2. Only 1 case, that of patient 8 has been solved, however strong candidate genes were found in 4 other cases while weaker candidates or variants of unknown significance were uncovered in an additional 5 patients. Table 3.3 lists the pathogenicity of these candidate genes as per the ACMG guidelines. The remaining cases, which are not presented in Table 3.3 either had no strong candidate genes after validation with PCR and Sanger sequencing or have yet to be validated.

## Table 3.2 Summary of candidate genes identified by WES analysis

| Patient # | Patient ID | Gene | Transcript | Genomic Position | Variant | Status |
|---|---|---|---|---|---|---|
| 1 | B5032 | *CDON* | NM_001243597 | 11:125848211 | c.G3344T; p.R1115L | *ZCCHC12* is a weak candidate |
| | | *ZCCHC12* | NM_001312891 | X:117960030 | c.G823A; p.E275K | |
| 2 | GB11.0 | *DEAF1* | NM_021008.2 | 11:694957 | c.G91A; p.A31T | *DEAF1* to be validated |
| 3 | GB22.0 | *ITPR1* | NM_002222 | 3:4825507 | c.G6329A; p.N1795N | *ABHD16A* is a strong candidate gene – to be validated, in an effort to find more patients, 3 other families were found (manuscript in preparation) |
| | | | | 3:4776960 | c.C5277T; p.R2110Q | |
| | | *MTA2* | NM_004739 | 11:62365580 | c.G406A; p.V136M | |
| | | | | 11:622364891 | c.T594-4G | |
| | | *TREX1* | NM_033629 | 3:48508931 | c.C877G; p.L293V | |
| | | *AP5Z1* | NM_014855 | 7:4829521 | c.C1766T; p.S589L | |
| | | | | | Low coverage in several exons | |
| | | *RGS3* | NM_017790 | 9:116268778 | c.749+6_749+7del | |
| | | | | 9:116346115 | c.C1580A; p.T527N | |
| | | *NPAS3* | NM_022123 | 14:34269812 | c.G2203A; p.G735R | |
| | | *ABHD16A* | NM_001177515 | 6:31664801 | c.G254A; p.R85H | |
| | | | | 6:31664801 | c.G254A; p.A85H | |
| 4 | GB24.0 | *ITPR1* | NM_002222 | 3:4825507 | c.G6329A; p.N1795N | *ABHD16A* is a strong candidate gene – to be validated, in an effort to find more patients, 3 other families were found (manuscript in preparation) |
| | | | | 3:4776960 | c.C5277T; p.R2110Q | |
| | | *MTA2* | NM_004739 | 11:62365580 | c.G406A; p.V136M | |
| | | | | 11:622364891 | c.T594-4G | |
| | | *TREX1* | NM_033629 | 3:48508931 | c.C877G; p.L293V | |
| | | *AP5Z1* | NM_014855 | 7:4829521 | c.C1766T; p.S589L | |
| | | | | | Low coverage in several exons | |
| | | *RGS3* | NM_017790 | 9:116268778 | c.749+6_749+7del | |
| | | | | 9:116346115 | c.C1580A; p.T527N | |
| | | *NPAS3* | NM_022123 | 14:34269812 | c.G2203A; p.G735R | |
| | | *ABHD16A* | NM_001177515 | 6:31664801 | c.G254A; p.R85H | |
| | | | | 6:31664801 | c.G254A; p.R85H | |
| 5 | GB47.0A | *TSPL6* | NM_001003937 | 2:54482116 | c.A1173C; p.R391S | *CGN* – variant of unknown significance. Put on GeneMatcher |
| | | | | 2:54482256 | c.C1033G; p.H345D | |
| | | *ABP1* | NM_001091 | 7:150554447 | c.G889A; p.V297M | |
| | | | | 7:150555982 | c.C1702T; p.R568C | |
| | | *COL4A2* | NM_001846 | 13:111114519 | c.C1655A; p.T552K | |
| | | *RNASET2* | NM_003730.4 | 6:167356584 | c.T262-7A | |
| | | *TOMM40* | NM_001128916 | 19:45394808 | c.G136A; p.A46T | |
| | | | | | Low coverage in several exons | |
| | | *ATRN* | NM_139321.5 | 20:3577018 | c.C3417-6T | |
| | | | | | Low coverage in several exons | |
| | | *ZNF536* | NM_014717 | 19:30936178 | c.G1709C; p.G570A | |
| | | *CGN* | NM_020770 | 1:151493128 | c.A1101C; p.K367N | |
| | | | | 1:151502430 | c.C2152T; p.R718W | |
| 6 | GB47.0B | *TSPL6* | NM_001003937 | 2:54482116 | c.A1173C; p.R391S | *CGN* – variant of unknown significance. Put on GeneMatcher |
| | | | | 2:54482256 | c.C1033G; p.H345D | |
| | | *ABP1* | NM_001091 | 7:150554447 | c.G889A; p.V297M | |
| | | | | 7:150555982 | c.C1702T; p.R568C | |
| | | *COL4A2* | NM_001846 | 13:111114519 | c.C1655A; p.T552K | |
| | | *RNASET2* | NM_003730.4 | 6:167356584 | c.T262-7A | |
| | | *TOMM40* | NM_001128916 | 19:45394808 | c.G136A; p.A46T | |
| | | | | | Low coverage in several exons | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | *ATRN* | NM_139321.5 | 20:3577018 | c.C3417-6T | |
| | | | | | Low coverage in several exons | |
| | | *ZNF536* | NM_014717 | 19:30936178 | c.G1709C; p.G570A | |
| | | *CGN* | NM_020770 | 1:151493128 | c.A1101C; p.K367N | |
| | | | | 1:151502430 | c.C2152T; p.R718W | |
| 7 | GB52.0 | *ACBD5* | NM_001271512 | 10:27499796 | c.G1145A; p.R382 | To be re-sent for WES or sent for WGS or RNA seq |
| | | | | | Low coverage exon 10 and 11 | |
| | | *PEX5* | NM_000319 | 12:7343109-7343153 | c.136_147del; p.46_49del | |
| | | | | 12:7343109-7343153 | c.136_147del; p.46_49del | |
| | | *SLAIN1* | NM_001242868 | 13:78272259 | c.C211T; p.P71S | |
| | | | | 13:78272259 | c.C211T; p.P71S | |
| | | *FAT1* | NM_005245 | 4:187584660 | c.A3373C; p.I1125L | |
| | | | | 4:187629951 | c.C1031Y; p.P344L | |
| | | *TMCC3* | NM_001301036 | 12:94965347 | c.C1205T; p.A402V | |
| | | | | 12:94975816 | c.C484T; p.L162F | |
| | | *DNAH6* | NM_001370 | 2:84771598 | c.C904A; p.Q302K | |
| | | *HSP90B* | NM_003299 | 12:104327743 | c.G421A; p.E141K | |
| | | *NCOR1* | NM_001190438 | 17:16004652 | c.G2323A; p.A775T | |
| | | *INST2* | NM_001330417 | 17:59984814 | c.C1136T; p.A379V | |
| 8 | GB53.0A | *ISCA2* | NM_19279 | 14:74960797 | c.C116A; p.S39X | *DYNC1H1* – solved, manuscript in preparation (see below) |
| | | | | 14:74961352 | g.C930T | |
| | | *B4GALNT1* | NM_001478 | 12:58022914 | c.A728G; p.E243G | |
| | | *TTF2* | NM_003594 | 1:117603184 | c.G131+5A | |
| | | | | 1:117605053 | c.A176G; p.Q59G | |
| | | *URB1* | NM_014825 | 21:33705646 | c.T940C; p.F1647S | |
| | | *CNTNAP2* | NM_014141 | 7:147926845 | c.C3355T; p.R1119C | |
| | | | | | Low coverage 5' UTR | |
| | | *PTPRZ1* | NM_002851 | 7:121653532 | c.T4432A; p.S1478T | |
| | | | | | Low coverage exon 21 | |
| | | *TMEM240* | NM_001114748 | 1:1471017 | c.G325A; p.G109S | |
| | | *RFT1* | NM_052859 | 3:53164412 | c.A5G; p.G2D | |
| | | | | | Low coverage 5'UTR | |
| | | *PLEKHG5* | NM_001042664 | 1:6533387 | c.A719G; p.D240G | |
| | | *DYNC1H1* | NM_001376 | 14:102446289 | c.G752T; p.R251L | |
| 9 | GB58.0 | *POLG* | NM_001126131 | 15:89862169 | c.A3266G; p.Q1089R | *LARGE1* – variant of unknown significance. Put on GeneMatcher. |
| | | | | | c.155_156insGCA; p.55_56insQ | |
| | | *NEB* | NM_004543 | 2:152522024 | c.G5061A; p.W1687* | |
| | | | | 2:152432753 | c.G11717A; p.R3906Q | |
| | | *SDK2* | NM_00114452 | 17:71357828 | c.G5462A; p.G1821E | |
| | | | | 17:71437028 | c.C648G; p.T216= | |
| | | *KIF1A* | NM_001244008.1 | 2:241696840 | c.2751_2753delGGA; p.917_918del | |
| | | | | 2:241662027 | c.C4511T; p.T1504I | |
| | | *VPS13C* | NM_017684 | 15:62208206 | c.C7942G; p.H2648D | |
| | | | | 15:62320573 | c.C432T | |
| | | *AGAP1* | NM_001037131 | 2:236579620 | c.C164-38203G | |
| | | *DNAH7* | NM_018897 | 2:196602782 | c.A11938G; p.T3980A | |
| | | *FAT2* | NM_001447 | 5:150922635 | c.C8066T; p.P2689L | |
| | | *LARGE1* | NM_004737 | 22:33780242 | c.A941T; p.Q314L | |
| | | | | 22:33673157 | c.G1962A; p.E654= | |
| 10 | GB60.0A | *LAD1* | NM_005558 | 1:201358354 | c.C158T; p.S53F | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 1:201351424 | c.A1546G; p.R516G | *ABHD16A* is a strong candidate gene – to be validated, in an effort to find more patients, 3 other families were found (manuscript in preparation) |
| | | *SNX2* | NM_003100 | 5:122110810 | c.G12C; p.E4D | |
| | | | | 5:122153010 | c.G948C; p.L199L | |
| | | *KMT2C* | NM_170606 | 7:151860675 | c.G9987A; p.M3329I | |
| | | | | 7:151896449 | p.K1396K | |
| | | *TRMT1* | NM_017722 | | Deletion exons 8-14 | |
| | | *AUTS2* | ENST00000406775.2 | 7:70252222 | c.A2264G; p.D755G | |
| | | *KANK1* | NM_015158 | 9:711518 | c.C752T; p.P251L | |
| | | | | 9:712343T | c.A1577G; p.M526R | |
| | | *BBS10* | NM_024685 | 12:76741078 | c.T687A; p.P229P | |
| | | | | 12:76740664 | c.G1101C; p.L367F | |
| | | *MAP4K3* | NM_03618 | 2:39526950 | c.A1120-8G | |
| | | | | 2:39553351 | c.G598T; p.V200L | |
| | | *IMPACT* | NM_018439 | 18:22006705 | c.G-16A | |
| | | *ABHD16A* | NM_001177515 | 6:31664801 | c.G254A; p.R85H | |
| | | | | 6:31664801 | c.G254A; p.R85H | |
| 11 | GB60.0B | *LAD1* | NM_005558 | 1:201358354 | c.C158T; p.S53F | *ABHD16A* is a strong candidate gene – to be validated, in an effort to find more patients, 3 other families were found (manuscript in preparation) |
| | | | | 1:201351424 | c.A1546G; R516G | |
| | | *SNX2* | NM_003100 | 5:122110810 | c.G12C; p.E4D | |
| | | | | 5:122153010 | c.G948C; p.L199L | |
| | | *KMT2C* | NM_170606 | 7:151860675 | c.G9987A; p.M3329I | |
| | | | | 7:151896449 | p.K1396K | |
| | | *TRMT1* | NM_017722 | | Deletion exons 8-14 | |
| | | *AUTS2* | ENST00000406775.2 | 7:70252222 | c.A2264G; p.D755G | |
| | | *KANK1* | NM_015158 | 9:711518 | c.C752T; p.P251L | |
| | | | | 9:712343T | c.A1577G; p.M526R | |
| | | *BBS10* | NM_024685 | 12:76741078 | c.T687A; p.P229P | |
| | | | | 12:76740664 | c.G1101C; p.L367F | |
| | | *MAP4K3* | NM_03618 | 2:39526950 | c.A1120-8G | |
| | | | | 2:39553351 | c.G598T; p.V200L | |
| | | *IMPACT* | NM_018439 | 18:22006705 | c.G-16A | |
| | | *ABHD16A* | NM_001177515 | 6:31664801 | c.G254A; p.R85H | |
| | | | | 6:31664801 | c.G254A; p.R85H | |
| 12 | GB111.0 | *APOB* | NM_000384 | 2:21230987 | c.A8753G; p.H2918R | To be re-analyzed with new pipeline |
| | | *ABCB7* | NM_004299 | X:74291494 | c.G1060A; p.A354T | |
| | | *WFS1* | NM_00605 | 4:6303211 | c.1689_1694del; p.563_565del | |
| 13 | ET/D247.0 | *IFIH1* | NM_022168 | 2:163167292 | c.G605A; p.C202Y | *MOCS1* to see if validated clinically |
| | | *ASTN1* | NM_001286164 | 1:176853546 | c.T3155C; p.V1052A | |
| | | | | 1:177030380 | c.A305C; p.D102A | |
| | | *SSPO* | NM_198455 | 7:149477992 | c.G1763A; p.C588Y | |
| | | | | 7:149493735-149493738 | c.6731_6734del; p.G2244fs | |
| | | *LHX4* | NM_033343 | 1:180243497 | c.C956A; p.S319Y | |
| | | *CAPN11* | NM_007058 | 6:44148395 | c.G1742A; p.G581E | |
| | | *C5orf42* | NM_023073 | 5:37157843 | c.C7886T; p.S2629L | |
| | | | | 5:37227490 | c.A1376C; p.K459T | |
| | | *MOCS1* | NM_005943 | 6: 39883820 | c.G575A; p.R192H | |
| | | | | 6: 39883820 | c.G575A; p.R192H | |
| 14 | U106.0 | *ZEB2* | NM_001171653 | 2:145156156 | c.G2526C; p.K842N | To be re-analyzed with new pipeline |
| | | *ATRX* | NM_138270 | X:76814157 | c.T6373A; p.Y2125N | |
| 15 | WM195.0 | *FADS6* | NM_178128 | 17:72889676 | c.17_18insGATGGAACCTACGGAGCCCATGGAACCTACGGAGCCCATGGAACCTACGGAGCC; | To be re-analyzed with new pipeline |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | p.P6delinsPMEPTEPMEPTEPMEPTEP | |
| | | | | 17:72889676 | c.17_18insGATGGAACCTACGGAGCCCATGGAACCTACGGAGCCCATGGAACCTACGGAGCC; p.P6delinsPMEPTEPMEPTEPMEPTEP | |
| | | *MROH8* | NM_152503 | 20:35807790 | c.93+1->AGTGCCGGCCGCGGGGCCCTGTCTATAAG | |
| | | | | 20:35807790 | c.93+1->AGTGCCGGCCGCGGGGCCCTGTCTATAAG | |
| | | *ZFC3H1* | NM_144982 | 12:72056858 | c.G533T; p.G178V | |
| | | *HEATR1* | NM_018072 | 1:236720571 | c.C5279A; p.A1760D | |
| | | *GPR12* | NM_005288 | 13:27333239 | c.C726A; p.H242Q | |
| | | *TP53BP1* | NM_001141979 | 15:43771595 | c.G788C; p.R263T | |
| 16 | WM279.0 B | *ZFYVE26* | NM_015346.3 | chr14:68215166 | c.G7607C; p.G2536A | Variants to be validated |
| | | | | chr14:68274306 | c.T695G; p.L232R | |
| | | *DMXL2* | NM_015263.3 | chr15:51791525 | c.C3896T; p.S1299L | |
| | | | | chr15:51868380 | c.88-3delT | |
| | | *TRMT44* | NM_152544.2 | chr4:8469891 | c.A1745G; p.E582G | |
| | | | | chr4:8477525 | c.G2071T; p.D691Y | |
| | | *HERC2* | NM_004667.5 | chr15:28419774 | c.G9832-8A | |
| | | | | chr15:28456220 | c.C6997T; p.L2333F | |
| | | *TGIF1* | NM_003244.3 | chr18:3457787 | c.G668A; p.G223E | |
| 17 | WM280.0 | *CRLS1* | NM_019095.4 | chr20:6011962 | c.T606G; p.D202E | Variants to be validated |
| | | | | chr20:6011962 | c.T606G; p.D202E | |
| | | *SAMHD1* | NM_015474.3 | chr20:35579880 | c.G167A; p.R56H | |
| | | | | chr20:35579880 | c.G167A; p.R56H | |
| | | *KMT2B* | NM_014727.1 | chr19:36214675 | c.A3101C; p.E1034A | |
| | | *ETFA* | NM_001127716.1 | chr15:76523730 | c.A679C; p.I227L | |
| | | | | chr15:76523730 | c.A679C; p.I227L | |
| | | *OBSCN* | NM_001271223.2 | chr1:228447246 | c.G4906A; p.V1636M | |
| | | | | chr1:228447246 | c.G4906A; p.V1636M | |
| | | *ARHGEF28* | NM_001177693.1 | chr5:73153483 | c.T1793G; p.I598S | |
| | | *SHMT1* | NM_004169.3 | chr17:18244085 | c.G562A; p.E188K | |
| | | | | chr17:18244085 | c.G562A; p.E188K | |
| | | *VWA8* | NM_001009814.1 | chr13:42465635 | c.A572G; p.D191S | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | chr13:42465635 | c.A572G; p.D191S | |
| 18 | WM283.0 | *TMEM67* | NM_001142301.1 | chr8:94772149 | c.T91C; p.Y31H | Variants to be validated |
| | | | | chr8:94800154 | c.G1252T; p.D418Y | |
| | | *DMD* | NM_004012.3 | chrX:31986576 | c.C2462A; p.T2161N | |
| | | *IFIH1* | NM_022168.3 | Chr2:16313097 | c.G2362A; p.A788T | |

## Table 3.3 Pathogenicity of the candidate genes as per ACMG Guidelines

| Patient Number | Patient ID | Gene | Transcript ID | Genomic Position | Variant | Zygosity | Interpretation of Pathogenicity | MAF | Criteria | In silico Predictions MT, Provean, SIFT, PolyPhen-2, CADD |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | B5032 | *ZCCHC12* | NM_001312891 | X:117960030 | c.G823A;pE275K | Hemizygous | Uncertain significance | 3.42e-5 2 hemizygotes | PM3, PP3 | MT: Polymorphism Provean: Neutral SIFT: Damaging PolyPhen-2: Probably damaging CADD: 25.4 |
| 2 | GB11.0 | *DEAF1* | NM_021008.2 | 11:694957 | c.G91A; p.A31T | Heterozygous | Uncertain significance – awaiting Sanger validation | 4.06e-5 | PM2, PM3, PP3 | MT: Disease causing Provean: Neutral SIFT: Damaging PolyPhen-2: Probably damaging CADD: 22.6 |
| 3 | GB22.0 | *ABHD16A* | NM_001177515 | 6:31664801 | c.G254A; p.Arg85His | Homozygous | Likely pathogenic – awaiting Sanger validation | 8.11e-6 | PM2, PM3, PP3, PP4 | MT: Disease causing Provean: Deleterious SIFT: Damaging CADD: 29.8 |
| 4 | GB24.0 | *ABHD16A* | NM_001177515 | 6:31664801 | c.G254A; p.Arg85His | Homozygous | Likely pathogenic – awaiting Sanger validation | 8.11e-6 | PM2, PM3, PP3, PP4 | MT: Disease causing Provean: Deleterious SIFT: Damaging CADD: 29.8 |
| 5 | GB47.0A | *CGN* | NM_020770 | 1:151493128 | c.A1101C;p.K367N | Compound heterozygous | Likely pathogenic | Not reported | PM2, PM3. PP3, PP4 | MT: Polymorphism Provean: Neutral SIFT: Damaging CADD: 21.5 |
| | | *CGN* | NM_020770 | 1:151502430 | c.C2152T;p.R718W | Compound heterozygous | Uncertain significance | 1.23e-3 4 homozyotes | PM3, PP3, PP4 | MT: Polymorphism Provean: Deleterious SIFT: Damaging CADD: 18.31 |
| 6 | GB47.0B | *CGN* | NM_020770 | 1:151493128 | c.A1101C;p.K367N | Compound heterozygous | Likely pathogenic | Not reported | PM2, PM3. PP3, PP4 | MT: Polymorphism Provean: Neutral SIFT: Damaging CADD: 21.5 |
| | | *CGN* | NM_020770 | 1:151502430 | c.C2152T;p.R718W | Compound heterozygous | Uncertain significance | 1.23e-3 4 homozyotes | PM3, PP3, PP4 | MT: Polymorphism Provean: Deleterious SIFT: Damaging CADD: 18.31 |
| 8 | GB53.0A | *DYNC1H1* | NM_001376 | 14:102446289 | c.G752T;p.R251L | Heterozygous | Pathogenic | Not reported | PS2, PM1, PM2, PP2, PP3, PP4 | MT: Disease causing Provean: Deleterious SIFT: Tolerated PolyPhen-2: Probably damaging CADD: 25.2 |
| 9 | GB58.0 | *LARGE1* | NM_004737 | 22:33780242 | c.A941T;p.Q314L | Compound heterozygous | Likely pathogenic | 4.24e-5 | PM2, PM3, PP2, PP3, PP4 | MT: Disease causing Provean: Deleterious SIFT: Tolerated PolyPhen-2: Benign CADD: 23.1 |
| | | *LARGE1* | NM_004737 | 22:33673157 | c.G1962A;p.E654= | Compound heterozygous | Uncertain significant | 6.26e-4 | PM2, PM3, PP4 | MT: Disease causing Provean: Neutral SIFT: Tolerated CADD: 9.559 |
| 11 | GB60.0A | *ABHD16A* | NM_001177515 | 6:31664801 | c.G254A; | Homozygous | Likely pathogenic – | 8.11e-6 | PM2, PM3, | MT: Disease causing Provean: Deleterious |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | p.Arg85His | | awaiting Sanger validation | | PP3, PP4 | SIFT: Damaging<br>CADD: 29.8 |
| 12 | GB60.0B | *ABHD16A* | NM_001177515 | 6:31664801 | c.G254A; p.Arg85His | Homozygous | Likely pathogenic – awaiting Sanger validation | 8.11e-6 | PM2, PM3, PP3, PP4 | MT: Disease causing<br>Provean: Deleterious<br>SIFT: Damaging<br>CADD: 29.8 |
| 13 | ET/D247.0 | *MOCS1* | NM_005943 | 6:39883820 | c.G575A;p.R192H | Homozygous | Uncertain significance – awaiting Sanger validation | 5.13e-5 | PM2, PP2, PP3 | MT: Disease causing<br>Provean: Deleterious<br>SIFT: Damaging<br>PolyPhen-2: Probably damaging<br>CADD: 32 |

# Chapter 4: DYNC1H1 Case Report

## 4.0 Preface

*De novo* variants in the gene *DYNC1H1* have been found to cause spinal muscular atrophy with lower extremity predominance, type 1 (SMALED1), Charcot-Marie-Tooth disease, axonal type 20 (CMT) and malformations in cortical development (MCD)[93,94]. This chapter has been adapted from the manuscript which describes the case of twin girls who present with a large phenotypic discordance and a novel *de novo* variant *DYNC1H1*: Derksen A^, Mirchi A^, Tran LT, Cao L, Oskoui M, Srour M, Poulin C, Care4Rare Canada Consortium, and Bernard G^^. A novel *de novo* variant in *DYNC1H1* causes spinal muscular atrophy lower extremity predominant in identical twins with discordant phenotypes. To be submitted shortly to Pediatric Neurology.

## 4.1 Title Page

**A novel *de novo* variant in *DYNC1H1* causes spinal muscular atrophy lower extremity predominant in identical twins with discordant phenotypes**

*Abbreviated Title:* Novel *DYNC1H1* variant causes SMALED1

Alexa Derksen [a, b, †], Amytice Mirchi, MD [b, c, †], Luan T. Tran [a, b], Lei Cao [a, b], Maryam Oskoui [b, c], MD, Myriam Srour [b, c], MD, Chantal Poulin [b, c], MD, Care4Rare Canada Consortium, Genevieve Bernard, MD, MSc, FRCP(c) [a, b, c, d, e, *]
[a] Child Health and Human Development Program, Research Institute of the McGill University Health Centre, Montreal, Canada
[b] Department of Neurology and Neurosurgery, McGill University, Montreal, Canada
[c] Department of Pediatrics, McGill University, Montreal, Canada
[d] Department of Human Genetics, McGill University, Montreal, Canada
[e] Division of Medical Genetics, Department of Specialized Medicine, McGill University Health Centre, Montreal, Canada

Conflict of interest: None
[†] These authors were equally responsible for the work described in this paper
* Communications should be addressed to:

Geneviève Bernard, MD, MSc, FRCPc
MyeliNeuroGene Laboratory

Research Institute of the McGill University Health Centre
1001 boul Décarie
Site Glen Pavilion E / Block E
*CHHD* Mail Drop Point #EM03211 (cubicle C)
Montréal, QC H4A 3J1
Canada
Tel:514-412-4400 ext.:23380
Fax:514-933-4149
genevieve.bernard@mcgill.ca

## 4.2 Abstract

**Background:** Spinal muscular atrophy lower extremity predominant 1 (SMALED1) is an autosomal dominant genetic neuromuscular disorder characterized by degeneration of spinal cord motor neurons which results in muscle weakness. Mutations in *DYNC1H1* have been associated with SMALED1.

**Methods:** Retrospective medical records review was performed for both patients. Whole exome sequencing was performed on one sibling and Sanger sequencing to confirm the presence of the variant and confirm the *de novo* inheritance in both twins.

**Patient Description:** Twin girls with a large discordance in disease severity were found to have a novel *de novo* heterozygous missense mutation c.G752T; p.Arg251Leu in *DYNC1H1*.

**Conclusions:** A novel *de novo* mutation in *DYNC1H1* causes SMALED1 in twin girls. Epigenetic factors, mosaicism, and the events that Twin A experienced at birth are some likely explanations for the phenotypic discordance.

**Key words:**
*DYNC1H1*
Whole exome sequencing
WES
Spinal muscular atrophy with lower extremity predominant type 1
SMALED1

## 4.3 Introduction

Spinal muscular atrophy lower extremity predominant type 1 (SMALED1) is an autosomal dominant genetic neuromuscular disorder characterized by congenital or very early onset static or slowly progressive muscle weakness primarily in the lower limbs without sensory abnormalities.[1] Mutations in the gene *DYNC1H1* (Online Mendelian Inheritance in Man (OMIM) #600112) have been associated with SMALED1 (OMIM #158600), as well as several other diseases including malformation of cortical development (MCD) (OMIM #614563) and Charcot-Marie-Tooth disease, axonal type 2O (OMIM #614228).[1,2] This gene has also been implicated in different neuropsychiatric disorders including autism spectrum disorder (ASD), intellectual disability (ID) and in a handful of cases with epileptic encephalopathy (EE).[3] *DYNC1H1* (NM_001376) is located on chromosome 14q32 and encodes the heavy chain of cytoplasmic dynein-1, the motor protein complex which is critical for the transport of organelles, vesicles and macromolecules towards the minus end of microtubules.[4] Neurons as a result of their long processes, are extremely reliant on the functional operation of these microtubules transport systems to traffic cellular components to and from their axons and dendrites.[4] Here we present twin girls with SMALED1 but with a large phenotypic discordance and report a novel *de novo* mutation in the gene *DYNC1H1* at position c.G752T; p.Arg251Leu.

## 4.4 Methods

All experiments involving human participants or data were conducted in compliance with all relevant ethical regulations. Informed consent was obtained from all participants. Approval for human subjects' research was obtained at the McGill University Health Centre Research Institute Research Ethics Board (project number 11-105-PED and 2019-4972). DNA was extracted from whole blood of Twin A, Twin B and unaffected parents. WES was performed by Care4Rare

according to previously published methods[5] using DNA from Twin A and analysis was conducted in accordance with the American College of Medical Genetics (ACMG) guidelines[6]. Candidate genes were validated in Twin A, Twin B and both unaffected parents using Sanger sequencing. A thorough review of the twin's medical records was also conducted, including their clinical evaluation, neuroradiological findings and metabolic investigations.

**Patient Description**

**Clinical findings**

The twin girls, now aged 10, were delivered at 35 weeks and 2 days of gestational age via an induced vaginal delivery due to weight discordance on the follow-up ultrasound and the possibility of twin-twin transfusion syndrome. At birth, this hypothesis was refuted. Twin A weighed 2525 grams at birth, with an Apgar score of 9-10-10. No cord pH was obtained. Her sister, Twin B, had a birth weight of 1800 grams with an Apgar score of 9-10-10.

Shortly after birth, Twin A developed respiratory distress secondary to transient tachypnea of the newborn and required continuous positive airway pressure ventilation (CPAP) for a few hours. However, at day of life 2, she was found to have persistent metabolic acidosis with increased anion gap. Lactates reached a maximal value of 16 on day of life 3 and she was described as encephalopathic, with poor suck, hypotonia and increased drowsiness. She required to be placed back on CPAP for a few hours that day. She was also found to have neonatal seizures and was started on phenobarbital. She received a total of 3 doses of bicarbonate and by day of life 4, the metabolic acidosis had resolved. In the absence of a sentinel anoxic event, extensive metabolic investigations were done and are all negative to this day. She went on to develop infantile spasms, spastic quadriparesis, severe global developmental delay, cortical blindness, moderate to severe dysarthria and acquired microcephaly. Nerve conductions studies have been within normal limits.

However, electromyography studies have shown signs of a chronic denervating process suggestive of a motor neuron or root abnormality.

Twin B presents a much milder phenotype with spastic paraparesis and Trendelenburg gait as well as gross motor delay. She has never had any seizures, is able to walk by herself without support and attends a regular school although she is known for attention deficit disorder and learning disability. She also has normal nerve conduction studies although electromyography has shown neurogenic changes that have evolved over time and that involve mostly the lower limbs.

**Neuroimaging findings**

Brain imaging studies on Twin A have revealed extensive changes in both cerebral hemispheres. Ventricular enlargement has been noted (Figure 4.1). This has been associated with a loss of white matter substance posteriorly and abnormal signal of the white matter in T2-weighted and FLAIR images indicating periventricular leukomalacia. No anomalies were noted on the spine imaging studies.

Brain imaging study for Twin B performed at 3 years of age is unremarkable aside from an increased signal intensity on T2 sequence in the peritrigonal area and subcortical U fibers, likely representing a terminal zone of myelination.

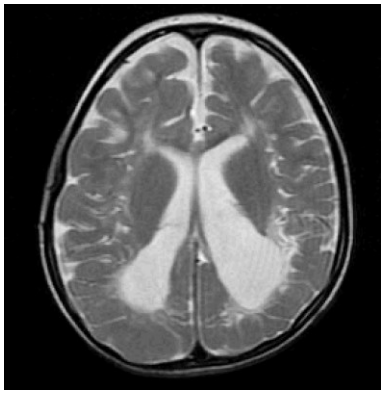
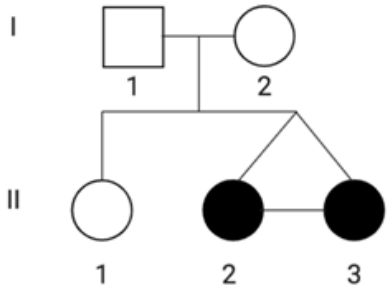


**Figure 4.1 Brain MRI of Twin A done at 3 years of age.** Axial T2 weighted image showing ventricular enlargement and asymmetry as well as extensive changes in both cerebral hemispheres and cerebral atrophy.

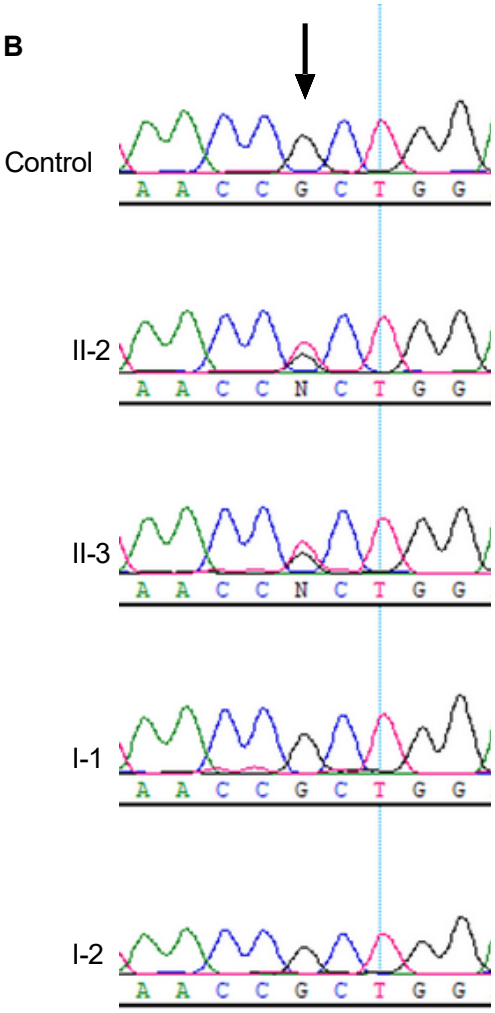
**Genetic findings**

WES analysis was performed on Twin A as previously described[5] and revealed a heterozygous variant in *DYNC1H1* at position c.G752T; p.Arg251Leu. Sanger sequencing confirmed the presence of this variant in both Twin A, Twin B and its absence in the healthy parents and controls (Figure 4.2A & B). This mutation is predicted to be pathogenic by numerous *in silico* prediction tools (Figure 4.2C), is found at a highly conserved amino acid residue (Figure 4.2D), and the ACMG guidelines predict it to be likely pathogenic.

A




B




C

| Predictive Software | Score | Prediction |
|---|---|---|
| PolyPhen-2 | 0.994 | Probably Damaging |
| Provean | -4.69 | Deleterious |
| MutationTaster | 0.999 | Disease Causing |
| CADD | 25.2 | Deleterious |

D

| | | | | | |
|---|---|---|---|---|---|
| Human | V | N | R | W | I |
| Rhesus | V | N | R | W | I |
| Mouse | V | N | R | W | I |
| Dog | V | N | R | W | I |
| Elephant | V | N | R | W | I |
| Chicken | V | N | R | W | I |

**Figure 4.2 Detection of *DYNC1H1* variant**. (A) Pedigree of family, black circles represent the affected children, II-2 is Twin A and II-3 Twin B. (B) Sanger sequencing confirmation of novel *de novo DYNC1H1* variant (NM_001376: c.G752T; p.Arg251Leu) in Twin A and Twin B and absence from unaffected parents. (C) Pathogenicity predictions for the c.G752T; p.Arg251Leu variant using PolyPhen-2, Provean, MutationTaster and CADD prediction software. (D) Arginine (R) at position 251 in *DYNC1H1* is a highly conserved amino acid residue.

## 4.5 Discussion

In this study we report for the first time a *de novo* missense mutation in *DYNC1H1* at position c.G752T; p.Arg251Leu to cause SMALED1. The ACMG guidelines predict this variant to be likely pathogenic and its location on the DYNC1H1 protein is consistent with other mutations

known to cause SMALED1. This is important as there appears to be a correlation between the location of the mutation in the DYNC1H1 protein and the associated disease phenotype.[4,7,8] Variants in the stem domain of this protein particularly those clustered in the important dimerization domain are associated with CMT and SMALED1 (Figure 4.3).[4,7,8] This novel variant is located just adjacent to the dimerization domain of this protein and is clustered amongst other variants reported to cause SMALED1 (Figure 4.3B).

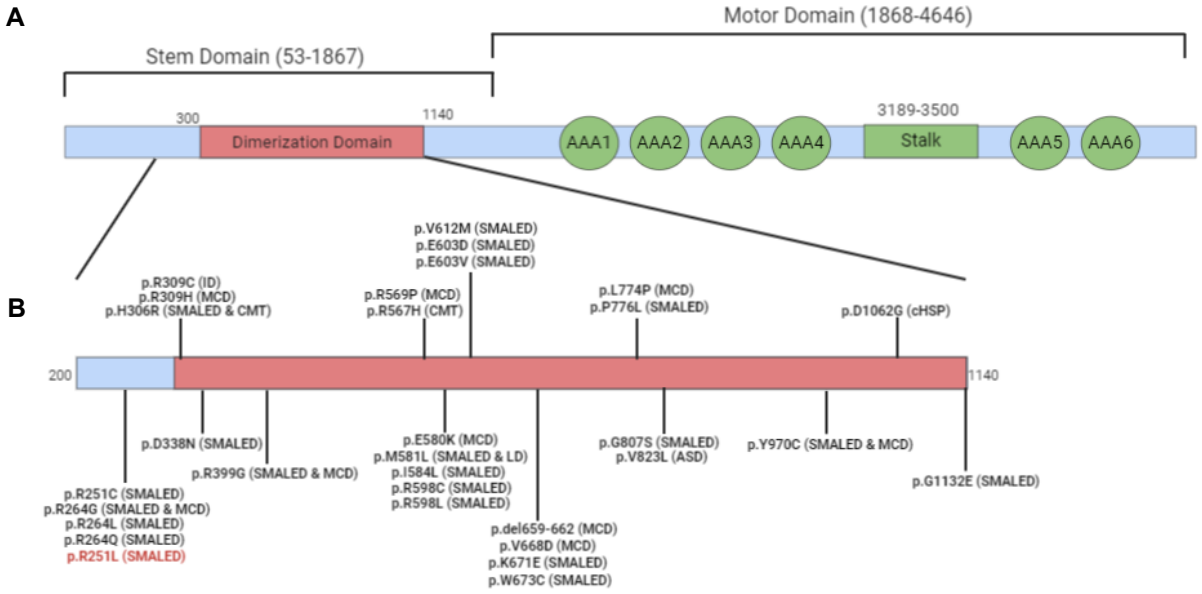



**Figure 4.3 Locations of reported disease-causing variants on DYNC1H1.** (A) Domains of DYNC1H1 protein. (B) Location of variants known to cause disease. Our novel variant p.R251L is indicated in red.
LD = learning disability, MCD = malformations in cortical development, CMT = Charcot-Marie-Tooth disease, SMALED = Spinal muscular atrophy lower extremity predominant, cHSP = complex hereditary spastic paraplegia, ASD = autism spectrum disorder.

Of interest in this case is the discordance in phenotypic presentation of the disease in these monozygotic twins. Mild variability in disease onset as well as disease phenotype has been seen from one generation to the next in family pedigrees of patients with SMALED1.[7,9] This case, however, appears to be the most striking in terms of discordance reported in the literature and is

even more interesting as the patients are identical twins. One contributing factor could be the respiratory distress that Twin A experienced at birth. However, the lack of an acute perinatal hypoxic event, adequate Apgar scores, and no history of encephalopathy until the third day of life in Twin A weaken this hypothesis. Furthermore, the white matter changes seen on imaging studies for Twin A are out of proportion to the severity of the event. As such, we believe that the MRI changes are at least in part secondary to *DYN1CH1*-related SMALED1. If there was a component of hypoxic-ischemic insult that went unnoticed, such as an in-utero insult, or if the post-natal respiratory distress led to some degree of hypoxia, one can wonder whether Twin A was predisposed due to an underlying more severe neuromuscular disease. The role of *DYNC1H1* in neuronal function of the central as well as peripheral nervous system has been demonstrated and mutations in *DYN1CH1* have also been associated with brain cortical malformations and ventricular dilatation.[4,10] Taking all of this to account, other avenues to describe this discordance were explored.

A disease in which monozygotic twin discordance has been well described is that of Beckwith-Wiedermann syndrome (BWS).[11,12] This discordance has been linked to imprinting defects or more broadly epigenetic defects that appear to be unique to the twinning event.[11] It is thought that the unequal splitting of the inner cell mass which contains the DNA methylation enzymes occurs during twinning resulting in differential methylation statuses of various genes throughout the genome.[13] As such, one possibility in our case is that differences at the epigenetic level may account for the discordance seen in the twins. Other cases of twins with highly discordant phenotypes have also reported in both Rett and Joubert syndromes.[14] These discordances have been attributed to the unequal allocation of the blastomeres at the time of separation, differences in attachment to the placenta or type of chorion, as well as the unequal

sharing of venous return from the monochorionic placenta.[11,14] It is also possible that the unequal allocation of cells during splitting could result in mosaicism[12,15], that is, Twin A could have the variant *DYNC1H1* in a higher percent of tissues compared to Twin B. These are some examples of potential explanations why our monozygotic twins have such discordant phenotypes.

This clinical case adds to the growing list of disease-causing variants in the gene *DYNC1H1*. It is unique due to the discordance in the phenotypic presentation in a set of monozygotic twins. Although the reasons for this discordance are still uncertain, an unnoticed hypoxic event could have played a role, and such event may have occurred because of a more severe phenotype to start with, and therefore predisposing the patient. Other factors are at play, and we hypothesize that they could be due to mosaicism and/or epigenetic factors.

## 4.6 References


1 Beecroft, S. J. *et al.* Expanding the phenotypic spectrum associated with mutations of DYNC1H1. *Neuromuscul Disord* **27**, 607-615, doi:10.1016/j.nmd.2017.04.011 (2017).

2 Strickland, A. V. *et al.* Mutation screen reveals novel variants and expands the phenotypes associated with DYNC1H1. *J Neurol* **262**, 2124-2134, doi:10.1007/s00415-015-7727-2 (2015).

3 Lin, Z. *et al.* Whole-exome sequencing identifies a novel de novo mutation in DYNC1H1 in epileptic encephalopathies. *Sci Rep* **7**, 258, doi:10.1038/s41598-017-00208-6 (2017).

4 Hoang, H. T., Schlager, M. A., Carter, A. P. & Bullock, S. L. DYNC1H1 mutations associated with neurological diseases compromise processivity of dynein-dynactin-cargo adaptor complexes. *Proc Natl Acad Sci U S A* **114**, E1597-E1606, doi:10.1073/pnas.1620141114 (2017).

5 Hamilton, A. *et al.* Concordance between whole-exome sequencing and clinical Sanger sequencing: implications for patient care. *Mol Genet Genomic Med* **4**, 504-512, doi:10.1002/mgg3.223 (2016).

6 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).

7 Tsurusaki, Y. *et al.* A DYNC1H1 mutation causes a dominant spinal muscular atrophy with lower extremity predominance. *Neurogenetics* **13**, 327-332, doi:10.1007/s10048-012-0337-6 (2012).

8 Poirier, K. *et al.* Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly. *Nat Genet* **45**, 639-647, doi:10.1038/ng.2613 (2013).

9 Harms, M. B. *et al.* Dominant spinal muscular atrophy with lower extremity predominance. *neurology* **10** (2010).

10 Chan, S. H. S. *et al.* A recurrent de novo DYNC1H1 tail domain mutation causes spinal muscular atrophy with lower extremity predominance, learning difficulties and mild brain abnormality. *Neuromuscul Disord* **28**, 750-756, doi:10.1016/j.nmd.2018.07.002 (2018).

11 Matias, A., Silva, S., Martins, Y. & Blickstein, I. Monozygotic twins: Ten reasons to be different. *Diagnóstico Prenatal* **25**, 53-57, doi:10.1016/j.diapre.2013.09.003 (2014).

12 Zwijnenburg, P. J., Meijers-Heijboer, H. & Boomsma, D. I. Identical but not the same: the value of discordant monozygotic twins in genetic research. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 1134-1149, doi:10.1002/ajmg.b.31091 (2010).

13 Wong, A. H., Gottesman, II & Petronis, A. Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum Mol Genet* **14 Spec No 1**, R11-18, doi:10.1093/hmg/ddi116 (2005).

14 Burnham, K. Discordant Symptoms in Monozygotic Twins with Huntington's Disease, Spinocerebellar Ataxia and other Neurodegenerative and Mendelian Disorders. *J Mol Genet Med* **12**, 1-4 (2018).

15 Castillo-Fernandez, J. E., Spector, T. D. & Bell, J. T. Epigenetics of discordant monozygotic twins: implications for disease. *Genome Medicine* **6**, 1-16 (2016).

# Chapter 5: Project 2 - Biallelic mutations in *LSM7* impair assembly of LSM complexes and lead to a rare neurodevelopmental and neurodegenerative disorder

## 5.0 Preface

Biallelic pathogenic variants in *LSM7,* a novel disease-causing gene were found in two unrelated patients via WES. This chapter describes the methods used to quantify the levels of U6 snRNA, the principle interactor of the LSM2-8 complex, in patient GB114.0 (p.Asp41Asn) compared to healthy controls and displays the subsequent results. The methods and results included in this chapter are absent from the manuscript found in chapter 6 of this thesis.

## 5.1 Materials and Methods

### 5.1.1 Primary Fibroblasts Cell Cultures

Fibroblasts were obtained via skin punches from patient GB114.0 and age-sex-matched controls and expanded according to a previously reported protocol[26]. Cells were grown in Dulbecco Modified Eagle Media (DMEM) supplemented with 10% fetal bovine serum (FBS). Cell lines were maintained at 37°C in a humidified 5% $CO_2$ atmosphere and were collected at the same time for further analysis once confluent and at similar passages.

### 5.1.2 Quantification of U6 snRNA Levels by RT-qPCR

Total RNA was extracted from fibroblast cells in TRIzol according to the manufacturer's instructions (Invitrogen). RNA concentration and integrity were determined using a NanoDrop Spectrophotometer ND-1000. 30µg of RNA was DNase I (Promega) treated for 30 minutes at 37°C. For each sample 1µg of total RNA was reverse transcribed using M-MLV (Promega) and random hexamers (Invitrogen) according to manufacturer's protocol (Promega). cDNAs were

then amplified by PCR using the following conditions: 95 °C for 10 seconds, 60 °C for 30 seconds, 72 °C for 30 seconds and repeated for 45 cycles. Primers for RT-qPCR were designed, and efficiency tested according to the MIQE guidelines. Samples were performed in triplicates with a 1:4 dilution of cDNA using SsoAdvanced Universal SYBR Green Supermix (Biorad) on the Roche LightCycler 96. All quantification data were normalized to *SDHA* or *RPL30* using the ΔΔCt method in accordance with the MIQE guidelines. RNA extractions and RT-qPCR analyses were performed with at least 6 biological replicates with each of these biological replicates analyzed by at least 3 technical replicates. The primers used for RT-qPCR were U6 snRNA forward 5'-CTCGCTTCGGCAGCACA-3', U6 snRNA reverse 5'-AACGCTTCACGAATTTGCGT-3'. *RPL30* forward 5'-AAGGCAGGAAGATGGTGGCC-3', *RPL30* reverse 5'-GAGTCTGCTTGTACCCCAGGAC-3', and *SDHA* forward 5'-CAGCATGTGTTACCAAGCTG-3', *SDHA* reverse 5'-GGTGTCGTAGAAATGCCAAC-3'.

### 5.1.3 Statistical Analysis

qPCR data were analyzed following the minimum information for publication of quantitative real-time PCR experiments (MIQE) guidelines and expressed as average relative normalized expression with appropriate error bars (SEM). Normality of all data were assessed using the Shapiro-Wilks Test of Normality and the equality of variances were assessed using Levene's Test. Statistical significance was set at $p < 0.05$.

## 5.2 Results

### 5.2.1 U6 snRNA Levels are Increased in our Patient's Fibroblasts

The levels of U6 snRNA, the principle interactor of the LSM2-8 complex was measured via RT-qPCR and found to be increased in GB114.0 (p.Asp41Asn) compared to the age- and sex-matched healthy controls ($p<0.001$).

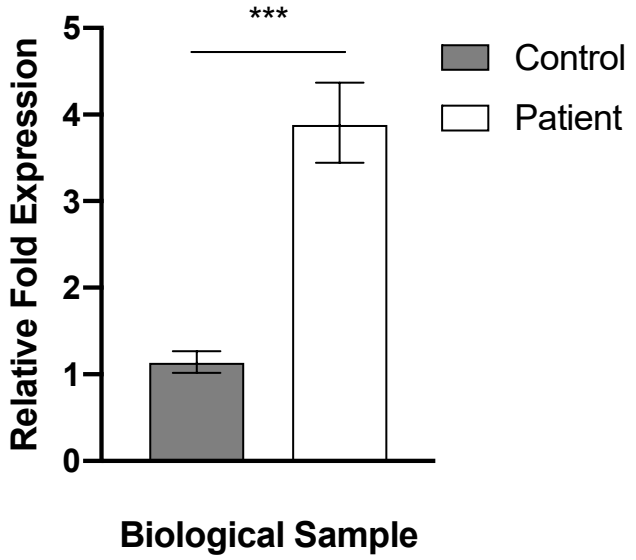



**Figure 5.2 RT-qPCR analysis of U6 snRNA levels in patient and age-sex-matched controls**

# Chapter 6: LSM7 Manuscript

## 6.0 Preface

This chapter is adapted from a manuscript which contains the results from the functional experiments used to determine the native functions of LSM7, as well as the two variant LSM7 proteins, and their respective roles in both the LSM1-7 and LSM2-8 complexes. It also describes the role of Lsm7 in the nervous system development of the model organism, *Danio rerio*.

Derksen A^, Shih A^, Forget D, Darbelli L, Tran LT, Poitras C, Guerrero K, Sundaresan T, Alkuray F, Kurdi WIY, Nguygen CTE, Laberge AM, Si Y, Gauthier MS, Bonkowsky J^^, Coulombe B^^ and Bernard G^^. Biallelic pathogenic variants in *LSM7* impair assembly of LSM complexes and lead to a rare neurodevelopmental and neurodegenerative disorder. Intended to be submitted for publication by September 2020 to Nature Genetics.

## 6.1 Title Page

**Biallelic pathogenic variants in *LSM7* impair assembly of LSM complexes and lead to a rare neurodevelopmental and neurodegenerative disorder**

Alexa Derksen[1,2,3,16], Hung-Yu Shih[4,16], Diane Forget[2], Lama Darbelli[1], Luan T. Tran[1], Christian Poitras[2], Kether Guerrero[1], Sundaresan Tharun[5], Fowzan S. Alkuraya[6], Wesam I. Kurdi[7], Cam-Tu Emilie Nguyen[8], Anne-Marie Laberge[9], Yue Si[10,] Marie-Soleil Gauthier[2], Joshua L. Bonkowsky[4,11,17], Benoit Coulombe[2,13,17], Geneviève Bernard*[1,3,12,14,15,17]

1 Child Health and Human Development Program, Research Institute of the McGill University Health Centre, Montréal, Canada
2 Translational Proteomics Laboratory, Institut de Recherches Cliniques de Montréal, Montréal, Canada
3 Department of Neurology and Neurosurgery, McGill University, Montréal, Canada
4 Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, Utah, USA
5 Department of Biochemistry, Uniformed Services University of Health Sciences (USUHS), Bethesda, MD, USA
6 Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia
7 Department of Obstetrics and Gynecology, Maternal Fetal Medicine, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia

8 Neurosciences Department, Université de Montréal, Montréal, Canada
9 Service de Génétique Médical, CHU Sainte-Justine, Montréal, Canada
10 GeneDx, Gaithersburg, Maryland, USA
11 Brain and Spine Center and Primary Children's Center for Personalized Medicine, Primary Children's Hospital, Salt Lake City, Utah, USA
12 Department of Pediatrics, McGill University, Montréal, Canada
13 Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, Canada
14 Department of Human Genetics, McGill University, Montréal, Canada
15 Division of Medical Genetics, Department of Specialized Medicine, McGill University Health Centre, Montréal, Canada
16 These authors contributed equally: Alexa Derksen and Hung-Yu Shih
17 These authors jointly supervised this work: Joshua Bonkowsky, Benoit Coulombe and Genevieve Bernard
* Correspondence should be addressed to Dr. Geneviève Bernard (genevieve.bernard@mcgill.ca)

**CORRESPONDING AUTHOR:**
Geneviève Bernard, MD, MSc, FRCPc
MyeliNeuroGene Laboratory
Research Institute of the McGill University Health Centre
1001 boul Décarie
Site Glen Pavilion E / Block E
*CHHD* Mail Drop Point #EM03211 (cubicle C)
Montréal, QC H4A 3J1
Canada
Tel:514-412-4400 ext.:23380
Fax:514-933-4149
genevieve.bernard@mcgill.ca

## 6.2 Abstract

Leukodystrophies, genetic neurodevelopmental and/or neurodegenerative disorders of cerebral white matter, have recently been associated with variants in numerous transcription- and translation-related genes. Here, we identify two patients with biallelic variants in *LSM7*, leading to leukodystrophy or *in utero* death. LSM7 is part of two complexes involved in mRNA decay and splicing. Affinity purification-mass spectrometry of the LSM7 variants showed defects in LSM complexes assembly. *Lsm7* knockdown in zebrafish led to central nervous system defects, including impaired oligodendrocyte development and motor behaviour. Our findings add *LSM7* to the list of leukodystrophy-causing genes and identify a novel category of disease mechanisms.

## 6.3 Introduction

Leukodystrophies encompass a group of rare genetic neurodevelopmental and neurodegenerative disorders affecting cerebral white matter. Recent advances in genetics and neuroimaging have led to the discovery of many subtypes that are clinically and genetically heterogeneous[1-4]. These diseases can be classified based on specific brain magnetic resonance imaging (MRI) patterns as either hypomyelinating or non-hypomyelinating. Hypomyelinating leukodystrophies are associated with abnormal myelin formation during development, whereas non-hypomyelinating leukodystrophies are associated with abnormal myelin homeostasis[5,6]. These heterogenous diseases collectively affect as many as 1 in 7,600 individuals with a broad range of onset spanning from prenatal life to late adulthood[7-9]. Although the era of next generation sequencing (NGS) has facilitated rare disease diagnosis, 20-30% of patients with leukodystrophies remain genetically undiagnosed[2,10,11]. These patients and their families often

undergo extensive diagnostic odysseys, which has an impact on their wellbeing and healthcare plans[12,13].

Leukodystrophies were initially identified and classified predominantly as disorders of myelin production or turnover, with the classical examples being Pelizaeus-Merzbacher Disease and Metachromatic Leukodystrophy[9]. Recent advances in research have revealed that pathogenic variants in many transcription- and translation-related genes are implicated in white matter pathologies. For example, it is well-established that pathogenic variants in the genes encoding RNA polymerase III subunits (i.e. *POLR3A*, *POLR3B*, *POLR1C* and *POLR3K*) as well as tRNA synthetase genes (i.e. *EPRS1*, *RARS1*, *VARS1*, etc.), are associated with white matter disorders[14-20].

LSM proteins are a family of proteins that form multi-subunit complexes that interact directly with RNA[21-25]. The 8 main LSM proteins (LSM1-8) form two heteroheptameric ring shaped complexes comprised of different subunits. More specifically, the LSM1 through LSM7 proteins form the LSM1-7 complex and the LSM2 through LSM8 proteins form the LSM2-8 complex. The LSM1-7 complex interacts with the mRNA processing and degradation factor protein PATL1, localizing to the cytoplasmic Processing bodies (P-bodies), and plays a role in mRNA decay[25-27]. Eukaryotic cytosolic mRNA decay, a well characterized and highly conserved process occurs via one of two pathways, both beginning with the shortening of the poly(A) tail and followed by degradation in either the 3'-5' or 5'-3' direction[22,24,25]. Studies in yeast have revealed that the LSM1-7 complex likely acts as an enhancer of decapping in the 5'-3' pathway, as cells with mutations in Lsm1p through Lsm7p block mRNA decay at the decapping step[22,28,29]. Conversely, the other LSM2-8 complex is localized to the nucleus, where it interacts with U6 snRNA, which is the catalytic component of the primary spliceosome[21,23,24]. Splicing is a multi-

step reaction involving the interaction and subsequent rearrangements of the 5 primary spliceosomal snRNAs (U1, U2, U4, U5 and U6) and their protein interactors[28-32]. The LSM2-8 complex acts as a chaperone for U6 snRNA, supporting its rearrangements during splicing and its recycling following the spliceosome reaction[33,34]. Thus, the LSM1-7 and LSM2-8 complexes play critical roles in the highly regulated cellular mechanisms of mRNA decay and splicing, respectively[21].

Here, we report that biallelic pathogenic variants in *LSM7* cause an ultra-rare neurodevelopmental and neurodegenerative disease. The novel *LSM7* variants were investigated on a molecular and functional level to evaluate their impact on the LSM complexes. Using a zebrafish model, we studied the impact of LSM7 hypofunction on central nervous system (CNS) development. This is the first report linking *LSM7* to any human disease. Furthermore, this work adds mRNA decay and splicing regulation to the categories of cell dysfunction implicated in causing leukodystrophies, neurodegenerative and neurodevelopmental diseases.

## 6.4 Results

### Identification of *LSM7* biallelic variants

Patient 1 was referred at the age of 3 years for neurological assessment at the McGill University Health Center. He presented with spastic quadriparesis, white matter anomalies, developmental delay, progressive sensorineural hearing loss, hypospadias and feeding difficulties. He is currently 7 years old, and wheelchair bound. His MRIs showed cerebellar hypoplasia and progressive white matter signal abnormalities between 17 and 35 months, indicative of a leukodystrophy (Figure 6.1). To determine the genetic cause, exome sequencing was performed on patient 1 and his parents. The mean exome coverage for patient 1 was 98.4% with mean depth of coverage of 88 reads. Analysis was completed using in-house filters and identified a

homozygous variant in *LSM7* (NM_016199.2; GRCh37/hg19) at position c.121G>A; p.Asp41Asn as a strong candidate for pathogenicity (Figure 6.2a-b). Targeted Sanger sequencing confirmed the presence of the variant as homozygous in the patient, with both parents being carriers (Figure 6.2d). Subsequently, a second patient with an assumed homozygous variant in *LSM7* was found through GeneMatcher[35,36]. This patient died *in utero*, but duo exome sequencing of the parents revealed that each carried an identical variant in *LSM7* (c.206G>C; p.Arg69Pro). Both *LSM7* variants are present in highly conserved amino acid residues (Figure 6.2c), and classified as rare, as the p.Asp41Asn variant was observed only 9 times as heterozygous in 232,800 alleles in large population cohorts[37] (minor allele frequency 3.87e-5), while the p.Arg69Pro variant has never been reported in large population control databases. *In silico* genetic analysis softwares predict both variants to be deleterious on protein structure and function (p.Asp41Asn: Mutation Taster 0.99, CADD 28.8, Provean -4.61, SIFT 0.001; p.Arg69Pro: Mutation Taster 0.99, CADD 24.9, Provean -7.00, SIFT 0.00)[38-42] (Supplementary Table 6.1).

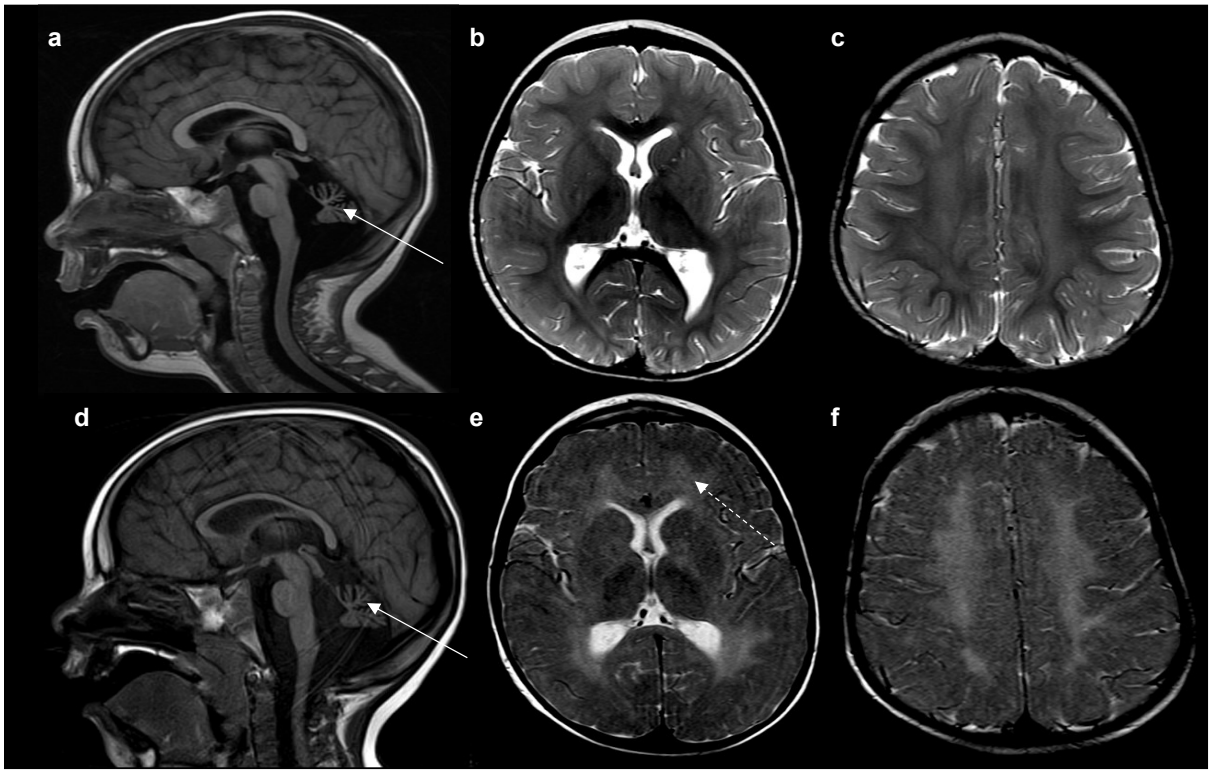



**Figure 6.1 MRI characteristics of the rare neurodevelopmental and neurodegenerative disorder caused by a biallelic pathogenic variant in *LSM7*.** Brain MRI of patient 1 at 17 months (**a-c**) and 35 months of age (**d-f**). **a** and **d**, Sagittal T1-weighted images at the midline showing cerebellar hypoplasia (thin arrows). **b**, **c**, **e** and **f**, Axial T2-weighted images at the level of the basal ganglia (**b**,**e**) and central semi-ovale (**c**,**f**) showing progressive diffuse demyelination (dotted arrow).

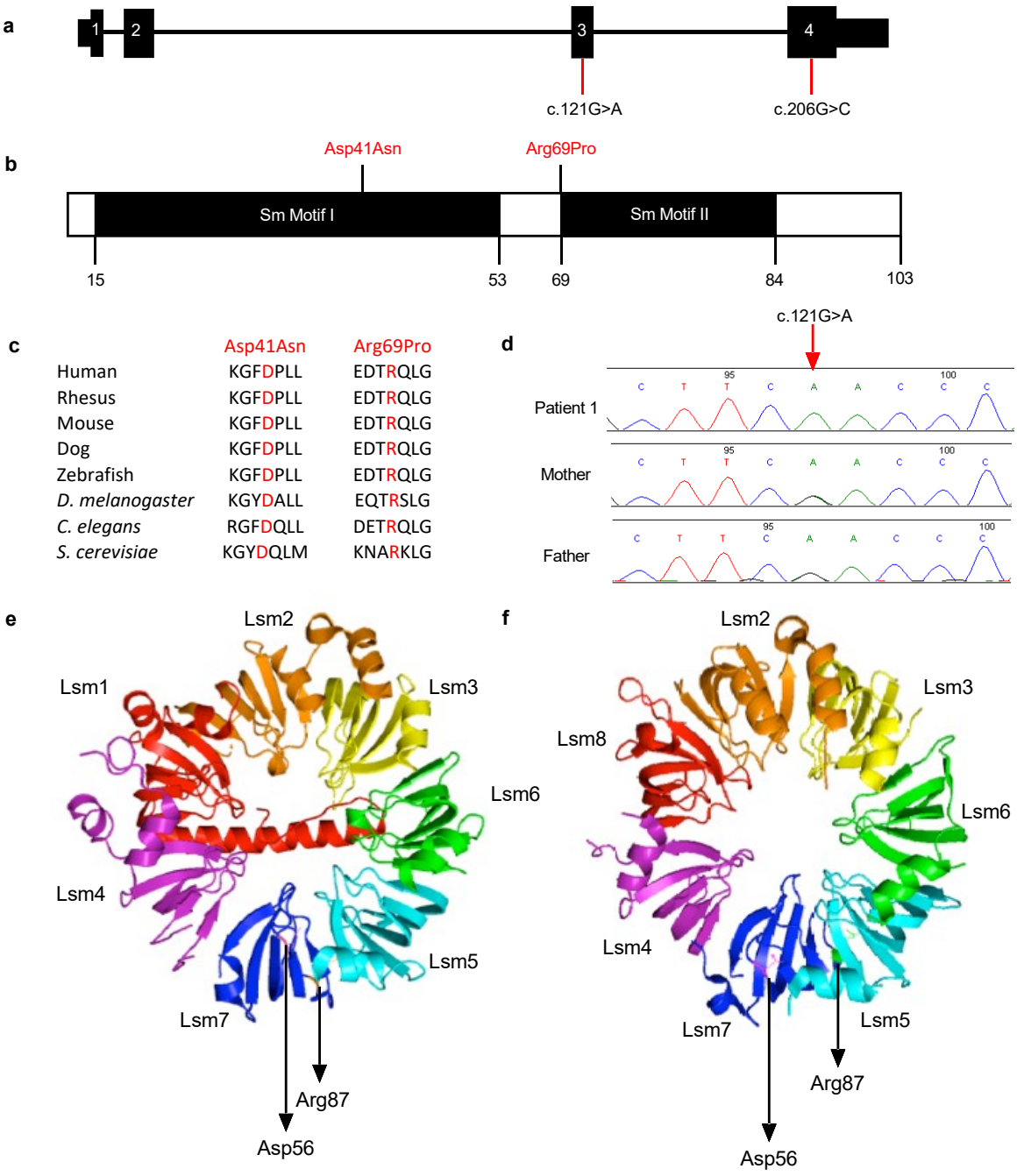



**Figure 6.2 Biallelic pathogenic variants in *LSM7* lead to a rare neurodevelopmental and neurodegenerative disorder. a,** Genomic organization of *LSM7* in humans (UCSC Genome Browser hg19) with the position of the pathogenic variants within the *LSM7* gDNA are indicated. **b,** Major motifs of the 103 amino acid LSM7 protein with position of variants indicated. **c,** *LSM7* variants in patients with leukodystrophy affect amino acids that are conserved in different species. **d,** *LSM7* gene sequencing of patient 1 and parents. **e-f,** 3D representations (created with PyMOL) of yeast **e,** LSM1-7 and f, LSM2-8 complexes displaying location of variants.

**The p.Asp41Asn variant leads to reduced LSM7 mRNA and protein levels**

We extracted RNA and protein from fibroblasts (n=6) obtained from patient 1 (p.Asp41Asn) as well as two age- and sex-matched healthy controls. RT-qPCR and Western blot revealed reduced levels of *LSM7* mRNA ($p<0.01$) and LSM7 protein ($p<0.001$) in patient 1 compared to controls (Figure 6.3a-c). Samples from patient 2 (p.Arg69Pro), who died *in utero,* were not available for analysis.

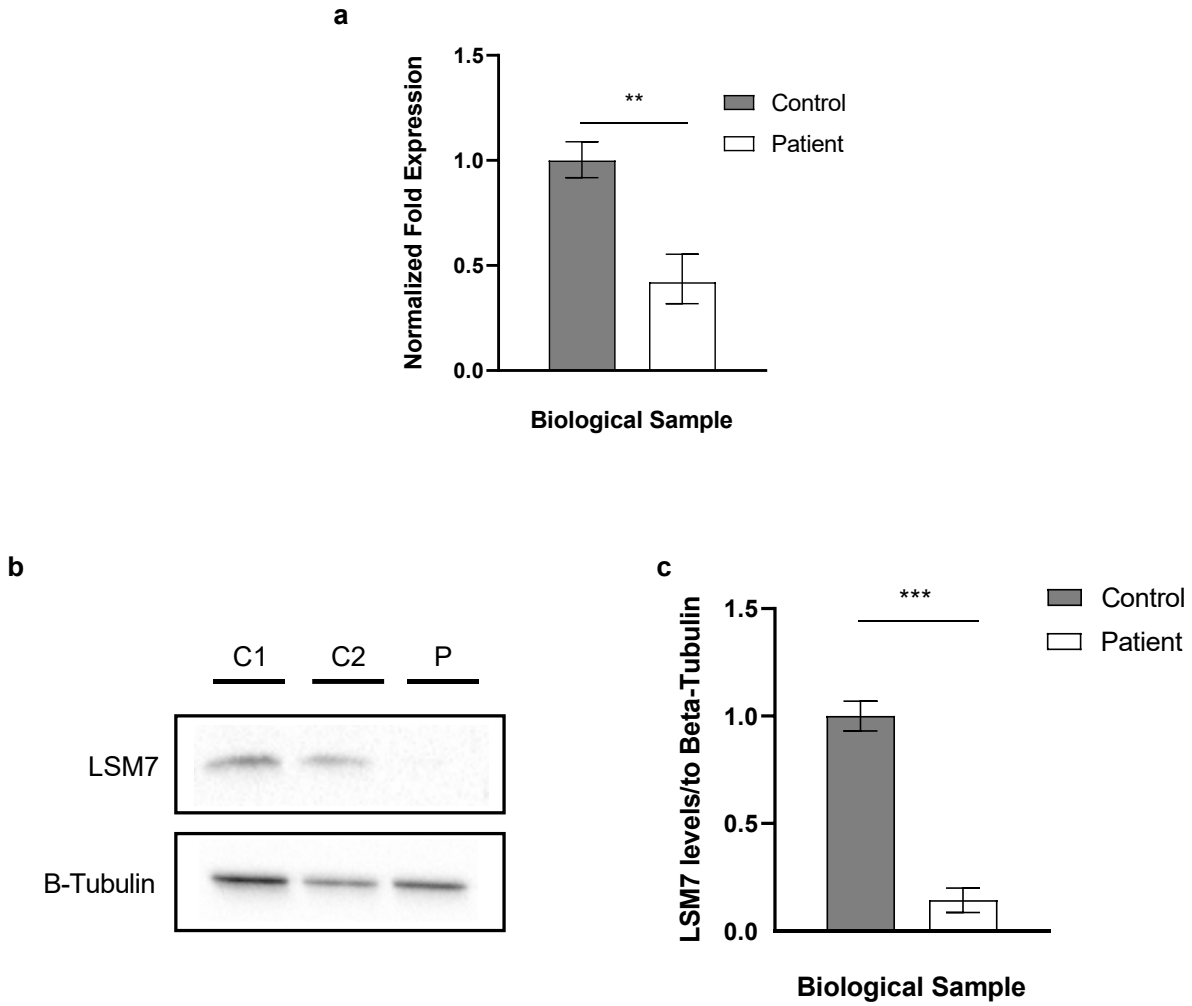



**Figure 6.3 LSM7 mRNA and protein levels are decreased in patient 1 compared to control**. **a,** RT-qPCR analysis of *LSM7* mRNA extracted from patient 1's fibroblasts and two age-sex-matched controls (n=6). The results are represented in terms of fold change after normalizing to *RPL30* and *SDHA* mRNAs. Each value represents the mean +/- SEM (unpaired *t*-test (two tailed) ***$p < 0.001$)). **b,** Total protein lysates extracted from patient 1 and age-sex-matched control fibroblasts were immunoblotted with anti-LSM7 and anti-B-tubulin. Molecular mass markers are shown on the left in kilodaltons. **c,** The pixel densities of immunoblot bands from at least 3 independent replicates were quantified using ImageJ software, normalized over β-tubulin immunoblot and represented as mean +/- SEM (unpaired *t*-test (two tailed)), ***$p<0.001$).

### *LSM7* variants are predicted to affect protein and complex structure

All the LSM proteins (LSM1 through LSM8) are highly conserved from yeast to humans. Although the crystal structure has not been resolved for the human LSM1-7 complex, yeast structures have been published for both the Lsm1-7 and Lsm2-8 complexes[25,43] (Figure 6.2e). Based on their homology, we mapped the human variants at positions p.Asp41 and p.Arg69 to their corresponding yeast residues at p.Asp56 and p.Arg87, respectively. Using PyMOL, we modeled the effects of the predicted deleterious variants, p.Asp56Asn and p.Arg87Pro in yeast in the Lsm1-7 (PDB ID 4M75) and Lsm2-8 complexes (PDB ID 4M77) (Figure 6.4a). The first variant, p.Asp56Asn in yeast, is predicted to result in the loss of a polar contacts with other Lsm7 residues in both the Lsm1-7 and Lsm2-8 complexes (Figure 6.4a-d). Meanwhile, the second variant, p.Arg87Pro in yeast, is predicted to result in the loss of polar contacts with Lsm7 residues and a residue on the neighbouring Lsm5 protein in both the Lsm1-7 and Lsm2-8 complexes (Figure 6.4e-h). Based on this modelling, we predict that each of these substitutions will lead to a compromised ability of Lsm7 to interact with the other Lsm subunits and consequently, the assembly of the Lsm1-7 and Lsm2-8 complexes will be significantly impaired.

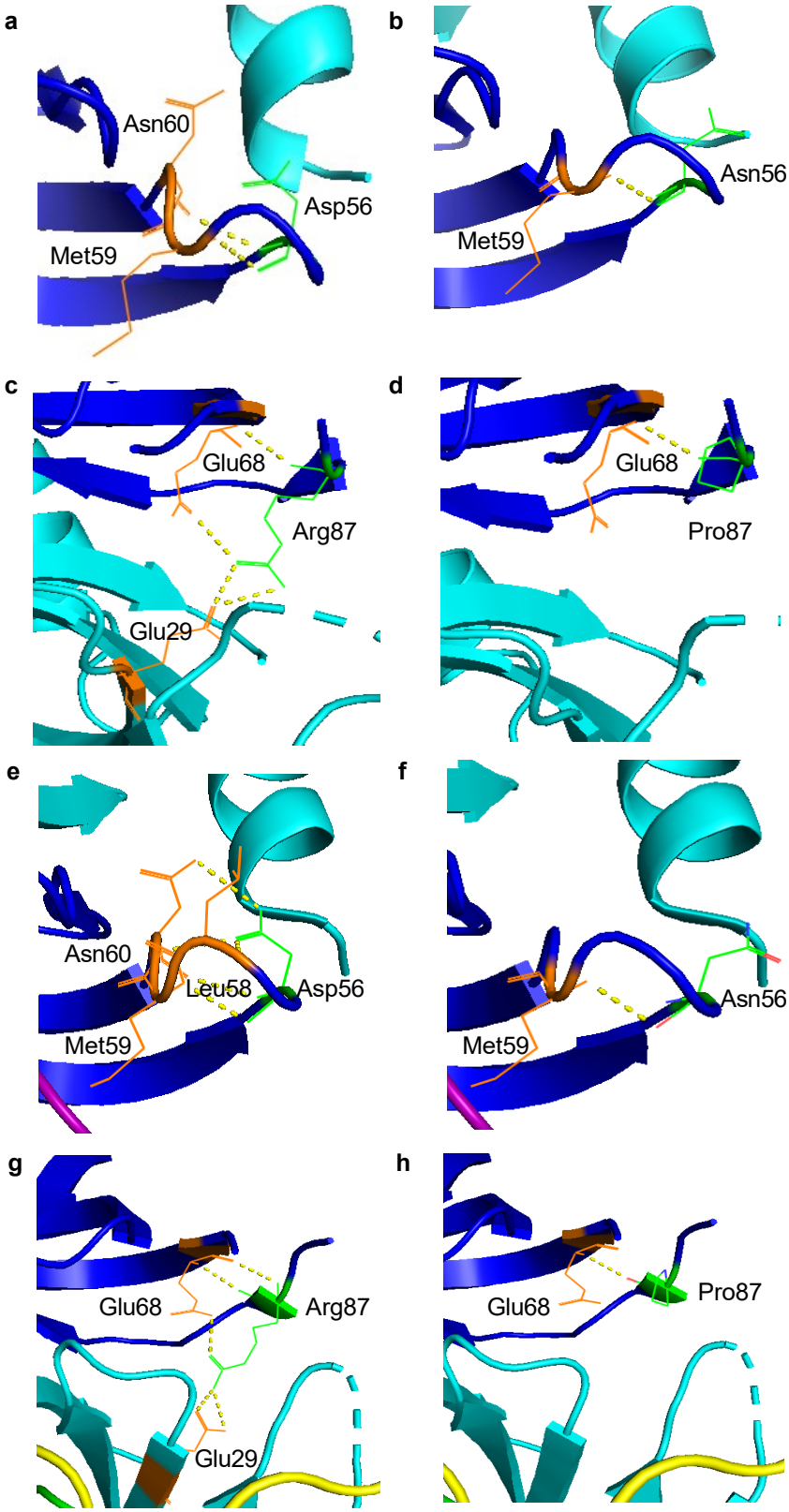

a
Asn60
Asp56
Met59
b
Asn56
Met59
c
Glu68
Arg87
Glu29
d
Glu68
Pro87
e
Asn60
Leu58
Asp56
Met59
f
Asn56
Met59
g
Glu68
Arg87
Glu29
h
Glu68
Pro87

**Figure 6.4 3D representation of predicted mutational impact on LSM1-7 and LSM2-8 structures created with PyMOL. a-d**, Pair-wise comparisons between the wild-type (left) and mutant (right) residues in LSM1-7 complex. **a**, Asp56 residue on yeast Lsm7 has polar contacts with Lsm7 residues Met59 and Asn60 in LSM1-7 complex. **b**, Asp56Asn mutant residue on yeast Lsm7 only has polar contacts with the Lsm7 residue Met59 in the LSM1-7 complex and has lost its interaction with the Asn60 residue. **c**, Arg87 residue on yeast Lsm7 has polar contacts with Lsm7 residue Glu68 and Lsm5 residue Glu29 in the LSM1-7 complex. **d**, Arg87Pro mutant residue on yeast Lsm7 only has polar contacts with Lsm7 residue E68 in the LSM1-7 complex and has lost its interaction with the Glu29 residue of Lsm5. **e-h**, Pair-wise comparisons between the wild-type (left) and mutant (right) residues in LSM2-8 complex. **e**, Asp56 residue on yeast Lsm7 has polar contacts with Lsm7 residues Leu58, Met59 and Asn60 in the LSM2-8 complex. **f**, Asp56Asn mutant residue on yeast Lsm7 only has polar contacts with Lsm7 residue Met59 in the LSM2-8 complex and has lost its interaction with Met59 and Asn60. **g**, Arg87 residue on yeast Lsm7 has polar contacts with Lsm7 residue Glu68 and Lsm5 residue Glu29 in the LSM2-8 complex. **h**, Arg87Pro mutant residue on yeast Lsm7 only has polar contacts with Lsm7 residue Glu68 in the LSM2-8 complex and has lost its interaction with the Lsm5 residue Glu29.

**p.Asp41Asn and p.Arg69Pro variants lead to decreased LSM7 interactions with its partners**

In order to determine the potential pathogenic role of these variants, we assessed their impact on the assembly of the LSM1-7 and LSM2-8 complexes in human cells. A FLAG-tagged version of the wild-type (WT) LSM7, as well as that containing either the p.Asp41Asn or p.Arg69Pro variant, were transiently expressed in HEK293 cells. Anti-FLAG affinity purification was conducted on cell extracts and purified proteins were analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) in quadruplicate. Pull down of FLAG-tagged protein from cells expressing the wild-type LSM7 protein resulted in the co-precipitation of all the LSM1-7 and LSM2-8 complexes subunits, along with multiple other known interactors of those complexes (Figure 6.5a-b). However, pull downs from lysates of HEK293 cells expressing the p.Asp41Asn and p.Arg69Pro variants revealed dramatically reduced association of the variant LSM7 proteins with the other LSM proteins constituting the LSM1-7 and LSM2-8 complexes. Consistent with this defect, most of the known interactors of these complexes were also not pulled down in the mutant strains (Figure 6.5c). The LSM1-7 and LSM2-8 complexes are known to

interact with many mRNA decay and splicing factors, respectively. Relative to the FLAG-tagged wild type LSM7, the p.Asp41Asn and p.Arg69Pro mutants showed reduced interaction with the mRNA decay factors PATL1 and EDC4 as well as with tri-snRNP factors including PRPF3, PRPF4, SART3 and SNRNP200. These splicing factors are required for the formation of the U4/U6/U5 tri-snRNP complex, an important precursor in the major spliceosome reaction[44-46]. Similarly, PATL1 is a highly conserved interacting partner of the LSM1-7 complex that plays a key role in mRNA decay[47]. Thus, these findings suggest that both variants lead to defects in the assembly of the LSM1-7 and LSM2-8 complexes and result in the subsequent loss of key complex interactors.

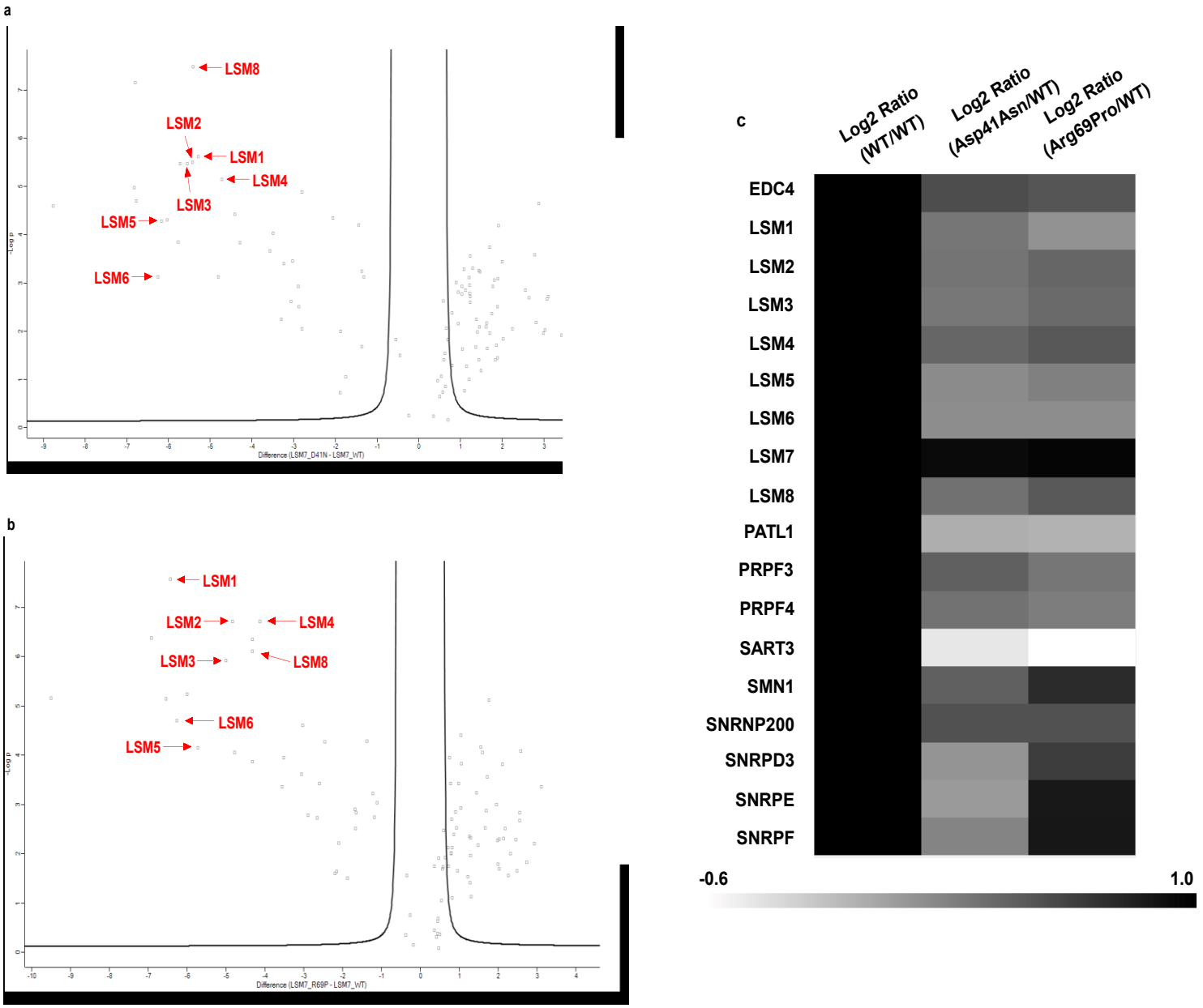



**Figure 6.5. Impact of LSM7 variants on LSM1-7 and LSM2-8 complex assembly. a-b**, Volcano plots of the $log_2$-transformed MaxQuant ratios of the p.Asp41Asn or p.Arg69Pro/Wild-type (x axis) and the -$log_{10}$- transformed p values (adjusted with permutation-based false discovery rate) resulting from the two-tailed one-sample t-tests. **c**, Heatmap contains the MaxQuant ratios for the subunits of the LSM complexes and known important complex interactors.

**Zebrafish Lsm7 shows amino acid conservation with human LSM7 and is expressed in the nervous system**

The zebrafish ortholog *lsm7* (ENSDART00000081188.7) and human *LSM7* share significant homology. The zebrafish Lsm7 protein has 95% amino acid identity with human LSM7, including conservation of the two amino acid residues sites of the variants in our patients (p.Asp41Asn and p.Arg69Pro) (Figure 6.6a). We evaluated *lsm7* gene expression from 1 day-post-fertilization (dpf) to 5 dpf. At 1 dpf and 2 dpf, the *lsm7* transcript was detected in the developing brain, spinal cord, and eye (Figure 6.6b). By 3 dpf and persisting through 5 dpf, expression is present in olfactory bulb, cerebellum, and hindbrain (Figure 6.6b).

**Zebrafish Lsm7 is necessary for nervous system development**

In order to investigate the physiological function of Lsm7 in an animal model of neurodevelopment, we used CRISPR-Cas9 gene editing to develop Lsm7-Crispant zebrafish lacking functional Lsm7 protein. Single guide RNAs (sgRNAs) targeting exon 3 of *lsm7* were designed for this purpose. Co-injection of *lsm7* sgRNA and Cas9 protein efficiently induced insertions and deletions as determined by high resolution melting analysis (HRMA) PCR (Supplemental Figure 6.1).

Embryos injected with *lsm7* sgRNA and Cas9 protein showed impaired nervous system development and low lethality (1.6%, n=128). We categorized the surviving embryos into two phenotypes according to the severity: the mild phenotype displayed a normal trunk morphology with brain and eye size reduction while the moderate phenotype exhibited a bent trunk, edema in the pericardial sac, and decreased eye and brain size (Figure 6.6c-e).

We quantified effects on eye size by measuring the horizontal axis (Figure 6.6e, dashed line) in Lsm7-Crispant (injected with *lsm7* sgRNA and Cas9 protein) larvae at 5 dpf. Eye diameter

was found to be significantly reduced in Lsm7-Crispants ($p<0.001$) (Figure 6.6e-f). Interestingly, the eye phenotype could be rescued by concomitant injection of human $LSM7^{full\text{-}length}$ and $LSM7^{Arg69Pro}$ cRNAs but not the $LSM7^{Asp41Asn}$ cRNA (Figure 6.6e-f).

We evaluated motor behaviour of Lsm7-Crispants by analyzing response to light stimuli at 7 dpf. Animals were adapted to a behaviour recording chamber for 30 minutes, and then recorded for 2 minutes, after which a single white light flash was provided for stimulus, and behaviour was recorded for an additional minute. The response to light stimuli was significantly reduced in the Lsm7-Crispants ($p<0.001$) and could be rescued by co-injection of the $LSM7^{full\text{-}length}$ cRNA, but not by co-injection with either $LSM7^{Asp41Asn}$ or $LSM7^{Arg69Pro}$ cRNAs (Figure 6.6g).

To further characterize the observed neurological defects in Lsm7-Crispants at the cellular level, we analyzed cell death, as well as neuron and oligodendrocyte generation in Lsm7-Crispants. We used the transgenic line Tg(*olig2*:*dsRed*) to label oligodendrocytes, oligodendrocyte precursor cells (OPCs) and motor neurons, and compared different cell numbers in larvae injected with wild-type $LSM7^{full\text{-}length}$ cRNA, in comparison to $LSM7^{Asp41Asn}$ or $LSM7^{Arg69Pro}$ cRNAs. Apoptosis of oligodendrocytes was significantly increased in Lsm7-Crispants ($p<0.01$) (Figure 6.6h-i) and could be rescued by concomitant injection of $LSM7^{full\text{-}length}$ cRNA, but not the $LSM7^{Asp41Asn}$ or $LSM7^{Arg69Pro}$ cRNAs (Figure 6.6h-i).

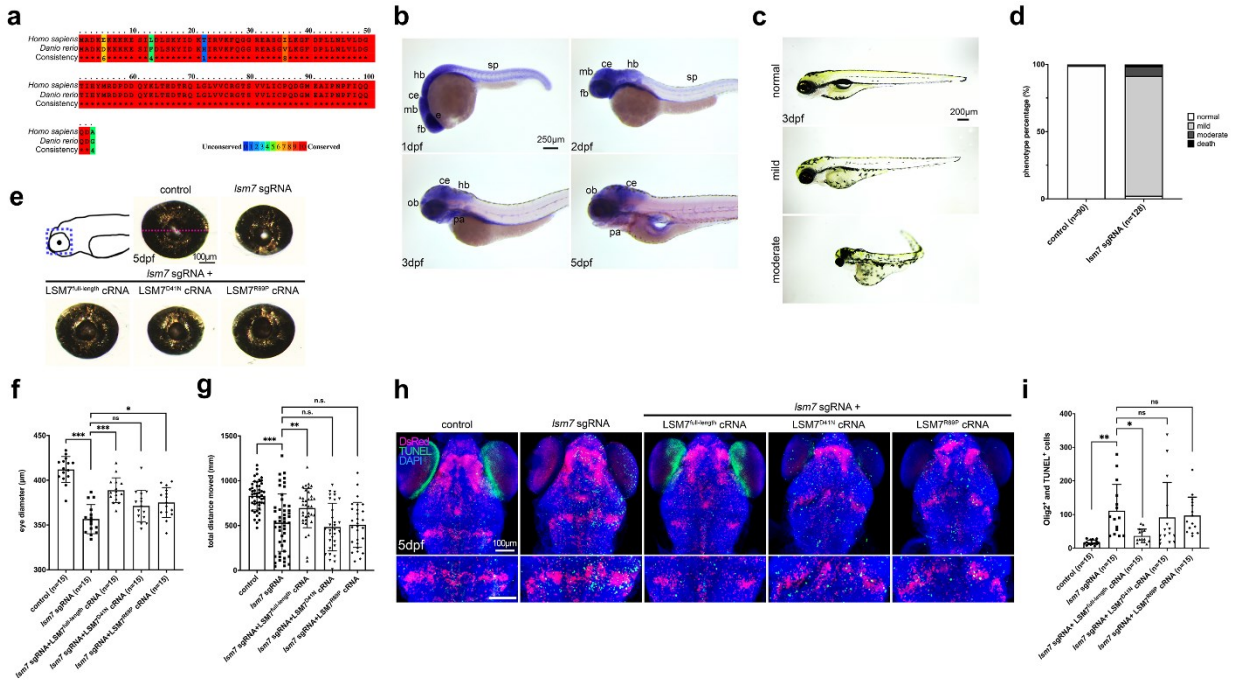


**Figure 6.6 The zebrafish *Lsm7* ortholog is highly conserved with human LSM7, is expressed in developing nervous system, and is necessary for embryonic morphology, eye size, motor behaviour, and survival of oligodendrocytes.** **a**, Amino acid alignment of human and zebrafish LSM7 sequences. Residues identical in all proteins are marked by red boxes, and the percent similarity is shown by different colors. **b**, In situ hybridization was performed for detecting *lsm7* expression during zebrafish embryogenesis. Embryo stages are shown in the bottom left of each panel. All panels display the lateral view, with dorsal to the top and rostral to the left. **c**, Phenotypes of *lsm7*-Crispant. **d**, Quantification of control and *lsm7*-Crispants by phenotype severity. **e**, Top left panel shows a cartoon representation of the dissected region of larvae (blue dashed line rectangle). Images of eyes of control, *lsm7*-Crispant, LSM7$^{\text{full-length}}$ cRNA-, LSM7$^{\text{R69P}}$ cRNA-, and LSM7$^{\text{D41N}}$ cRNA-rescued larvae are shown. The eye diameter was measured as indicated by the magenta dashed line shown on control eye. **f**, Quantification of eye size. **g**, Motor swimming behavior of *lsm7*-Crispant, and LSM7$^{\text{full-length}}$ cRNA-, LSM7$^{\text{R69P}}$ cRNA-, and LSM7$^{\text{D41N}}$ cRNA-rescued larvae at 7 dpf. **h**, Confocal images of dorsal view of the control, *lsm7*-Crispant, LSM7$^{\text{full-length}}$ cRNA-, LSM7$^{\text{R69P}}$ cRNA-, and LSM7$^{\text{D41N}}$ cRNA-rescued larvae with rostral to the top. Lower panels are enlarged region from upper panels for apoptotic cell quantification. **i**, Quantification of the number of TUNEL and DsRed double-positive cells of **h**, Abbreviations: ce, cerebellum; dpf, days-post-fertilization; fb, forebrain; hb, hindbrain; mb, midbrain; ob, olfactory bulb; pa, pharyngeal arches; sp, spinal cord. *$p < 0.05$; **$p < 0.01$; ***, $p < 0.001$; n.s., not significant.

## 6.5 Discussion

Historically, leukodystrophies were thought to be caused by variants in myelin- or oligodendrocyte-specific genes, until approximately 10 years ago, when a subset of leukodystrophies caused by mutations in genes important for transcription and translation were uncovered[5,9,11,14-20]. Here, we identified *LSM7,* a gene involved in the processes of 5'-3' mRNA decay and splicing[27], as the cause of an ultra-rare leukodystrophy. In recent years, gene discovery has shifted from identifying causal genes for rare diseases to ultra-rare entities with only a handful of affected individuals. Indeed, this is supported by several newly uncovered leukodystrophy genes, in which only a few patients were identified and published, including *EPRS1, POLR3K, RARS1*, *TMEM106B*, *CNP,* and *NKX6-2*[14,20,48-51].

*LSM7,* which encodes a subunit of both the LSM1-7 and LSM2-8 complexes, has never been associated with human disease. However, other proteins involved in the processes of mRNA decay and splicing, the primary functions of the LSM1-7 and LSM2-8 complexes, respectively, have been implicated in a variety of diseases[24,26,33,52-68]. Cellular decay pathways play a critical role in regulating mRNA levels and alterations in these pathways could be the principle factor underlying many diseases[57-67,69,70], including neurological disorders[55,56]. Similarly, mutations in genes encoding proteins involved in the various steps of the splicing reaction have also been linked to numerous neurodevelopmental and neurodegenerative disorders[44,71-76]. The identification and functional validation of *LSM7* adds mRNA decay and splicing regulation to the categories of cell dysfunction implicated in causing leukodystrophies, neurodevelopmental and neurodegenerative diseases.

Two major pathways of bulk mRNA decay exist in eukaryotes following the initial deadenylation stage. mRNAs can be either degraded via 3'-5' exonucleolysis by the conserved

RNA exosome (3'-5' pathway) or decapped and then degraded via 5'-3' exonucleolysis (5'-3' pathway)[59,77-79]. Interestingly, pathogenic variants in *EDC3*, an enhancer of decapping in the 5'-3' pathway, have been linked to cognitive disability[80]. Studies also show that the EDC4 protein, another enhancer of decapping, interacts directly with the LSM1-7 complex[29,81,82]. This corresponds to our LC-MS/MS experimental data which show that LSM1-7-EDC4 association is decreased in cells expressing FLAG-tagged LSM7 mutants relative to WT (Figure 6.5c). During the mRNA decay process, the LSM1-7 complex binds to the 3' ends of oligoadenylated mRNAs and facilitates their 5'- 3' decay[83-85]. Several studies support the idea that the LSM1-7 complex can also inhibit the RNA exosome mediated 3'-5' decay by binding to the 3' end of the mRNA and blocking its access to the RNA exosome[26,54,86,87]. Thus, the mRNAs that are optimal binding substrates for the LSM1-7 complex are likely to be poor targets for RNA exosome mediated decay. Therefore, in principle, disruption of the LSM1-7 complex assembly could cause deregulation of not only the 5'-3' but also 3'-5' mRNA decay pathway. This has important implications considering that disruption of the RNA exosome function can lead to neurodegenerative diseases, such as pontocerebellar hypoplasia, which is caused by pathogenic variants in genes encoding the core exosome components (*EXOSC3*, *EXOSC8* and *EXOSC9*)[69,88-91].

Autophagy is another important cellular process implicated in the pathogenesis of several neurodegenerative disorders, including leukodystrophies[92-96]. Recent studies have shown that the Lsm1-7-Pat1 complex facilitates this process by protecting and stabilizing the ATG mRNAs, which encode proteins involved in autophagy, from RNA exosome mediated 3'-5' degradation[97]. Thus, by enhancing susceptibility to the RNA exosome mediated 3'-5' decay of specific mRNAs, mutations that disable the Lsm1-7-Pat1 complex could potentially lead to the dysregulation of several cellular processes, including autophagy[97], causing leukodystrophy.

Shifting perspectives to splicing, the principle role of the LSM2-8 complex, it has been shown that several interactors of this complex, namely proteins with splicing-related functions, have been implicated in disease. In the past, extensive work has been performed to study the impact of mutations on the spliceosomal machinery and their implications in disease[44,71-73]. A prime example is spinal muscular atrophy (SMA) caused by mutations in the *SMN1* gene[30,44,72,98]. The *SMN1* gene encodes the protein SMN, which is a component of the SMN multi-protein complex required for loading the Sm ring structure onto U-snRNAs (excluding U6 and U6atac) of the major and minor spliceosomes[30,44,72,73,98,99]. This complex promotes the stabilization and maturation of snRNPs so they can effectively partake in the splicing reaction[73,100]. The SMN complex has also been found to interact directly with LSM4 and LSM6, suggesting that it may also play a role in the assembly and/or regeneration of LSM-containing complexes[30]. Interestingly, we found reduced amounts of SMN1 were pulled down in our LC-MS/MS experiments in each of the FLAG-tagged mutants compared to WT (Figure 6.5c).

Pathogenic variants in *PRPF4*, *PRPF3* and *SNRNP200*, which encode components of the tri-snRNP complexes (U4, U5, U6 of the major spliceosome, and U4atac, U5 and U6atac of the minor spliceosome) have been found to cause retinitis pigmentosa (RP)[101-104]. RP is a well-studied disease which clinically results in severe vision impairment due to progressive degeneration of the rod photoreceptor cells in the retina[44]. Our results reveal all three of these proteins (PRPF4, PRPF3 and SNRNP200*)* were pulled down in lower amounts in the two FLAG-tagged mutants compared to WT (Figure 6.5c). Other notable diseases include cerebrocostomandibular syndrome (CCMS), characterized by characterized by micrognathia, rib defects and intellectual disability[75,76], and hypotrichosis-11, a rare form of hair loss[74]. These disorders are caused by pathogenic variants in

*SNRPB* and *SNRPE,* respectively,[74-76] which both encode for Sm proteins that share many structural and functional similarities with LSM proteins[23,27,105].

Overall, hypomorphic mutations in *LSM7* could, in principle, interfere with 5'-3' decay, 3'-5' decay, autophagy and/or splicing of cellular mRNAs. The above examples highlight the importance of splicing and demonstrate how mutations in genes encoding components of the spliceosomal machinery can lead to disease. These examples offer precedence for genes related to the LSM complex causing disease and support the role of LSM7 as the cause of an ultra-rare leukodystrophy.

*In silico* modelling of *LSM7* allowed us to predict that both identified variants cause conformational changes in the structure of the yeast Lsm7 protein, and ultimately result in complex assembly defects. We confirmed that human LSM7 harbouring either of the mutated amino acids pulled down less of the other LSM proteins and their interactors, which suggests complex formation defects. Interestingly, the levels of LSM7 protein were found to be decreased in patient 1. We predict that the defect in LSM1-7 and LSM2-8 complex formation could be further exacerbated by reduced starting levels of the mutant LSM7 protein in this patient. In our LC-MS/MS experiments, when the levels of FLAG-tagged mutant LSM7 were normalized to that of the FLAG-tagged WT, we found that both mutated subunits pulled down relatively lower amounts of the other LSM proteins. Therefore, reduced protein levels of mutant LSM7, as seen in patient 1, could potentially intensify the defects we observed in our LC-MS/MS experiments. We also saw reduced levels of LSM7 mRNA in patient 1 compared to controls (Figure 6.3a). Since the levels of the LSM2-8 complex are expected to be lower in this patient (Figure 6.5c), reduced LSM7 transcript levels could result from inefficient splicing of LSM7 pre-mRNA. Several other studies have shown that both missense and synonymous variants can impact RNA structure, stability,

folding, miRNA binding, splicing regulatory sites and/or translation efficiency[106-108]. Therefore, we hypothesize that this missense variant could impact some aspect of the LSM7 pre-RNA, further targeting it for degradation, and thereby reducing transcript levels.

Since patient 1 (p.Asp41Asn) exhibited cerebellar hypoplasia, ocular abnormalities, and spastic quadriparesis, zebrafish was used as a model organism to investigate the role of LSM7 in the CNS. Unfortunately, phenotypic comparisons to patient 2 (p.Arg69Pro), were not possible as death occurred *in utero* and autopsy was not performed. Our zebrafish model revealed *lsm7* is expressed in the CNS, specifically in the developing brain, spinal cord, and eye. Expression persisted at day 5 post fertilization in the olfactory bulb, cerebellum, and hindbrain. We also found that Lsm7-Crispants had a reduced eye diameter, which could be rescued by injection with the human $LSM7^{full\text{-}length}$ and the $LSM7^{Arg69Pro}$ cRNAs but not by the $LSM7^{Asp41Asn}$ cRNA. Since the p.Asp41Asn variant is located on a conserved RNA contacting residue in the Sm motif-I of LSM7, the inability of the $LSM7^{Asp41Asn}$ cRNA to rescue the eye size phenotype of the Lsm7-Crispants implies that RNA processing activity is critical for eye development. Furthermore, movement deficits were also observed in the Lsm7-Crispants and could only be rescued with the $LSM7^{full\text{-}length}$ cRNA but not by either of the mutants cRNAs ($LSM7^{Arg69Pro}$ or $LSM7^{Asp41Asn}$). These findings demonstrate that *lsm7* knock-down zebrafish have phenotypic characteristics similar to those observed in patient 1.

On the cellular level, our zebrafish studies also aimed to explore the role of *LSM7* in oligodendrocytes, the myelin producing cells of the CNS primarily implicated in leukodystrophies. Our results revealed apoptosis of oligodendrocytes was significantly up-regulated in Lsm7-Crispants, which could be rescued by concomitant injection of the $LSM7^{full\text{-}length}$ cRNA but not by either of the mutant cRNAs. This suggests a role for *LSM7* in oligodendrocyte survival and

demonstrates that both variants are associated with an increase of apoptosis of oligodendrocytes. Interestingly, the hypoplastic cerebellum in patient 1 corresponded to the small cerebellar phenotype observed in Lsm7-Crispants, which could indicate a similar role for apoptosis in cells of the human cerebellum, and potentially a distinct role for LSM7 in human cerebellar development. We thereby hypothesize that deficiency or dysregulation of LSM7 in the cerebellum of patient 1 at key developmental stages could contribute to the observed cerebellar hypoplasia. Overall, this zebrafish model has allowed us to identify a relationship between LSM7 and CNS development, as well as oligodendrocyte survival, thus providing further evidence of association with the ultra-rare leukodystrophy observed in our patients.

Unfortunately, our results do not provide evidence as to why one missense variant (p.Arg69Pro) led to the *in utero* death of one patient while the other (p.Asp41Asn) appears to be less severe, with patient 1 currently aged 7 years. Samples were unavailable from patient 2 for mRNA and protein analysis, however we can hypothesize that the p.Arg69Pro variant may lead to splice site changes, as predicted by the *in silico* prediction tool MutationTaster[42]. The p.Arg69Pro, but not the p.Asp41Asn variant, is predicted to be located at an exon-intron boundary which could impair or cause dysfunction in pre-mRNA splicing. For this reason, we hypothesize that the p.Arg69Pro variant could result in splice site changes that are detrimental and cause embryonic lethality. Specific splicing variants can cause cell-specific defects, as in the case of patients presenting with a unique phenotype of basal ganglia abnormalities and the same pathogenic splicing variant in *POLR3A*[109-113]. This further supports a possible association between distinct splicing defects and phenotypic heterogeneity within patients with pathogenic variants in the same gene.

Although our studies provide compelling evidence of pathogenicity, the mechanism through which the variants in *LSM7* cause this ultra-rare leukodystrophy, with such tissue specific defects, remains unclear. As in many cases of genetic disorders, it is challenging to delineate genotype-phenotype relationships, and determine the pathogenic mechanisms underlying how and why mutations in certain genes cause leukodystrophies. For example, it is still not understood how mutations genes encoding POLR3 subunits, including *POLR3A, POLR3B, POLR1C* and *POLR3K*, or tRNA synthetases, including *EPRS1, RARS1*, and *VARS1*, can lead to specific myelin defects[14-20]. It has been well established that pathogenic variants in these genes cause leukodystrophies, however, their pathophysiology is yet to be uncovered. The most widely accepted hypothesis is that mutations in these POLR3 subunits result in a decrease in the levels of tRNAs which are critical for the synthesis of essential CNS proteins[14,17,114]. Similarly, it is believed that alterations in the various stages of the tRNA-synthetase pathway could lead to abnormal translation and thus the insufficient production of specific proteins at critical stages in development[14,16]. What remains unclear is why and even if these abundant CNS myelin proteins are the most affected. Thereby, the mechanisms through which these genes, which are widely accepted as causing leukodystrophies, are unfortunately still not well understood. As such, future studies are needed to shed light on the pathogenesis of these diseases, as well as how *LSM7* and many other leukodystrophy-causing genes, when mutated, lead to such tissue-specific diseases.

## 6.6 Methods

### Patient ascertainment

All experiments involving human participants or data were conducted in compliance with all relevant ethical regulations. Informed consent was obtained from all participants. Approval for human subjects' research was obtained at the McGill University Health Centre Research Institute Research Ethics Board (project number PED-11-105, 2019-4972). Patient 1 was clinically evaluated by a pediatric neurologist and a geneticist. Clinical brain MRIs were obtained for patient 1 and longitudinal assessment performed in order to determine disease course. Saliva samples were collected for DNA isolation from patient 1 and his mother and father. A dermal skin punch was also obtained from patient 1 after consent and was cultured in 10% fetal bovine serum and DMEM to obtain low passage primary fibroblasts. Consent for clinical next generation sequencing of both parents of patient 2 was obtained at the King Faisal Specialist Hospital and Research Centre in Riyadh, Saudi Arabia.

### Whole exome sequencing and variant identification

Clinical whole exome sequencing was performed on affected patient 1, as a trio with biological parents by GeneDx. The exonic regions and flanking splice junctions of the genome were captured using a proprietary system developed by GeneDx and sequenced by massive parallel (NextGen) sequencing on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg 19 and analyzed for sequence variants using a custom developed analysis tool (Xome Analyzer). Additional sequencing technology and variant interpretation protocols have been previously described[115]. Capillary sequencing or another appropriate method was used to confirm all potentially pathogenic variants identified in patient 1

and their relative samples. The general assertion criteria for variant classification are publicly available on GeneDx ClinVar submission page (http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957). A second analysis of the sequence variants was conducted in-house at the Research Institute of the McGill University Health Centre. Sequence variants were prioritized using the standard and guidelines for interpretation set out by the American College of Medical Genetics (ACMG)[116]. Clinical whole exome sequencing was performed on the parents of patient 2 as previously described[117].

**Cell culture**

Fibroblasts from patient 1 and two age- and sex-matched controls were grown in DMEM supplemented with 10% fetal bovine serum and maintained at 37°C and 5% $CO_2$. Each cell line was expanded and harvested at similar passages at least three consecutive times for use in molecular experiments. HEK293T cells were grown in DMEM media supplemented with 10% fetal bovine serum, 2mM glutamine and 1% penicillin-streptomycin and maintained at 37°C and 5% $CO_2$.

**RT-qPCR**

RNA was extracted from fibroblasts using TRIzol reagent according to manufacturer's instructions (Invitrogen). For each sample, 1µg of total RNA was reverse transcribed using M-MLV reverse transcriptase (Promega) and random hexamers (Invitrogen) according to manufacturer's protocol (Promega). Primers for RT-qPCR were designed and efficiency tested according to the MIQE guidelines[118,119]. Primers were as followed for LSM7: forward 5'-GAAGCCAGTGGAATCCTGAAGG-3'; LSM7 reverse: 5'-CCTCCGTGAGCTTGTACTGG-

3’; U6 snRNA forward: 5’-CTCGCTTCGGCAGCACA-3’,; U6 snRNA reverse: 5’-AACGCTTCACGAATTTGCGT-3’; RPL30 forward: 5’-AAGGCAGGAAGATGGTGGCC-3’; RPL30 reverse: 5’-GAGTCTGCTTGTACCCCAGGAC-3’; and SDHA forward: 5’-CAGCATGTGTTACCAAGCTG-3’; SDHA reverse: 5’-GGTGTCGTAGAAATGCCAAC-3’. Reverse transcription was performed in triplicate with a 1:4 dilution of cDNA using SsoAdvanced Universal SYBR Green Supermix (Biorad) on the Roche LightCycler 96. All data were normalized to *SDHA* or *RPL30* using the ΔΔCt method in accordance with the MIQE guidelines[118,119].

**Western blot**

Total protein was extracted from fibroblasts in ice cold RIPA buffer with protease inhibitors (Roche) and spun down for 20 minutes at maximum speed. Protein concentration was determined using the Bradford protein assay in cuvettes according to manufacturer’s instructions (Biorad). 30ug of protein were separated on a 15% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane using the Bio-Rad Trans-Blot Turbo Transfer System. The membranes were first blocked with 1X *TBST* with *5*% w/v non-fat dry *milk and then probed with primary antibodies* against LSM7 (Abcam 241656, dilution 1:10,000) and against beta-tubulin (Abcam 6046, dilution 1:20,000) diluted in primary antibody buffer overnight at 4℃. The next day, a secondary polyclonal goat-anti rabbit IgG (H+L) (Novus biologicals, 1:5,000) antibody was used in 5% non-fat milk in TBST and membranes were subsequently visualized with Amersham ECL Western Blot Detecting Reagent according to manufacturer’s protocol (GE Life Sciences).

**Construction of Lsm1-7 and Lsm2-8 structural models**

Structural models of the wild-type and mutant yeast Lsm1-7 (PDB ID 4M75) and Lsm2-8 complexes (PDB ID 4M77) were generated using PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC (http://www.pymol.org).

**Protein affinity purification coupled to mass spectrometry**

Plasmids for 3xFLAG-tagged LSM7 wild-type and the two mutants (p.Asp41Asn and p.Arg69Pro) were generated using the p3xFLAG-CMV-14 plasmid (Sigma). 0.4ug of the plasmids were transfected into HEK293 cells seeded at 80% confluence in 6-well plates using jetPRIME DNA transfection reagent kit according to manufacturer's protocol (Polyplus-Transfection). The HEK293 cells were maintained in culture in DMEM media supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin and 1% L-glutamine. Forty-eight hours after transfection, protein lysates were collected using a gentle lysis buffer to preserve protein-protein interactions and quantified in a 96-well plate using Bradford Reagent according to manufacturer's protocol (LifeTechnologies). Equal amounts of protein were purified using FLAG M2 magnetic beads according to standard procedures in 4 independent replicate experiments[120,121]. The protein extracts were dried down in a speed-vac before being resolubilized and digested with trypsin. The tryptic peptides were then purified and identified using tandem mass spectrometry (LC-MS/MS) with an HPLC system coupled to Orbitrap fusion mass spectrometer (Thermo Scientific) through a Nanospray Flex Ion Source. Protein database searching was performed using MaxQuant version 1.6.10.43[122-127] against the SwissProt human protein database downloaded on 4 April 2019. Default MaxQuant parameters were modified as follows: trypsin was used as the digestion enzyme, LFQ computation and "Match between runs" were activated. Known AP-MS protein

contaminants including keratins were excluded. Protein intensities were analysed using Perseus version 1.6.10.43[128-131]. Proteins marked by MaxQuant as being "Only identified by site" and "Reverse" were excluded. Proteins not present in at least 3 out of 4 replicates of WT, Asp41Asn, or Arg69Pro were excluded. Missing LFQ values were replaced by normally distributed values with a mean downshifted by 1.8 and a standard deviation of 0.3 times the non-missing values. WT, Asp41Asn, and Arg69Pro proteins were compared against FLAG empty vector control samples and were labeled as high-confidence interactors when their adjusted p-values was under 0.05 and their intensity ratio was over 2. P-values were corrected using permutated based approaching using an s0 correction factor of 0.1 with 10,000 iterations[132]. A Student's t-test was performed between Asp41Asn and Arg69Pro against the WT samples. P-values were corrected using a permutated based approach using an s0 correction factor of 0.1 with 10,000 iterations. For the generation of the heatmap, log2 ratios were computed between the means of the WT, Asp41Asn, and Arg69Pro against WT for known interactors of LSM7.

**Zebrafish ethics statement**

Zebrafish experiments were performed in strict accordance of guidelines from the University of Utah Institutional Animal Care and Use Committee (IACUC), regulated under federal law (the Animal Welfare Act and Public Health Services Regulation Act) by the U.S. Department of Agriculture (USDA) and the Office of Laboratory Animal Welfare at the NIH, and accredited by the Association for Assessment and Accreditation of Laboratory Care International (AAALAC).

**Fish stocks and embryo raising**

Adult fish were bred according to standard methods. Embryos were raised at 28.5°C in E3 embryo medium and staged by time and morphology. For *in situ* staining and immunohistochemistry, embryos were fixed in 4% paraformaldehyde (PFA) in 1x PBS overnight at 4°C, washed briefly in 1xPBS with 0.1% Tween-20, serially dehydrated, and stored in 100% MeOH at -20°C until use. Transgenic fish line and alleles used in this paper were the following: Tg(*olig2:dsRed)*$^{vu19}$ [133].

***Lsm7* sequence analysis and construct generation**

Human and zebrafish Lsm7 amino acid sequences were aligned using PRALINE (http://www.ibi.vu.nl/programs/pralinewww/). The construct for the *lsm7* riboprobe was amplified by PCR using Phusion® High-Fidelity DNA Polymerases (New England Biolabs, NEB) with primers: *lsm7* forward (5'-GGATCCGCCGCCACCATGGCGGACAAAGACAAGAAAAAGAAGGAG-3') and *lsm7* reverse (5'- CTCAAGGCGTCCAGTAACCAAAGCCC-3').

The constructs of human LSM7$^{full\text{-}length}$, LSM7$^{D41N}$, and LSM7$^{R69P}$ for capped RNA synthesis were amplified from p3XFlag-CMV-LSM7$^{full\text{-}length}$, p3XFlag-CMV-LSM7$^{D41N}$, and p3XFlag-CMV-LSM7$^{R69P}$, respectively, by PCR using Phusion® High-Fidelity DNA Polymerases with the following primers: LSM7 forward (5'-GAATTCGCCGCCACCATGGCGGATAAGGAGAAGAAGAAAAAG-3') and LSM7 reverse (5'- GAATTCCTACTTGTCATCGTCATCCTTGTAGTC -3'), and then subcloned into pCS2$^{+}$ vector.

### *lsm7* CRISPR sgRNA construction and injection

We designed sgRNA target sites for the zebrafish *lsm7* gene (Ensembl Zv11: ENSDART00000081188.7) by looking for sequences corresponding to $GGN_{18}nGG$ on the sense or antisense strand of the DNA using the CRISPR design program CHOPCHOP (http://chopchop.cbu.uib.no/)[134]. Off-target effects were checked with NIH BLAST tool applied to the zebrafish genome (zv11). Off-target sequences that had significant matches of the final 23 nt of the target and NGG PAM sequence were discarded.

DNA template for *lsm7 exon3* sgRNA was amplified by PCR using Phusion® High-Fidelity DNA Polymerases with *lsm7 exon3* specific forward primer (5'-GAAATTAATACGACTCACTATACCCGTTGTTGAATCTTGTGTGTTTTAGAGCTAGAA ATAGC -3') and universal reverse primer (5'-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAA CTTGCTATTTCTAGCTCTAAAAC -3'). The PCR program used for template synthesis was 98°C 30 seconds, 35 cycles of [98°C 10 seconds, 60°C 30 seconds, 72°C 15 seconds], 72°C 10 minutes, 10°C end. Subsequently, the DNA template was purified by QIAquick PCR Purification Kit (Qiagen). The *lsm7 exon3* sgRNA was synthesized by the HiScribe T7 RNA Synthesis kit (BioLabs) followed by RNA purification with RNA Clean & Concentrator-5 (Zymo research).

The *lsm7* exon3 sgRNA (460 pg) and Cas9 protein (460 pg, PNA BIO) were injected into one cell stage embryos. CRISPR efficiency was evaluated on individual 24 hpf injected embryos after DNA extraction, PCR amplification of the target locus, and HRMA analysis. A *lsm7* exon3 sgRNA dose of 460 pg resulted in >90% mutagenesis in 24 hpf embryos, assayed by high-resolution melt analysis (Xing et al., 2014). Embryos used for injection were derived from either wild-type AB parents or Tg(*olig2:dsRed)*$^{vu19}$ as mentioned in the text.

**Capped RNA synthesis**

Capped RNA encoding the full coding sequence of LSM7$^{full\text{-}length}$, LSM7$^{Asp41Asn}$, or LSM7$^{Arg69Pro}$ were prepared as per manufacturer's instructions and purified with Micro Bio-spin 6 columns (BioRad). Capped RNAs (460 pg) were injected for rescue experiments.

**Immunohistochemistry and in situ hybridization**

Immunohistochemistry was performed as previously described[135]. Antibodies used were rabbit anti-dsRed 1:250 (Clontech), Alexa 555 goat anti-rabbit 1:400 (ThermoFisher Scientific), and 4',6-diamidino-2-phenylindole (DAPI). The antisense digoxigenin-UTP labeled riboprobe for detecting *lsm7* transcript was synthesized according to manufacturer's instructions (Roche), and *in situ* hybridizations were performed as described previously[136]. The color reaction was carried out using NBT/BCIP substrate (Roche).

**Morphological analysis**

5 dpf fish larvae were fixed in 4% PFA in 1x PBS overnight at 4°C, washed briefly in 1xPBS with 0.1% Tween-20. One eye of an individual larva was dissected out and mounted laterally. Eyes were imaged, and the length was measured from anterior to posterior along the longest axis by ImageJ software.

**Behaviour Analysis**

Larval behaviour analysis was performed on 7 dpf larvae in 96-well square bottom plates (Krackeler Scientific) using video analysis software (Noldus EthoVision). For spontaneous behaviour, animals were transferred at 6 dpf to a 96-well plate and kept at 28.5˚C overnight. At 7

dpf the plate was placed on the video imaging system and animals were allowed to adapt in the dark for 10 minutes, and then recording was performed for 5 minutes (1 minute dark and 4 minutes light).

**Microscopy and image analysis**

Immunostained embryos were transferred step-wise into 90% glycerol/10% PBS, mounted on a glass slide with a #0 coverslip, and imaged on a confocal microscope. Confocal stacks were projected in ImageJ, and images composed with Adobe Photoshop and Illustrator.

**TUNEL quantification**

Terminal deoxynucleotidyl transferase dUTP nick-end labeling (TUNEL) was performed on whole-mount larvae (ApopTag Fluorescein *In Situ* Apoptosis Detection Kit; Millipore) as previously described[137]. Confocal imaging was performed, and images were rendered in ImageJ by compiling an average intensity projection of 120 μm (step size 5 μm) into a single z-stack image, for cell counting using Photoshop's (Adobe) count tool.

**Statistical analysis**

qPCR data were analyzed following the MIQE guidelines and expressed as average relative normalized expression with appropriate error bars (SEM). Normality of all data were assessed using the Shapiro-Wilks Test of Normality and the equality of variances were assessed using Levene's Test. Changes in LSM7 mRNA levels in a patient compared to controls were analyzed using a parametric, unpaired *t*-test (two-tailed). Changes in the levels of U6 snRNA in a patient compared to controls were analyzed with the parametric, unpaired Welch's *t*-test (two-tailed)

rather than a standard *t*-test as the equality of variances assumption was violated. Changes in LSM7 protein levels in a patient compared to controls were analyzed using a parametric, unpaired *t*-test (two tailed). Statistical significance was set at $p < 0.05$. Statistical analyses for zebrafish experiments were performed using Prism8 software (GraphPad). Student's *t*-test was used for two-way comparisons; comparisons between three or more groups was performed with ANOVA with post-hoc Tukey's HSD between individual means.

## 6.7 Acknowledgements

First and foremost, we would like to thank the patients and their families for their participation. This study was supported by grants from the Canadian Institute of Health Research (CIHR)(…), Fondation Les Amis d'Elliot, Fondation le Tout pour Loo and Réseau de Médicine Génétique Appliquée of the Fonds de Recherche du Québec – Santé. This research was enabled in part by support provided by Compute Canada (www.computecanada.ca). A. Derksen is supported by the Canadian Institutes of Health Research (CIHR) Canadian Graduates Scholarships – Master's, the Fondation du Grand Defi Pierre Lavoie Master's Scholarship, and Heathy Brains for Healthy Lives Masters Fellowship. G. Bernard has received a Research Scholar Junior 1 award from the Fonds de Recherche du Quebec – Sante (FRQS) (2012-2016) and the New Investigator Salary Award from the Canadian Institutes of Health Research (2017-2022). B. Coulombe holds a Bell-Bombardier Research Chair awarded by the IRCM. We would also like to thank the McGill University and Genome Quebec Innovation Center, the IRCM Molecular Biology and Functional Genomics Platform and Denis Faubert, Josee Champagne, and Sylvain Tessier of the IRCM Proteomics Discovery Platform for their services.

## 6.8 Author Contributions

A.D. performed the experiments, collected and analyzed the data and wrote the first draft of the manuscript. H.Y.S. performed the zebrafish experiments, collected and analyzed the data and helped in the writing of the first draft of the manuscript. D.F. helped in the design of the LC-MS/MS experiments and reviewed the manuscript. L.D. helped in the design of the molecular experiments and reviewed the manuscript. L.T. helped in the collection of clinical information and in the review of the manuscript. C.P. performed the statistical analyses of the LC-MS/MS data and helped with the generation of figures. M.S.G. helped in the design and analysis of the LC-MS/MS experiments and reviewed the manuscript. K.G. helped in the design and molecular experiments. G.B., C.N., A.L., F.A., and W.K. recruited the patients, contributed to their phenotyping and reviewed the manuscript. S.T. provided guidance and expertise in the field. J.L.B. designed and supervised the zebrafish experiments, helped draft manuscript, and reviewed the manuscript. B.C. designed and supervised the functional LC-MS/MS studies and reviewed the manuscript. G.B. designed and supervised the study and reviewed the manuscript.

## 6.9 Competing Interests

Y.S. is an employee of GeneDx, Inc. H.Y.S reports no conflicts of interest. JLB reports board of director for wFluidx; stock ownership of Orchard Therapeutics; consulting for Bluebird, Calico, Denali, Enzyvant, Neurogene, and Passage Bio; and royalties from BioFire (spouse).

## 6.10 References


1 Vanderver, A. *et al.* Case definition and classification of leukodystrophies and leukoencephalopathies. *Mol Genet Metab* **114**, 494-500, doi:10.1016/j.ymgme.2015.01.006 (2015).

2 Parikh, S. *et al.* A clinical approach to the diagnosis of patients with leukodystrophies and genetic leukoencephelopathies. *Mol Genet Metab* **114**, 501-515, doi:10.1016/j.ymgme.2014.12.434 (2015).

3 Vanderver, A., Tonduti, D., Schiffmann, R., Schmidt, J. & van der Knaap, M. S. in *GeneReviews* (eds M. P. Adam *et al.*) (2014).

4 Ashrafi, M. R. & Tavasoli, A. R. Childhood leukodystrophies: A literature review of updates on new definitions, classification, diagnostic approach and management. *Brain Dev* **39**, 369-385, doi:10.1016/j.braindev.2017.01.001 (2017).

5 Schiffmann, R. & van der Knaap, M. S. Invited article: an MRI-based approach to the diagnosis of white matter disorders. *Neurology* **72**, 750-759, doi:10.1212/01.wnl.0000343049.00540.c8 (2009).

6 Steenweg, M. E. *et al.* Magnetic resonance imaging pattern recognition in hypomyelinating disorders. *Brain* **133**, 2971-2982, doi:10.1093/brain/awq257 (2010).

7 Bonkowsky, J. L. *et al.* The burden of inherited leukodystrophies in children. *Neurology* **75**, 718-725, doi:10.1212/WNL.0b013e3181eee46b (2010).

8 Adang, L. A. *et al.* Revised consensus statement on the preventive and symptomatic care of patients with leukodystrophies. *Mol Genet Metab* **122**, 18-32, doi:10.1016/j.ymgme.2017.08.006 (2017).

9 van der Knaap, M. S. & Bugiani, M. Leukodystrophies: a proposed classification system based on pathological changes and pathogenetic mechanisms. *Acta Neuropathol* **134**, 351-382, doi:10.1007/s00401-017-1739-1 (2017).

10 Vanderver, A. *et al.* Whole exome sequencing in patients with white matter abnormalities. *Ann Neurol* **79**, 1031-1037, doi:10.1002/ana.24650 (2016).

11 Kevelam, S. H. *et al.* Update on Leukodystrophies: A Historical Perspective and Adapted Definition. *Neuropediatrics* **47**, 349-354, doi:10.1055/s-0036-1588020 (2016).

12 Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* **14**, 681-691, doi:10.1038/nrg3555 (2013).

13 Sawyer, S. L. *et al.* Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet* **89**, 275-284, doi:10.1111/cge.12654 (2016).

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

15 Rezaei, Z. *et al.* Hypomyelinating Leukodystrophy with Spinal Cord Involvement Caused by a Novel Variant in RARS: Report of Two Unrelated Patients. *Neuropediatrics*, doi:10.1055/s-0039-1679911 (2019).

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

17 Thiffault, I. *et al.* Recessive mutations in POLR1C cause a leukodystrophy by impairing biogenesis of RNA polymerase III. *Nat Commun* **6**, 7623, doi:10.1038/ncomms8623 (2015).
18 Bernard, G. *et al.* Mutations of POLR3A encoding a catalytic subunit of RNA polymerase Pol III cause a recessive hypomyelinating leukodystrophy. *Am J Hum Genet* **89**, 415-423, doi:10.1016/j.ajhg.2011.07.014 (2011).
19 Tetreault, M. *et al.* Recessive mutations in POLR3B, encoding the second largest subunit of Pol III, cause a rare hypomyelinating leukodystrophy. *Am J Hum Genet* **89**, 652-655, doi:10.1016/j.ajhg.2011.10.006 (2011).
20 Dorboz, I. *et al.* Mutation in POLR3K causes hypomyelinating leukodystrophy and abnormal ribosomal RNA regulation. *Neurol Genet* **4**, e289, doi:10.1212/NXG.0000000000000289 (2018).
21 Sundaresan, T. Roles of eukaryotic Lsm proteins in the regulation of mRNA function. *Int Rev Cell Mol Biol* **272**, 149-189, doi:10.1016/S1937-6448(08)01604-3 (2009).
22 Sundaresan, T. Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay? *RNA Biol* **6**, 228-232 (2009).
23 Beggs, J. D. Lsm proteins and RNA processing. *Biochem Soc Trans* **33**, 433-438, doi:10.1042/BST0330433 (2005).
24 Kufel, J., Bousquet-Antonelli, C., Beggs, J. D. & Tollervey, D. Nuclear pre-mRNA decapping and 5' degradation in yeast require the Lsm2-8p complex. *Mol Cell Biol* **24**, 9646-9657, doi:10.1128/MCB.24.21.9646-9657.2004 (2004).
25 Sharif, H. & Conti, E. Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. *Cell Rep* **5**, 283-291, doi:10.1016/j.celrep.2013.10.004 (2013).
26 Tharun, S. Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay? *RNA Biol* **6**, 228-232 (2009).
27 Tharun, S. Roles of eukaryotic Lsm proteins in the regulation of mRNA function. *Int Rev Cell Mol Biol* **272**, 149-189, doi:10.1016/S1937-6448(08)01604-3 (2009).
28 Chowdhury, A., Mukhopadhyay, J. & Tharun, S. The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* **13**, 998-1016, doi:10.1261/rna.502507 (2007).
29 Wu, D. *et al.* Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. *Cell Res* **24**, 233-246, doi:10.1038/cr.2013.152 cr2013152 [pii] (2014).
30 Friesen, W. J. & Dreyfuss, G. Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN). *J Biol Chem* **275**, 26370-26375, doi:10.1074/jbc.M003299200 (2000).
31 Will, C. L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**, doi:10.1101/cshperspect.a003707 (2011).
32 Chen, W. & Moore, M. J. Spliceosomes. *Curr Biol* **25**, R181-183, doi:10.1016/j.cub.2014.11.059 (2015).
33 Verdone, L., Galardi, S., Page, D. & Beggs, J. D. Lsm proteins promote regeneration of pre-mRNA splicing activity. *Curr Biol* **14**, 1487-1491, doi:10.1016/j.cub.2004.08.032 (2004).
34 Wilusz, C. J. & Wilusz, J. Lsm proteins and Hfq: Life at the 3' end. *RNA Biol* **10**, 592-601, doi:10.4161/rna.23695 (2013).

35 Sobreira, N., Schiettecatte, F., Boehm, C., Valle, D. & Hamosh, A. New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. *Hum Mutat* **36**, 425-431, doi:10.1002/humu.22769 (2015).
36 Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat* **36**, 928-930, doi:10.1002/humu.22844 (2015).
37 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
38 Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812-3814, doi:10.1093/nar/gkg509 (2003).
39 Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745-2747, doi:10.1093/bioinformatics/btv195 (2015).
40 Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-315, doi:10.1038/ng.2892 (2014).
41 Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894, doi:10.1093/nar/gky1016 (2019).
42 Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **11**, 361-362, doi:10.1038/nmeth.2890 (2014).
43 Zhou, L. *et al.* Crystal structure and biochemical analysis of the heptameric Lsm1-7 complex. *Cell Res* **24**, 497-500, doi:10.1038/cr.2014.18 (2014).
44 Vazquez-Arango, P. & O'Reilly, D. Variant snRNPs: New players within the spliceosome system. *RNA Biol* **15**, 17-25, doi:10.1080/15476286.2017.1373238 (2018).
45 Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**, 108-121, doi:10.1038/nrm3742 (2014).
46 Didychuk, A. L., Butcher, S. E. & Brow, D. A. The life of U6 small nuclear RNA, from cradle to grave. *RNA* **24**, 437-460, doi:10.1261/rna.065136.117 (2018).
47 Vindry, C. *et al.* Dual RNA Processing Roles of Pat1b via Cytoplasmic Lsm1-7 and Nuclear Lsm2-8 Complexes. *Cell Rep* **20**, 1187-1200, doi:10.1016/j.celrep.2017.06.091 (2017).
48 Ito, Y. *et al.* Lysosomal dysfunction in TMEM106B hypomyelinating leukodystrophy. *Neurol Genet* **4**, e288, doi:10.1212/NXG.0000000000000288 (2018).
49 Wolf, N. I. *et al.* Mutations in RARS cause hypomyelination. *Ann Neurol* **76**, 134-139, doi:10.1002/ana.24167 (2014).
50 Al-Abdi, L. *et al.* CNP deficiency causes severe hypomyelinating leukodystrophy in humans. *Hum Genet* **139**, 615-622, doi:10.1007/s00439-020-02144-4 (2020).
51 Chelban, V. *et al.* Mutations in NKX6-2 Cause Progressive Spastic Ataxia and Hypomyelination. *Am J Hum Genet* **100**, 969-977, doi:10.1016/j.ajhg.2017.05.009 (2017).
52 Kufel, J., Allmang, C., Petfalski, E., Beggs, J. & Tollervey, D. Lsm Proteins are required for normal processing and stability of ribosomal RNAs. *J Biol Chem* **278**, 2147-2156, doi:10.1074/jbc.M208856200 (2003).

53 Ingelfinger, D., Arndt-Jovin, D. J., Luhrmann, R. & Achsel, T. The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. *RNA* **8**, 1489-1501 (2002).
54 He, W. & Parker, R. The yeast cytoplasmic LsmI/Pat1p complex protects mRNA 3' termini from partial degradation. *Genetics* **158**, 1445-1455 (2001).
55 Hollams, E. M., Giles, K. M., Thomson, A. M. & Leedman, P. J. MRNA stability and the control of gene expression: implications for human disease. *Neurochem Res* **27**, 957-980, doi:10.1023/a:1020992418511 (2002).
56 Linder, B., Fischer, U. & Gehring, N. H. mRNA metabolism and neuronal disease. *FEBS Lett* **589**, 1598-1606, doi:10.1016/j.febslet.2015.04.052 (2015).
57 Vlasova-St Louis, I., Dickson, A. M., Bohjanen, P. R. & Wilusz, C. J. CELFish ways to modulate mRNA decay. *Biochim Biophys Acta* **1829**, 695-707, doi:10.1016/j.bbagrm.2013.01.001 (2013).
58 Eberhardt, W., Doller, A., Akool el, S. & Pfeilschifter, J. Modulation of mRNA stability as a novel therapeutic approach. *Pharmacol Ther* **114**, 56-73, doi:10.1016/j.pharmthera.2007.01.002 (2007).
59 Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* **13**, 246-259, doi:10.1038/nrg3160 (2012).
60 Cheneval, D., Kastelic, T., Fuerst, P. & Parker, C. N. A review of methods to monitor the modulation of mRNA stability: a novel approach to drug discovery and therapeutic intervention. *J Biomol Screen* **15**, 609-622, doi:10.1177/1087057110365897 (2010).
61 Anderson, P. Post-transcriptional regulons coordinate the initiation and resolution of inflammation. *Nat Rev Immunol* **10**, 24-35, doi:10.1038/nri2685 (2010).
62 Frischmeyer, P. A. & Dietz, H. C. Nonsense-mediated mRNA decay in health and disease. *Hum Mol Genet* **8**, 1893-1900, doi:10.1093/hmg/8.10.1893 (1999).
63 Benjamin, D. & Moroni, C. mRNA stability and cancer: an emerging link? *Expert Opin Biol Ther* **7**, 1515-1529, doi:10.1517/14712598.7.10.1515 (2007).
64 Khajavi, M., Inoue, K. & Lupski, J. R. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. *Eur J Hum Genet* **14**, 1074-1081, doi:10.1038/sj.ejhg.5201649 (2006).
65 Miller, J. N. & Pearce, D. A. Nonsense-mediated decay in genetic disease: friend or foe? *Mutat Res Rev Mutat Res* **762**, 52-64, doi:10.1016/j.mrrev.2014.05.001 (2014).
66 Pashler, A. L., Towler, B. P., Jones, C. I. & Newbury, S. F. The roles of the exoribonucleases DIS3L2 and XRN1 in human disease. *Biochem Soc Trans* **44**, 1377-1384, doi:10.1042/BST20160107 (2016).
67 Morita, M. *et al.* Hepatic posttranscriptional network comprised of CCR4-NOT deadenylase and FGF21 maintains systemic metabolic homeostasis. *Proc Natl Acad Sci U S A* **116**, 7973-7981, doi:10.1073/pnas.1816023116 (2019).
68 Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A. & Skotheim, R. I. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **35**, 2413-2427, doi:10.1038/onc.2015.318 (2016).
69 Weskamp, K. & Barmada, S. J. RNA Degradation in Neurodegenerative Disease. *Adv Neurobiol* **20**, 103-142, doi:10.1007/978-3-319-89689-2_5 (2018).
70 Kapur, M., Monaghan, C. E. & Ackerman, S. L. Regulation of mRNA Translation in Neurons-A Matter of Life and Death. *Neuron* **96**, 616-637, doi:10.1016/j.neuron.2017.09.057 (2017).

71 Padgett, R. A. New connections between splicing and human disease. *Trends Genet* **28**, 147-154, doi:10.1016/j.tig.2012.01.001 (2012).
72 Faustino, N. A. & Cooper, T. A. Pre-mRNA splicing and human disease. *Genes Dev* **17**, 419-437, doi:10.1101/gad.1048803 (2003).
73 Jutzi, D., Akinyi, M. V., Mechtersheimer, J., Frilander, M. J. & Ruepp, M. D. The emerging role of minor intron splicing in neurological disorders. *Cell Stress* **2**, 40-54, doi:10.15698/cst2018.03.126 (2018).
74 Pasternack, S. M. *et al.* Mutations in SNRPE, which encodes a core protein of the spliceosome, cause autosomal-dominant hypotrichosis simplex. *Am J Hum Genet* **92**, 81-87, doi:10.1016/j.ajhg.2012.10.022 (2013).
75 Lynch, D. C. *et al.* Disrupted auto-regulation of the spliceosomal gene SNRPB causes cerebro-costo-mandibular syndrome. *Nat Commun* **5**, 4483, doi:10.1038/ncomms5483 (2014).
76 Bacrot, S. *et al.* Mutations in SNRPB, encoding components of the core splicing machinery, cause cerebro-costo-mandibular syndrome. *Hum Mutat* **36**, 187-190, doi:10.1002/humu.22729 (2015).
77 Balagopal, V., Fluch, L. & Nissan, T. Ways and means of eukaryotic mRNA decay. *Biochim Biophys Acta* **1819**, 593-603, doi:10.1016/j.bbagrm.2012.01.001 S1874-9399(12)00007-7 [pii] (2012).
78 Wu, X. & Brewer, G. The regulation of mRNA stability in mammalian cells: 2.0. *Gene* **500**, 10-21, doi:10.1016/j.gene.2012.03.021 (2012).
79 Mugridge, J. S., Coller, J. & Gross, J. D. Structural and molecular mechanisms for the control of eukaryotic 5'-3' mRNA decay. *Nat Struct Mol Biol* **25**, 1077-1085, doi:10.1038/s41594-018-0164-z (2018).
80 Ahmed, I. *et al.* Mutations in DCPS and EDC3 in autosomal recessive intellectual disability indicate a crucial role for mRNA decapping in neurodevelopment. *Hum Mol Genet* **24**, 3172-3180, doi:10.1093/hmg/ddv069 (2015).
81 Tritschler, F. *et al.* A divergent Sm fold in EDC3 proteins mediates DCP1 binding and P-body targeting. *Mol Cell Biol* **27**, 8600-8611, doi:10.1128/MCB.01506-07 (2007).
82 Siwaszek, A., Ukleja, M. & Dziembowski, A. Proteins involved in the degradation of cytoplasmic mRNA in the major eukaryotic model systems. *RNA Biol* **11**, 1122-1136, doi:10.4161/rna.34406 (2014).
83 Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat Struct Mol Biol* **20**, 127-133, doi:10.1038/nsmb.2468 nsmb.2468 [pii] (2013).
84 Tharun, S. & Parker, R. Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol Cell* **8**, 1075-1083, doi:10.1016/s1097-2765(01)00395-1 (2001).
85 Mullen, T. E. & Marzluff, W. F. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev* **22**, 50-65, doi:10.1101/gad.1622708 (2008).
86 Tharun, S., Muhlrad, D., Chowdhury, A. & Parker, R. Mutations in the Saccharomyces cerevisiae LSM1 gene that affect mRNA decapping and 3' end protection. *Genetics* **170**, 33-46, doi:10.1534/genetics.104.034322 (2005).
87 Song, M. G. & Kiledjian, M. 3' Terminal oligo U-tract-mediated stimulation of decapping. *Rna* **13**, 2356-2365 (2007).

88 Burns, D. T. *et al.* Variants in EXOSC9 Disrupt the RNA Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy. *Am J Hum Genet* **102**, 858-873, doi:10.1016/j.ajhg.2018.03.011 (2018).
89 Ryan, M. M., Cooke-Yarborough, C. M., Procopis, P. G. & Ouvrier, R. A. Anterior horn cell disease and olivopontocerebellar hypoplasia. *Pediatr Neurol* **23**, 180-184, doi:10.1016/s0887-8994(00)00166-1 (2000).
90 Boczonadi, V. *et al.* EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. *Nat Commun* **5**, 4287, doi:10.1038/ncomms5287 (2014).
91 Fasken, M. B. *et al.* The RNA Exosome and Human Disease. *Methods Mol Biol* **2062**, 3-33, doi:10.1007/978-1-4939-9822-7_1 (2020).
92 Lin, D. S. *et al.* Impairment of Proteasome and Autophagy Underlying the Pathogenesis of Leukodystrophy. *Cells* **9**, doi:10.3390/cells9051124 (2020).
93 Del Grosso, A. *et al.* Dysregulated autophagy as a new aspect of the molecular pathogenesis of Krabbe disease. *Neurobiology of disease* **129**, 195-207, doi:10.1016/j.nbd.2019.05.011 (2019).
94 Del Grosso, A. *et al.* Lithium improves cell viability in psychosine-treated MO3.13 human oligodendrocyte cell line via autophagy activation. *J Neurosci Res* **94**, 1246-1260, doi:10.1002/jnr.23910 (2016).
95 Ebrahimi-Fakhari, D., Wahlster, L., Hoffmann, G. F. & Kölker, S. Emerging role of autophagy in pediatric neurodegenerative and neurometabolic diseases. *Pediatric research* **75**, 217-226, doi:10.1038/pr.2013.185 (2014).
96 Karnati, R., Talla, V., Peterson, K. & Laurie, G. W. Lacritin and other autophagy associated proteins in ocular surface health. *Experimental eye research* **144**, 4-13, doi:10.1016/j.exer.2015.08.015 (2016).
97 Gatica, D. *et al.* The Pat1-Lsm Complex Stabilizes ATG mRNA during Nitrogen Starvation-Induced Autophagy. *Mol Cell* **73**, 314-324 e314, doi:10.1016/j.molcel.2018.11.002 (2019).
98 Pellizzoni, L., Charroux, B. & Dreyfuss, G. SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. *Proc Natl Acad Sci U S A* **96**, 11167-11172, doi:10.1073/pnas.96.20.11167 (1999).
99 Verma, B., Akinyi, M. V., Norppa, A. J. & Frilander, M. J. Minor spliceosome and disease. *Semin Cell Dev Biol* **79**, 103-112, doi:10.1016/j.semcdb.2017.09.036 (2018).
100 Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777-793, doi:10.1016/j.cell.2009.02.011 (2009).
101 Liu, T. *et al.* A novel missense SNRNP200 mutation associated with autosomal dominant retinitis pigmentosa in a Chinese family. *PLoS One* **7**, e45464, doi:10.1371/journal.pone.0045464 (2012).
102 Li, N., Mei, H., MacDonald, I. M., Jiao, X. & Hejtmancik, J. F. Mutations in ASCC3L1 on 2q11.2 are associated with autosomal dominant retinitis pigmentosa in a Chinese family. *Invest Ophthalmol Vis Sci* **51**, 1036-1043, doi:10.1167/iovs.09-3725 (2010).
103 Chen, X. *et al.* PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum Mol Genet* **23**, 2926-2939, doi:10.1093/hmg/ddu005 (2014).
104 Chakarova, C. F. *et al.* Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet* **11**, 87-92, doi:10.1093/hmg/11.1.87 (2002).

105 Achsel, T., Stark, H. & Luhrmann, R. The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proc Natl Acad Sci U S A* **98**, 3685-3689, doi:10.1073/pnas.071033998 (2001).
106 Sabarinathan, R. *et al.* RNAsnp: efficient detection of local RNA secondary structure changes induced by SNPs. *Hum Mutat* **34**, 546-556, doi:10.1002/humu.22273 (2013).
107 Sharma, Y. *et al.* A pan-cancer analysis of synonymous mutations. *Nat Commun* **10**, 2569, doi:10.1038/s41467-019-10489-2 (2019).
108 Salari, R., Kimchi-Sarfaty, C., Gottesman, M. M. & Przytycka, T. M. Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: application to disease studies. *Nucleic Acids Res* **41**, 44-53, doi:10.1093/nar/gks1009 (2013).
109 Minnerop, M. *et al.* Hypomorphic mutations in POLR3A are a frequent cause of sporadic and recessive spastic ataxia. *Brain : a journal of neurology* **140**, 1561-1578, doi:10.1093/brain/awx095 (2017).
110 Harting, I. *et al.* POLR3A variants with striatal involvement and extrapyramidal movement disorder. *Neurogenetics*, 1-13 (2020).
111 Hiraide, T. *et al.* POLR3A variants in striatal involvement without diffuse hypomyelination. *Brain and Development* (2020).
112 Perrier, S. *et al.* Expanding the phenotypic and molecular spectrum of RNA polymerase III–related leukodystrophy. *Neurology Genetics* **6**, e425, doi:10.1212/nxg.0000000000000425 (2020).
113 Wu, S. *et al.* Novel mutations of the POLR3A gene caused POLR3-related leukodystrophy in a Chinese family: a case report. *BMC Pediatrics* **19**, 289, doi:10.1186/s12887-019-1656-7 (2019).
114 Cayami, F. K. *et al.* 4H Leukodystrophy: Lessons from 3T Imaging. *Neuropediatrics* **49**, 112-117, doi:10.1055/s-0037-1608780 (2018).
115 Retterer, K. *et al.* Clinical application of whole-exome sequencing across clinical indications. *Genet Med* **18**, 696-704, doi:10.1038/gim.2015.148 (2016).
116 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).
117 Monies, D. *et al.* Lessons Learned from Large-Scale, First-Tier Clinical Exome Sequencing in a Highly Consanguineous Population. *Am J Hum Genet* **105**, 879, doi:10.1016/j.ajhg.2019.09.019 (2019).
118 Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M. & Nguyen, M. A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. *Methods* **50**, S1-5, doi:10.1016/j.ymeth.2010.01.005 (2010).
119 Taylor, S. C. & Mrkusich, E. M. The state of RT-quantitative PCR: firsthand observations of implementation of minimum information for the publication of quantitative real-time PCR experiments (MIQE). *J Mol Microbiol Biotechnol* **24**, 46-52, doi:10.1159/000356189 (2014).
120 Chen, G. I. & Gingras, A. C. Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods* **42**, 298-305, doi:10.1016/j.ymeth.2007.02.018 (2007).

121 Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature methods* **10**, 730-736, doi:10.1038/nmeth.2557 (2013).
122 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).
123 Cox, J., Michalski, A. & Mann, M. Software lock mass by two-dimensional minimization of peptide mass errors. *J Am Soc Mass Spectrom* **22**, 1373-1380, doi:10.1007/s13361-011-0142-8 (2011).
124 Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526, doi:10.1074/mcp.M113.031591 (2014).
125 Schaab, C., Geiger, T., Stoehr, G., Cox, J. & Mann, M. Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol Cell Proteomics* **11**, M111 014068, doi:10.1074/mcp.M111.014068 (2012).
126 Tyanova, S. *et al.* Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453-1456, doi:10.1002/pmic.201400449 (2015).
127 Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319, doi:10.1038/nprot.2016.136 (2016).
128 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740, doi:10.1038/nmeth.3901 (2016).
129 Rudolph, J. D. & Cox, J. A Network Module for the Perseus Software for Computational Proteomics Facilitates Proteome Interaction Graph Analysis. *J Proteome Res* **18**, 2052-2064, doi:10.1021/acs.jproteome.8b00927 (2019).
130 Tyanova, S. *et al.* Proteomic maps of breast cancer subtypes. *Nat Commun* **7**, 10259, doi:10.1038/ncomms10259 (2016).
131 Cox, J. & Mann, M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13 Suppl 16**, S12, doi:10.1186/1471-2105-13-S16-S12 (2012).
132 Choquet, K. *et al.* The leukodystrophy mutation Polr3b R103H causes homozygote mouse embryonic lethality and impairs RNA polymerase III biogenesis. *Mol Brain* **12**, 59, doi:10.1186/s13041-019-0479-7 (2019).
133 Kucenas, S. *et al.* CNS-derived glia ensheath peripheral nerves and mediate motor root development. *Nat Neurosci* **11**, 143-151, doi:10.1038/nn2025 (2008).
134 Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171-W174, doi:10.1093/nar/gkz365 (2019).
135 Bonkowsky, J. L. *et al.* Domain-specific regulation of foxP2 CNS expression by lef1. *BMC Dev Biol* **8**, 103, doi:10.1186/1471-213X-8-103 (2008).
136 Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* **3**, 59-69, doi:10.1038/nprot.2007.514 (2008).
137 Lambert, A. M., Bonkowsky, J. L. & Masino, M. A. The conserved dopaminergic diencephalospinal tract mediates vertebrate locomotor development in zebrafish larvae. *J Neurosci* **32**, 13488-13500, doi:10.1523/JNEUROSCI.1638-12.2012 (2012).

## 6.11 Supplemental

**Table 6.1 *In silico* analysis using damage prediction algorithms and highest minor allele frequency in GnomAD**

| Variant | Mutation Taster | SIFT | CADD | Provean | PolyPhen2 | MAF (GnomAD) |
|---|---|---|---|---|---|---|
| c.121G>A (p.Asp41Asn) | Disease Causing | Deleterious | Deleterious | Deleterious | Probably Damaging | 3.87e-5 |
| c.206G>C (p.Arg69Pro) | Disease Causing | Deleterious | Deleterious | Deleterious | Probably Damaging | Absent |

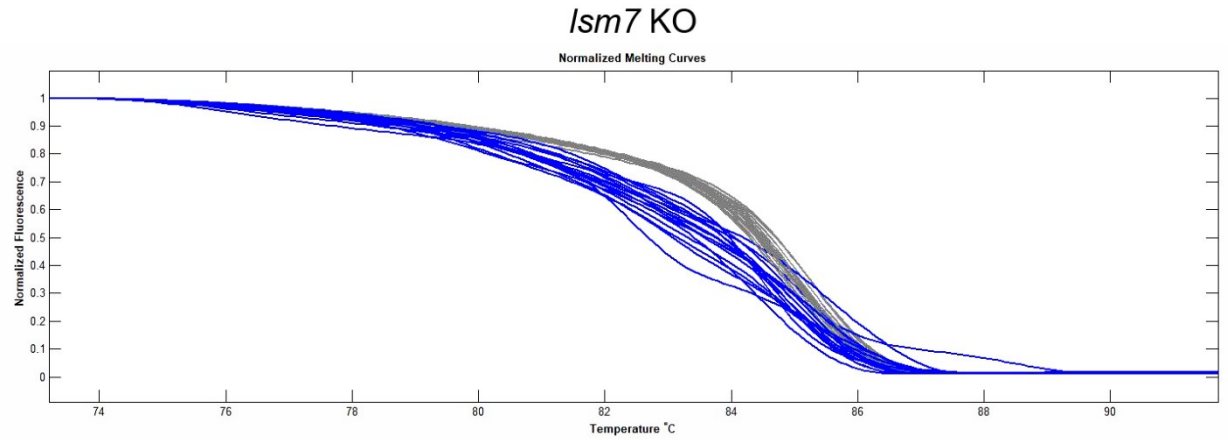



**Figure 6.1 Lsm7 sgRNA could efficiently induce insertion and/or deletion formation of lsm7 loci.** HRMA analysis of wild-type embryos and Lsm7-Crispants. Gray curves, wild-type; blue curves, Lsm7-Crispants. X-axis, melt temperature (˚C); y-axis, normalized change in fluorescence with temperature.

# Chapter 7: Discussion and Conclusion

## 7.0 Preface

This chapter acts as a general discussion for the results that were not discussed in either of the manuscripts. The first section of this chapter considers the results from Project 1, the whole exome sequencing (WES) portion of this thesis and discusses the advantages and limitations of this technique. The second section explores the results from the RT-qPCR experiments measuring U6 snRNA and discusses future experiments to study the role of *LSM7*. Finally, this chapter concludes up with a short conclusion detailing the major findings of this thesis.

## 7.1 WES Analysis Discussion

Prior to the advent of next generation sequencing, molecular testing in a research setting relied primarily on single or multi-gene panel testing[13,95,96]. To date, in the province of Quebec, these single and multi-gene panel methods are still the preferred first line of testing because of their low sequencing cost, short turnaround time, and low rate of unspecific or incidental findings[95,96]. Unfortunately, upwards of 50% of these patients remain without a molecular diagnosis and as a result many undergo long, costly, and invasive diagnostic odysseys[13,96,97]. For those cases that remain unsolved, whole exome sequencing (WES) is the logical next step[95,96].

In general, WES analysis is performed following a negative single or multi-gene panel test in cases that are genetically heterogeneous but associated with a clinical diagnosis or in cases where there is an atypical presentation of a known clinically recognized disease[8]. The goal of these next generation sequencing techniques is to increase the solve rate and decrease the time from testing to genetic diagnosis, thereby shortening the diagnostic odyssey. This has proved to be

successful, particularly in our patient group of interest, as studies have shown that WES allowed for a molecular diagnosis in 70-80%of patients with white matter abnormalities[9,44].

In this thesis, 12 of the 18 cases (67%) analyzed were second analyses performed at least 6 months after the initial analysis. We solved one case, patient 8, who had an identical twin sister, whose case was not sent for WES. It was found that both these twins had a novel *de novo* mutation in a known disease-causing gene, *DYNC1H1.* Four of the 18 (22%) remaining cases (patients 3, 4, 11, and 12) had a strong candidate gene however, this gene has not previously been associated with any human disease and as such, clinical validation and functional testing will need to be performed to demonstrate its pathogenicity. Six of the other 18 cases (33%) revealed variants of unknown significance, 3 of which have been placed on the matching platform GeneMatcher[98,99]. Through this platform we hope to be matched with clinicians and researchers from across the globe who report variants in the same genes with overlapping phenotypes. The remaining seven (7/18) patients (38%) in this cohort are still unsolved. For these cases, we propose running the data files through a new pipeline that has been developed at the Research Institute McGill University Health Centre (RI-MUHC) and re-analyzing them. If patients remain without any strong candidate gene, we will send them for WGS and RNA sequencing.

The current major debate is whether one should perform WES or WGS[95]. There are indeed limitations to WES as only about 1% of the genome, or the coding portion is covered[12,13,47]. Additionally, WES is unable to detect certain types of variants, including copy number variants, translocations, and inversions which are frequently found to cause disease[54,100,101]. There is also the risk that certain protein coding regions may not be covered due to incomplete annotation[54,95,100]. WGS addresses many of these limitations, and by covering the entire genome it can pick up variants that would have otherwise been absent in WES[54,95]. Some studies argue that the main

advantage of WGS is not that it provides coverage of non-coding portions but that it provides more complete exomic coverage, that is, it is simply better than WES[54,95]. However, even with these claims the success rates of identification for both WES and WGS remains similar at between 30-50% for the first analysis[95,96,102-107]. Furthermore, there are indeed several limitations to WGS, first and foremost being its cost which has been found to be upwards of three times as expensive as WES[47]. Another major drawback of this method is the time required for the analysis of WGS data which is significantly longer than that of WES due to the upwards of 4 million variants that can be identified[47].

## 7.2 LSM7 General Discussion

The levels of U6 snRNA, which relies on the LSM2-8 complex as a chaperone to maintain its stability throughout the numerous rearrangements of the splicing reaction was found to be increased in patient GB114.0 (p.Asp41Asn) compared to age-sex-matched healthy controls[74,79]. This result was unexpected as the levels of U6 snRNA have been found to be reduced in yeast with mutant Lsm2 through Lsm8 proteins[78]. Thus, why the levels of U6 snRNA were increased in this patient's cells when one of the components of its chaperoning complex was found to be depleted was puzzling. Interestingly, one study found that overexpression of U6 snRNA is able to rescue some of the defects seen in Lsm1 overexpressing cells which are normally U6 deficient[108]. It has also been shown that the overproduction of U6 snRNA can to some extent compensate for the loss of the Lsm7 in yeast[78]. This suggests that perhaps the overexpression of U6 snRNA may act as a compensatory mechanism in cells with reduced levels of LSM7, as is the case in the patient with the p.Asp41Asn variant. Another potential explanation for this difference could be that in yeast there is a single genomic locus for U6 snRNA, the *SNR6* gene on chromosome XII while in humans

there are at least 4 U6 genes which encode identical RNAs that are transcriptionally active to various degrees[82]. The presence of multiple U6 genes in humans with varying levels of transcriptional activity complicates their study and could potentially explain our results[82]. Furthermore, human mature U6 snRNA possesses a γ-monomethyl phosphate 5′ modification, whose function could be important for stability[82]. It is not clear if yeast U6 snRNA possesses a cap potentially making it inheritably less stable than its human counterparts[82]. A final reason for why our results differ from what is reported in the literature for yeast could be that there are tissue-specific differences in U6 expression levels and thus the levels of U6 snRNA could differ depending on the tissue type selected for study[82]. Overall, however, we report the novel finding that the levels U6 snRNA are increased in the patient with the p.Asp41Asn variant compared to healthy controls, perhaps as a compensatory measure.

Additional functional testing will be useful in the greater understanding of the pathogenesis of this *LSM7*-associated disease and for the eventual development of therapeutic strategies. Splicing assays, are one such example and can be conducted in order to determine the effects, if any the mutants have on the splicing abilities of both the major and minor spliceosome targets[109-111]. Such splicing assays could allow for the study of the major and minor spliceosome assembly, as well as for the determination of abnormal splicing events which could cause human disease[109-111]. As previously described the LSM2-8 complex acts as a chaperone and stabilizes U6 snRNA. Since we found that both mutants led to defects in the assembly of this complex, it should be determined if the limited LSM2-8 complex that is formed is sufficient to support the splicing reaction. Alternatively, there may be another complex, one in which perhaps LSM7 is replaced by another LSM protein that functions in supporting the rearrangements of U6 snRNA. It is also possible that the LSM2-8 complex is not required for, but enhances the splicing reaction, and that

perhaps other compensatory factors such as the overexpression of U6 snRNA or the presence of other complexes can make up for its decreased levels. These are just several potential theories that require follow up experiments in order to determine the true effects that these variants have on splicing within the cell.

As LSM7 is a member of both the LSM1-7 and LSM2-8 complexes the effect of these variants on the LSM1-7 complex should also be studied in more depth. Such experiments have been performed in yeast models and use the stable mRNA decay reporter *MFA2pG* to determine the impact of various Lsm variants on the cytoplasmic 5' to 3' mRNA decay[61,81]. We suggest performing similar experiments using our disease-causing variants by either mutating the yeast genome at the conserved residues of interest or if possible transfecting Lsm7Δ yeast strains with human *LSM7* plasmids bearing the variants of interest. These experiments would allow us to analyze the 5' to 3' mRNA decay via the reporter, *MFA2pG* in wild type and variant containing yeast and provide insight into how these disease-causing variants affect the function of the LSM1-7 complex.

The LC-MS/MS results as discussed in the LSM7 manuscript in chapter 6 showed that FLAG-tagged mutant LSM7 proteins had significantly reduced association with the other LSM proteins constituting the LSM1-7 and LSM2-8 complexes as well as their known interactors. While these results are significant, they do not reveal whether or not the LSM1-7 and LSM2-8 complexes are intact. In order to determine this, 2-dimensional blue native/sodium dodecyl sulfate-polyacrylamide gel electrophoresis (2D BN/SDS-PAGE) could be used[112,113]. The BN-PAGE technique allows for the separation of multiprotein complexes in their native conformation from cellular lysates with high resolution. By using a combination of BN- in the first dimension and SDS-PAGE in the second dimension, it is possible to further subdivide the multiprotein complexes into their individual components[112,113]. Such an experiment would allow for the determination of whether the mutant LSM

complexes are indeed intact. Unfortunately, this suggested experiment would only be able to be performed on patient 1 as no cells from patient 2, who died *in utero* are available.

Other future experiments that could reveal valuable information include localization studies. That is, using staining to determine the location of the LSM7 mutant proteins in the cell. Co-staining with LSM1 and LSM8 could also reveal the location of the LSM1-7 and LSM2-8 complexes, respectively. Any differences detected in the location of the mutant protein and/or complexes compared to wild type could provide important information surrounding the mechanism through which these mutants may cause disease. Finally, we also propose using cells that are perhaps more vulnerable to *LSM7* mutations, namely, neurons and/or oligodendrocytes. Creating induced pluripotent stem cells from patient fibroblasts and using them to carry out functional studies such as myelination assays could allow us to gain a better understanding of how these variants lead to the detrimental effects seen in the central nervous system (CNS).

## 7.3 Concluding Remarks

In this thesis we used WES to identify the molecular etiology of the patients with rare genetically determined leukoencephalopathies or other rare genetic diseases. 78% of these cases were 2$^{nd}$ or 3$^{rd}$ analyses conducted 6-12 months following a previous negative analysis. We were able to solve one case and find strong candidate genes in 4 additional cases (22%). These numbers give a solved rate of approximately 27% which exceeds the solved rate for re-analyses reported in the literature (5-11%)[49-52].

In this thesis, we demonstrated that biallelic variants in *LSM7* cause an ultra-rare neurodevelopmental and neurodegenerative disease. Indeed, we carried out functional testing to demonstrate that the p.Asp41Asn variant led to changes in LSM7 mRNA and protein levels. We

also showed that both variants (p.Asp41Asn and p.Arg69Pro) led to defects in the formation of both the LSM1-7 and LSM2-8 complexes. A zebrafish animal model demonstrated that Lsm7 is important for the development of the nervous system and has a critical role in oligodendrocyte survival. Additionally, the levels of U6 snRNA, a key interactor of the LSM2-8 complex were upregulated in patient GB114.0 pointing towards possible dysfunction in splicing, which would require further analysis to confirm. These results represent potential disease-causing mechanisms that require additional investigations to determine exactly how variants in this seemingly widely expressed protein involved in the general cellular processes of mRNA decay and splicing have such a devastating effect on the CNS and myelin production. Nonetheless, these findings have been essential as they will allow families to have a molecular diagnosis, as well as give them the opportunity to obtain genetic counselling to make reproductive decisions. The discovery of novel genes, such as *LSM7* and the subsequent functional studies are the first steps in understanding their pathogenesis and eventually in identifying therapeutic avenues.

## References


1 Vindry, C. *et al.* Dual RNA Processing Roles of Pat1b via Cytoplasmic Lsm1-7 and Nuclear Lsm2-8 Complexes. *Cell Rep* **20**, 1187-1200, doi:10.1016/j.celrep.2017.06.091 (2017).

2 Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol* **18**, 655-670, doi:10.1038/nrm.2017.86 (2017).

3 Vanderver, A. *et al.* Case definition and classification of leukodystrophies and leukoencephalopathies. *Mol Genet Metab* **114**, 494-500, doi:10.1016/j.ymgme.2015.01.006 (2015).

4 Schiffmann, R. & van der Knaap, M. S. Invited article: an MRI-based approach to the diagnosis of white matter disorders. *Neurology* **72**, 750-759, doi:10.1212/01.wnl.0000343049.00540.c8 (2009).

5 Steenweg, M. E. *et al.* Magnetic resonance imaging pattern recognition in hypomyelinating disorders. *Brain* **133**, 2971-2982, doi:10.1093/brain/awq257 (2010).

6 Bonkowsky, J. L. *et al.* The burden of inherited leukodystrophies in children. *Neurology* **75**, 718-725, doi:10.1212/WNL.0b013e3181eee46b (2010).

7 van der Knaap, M. S. & Bugiani, M. Leukodystrophies: a proposed classification system based on pathological changes and pathogenetic mechanisms. *Acta Neuropathol* **134**, 351-382, doi:10.1007/s00401-017-1739-1 (2017).

8 Adang, L. A. *et al.* Revised consensus statement on the preventive and symptomatic care of patients with leukodystrophies. *Mol Genet Metab* **122**, 18-32, doi:10.1016/j.ymgme.2017.08.006 (2017).

9 Kevelam, S. H. *et al.* Update on Leukodystrophies: A Historical Perspective and Adapted Definition. *Neuropediatrics* **47**, 349-354, doi:10.1055/s-0036-1588020 (2016).

10 Biffi, A., Lucchini, G., Rovelli, A. & Sessa, M. Metachromatic leukodystrophy: an overview of current and prospective treatments. *Bone Marrow Transplant* **42 Suppl 2**, S2-6, doi:10.1038/bmt.2008.275 (2008).

11 Escolar, M. L. *et al.* Transplantation of umbilical-cord blood in babies with infantile Krabbe's disease. *N Engl J Med* **352**, 2069-2081, doi:10.1056/NEJMoa042604 (2005).

12 Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* **14**, 681-691, doi:10.1038/nrg3555 (2013).

13 Sawyer, S. L. *et al.* Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet* **89**, 275-284, doi:10.1111/cge.12654 (2016).

14 Parikh, S. *et al.* A clinical approach to the diagnosis of patients with leukodystrophies and genetic leukoencephelopathies. *Mol Genet Metab* **114**, 501-515, doi:10.1016/j.ymgme.2014.12.434 (2015).

15 Ashrafi, M. R. & Tavasoli, A. R. Childhood leukodystrophies: A literature review of updates on new definitions, classification, diagnostic approach and management. *Brain Dev* **39**, 369-385, doi:10.1016/j.braindev.2017.01.001 (2017).

16 Vanderver, A., Tonduti, D., Schiffmann, R., Schmidt, J. & van der Knaap, M. S. in *GeneReviews* (eds M. P. Adam *et al.*) (2014).

17 Tetreault, M. *et al.* Recessive mutations in POLR3B, encoding the second largest subunit of Pol III, cause a rare hypomyelinating leukodystrophy. *Am J Hum Genet* **89**, 652-655, doi:10.1016/j.ajhg.2011.10.006 (2011).
18 Mendes, M. I. *et al.* Bi-allelic Mutations in EPRS, Encoding the Glutamyl-Prolyl-Aminoacyl-tRNA Synthetase, Cause a Hypomyelinating Leukodystrophy. *Am J Hum Genet* **102**, 676-684, doi:10.1016/j.ajhg.2018.02.011 (2018).
19 Bernard, G. & Vanderver, A. in *GeneReviews((R))* (eds M. P. Adam *et al.*) (2012).
20 Heim, P. *et al.* Leukodystrophy incidence in Germany. *Am J Med Genet* **71**, 475-478 (1997).
21 Bernard, G. *et al.* Mutations of POLR3A encoding a catalytic subunit of RNA polymerase Pol III cause a recessive hypomyelinating leukodystrophy. *Am J Hum Genet* **89**, 415-423, doi:10.1016/j.ajhg.2011.07.014 (2011).
22 Perrier, S. *et al.* Expanding the phenotypic and molecular spectrum of RNA polymerase III–related leukodystrophy. *Neurol Genet* **6**, 1-12, doi:doi:10.1212/NXG.0000000000000425 (2020).
23 Simons, C. *et al.* Loss-of-function alanyl-tRNA synthetase mutations cause an autosomal-recessive early-onset epileptic encephalopathy with persistent myelination defect. *Am J Hum Genet* **96**, 675-681, doi:10.1016/j.ajhg.2015.02.012 (2015).
24 Thiffault, I. *et al.* Recessive mutations in POLR1C cause a leukodystrophy by impairing biogenesis of RNA polymerase III. *Nat Commun* **6**, 7623, doi:10.1038/ncomms8623 (2015).
25 Cayami, F. K. *et al.* 4H Leukodystrophy: Lessons from 3T Imaging. *Neuropediatrics* **49**, 112-117, doi:10.1055/s-0037-1608780 (2018).
26 Friedman, J. *et al.* Biallelic mutations in valyl-tRNA synthetase gene VARS are associated with a progressive neurodevelopmental epileptic encephalopathy. *Nat Commun* **10**, 707, doi:10.1038/s41467-018-07067-3 (2019).
27 van Rappard, D. F., Boelens, J. J. & Wolf, N. I. Metachromatic leukodystrophy: Disease spectrum and approaches for treatment. *Best Pract Res Clin Endocrinol Metab* **29**, 261-273, doi:10.1016/j.beem.2014.10.001 (2015).
28 Sevin, C., Aubourg, P. & Cartier, N. Enzyme, cell and gene-based therapies for metachromatic leukodystrophy. *J Inherit Metab Dis* **30**, 175-183, doi:10.1007/s10545-007-0540-z (2007).
29 Matthes, F. *et al.* Efficacy of enzyme replacement therapy in an aggravated mouse model of metachromatic leukodystrophy declines with age. *Hum Mol Genet* **21**, 2599-2609, doi:10.1093/hmg/dds086 (2012).
30 Spratley, S. J. & Deane, J. E. New therapeutic approaches for Krabbe disease: The potential of pharmacological chaperones. *J Neurosci Res* **94**, 1203-1219, doi:10.1002/jnr.23762 (2016).
31 Kato, K. *et al.* Allogeneic stem cell transplantation with reduced intensity conditioning for patients with adrenoleukodystrophy. *Mol Genet Metab Rep* **18**, 1-6, doi:10.1016/j.ymgmr.2018.11.001 (2019).
32 Resnick, I. B. *et al.* Treatment of X-linked childhood cerebral adrenoleukodystrophy by the use of an allogeneic stem cell transplantation with reduced intensity conditioning regimen. *Clin Transplant* **19**, 840-847, doi:10.1111/j.1399-0012.2005.00411.x (2005).

33 van der Knaap, M. S. *et al.* Mutations in each of the five subunits of translation initiation factor eIF2B can cause leukoencephalopathy with vanishing white matter. *Ann Neurol* **51**, 264-270, doi:10.1002/ana.10112 (2002).
34 Leegwater, P. A. *et al.* Subunits of the translation initiation factor eIF2B are mutant in leukoencephalopathy with vanishing white matter. *Nat Genet* **29**, 383-388, doi:10.1038/ng764 (2001).
35 Li, W., Wang, X., Van Der Knaap, M. S. & Proud, C. G. Mutations linked to leukoencephalopathy with vanishing white matter impair the function of the eukaryotic initiation factor 2B complex in diverse ways. *Mol Cell Biol* **24**, 3295-3306, doi:10.1128/mcb.24.8.3295-3306.2004 (2004).
36 Steenweg, M. E. *et al.* Leukoencephalopathy with thalamus and brainstem involvement and high lactate 'LTBL' caused by EARS2 mutations. *Brain* **135**, 1387-1394, doi:10.1093/brain/aws070 (2012).
37 Taskin, B. D. *et al.* Early-Onset Mild Type Leukoencephalopathy Caused by a Homozygous EARS2 Mutation. *J Child Neurol* **31**, 938-941, doi:10.1177/0883073816630087 (2016).
38 Scheper, G. C. *et al.* Mitochondrial aspartyl-tRNA synthetase deficiency causes leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. *Nat Genet* **39**, 534-539, doi:10.1038/ng2013 (2007).
39 Nahhas, N., Conant, A., Orthmann-Murphy, J., Vanderver, A. & Hobson, G. in *GeneReviews((R))* (eds M. P. Adam *et al.*) (1993).
40 Dorboz, I. *et al.* Mutation in POLR3K causes hypomyelinating leukodystrophy and abnormal ribosomal RNA regulation. *Neurol Genet* **4**, e289, doi:10.1212/NXG.0000000000000289 (2018).
41 Nguengang Wakap, S. *et al.* Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet*, doi:10.1038/s41431-019-0508-0 (2019).
42 McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* **80**, 588-604, doi:10.1086/514346 (2007).
43 Ayme, S., Urbero, B., Oziel, D., Lecouturier, E. & Biscarat, A. C. [Information on rare diseases: the Orphanet project]. *Rev Med Interne* **19 Suppl 3**, 376S-377S (1998).
44 Vanderver, A. *et al.* Whole exome sequencing in patients with white matter abnormalities. *Ann Neurol* **79**, 1031-1037, doi:10.1002/ana.24650 (2016).
45 Singleton, A. B. Exome sequencing: a transformative technology. *Lancet Neurol* **10**, 942-946, doi:10.1016/S1474-4422(11)70196-X (2011).
46 Rehm, H. L. *et al.* ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* **15**, 733-747, doi:10.1038/gim.2013.92 (2013).
47 Sun, Y. *et al.* Next-generation diagnostics: gene panel, exome, or whole genome? *Hum Mutat* **36**, 648-655, doi:10.1002/humu.22783 (2015).
48 Kong, S. W., Lee, I. H., Liu, X., Hirschhorn, J. N. & Mandl, K. D. Measuring coverage and accuracy of whole-exome sequencing in clinical context. *Genet Med* **20**, 1617-1626, doi:10.1038/gim.2018.51 (2018).
49 Williams, E., Retterer, K., Cho, M., Richard, G. & Juusola, J. in *GeneDx.*
50 Ewans, L. J. *et al.* Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genet Med* **20**, 1564-1574, doi:10.1038/gim.2018.39 (2018).

51 Jalkh, N. *et al.* The added value of WES reanalysis in the field of genetic diagnosis: lessons learned from 200 exomes in the Lebanese population. *BMC Med Genomics* **12**, 11, doi:10.1186/s12920-019-0474-y (2019).
52 Epilepsy Genetics, I. The Epilepsy Genetics Initiative: Systematic reanalysis of diagnostic exomes increases yield. *Epilepsia* **60**, 797-806, doi:10.1111/epi.14698 (2019).
53 Gonorazky, H. D. *et al.* Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *Am J Hum Genet* **104**, 1007, doi:10.1016/j.ajhg.2019.04.004 (2019).
54 Fresard, L. & Montgomery, S. B. Diagnosing rare diseases after the exome. *Cold Spring Harb Mol Case Stud* **4**, doi:10.1101/mcs.a003392 (2018).
55 Cho, A. *et al.* Fundoscopy-directed genetic testing to re-evaluate negative whole exome sequencing results. *Orphanet J Rare Dis* **15**, 32, doi:10.1186/s13023-020-1312-1 (2020).
56 Reeves, W. H., Narain, S. & Satoh, M. Henry Kunkel, Stephanie Smith, clinical immunology, and split genes. *Lupus* **12**, 213-217, doi:10.1191/0961203303lu360xx (2003).
57 Veretnik, S., Wills, C., Youkharibache, P., Valas, R. E. & Bourne, P. E. Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome. *PLoS Comput Biol* **5**, e1000315, doi:10.1371/journal.pcbi.1000315 (2009).
58 Tharun, S. Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay? *RNA Biol* **6**, 228-232 (2009).
59 Beggs, J. D. Lsm proteins and RNA processing. *Biochem Soc Trans* **33**, 433-438, doi:10.1042/BST0330433 (2005).
60 Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**, 108-121, doi:10.1038/nrm3742 (2014).
61 Tharun, S. Roles of eukaryotic Lsm proteins in the regulation of mRNA function. *Int Rev Cell Mol Biol* **272**, 149-189, doi:10.1016/S1937-6448(08)01604-3 (2009).
62 Perea-Resa, C., Hernandez-Verdeja, T., Lopez-Cobollo, R., del Mar Castellano, M. & Salinas, J. LSM proteins provide accurate splicing and decay of selected transcripts to ensure normal Arabidopsis development. *Plant Cell* **24**, 4930-4947, doi:10.1105/tpc.112.103697 (2012).
63 Achsel, T. *et al.* A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J* **18**, 5789-5802, doi:10.1093/emboj/18.20.5789 (1999).
64 Achsel, T., Stark, H. & Luhrmann, R. The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proc Natl Acad Sci U S A* **98**, 3685-3689, doi:10.1073/pnas.071033998 (2001).
65 Sharif, H. & Conti, E. Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. *Cell Rep* **5**, 283-291, doi:10.1016/j.celrep.2013.10.004 (2013).
66 Kufel, J., Bousquet-Antonelli, C., Beggs, J. D. & Tollervey, D. Nuclear pre-mRNA decapping and 5' degradation in yeast require the Lsm2-8p complex. *Mol Cell Biol* **24**, 9646-9657, doi:10.1128/MCB.24.21.9646-9657.2004 (2004).
67 Herrero, A. B. & Moreno, S. Lsm1 promotes genomic stability by controlling histone mRNA decay. *EMBO J* **30**, 2008-2018, doi:10.1038/emboj.2011.117 (2011).
68 Tharun, S. Purification and analysis of the decapping activator Lsm1p-7p-Pat1p complex from yeast. *Methods Enzymol* **448**, 41-55, doi:10.1016/S0076-6879(08)02603-7 (2008).

69 Caponigro, G. & Parker, R. Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. *Genes Dev* **9**, 2421-2432, doi:10.1101/gad.9.19.2421 (1995).
70 Wu, D. *et al.* Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. *Cell Res* **24**, 233-246, doi:10.1038/cr.2013.152 (2014).
71 Chowdhury, A., Mukhopadhyay, J. & Tharun, S. The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* **13**, 998-1016, doi:10.1261/rna.502507 (2007).
72 Tharun, S. *et al.* Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* **404**, 515-518, doi:10.1038/35006676 (2000).
73 Chowdhury, A., Kalurupalle, S. & Tharun, S. Pat1 contributes to the RNA binding activity of the Lsm1-7-Pat1 complex. *RNA* **20**, 1465-1475, doi:10.1261/rna.045252.114 (2014).
74 Verdone, L., Galardi, S., Page, D. & Beggs, J. D. Lsm proteins promote regeneration of pre-mRNA splicing activity. *Curr Biol* **14**, 1487-1491, doi:10.1016/j.cub.2004.08.032 (2004).
75 Friesen, W. J. & Dreyfuss, G. Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN). *J Biol Chem* **275**, 26370-26375, doi:10.1074/jbc.M003299200 (2000).
76 Will, C. L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**, doi:10.1101/cshperspect.a003707 (2011).
77 Chen, W. & Moore, M. J. Spliceosomes. *Curr Biol* **25**, R181-183, doi:10.1016/j.cub.2014.11.059 (2015).
78 Mayes, A. E., Verdone, L., Legrain, P. & Beggs, J. D. Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *The EMBO Journal* **18**, 4321-4331 (1999).
79 Wilusz, C. J. & Wilusz, J. Lsm proteins and Hfq: Life at the 3' end. *RNA Biol* **10**, 592-601, doi:10.4161/rna.23695 (2013).
80 Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Res* **28**, 1129-1140, doi:10.1038/s41422-018-0094-7 (2018).
81 Tharun, S., Muhlrad, D., Chowdhury, A. & Parker, R. Mutations in the Saccharomyces cerevisiae LSM1 gene that affect mRNA decapping and 3' end protection. *Genetics* **170**, 33-46, doi:10.1534/genetics.104.034322 (2005).
82 Didychuk, A. L., Butcher, S. E. & Brow, D. A. The life of U6 small nuclear RNA, from cradle to grave. *RNA* **24**, 437-460, doi:10.1261/rna.065136.117 (2018).
83 Stanek, D., Rader, S. D., Klingauf, M. & Neugebauer, K. M. Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J Cell Biol* **160**, 505-516, doi:10.1083/jcb.200210087 (2003).
84 Fallini, C. *et al.* The survival of motor neuron (SMN) protein interacts with the mRNA-binding protein HuD and regulates localization of poly(A) mRNA in primary motor neuron axons. *J Neurosci* **31**, 3914-3925, doi:10.1523/JNEUROSCI.3631-10.2011 (2011).
85 Jimenez-Lopez, D. & Guzman, P. Insights into the evolution and domain structure of ataxin-2 proteins across eukaryotes. *BMC Research Notes* **7**, 453 (2014).

86 Ostrowski, L. A., Hall, A. C. & Mekhail, K. Ataxin-2: From RNA Control to Human Health and Disease. *Genes (Basel)* **8**, doi:10.3390/genes8060157 (2017).
87 Tharun, S. & Parker, R. Analysis of mutations in the yeast mRNA decapping enzyme. *Genetics* **151**, 1273-1285 (1999).
88 Hamilton, A. *et al.* Concordance between whole-exome sequencing and clinical Sanger sequencing: implications for patient care. *Mol Genet Genomic Med* **4**, 504-512, doi:10.1002/mgg3.223 (2016).
89 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).
90 Lim, E. T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet* **10**, e1004494, doi:10.1371/journal.pgen.1004494 (2014).
91 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
92 Philippakis, A. A. *et al.* The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat* **36**, 915-921, doi:10.1002/humu.22858 (2015).
93 Beecroft, S. J. *et al.* Expanding the phenotypic spectrum associated with mutations of DYNC1H1. *Neuromuscul Disord* **27**, 607-615, doi:10.1016/j.nmd.2017.04.011 (2017).
94 Strickland, A. V. *et al.* Mutation screen reveals novel variants and expands the phenotypes associated with DYNC1H1. *J Neurol* **262**, 2124-2134, doi:10.1007/s00415-015-7727-2 (2015).
95 Meienberg, J., Bruggmann, R., Oexle, K. & Matyas, G. Clinical sequencing: is WGS the better WES? *Hum Genet* **135**, 359-362, doi:10.1007/s00439-015-1631-9 (2016).
96 Schwarze, K., Buchanan, J., Taylor, J. C. & Wordsworth, S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet Med* **20**, 1122-1130, doi:10.1038/gim.2017.247 (2018).
97 Trujillano, D. *et al.* Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet* **25**, 176-182, doi:10.1038/ejhg.2016.146 (2017).
98 Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat* **36**, 928-930, doi:10.1002/humu.22844 (2015).
99 Sobreira, N., Schiettecatte, F., Boehm, C., Valle, D. & Hamosh, A. New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. *Hum Mutat* **36**, 425-431, doi:10.1002/humu.22769 (2015).
100 Goh, G. & Choi, M. Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. *Genomics Inform* **10**, 214-219, doi:10.5808/GI.2012.10.4.214 (2012).
101 Lupski, J. R. Structural variation mutagenesis of the human genome: Impact on disease and evolution. *Environ Mol Mutagen* **56**, 419-436, doi:10.1002/em.21943 (2015).
102 Clark, M. M. *et al.* Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med* **3**, 16, doi:10.1038/s41525-018-0053-8 (2018).

103 Veeramah, K. R. *et al.* Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia* **54**, 1270-1281, doi:10.1111/epi.12201 (2013).
104 McInerney-Leo, A. M. *et al.* Whole exome sequencing is an efficient, sensitive and specific method of mutation detection in osteogenesis imperfecta and Marfan syndrome. *Bonekey Rep* **2**, 456, doi:10.1038/bonekey.2013.190 (2013).
105 Wortmann, S. B., Koolen, D. A., Smeitink, J. A., van den Heuvel, L. & Rodenburg, R. J. Whole exome sequencing of suspected mitochondrial patients in clinical practice. *J Inherit Metab Dis* **38**, 437-443, doi:10.1007/s10545-015-9823-y (2015).
106 Powis, Z. *et al.* Exome sequencing in neonates: diagnostic rates, characteristics, and time to diagnosis. *Genet Med* **20**, 1468-1471, doi:10.1038/gim.2018.11 (2018).
107 Taylor, J. C. *et al.* Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet* **47**, 717-726, doi:10.1038/ng.3304 (2015).
108 Luhtala, N. & Parker, R. LSM1 over-expression in Saccharomyces cerevisiae depletes U6 snRNA levels. *Nucleic Acids Res* **37**, 5529-5536, doi:10.1093/nar/gkp572 (2009).
109 Vazquez-Arango, P. & O'Reilly, D. Variant snRNPs: New players within the spliceosome system. *RNA Biol* **15**, 17-25, doi:10.1080/15476286.2017.1373238 (2018).
110 Movassat, M., Mueller, W. F. & Hertel, K. J. In vitro assay of pre-mRNA splicing in mammalian nuclear extract. *Methods Mol Biol* **1126**, 151-160, doi:10.1007/978-1-62703-980-2_11 (2014).
111 Whiley, P. J. *et al.* Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin Chem* **60**, 341-352, doi:10.1373/clinchem.2013.210658 (2014).
112 Camacho-Carvajal, M. M., Wollscheid, B., Aebersold, R., Steimle, V. & Schamel, W. W. Two-dimensional Blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach. *Mol Cell Proteomics* **3**, 176-182, doi:10.1074/mcp.T300010-MCP200 (2004).
113 Fiala, G. J., Schamel, W. W. & Blumenthal, B. Blue native polyacrylamide gel electrophoresis (BN-PAGE) for analysis of multiprotein complexes from cellular lysates. *J Vis Exp*, doi:10.3791/2164 (2011).