Identification, Validation and Characterization of the Mutation on Chromosome 18p which is Responsible for Causing Myoclonus-Dystonia

By

Megan Vanstone

Thesis
Submitted to the Department of Biochemistry, Microbiology and Immunology
Graduate Collaborative Program in Human and Molecular Genetics
Faculty of Medicine
University of Ottawa
2012

© Megan Vanstone, Ottawa, Canada, 2012

Abstract

Myoclonus-Dystonia (MD) is an inherited, rare, autosomal dominant movement disorder characterized by quick, involuntary muscle jerking or twitching (myoclonus) and involuntary muscle contractions that cause twisting and pulling movements, resulting in abnormal postures (dystonia). The first MD locus was mapped to 7q21-q31 and called DYT11; this locus corresponds to the *SGCE* gene. Our group previously identified a second MD locus (DYT15) which maps to a 3.18 Mb region on 18p11. Two patients were chosen to undergo next-generation sequencing, which identified 2,292 shared novel variants within the critical region. Analysis of these variants revealed a 3 bp duplication in a transcript referred to as CD108131, which is believed to be a long non-coding RNA. Characterization of this transcript determined that it is 863 bp in size, it is ubiquitously expressed, with high expression in the cerebellum, and it accounts for ~3% of MD cases.

Acknowledgements

First and foremost I would like to thank my supervisor Dr. Dennis Bulman for accepting me as a student in his lab and for guiding me through to the completion of this thesis project. Not only has he provided me with the tools, training, and knowledge to complete this Master's program, he has provided me with the support and the encouragement to aim even higher. Whether my concerns involved research questions, career advice, or personal challenges, Dennis' door was always open; for that, I am grateful. I also owe a huge thank you to Dr. Kym Boycott for her involvement in the lab and my project. Her insights during lab meetings, participation on my advisory committee, and generosity with her time and her resources have helped make my project a success.

I also owe a huge thank you to all of the members of the Bulman/Boycott lab for making my experience a memorable one. The laughter, lunch dates, and time that we've shared together made every day in the lab enjoyable. Thank you to Dr. Zoe Prichard for her help with all of the bioinformatics required for my project, as well as for the encouragement she provided along the way. Thanks also to Chandree Beaulieu and Jeremy Schwartzentruber for their assistance with sample preparation and data analysis for exome sequencing.

I would also like to acknowledge Dr. David Grimes for his involvement in this project. His clinical expertise, collaboration with movement disorder clinics, recruitment of patients, and participation on my advisory committee were all indispensable. Dr. Bernard Jasmin, the third member of my advisory committee, has also provided me with great insight and support throughout my degree. It is also important that I acknowledge the families who donated samples for this project. Without their participation, this thesis and the progress made so far towards identifying the genetic cause of Myoclonus-Dystonia would not have been possible.

Thanks are also owed to Dr. Bulman, CIHR and OGS for funding this project. The FGPS, Department of BMI, and GSAED also provided me with travel funding, which allowed me to attend an international conference organized by the American Society of Human Genetics.

Finally, I would like to extend my gratitude to my family and friends for their love and support throughout (and leading up to) this Master's degree. To my parents, your encouragement, patience, and understanding have motivated and inspired me. To my husband, Jason, thank you for opening my eyes to the world of research and for being the role model that you are with your determination and passion for science. Thank you for believing in my potential, for supporting my goals, and for making every day of my life a joy.

Table of Contents

Page

List of Abbreviations

µg	microgram
µl	microlitre
µM	micromoles per litre
18p11	chromosome 18, short arm, region 1, band 1
7q21	chromosome 7, long arm, region 2, band 1
AA	amino acid
AAA	ATPases associated with diverse cellular activities
ACTB	actin, beta
Air	acute insulin response
ALAS2	aminolevulinate, delta-, synthase 2
ALS	amyotrophic lateral sclerosis
ANK2	ankyrin 2
ATP1A3	ATPase alpha-3 subunit
BH4	tetrahydrobiopterin
BHC	benign hereditary chorea
bp	base pair

cDNA	complementary DNA
ChIRP-seq	chromatin isolation by RNA purification and sequencing
Chr	chromosome
cm	centimetre
cM	centimorgans
Ctrl	control
cpm	counts per minute
CRIPAK	cysteine-rich PAK1 inhibitor
D2R	dopamine D2 receptor
dbSNP	Single Nucleotide Polymorphism database
ddNTP	dideoxyribonucleoside triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
DRD	dopamine-responsive dystonia
DRD2	dopamine receptor D2
EBV	Epstein-Barr virus
EPB41L3	erythrocyte membrane protein band 4.1-like 3

EST	expressed sequence tag
FBS	fetal bovine serum
FORGE	Finding of Rare Disease Genes
g	gram
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GCH1	GTP cyclohydrolase 1
GLUT1	glucose transporter type 1
GPi	globus pallidum internum
Het	heterozygous
hg19	human genome version 19
Hom	homozygous
HOTAIR	HOX transcript antisense RNA
IDT	Integrated DNA Technologies
kb	kilobase
lncRNA	long non-coding RNA
lod	logarithm of the odds ratio
M	mutant

Mb	megabase
MD	Myoclonus-Dystonia
MF1	Myoclonus Family 1
MF2-38	Myoclonus Family 2 – Myoclonus Family 38
miRNA	microRNA
mL	millilitres
mM	millimoles per litre
MR1	major histocompatibility complex, class 1-related
mRNA	messenger RNA
ncRNA	non-coding RNA
ng	nanogram
NGS	next-generation sequencing
No RT (-RT)	no reverse transcriptase
nt	nucleotides
PCR	polymerase chain reaction
PNKD1	paroxysmal nonkinesigenic dyskinesia 1
PRKRA	protein kinase, interferon-inducible double stranded RNA dependent activator

qRT-PCR	Quantitative Real-Time PCR
RDP	rapid-onset dystonia-parkinsonism
REFSEQ	Reference Sequence
RNA	ribonucleic acid
rpm	revolutions per minute
RT-PCR	reverse-transcriptase PCR
SDK1	sidekick cell adhesion molecule 1
SGCE	epsilon-sarcoglycan
siRNA	small interfering RNA
SLC2A1	glucose transporter type 1
SNAI1	snail homolog 1
SNP	single nucleotide polymorphism
SPR	sepiapterine reductase
SYCP1	synaptonemal complex protein 1
TAF1	TATA box-binding protein-associated factor 1
TCAG	The Centre for Applied Genomics
TH	tyrosine hydroxlase

THAP1	Thanatos-associated protein domain containing apoptosis-associated protein 1
TITF-1	thyroid nuclear factor 1
T_M	melting temperature
TOR1A	torsin A
TTN	titin
UCSC	University of California Santa Cruz
UTRN	utrophin
V	volt
-ve C	negative control
VTCN1	V-set domain containing T cell activation inhibitor 1
WT	wild-type
Xist	X-inactive specific transcript
ZFP161	zinc finger protein 161 homolog (mouse)

List of Figures

List of Tables

Chapter 1: Introduction

1.1. Dystonic movement disorders: classification and genetic loci

Dystonia is the third most common movement disorder next to Parkinson's disease and essential tremor (Pisani, 2011). Diseases where there is no obvious brain damage and where dystonia is the only or major symptom are referred to as primary dystonias (Fahn *et al*., 1998). The primary dystonias can be further divided into three subgroups: pure dystonia, dystonia-plus syndromes, and paroxysmal dystonia (Albanese *et al*., 2011). The complexity of these movement disorders is highlighted by the fact that there are currently 21 different loci (DYT1-21) linked to these various forms of dystonia (Norgren *et al*., 2011). Of the 21 dystonia loci, 10 (DYT3, DYT5a/DYT14, DYT5b, DYT6, DYT8, DYT11, DYT12, DYT16, DYT18) have been linked to mutations in protein-coding genes, including: *TAF1, GCH1, TH, THAP1, PNKD1/MR1, SGCE, ATP1A3, PRKRA*, and *SLC2A1* respectively (Bragg *et al*., 2011).

When dystonia occurs without any additional symptoms, it is classified as pure dystonia (Albanese *et al*., 2011). There are currently 8 known genetic loci associated with pure dystonia: DYT1, DYT2, DYT4, DYT6, DYT7, DYT13, DYT17, and DYT21 (Spatola & Wider, 2012). The most common form of pure dystonia is primary generalized torsion dystonia, which occurs at a frequency of 1/200,000 in the general population (Spatola & Wider, 2012). The gene associated with this form of the disease was the first dystonia gene (DYT1) to be identified in 1997 and encodes the protein torsin A (TOR1A), which is an endoplasmic reticulum ATPase and a member of the AAA+ superfamily (Ozelius *et al*., 1997). TOR1A is widely distributed in the brain and mutant TOR1A has been shown to result in impaired dopamine release as well as interfere with cytoskeletal events, affecting the proper development of neuronal pathways in the

brain (Bressman, 2007). A second form of pure dystonia is mixed type primary torsion dystonia, resulting from mutations in the Thanatos-associated protein domain containing apoptosis-associated protein1 (*THAP1*) gene (DYT6; Fuchs *et al*., 2009). The protein encoded by this gene is involved in apoptosis and the cell cycle (Spatola & Wider, 2012). There is also evidence that THAP1 down-regulates the TOR1A core promoter and that mutations in the *THAP1* gene can result in an increase of torsin A expression (Spatola & Wider, 2012).

Dystonia-plus syndromes are non-degenerative diseases that are characterized by dystonia as well as additional clinical features such as Parkinsonism or myoclonus (Albanese *et al*., 2011). There are currently 7 known genetic loci associated with the dystonia-plus syndromes: DYT3, DYT5a (DYT14), DYT5b, DYT11, DYT12, DYT15, and DYT16 (Spatola & Wider, 2012). There are three major disorders characterized as dystonia-plus syndromes: dopa-responsive dystonia (DRD), Myoclonus-Dystonia (MD), and dystonia-parkinsonism (Asmus & Gasser, 2010). DRD, also referred to as Segawa Disease, typically presents in childhood as postural dystonia of a single limb and eventually spreads to all limbs (Gordon, 2008). An important distinction of this disorder from other dystonia-plus syndromes is that symptoms can be alleviated by treatment with levodopa (Gordon, 2008). Three key genes have been associated with DRD: GTP-cyclohydrolase I (*GCH1*; DYT5a/DYT14), which encodes an enzyme that catalyzes the first step of tetrahydrobiopterin (BH4) synthesis (which is a cofactor in the dopamine synthesis pathway); tyrosine hydroxylase (*TH*; DYT5b), which encodes an enzyme involved in the synthesis of catecholamines; and sepiapterine reductase (*SPR*; DYT5b), which encodes an enzyme that catalyzes the final step in the synthesis of BH4 (Ichinose *et al*., 1994; Lüdecke *et al*., 1995; Steinberger *et al*., 2004). Mutations in any of these three genes result in deficient dopamine synthesis, which explains why treatment with levodopa can aid in alleviating

symptoms (Schneider & Bhatia, 2010). Rapid-onset dystonia-parkinsonism (RDP) is characterized by its abrupt symptom onset (Asmus & Gasser, 2010). RDP results from mutations in the Na^+/K^+ ATPase alpha-3 subunit (*ATP1A3*; DYT12), which is responsible for maintaining the ion gradient across resting neuronal membranes (De Carvalho Aguiar *et al*., 2004). An X-linked form of dystonia-parkinsonism referred to as "Lubag" is believed to result from mutations in the TATA box-binding protein-associated factor 1 (*TAF1*) gene (Schneider & Bhatia, 2010). A form of young-onset dystonia-parkinsonism has been linked to mutations in the protein kinase, interferon-inducible double-stranded RNA-dependent activator (*PRKRA*; DYT16) gene (Camargos *et al*., 2008). *PRKRA* is a stress-response gene, however the mechanism by which it causes dystonia-parkinsonism is unknown (Asmus & Gasser, 2010; Schneider & Bhatia, 2010). The third sub-type of dystonia-plus syndromes, Myoclonus-Dystonia, will be discussed in detail in section 1.2.

Paroxysmal dystonia involves episodes of dystonic movements separated by periods of normalcy (Albanese *et al*., 2011). There are currently 5 known genetic loci associated with paroxysmal dystonia: DYT8, DYT10, DYT18 (DYT9), DYT19, and DYT20 (Spatola & Wider, 2012). This form of dystonia has been associated with mutations in the glucose transporter type 1 (*GLUT1*) gene, also referred to as *SLC2A1* (DYT18/DYT9; Weber *et al*., 2008). Mutations in this gene result in GLUT1 deficiency; however, the resultant pathophysiology leading to dystonia is still unknown (Spatola & Wider, 2012). Mutations in the brain-specific isoform of the Myofibrillogenesis Regulator 1 (*MR1*; DYT8) gene have also been found in patients with paroxysmal dystonia, however the function of the protein encoded by this gene is yet to be determined (Rainier *et al*., 2004).

Despite the prevalence and diverse genetic cause of dystonia-based disorders, effective

treatments and therapies remain elusive. A great deal of the underlying mechanism leading to dystonia is uncharacterized; for this reason, treatments of the disease primarily involve the alleviation of symptoms including pain and abnormal movements (Lalli *et al*., 2011). The two major forms of treatment involve drug administration and surgical intervention (Lalli *et al*., 2011). Drugs such as botulinum toxin (for focal dystonia), anticholinergics, levodopa (in the case of DRD), dopamine blockers, baclofen (for generalized or secondary dystonia), and benzodiazepines are used (Lalli *et al*., 2011; Ozelius *et al*., 2011). Surgical intervention is typically used for the treatment of generalized dystonia and involves implanting electrical stimulators of the globus pallidum internum (GPi) for long-term deep brain stimulation (Lalli *et al*., 2011). While the aforementioned treatments provide relief of symptoms for some patients, they are not without side effects and they do not work in all instances. For this reason, improvements must be made towards finding alternative treatment options that are more effective and that address the large amount of diversity seen in dystonia phenotypes.

1.2. Myoclonus-Dystonia

Inherited Myoclonus-Dystonia (MD) is an autosomal-dominant dystonia-plus movement disorder (Albanese *et al*., 2011). While Dystonias as a whole are fairly common, MD is very rare, occurring at a frequency of approximately 2 out of 1,000,000 in Europe (Asmus & Gasser, 2010). The two main clinical characteristics of the disorder are myoclonus and dystonia, thus its name. Myoclonus tends to be the predominant clinical feature and it typically manifests itself as lightning-fast myoclonic jerks, most commonly in the arms, shoulders, and neck (Furukawa & Rajput, 2002; Asmus & Gasser, 2010). The two major symptoms of dystonia are torticollis (muscle spasm in the neck) and writer's cramp, which result from sustained muscle contractions

(Furukawa & Rajput, 2002; Albanese *et al*., 2011).

Onset of the disorder usually occurs within the first two decades of life, with a mean age of onset of 6 years (Asmus & Gasser, 2010). The symptoms of MD are often alcohol responsive and psychiatric problems such as alcohol abuse, depression, obsessive compulsive disorder and anxiety are common features of the disorder (Furukawa & Rajput, 2002; Ritz *et al*., 2009).

Myoclonus-Dystonia is a heterogenous disorder (Grimes *et al*., 2001). There are currently two loci associated with MD, DYT11 and DYT15 (Spatola & Wider, 2012). The DYT11 locus was the first to be identified and pertains to chromosome 7q21-q31 (Nygaard *et al*., 1999). It was later determined that the disease gene within the DYT11 locus is *epsilon-sarcoglycan* (Zimprich *et al*., 2001). However, mutations in *SGCE* only account for about 30-50% of MD cases (Grünewald *et al*., 2008; Han *et al*., 2007; Ritz *et al*., 2011). In 2002, Grimes *et al.* identified a novel locus for inherited Myoclonus-Dystonia on 18p11, DYT15 (Grimes *et al*., 2002). Most recently, a third locus has been associated with MD: the tyrosine hydroxylase (*TH*) gene (Stamelou *et al*., 2012). This gene was initially identified as DYT5b, responsible for causing dopa-responsive dystonia (Lüdecke *et al*., 1995). For the first time, however, Stamelou *et al*. (2012) have identified a family in which three siblings are affected with levodopa-responsive MD as a result of compound heterozygosity of a point mutation in the promoter region and a nonsynonymous substitution in exon 12 of the *TH* gene. These results suggest that there could be a higher degree of mechanistic overlap underlying the different dystonia phenotypes than previously thought.

1.3. Mutations in the *SGCE* gene cause Myoclonus-Dystonia

Zimprich *et al*. (2001) were the first to identify MD-causing mutations in *SGCE*.

Through the use of positional cloning, they identified five separate heterozygous *SGCE* loss-of-function mutations in a German cohort (Zimprich *et al*., 2001). Among these mutations were two nonsense changes, a deletion leading to a frameshift which results in a premature stop codon, a splice-site mutation, and a deletion including 15 bp of exon 4 (Zimprich *et al*., 2001). Since then, whole and partial *SGCE* deletions have been described, as well as greater than 35 point mutations, most of which result in premature termination of the protein (Peall *et al*., 2011).

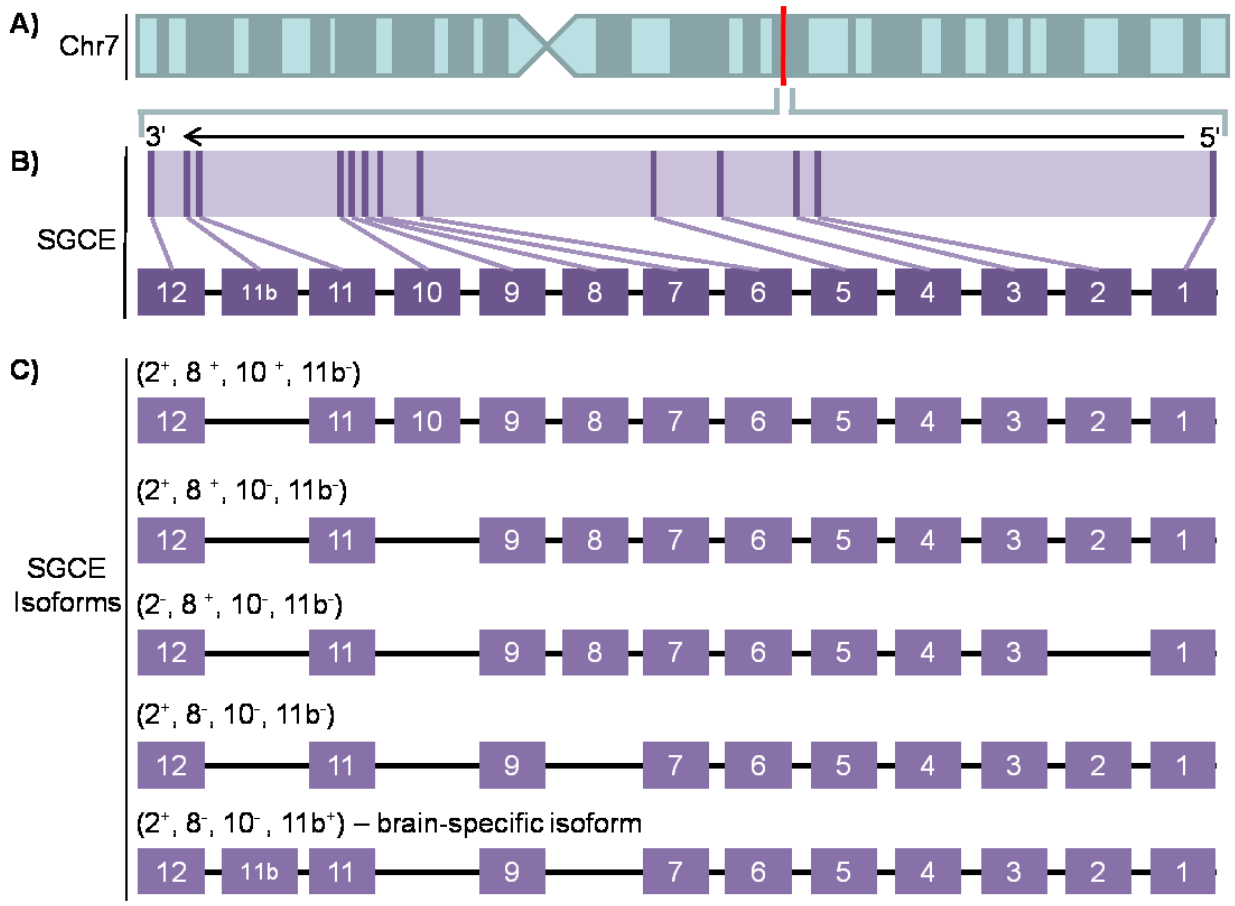
1.4. Characterizing the *SGCE* gene

Epsilon-sarcoglycan is a 70,986 bp-long gene on chromosome 7q21.3 (Fig. 1.1). It has 13 exons numbered 1-11, 11b, and 12. *SGCE* has four exons that undergo alternative splicing (2, 8, 10, and 11b), resulting in 5 known protein-coding isoforms of the gene (Ritz *et al*., 2011). The first isoform (2^{+}, 8^{+}, 10^{+}, $11b^{-}$) is 462 amino acids long, the second isoform (2^{+}, 8^{+}, 10^{-}, $11b^{-}$) is 437 amino acids long, the third isoform (2^{-}, 8^{+}, 10^{-}, $11b^{-}$) is 396 amino acids long, the fourth isoform (2^{+}, 8^{-}, 10^{-}, $11b^{-}$) is 428 amino acids long, and the fifth isoform (2^{+}, 8^{-}, 10^{-}, $11b^{+}$) is 451 amino acids long (Fig. 1.1). All of the isoforms are widely expressed in regions including skeletal muscle, heart, kidney, and liver, with the exception of (2^{+}, 8^{-}, 10^{-}, $11b^{+}$), which is a brain-specific isoform (Ritz *et al*., 2011). In general, all isoforms are expressed most prominently in the neurons of the brain, skeletal muscle, heart, kidney, and liver (Ritz *et al*., 2011). While the brain-specific isoform is expressed throughout the entire brain, the level of expression differs between various regions (Ritz *et al*., 2011). The highest levels of expression are seen in the primary somatosensory cortex and the motor cortex while the lowest levels of expression are seen in the globus pallidus (Ritz *et al*., 2011). Intermediate expression is observed in the caudate nucleus and substantia nigra (Ritz *et al*., 2011). There is a high degree

of variability in cerebellar and putamen expression between individuals (Ritz *et al*, 2011). The function of the brain-specific *SGCE* isoform is likely the key to the exclusively neurological phenotype of MD, however this function remains unclear (Ritz *et al*., 2011).

Reduced penetrance is also a common feature of MD caused by mutations in the epsilon-sarcoglycan gene (Muller *et al*., 2002). This reduced penetrance can be explained by *SGCE* imprinting in all tissues including brain (Grabowski *et al*., 2003). The maternal allele of epsilon-sarcoglycan is methylated, resulting in paternal allele expression only (Grabowski *et al*., 2003). Those individuals with a paternally inherited loss-of-function mutation exhibit MD, whereas individuals with a maternally inherited mutation are asymptomatic (Grabowski *et al*., 2003).

Figure 1.1. Characterization of the ε-sarcoglycan (*SGCE*) gene. A) *SGCE* occupies 71 kb on chromosome 7q21.3. B) *SGCE* is composed of 13 exons referred to as exon 1-11, 11b, and 12. The 5′ end of the gene is telomeric to the 3′ end. C) There are 5 major isoforms of *SGCE*. All isoforms share 9 of the same exons, with exons 2, 8, 10, and 11b being alternatively spliced. The (2+, 8 +, 10 +, 11b-), (2+, 8 +, 10-, 11b-), (2-, 8 +, 10-, 11b-), and (2+, 8-, 10-, 11b-) isoforms are ubiquitously expressed, whereas the (2+, 8-, 10- , 11b+) isoform is brain-specific.


A)
Chr7
B)
3'
5'
SGCE
12
11b
11
10
9
8
7
6
5
4
3
2
1
C)
SGCE
Isoforms
(2+, 8 +, 10 +, 11b-)
(2+, 8 +, 10-, 11b-)
(2-, 8 +, 10-, 11b-)
(2+, 8-, 10-, 11b-)
(2+, 8-, 10- , 11b+) – brain-specific isoform

1.5. Functional characterization of the SGCE protein

SGCE is an N-glycosylated transmembrane protein belonging to the sarcoglycan family. Other sarcoglycan family members include α-, β-, γ-, δ-, and ζ-sarcoglycan. Very little is known about the tissue-specific functions of each of the sarcoglycans, particularly within the brain. However, in muscle the sarcoglycans have been shown to form a heterotetrameric complex that is a member of the dystrophin-associated protein complex, which mediates the structural stability of the plasma membrane and forms interactions with the extracellular matrix and cytoskeleton (Waite *et al*., 2009). Limb-girdle muscular dystrophies (caused by instability of the muscle cell membrane) result from mutations in α-, β-, γ-, or δ-sarcoglycan (Nishiyama *et al*., 2004).

Epsilon-sarcoglycan was initially identified as a homologue of α-sarcoglycan (Ettinger *et al*., 1997; McNally *et al*., 1998). The coding region of α- and ε-sarcoglycan shares 47% homologous nucleotide sequence and 62% homologous amino acid sequence (Ozawa *et al*., 2005). Whereas the expression of α-sarcoglycan is restricted to striated muscle, ε-sarcoglycan is expressed in many different tissues including muscle, kidney, lung, liver, spleen, testis, and brain (Nishiyama *et al*., 2004). Epsilon-sarcoglycan expression occurs earlier than that of α-sarcoglycan during myogenesis, however its expression decreases concomitantly with the increase in expression of α-sarcoglycan during developmental progression (Ozawa *et al*., 2005). Similar to α-sarcoglycan, ε-sarcoglycan has also been shown to form complexes with β-, γ-, and δ-sarcoglycan in muscle; in the Schwann cells surrounding peripheral nerve fibers it has been shown to form a complex with β- and δ-sarcoglycan (Nishiyama *et al*., 2004). In 2004, Nishiyama *et al*. determined that there appears to be a particular enrichment of ε-sarcoglycan in the pre- and post-synaptic membranes of neuronal cells in the mouse brain, suggesting a synaptic function for the protein in the central nervous system.

In order to obtain a representative disease model for Myoclonus-Dystonia, Yokoi *et al*. (2006) generated a paternally-inherited epsilon-sarcoglycan (*Sgce)* heterozygous knockout mouse. The *Sgce* knockout mouse exhibited myoclonus, impaired motor coordination, learning, and balance, as well as anxiety and depression, all of which are symptoms of MD in humans (Yokoi *et al*., 2006). In addition, these mice exhibited high serotonin turnover and hyperdopaminergic striatum (Yokoi *et al*., 2006). More recently, Zhang *et al*. (2012) demonstrated that the *Sgce* knockout mouse has significantly reduced levels of the striatal dopamine D2 receptor (D2R), as well as an increase in striatal dopamine discharge upon amphetamine injection. Later work by this same group revealed that the cerebellar Purkinje cells of the *Sgce* knockout mouse have abnormal nuclear envelopes (Yokoi *et al*., 2012). Abnormal nuclear envelopes were also found in the striatal medium spiny neurons in a separate study involving the *Sgce* knockout mouse (Yokoi *et al*., 2012). However, abnormal nuclear envelopes were not observed when *Sgce* was knocked out in Purkinje cells alone or striatal medium spiny neurons alone (although motor deficits did occur); these results suggest that the MD phenotype is a result of the dysfunction of multiple cell types within the brain (Yokoi *et al*., 2012).

While the precise function of the ε-sarcoglycan protein remains unclear, research findings suggest that Myoclonus-Dystonia results from the dysfunction of its brain-specific role; that brain-specific role appears to include the synaptic transmission of dopamine.

1.6. Identification and refinement of the DYT15 locus

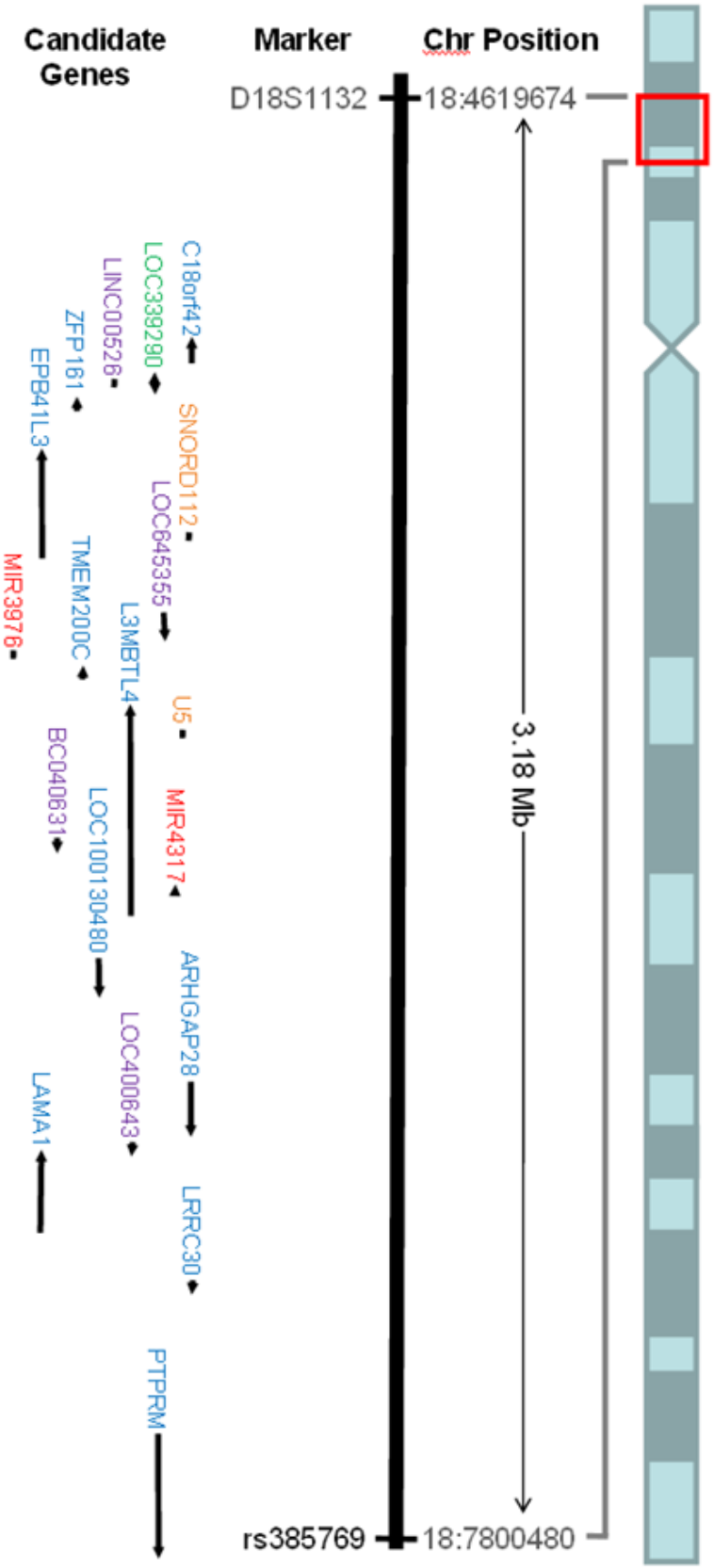
In 2001, Grimes *et al*. described a large, 5 generation Canadian family with inherited Myoclonus-Dystonia and no mutations in the previously characterized dystonia gene, *DRD2* (also known as D2R) (Grimes *et al*., 2001). By 2002, this group had also eliminated mutations

in *SGCE* as the cause of the disease in this family and had identified a novel locus for inherited Myoclonus-Dystonia on 18p11, DYT15 (Grimes *et al*., 2002). Two-point and multipoint genome-wide linkage analysis localized a gene for MD in this family to a 17-cM region on chromosome 18p11 (Grimes *et al*., 2002). Further refinement of the locus to a 3.18 Mb region was achieved in 2007 by using single nucleotide polymorphisms for fine-mapping (Fig. 1.2; Han *et al*., 2007). The identification of the causative gene for MD in this family, referred to as Myoclonus Family 1 (MF1), became the main focus of this research group and continues to be the main focus of this thesis project. MD in this family exhibits an autosomal dominant mode of inheritance with incomplete penetrance.

Once the critical region on 18p11 was defined, the exons and exon-intron boundaries of all of the known REFSEQ genes within that region were sequenced and screened for mutations. In total, 358 novel SNPs, insertions, and deletions were found, however all were detected in control samples and therefore no causative coding or splice site mutation was found (unpublished data). In addition to direct sequencing, a genome-wide gene dosage analysis was performed using a 500K SNP Array in order to search for potential large deletions or duplications; no significant dosage alterations were detected.

With the passage of time, new technologies become available that aid in the detection of genetic mutations. While Sanger sequencing and gene dosage analysis methods have been used to attempt to identify the causative mutation for MD on 18p11, this mutation remains elusive. More recently, next-generation sequencing technologies have become available that allow for comprehensive screening of the entire disease-linked region. It is through the use of these advancing technologies that we hope to identify the MD-causing mutation in MF1.

Figure 1.2. Schematic representation of DYT15, a 3.18 Mb region on 18p11. Candidate genes and their physical distances were determined using UCSC (http://genome.ucsc.edu/). The genes are categorized based on function: blue denotes protein-coding genes, green denotes mRNAs with a hypothetical protein sequence, purple denotes non-coding RNAs, orange denotes small nucleolar and small nuclear RNAs, and red denotes microRNAs. This map is based on the Feb. 2009 (GRCh37/hg19) build.


Candidate Genes
Marker
Chr Position
D18S1132
18:4619674
3.18 Mb
rs385769
18:7800480
C18orf42
LOC339290
LINC00526
ZFP161
EPB41L3
MIR3976
SNORD112
LOC645355
L3MBTL4
TMEM200C
U5
BC040631
MIR4317
LOC100130480
ARHGAP28
LOC400643
LAMA1
LRRC30
PTPRM

1.7. Next-generation sequencing

In 1977, Sanger *et al*. developed the first practical method for DNA sequencing. The Sanger method involves the use of chain-terminating inhibitors during DNA sequencing (Sanger *et al*., 1977). The use of dideoxyribonucleoside triphosphates (ddNTP) in place of deoxyribonucleoside triphosphates (dNTP) results in chain termination during strand synthesis by DNA polymerase I (Sanger *et al*., 1977). By incubating a template, primer, and DNA polymerase I along with a mixture of labelled, terminating ddNTPs and non-labelled, non-terminating dNTPs, a series of fragments can be generated that all share the same 5′ end but differ in the point at which a ddNTP terminated the reaction at the 3′ end (Sanger *et al*., 1977). By separating all of the fragments from each of the four deoxyribonucleoside triphosphate reactions into different lanes on a denaturing acrylamide gel, one can decipher the DNA sequence based on the banding pattern (Sanger *et al*., 1977). Modifications and improvements to the method have led to a 10,000-fold reduction in cost over the last 20 years and have resulted in sequence read lengths of up to 1,000 bases (Pettersson *et al*., 2009). For several decades the Sanger sequencing method remained the most reliable technique; it is only within the last 10 years that new commercially available platforms have been introduced that rival Sanger *et al*.'s method (Moorthie *et al*., 2011).

Currently, the sequencing technologies can be divided into three generations (Moorthie *et al*., 2011). Sanger sequencing represents the first generation of sequencing methods, where a given reaction reveals the allelic content of a predefined target (Moorthie *et al*., 2011). The later generations of sequencing methods are defined by their ability to accomplish massive parallelization (Moorthie *et al*., 2011). The second generation of sequencing refers to "next-generation sequencing (NGS)," where clonal amplification of the target precedes the sequencing

steps (Moorthie *et al*., 2011). There are currently three main NGS platforms that use three different sequencing methodologies: Illumina/Solexa which utilizes a reversible termination method; Roche 454 which utilizes a pyrosequencing method; and Life Technologies/Applied Biosystems which utilizes a sequencing-by-ligation method (Moorthie *et al*., 2011).

The reversible termination method employed by Illumina/Solexa was introduced by Bentley *et al*. in 2008 and is the most closely related to the Sanger sequencing-by-synthesis method (Moorthie *et al*., 2011). In the reversible termination method, individual fragmented DNA molecules are hybridized to a glass flow cell, where they undergo bridge amplification followed by sequencing (Bentley *et al*., 2008; Moorthie *et al*., 2011). The sequencing-by-synthesis reaction occurs similarly to Sanger sequencing: the template undergoes chain extension by DNA polymerase I using labelled terminating ddNTPs (each of the four nucleotides are differentially labelled with fluorescent dyes); once each template on the flow cell has been extended by one base pair, color imaging can be used in order to determine which base was incorporated on each target (Bentley *et al*., 2008; Moorthie *et al*., 2011). The key difference between the reversible termination method and Sanger sequencing is that once a signal has been read to determine which base was incorporated, the dye and terminating group of the incorporated ddNTP can be chemically cleaved and another round of base incorporation and signal imaging can begin on the same targets as the previous reaction (Bentley *et al*., 2008; Moorthie *et al*., 2011). This method effectively increases the throughput and decreases the cost compared to the Sanger sequencing method (Bentley *et al*, 2008). Unlike Sanger sequencing, however, read length is limited to approximately 75 base pairs (Tucker *et al*., 2009).

The pyrosequencing method employed by Roche 454 was first introduced by Margulies *et al*. (2005) and is also a sequencing-by-synthesis technique. Unlike the termination methods,

however, pyrosequencing measures the amount of light emitted upon base incorporation rather than the fluorescent signal of a given nucleotide (Margulies *et al*., 2005; Moorthie *et al*., 2011). The Roche 454 method will be described in more detail in the materials and methods section, given that it was the sequencing method of choice for this project. In summary, template fragments are first separated and clonally amplified by emulsion PCR (Margulies *et al*., 2005). The amplified fragments then undergo sequencing, whereby nucleotides are added one at a time and the successful incorporation of a given nucleotide is measured by the amount of light emitted following an enzymatic cascade initiated by the release of an inorganic phosphate ion (Margulies *et al*., 2005). The sequential addition of nucleotides and detection of light emission allows for the elucidation of template sequence. While the pyrosequencing method is the most expensive of the second-generation sequencing platforms, it generates the longest read lengths (approximately 400 bp), making it extremely advantageous (Tucker *et al*., 2009).

The sequencing-by-ligation method employed by Life Technologies/Applied Biosystems is different from the sequencing-by-synthesis methods because it does not involve strand synthesis by DNA polymerase (Moorthie *et al*., 2011). Instead, this method depends on the ligation of fluorescently labelled hybridization probes in order to reveal the sequence of the template (Moorthie *et al*., 2011). This method was first described by Shendure *et al*. in 2005 and is now referred to as the SOLiD (Sequencing by Oligonucleotide Ligation and Detection) system. The first steps of SOLiD sequencing are similar to those of the Roche 454 method: template DNA is fragmented, adaptors are ligated to the fragments, the fragments are hybridized to beads and then amplification of the fragments occurs by emulsion PCR (Moorthie *et al*., 2011). Once the amplification steps are completed, the emulsion is broken and the beads are transferred to a glass slide (Moorthie *et al*., 2011). A universal primer then binds to the adaptor

ligated to each template fragment and fluorescently labelled oligonucleotide probes are added (Shendure *et al*., 2005). Each oligonucleotide probe contains two bases that define its sequence specificity; when the bases of a given probe match the sequence on the template strand, DNA ligase links that probe to the 3′ end of the growing strand (which began with the universal primer) (Moorthie *et al*., 2011). The unbound probes can then be washed away and the incorporated probe sequence (and therefore the template sequence) can be determined based on the fluorescent signal detected (Moorthie *et al*., 2011). Cleavage of the fluorescent dye and degenerate bases from the incorporated probe then allows for the cycle to be repeated, allowing the next two bases of the template strand to be read and so on (Shendure *et al*., 2005; Moorthie *et al*., 2011). SOLiD is the cheapest method of next-generation sequencing, however the read length is the shortest at approximately 50 base pairs (Tucker *et al*., 2009).

Currently scientists are working towards the "next-next" or third generation of sequencing. The key features of these new methods will be the ability to sequence single molecules of DNA (eliminating the need for clonal amplification) in real-time (Moorthie *et al*., 2011). These methods will use a wide range of detection methods including capturing fluorescent, electronic, and atomic signals (Moorthie *et al*., 2011).

Within the last decade the field of genetics has undergone unprecedented advancement in the technology available for DNA sequencing. This has allowed for a significant improvement in our ability to identify and characterize novel genes, discover disease-causing mutations, and track polymorphisms within specific populations. It was through the use of this technology that we hoped to ascertain the DYT15 mutation that had eluded us for so long.

1.8. Rationale, Hypothesis, and Aims

1.8.1. Rationale

There is strong evidence for genetic heterogeneity in Myoclonus-Dystonia, given that mutations in epsilon-sarcoglycan only account for 30-50% of MD cases and that a second locus, DYT15, has been identified. Our group alone has recruited 36 MD families to our study, 26 of which are lacking mutations in *SGCE*; the frequency of *SGCE* mutations in our cohort is consistent with the literature at 27%. For these reasons, there is likely to be at least one additional gene responsible for causing MD in the remaining 50-70% of patients lacking mutations in the epsilon-sarcoglycan gene. The Myoclonus Family 1 linkage to chromosome 18p11 provides a 3.18 Mb region where DYT15 lies. Given that the majority of the genes within the linked region have already been sequenced without the identification of a disease-causing variant, the DYT15 mutation is likely non-coding.

1.8.2. Hypothesis

I hypothesize that mutations in at least one additional gene are responsible for causing Myoclonus-Dystonia and that one of these mutations exists on 18p11 which is responsible for causing MD in Myoclonus Family 1.

1.8.3. Aims

In order to address my hypothesis, I will achieve the following objectives:

1. To identify using next-generation sequencing the sequence variants likely causing MD in Myoclonus Family 1.

Through the use of targeted capture of the critical region and massively-parallel sequencing by Roche 454, we sequenced all coding and non-coding bases within the linked region on 18p11 for two members of MF1. Each patient's sequence reads were then aligned to a reference genome in order to identify individual sequence variants that are shared between the two affected family members.

2. To assess the likelihood that a given variant is disease-causing.

In order to prioritize the variants identified by next-generation sequencing in terms of their disease-causing potential, they were categorized based on their proposed function. For example, exonic mutations were considered to be more likely to be disease-causing than intronic mutations given their direct implications for protein alteration. Following the categorization and subsequent prioritization of each of the shared variants on 18p11, their disease-causing potential was assessed based on the following criteria: the variant had to validate by Sanger sequencing; the variant had to co-segregate with the disease in the family; and the variant had to be absent in unaffected control individuals.

3. To characterize the function of the mutation and determine the mechanism by which is causes MD.

Upon finding a likely candidate for the disease-causing mutation in MF1, the functional implications of that mutation had to be determined. Given that the mutation I identified occurred within a reported EST, I first had to confirm that that EST was in fact an expressed portion of the genome, resulting in an RNA transcript. I then had to characterize that RNA transcript by elucidating its size, expression pattern, and potential function.

4. To screen for the MD-causing mutation identified in Myoclonus Family 1 in other *SGCE* mutation-negative families.

The identification of a novel MD-causing mutation in MF1 provides a potential alternative to *SGCE* in other MD families. Our MD cohort includes 26 patients outside of MF1 that tested negative for mutations in epsilon-sarcoglycan. All 26 patients were screened for the same mutation identified in MF1, as well as for novel mutations in the same transcript as the mutation found in MF1.

Chapter 2: Subjects and General Methodology

2.1. Myoclonus-Dystonia patient cohort, screening history, and phenotypic characteristics.

Patients were recruited by our collaborating neurologist, Dr. David A. Grimes, from his practice as well as from other movement disorder clinics across Canada and the United States. All recruited patients were assessed by a movement disorder neurologist and signed consent forms approved by the Ottawa Hospital Research Ethics Board. Parental consent was obtained for patients under the age of 16. Diagnosis of Myoclonus-Dystonia was made according to the criteria outlined by Grimes *et al*. (2001).

Myoclonus-Dystonia Family 1 (MF1) includes 46 members, 36 of which were initially examined, videotaped, and had DNA extracted from venous blood drawn (Grimes *et al*., 2002; Fig. 2.1). There are 14 affected individuals and 7 individuals that carry the disease haplotype but are unaffected (Fig. 2.1). Myoclonus-Dystonia Family 1 has been screened for mutations in *SGCE*, as well as the D2 dopamine receptor gene; no mutations were found (Grimes *et al*., 2001).

In addition to MF1, 36 MD families have been recruited to our study (MF2, MF3, MF4, MF5, MF6, MF7, MF8, MF9, MF11, MF12, MF13, MF14, MF15, MF16, MF17, MF18, MF19, MF20, MF21, MF22, MF23, MF24, MF25, MF26, MF27, MF28, MF29, MF30, MF31, MF32, MF33, MF34, MF35, MF36, MF37, and MF38). The cohort families are numbered according to the order in which they were recruited into the study. All affected members of each family were screened for mutations in *SGCE*. Of the 36 families, 10 tested positive for mutations in *SGCE*: MF2, MF3, MF4, MF6, MF7, MF15, M21, MF28, MF29, and MF34. The *SGCE* mutation-

positive families have been excluded from further analysis; the remaining 26 *SGCE* mutation-negative families will be used to validate the mutation found on 18p11 (Fig. 2.2).

The phenotypic characteristics of each of the patients included in the Myoclonus-Dystonia patient cohort are summarized in Table 2.1.

2.2. DNA extraction

Whole blood was drawn from patients using yellow top ACD Vacutainer tubes. Genomic DNA was then extracted from whole blood using the QIAamp® DNA Mini and Blood Mini Kit (QIAGEN). The DNA Purification from Blood Spin Protocol was followed as per the kit handbook. In brief, 20 µl of protease was combined with 200 µl of whole blood, then 200 µl of Buffer AL (lysis buffer) was added and the sample was vortexed for 15 seconds. Samples were incubated for 10 minutes at 56°C, followed by the addition of 200 µl of 100% ethanol, and vortexing for 15 seconds. The sample was then loaded into the QIAamp Mini spin column and centrifuged for 1 minute at 8000 rpm. Following the transfer of the column to a clean collection tube, the filtrate was discarded and 500 µl of Buffer AW1 was added to the column. The column was then centrifuged at 8000 rpm for 1 minute. A second wash with Buffer AW2 was performed in the same way as described for Buffer AW1, however the sample was centrifuged for 3 minutes at 14000 rpm. After discarding the filtrate and placing the column in a new collection tube, the column was spun at full speed for an additional minute in order to ensure that no carryover of Buffer AW2 occurred. The DNA sample was eluted from the column using 100 µl of Buffer AE and the concentration of the sample was determined using a NanoDrop spectrophotometer (Thermo Scientific).

2.3. Targeted capture and Roche 454 sequencing

Two meiotically distant affected members of MF1 (MF1-3 and MF1-5; Fig. 2.1) were chosen to undergo targeted capture followed by Roche 454 sequencing. NimbleGen Sequence Capture technology was used to enrich for DNA from the chr18:4,619,674-7,800,480 critical region plus 50 kb flanking at 323 and 435 fold for MF1-3 and MF1-5, respectively. The array design covered >97% of the target sequence. The sequence capture method involves the following: first, the genomic DNA samples are fragmented and linkers are ligated to the ends of each fragment; second, the fragments are hybridized to a custom array containing oligonucleotide probes specific to the region of interest; third, unwanted genomic fragments are washed away as only those fragments from the target region will be hybridized to the probes on the array; fourth, the target fragments are eluted; and fifth, the fragments of interest are amplified using universal primers specific to the linkers ligated to the fragments early on in the protocol. Once an amplified sample of the target region is obtained, that sample can undergo sequencing.

The Roche 454 method can be summarized in the following steps: 1) template DNA is fragmented and universal adaptors are ligated to the ends; 2) the template fragments are bound to beads (so that there is a single fragment hybridized to each individual bead), which are then captured in oil droplets containing all of the appropriate PCR reagents, resulting in individual oil reaction chambers for each template; 3) each template fragment undergoes amplification by emulsion PCR, resulting in 10 million copies of each fragment per bead; 4) the oil emulsion is broken and the beads are loaded into the wells of a fibre-optic "PicoTitre Plate" along with separate beads containing the necessary reagents for the sequencing reaction (with the exception of the nucleotides); 5) nucleotides are added one at a time, if a given nucleotide is successfully incorporated an inorganic phosphate ion will be released and an enzyme cascade will be initiated

that results in the emission of light; 6) the intensity and location of light emission across the plate can be measured in order to determine if a base was incorporated; by repeating the addition of each of the four bases in a cyclical fashion, the template sequence for each fragment can be deduced (Margulies *et al*., 2005; Moorthie *et al*., 2011).

2.4. Primer design and PCR

For each variant candidate, the genomic DNA sequence (assembly GRCh27/hg19) was obtained from the University of California Santa Cruz (UCSC) Genome Brower website (http://www.genome.ucsc.edu/). Primers were then designed using the Integrated DNA Technologies (IDT) PrimerQuest^{SM} tool (http://www.idtdna.com/Scitools/Applications/Primerquest/) with an optimum primer size of 24 nt, an optimum melting temperature (T_M) of 60°C, and an optimum G/C content of 50%.

Polymerase Chain Reaction (PCR) was performed using the HotStarTaq DNA Polymerase and PCR Buffer from QIAGEN. Each primer set was optimized individually by including or excluding Q-Solution (QIAGEN), or by varying the annealing temperature, extension time, and/or cycle number. Following PCR, products were separated by agarose gel electrophoresis in order to confirm amplicon size and specificity.

2.5. Sanger sequencing

Sanger sequencing was performed by the McGill University and Génome Québec Innovation Centre (Montreal, QC). In preparation for Sanger sequencing of a given sample, 12.5-500 ng (depending on the length of the amplicon) of unpurified PCR product was sent to the Innovation Centre, along with a 10 µl volume of the appropriate forward and reverse primers

at a concentration of 5 µM. Each sample required two separate submissions, one for a forward sequencing reaction and one for a reverse sequencing reaction. PCR product sequences could then be retrieved from the Innovation Centre portal, Nanuq, and aligned to a reference sequence for analysis. Sequence alignment and variant analysis of Sanger results were performed using the Geneious™ Sequence Analysis Software from Biomatters Ltd.

2.6. RNA extraction and cDNA synthesis

RNA was extracted from human lymphoblast cells using the RNeasy® Mini Kit (QIAGEN). The Purification of Total RNA from Animal Cells Using Spin Technology Protocol was followed as outlined in the kit handbook. Briefly, lymphoblast cells grown in suspension were pelleted by centrifugation at 300 x g for 5 minutes. The supernatant was removed and the cells were resuspended in 600 µl of Buffer RLT (lysis buffer). The sample was then homogenized using a QIAshredder spin column, which was placed in a collection tube and centrifuged at full speed for 2 minutes. Following homogenization, 1 volume of 70% ethanol was added to the lysate and the sample was applied to an RNeasy spin column. The column was centrifuged at ≥8000 x g for 15 seconds and the flow-through was discarded. 700 µl of Buffer RW1 was then applied to the spin column to wash the membrane and the sample was centrifuged at ≥8000 x g for 15 seconds. After the flow-through was discarded, 500 µl of a second wash solution, Buffer RPE, was applied to the column and the same 15 second centrifugation was repeated. A second 500 µl volume of Buffer RPE was added to the column once the flow-through had been removed and the sample was centrifuged at ≥8000 x g for 2 minutes. The column was transferred to a clean collection tube and an additional centrifugation at full speed was performed for 1 minute to ensure that no residual Buffer RPE remained in the column. The

column was then placed in a clean collection tube and the addition of 30 µl of RNase-free water followed by a 1 minute centrifugation at ≥8000 x g was used to elute the RNA. The concentration of the sample was determined using a NanoDrop spectrophotometer (Thermo Scientific).

First-strand cDNA synthesis from the RNA samples was performed using the SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen). Each individual cDNA synthesis reaction included 4 µl of 5X VILO™ Reaction Mix, 2 µl of 10X SuperScript® Enzyme Mix, 1-2 µg of RNA, and enough DEPC-treated water to bring the volume to 20 µl. Once all reagents were added, the sample underwent the following incubations: 25°C for 10 minutes; 42°C for 60 minutes; and 85°C for 5 minutes. cDNA samples were stored at -20°C between uses.

2.7. Northern Blot

The Northern Blot experiment was performed according to the protocol outlined in the formaldehyde-based Ambion® NorthernMax® Kit (Life Technologies). First, a denaturing agarose gel was made by melting 0.4 g of agarose in 36 mL of RNase-free water, then adding 4 mL of 10X denaturing buffer. The gel was poured to a thickness of 0.6 cm and allowed to solidify. The molecular weight marker (a 0.24-9.5 Kb RNA ladder from Invitrogen), human brain, patient lymphoblast, and control lymphoblast RNA samples were prepared by combining 3 µg of RNA with 3 volumes of Formaldehyde Load Dye and then incubating them at 65°C for 15 minutes. The gel was then set up in the electrophoresis chamber and covered with 1X MOPS Gel Running Buffer. The RNA samples were loaded onto the gel, which was then run at 90 V until the dye front reached the bottom of the gel. Once the RNA samples had been separated on the agarose gel, they were transferred to an Ambion® BrightStar®-Plus membrane by capillary

action (Southern, 1975). To complete the transfer, a stack of materials was constructed in the following order (from bottom to top): 3 cm of paper towels cut to be slightly larger than the size of the gel; 3 dry pieces of Whatman filter paper (cut to the same size as the gel); 2 pieces of filter paper soaked in Transfer Buffer; a piece of BrightStar-Plus membrane (cut to the same size as the gel), made wet with Transfer Buffer; the agarose gel (with the area including and above the wells trimmed away); 3 more pieces of wet filter paper; 3 pieces of wet filter paper bridges that extend from the stack of materials into the Transfer Buffer reservoir; a piece of plastic (the empty agarose gel tray); and a 150 g weight. The transfer occurred over 3 hours, after which the transfer setup was disassembled and the RNA was crosslinked by using a commercial ultraviolet crosslinker. The membrane was then wrapped in plastic and stored at -20°C until the prehybridization and hybridization steps one week later.

In order to make the CD108131 probe, a cDNA template (made from a control lymphoblast cell line, see RNA extraction and cDNA synthesis protocol above) was used to generate an 824 bp product specific to the CD108131 transcript. The forward primer sequence was 5′-GTCTGAATTTGCAGTTACTGAGAGGC-3′ and the reverse primer sequence was 5′-GTGAGGTTGACACTAACAAATTCACCC-3′. The PCR was performed using the HotStarTaq DNA Polymerase and PCR Buffer from QIAGEN and the reagent concentrations in the mastermix were determined according to the protocol outlined in the HotStarTaq instructions. An annealing temperature of 58°C and 35 PCR cycles were used. The size of the PCR product was then verified on a 1% agarose gel to ensure specificity, extracted from the gel using the QIAquick Gel Extraction Kit from QIAGEN, and eluted from the QIAquick column with 30 µl of ddH_2O. The CD108131 PCR product was then used to generate an α-^{32}P-CTP-labeled probe according to the NEBlot® Kit protocol from New England BioLabs Inc. The

probe was then purified using the AutoSeq G-50 spin columns from Amersham Biosciences according to the manufacturer's instructions.

Following probe synthesis, the ULTRAhyb Buffer was preheated to 68°C. The membrane was then prehybridized in approximately 10 mL of the buffer for 30 minutes at 42°C. The probe was diluted 10-fold with 10 mM EDTA and heat denatured at 90°C for 10 minutes. After the 30 minute incubation, the denatured probe was added directly to the buffer used during the prehybridization step at a concentration of 10^6 cpm per mL and hybridized to the membrane overnight at 42°C. The following day, the membrane was washed twice in 20 mL of Low Stringency Wash Solution #1 for 5 minutes at room temperature (with agitation) each time. Two 15 minute washes with High Stringency Wash Solution #2 at 42°C (with agitation) were then performed. Once the washes were complete, the membrane was sealed in plastic wrap, placed in an exposing cassette (containing intensifying screens) along with X-ray film, and stored at -80°C overnight. The following day the film was developed to visualize the results.

In order to visualize an endogenous control on the membrane, a β-actin probe was generated in the same way as the CD108131 probe (described above) using a PCR product included within the NorthernMax Kit. Once the film had been developed to visualize the CD108131 results, the membrane was stripped by incubating it in the High Stringency Wash Solution #2 for 15 minutes at 72°C. The prehybridization, hybridization, and visualization steps were then performed as described above, however this time the β-actin probe was used.

2.8. Quantitative Real-Time PCR

All Quantitative Real-Time PCR (qRT-PCR) experiments were performed according to the TaqMan® Gene Expression Assays Protocol from Applied Biosystems. The *ZFP161*, *ACTB*,

and *SGCE* probes were all pre-designed TaqMan® Gene Expression Assays (catalog #s Hs01924092_s1, Hs01060665_g1, and Hs00938948_m1 respectively). A custom TaqMan® Gene Expression Assay was obtained, based on the transcript sequence of CD108131 in the UCSC Genome Browser. Each qRT-PCR reaction contained 1X TaqMan® Gene Expression Assay, 1X TaqMan® Gene Expression Master Mix, 1-100 ng of cDNA template, and 5 µl of RNase-free water. Standard curves were generated by pooling all cDNA templates and performing qRT-PCR on 5, 4-fold serial dilutions of the pooled samples. Each reaction was performed in triplicate or quadruplicate. No Reverse-Transcriptase (RNA samples prior to cDNA synthesis) and H_2O negative controls were run in conjunction with the cDNA template reactions to ensure the absence of genomic (or other) contamination. The qRT-PCR was run according to the following protocol: 50°C for 2 minutes; 95°C for 10 minutes; and 40 cycles of 95°C for 15 seconds followed by 60°C for 1 minute. Following all qRT-PCR experiments, the products were separated on an agarose gel to confirm specificity.

Figure 2.1. Pedigree and chr18p11 haplotypes of Myoclonus-Dystonia Family 1 (MF1). 46 family members are illustrated in this pedigree, 14 of which are symptomatic and 7 of which are asymptomatic but have the chr18p11 disease haplotype, which is indicative of incomplete penetrance. Haplotypes are listed for the chr18p11 microsatellite markers; the blue markers fall within the 3.18 Mb critical region. The disease haplotype is identified by a shaded rectangle. Circles, female; squares, male; slash, individual deceased; shaded, affected; +, blood collected from individual; arrows, recombination event within the region allowed for refinement of the disease haplotype. The two individuals indicated by red boxes were sent for targeted capture of the critical region followed by Roche454 sequencing.

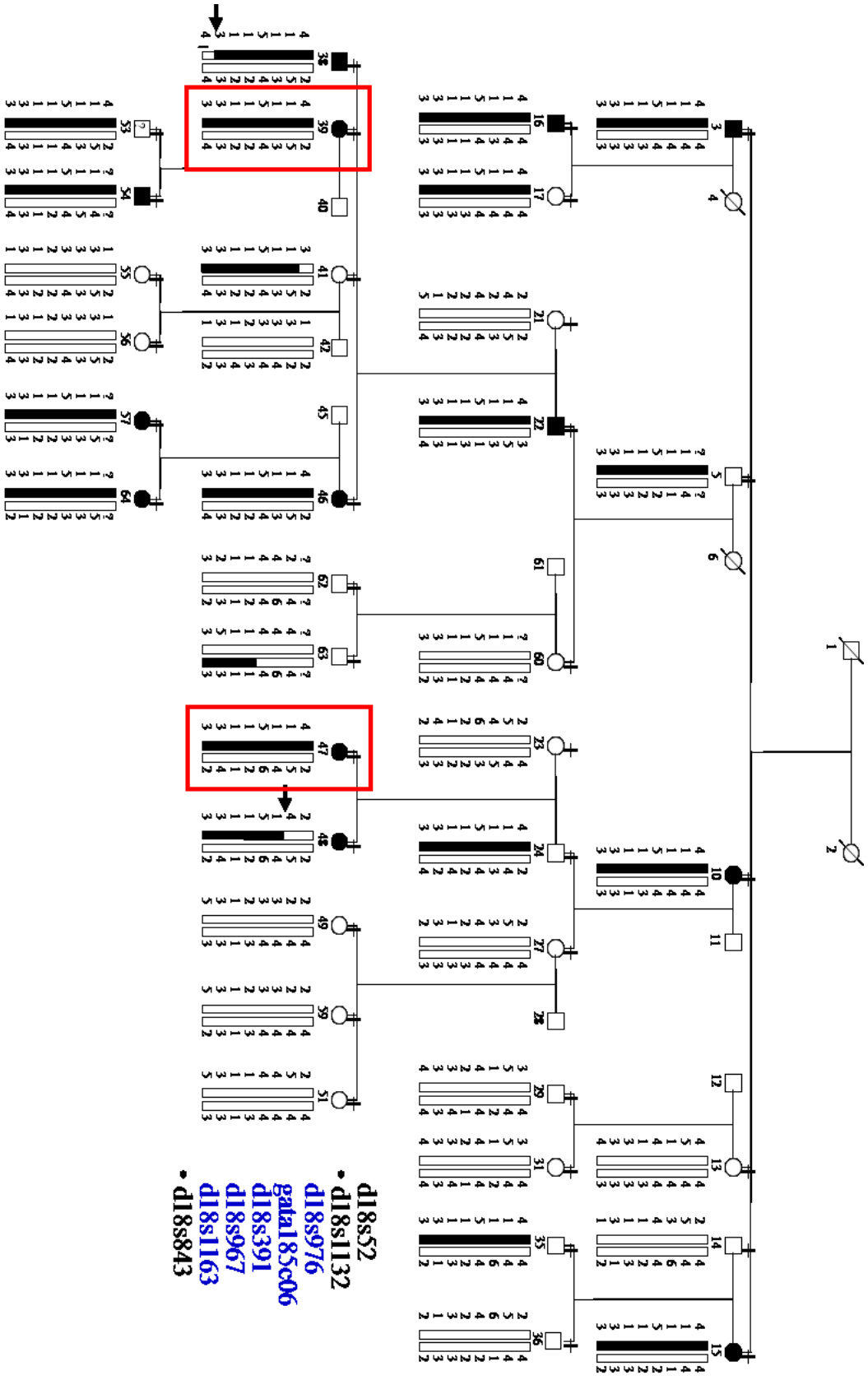
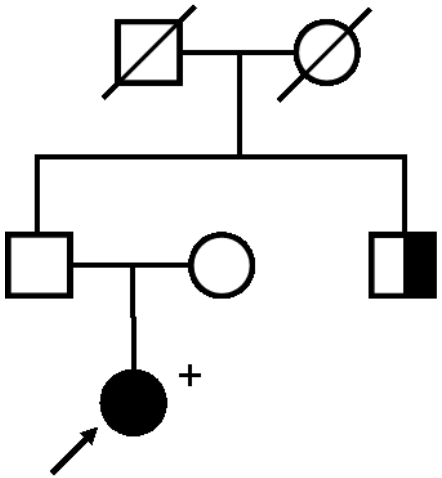
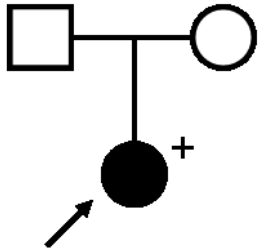

d18s52
• d18s1132
d18s976
gata185c06
d18s391
d18s967
d18s1163
• d18s843

Figure 2.2. Pedigrees for Myoclonus-Dystonia patient cohort families 5, 8, 9, 11, 12, and 13. Circles, female; squares, male; slash, individual deceased; shaded, affected; +, blood collected from individual; arrows, proband; ?, possibly affected; half-shaded symbols represent individuals who are believed to be affected but have not been seen by a neurologist.

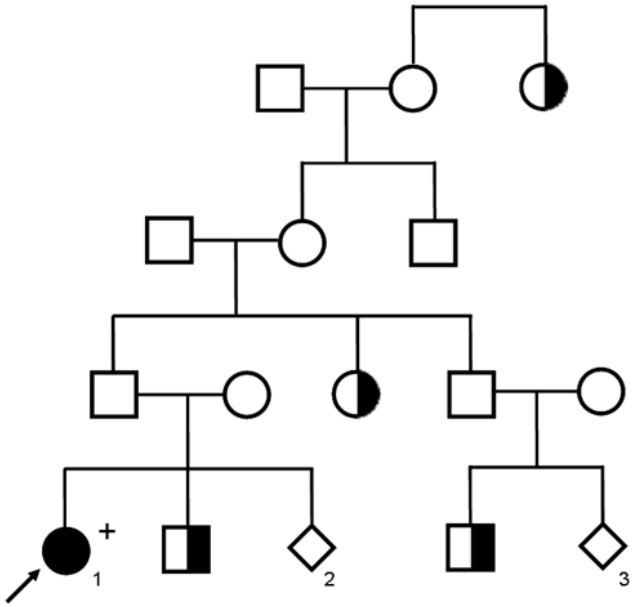
MF5



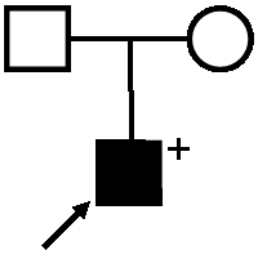
MF8



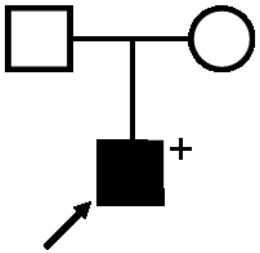
MF9



MF11



MF12



MF13

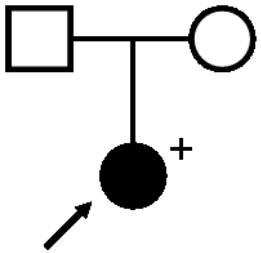


Figure 2.3. Pedigrees for Myoclonus-Dystonia patient cohort families 14, 16, 17, 18, 19, and 20. Circles, female; squares, male; slash, individual deceased; shaded, affected; +, blood collected from individual; arrows, proband; ?, possibly affected; half-shaded symbols represent individuals who are believed to be affected but have not been seen by a neurologist.

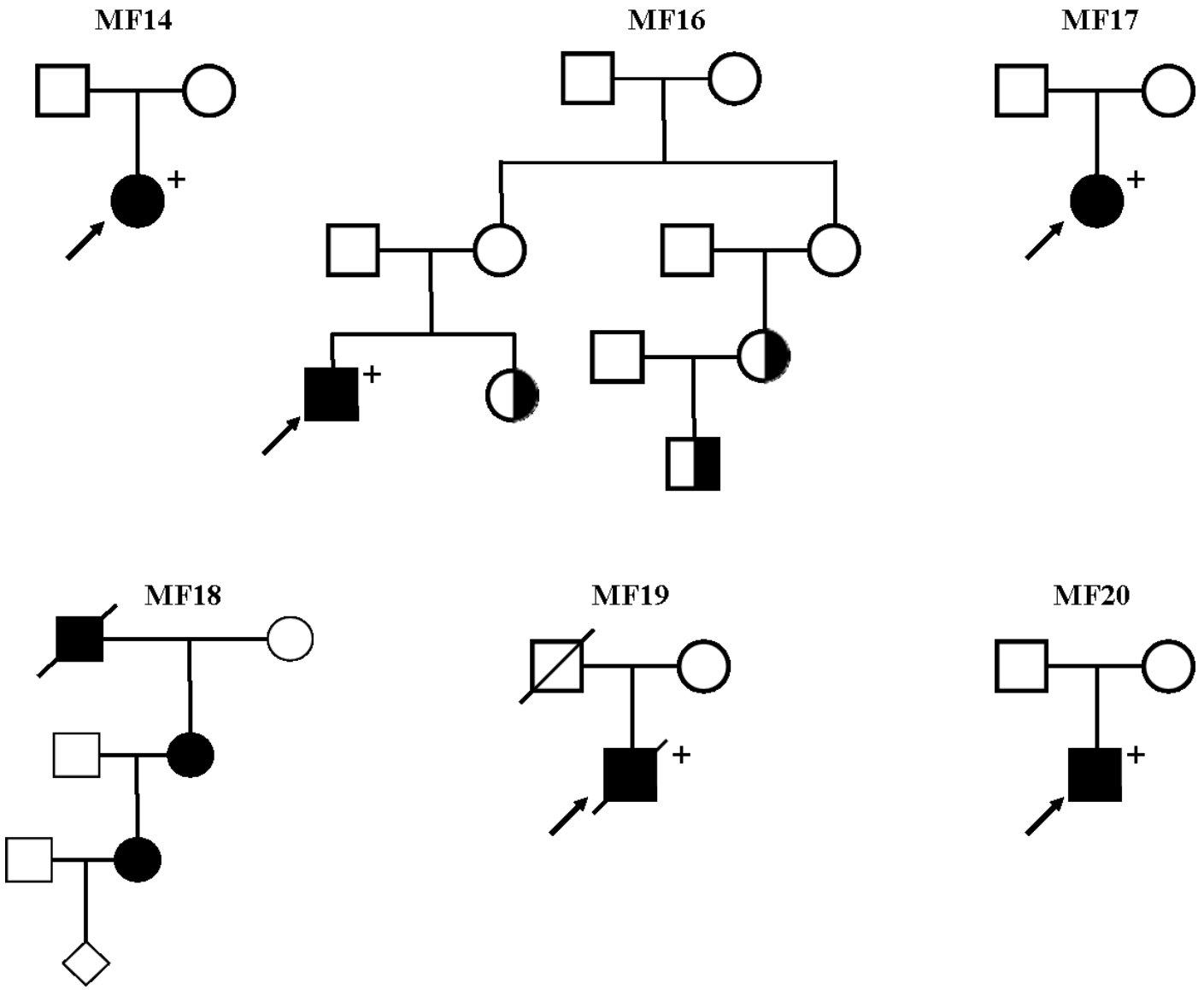
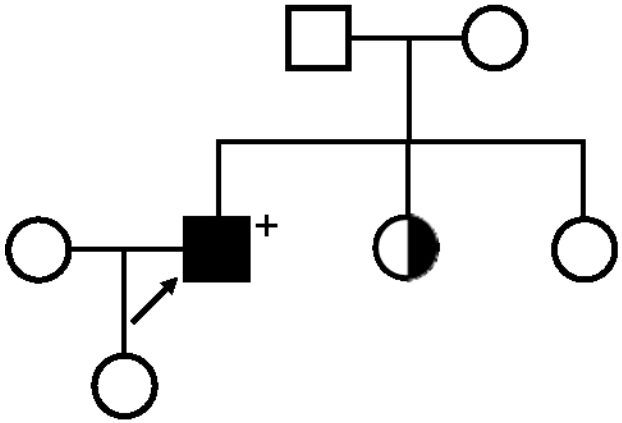
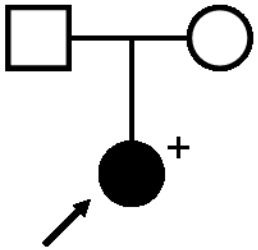

MF14
MF16
MF17
MF18
MF19
MF20

Figure 2.4. Pedigrees for Myoclonus-Dystonia patient cohort families 22-27. Circles, female; squares, male; slash, individual deceased; shaded, affected; +, blood collected from individual; arrows, proband; ?, possibly affected; half-shaded symbols represent individuals who are believed to be affected but have not been seen by a neurologist.

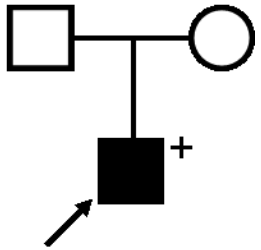
MF22



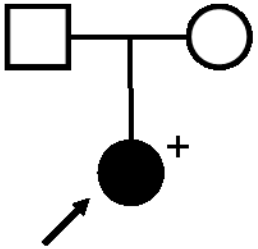
MF23



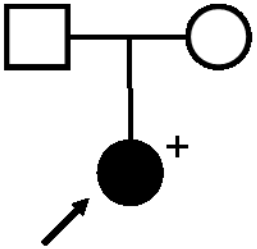
MF24



MF25



MF26



MF27

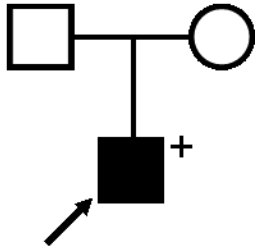
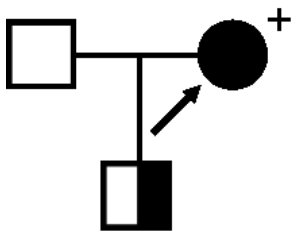
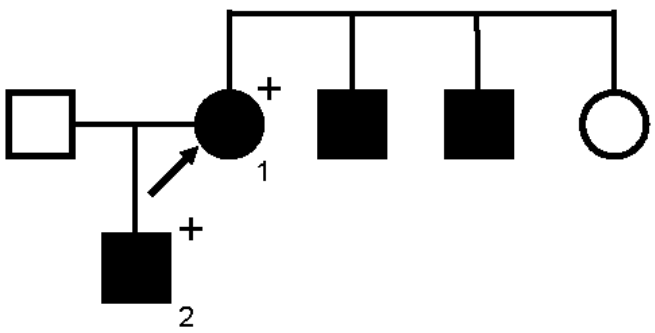


Figure 2.5. Pedigrees for Myoclonus-Dystonia patient cohort families 30-33, 35, and 36. Circles, female; squares, male; slash, individual deceased; shaded, affected; +, blood collected from individual; arrows, proband; ?, possibly affected; half-shaded symbols represent individuals who are believed to be affected but have not been seen by a neurologist.

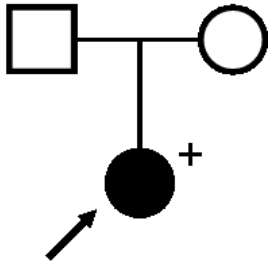
MF30



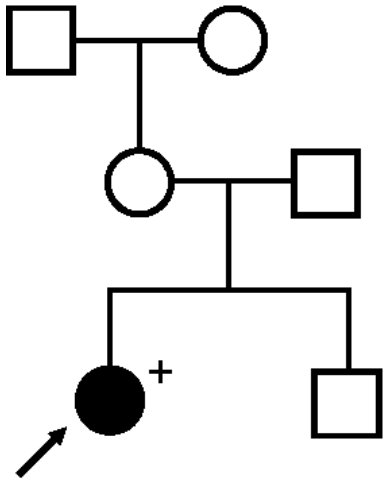
MF31



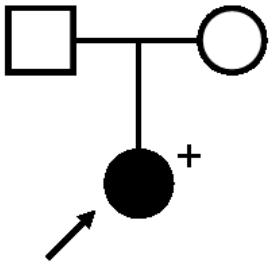
MF32



MF33



MF35



MF36

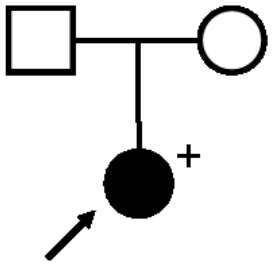
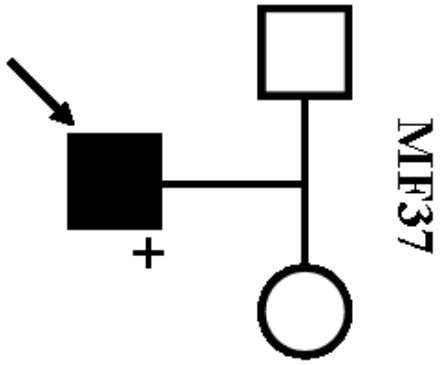


Figure 2.6. Pedigrees for Myoclonus-Dystonia patient cohort families 37 and 38. Circles, female; squares, male; slash, individual deceased; shaded, affected; +, blood collected from individual; arrows, proband; ?, possibly affected; half-shaded symbols represent individuals who are believed to be affected but have not been seen by a neurologist.


MF37
+

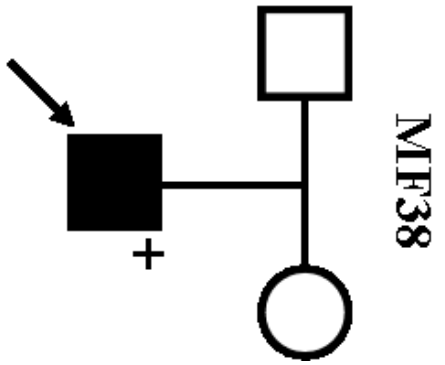

MF38
+

Table 2.1. Clinical characteristics of the probands from Myoclonus Family 1 and the 26 Myoclonus-Dystonia cohort families.

Pedigree/Family Number (MF#)	Number of affected individuals	Mean age of onset in years (range)	Location of Myoclonus	Location of Dystonia	Alcohol Responsive?	Additional Features
1	14	9.1 (3-15)	Mainly UE, face, trunk	UE, LE, axial	Yes	
5	1 (2?)	15	Face, neck, UE	None	Yes	
8	1	1	Mild trunk, UE	None	Unknown	Dystonia more prominent than myoclonus
9	1 (5?)	5	Left arm, UE, trunk, axial	Subtle left arm	Unknown	Kinetic Tremor in UE (left moreso than right)
11	1	15	Neck, shoulders, upper trunk	Neck, trunk, right UE	Unknown	Irregular postural and kinetic tremor in trunk muscles (right moreso than left), postural tremor in right LE
12	1	5	UE (left moreso than right)	Neck	Yes	Dysarthria and irregular tremor
13	1	15	LE	None	Yes	Tremors
14	1	13	Face,UE	Neck, right arm	No	Postural and kinetic tremor of UE and LE
16	1 (4?)	25	Generalized	Eyes, neck	Yes	Vocal tremor
17	1	30	UE	None	Yes	
18	3	36	Face, UE (left moreso than right)	UE	Unknown	
19	1	12	Arms, legs, face, post-cervical, axial	Unknown	Unknown	
20	1	< 1	Right hand, head	Unknown	Yes	Tremor, difficulty walking as a child
22	1 (2?)	12	Left arm, head, generalized	Arms, head	Yes	Tremor of jaw, lips, tongue
23	1	18	Axial	None	Unknown	
24	1	6	Arms, face, neck	None	No	
25	1	23	Generalized, subtle	None	No	Tremor, parasthesisas in hands and feet
26	1	1.5	Arms, legs, axial	Unknown	Unknown	Tremor in hand
27	1	1.5	Arms, legs, axial	Present on stressed gait	Unknown	
30	1 (2?)	childhood	Arms, face	Hand	No	
31	4	34 (13-55)	Right leg	Unknown	Unknown	Chorea and tremor
32	1 (2?)	6	Arms, face	UE	No	
33	1	1	Arms, legs, axial	Legs	Unknown	Postural tremor
35	1	Unknown	Unknown	Unknown	Unknown	
36	1	14	Arms	Arms	Yes	Tremor
37	1	8	Generalized, most severe in arms (R > L)	Neck, right hand	No	Rest and postural tremor
38	1	7	Hands	None	No	

Chapter 3: Targeted Sequencing and Variant Analysis of the DYT15 Locus

3.1. Introduction

Linkage analysis performed on our 5 generation Canadian Myoclonus-Dystonia family (MF1) revealed a 3.18 Mb disease-linked region on chromosome 18p11 (Grimes *et al*., 2002; Han *et al*., 2007). Given that targeted sequencing of the known REFSEQ coding exons within the critical region did not result in the identification of a causative mutation for MD in MF1, it was determined that the mutation is likely non-coding. For this reason, it was decided that a more appropriate method for mutation discovery would be to send the DNA from affected patients for next-generation sequencing in order to obtain sequence information for all bases (coding and non-coding) within the critical region.

Two affected members of MF1 (MF1-3 and MF1-5) were chosen to undergo targeted capture of the disease-linked region followed by sequencing by Roche 454. These two patients were chosen because they are meiotically distant from each other; this helps to reduce the number of common familial variants. The 454 sequencing platform was chosen based on its ability to produce much longer read lengths (average of 400 bp) than other sequencing platforms such as SOLiD and Illumina (50 bp and 75 bp, respectively) (Tucker *et al*., 2009). Sequencing of the entire critical region, including introns and intergenic sequence, allows for comprehensive screening of mutations. Given that the majority of the coding sequence of the genes within the disease-linked region has already been analyzed, we believe that the causative mutation on 18p11 may be non-coding. Once sequencing of the disease-linked regions for patient 1 (MF1-5) and patient 2 (MF1-3) were complete, the sequence reads could be assembled to a reference genome and variant analysis could be performed. Read alignment was performed using

NextGENe™ software from SOFTGENETICS.

In order for a variant to be considered as a potential disease-causing mutation, it must meet several criteria: first, the variant must be heterozygous, given that the mode of inheritance is autosomal dominant; second, the variant must not be a reported SNP (obtained from various SNP databases such as dbSNP build 131, Personal Genome Variants, and the International Hap Map project), considering that MD is a rare disorder; third, the variant must validate by Sanger sequencing to ensure that it was not a false-positive; fourth, the variant must co-segregate with the disease in the family, meaning that the variant must be present in all affected members and obligate carriers, but not in any unaffected members of the family; and finally, the variant must be absent in a sufficient number of control chromosomes (~300) to ensure that it is specific to MD-affected individuals.

3.2. Materials and Methods

3.2.1. Patients chosen for sequencing

Two meiotically distant family members, MF1-3 (number on pedigree: 39) and MF1-5 (number on pedigree: 47), were chosen to undergo targeted enrichment of the 18p critical region followed by Roche 454 sequencing. One of the two patients was also genotyped using the 500K SNP chip, which allowed for SNP confirmation and assurance against sample mix-up. The DNA used for sequencing was extracted from venous blood, according to the DNA extraction protocol outlined in section 2.2 of Chapter 2.

3.2.2. Targeted capture and sequencing methodology

Targeted capture and Roche 454 sequencing was performed by Roche Diagnostics

(Laval, QC). First, the NimbleGen Sequence Capture 385K Version 2.0 Array was used to enrich for DNA from the chr18:4,619,674-7,800,480 critical region plus 50 kb flanking. The array design covered >97% of the target sequence. Following targeted enrichment, 454 sequencing was performed. The Roche 454 sequencing method is summarized in section 2.3. of Chapter 2.

3.2.3. Sequence alignment procedure

The sequence data was analyzed with NextGENe™ software v2.10 from SOFTGENETICS. Raw sequence files were returned from Roche Diagnostics in .sff format. These files were converted to .fna and .qual formats using the NextGENe software. Following file conversion and prior to alignment, all sequence reads had to be trimmed of the adaptor sequences used during the amplification and sequencing phases of the Roche 454 protocol; the enrichment primers flank the target sequence fragments and the sequencing primer flanks the enrichment primers. The adaptor sequences were removed using the "trim by sequence" function in the NextGENe software; first, the sequencing primer was removed (5′-AAGGCACACAGGGGATAGG-3′), then the two enrichment primer sequences were removed (5′-CTCGAGAATTCTGGATCCTC-3′ and 5′-GAGGATCCAGAATTCTCGAGTT-3′). The trimmed sequence reads were then aligned to the entire human reference sequence Genome Reference Consortium: Human build 37.

Separate alignments for each patient were performed using the following parameters: matching base percentage ≥ 90.0; mutation percentage ≤ 20.00; allowable mismatched bases = 4; allowable ambiguous alignments = 1; seed = 17 bases; move step = 7 bases.

3.2.4. Variant identification and prioritization

Once each patient's alignment was complete, separate mutation reports were generated by the software. The variant reports were then filtered for known SNPs and compared for shared novel variants using the open, web-based platform Galaxy (main.g2.bx.psu.edu). All homozygous variants were then removed from the shared variant list. Novel, heterozygous variants which were shared between the two patients were identified and considered as potential mutations. Variants were assigned a priority based on their position along the chromosome; the following variant categories are listed from top priority to lowest priority: coding variants, variants within splice sites, intronic variants, structural variants, variants occurring within expressed sequence tags (ESTs), and intergenic variants.

3.2.5. Structural variant analysis

Structural variants such as large deletions, insertions, or duplications, as well as rearrangements require separate analysis because a single read is not long enough to include unique sequence on either end of, for example, a 1 kb duplication. The NextGENe software considers any read with a high number of mutations in a localized region (particularly on one side of the read) a potential structural variant (SOFTGENETICS, 2010). Given that these types of variants can only be detected through the allowance of alignment of long stretches of mismatched bases, the stringency of the alignment must be relaxed. Therefore, I performed a separate alignment in order to detect structural variants alone. Rather than performing separate alignments for person 1 and person 2, both of their reads were aligned together to the reference genome. A matching base percentage of 45 and a mutation percentage of 10 were used. The difficulty with this is that due to the extremely large amount of reads generated by next-

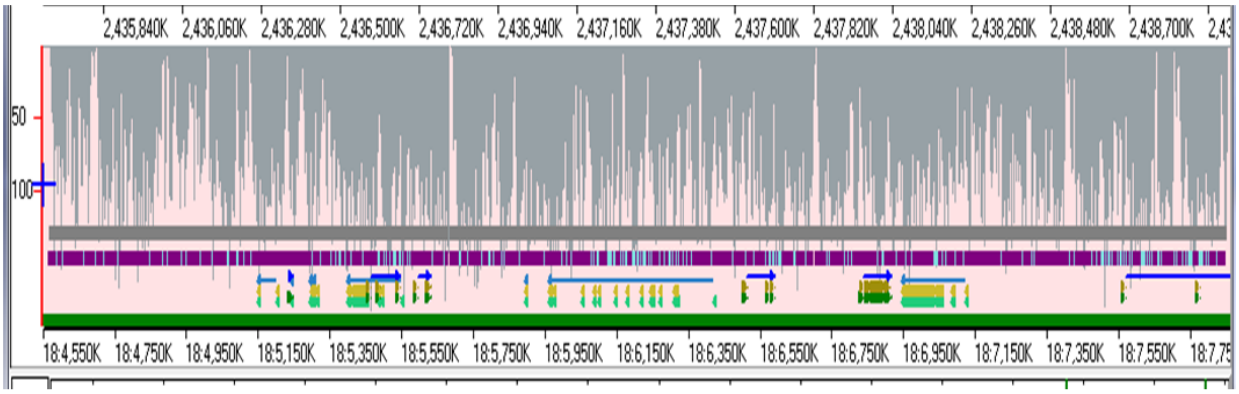
generation sequencing, as well as the potential for alignment of untrimmed reads, improper alignment of a read due to homology with another region of the genome, or difficulty with alignment of tandem repeats, the number of structural variants listed within the report is extremely high (the alignment of person 1 and 2 at 45 percent matching base resulted in a structural variant report with 4114 candidates). As an alternative to the software-generated structural variant report, and after several attempts at optimizing the alignment parameters, I developed the following method: I took the unmatched reads from the 90 percent matching base file (when a read does not meet the criteria for percent matching base, the NextGENe software places that read in an "unmatched" file) and aligned them at 45 percent matching base. By doing this, I eliminated all of the reads that align well and are of no interest to the detection of structural variants. What I obtained as a result of this was a file of aligned reads with peaks in coverage at regions with a high number of reads showing misalignment (Fig. 3.1). This method allowed me to quickly locate structural variants that were a top priority for analysis.

3.2.6. Variant validation by Sanger sequencing and control screening

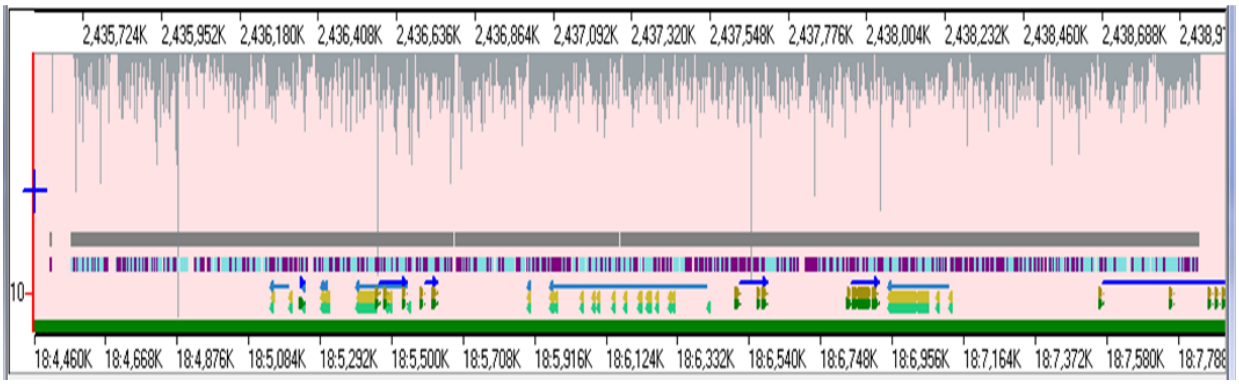
Each variant candidate was validated by Sanger sequencing according to the protocol outlined in section 2.5 of Chapter 2. Variants that were confirmed by Sanger sequencing were screened-for in control individuals; if a variant candidate was identified in a control sample, that variant was eliminated from the list of potential disease-causing candidates.

Figure 3.1. The identification of structural variants using a unique alignment method. A) Screenshot of coverage profile from reads from Person 1 and Person 2 aligned together at 45 percent matching base using NextGENe software. The deeper the grey bars at any given position, the greater the coverage. B) Screenshot of coverage profile where the unmatched reads generated from an alignment using 90 percent matching base for Person 1 and Person 2 were aligned at 45 percent matching base. The peaks in coverage correspond with positions at which a structural variant candidate occurs.

A)


2,435,840K 2,436,060K 2,436,280K 2,436,500K 2,436,720K 2,436,940K 2,437,160K 2,437,380K 2,437,600K 2,437,820K 2,438,040K 2,438,260K 2,438,480K 2,438,700K
50
100
18:4,550K 18:4,750K 18:4,950K 18:5,150K 18:5,350K 18:5,550K 18:5,750K 18:5,950K 18:6,150K 18:6,350K 18:6,550K 18:6,750K 18:6,950K 18:7,150K 18:7,350K 18:7,550K

B)


2,435,724K 2,435,952K 2,436,180K 2,436,408K 2,436,636K 2,436,864K 2,437,092K 2,437,320K 2,437,548K 2,437,776K 2,438,004K 2,438,232K 2,438,460K 2,438,688K
10
18:4,460K 18:4,668K 18:4,876K 18:5,084K 18:5,292K 18:5,500K 18:5,708K 18:5,916K 18:6,124K 18:6,332K 18:6,540K 18:6,748K 18:6,956K 18:7,164K 18:7,372K 18:7,580K

3.3. Results & Discussion

3.3.1. Sequence alignment within the critical region

Coverage of the target region was >97% and the average depth was >50 per patient. A statistical summary of the regions with low coverage is provided in Table 3.1. All but one of the regions with low coverage contain repeating elements; the capture arrays and NextGENe™ software have difficulty with these regions because of their homology to other regions of the genome. These regions are considered a low priority for disease-causing mutations and therefore their low coverage is not an immediate concern.

3.3.2. Summary of variants within the critical region

The total number of variants for persons 1 and 2, before and after filtering, are summarized in Table 3.2. A breakdown of variant type and status is summarized in Table 3.3.

Coding and Splice Site Variants

In total, 15 shared coding/splice site variants were detected. All variants but one were eliminated via Sanger sequencing either because they were false positives or because they occurred within unaffected individuals who did not share the disease haplotype. The one remaining candidate was an A to C change 9 bases before exon 10 of *EPB41L3* (c.1066-9A>C). This variant was confirmed to be real but it was found in 6 out of 88 control chromosomes and was therefore eliminated as a disease-causing candidate (Fig. 3.2). For a complete list of the coding and splice site variant candidates, refer to Appendix 1.1. These results indicate that the causative mutation for MD in our family is not within the protein-coding sequence for any *known* REFSEQ gene. This allows me to restate my hypothesis: a <u>non-coding</u> mutation exists on

Table 3.1. Statistics regarding regions of low coverage in the critical region of 18p11.

Regions of Low Coverage (<10x)	
Total Number of Regions with <10x Coverage	95
Total Number of Base Pairs with <10x Coverage (bp)	78,190
Percentage of Targeted Region on 18p11 with <10x Coverage (%)	3.52

Table 3.2. Number of variants identified in each of the two sequenced patients from Myoclonus Family 1 following the application of several filters.

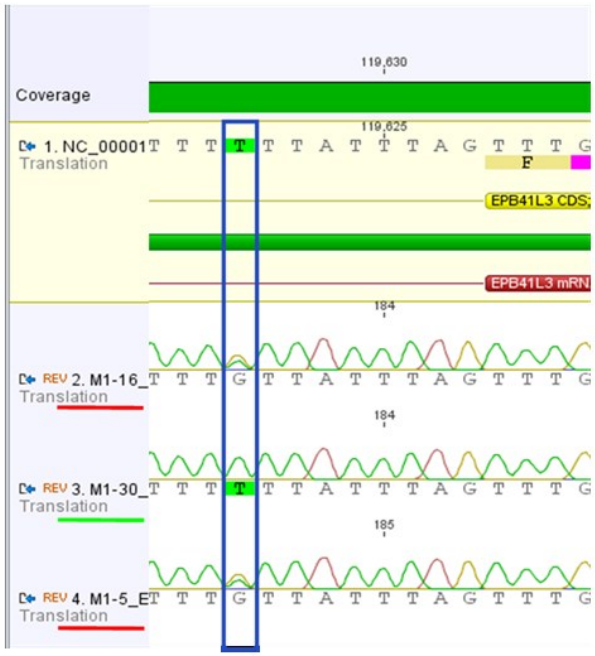
	Person 1	Person 2
Total Number of Variants	14260	16968
Total Number of Variants in Critical Region	10754	12350
Minus Known SNPs	5794	7083
Total Number of Shared Novel Variants in Targeted Region	3140	
Minus Homozygous	3100	
Consecutive Variants Merged	2292	

Table 3.3. Statistics of shared unique or unknown variants from the DYT15 critical region. Variants were categorized based on their position along the chromosome in order to determine a priority ranking for analysis. (*) the 31 remaining candidates do not affect splice site predictions. All intronic variants causing potential splice site alterations have been eliminated. (**) the remaining candidate is the ACC duplication in CD108131

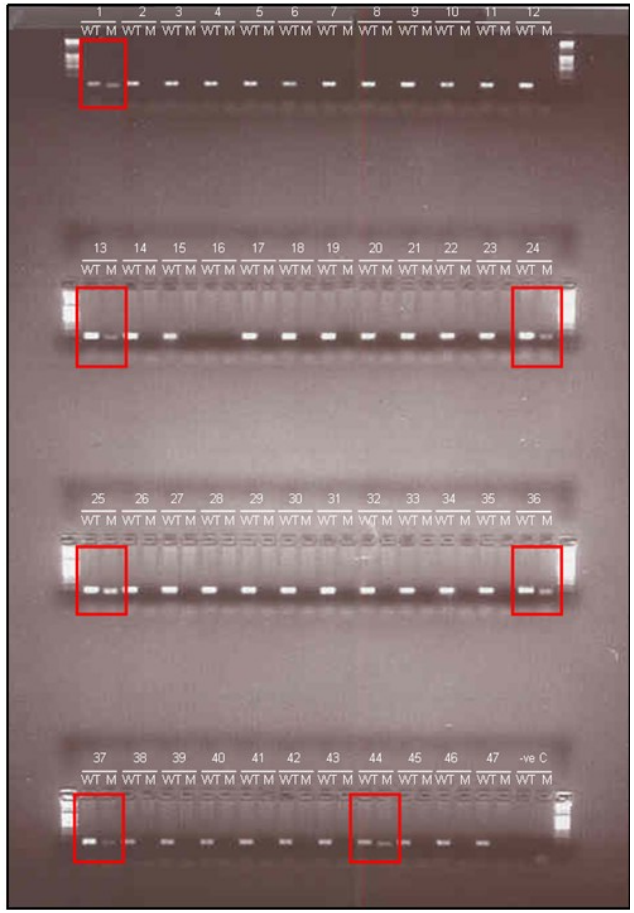
Type of Variant	Number of Variants	Number Eliminated	Number Remaining
Coding	12	12	0
Splice Sites	3	3	0
Intronic Variants	914	883	31*
Structural Variants	55	55	0
Variants Within an EST	3	2	1**
Intergenic Variants	1305	26	1279
Total	2292	981	1311

Figure 3.2. Identification of an A to C change 9 bases before exon 10 of *EPB41L3* (c.1066-9A>C). A) The c.1066-9A>C mutation was confirmed by Sanger sequencing. Sequenced patients underlined in red (M1-16 and M1-5) are affected members of MF1; the patient underlined in green (M1-30) is an unaffected member of MF1. The blue rectangle indicates the position of the A>C mutation. B) Agarose gel of allele-specific PCR products for the *EPB41L3* mutation. Every two lanes are PCR products from different individual templates (numbered 1-47). Individual 1 is patient M1-5, an affected member of MF1; individual 2 is M1-30, an unaffected member of MF1; and individuals 3-47 are controls external to MF1. The first of the two lanes for each individual contains the PCR product for the wild-type (WT) allele; the second of the two lanes for each individual contains the PCR product for the mutant (M) allele. The final two lanes on the gel contain the negative controls (-ve C) for the wild-type and mutant PCR reactions. Red boxes indicate those individuals who tested positive for a wild-type and a mutant allele. The wild-type and mutant alleles failed to amplify for control individual 16, therefore this individual was not included in the total count of control chromosomes screened.

A)


119,630
Coverage
119,625
1. NC_00001
Translation
T T T T T A T T T A G T T T G
F
EPB41L3 CDS;
EPB41L3 mRN
184
REV 2. M1-16_
Translation
T T T G T T A T T T A G T T T G
184
REV 3. M1-30_
Translation
T T T T T A T T T A G T T T G
185
REV 4. M1-5_E
Translation
T T T G T T A T T T A G T T T G

B)


1 2 3 4 5 6 7 8 9 10 11 12
WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M
13 14 15 16 17 18 19 20 21 22 23 24
WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M
25 26 27 28 29 30 31 32 33 34 35 36
WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M
37 38 39 40 41 42 43 44 45 46 47 -ve C
WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M WT M

18p11 which is responsible for causing MD in MF1 and probably other *SGCE* mutation-negative families.

Intronic Variants

Once I eliminated coding variants from my list of potential disease-causing mutations, I focused my analysis on intronic variants. A large number of the intronic variants occurred within homopolymeric regions. Roche 454 sequencing has difficulty sequencing through homopolymeric tracks and therefore a large number of false positive deletions and insertions are called within these regions. For this reason, I removed any intronic variants occurring within a homopolymeric track from my list of variants to be analyzed. I then used the online splice site predictor NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/) to determine if any of the intronic variants occur within a predicted splice site. The intronic variants that did not occur within or create novel predicted splice sites were not evaluated further given their unlikely disease-causing potential. The intronic variants that did occur within or create novel predicted splice sites were confirmed by Sanger sequencing and screened in controls; all of these candidates have been eliminated as potential disease-causing mutations (Table 3.3). For a complete list of intronic variant candidates, refer to Appendix 1.2.

Structural Variants

In all, I identified 55 structural variants requiring further analysis. I eliminated candidates based on the following criteria: if it was a documented polymorphism, if it was simply a misalignment of a tandem repeat, if it only occurred in one of the two affected patients, and if there was no consistent pattern/shared common sequence at the breakpoint. I was able to

eliminate 43 candidates based on the aforementioned criteria. I designed primers for the remaining 12 candidates and attempted to validate them by PCR and Sanger sequencing. I eliminated all 12 of the remaining candidates because they were either false positives or they were not unique to affected individuals (Table 3.3). Based on these results, I do not believe that my disease-causing mutation is a structural variant. For a complete list of structural variant candidates, refer to Appendix 1.3.

Variants within ESTs

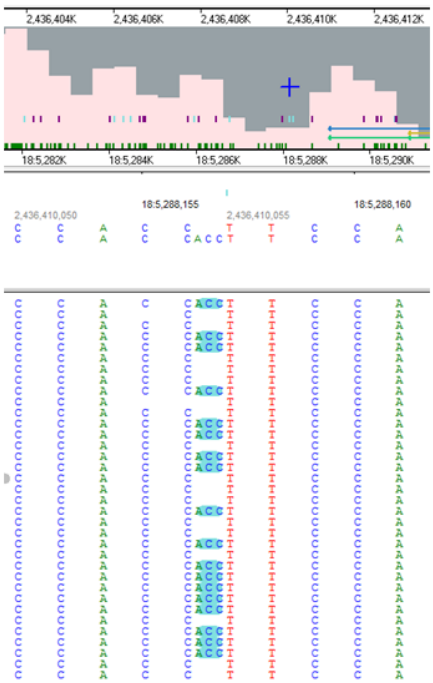
The next priority in my variant analysis were those variants that occur within expressed sequence tags (ESTs) reported in the UCSC database. ESTs are cDNA sequences that can be used to identify novel transcripts and genes. There were 3 variant candidates that occurred within reported EST sequence (Appendix 1.4.); 2 of those candidates were eliminated but 1 candidate, a 3 bp ACC duplication (Fig. 3.3a) in an unspliced human EST called CD108131 (GenBank: CD108131.1), remains (Table 3.3).

3.3.3. Validation of the ACC duplication

I have confirmed by Sanger sequencing the presence of this ACC duplication (Fig. 3.3b) and shown by allele-specific PCR that only the affected and unaffected-disease-haplotype-carrying members of Myoclonus Family 1 (MF1) have the duplication (Appendix 1.5). All of the unaffected non-carriers and married-in individuals tested negative for the duplication (Appendix 1.5). I also screened 524 control chromosomes for the duplication and all were negative. Figure 3.3c summarizes the results of the allele-specific PCR with representative affected, unaffected, and control individuals.

Figure 3.3. Identification of an ACC duplication. A) Screenshot of the ACC variant call identified using NextGENe software. B) Screenshot of the ACC variant confirmation by Sanger sequencing. Samples underlined in green are unaffected individuals; samples underlined in red are affected individuals. C) Results of allele-specific PCR for the ACC variant to test for co-segregation with the disease. M1-5 and M1-6 are affected members of MF1, M1-18 is an unaffected member of MF1, M1-33 is a married-in member of MF1, and Q66 is an external control. The 187 bp product represents the wild-type allele and the 161bp product represents the mutant allele.

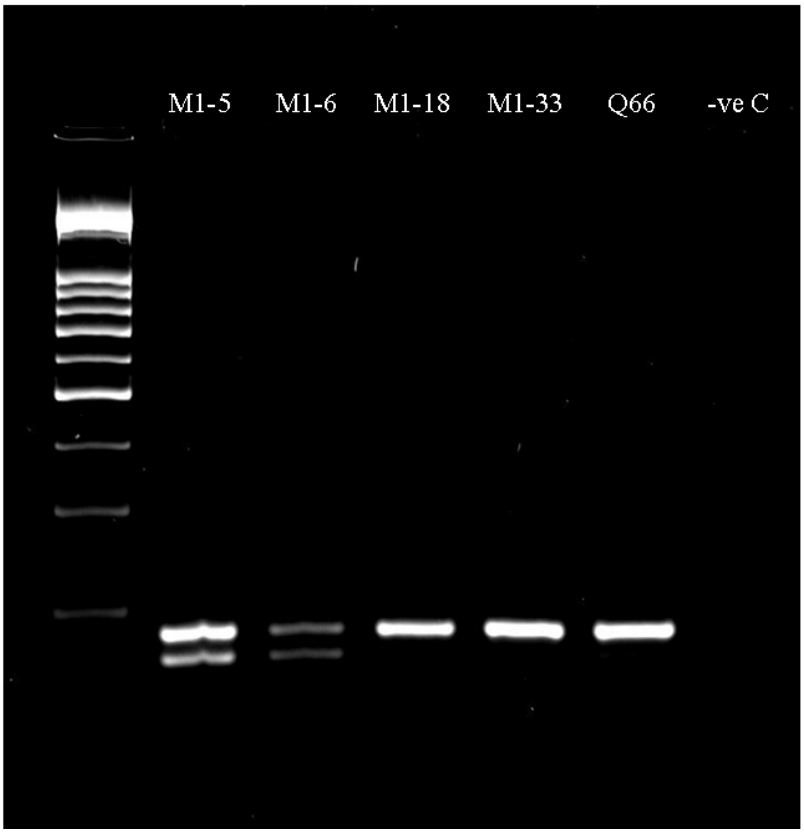
A)


2,436,404K
2,436,406K
2,436,408K
2,436,410K
2,436,412K
18:5,282K
18:5,284K
18:5,286K
18:5,288K
18:5,290K
18:5,288,155
18:5,288,160
2,436,410,050
2,436,410,055

B)


Coverage
1. hg19_dna
Translation
FWD 2. ZFP161 - 7541
Translation
REV 3. M1-7_Mut69_R_...
Translation
REV 4. M1-5_Mut69_R_...
Translation
REV 5. M1-20_Mut69_R...
Translation
FWD 6. M1-7_Mut69_F_...
Translation

C)


M1-5
M1-6
M1-18
M1-33
Q66
-ve C

3.4. Conclusions

In total, 2292 shared novel variants were identified as potential disease-causing candidates in MF1 (Table 3.2). Of these 2292 variant candidates, 974 have been eliminated (Table 3.3). All coding variants, splice site variants, intronic variants that affect splicing, structural variants and variants occurring within ESTs have been eliminated (Table 3.3). A 3 bp ACC duplication that occurs within an EST has been identified as a likely disease-causing mutation. It has been validated by Sanger sequencing and co-segregates with the disease in MF1. In addition, 524 control chromosomes have been screened for the duplication and they all tested negative. The ACC duplication is the only variant analyzed to date that meets all disease-causing mutation criteria: it is heterozygous, it is not a known SNP, it was confirmed by Sanger sequencing, it co-segregates with the disease in the family, and it was not seen in any control chromosomes tested. The remaining 1311 variant candidates in the 18p11 critical region are considered to be of lower functional relevance than the ACC duplication.

Chapter 4: Characterization of the CD108131 Non-Coding RNA

4.1. Introduction

From the 2292 variant candidates in the DYT15 locus, an ACC duplication at chromosome position 18:5288157 has been identified. According to the "Human ESTs Including Unspliced" track in the UCSC Genome Browser, the ACC duplication lies within the CD108131 transcript. This EST was identified from pituitary tissue as a part of the NIH-MGC EST Sequencing Project (http://mgc.nci.nih.gov/ESTSequences). The EST is 863 bp long (refer to Appendix 2.1 for nucleotide sequence) and has never been further characterized beyond its initial identification. The fact that the ACC duplication occurs within CD108131 makes it a more appealing disease-causing candidate, especially if this EST turns out to be a functional piece of DNA, such as a long non-coding RNA (lncRNA). The transcript does not appear to be protein-coding due to its lack of open reading frames (Appendix 2.2), therefore its role as a ncRNA with a regulatory function is more likely. Given the lack of characterization of this EST, I first tested to see if I could confirm its expression in the human brain. I then attempted to confirm its identity as a unique transcript, as well as its size.

Once I was able to confirm that CD108131 is a unique transcript, I attempted to identify its regulatory target. While lncRNAs do not encode for proteins, they play important roles in the regulation of transcription; the specificity of this regulation is determined by the ncRNA's nucleotide sequence or structural conformation (Ørom and Shiekhattar, 2011). Examples have been found for both the inhibition and activation of transcription by lncRNAs (Ørom and Shiekhattar, 2011). The X-inactive specific transcript (*Xist*) is a lncRNA that silences gene expression in *cis* on the X chromosome, resulting in X-inactivation (Chow and Heard, 2009).

Other inhibitory ncRNAs, such as *Air*, silence genes through imprinting (Sleutels *et al*., 2002). Some ncRNAs, such as *HOTAIR*, can even silence genes in *trans*, demonstrating that the regulatory targets of ncRNAs are not limited to the same chromosome (Ørom and Shiekhattar, 2011). In contrast to these examples involved in inhibition, a proximally located lncRNA has been shown to be required for the transcriptional activation of the Snail homolog 1 (*SNAI1*) gene (Ørom and Shiekhattar, 2011).

I expect that the mutation in CD108131 affects its ability to bind its regulatory target, not its own level of expression. To test this, I performed a Quantitative Real-Time PCR (qRT-PCR) experiment to compare the level of CD108131 expression in an MD patient from MF1 to that of a control individual. Non-coding RNAs often regulate genes in *cis*, meaning that their targets often lay up- or downstream of them on the chromosome (Ørom and Shiekhattar, 2011). Therefore, I wanted to determine if *ZFP161*, a gene positioned approximately 750 bp upstream of CD108131, could be the regulatory target of the CD108131 ncRNA. To test this, I performed a second qRT-PCR experiment, this time to compare the level of *ZFP161* expression in an MD patient from MF1 compared to a control individual. Finally, *SGCE* expression was compared by qRT-PCR between an MD patient and control individuals. *SGCE* was chosen as a potential regulatory target of CD108131 due to its previously characterized role in MD pathology.

4.2. Materials and Methods

4.2.1. Patient and control cell lines

All lymphoblastoid cell lines are Epstein-Barr Virus (EBV) transformed and were obtained from the Department of Genetics Cell Culture Service at The Centre for Applied Genomics (TCAG), operated by The Hospital for Sick Children in Toronto. The patient cell line

for MF1 belongs to the index patient of the family, M1-16, an affected male (number on pedigree: 16). Control cell lines 1-8 were obtained from individuals outside of MF1 that are not affected with MD.

4.2.2. Cell line maintenance

Lymphoblast cells were obtained from liquid nitrogen storage and thawed at 37°C. Thawed cultures were then transferred to separate 15 mL conical tubes and centrifuged at 1500 rpm for 2 minutes. Freezing media was aspirated and discarded. The cell pellet was resuspended in 2 mL of fresh RPMI 1640 media, supplemented with 10% FBS and 1% penicillin/streptomycin. 1 mL of the resuspended cells was then added to a 50 mL culture flask containing 4 mL of fresh RPMI 1640 media (10% FBS and 1% penicillin/streptomycin). Cell cultures were then incubated at 37°C and 5% CO_2. 1 volume of fresh RPMI 1640 media (10% FBS and 1% penicillin/streptomycin) was added every 2-5 days, once the pH of culture medium dropped (turning the culture medium from pink to orange/yellow). Once the culture volume reached 20 mL, the culture was pelleted, resuspended in fresh media, and split into two culture flasks.

4.2.3. RNA extraction from lymphoblast cell lines and cDNA synthesis

RNA was extracted and cDNA was synthesized from the patient and control lymphoblast cell lines according to the protocol in section 2.6 of Chapter 2.

4.2.4. Additional RNA samples from human tissues

MVP™ Total RNA, isolated from human brain, was obtained from Agilent Technologies

(Catalog Number 540005). A separate Human Total RNA Master Panel was obtained from Clontech (Catalog Number 636643). The Master Panel contains total RNA samples isolated from the following tissues: brain (cerebellum), brain (whole), fetal brain, fetal liver, heart, kidney, adrenal gland, lung, placenta, prostate, salivary gland, skeletal muscle, spleen, testis, thymus, thyroid gland, trachea, uterus, stomach, and small intestine. First-strand cDNA synthesis was performed on all RNA samples according to the protocol outlined in section 2.6 of Chapter 2.

4.2.5. Transcript analysis

In order to test for the expression of CD108131 in human brain, primers were designed to amplify 824 bp of the CD108131 transcript:

Forward Primer: 5′-GTCTGAATTTGCAGTTACTGAGAGGC-3′

Reverse Primer: 5′-GTGAGGTTGACACTAACAAATTCACCC-3′

PCR was then performed on human brain cDNA (made from the human brain total RNA obtained from Agilent) in conjunction with a No Reverse Transcriptase (No RT) Control (RNA template) and a Negative Control (H_2O). A separate PCR using *GAPDH* primers served as a positive control for cDNA synthesis. The PCR was run for 35 cycles with an annealing temperature of 58°C.

To determine whether CD108131 is a unique transcript, separate from the upstream *ZFP161* gene and the downstream AK098498 mRNA, primers were designed to amplify from the cDNA of each of the 3 transcripts separately, as well as from any transcript that may be a hybrid of any of the 3 transcripts. The primer sequences were as follows:

AK098498cDNA_FWD: 5′-TACAGGCTGCCCCCACGTGAT-3′

AK098498cDNA_REV: 5′-AAGGCGGCAAGGAGAGGAGAGT-3′

CD108131_FWD: 5′-GTCTGAATTTGCAGTTACTGAGAGGC-3′

CD108131_REV: 5′-GTGAGGTTGACACTAACAAATTCACCC-3′

ZFP161cDNA2_FWD: 5′-TGCTGACACCCAGGATGATGATGT-3′

ZFP161cDNA2_REV: 5′-TTAGGTGTTCTTTCAGGTGGGCCT-3′

AK098498_FWD6: 5′-TTTGCTTGGCTCGCCACATAATGG-3′

CD108131_REV2: 5′-ACGGAAAGCTCTCACTCCCTTTGA-3′

CD108131_FWD3: 5′-TAGGACTCAACCCTGCACCACAAT-3′

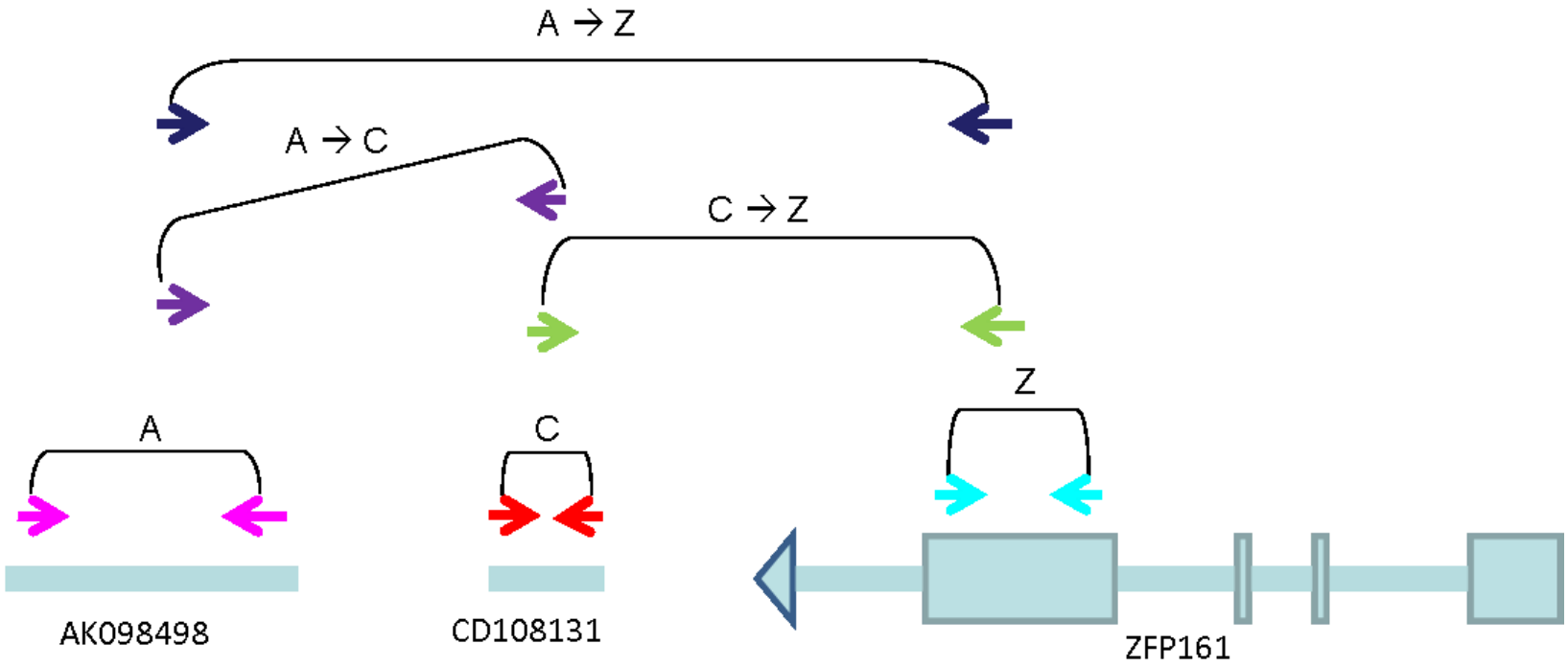
ZFP161_REV2: 5′-GCCACATGTGTGACAAAGCCTTCA-3′

AK098498_FWD5: 5′-TTTGCTTGGCTCGCCACATAATGG-3′

ZFP161_REV1: 5′-GCCACATGTGTGACAAAGCCTTCA-3′

Figure 4.1 summarizes the experimental design for PCR amplification of the 3 transcripts of interest. PCR for all of the primer sets listed above included 35 cycles with an annealing temperature of 58°C. cDNA from patient M1-16 and human brain total RNA were used as templates.

Figure 4.1. Primer design for transcript analysis of CD108131. Schematic of the position of the primers used to amplify the AK098498, CD108131, and *ZFP161* cDNAs, as well as any hybrid transcripts of the three. A: primers AK098498cDNA_FWD and AK098498cDNA_REV; C: primers CD108131_FWD and CD108131_REV; Z: primers ZFP161cDNA2_FWD and ZFP161cDNA2_REV; A→C: primers AK098498_FWD6 and CD108131_REV2; C→Z: primers CD108131_FWD3 and ZFP161_REV2; A→Z: primers AK098498_FWD5 and ZFP161_REV1.


A → Z
A → C
C → Z
A
C
Z
AK098498
CD108131
ZFP161

4.2.6. Northern Blot

The Northern Blot experiment was conducted as outlined in section 2.7 of Chapter 2. The samples used for the Northern Blot experiment include human brain total RNA (obtained from Agilent), patient M1-16 RNA, and Control 1 RNA (extracted from a lymphoblast cell line according to section 2.6 of Chapter 2).

4.2.7. Quantitative Real-Time PCR

Quantitative Real-Time PCR (qRT-PCR) was performed according to the protocol outlined in section 2.8 of Chapter 2. The CD108131 and *ZFP161* qRT-PCRs were performed in triplicate on patient M1-16 and control 1 cDNA. The *SGCE* qRT-PCR was performed in 3 to 7 replicates on patient M1-16 and controls 1-8. All qRT-PCR experiments were run in conjunction with *ACTB* and all CD108131, *ZFP161*, and *SGCE* values were normalized to *ACTB* expression.

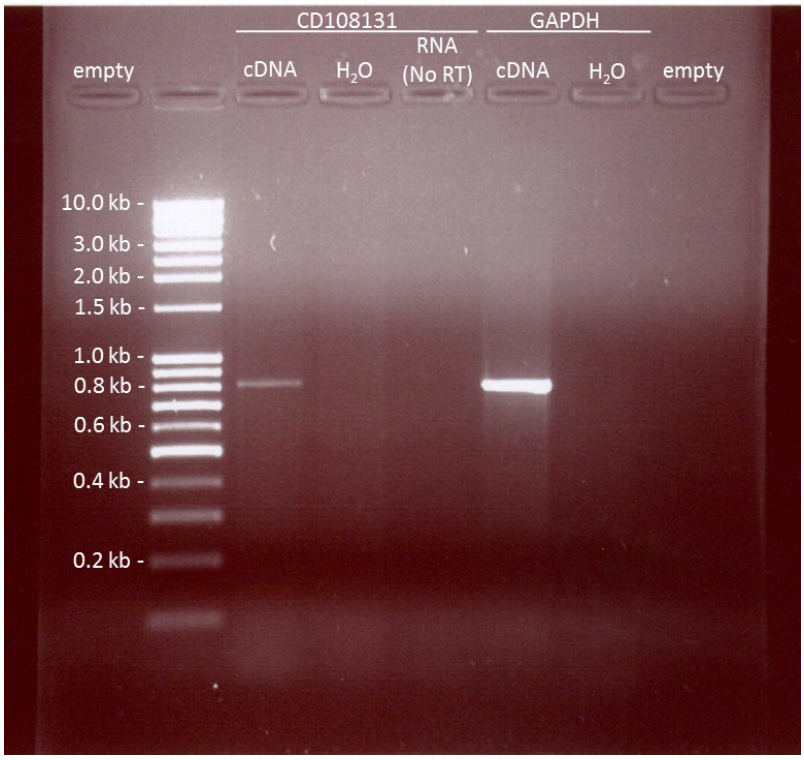
4.3. Results

4.3.1. Characterizing the CD108131 transcript

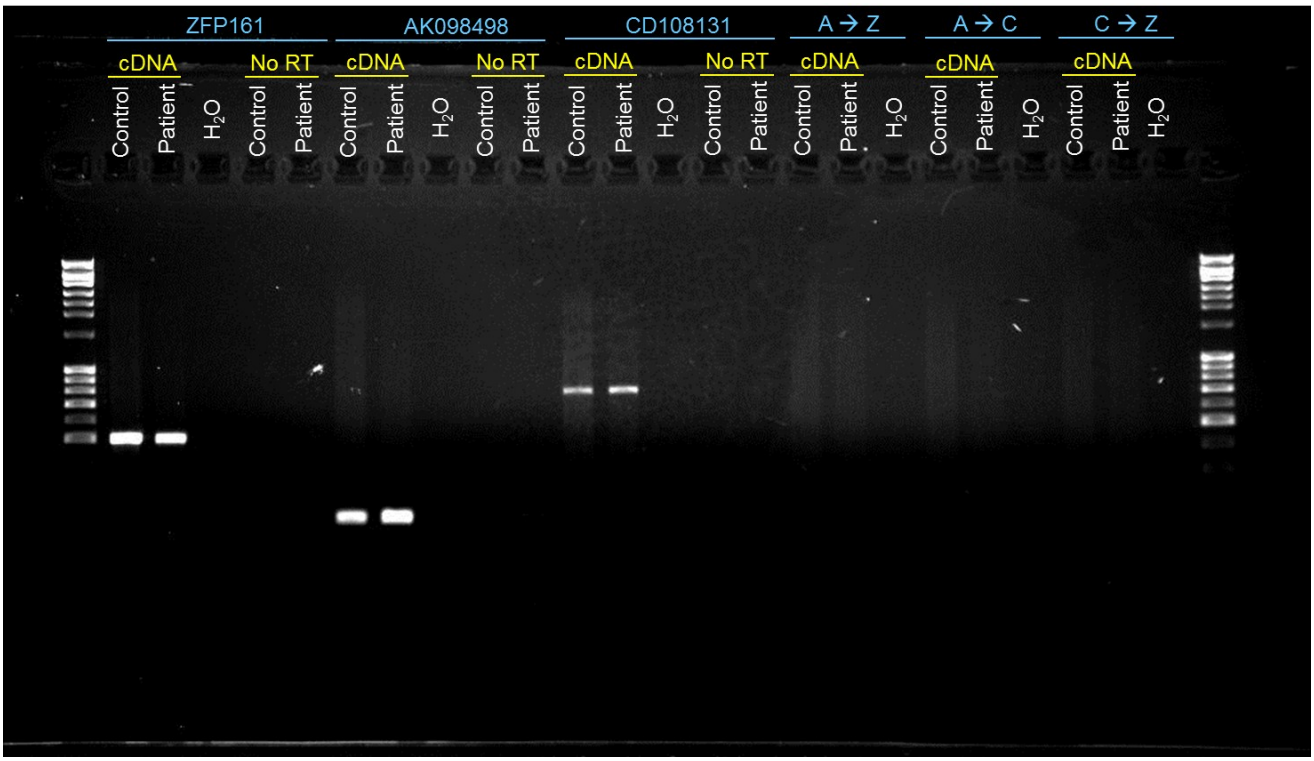
PCR of human brain cDNA using primers specific to CD108131 revealed that it is expressed in human brain (Fig. 4.2A). PCR using transcript-specific primer sets (as described in Fig. 4.1) confirmed that the AK098498 mRNA, CD108131, and *ZFP161* are all expressed uniquely from each other in both brain and lymphoblast cells (Fig. 4.2B). AK098498, CD108131, and *ZFP161* were all successfully amplified from patient and control cDNA, however a single transcript including AK098498 and CD108131, CD108131 and *ZFP161*, or AK098498 and *ZFP161* failed to amplify (Fig. 4.2B).

Figure 4.2. Amplification of the CD108131 transcript from human brain total RNA and confirmation that CD108131 is a unique transcript. A) Reverse Transcriptase PCR was performed to generate cDNA from human brain total RNA. The cDNA was then used as template for a standard PCR in which primers specific to CD108131 and *GAPDH* were used. Lane 2 contains the molecular size standard (HighRanger Plus – 100 bp DNA Ladder from Norgen Biotek Corporation). H_2O was used as a negative control and RNA was used as a control for genomic contamination. No RNA control was used for *GAPDH* because the size of the cDNA product is different from that of the genomic product. B) Results of the PCR experiment described in Fig.4.1. The blue headings indicate the target transcript being amplified. The control sample was derived from human brain total RNA obtained from Agilent Technologies; the patient sample was derived from human lymphoblast RNA obtained from patient M1-16. The H_2O sample serves as a negative control for primer contamination and the No RT controls serve as a negative control for genomic contamination of the RNA. A→C indicates an attempt to amplify a hybrid transcript of AK098498 and CD108131; C→Z indicates an attempt to amplify a hybrid transcript of CD108131 and ZFP161; and A→Z indicates an attempt to amplify a hybrid transcript of AK098498 and ZFP161.

A)


CD108131
GAPDH
empty
cDNA
H2O
RNA (No RT)
cDNA
H2O
empty
10.0 kb -
3.0 kb -
2.0 kb -
1.5 kb -
1.0 kb -
0.8 kb -
0.6 kb -
0.4 kb -
0.2 kb -

B)


ZFP161
AK098498
CD108131
A → Z
A → C
C → Z
cDNA
No RT
Control
Patient
H2O

To further confirm that the CD108131 transcript is 863 bp in size, as suggested by the NIH-MGC EST Sequencing Project, I performed a Northern Blot experiment on total brain RNA as well as RNA from an MD patient and a control individual, isolated from lymphoblast cell lines. Probing for the CD108131 transcript using a ^{32}P-labeled PCR product probe revealed that the size of the CD108131 transcript is as expected, approximately 860 bp (Fig.4.3).

4.3.2. CD108131 is expressed in multiple tissue types

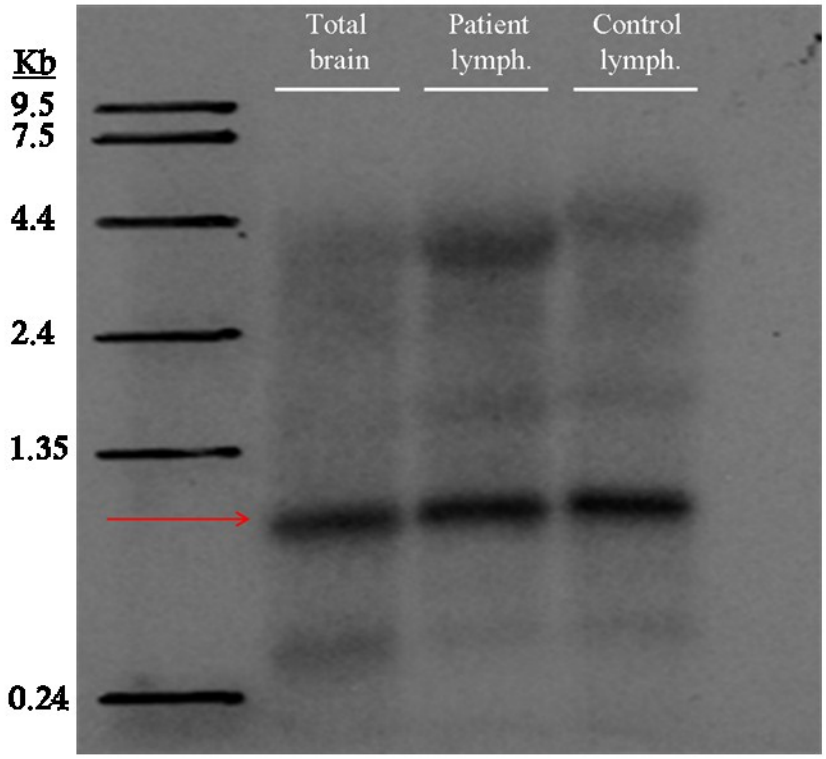
In order to determine where CD108131 is expressed, a total RNA master panel containing samples from 20 different tissues was obtained. Following cDNA synthesis, amplification of the CD108131 transcript was attempted on each of the samples by PCR. Results show that CD108131 is expressed in the cerebellum, whole brain, kidney, adrenal gland, lung, placenta, prostate, salivary gland, skeletal muscle, spleen, testis, thymus, thyroid gland, trachea, uterus, stomach, and small intestine (Fig. 4.4). No CD108131 was detected in the fetal brain sample, and very little was amplified from the fetal liver and heart samples (Fig. 4.4). Expression appears to be greatest in the cerebellum, followed by the thymus (Fig. 4.4).

4.3.3. Expression analysis of CD108131 and its potential regulatory targets

The qRT-PCR experiments revealed that there is no difference in the expression of CD108131 between the MD patient and the control (Fig. 4.5A) and that there is no difference in the expression of *ZFP161* between the MD patient and the control (Fig. 4.5B). There is, however, a significant decrease in *SGCE* expression in the MD patient when compared to 8 different control individuals (Fig. 4.5C). *SGCE* expression was variable between individuals, however the amount of *SGCE* in the affected patient, M1-16 was consistently lowest (Fig. 4.5C).

Figure 4.3. Northern Blot to determine the size of the CD108131 transcript. Brain total RNA was obtained from Agilent Technologies and the patient and control RNA was isolated from blood-derived lymphoblast cell lines. 3 µg of RNA was loaded onto the gel for each of the samples. A) The red arrow indicates the transcript of interest, CD108131, at approximately 860 bp (size markers are indicated to the left of the blot). B) This image represents the same membrane used to determine the size of the CD108131 transcript, however the membrane has been stripped and re-hybridized with an *ACTB*-specific probe. This control serves to demonstrate the presence and quality of RNA for each of the samples.

A)


Kb
9.5
7.5
4.4
2.4
1.35
0.24
Total brain
Patient lymph.
Control lymph.

B)

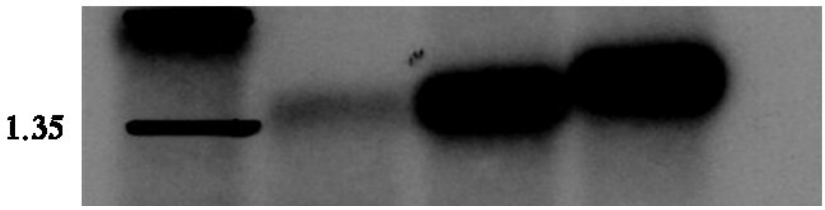
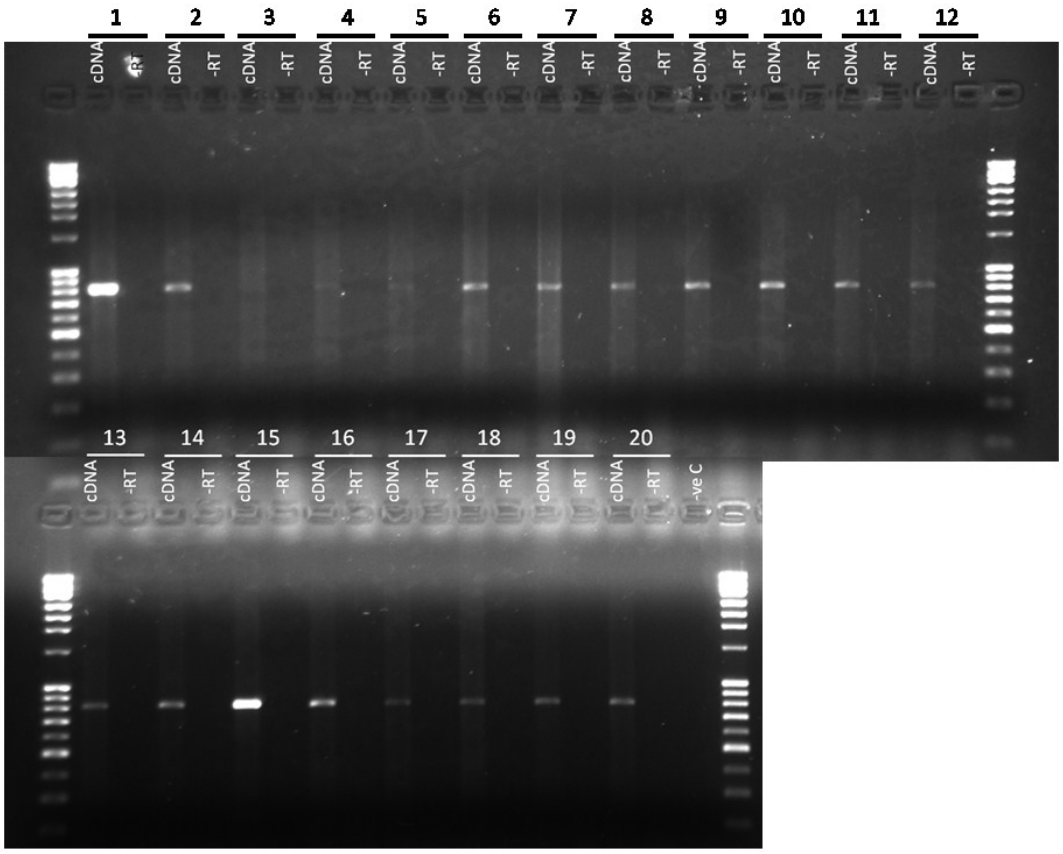

1.35

Figure 4.4. CD108131 expression analysis in 20 human tissues. Gel photos of the PCR products from (A) CD108131 and (B) *GAPDH* amplification of 20 different cDNA templates. The -RT controls are the total RNA samples prior to cDNA synthesis to ensure a lack of genomic DNA contamination. *GAPDH* serves as a positive control for successful cDNA synthesis. Total RNA was extracted from the following 20 tissues: 1 = cerebellum; 2 = whole brain; 3 = fetal brain; 4 = fetal liver; 5 = heart; 6 = kidney; 7 = adrenal gland; 8 = lung; 9 = placenta; 10 = prostate; 11 = salivary gland; 12 = skeletal muscle; 13 = spleen; 14 = testis; 15 = thymus; 16 = thyroid gland; 17 = trachea; 18 = uterus; 19 = stomach; 20 = small intestine.

A)


1 2 3 4 5 6 7 8 9 10 11 12
cDNA -RT
13 14 15 16 17 18 19 20
cDNA -RT
-ve C

B)

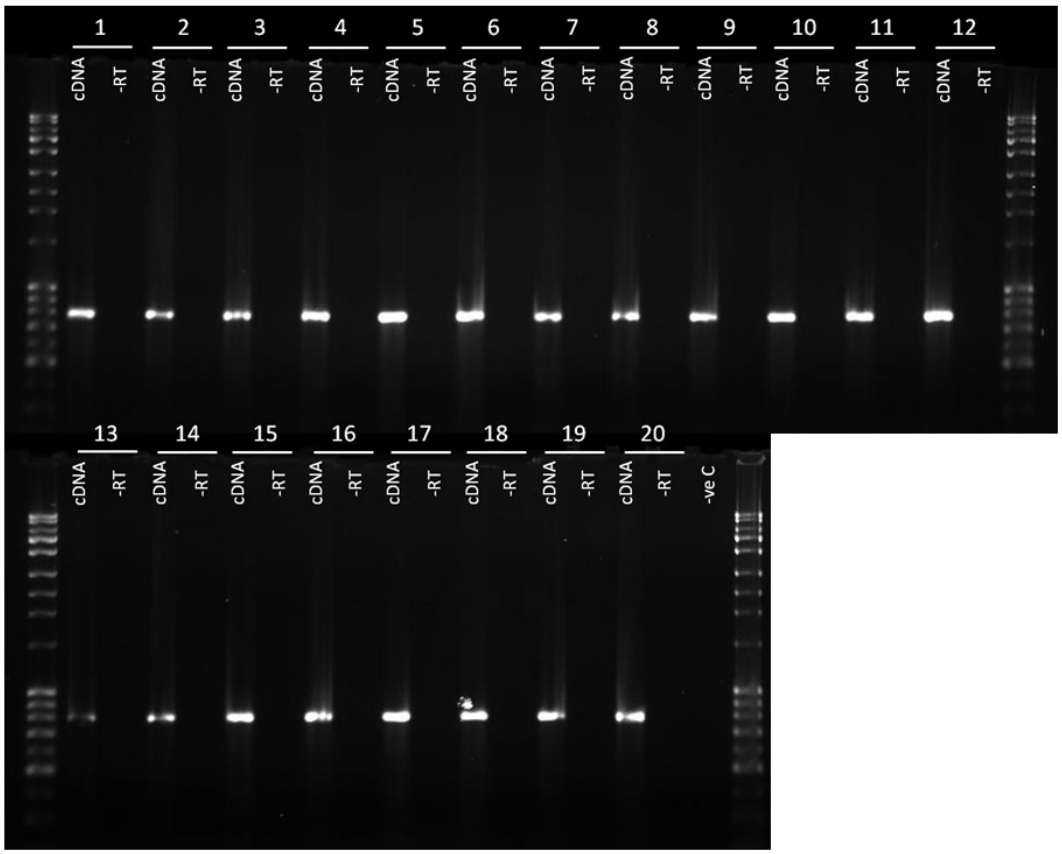
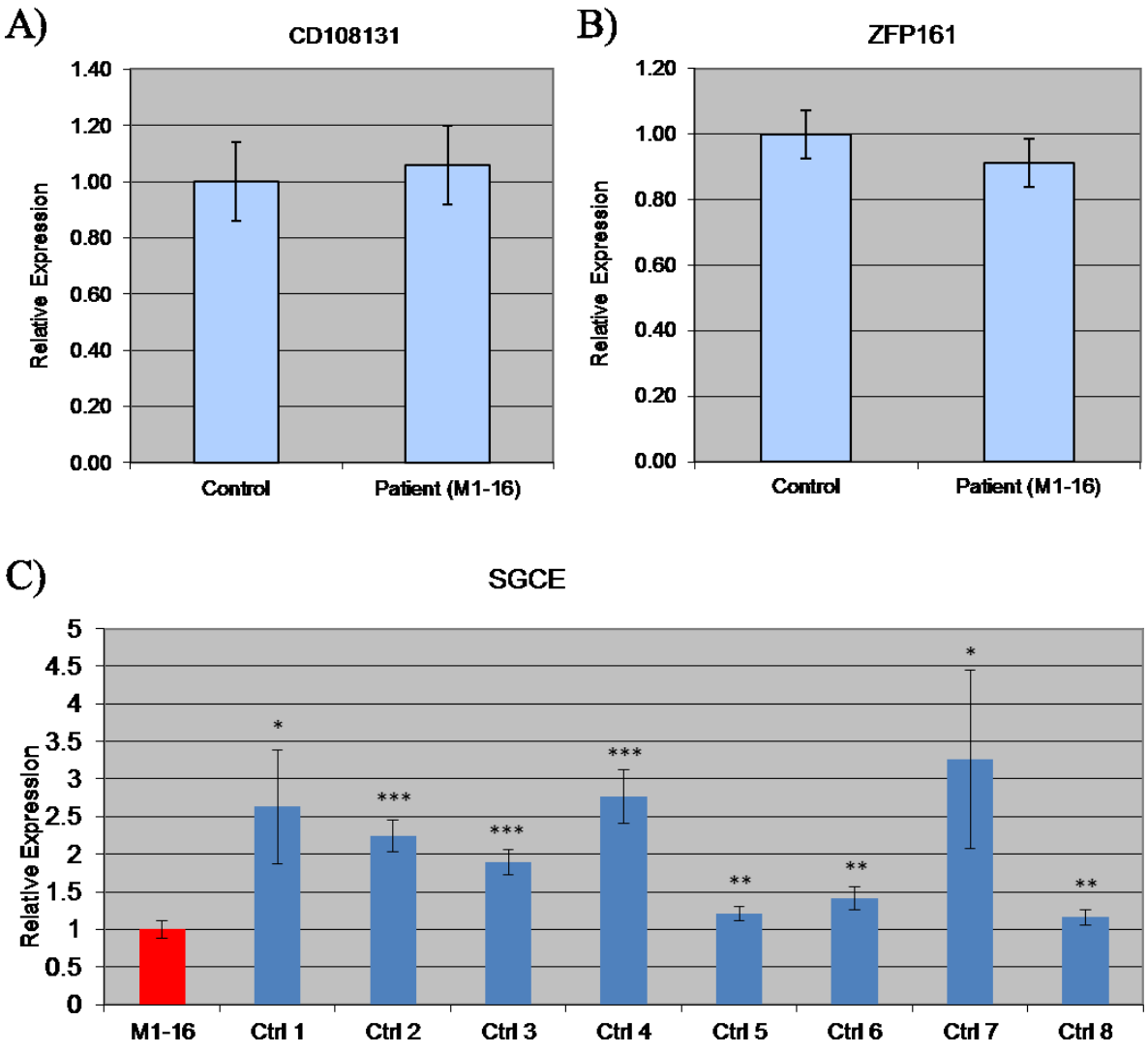

1 2 3 4 5 6 7 8 9 10 11 12
cDNA -RT
13 14 15 16 17 18 19 20
cDNA -RT
-ve C

Figure 4.5. Comparing the relative expression of CD108131, *ZFP161*, and *SGCE* in patient and control cell lines. Results from qRT-PCR of patient M1-16 and control lymphoblast cDNA to quantify the expression of A) the CD108131 transcript; B) *ZFP161*; and C) *SGCE*. For the CD108131 and *ZFP161* analysis, the control and patient values represent an average of three replicates; the patient value is expressed as abundance relative to the control. Expression levels were normalized to that of β-actin. The difference in CD108131 and *ZFP161* expression between the patient and control samples is not significant. For the *SGCE* analysis, all samples in blue are controls external to MF1; the patient M1-16 in red is an affected member of MF1. The Ctrl 1, Ctrl 2, Ctrl 5, Ctrl 7, and Ctrl 8 values represent an average of three replicates; the Ctrl 3 and Ctrl 4 values represent and average of 4 replicates, and the M1-16 and Ctrl 6. values represent an average of 7 replicates. The control patient values are represented as a relative abundance to the MD patient, M1-16. Expression levels were normalized to a β-actin endogenous control. The significance of the difference in expression between each control and M1-16 is represented by asterix (*): p-value ≤ 0.05 (*); p-value ≤ 0.03 (**); p-value ≤ 0.003 (***).


A)
CD108131
Relative Expression
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Control
Patient (M1-16)
B)
ZFP161
Relative Expression
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Control
Patient (M1-16)
C)
SGCE
Relative Expression
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
*
***
***
***
**
**
*
**
M1-16
Ctrl 1
Ctrl 2
Ctrl 3
Ctrl 4
Ctrl 5
Ctrl 6
Ctrl 7
Ctrl 8

4.4. Discussion

The results of the transcript characterization experiments confirm that CD108131 is a novel, distinctive transcript rather than a component of the upstream *ZFP161* gene or the downstream AK098498 mRNA. Considering the evidence that CD108131 is a unique, transcribed piece of RNA with expression in multiple tissue types, and given that it has no open reading frames for translation into protein, CD108131 is likely a long non-coding RNA (lncRNA) involved in the regulation of expression of another gene(s). The Northern Blot experiment revealed that CD108131 is ~860 bp in size, which is consistent with the 1 kb average length of most known lncRNAs (Ørom and Shiekhattar, 2011). This lncRNA is also very poorly conserved among species outside of primates (Table. 4.1), which is consistent with the finding that the majority of long non-coding RNAs are poorly conserved, exhibiting sequence divergence across phyla (Pang *et al*., 2006). For decades researchers have believed that protein-coding genes are the most important elements of genomic DNA, and that the non-coding regions of the genome are simply "junk." This theory is changing quickly, however, now that we are gaining insight into the complex functional contributions of non-coding RNAs. While many organisms share a similar number of protein-coding genes (~21,561 in humans and 21,839 in mice), the non-protein-coding contribution of the genome appears to increase with an organism's complexity (Prasanth and Spector, 2007). The lack of conservation of lncRNAs (including CD108131) does not mean that these elements are not functionally relevant, rather it implies that these elements could in fact be the key to higher eukaryotic organism complexity (Pang *et al*., 2006; Prasanth and Spector, 2007).

Myoclonus-Dystonia has an exclusively neurological phenotype and a brain-specific function for *SGCE* is believed to explain why mutations in this gene lead to the disease (Ritz *et*

Table 4.1. Sequence conservation of the CD108131 lncRNA across 11 species. Results summarized from a nucleotide BLAST of 840 bp of CD108131 sequence obtained from the UCSC Genome Browser against 11 different species.

Species	% Identity to CD108131	% Coverage of CD108131	Chromosome
Chimp	99.3	100	18
Gorilla	98.7	100	18
Rhesus	92.4	97.7	18
Elephant	82.4	37.1	4
Rabbit	76.7	34.4	9
Sheep	83.7	18.8	23
Cow	80.7	18.8	24
Pig	96.6	7.7	10
Mouse	92.6	3.6	3
Rat	76	3.1	3
Chicken	80.8	3.1	4

al., 2011). While, similar to *SGCE*, the CD108131 ncRNA appears to be expressed ubiquitously in human tissues, there is likely a brain-specific function related to MD pathology. CD108131 expression appears to be highest in the cerebellum (Fig. 4.4). This result is also consistent with that of Ritz *et al*. (2011), who determined that the *SGCE* brain-specific isoform's expression is greatest in the cerebellum. These results complement the increasing evidence that the cerebellum plays a role in Myoclonus-Dystonia (Hubsch *et al*., 2011). Cerebellar abnormalities in MD patients have been demonstrated by imaging studies, neurophysiological experiments,

and sensorimotor adaptation studies (Beukers *et al*., 2010; Teo *et al.,* 2009; Hubsch *et al*., 2011). A greater understanding of the role of *SGCE* and CD108131 in the cerebellum could be the key to understanding MD pathology.

Considering that CD108131 is believed to be a lncRNA, I expected that the mutation in CD108131 would affect its ability to bind its regulatory target, not its own level of expression. The results of the qRT-PCR experiment confirm this; there was no difference in the level of expression of CD108131 in the affected patient compared to the control (Fig. 4.5A). Given that some ncRNAs have been found to regulate the genes most proximal to them on the chromosome, I tested to see if *ZFP161* could be a potential regulatory target for the CD108131 ncRNA. Results of the qRT-PCR experiment revealed that there is no difference in the level of expression of *ZFP161* in the affected patient compared to the control (Fig. 4.5B); *ZFP161* is therefore not a likely regulatory target of CD108131 in relation to Myoclonus-Dystonia. *SGCE*, however, was expressed at a significantly lower level in the MD patient compared to 8 different control individuals (Fig. 4.5C). There was a great deal of variability in *SGCE* expression from one individual to the next, which is consistent with the findings of Ritz *et al*. (2011) that *SGCE* expression levels vary among individuals. Despite this variability, however, the MD patient consistently had lower levels of *SGCE* expression when compared to controls (Fig. 4.5C). This result could suggest that CD108131 plays a role in positively regulating *SGCE* and that the mutation occurring in MF1 disrupts its ability to do so, resulting in decreased expression of *SGCE* and subsequent Myoclonus-Dystonia.

Chapter 5: Screening for *TITF-1* Mutations in the Myoclonus-Dystonia Patient Cohort

5.1. Introduction

In addition to the mutation screening that I performed on MF1, I screened our 26 other *SGCE* mutation-negative MD patients for mutations in the *TITF-1* (also known as *NKX2-1*) gene. Mutations in the *TITF-1* gene have been associated with benign hereditary chorea (BHC), a non-progressive chorea with onset in infancy or childhood (Asmus *et al*., 2007). Due to their phenotypic overlap, Myoclonus-Dystonia and BHC can be confused and an improper diagnosis can be made (Asmus *et al*., 2007). To ensure that none of my *SGCE* mutation-negative MD patients have a mutation in the *TITF-1* gene (and therefore that they are useful candidates for validating my MF1 mutation), I sequenced all four coding exons of the gene in each of the affected individuals.

5.2. Materials and Methods

5.2.1. Primer design, PCR, and sequencing the *TITF-1* gene

Primers were designed for each of the 4 exons of the *TITF-1* gene using the reference sequence obtained from the UCSC Genome Browser. Primer sequences are as follows:

TITF-Ex1-F: 5′-CTCGGATTCTCTCCGGTAGG-3′

TITF-Ex1-R: 5′-GAAGGAAGGTGAATGCTGCT-3′

TITF-Ex2-F: 5′-GAAATGCTTTGGGTCTCGTC-3′

TITF-Ex2-R: 5′-GGAGGAAGGAAGAGGAGGAA-3′

TITF-Ex3-F: 5′-CCAACAAGATCGGCGTTAAG-3′

TITF-Ex3-R: 5′-GCCTTCTGGACGGCTCTC-3′

TITF-Ex4-1-F: 5′-GCTAGGCTGCCTGGGTCA-3′

TITF-Ex4-1-R: 5′-GTTTGCCGTCTTTCACCAG-3′

TITF-Ex4-2-F: 5′-CAACAGGCTCAGCAGCAGT-3′

TITF-Ex4-2-R: 5′-AAAGACGTCCAGCAGTTTGG-3′

All PCR was conducted according to the protocol outlined in section 2.4 of Chapter 2. All 5 primer sets underwent 35 cycles with an annealing temperature of 58°C.

PCR products for each of the 4 exons from each of the 26 MD patients (see section 2.1 of Chapter 2 for patient details) underwent Sanger sequencing according to the protocol outlined in section 2.5 of Chapter 2.

5.3. Results

Sequencing revealed that there were no novel mutations in any of the 4 exons of the *TITF-1* gene for any of the 26 patients screened.

5.4. Discussion

Given that none of the 26 patients belonging to the MD cohort tested positive for mutations in the *TITF-1* gene, they are all appropriate candidates for screening for the ACC duplication in the CD108131 transcript identified in MF1.

Chapter 6: Screening the Myoclonus-Dystonia Patient Cohort for Mutations in CD108131

6.1. Introduction

Given that a novel mutation has been found in Myoclonus Family 1 that is believed to be disease-causing, other MD patients must be screened for the same mutation in order to attempt to identify the genetic cause of the disease in these individuals. If additional MD patients share the same mutation as that identified in MF1, it provides support for that mutation's disease-causing potential, as well as provides a diagnostic screening tool in those families. Considering that mutations in *SGCE* only account for ~40% of MD cases, the CD108131 ncRNA provides an additional target for the identification of MD-causing mutations in the remaining ~60% of patients.

In order to determine whether any additional MD patients external to MF1 share the ACC duplication in the CD108131 ncRNA, PCR amplification and Sanger sequencing of the CD108131 transcript was performed on all 26 of the individuals belonging to our MD patient cohort. In addition, 48 *SGCE* mutation-negative MD patients obtained from a German cohort were screened for mutations in CD108131.

6.2. Materials and Methods

6.2.1. Primer design and PCR of the CD108131 transcript

The following primers were designed to amplify CD108131 by PCR:

Forward Primer: 5′-GTCTGAATTTGCAGTTACTGAGAGGC-3′

Reverse Primer: 5′-GTGAGGTTGACACTAACAAATTCACCC-3′

PCR was performed on all 26 members of the MD patient cohort (described in section 2.1

of Chapter 2), as well as on all 48 members of a German MD patient cohort (obtained from Dr. Christine Klein, Department of Neurology, University of Lübeck, Lübeck, Germany), totalling 74 *SGCE* mutation-negative MD patients. The PCR was run for 35 cycles with an annealing temperature of 58°C.

6.2.2. Sequencing the CD108131 ncRNA

The same primers used for CD108131 amplification were used for Sanger sequencing of the PCR product. Sanger sequencing was performed according to the protocol outlined in section 2.5 of Chapter 2.

6.3. Results

All 74 of the *SGCE* mutation-negative MD patients from the two cohorts were screened for the ACC duplication; they all tested negative. However one individual, M36-1, had a different novel mutation in CD108131: a heterozygous T to A change that corresponds with chr18:5287648 (CD108131.1(CD108131_v001):c.642T>A; Fig. 6.1a). This mutation was confirmed by Sanger sequencing and was absent in the 186 control chromosomes tested.

6.4. Discussion

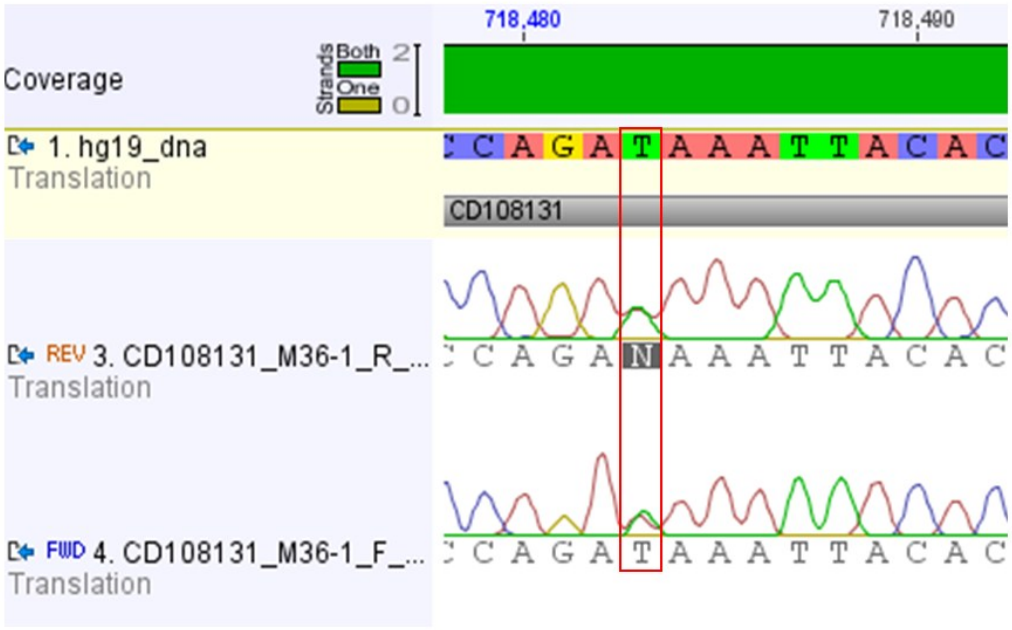
The ACC duplication identified in MF1 appears to be an extremely rare, familial mutation unique to affected individuals and carriers in that family. While the screening of 74 *SGCE* mutation-negative MD patients did not result in the identification of an ACC duplication in the CD108131 ncRNA outside of MF1, it did result in the identification of a second novel variant in the same ncRNA (Fig. 6.1). This result suggests that while the occurrence is rare, MD

can be attributed to different mutations in the CD108131 ncRNA.

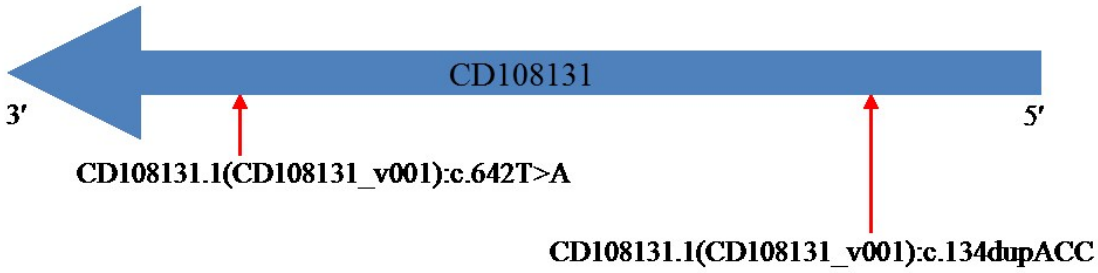
Out of 75 MD families that tested negative for mutations in *SGCE*, 2 tested positive for mutations in CD108131, suggesting that ~3% of MD cases can be attributed to mutations in the CD108131 ncRNA. This implies that while *SGCE* and CD108131 are key players in MD pathology, there are additional genes that, when mutated, result in the same disease.

Figure 6.1. Mutation sites identified in the CD108131 non-coding RNA. A) Screenshot of the Sanger sequencing results of a heterozygous T>A mutation occurring in patient M36-1 (CD108131.1(CD108131_v001):c.642T>A)). This mutation was absent in the 186 control chromosomes tested. B) Two unique mutations have been identified in the CD108131 non-coding RNA, one 134 bases from the 5′ end of the transcript (seen in MF1) and one 221 bases from the 3′ end of the transcript (seen in M36-1).

A)


718,480
718,490
Coverage
Strands
Both
One
2
0
1. hg19_dna
Translation
C C A G A T A A A T T A C A C
CD108131
REV 3. CD108131_M36-1_R_...
Translation
C C A G A N A A A T T A C A C
FWD 4. CD108131_M36-1_F_...
Translation
C C A G A T A A A T T A C A C

B)


CD108131
3′
5′
CD108131.1(CD108131_v001):c.642T>A
CD108131.1(CD108131_v001):c.134dupACC

Chapter 7: Exome Sequencing the Myoclonus-Dystonia Patient Cohort

7.1. Introduction

Given that only ~40% of MD cases can be linked to mutations in *SGCE*, and that only ~3% of cases can be linked to mutations in the CD108131 ncRNA, there are clearly additional genes implicated in this disease. In order to determine what this (these) gene(s) could be, patients from the MD cohort with unidentified mutations are being sent for exome sequencing. Exome sequencing is a method of next-generation sequencing whereby only the protein-coding portion of the genome is captured and sequenced (Gilissen *et al*., 2012). This allows for the comprehensive screening of all candidate variants resulting in direct amino acid or splicing alterations. Exome sequencing is especially useful when insufficient family information exists to perform genetic mapping studies or when little is known about a given disease pathway, making candidate disease-genes difficult to predict (Gilissen *et al*., 2012). Considering that very little is known about the function of *SGCE* or the identity of its interacting partners, exome sequencing provides a means for which novel genes commonly mutated in MD patients can be identified.

In the case of our MD cohort, the majority of patients were recruited as single individuals with very little familial information and/or no parental DNA. This reality makes genetic mapping and the identification of familial variants impossible; however, by comparing the genetic variants between patients affected with MD, the gene(s) responsible for the disease can be identified through common mutations. Through the identification of mutations that are unique to individuals affected with the disease (or through the identification of multiple unique mutations in the same gene), additional disease-linked genes can be elucidated. In order to accomplish this using our MD patient cohort, 4 individuals were selected to undergo exome

sequencing: M9-1, M11-1, M13-1, and M23-1. Once sequencing was complete, each individual's variants could be compared and commonly occurring mutations or commonly mutated genes could be identified.

While the linkage analysis performed on MF1 resulted in strong evidence for a novel MD locus on chromosome 18p11 (lod score of 3.9; Grimes, *et al*., 2002), the ability to perform exome sequencing provides an additional means to exclude the possibility of disease-causing variants elsewhere in the genome. For this reason, four affected members of MF1 (M1-1, M1-5, M1-15, and M1-37) were also sent for exome sequencing in order to eliminate the possibility of a disease-causing mutation in a protein-coding sequence outside of the region of interest on 18p11.

7.2. Materials and Methods

7.2.1. DNA extraction and sample preparation

DNA was extracted from the blood of patients M9-1, M11-1, M13-1, M23-1, M1-1, M1-5, M1-15, and M1-37 according to the protocol outlined in section 2.2 of Chapter 2. DNA samples were run on a 1% agarose gel in order to ensure that no degradation had occurred. In order to meet quality standards for exome sequencing, a minimum of 3 μg of DNA had to be supplied with a minimum 260/280 ratio of 1.8 and 260/230 ratio of 1.5.

7.2.2. Targeted capture and sequencing methods

DNA samples were sent to the McGill University and Génome Québec Innovation Centre (Montreal, QC) for exome capture followed by next-generation sequencing. The samples were processed as part of the Finding of Rare Disease Genes (FORGE) Canada Consortium, for which

the study design was approved by the institutional research ethics board at the Children's Hospital of Eastern Ontario. The capture and sequencing protocols were performed as described in Lines *et al*. (2012). The Agilent SureSelect 50 Mb All Exon Kit was used for target enrichment and sequencing was performed using the Illumina Hiseq (Lines *et al*., 2012).

7.2.3. Sequence alignment and variant identification

Bioinformatic analysis was performed according to the FORGE pipeline, as described in Lines *et al*. (2012). In brief, the sequence reads were aligned to the hg19 reference genome using the Burrows-Wheeler Aligner, variants were called using SAMtools pileup and varaFilter, and variants were annotated using Annovar (Lines *et al*., 2012).

7.3. Results

7.3.1. Exome results for Myoclonus Family 1

Exome sequencing and bioinformatic analysis identified 2 variants that were shared between M1-1, M1-5, M1-15, and M1-37 (Fig. 7.1). These 2 variants are p.T179I in the *SYCP1* gene and p.V120M in the *VTCN1* gene. Sanger sequencing confirmed that the two aforementioned variants are real (Fig. 7.2), however they do not co-segregate with the disease in the family (Table 7.1). The affected family members M1-6, M1-16, M1-20, and M1-23, as well as the obligate carrier M1-14, do not possess the mutant allele in *SYCP1* (Table 7.1). The affected family member M1-23 and the obligate carrier M1-14 are lacking the mutant allele in *VTCN1* (Table 7.1). No protein-coding variant exists in MF1 that is shared by all affected members and obligate carriers.

Figure 7.1. Number of shared variants in Myoclonus Family 1 (MF1) as determined by exome sequencing. Venn diagram summarizing the number of variants identified by exome sequencing for 4 members of MF1. There are two variants shared by all 4 individuals, a p.T179I in the *SYCP1* gene and p.V120M in the *VTCN1* gene.

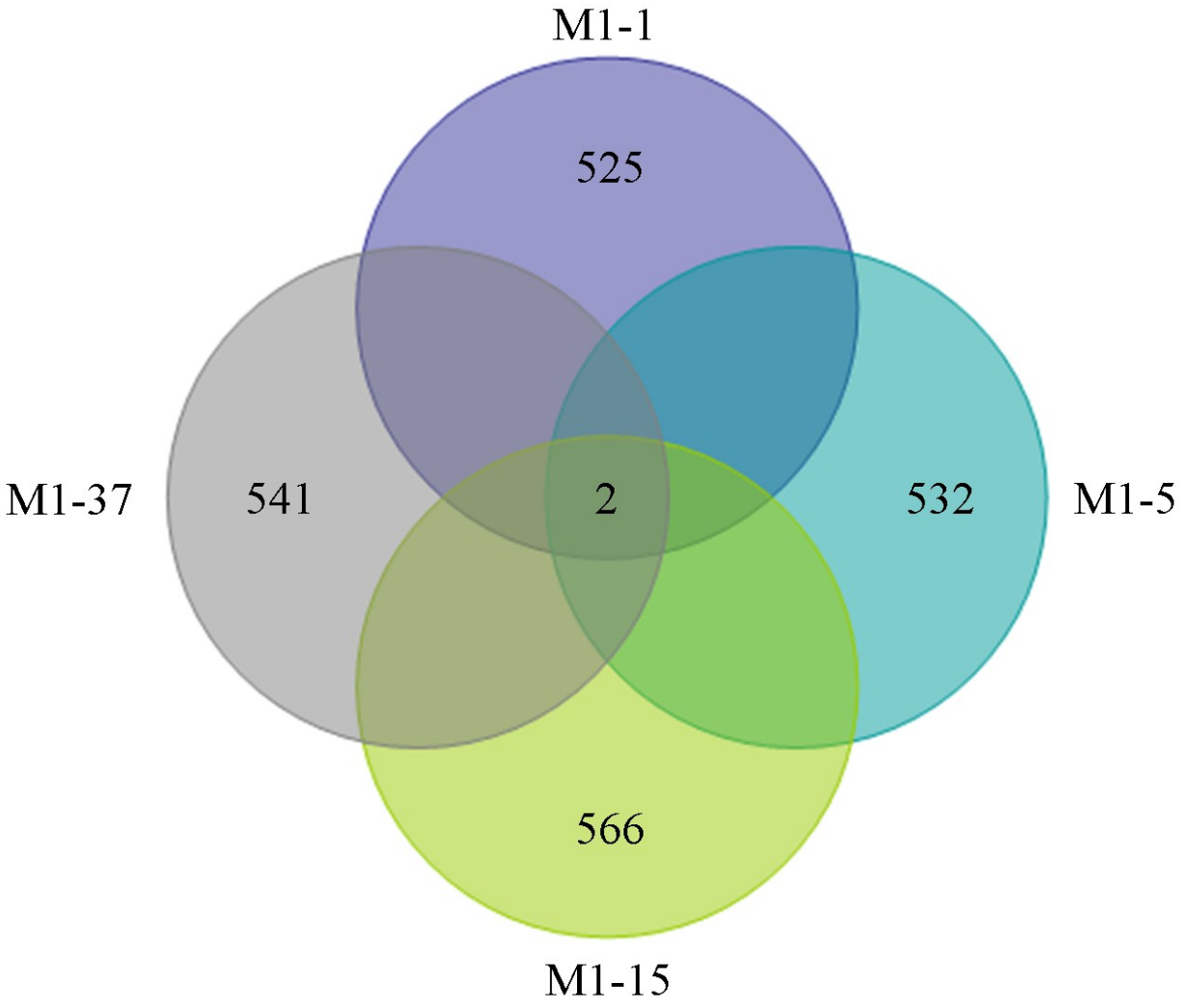

M1-1
525
M1-37
541
2
532
M1-5
566
M1-15

Figure 7.2. Validating the *SYCP1* and *VTCN1* variants by Sanger sequencing. Screenshots of the Sanger sequencing results for the (A) *SYCP1* and (B) *VTCN1* variants identified by exome sequencing and shared by M1-1, M1-5, M1-15, and M1-37.


A)
Coverage
Strands Both One
5,910
5,920
1. SYCP1 - 6847
Translation
Complement
SYCP1 mRNA
SYCP1 CDS; 1; synaptonemal complex protein 1
SYCP1 gene
REV 2. M1-37_SYCP1...
REV 3. M1-5_SYCP1_...
REV 4. M1-15_SYCP1...
REV 5. M1-1_SYCP1_...

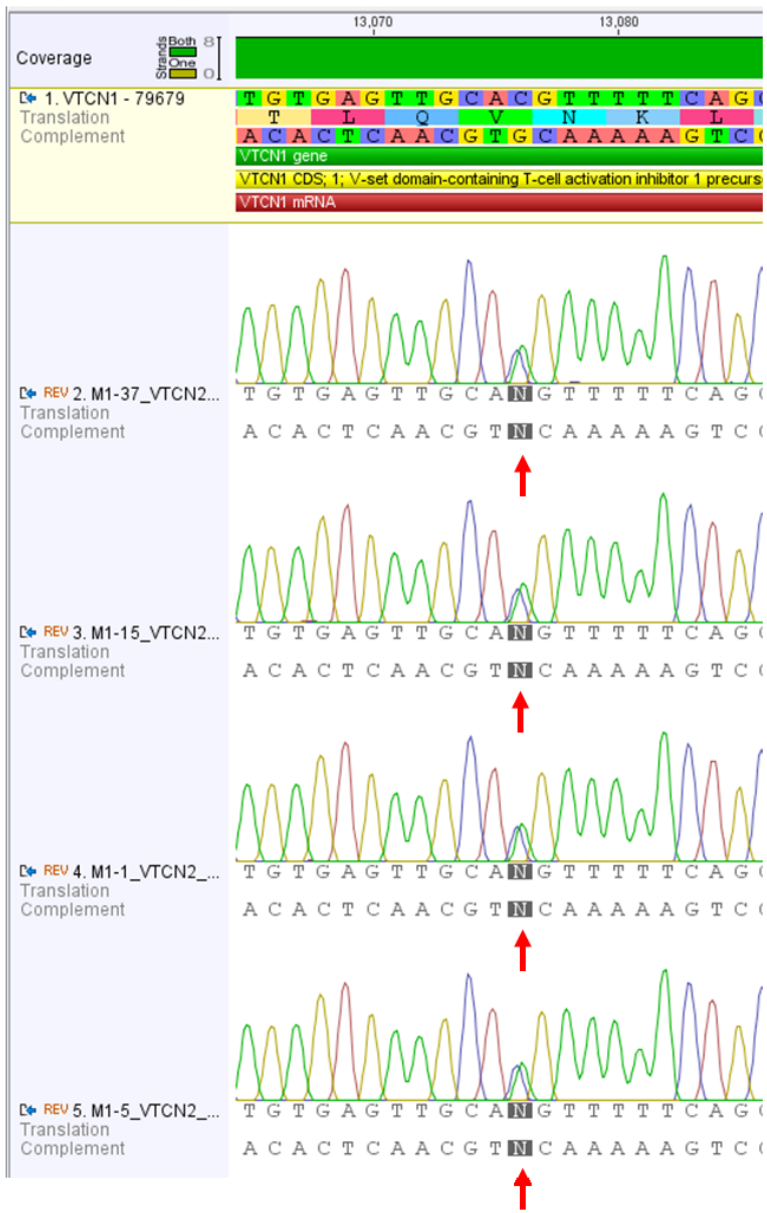

B)
Coverage
Strands Both One
13,070
13,080
1. VTCN1 - 79679
Translation
Complement
VTCN1 gene
VTCN1 CDS; 1; V-set domain-containing T-cell activation inhibitor 1 precurs
VTCN1 mRNA
REV 2. M1-37_VTCN2...
REV 3. M1-15_VTCN2...
REV 4. M1-1_VTCN2_...
REV 5. M1-5_VTCN2_...

Table 7.1. Testing for co-segregation of the *SYCP1* and *VTCN1* variants with the disease in Myoclonus Family 1 (MF1). Presence (+) or absence (-) of the wild-type and mutant alleles was determined by Sanger sequencing. Affected individuals are indicated by red font, unaffected carriers are indicated by blue font. All unaffected non-carriers are indicated by black font. Cells highlighted in yellow indicate affected individuals or obligate carriers who lack the mutant allele, while unaffected (non-obligate carriers) individuals who have the mutated allele are highlighted in orange.

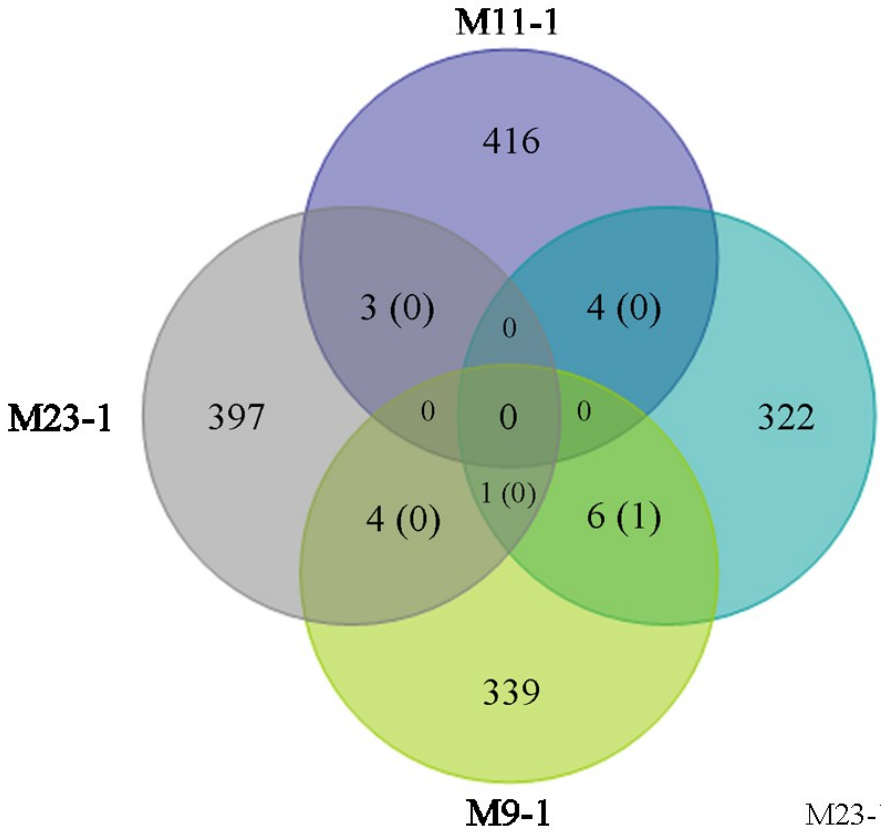
Family Member	SYCP1 Wild-Type Allele	SYCP1 Mutant Allele	VTCN1 Wild-Type Allele	VTCN1 Mutant Allele
M1-1	+	+	+	+
M1-2	+	-	+	-
M1-3	+	+	+	+
M1-4	+	-	+	-
M1-5	+	+	+	+
M1-6	+	-		
M1-7	+	-	+	-
M1-8	+	+	+	+
M1-9	+	-	+	-
M1-11	+	-	+	-
M1-12	+	-	+	-
M1-13	+	+	+	+
M1-14	+	-	+	-
M1-15	+	+	+	+
M1-16	+	-	+	+
M1-17	+	+	+	+
M1-18	+	+	+	+
M1-19	+	+	+	+
M1-20	+	-	+	+
M1-21	+	+	+	+
M1-22	+	+	+	+
M1-23	+	-	+	-
M1-24	+	-	+	-
M1-25	+	+	+	+
M1-26	+	-	+	-
M1-27	+	+	+	+
M1-28	+	+	+	+
M1-29	+	+	+	+
M1-30	+	-	+	-
M1-31	+	-	+	-
M1-32	+	-	+	-
M1-33	+	-	+	-
M1-34	+	+	+	+
M1-35	+	-	+	-
M1-36	+	-	+	-
M1-37	+	+	+	+

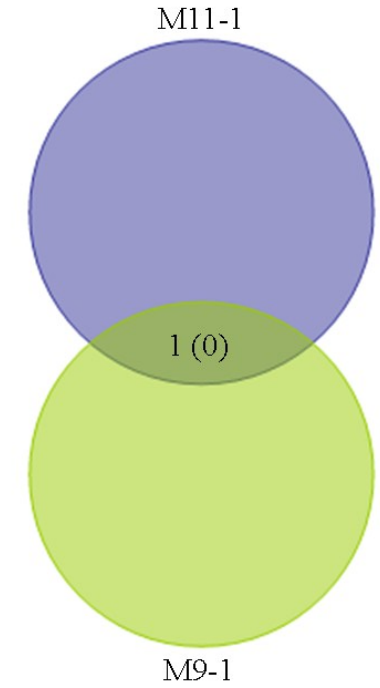
7.3.2. Exome results for M11-1, M13-1, M9-1, and M23-1

When considering all variants (excluding synonymous changes and those variants found in dbSNP), exome sequencing identified 416, 322, 339, and 397 variants in M11-1, M13-1, M9-1, and M23-1 respectively (Fig. 7.3). There was no single variant that occurred in all four individuals (Fig. 7.3). There was a single variant, a p.109_119del change in the *CRIPAK* gene that occurred in M13-1, M9-1, and M23-1; however it did not meet the filter criteria for rarity as it has occurred in control exomes (Fig. 7.3). While there were 25 variants shared by 2 of the 4 patients that underwent exome sequencing, only 1 of these met the filter criteria for rarity: a heterozygous p.P483L change in the *ALAS2* gene occurring in both M13-1 and M9-1 (Fig. 7.3).

When considering the number of genes containing variants (excluding synonymous changes and those variants found in dbSNP), exome sequencing identified 404, 308, 332, and 382 genes in M11-1, M13-1, M9-1, and M23-1 respectively (Fig. 7.4). When comparing between the 4 individuals, there was no single gene that contained mutations for M11-1, M13-1, M9-1, and M23-1 (Fig. 7.4). There was a single gene, *TTN*, that carried separate mutations in M11-1, M13-1, and M9-1, however these mutations did not all pass the filter criteria for rarity (Fig. 7.4). When considering all of the genes with separate mutations in 2 out of the 4 patients that underwent exome sequencing, there were 48 candidates, 35 of which met the criteria for rarity (Fig. 7.4); these 35 candidates are summarized in Table 7.2.

Figure 7.3. Variants identified by exome sequencing for M11-1, M13-1, M9-1, and M23-1. Venn diagram summarizing the number of variants identified by exome sequencing for 4 members of the *SGCE* mutation-negative MD cohort. Values represent the number of variants identified excluding synonymous mutations and mutations found in dbSNP. Values in brackets represent the number of variants that pass the filter for rarity (have not been observed in control exomes).


M11-1
416
3 (0)
4 (0)
0
M23-1
397
0
0
0
322
M13-1
1 (0)
4 (0)
6 (1)
339
M9-1


M11-1
1 (0)
M9-1

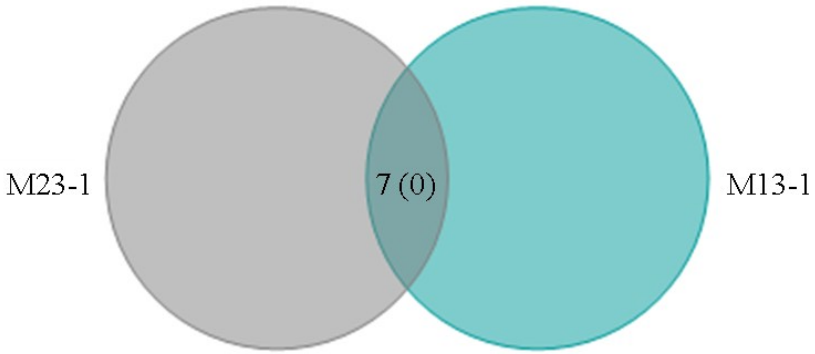
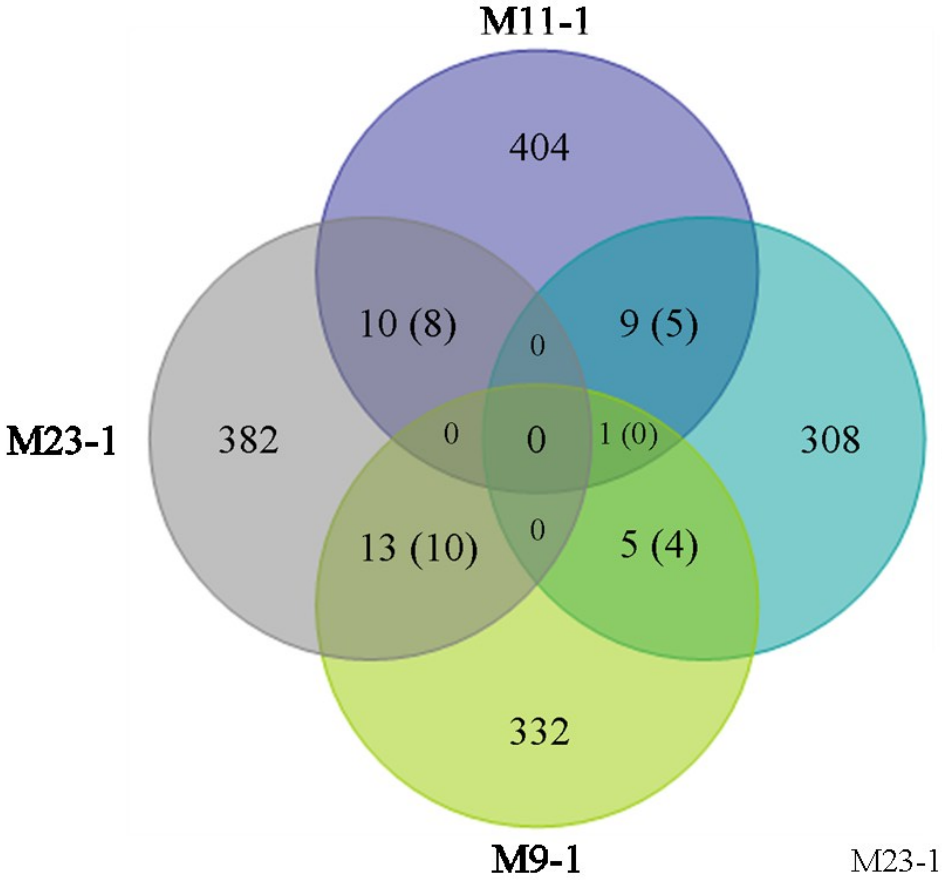
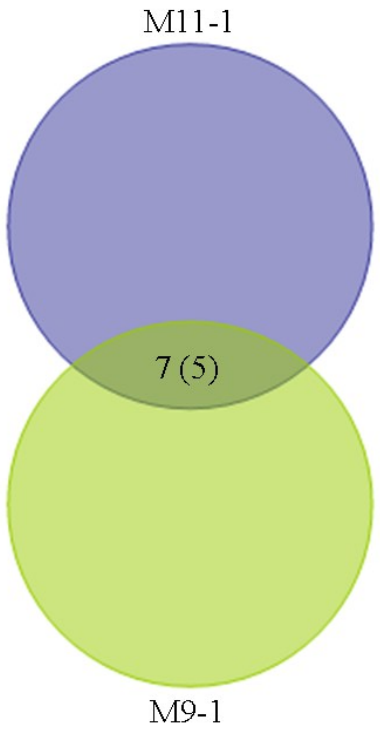

M23-1
7 (0)
M13-1

Figure 7.4. Genes containing rare variants identified by exome sequencing for M11-1, M13-1, M9-1, and M23-1. Venn diagram summarizing the number of genes containing variants identified by exome sequencing for 4 members of the *SGCE* mutation-negative MD cohort. Values represent the number of genes containing variants (excluding synonymous mutations and mutations found in dbSNP). Values in brackets represent the number of genes containing variants that pass the filter for rarity (have not been observed in control exomes).


M11-1
404
10 (8)
9 (5)
0
M23-1
382
0
0
1 (0)
308
M13-1
13 (10)
0
5 (4)
332
M9-1


M11-1
7 (5)
M9-1

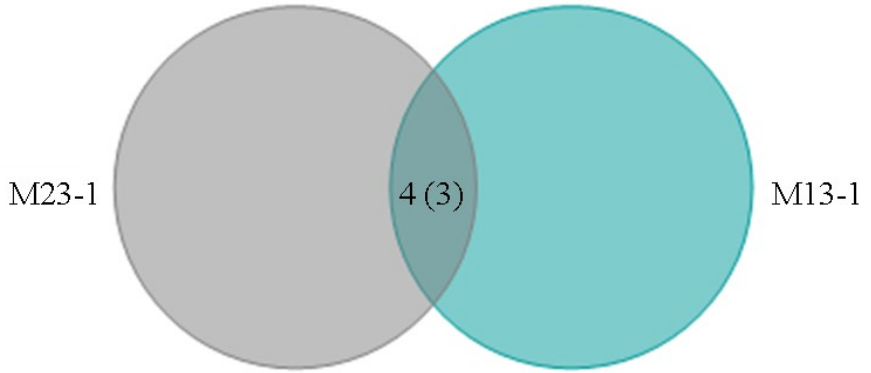

M23-1
4 (3)
M13-1

Table 7.2. Genes containing at least 2 novel variants indentified in patients M11-1, M13-1, M9-1, or M13-1. AA = Amino Acid; het = heterozygous genotype; hom = homozygous genotype.

Position	Gene	Variation	AA Change	M11-1	M13-1	M9-1	M23-1
chr1:247058005	AHCTF1	c.C1787T	p.T596M	-	-	het	-
chr1:247065917	AHCTF1	c.A1054G	p.R352G	-	-	-	het
chr4:114163351	ANK2	c.G877A	p.D293N	-	-	het	-
chr4:114274230	ANK2	c.G4456C	p.V1486L	-	-	-	het
chr12:27533227	ARNTL2	c.A374G	p.N125S	-	-	het	-
chr12:27573415	ARNTL2	c.G1861T	p.G621W	-	-	-	het
chr2:25965362	ASXL2	c.A3844T	p.I1282F	-	-	het	-
chr2:25965390	ASXL2	c.G3816T	p.Q1272H	-	-	het	-
chr2:25991689	ASXL2	c.A553G	p.S185G	-	-	-	het
chr6:30615304	C6orf136	c.G296T	p.R99M	-	-	-	het
chr6:30615550	C6orf136	c.C542T	p.P181L	-	-	hom	-
chr3:50645082	CISH	c.C733G	p.R245G	het	-	-	-
chr3:50645812	CISH	c.G233A	p.R78Q	-	-	-	het
chr1:205033774	CNTN2	c.C1415A	p.T472N	het	-	-	-
chr1:205038648	CNTN2	c.G2155A	p.G719R	-	het	-	-
chr4:166300588	CPE	c.C215T	p.A72V	het	-	-	-
chr4:166405657	CPE	c.A874G	p.M292V	-	-	het	-
chr19:11327648	DOCK6	c.C3836G	p.P1279R	-	het	-	-
chr19:11327649	DOCK6	c.C3835T	p.P1279S	-	het	-	-
chr19:11364374	DOCK6	c.G73T	p.V25L	-	-	-	het
chr2:55130214	EML6	c.A3158T	p.K1053I	-	-	-	het
chr2:55185121	EML6	c.G4681A	p.V1561M	-	-	het	-
chr16:67575421	FAM65A	c.T902G	p.V301G	-	-	-	het
chr16:67575590	FAM65A	c.T997G	p.S333A	het	-	-	-
chr9:4118061	GLIS3	c.C952T	p.L318F	het	-	-	-
chr9:4286272	GLIS3	c.G154A	p.A52T	-	-	-	het
chr11:118959417	HMBS	exonic;splicing		het	-	-	-
chr11:118963869	HMBS	c.G962A	p.R321H	-	-	-	het
chr3:113376130-113376137	KIAA2018	c.4392_4399del	p.1464_1467del	-	het	-	-
chr3:113376137-113376144	KIAA2018	c.4385_4392del	p.1462_1464del	-	-	het	-
chr17:38938525	KRT27	c.G221A	p.G74E	-	-	-	het
chr17:38938724	KRT27	c.A22G	p.T8A	het	-	-	-
chr19:9019278	MUC16	c.G37609A	p.A12537T	het	-	-	-
chr19:9066036	MUC16	c.C21410T	p.T7137I	-	-	het	-
chr11:1264346	MUC5B	c.C6236A	p.T2079K	-	-	het	-
chr11:1267052	MUC5B	c.C8942T	p.A2981V	-	het	-	-
chr2:203990219	NBEAL1	c.G2740A	p.V914I	-	-	-	het
chr2:204075826	NBEAL1	c.A7844G	p.H2615R	-	het	-	-
chr3:47044189	NBEAL2	c.A5356T	p.T1786S	-	-	-	het
chr3:47045708	NBEAL2	c.C6023T	p.T2008M	-	-	het	-
chr1:902104	PLEKHN1	c.C104T	p.S35L	hom	-	-	-
chr1:906298	PLEKHN1	c.G524A	p.R175Q	-	het	-	-

chr8:10466743	RP1L1	c.G4865A	p.R1622Q	-	-	-	het
chr8:10480107	RP1L1	c.A605G	p.K202R	-	-	het	-
chr7:4119163	SDK1	c.C3271T	p.R1091W	-	-	-	het
chr7:4245579	SDK1	c.G5167A	p.D1723N	-	-	het	-
chr14:65236330	SPTB	c.G5915A	p.R1972Q	-	-	het	-
chr14:65263313	SPTB	c.G1303C	p.E435Q	-	het	-	-
chr15:99670730	SYNM	c.C2164T	p.P722S	-	-	het	-
chr15:99673100	SYNM	c.G4534A	p.E1512K	het	-	-	-
chr1:36551441	TEKT2	c.A287G	p.K96R	-	-	het	-
chr1:36552372	TEKT2	c.G556A	p.D186N	het	-	-	-
chr9:75435805	TMC1	c.G1811A	p.R604Q	-	het	-	-
chr9:75450866	TMC1	c.2261-1G>A		het	-	-	-
chr1:43738615-	TMEM125	c.222_223del	p.74_75del	-	-	-	het
chr1:43739021	TMEM125	c.G628A	p.E210K	-	-	het	-
chr18:72997924	TSHZ1	c.A427G	p.T143A	-	-	-	het
chr18:72999734	TSHZ1	c.C2237G	p.P746R	het	-	-	-
chr2:74717790	TTC31	c.A470G	p.N157S	-	het	-	-
chr2:74720326	TTC31	c.A1541G	p.H514R	-	-	het	-
chr7:6194034	USP42	c.G2849A	p.R950Q	-	het	-	-
chr7:6196638	USP42	c.C3895T	p.R1299W	het	-	-	-
chr6:144768360	UTRN	c.C1628T	p.A543V	-	-	-	het
chr6:145069426	UTRN	c.C7984T	p.R2662C	het	-	-	-
chr15:62214756	VPS13C	c.T6686C	p.I2229T	-	-	het	-
chr15:62237973	VPS13C	c.A4960T	p.S1654C	het	-	-	-
chr16:2049508	ZNF598	c.C2044T	p.R682C	-	het	-	-
chr16:2049660	ZNF598	c.C1892T	p.P631L	-	-	-	het
chr16:2049898	ZNF598	c.1654_1655insAGG	p.R552delinsRR	-	-	-	hom
chr2:71591119	ZNF638	c.G1454A	p.R485Q	-	-	-	het
chr2:71623300	ZNF638	c.G2407T	p.A803S	het	-	-	-
chr2:71645749	ZNF638	c.C3279A	p.D1093E	het	-	-	-
chr7:63808506	ZNF736	c.C265T	p.H89Y	het	-	-	-
chr7:63809227	ZNF736	c.G986A	p.S329N	-	het	-	-

7.4. Discussion

Exome sequencing of four affected individuals from MF1 revealed that there is no protein-coding variant responsible for causing MD in this family. This data supports our linkage to chromosome 18p11 and the identification of a 3 bp duplication in the CD108131 ncRNA as the disease-causing mutation in this family.

Exome sequencing of four unrelated affected members of the *SGCE* mutation-negative MD cohort— M11-1, M13-1, M9-1, and M37-1— did not reveal a single disease-causing mutation, or a single disease-causing gene with individual mutations. Given that there were no shared mutations that met the filter criteria for rarity in 2 or more of the individuals, MD in these patients can likely be explained by individual mutations in the same gene(s). There is likely no single gene that can account for all of the MD cases where mutations in *SGCE* and CD108131 are absent. This is supported by the exome results; there was no one gene containing individual mutations that met the criteria for rarity in greater than 2 out of the 4 individuals. There were 35 gene candidates with individual mutations in 2 out of the 4 patients sent for exome sequencing (Table 7.2). There are several candidates within this list of 35 that are potentially very relevant to Myoclonus-Dystonia, due to their functional similarities to what is known about *SGCE*; these candidates include: *ANK2*, *SDK1*, and *UTRN* (Table 7.2).

Ankyrin 2 (*ANK2*), also known as ankyrin-B, is part of the ankyrin protein family, which are responsible for linking the integral membrane proteins to the underlying spectrin-actin cytoskeleton. Through its interactions with β2 spectrin and dynactin-4, *ANK2* localizes dystrophin, dystroglycan, and microtubules to costameres in muscle (Ayalon *et al*., 2011). *ANK2* also binds to and mediates the sarcolemmal association of dystrophin, dynactin-4, and microtubules (Ayalon *et al*., 2011). Dystrophin binds membrane-associated glycoproteins

(including the family members of *SGCE*) to form the dystrophin-glycoprotein complex; this functional link to *SGCE* makes *ANK2* an attractive candidate from the exome sequencing results.

Sidekick cell adhesion molecule 1 (*SDK1*) is a transmembrane cell adhesion protein that, in developing neurons, functions in guiding axonal terminals to specific synapses. So far, *SDK1* characterization has occurred through analysis of the retina, where it has been determined to interact with synaptic scaffolding proteins in order to form lamina-specific synapses (Yamagata and Sanes, 2010). If this role for *SDK1* is consistent in other regions of the central nervous system, it could prove to be extremely important for synaptic transmission and the proper formation of neural networks, both of which are believed to be related to Myoclonus-Dystonia.

Utrophin (*UTRN*) is a homolog of dystrophin; the two proteins function similarly in binding cytoskeletal actin to the dystroglycoprotein complex at the sarcolemma membrane, however utrophin is specific to fetal and regenerating muscle (Lin *et al*., 2012). By mediating the interactions between the cytoskeleton, cell membrane, and extracellular matrix, dystrophin is involved in the structural organization of ion channels and postsynaptic membrane receptors during synaptogenesis (Perronnet and Vaillend, 2010). Utrophin expression is not limited to muscle, however. In fact, utrophin is expressed ubiquitously, including within the central nervous system (Perronnet and Vaillend, 2010). Considering the dystrophin-dystroglycan association with the heterotetrameric α-, β-, γ-, and δ-sarcoglycan complex in muscle, many speculate a brain-specific association involving either dystrophin or utrophin and ε-sarcoglycan (Perronnet and Vaillend, 2010). This link between utrophin and ε-sarcoglycan could be crucial to understanding the strictly neurological phenotype of Myoclonus-Dystonia.

Analysis of each of these 35 candidates (including the top three summarized above) must be performed in order to determine their disease-causing potential. The mutations identified

must validate by Sanger sequencing and they must co-segregate with the disease. A limitation of this analysis, however, is that no parental DNA exists for any of these patients. Without parental DNA and sufficient familial information to determine disease or carrier status, it cannot be determined whether each of the variants identified is unique to disease-carrying individuals. Therefore, an alternative approach for exome analysis is required. This approach must involve the exome sequencing of additional MD patients from the cohort; the greater the number of sequenced patients, the fewer the number of gene candidates containing variants among affected individuals. Once a gene can be identified with a higher-than-normal frequency of mutations among individuals affected with MD, functional characterization of that gene can occur.

Chapter 8: Conclusions and Future Directions

8.1. Conclusions

I have identified the disease-causing mutation in MF1: a 3 bp duplication that lies within an uncharacterized lncRNA, CD108131. This mutation co-segregates with the disease in the family and has not been identified in any control samples. There is one additional patient affected with MD, outside of MF1, that has a separate, unique mutation in CD108131. This suggests that mutations in this lncRNA could account for ~3% of MD cases. The exact function of this ncRNA and its role in MD pathology is yet to be elucidated. So far, I have determined that it is an 863 bp RNA that is expressed in many tissues, with the highest level of expression in the cerebellum. Expression of the lncRNA itself does not differ between unaffected and affected individuals, however individuals affected with MD that carry the 3 bp duplication in CD108131 have lower expression of *SGCE*, suggesting a possible link between the lncRNA and the known MD gene.

More and more, researchers are identifying novel lncRNAs with important functions in central nervous system development and function. These regulatory RNAs have been implicated in neural differentiation, synaptic plasticity, and stress response (Qureshi *et al*., 2010). In addition, many neurodegenerative disorders are being linked to lncRNAs, including Alzheimer's disease, spinocerebellar ataxia, amyotrophic lateral sclerosis (ALS), and even Huntington's disease (Qureshi *et al*., 2010). Separate from neurodegenerative diseases, lncRNAs have been implicated in Multiple sclerosis, epilepsy, Restless Legs Syndrome, schizophrenia, bipolar disorder, major depression, and autistic spectrum disorders (Qureshi *et al*., 2010).

8.2. Future Directions

Future work on this project must involve further analysis of CD108131 function. The link between CD108131 and *SGCE* (as discussed in Chapter 4) must be further validated. This could be accomplished by (1) further qPCR experiments, (2) binding assays, or (3) knock-down experiments.

(1) Quantitative Real-Time PCR Experiments:

i. Additional screening of controls to obtain a more significant representation of "normal" *SGCE* expression, as well as additional patient screening within MF1 to ensure that the reduced level of *SGCE* expression is a common phenotype among the affected members of the family.

ii. *SGCE* expression analysis in M36-1 in order to determine if the mutation in CD108131 specific to that individual also results in reduced levels.

iii. Analysis of *SGCE* mutation-positive patients in order to compare levels of *SGCE* expression to patients with non-*SGCE* mutations.

(2) Binding Assay Experiments:

i. Chromatin Isolation by RNA Purification and sequencing (ChIRP-seq) could be performed as described by Chu *et al*. (2011). This method would allow for the retrieval and sequencing of the CD108131 ncRNA occupancy sites along the DNA (Chu *et al*., 2011).

(3) Knock-Down Experiments:

i. CD108131 expression could be knocked down in a human cell line by siRNA and the level of *SGCE* expression could be measured. This level of *SGCE* expression could be compared to a cell line with wild-type

CD108131 expression in order to determine whether knocking down CD108131 affects the expression of *SGCE*.

Should a regulatory connection between the CD108131 ncRNA and *SGCE* fail to validate, alternative regulatory targets for the ncRNA must be found. This could be accomplished by microarray analysis of genes that are up or downregulated upon knockdown of the CD108131 ncRNA.

In addition to the further characterization of the CD108131 ncRNA, future work on this project must involve the identification of additional genes involved in MD pathology. Considering that only ~40% of MD patients possess mutations in *SGCE*, only ~3% carry mutations in CD108131, and that of the 4 patients we sent for exome sequencing, there was no one gene with mutations in all 4 individuals, MD is clearly caused by more than one unidentified gene. In order to identify these genes, additional MD patients that test negative for mutations in *SGCE* and CD108131 should be subjected to exome sequencing.

References

Albanese, A., Asmus, F., Bhatia, K.P., Elia, A.E., Elibol, B., Filippini, G., Gasser, T., Krauss, J.K., Nardocci, N., Newton, A., and Valls-Solé, J. (2011). EFNS guidelines on diagnosis and treatment of primary dystonias. European Journal of Neurology *18,* 5-18.

Asmus, F., Devlin, A., Munz, M., Zimprich, A., Gasser, T., and Chinnery, P.F. (2007). Clinical differentiation of genetically proven benign hereditary chorea and myoclonus-dystonia. Movement Disorders *22,* 2104-2109.

Asmus, F., and Gasser, T. (2010). Dystonia-plus syndromes. European Journal of Neurology *17,* 37-45.

Ayalon, G., Hostettler, J.D., Hoffman, J., Kizhatil, K., Davis, J.Q., and Bennett, V. (2011). Ankyrin-B interactions with spectrin and dynactin-4 are required for dystrophin-based protection of skeletal muscle from exercise injury. J. Biol. Chem. *286,* 7370-7378.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R.*, et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456,* 53-59.

Beukers, R.J., Foncke, E.M.J., Van Der Meer, J.N., Nederveen, A.J., De Ruiter, M.B., Bour, L.J., Veltman, D.J., and Tijssen, M.A.J. (2010). Disorganized sensorimotor integration in mutation-positive myoclonus-dystonia: A functional magnetic resonance imaging study. Arch. Neurol. *67,* 469-474.

Bragg, D.C., Armata, I.A., Nery, F.C., Breakefield, X.O., and Sharma, N. (2011). Molecular pathways in dystonia. Neurobiol. Dis. *42,* 136-147.

Bressman, S.B. (2007). Genetics of dystonia: an overview. Parkinsonism and Related Disorders *13,*

Camargos, S., Scholz, S., Simón-Sánchez, J., Paisán-Ruiz, C., Lewis, P., Hernandez, D., Ding, J., Gibbs, J.R., Cookson, M.R., Bras, J.*, et al.* (2008). DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress-response protein PRKRA. The Lancet Neurology *7,* 207-215.

Chow, J., and Heard, E. (2009). X inactivation and the complexities of silencing a sex chromosome. Curr. Opin. Cell Biol. *21,* 359-366.

Chu, C., Qu, K., Zhong, F., Artandi, S., and Chang, H. (2011). Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Mol. Cell *44,* 667-678.

De Carvalho Aguiar, P., Sweadner, K.J., Penniston, J.T., Zaremba, J., Liu, L., Caton, M.,

Linazasoro, G., Borg, M., Tijssen, M.A.J., Bressman, S.B., *et al.* (2004). Mutations in the Na+/K+-ATPase α3 gene ATP1A3 are associated with rapid-onset dystonia parkinsonism. Neuron *43,* 169-175.

Ettinger, A.J., Feng, G., and Sanes, J.R. (1997). ε-sarcoglycan, a broadly expressed homologue of the gene mutated in limb-girdle muscular dystrophy 2d. J. Biol. Chem. *272,* 32534-32538.

Fahn, S., Bressman, S.B., and Marsden, C.D. (1998). Classification of dystonia. Adv. Neurol. *78,* 1-10.

Fuchs, T., Gavarini, S., Saunders-Pullman, R., Raymond, D., Ehrlich, M.E., Bressman, S.B., and Ozelius, L.J. (2009). Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia. Nat. Genet. *41,* 286-288.

Furukawa, Y., and Rajput, A.H. (2002). Inherited myoclonus-dystonia: How many causative genes and clinical phenotypes? Neurology *59,* 1130-1131.

Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2012). Disease gene identification strategies for exome sequencing. European Journal of Human Genetics *20,* 490-497.

Gordon, N. (2008). Segawa's disease: Dopa-responsive dystonia. Int. J. Clin. Pract. *62,* 943-946.

Grabowski, M., Zimprich, A., Lorenz-Depiereux, B., Kalscheuer, V., Asmus, F., Gasser, T., Meitinger, T., and Strom, T.M. (2003). The epsilon-sarcoglycan gene (SGCE), mutated in myoclonus-dystonia syndrome, is maternally imprinted. European Journal of Human Genetics *11,* 138-144.

Grimes, D.A., Bulman, D., St. George-Hyslop, P., and Lang, A.E. (2001). Inherited myoclonus-dystonia: Evidence supporting genetic heterogeneity. Movement Disorders *16,* 106-110.

Grimes, D.A., Han, F., Lang, A.E., St. George-Hyssop, P., Racacho, L., and Bulman, D.E. (2002). A novel locus for inherited myoclonus-dystonia on 18p11. Neurology *59,* 1183-1186.

Grünewald, A., Djarmati, A., Lohmann-Hedrich, K., Farrell, K., Zeller, J.A., Allert, N., Papengut, F., Petersen, B., Fung, V., Sue, C.M., *et al.* (2008). Myoclonus-dystonia: significance of large SGCE deletions. Hum. Mutat. *29,* 331-332.

Han, F., Racacho, L., Lang, A.E., Bulman, D.E., and Grimes, D.A. (2007). Refinement of the DYT15 locus in myoclonus dystonia. Movement Disorders *22,* 888-892.

Hubsch, C., Vidailhet, M., Rivaud-Péchoux, S., Pouget, P., Brochard, V., Degos, B., Pélisson, D., Golmard, J.-., Gaymard, B., and Roze, E. (2011). Impaired saccadic adaptation in DYT11 dystonia. Journal of Neurology, Neurosurgery and Psychiatry *82,* 1103-1106.

Ichinose, H., Ohye, T., Takahashi, E.-., Seki, N., Hori, T.-., Segawa, M., Nomura, Y., Endo, K., Tanaka, H., Tsuji, S., Fujita, K., and Nagatsu, T. (1994). Hereditary progressive dystonia with marked diurnal fluctuation caused by mutations in the GTP cyclohydrolase I gene. Nat. Genet. *8,* 236-242.

Lalli, S., Canavese, C., Zorzi, G., Nardocci, N., and Albanese, A. (2011). Diagnostic issues in childhood and adult dystonia. Expert Opinion on Medical Diagnostics *5,* 483-500.

Lin, A.Y., Prochniewicz, E., Henderson, D.M., Li, B., Ervasti, J.M., and Thomas, D.D. (2012). Impacts of dystrophin and utrophin domains on actin structural dynamics: Implications for therapeutic design. J. Mol. Biol. *420,* 87-98.

Lines, M.A., Huang, L., Schwartzentruber, J., Douglas, S.L., Lynch, D.C., Beaulieu, C., Guion-Almeida, M.L., Zechi-Ceide, R.M., Gener, B., Gillessen-Kaesbach, G.*, et al.* (2012). Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. Am. J. Hum. Genet. *90,* 369-377.

Ludecke, B., Dworniczak, B., and Bartholome, K. (1995). A point mutation in the tyrosine hydroxylase gene associated with Segawa's syndrome. Hum. Genet. *95,* 123-125.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-., Chen, Z.*, et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437,* 376-380.

McNally, E.M., Ly, C.T., and Kunkel, L.M. (1998). Human ε-sarcoglycan is highly related to α-sarcoglycan (adhalin), the limb girdle muscular dystrophy 2D gene. FEBS Lett. *422,* 27-32.

Moorthie, S., Mattocks, C.J., and Wright, C.F. (2011). Review of massively parallel DNA sequencing technologies. HUGO Journal *5,* 1-12.

Müller, B., Hedrich, K., Kock, N., Dragasevic, N., Svetel, M., Garrels, J., Landt, O., Nitschke, M., Pramstaller, P.P., Reik, W., *et al*. (2002). Evidence that paternal expression of the ε-Sarcoglycan gene accounts for reduced penetrance in myoclonus-dystonia. Am. J. Hum. Genet. *71*, 1301-1311.

Nishiyama, A., Endo, T., Takeda, S., and Imamura, M. (2004). Identification and characterization of ε-sarcoglycans in the central nervous system. Mol. Brain Res. *125,* 1-12.

Norgren, N., Mattson, E., Forsgren, L., and Holmberg, M. (2011). A high-penetrance form of late-onset torsion dystonia maps to a novel locus (DYT21) on chromosome 2q14.3-q21.3. Neurogenetics *12,* 137-143.

Nygaard, T.G., Raymond, D., Chen, C., Nishino, I., Greene, P.E., Jennings, D., Heiman, G.A., Klein, C., Saunders-Pullman, R.J., Kramer, P., Ozelius, L.J., and Bressman, S.B. (1999).

Localization of a gene for myoclonus-dystonia to chromosome 7q21-q31. Ann. Neurol. *46,* 794-798.

Ørom, U.A., and Shiekhattar, R. (2011). Noncoding RNAs and enhancers: Complications of a long-distance relationship. Trends in Genetics *27,* 433-439.

Ozawa, E., Mizuno, Y., Hagiwara, Y., Sasaoka, T., and Yoshida, M. (2005). Molecular and cell biology of the sarcoglycan complex. Muscle and Nerve *32,* 563-576.

Ozelius, L.J., Hewett, J.W., Page, C.E., Bressman, S.B., Kramer, P.L., Shalish, C., De Leon, D., Brin, M.F., Raymond, D., Corey, D.P*., et al.* (1997). The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein. Nat. Genet. *17,* 40-48.

Ozelius, L.J., Lubarr, N., and Bressman, S.B. (2011). Milestones in dystonia. Movement Disorders *26,* 1106-1126.

Pang, K.C., Frith, M.C., and Mattick, J.S. (2006). Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. Trends in Genetics *22,* 1-5.

Peall, K.J., Waite, A.J., Blake, D.J., Owen, M.J., and Morris, H.R. (2011). Psychiatric disorders, myoclonus dystonia, and the epsilon-sarcoglycan gene: A systematic review. Movement Disorders *26,* 1939-1942.

Perronnet, C., and Vaillend, C. (2010). Dystrophins, utrophins, and associated scaffolding complexes: Role in mammalian brain and implications for therapeutic strategies. Journal of Biomedicine and Biotechnology *2010,*

Pettersson, E., Lundeberg, J., and Ahmadian, A. (2009). Generations of sequencing technologies. Genomics *93,* 105-111.

Pisani, A. (2011). Towards a new era for dystonia, a high priority for biomedical research. Neurobiol. Dis. *42,* 125-126.

Prasanth, K.V., and Spector, D.L. (2007). Eukaryotic regulatory RNAs: An answer to the 'genome complexity' conundrum. Genes and Development *21,* 11-42.

Qureshi, I.A., Mattick, J.S., and Mehler, M.F. (2010). Long non-coding RNAs in nervous system function and disease. Brain Res. *1338,* 20-35.

Rainier, S., Thomas, D., Tokarz, D., Ming, L., Bui, M., Plein, E., Zhao, X., Lemons, R., Albin, R., Delaney, C., Alvarado, D., and Fink, J.K. (2004). Myofibrillogenesis regulator 1 gene mutations cause paroxysmal dystonic choreoathetosis. Arch. Neurol. *61,* 1025-1029.

Ritz, K., Gerrits, M.C.F., Foncke, E.M.J., Van Ruissen, F., Van Der Linden, C., Vergouwen, M.D.I., Bloem, B.R., Vandenberghe, W., Crols, R., Speelman, J.D., Baas, F., and Tijssen, M.A.J. (2009). Myoclonus-dystonia: Clinical and genetic evaluation of a large cohort.

Journal of Neurology, Neurosurgery and Psychiatry *80,* 653-658.

Ritz, K., Van Schaik, B.D.C., Jakobs, M.E., Van Kampen, A.H., Aronica, E., Tijssen, M.A., and Baas, F. (2011). SGCE isoform characterization and expression in human brain: Implications for myoclonus-dystonia pathogenesis. European Journal of Human Genetics *19,* 438-444.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. *74,* 5463-5467.

Schneider, S.A., and Bhatia, K.P. (2010). Rare causes of dystonia parkinsonism. Current Neurology and Neuroscience Reports *10,* 431-439.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Molecular biology: Accurate multiplex polony sequencing of an evolved bacterial genome. Science *309,* 1728-1732.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415,* 810-813.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98,* 503-517.

Spatola, M., and Wider, C. (2012). Overview of primary monogenic dystonia. Parkinsonism and Related Disorders *18,* S158-S161.

Stamelou, M., Mencacci, N.E., Cordivari, C., Batla, A., Wood, N.W., Houlden, H., Hardy, J., and Bhatia, K.P. (2012). Myoclonus-dystonia syndrome due to tyrosine hydroxylase deficiency. Neurology *79*, 435-441.

Steinberger, D., Blau, N., Goriuonov, D., Bitsch, J., Zuker, M., Hummel, S., and Müller, U. (2004). Heterozygous mutation in 5′-untranslated region of sepiapterin reductase gene (SPR) in a patient with dopa-responsive dystonia. Neurogenetics *5,* 187-190.

Teo, J.T.H., Van De Warrenburg, B.P.C., Schneider, S.A., Rothwell, J.C., and Bhatia, K.P. (2009). Neurophysiological evidence for cerebellar dysfunction in primary focal dystonia. Journal of Neurology, Neurosurgery and Psychiatry *80,* 80-83.

Tucker, T., Marra, M., and Friedman, J.M. (2009). Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine. Am. J. Hum. Genet. *85,* 142-154.

Waite, A., Tinsley, C.L., Locke, M., and Blake, D.J. (2009). The neurobiology of the dystrophin-associated glycoprotein complex. Ann. Med. *41,* 344-359.

Weber, Y.G., Storch, A., Wuttke, T.V., Brockmann, K., Kempfle, J., Maljevic, S., Margari, L., Kamm, C., Schneider, S.A., Huber, S.M.*, et al.* (2008). GLUT1 mutations are a cause of

paroxysmal exertion-induced dyskinesias and induce hemolytic anemia by a cation leak. J. Clin. Invest. *118,* 2157-2168.

Yamagata, M., and Sanes, J.R. (2010). Synaptic localization and function of sidekick recognition molecules require MAGI scaffolding proteins. Journal of Neuroscience *30,* 3579-3588.

Yokoi, F., Dang, M.T., Li, J., and Li, Y. (2006). Myoclonus, motor deficits, alterations in emotional responses and monoamine metabolism in ε-sarcoglycan deficient mice. J. Biochem. *140,* 141-146.

Yokoi, F., Dang, M.T., Yang, G., Li, J., Doroodchi, A., Zhou, T., and Li, Y. (2012). Abnormal nuclear envelope in the cerebellar Purkinje cells and impaired motor learning in DYT11 myoclonus-dystonia mouse models. Behav. Brain Res. *227,* 12-20.

Yokoi, F., Dang, M.T., Zhou, T., and Li, Y. (2012). Abnormal nuclear envelopes in the striatum and motor deficits in DYT11 myoclonus-dystonia mouse models. Hum. Mol. Genet. *21,* 916-925.

Zhang, L., Yokoi, F., Parsons, D.S., Standaert, D.G., and Li, Y. (2012). Alteration of striatal dopaminergic neurotransmission in a mouse model of DYT11 myoclonus-dystonia. PLoS ONE *7,*

Zimprich, A., Grabowski, M., Asmus, F., Naumann, M., Berg, D., Bertram, M., Scheidtmann, K., Kern, P., Winkelmann, J., Müller-Myhsok, B.*, et al.* (2001). Mutations in the gene encoding ε-sarcoglycan cause myoclonus-dystonia syndrome. Nat. Genet. *29,* 66-69.

Appendix 1.1. Coding and splice site variant candidates in the 18p11 critical region.

Chromosome Position	Gene	Coding Exon	SNP db_xref	Genotype	Mutation Call	AA Change
18:5424367	EPB41L3			CA	IVS1066-9A>AC	
18:5443853	EPB41L3	4		delT	c.513delT	FS
18:5629216	LOC100286986	4		insT	c.941_942insT	FS
18:6908955	ARHGAP28			delT	IVS1554-4delT	
18:6912057	ARHGAP28			insT	IVS1619-3_1619-2insT	
18:6943264	LAMA1	62	rs2016639	AG	c.8982G>AG	2994D>D
18:6980523	LAMA1	42	rs607230	CT	c.6004T>CT	2002K>E
18:6985631	LAMA1	38	rs617206	TC	c.5391C>CT	1797L>L
18:6986227	LAMA1	37	rs12607841	GA	c.5288G>AG	1763A>V
18:6999628	LAMA1	32	rs625106	CT	c.4479T>CT	1493S>S
18:7011413	LAMA1	25	rs619106	AG	c.3573G>AG	1191T>T
18:7012123	LAMA1	24	rs9946794	AG	c.3378A>AG	1126G>G
18:7015714	LAMA1		rs603258	TC	IVS3126+7T>CT	
18:7023371	LAMA1	19	rs684634	AG	c.2493A>AG	831C>C
18:7033037	LAMA1	15	rs621993	AG	c.2109G>AG	703A>A

Appendix 1.2. Intronic variant candidates in the 18p11 critical region, excluding SNPs and homozygous variants. Shaded-in variants have not yet been analyzed because they do not occur in (or create) predicted splice sites.

Chr	Position	Gene	Genotype	Mutation Call
18	5144292	LOC642597	delA	IVS210+1269delA
18	5144604	LOC642597	delA	IVS210+957delA
18	5145275	LOC642597	delT	IVS210+286delT
18	5148863	LOC642597	delA	IVS17-3109delA
18	5148871	LOC642597	delA	IVS17-3117delA
18	5149854	LOC642597	delT	IVS17-4100delT
18	5149862	LOC642597	delT	IVS17-4108delT
18	5153327	LOC642597	insT	IVS17-7574_17-7573insT
18	5153353	LOC642597	CA	IVS17-7599C>AC
18	5156262	LOC642597	delA	IVS17-10508delA
18	5157132	LOC642597	delT	IVS17-11378delT
18	5157703	LOC642597	delA	IVS17-11949delA
18	5159056	LOC642597	delT	IVS17-13302delT
18	5161234	LOC642597	delA	IVS17-15480delA
18	5174748	LOC642597	delA	IVS16+22289delA
18	5175822	LOC642597	delT	IVS16+21215delT
18	5176464	LOC642597	delA	IVS16+20573delA
18	5178562	LOC642597	delA	IVS16+18475delA
18	5180742	LOC642597	delT	IVS16+16295delT
18	5180750	LOC642597	delT	IVS16+16287delT
18	5181001	LOC642597	delA	IVS16+16036delA
18	5181300	LOC642597	delA	IVS16+15737delA
18	5182990	LOC642597	delT	IVS16+14047delT
18	5183590	LOC642597	delA	IVS16+13447delA
18	5186248	LOC642597	AT	IVS16+10789A>AT
18	5188026	LOC642597	delA	IVS16+9011delA
18	5191698	LOC642597	delA	IVS16+5339delA
18	5194433	LOC642597	delA	IVS16+2604delA
18	5239996	LOC339290	delTT	2436361894_2436361895delTT
18	5240043	LOC339290	delA	2436361941delA
18	5240168	LOC339290	delT	2436362066delT
18	5240179	LOC339290	delT	2436362077delT
18	5240537	LOC339290	delT	2436362435delT
18	5240597	LOC339290	delT	2436362495delT
18	5240639	LOC339290	delT	2436362537delT
18	5243371	LOC339290	delT	2436365269delT
18	5244038	LOC339290	delT	2436365936delT
18	5244964	LOC339290	insGT	2436366861_2436366862insGT
18	5246031	LOC339290	insA	2436367928_2436367929insA

18	5292219	ZFP161	delA	IVS4-16delA
18	5300266	LOC1001326	delT	IVS627-587delT
18	5300382	LOC1001326	insT	IVS627-704_627-703insT
18	5301688	LOC1001326	delT	IVS627-2009delT
18	5303429	LOC1001326	delT	IVS627-3750delT
18	5393699	EPB41L3	delT	IVS3264+983delT
18	5398511	EPB41L3	delA	IVS2350-369delA
18	5402293	EPB41L3	delT	IVS2350-4151delT
18	5403603	EPB41L3	delAAA	IVS2349+3173_2349+3175delAAA
18	5404627	EPB41L3	delA	IVS2349+2149delA
18	5405832	EPB41L3	delA	IVS2349+944delA
18	5406261	EPB41L3	delA	IVS2349+515delA
18	5406268	EPB41L3	delA	IVS2349+508delA
18	5407164	EPB41L3	delA	IVS2158-197delA
18	5408634	EPB41L3	insG	IVS2122-900_2122-899insG
18	5410229	EPB41L3	delAA	IVS2121+336_2121+337delAA
18	5412883	EPB41L3	delAAA	IVS2068-2265_2068-2263delAAA
18	5414601	EPB41L3	delA	IVS2067+1216delA
18	5416861	EPB41L3	delAA	IVS1507-484_1507-483delAA
18	5418691	EPB41L3	delA	IVS1506+1019delA
18	5420020	EPB41L3	AG	IVS1340-144A>AG
18	5422463	EPB41L3	delTTT	IVS1339+914_1339+916delTTT
18	5429228	EPB41L3	delT	IVS913-764delT
18	5430494	EPB41L3	delA	IVS913-2030delA
18	5430578	EPB41L3	delTT	IVS913-2114_913-2113delTT
18	5438985	EPB41L3	delT	IVS530-876delT
18	5441483	EPB41L3	delG	IVS529+2354delG
18	5441995	EPB41L3	delA	IVS529+1842delA
18	5442216	EPB41L3	delT	IVS529+1621delT
18	5442781	EPB41L3	delA	IVS529+1056delA
18	5445952	EPB41L3	delA	IVS382-709delA
18	5447446	EPB41L3	insT	IVS382-2204_382-2203insT
18	5447476	EPB41L3	insT	IVS382-2234_382-2233insT
18	5447937	EPB41L3	delT	IVS382-2694delT
18	5448376	EPB41L3	delT	IVS382-3133delT
18	5449649	EPB41L3	delA	IVS382-4406delA
18	5450448	EPB41L3	delA	IVS382-5205delA
18	5450650	EPB41L3	delA	IVS382-5407delA
18	5450994	EPB41L3	delT	IVS382-5751delT
18	5452777	EPB41L3	insA	IVS382-7535_382-7534insA
18	5454199	EPB41L3	delT	IVS382-8956delT
18	5454224	EPB41L3	delTTT	IVS382-8981_382-8979delTTT
18	5455304	EPB41L3	delA	IVS382-10061delA
18	5455630	EPB41L3	delT	IVS382-10387delT
18	5457548	EPB41L3	delA	IVS382-12305delA

18	5461033	EPB41L3	delT	IVS382-15790delT
18	5462076	EPB41L3	delT	IVS381+16164delT
18	5464666	LOC1002869	delA	IVS163+1988delA
18	5465995	LOC1002869	delA	IVS163+3317delA
18	5469447	LOC1002869	delT	IVS163+6769delT
18	5469449	LOC1002869	delCCTGGCTGTG	IVS163+6771_163+6780delCCTGGCTGTG
18	5469460	LOC1002869	delGTA	IVS163+6782_163+6784delGTA
18	5473929	LOC1002869	delA	IVS163+11251delA
18	5474664	LOC1002869	delA	IVS163+11986delA
18	5474998	LOC1002869	delT	IVS163+12320delT
18	5476054	LOC1002869	delT	IVS164-12927delT
18	5476407	LOC1002869	delA	IVS164-12574delA
18	5476414	LOC1002869	delA	IVS164-12567delA
18	5480967	LOC1002869	delA	IVS164-8014delA
18	5481431	LOC1002869	insA	IVS164-7551_164-7550insA
18	5482318	LOC1002869	delA	IVS164-6663delA
18	5483631	LOC1002869	delA	IVS164-5350delA
18	5484085	LOC1002869	insA	IVS164-4897_164-4896insA
18	5484281	LOC1002869	delA	IVS164-4700delA
18	5484477	LOC1002869	delA	IVS164-4504delA
18	5484951	LOC1002869	delAA	IVS164-4030_164-4029delAA
18	5486532	LOC1002869	delA	IVS164-2449delA
18	5492850	LOC1002869	delT	IVS373+3660delT
18	5493078	LOC1002869	delA	IVS373+3888delA
18	5493406	LOC1002869	delT	IVS373+4216delT
18	5493582	LOC1002869	delA	IVS373+4392delA
18	5493621	LOC1002869	delA	IVS373+4431delA
18	5493638	LOC1002869	insA	IVS373+4447_373+4448insA
18	5494666	LOC1002869	delC	IVS373+5476delC
18	5494680	LOC1002869	delAA	IVS373+5490_373+5491delAA
18	5494685	LOC1002869	delA	IVS373+5495delA
18	5495616	LOC1002869	delA	IVS373+6426delA
18	5499606	LOC1002869	delA	IVS373+10416delA
18	5499616	LOC1002869	delA	IVS373+10426delA
18	5499778	LOC1002869	AC	IVS373+10588A>AC
18	5499780	LOC1002869	delA	IVS373+10590delA
18	5502314	LOC1002869	delA	IVS373+13124delA
18	5502355	LOC1002869	delT	IVS373+13165delT
18	5503798	LOC1002869	delT	IVS373+14608delT
18	5504289	LOC1002869	delT	IVS373+15099delT
18	5504297	LOC1002869	delT	IVS373+15107delT
18	5507740	LOC1002869	delA	IVS373+18550delA
18	5507745	LOC1002869	delA	IVS373+18555delA
18	5513385	LOC1002869	delA	IVS373+24195delA
18	5517416	LOC1002869	delA	IVS374-26492delA

18	5517447	LOC1002869	delT	IVS374-26461delT
18	5520662	LOC1002869	delA	IVS374-23246delA
18	5522940	LOC1002869	delT	IVS374-20968delT
18	5523793	LOC1002869	delA	IVS374-20115delA
18	5523797	LOC1002869	delA	IVS374-20111delA
18	5524195	LOC1002869	delT	IVS374-19713delT
18	5527023	LOC1002869	delAAA	IVS374-16885_374-16883delAAA
18	5528344	LOC1002869	delT	IVS374-15564delT
18	5529287	LOC1002869	delA	IVS374-14621delA
18	5529960	LOC1002869	delT	IVS374-13948delT
18	5531755	LOC1002869	delA	IVS374-12153delA
18	5533006	LOC1002869	delA	IVS374-10902delA
18	5536660	LOC1002869	delAA	IVS374-7248_374-7247delAA
18	5536837	LOC1002869	insA	IVS374-7072_374-7071insA
18	5537435	LOC1002869	delA	IVS374-6473delA
18	5539015	LOC1002869	delTT	IVS374-4893_374-4892delTT
18	5539272	LOC1002869	delT	IVS374-4636delT
18	5541085	LOC1002869	AC	IVS374-2823A>AC
18	5541103	LOC1002869	insA	IVS374-2806_374-2805insA
18	5541645	LOC1002869	delA	IVS374-2263delA
18	5542474	LOC1002869	delC	IVS374-1434delC
18	5542482	LOC1002869	delC	IVS374-1426delC
18	5542861	LOC1002869	delG	IVS374-1047delG
18	5567187	EPB41L3	delT	IVS762+22891delT
18	5594495	LOC1002869	delA	IVS132+1227delA
18	5603731	LOC1002869	delA	IVS132+10463delA
18	5605389	LOC1002869	delT	IVS132+12121delT
18	5605393	LOC1002869	delT	IVS132+12125delT
18	5605402	LOC1002869	delT	IVS132+12134delT
18	5605600	LOC1002869	delT	IVS132+12332delT
18	5605607	LOC1002869	delT	IVS132+12339delT
18	5610705	LOC1002869	delAA	IVS133-17356_133-17355delAA
18	5612100	LOC1002869	delAAA	IVS133-15961_133-15959delAAA
18	5619369	LOC1002869	delT	IVS133-8692delT
18	5621591	LOC1002869	delA	IVS133-6470delA
18	5622820	LOC1002869	delT	IVS133-5241delT
18	5622979	LOC1002869	insT	IVS133-5083_133-5082insT
18	5623962	LOC1002869	delA	IVS133-4099delA
18	5624017	LOC1002869	delT	IVS133-4044delT
18	5625954	LOC1002869	delT	IVS133-2107delT
18	5629601	LOC1002869	delC	IVS1045-253delC
18	5629682	LOC1002869	delG	IVS1045-172delG
18	5955849	L3MBTL4	delT	IVS1872+370delT
18	5956918	L3MBTL4	TG	IVS1705-532T>GT
18	5956921	L3MBTL4	delTA	IVS1705-535_1705-534delTA

18	5957974	L3MBTL4	delA	IVS1705-1588delA
18	5958043	L3MBTL4	delGAG	IVS1705-1657_1705-1655delGAG
18	5958106	L3MBTL4	AG	IVS1705-1720A>AG
18	5958112	L3MBTL4	AG	IVS1705-1726A>AG
18	5958280	L3MBTL4	delA	IVS1704+1813delA
18	5960023	L3MBTL4	delA	IVS1704+70delA
18	5960081	L3MBTL4	delT	IVS1704+12delT
18	5961600	L3MBTL4	delA	IVS1642-1445delA
18	5964519	L3MBTL4	delT	IVS1642-4364delT
18	5968546	L3MBTL4	delA	IVS1641+846delA
18	5972882	L3MBTL4	delAGAA	IVS1472-3321_1472-3318delAGAA
18	5974813	L3MBTL4	delA	IVS1472-5252delA
18	5974980	L3MBTL4	TA	IVS1472-5419T>AT
18	5956505	L3MBTL4	delC	IVS1705-119delC
18	5975354	L3MBTL4	delA	IVS1472-5793delA
18	5981665	L3MBTL4	delAA	IVS1472-12104_1472-12103delAA
18	5983512	L3MBTL4	delA	IVS1472-13951delA
18	5983928	L3MBTL4	delT	IVS1472-14367delT
18	5987747	L3MBTL4	delA	IVS1472-18186delA
18	5992027	L3MBTL4	delATC	IVS1472-22466_1472-22464delATC
18	5992084	L3MBTL4	delT	IVS1472-22523delT
18	5993888	L3MBTL4	delA	IVS1472-24327delA
18	5995385	L3MBTL4	delT	IVS1472-25824delT
18	6000400	L3MBTL4	delA	IVS1472-30839delA
18	6001465	L3MBTL4	CA	IVS1472-31904C>AC
18	6001629	L3MBTL4	delA	IVS1472-32068delA
18	6002164	L3MBTL4	delA	IVS1472-32603delA
18	6005339	L3MBTL4	TA	IVS1472-35778T>AT
18	6005340	L3MBTL4	GA	IVS1472-35779G>AG
18	6005341	L3MBTL4	TA	IVS1472-35780T>AT
18	6005343	L3MBTL4	TA	IVS1472-35782T>AT
18	6005344	L3MBTL4	CA	IVS1472-35783C>AC
18	6006021	L3MBTL4	delA	IVS1472-36460delA
18	6007334	L3MBTL4	insA	IVS1472-37774_1472-37773insA
18	6007372	L3MBTL4	delA	IVS1472-37811delA
18	6011324	L3MBTL4	delA	IVS1471+35403delA
18	6012664	L3MBTL4	delA	IVS1471+34063delA
18	6012679	L3MBTL4	delA	IVS1471+34048delA
18	6015983	L3MBTL4	insA	IVS1471+30743_1471+30744insA
18	6017708	L3MBTL4	delT	IVS1471+29019delT
18	6021329	L3MBTL4	delT	IVS1471+25398delT
18	6023357	L3MBTL4	delT	IVS1471+23370delT
18	6031823	L3MBTL4	delTT	IVS1471+14904_1471+14905delTT
18	6031993	L3MBTL4	delT	IVS1471+14734delT
18	6032294	L3MBTL4	delAAAA	IVS1471+14433_1471+14436delAAAA

18	6033414	L3MBTL4	delT	IVS1471+13313delT
18	6036375	L3MBTL4	delT	IVS1471+10352delT
18	6038271	L3MBTL4	delG	IVS1471+8456delG
18	6041175	L3MBTL4	delG	IVS1471+5552delG
18	6041395	L3MBTL4	delC	IVS1471+5332delC
18	6041742	L3MBTL4	delT	IVS1471+4985delT
18	6041899	L3MBTL4	delTT	IVS1471+4828_1471+4829delTT
18	6044701	L3MBTL4	delA	IVS1471+2026delA
18	6046144	L3MBTL4	delA	IVS1471+583delA
18	6047085	L3MBTL4	delA	IVS1445-332delA
18	6047275	L3MBTL4	delA	IVS1445-522delA
18	6047454	L3MBTL4	delA	IVS1445-701delA
18	6050139	L3MBTL4	delT	IVS1445-3386delT
18	6054211	L3MBTL4	insT	IVS1445-7459_1445-7458insT
18	6054221	L3MBTL4	delT	IVS1445-7468delT
18	6057423	L3MBTL4	delT	IVS1445-10670delT
18	6057559	L3MBTL4	delT	IVS1445-10806delT
18	6060458	L3MBTL4	delA	IVS1445-13705delA
18	6062040	L3MBTL4	delA	IVS1445-15287delA
18	6063336	L3MBTL4	delG	IVS1445-16583delG
18	6065040	L3MBTL4	delT	IVS1444+15840delT
18	6067724	L3MBTL4	delT	IVS1444+13156delT
18	6069442	L3MBTL4	delT	IVS1444+11438delT
18	6070418	L3MBTL4	delA	IVS1444+10462delA
18	6070525	L3MBTL4	delAA	IVS1444+10355_1444+10356delAA
18	6070665	L3MBTL4	delA	IVS1444+10215delA
18	6070672	L3MBTL4	delA	IVS1444+10208delA
18	6070687	L3MBTL4	delA	IVS1444+10193delA
18	6071638	L3MBTL4	delA	IVS1444+9242delA
18	6071675	L3MBTL4	delA	IVS1444+9205delA
18	6071712	L3MBTL4	delGAA	IVS1444+9168_1444+9170delGAA
18	6073439	L3MBTL4	delA	IVS1444+7441delA
18	6073872	L3MBTL4	delA	IVS1444+7008delA
18	6077716	L3MBTL4	insA	IVS1444+3163_1444+3164insA
18	6077733	L3MBTL4	delAA	IVS1444+3147_1444+3148delAA
18	6078676	L3MBTL4	TC	IVS1444+2204C>CT
18	6079717	L3MBTL4	delT	IVS1444+1163delT
18	6082321	L3MBTL4	delT	IVS1374-1371delT
18	6084428	L3MBTL4	delA	IVS1374-3478delA
18	6090145	L3MBTL4	delA	IVS1373+3209delA
18	6090644	L3MBTL4	delT	IVS1373+2710delT
18	6091358	L3MBTL4	delA	IVS1373+1996delA
18	6095363	L3MBTL4	delT	IVS1200-1836delT
18	6095366	L3MBTL4	insT	IVS1200-1840_1200-1839insT
18	6095679	L3MBTL4	delA	IVS1200-2152delA

18	6096079	L3MBTL4	delTT	IVS1200-2552_1200-2551delTT
18	6103881	L3MBTL4	delA	IVS1200-10354delA
18	6105209	L3MBTL4	insT	IVS1200-11683_1200-11682insT
18	6108612	L3MBTL4	delA	IVS1200-15085delA
18	6108777	L3MBTL4	insA	IVS1200-15251_1200-15250insA
18	6110476	L3MBTL4	insG	IVS1200-16950_1200-16949insG
18	6111838	L3MBTL4	delT	IVS1200-18311delT
18	6113192	L3MBTL4	delA	IVS1200-19665delA
18	6116373	L3MBTL4	delTTTT	IVS1199+21820_1199+21823delTTTT
18	6116732	L3MBTL4	delT	IVS1199+21461delT
18	6117760	L3MBTL4	delA	IVS1199+20433delA
18	6117766	L3MBTL4	CT	IVS1199+20427T>CT
18	6121652	L3MBTL4	delT	IVS1199+16541delT
18	6124380	L3MBTL4	delA	IVS1199+13813delA
18	6125106	L3MBTL4	delA	IVS1199+13087delA
18	6126145	L3MBTL4	delA	IVS1199+12048delA
18	6128187	L3MBTL4	delA	IVS1199+10006delA
18	6128443	L3MBTL4	delA	IVS1199+9750delA
18	6130697	L3MBTL4	delA	IVS1199+7496delA
18	6137255	L3MBTL4	delT	IVS1199+938delT
18	6137274	L3MBTL4	delT	IVS1199+919delT
18	6137966	L3MBTL4	GC	IVS1199+227G>CG
18	6138563	L3MBTL4	delT	IVS1097-268delT
18	6142759	L3MBTL4	delA	IVS1097-4464delA
18	6142895	L3MBTL4	delA	IVS1097-4600delA
18	6143068	L3MBTL4	delA	IVS1097-4773delA
18	6143075	L3MBTL4	delA	IVS1097-4780delA
18	6145599	L3MBTL4	delT	IVS1097-7304delT
18	6145720	L3MBTL4	delT	IVS1097-7425delT
18	6148622	L3MBTL4	delG	IVS1097-10327delG
18	6149053	L3MBTL4	delT	IVS1097-10758delT
18	6149197	L3MBTL4	delC	IVS1097-10902delC
18	6152314	L3MBTL4	delT	IVS1097-14019delT
18	6152553	L3MBTL4	delA	IVS1097-14258delA
18	6155991	L3MBTL4	delA	IVS1096+15836delA
18	6157085	L3MBTL4	delT	IVS1096+14742delT
18	6157326	L3MBTL4	delT	IVS1096+14501delT
18	6160746	L3MBTL4	delAAAA	IVS1096+11081_1096+11084delAAAA
18	6162406	L3MBTL4	delA	IVS1096+9421delA
18	6163201	L3MBTL4	delA	IVS1096+8626delA
18	6163258	L3MBTL4	GT	IVS1096+8569T>GT
18	6163272	L3MBTL4	delG	IVS1096+8555delG
18	6171279	L3MBTL4	delA	IVS1096+548delA
18	6171401	L3MBTL4	delT	IVS1096+426delT
18	6174157	L3MBTL4	delAAA	IVS982-2216_982-2214delAAA

18	6174878	L3MBTL4	delAAAA	IVS982-2937_982-2934delAAAA
18	6175038	L3MBTL4	delA	IVS982-3097delA
18	6175392	L3MBTL4	delA	IVS982-3451delA
18	6175922	L3MBTL4	delA	IVS982-3981delA
18	6175931	L3MBTL4	delA	IVS982-3990delA
18	6179171	L3MBTL4	delT	IVS982-7230delT
18	6181333	L3MBTL4	delT	IVS982-9392delT
18	6187250	L3MBTL4	delT	IVS982-15309delT
18	6187792	L3MBTL4	delT	IVS982-15851delT
18	6187907	L3MBTL4	insT	IVS982-15967_982-15966insT
18	6188000	L3MBTL4	delA	IVS982-16059delA
18	6188012	L3MBTL4	delA	IVS982-16071delA
18	6190060	L3MBTL4	delA	IVS982-18119delA
18	6191992	L3MBTL4	insC	IVS982-20052_982-20051insC
18	6191996	L3MBTL4	insA	IVS982-20056_982-20055insA
18	6192016	L3MBTL4	delA	IVS982-20075delA
18	6192050	L3MBTL4	delA	IVS982-20109delA
18	6194110	L3MBTL4	delA	IVS981+19038delA
18	6194119	L3MBTL4	delA	IVS981+19029delA
18	6196358	L3MBTL4	GA	IVS981+16790G>AG
18	6205415	L3MBTL4	delT	IVS981+7733delT
18	6205422	L3MBTL4	delT	IVS981+7726delT
18	6207943	L3MBTL4	delA	IVS981+5205delA
18	6207954	L3MBTL4	delA	IVS981+5194delA
18	6207958	L3MBTL4	delA	IVS981+5190delA
18	6214537	L3MBTL4	delA	IVS870+1212delA
18	6220056	L3MBTL4	delA	IVS785-4222delA
18	6221033	L3MBTL4	delT	IVS785-5199delT
18	6221086	L3MBTL4	delT	IVS785-5252delT
18	6222890	L3MBTL4	delA	IVS785-7056delA
18	6224112	L3MBTL4	delA	IVS785-8278delA
18	6224118	L3MBTL4	delA	IVS785-8284delA
18	6224920	L3MBTL4	delA	IVS785-9086delA
18	6225435	L3MBTL4	delA	IVS785-9601delA
18	6225464	L3MBTL4	delA	IVS785-9630delA
18	6225752	L3MBTL4	delAAAA	IVS785-9918_785-9915delAAAA
18	6225841	L3MBTL4	delA	IVS785-10007delA
18	6226831	L3MBTL4	delA	IVS785-10997delA
18	6227845	L3MBTL4	delT	IVS784+10118delT
18	6230079	L3MBTL4	delT	IVS784+7884delT
18	6232762	L3MBTL4	delA	IVS784+5201delA
18	6232812	L3MBTL4	delT	IVS784+5151delT
18	6238929	L3MBTL4	delA	IVS707+788delA
18	6239361	L3MBTL4	insA	IVS707+355_707+356insA
18	6239402	L3MBTL4	delA	IVS707+315delA

18	6245585	L3MBTL4	delT	IVS220-998delT
18	6246900	L3MBTL4	insA	IVS220-2314_220-2313insA
18	6247106	L3MBTL4	delT	IVS220-2519delT
18	6247499	L3MBTL4	insT	IVS220-2913_220-2912insT
18	6247663	L3MBTL4	delT	IVS220-3076delT
18	6249578	L3MBTL4	delA	IVS220-4991delA
18	6249699	L3MBTL4	delT	IVS220-5112delT
18	6249706	L3MBTL4	delT	IVS220-5119delT
18	6251058	L3MBTL4	delA	IVS220-6471delA
18	6252326	L3MBTL4	delA	IVS220-7739delA
18	6252420	L3MBTL4	delT	IVS220-7833delT
18	6252431	L3MBTL4	delA	IVS220-7844delA
18	6255791	L3MBTL4	delAAA	IVS219+8155_219+8157delAAA
18	6256606	L3MBTL4	delG	IVS219+7340delG
18	6257668	L3MBTL4	insT	IVS219+6277_219+6278insT
18	6263280	L3MBTL4	delAAAA	IVS219+666_219+669delAAAA
18	6263293	L3MBTL4	delAA	IVS219+653_219+654delAA
18	6264196	L3MBTL4	CA	IVS128-159A>AC
18	6264199	L3MBTL4	insG	IVS128-163_128-162insG
18	6264796	L3MBTL4	delA	IVS128-759delA
18	6266105	L3MBTL4	delA	IVS128-2068delA
18	6269462	L3MBTL4	insA	IVS128-5426_128-5425insA
18	6271604	L3MBTL4	CT	IVS128-7567T>CT
18	6272645	L3MBTL4	AG	IVS128-8608A>AG
18	6272796	L3MBTL4	CG	IVS128-8759C>CG
18	6273090	L3MBTL4	insG	IVS128-9054_128-9053insG
18	6276675	L3MBTL4	insA	IVS128-12639_128-12638insA
18	6277147	L3MBTL4	delG	IVS128-13110delG
18	6277293	L3MBTL4	delA	IVS128-13256delA
18	6278876	L3MBTL4	delA	IVS128-14839delA
18	6282929	L3MBTL4	delA	IVS128-18892delA
18	6286512	L3MBTL4	delA	IVS127+15390delA
18	6289702	L3MBTL4	delT	IVS127+12200delT
18	6204513	L3MBTL4	delA	IVS981+8635delA
18	6289926	L3MBTL4	delT	IVS127+11976delT
18	6290286	L3MBTL4	delTT	IVS127+11616_127+11617delTT
18	6290296	L3MBTL4	delT	IVS127+11606delT
18	6290523	L3MBTL4	delT	IVS127+11379delT
18	6292159	L3MBTL4	delA	IVS127+9743delA
18	6292167	L3MBTL4	delA	IVS127+9735delA
18	6293474	L3MBTL4	delA	IVS127+8428delA
18	6293552	L3MBTL4	delA	IVS127+8350delA
18	6298723	L3MBTL4	delA	IVS127+3179delA
18	6299685	L3MBTL4	delT	IVS127+2217delT
18	6300059	L3MBTL4	delT	IVS127+1843delT

18	6301018	L3MBTL4	delA	IVS127+884delA
18	6302833	L3MBTL4	delA	IVS73-877delA
18	6303121	L3MBTL4	delT	IVS73-1165delT
18	6303132	L3MBTL4	delT	IVS73-1176delT
18	6304013	L3MBTL4	insA	IVS73-2058_73-2057insA
18	6304026	L3MBTL4	delAAA	IVS73-2070_73-2068delAAA
18	6304038	L3MBTL4	delAA	IVS73-2082_73-2081delAA
18	6304176	L3MBTL4	delT	IVS73-2220delT
18	6307434	L3MBTL4	delC	IVS72+4119delC
18	6307450	L3MBTL4	delAA	IVS72+4103_72+4104delAA
18	6308659	L3MBTL4	insA	IVS72+2893_72+2894insA
18	6308761	L3MBTL4	delT	IVS72+2792delT
18	6311113	L3MBTL4	insT	IVS72+439_72+440insT
18	6313835	L3MBTL4	delA	IVS1-2211delA
18	6314082	L3MBTL4	delTAG	IVS1-2458_1-2456delTAG
18	6314351	L3MBTL4	delA	IVS1-2727delA
18	6319657	L3MBTL4	delA	IVS1-8033delA
18	6321389	L3MBTL4	delA	IVS1-9765delA
18	6322207	L3MBTL4	delA;TA	IVS1-10583delA;IVS1-10583A>AT
18	6325196	L3MBTL4	delA	IVS1-13572delA
18	6329325	L3MBTL4	delT	IVS1-17701delT
18	6331006	L3MBTL4	delA	IVS1-19382delA
18	6331953	L3MBTL4	delAAA	IVS1-20329_1-20327delAAA
18	6333594	L3MBTL4	delA	IVS1-21970delA
18	6337194	L3MBTL4	delA	IVS1-25570delA
18	6337874	L3MBTL4	delA	IVS1-26250delA
18	6338029	L3MBTL4	delA	IVS1-26405delA
18	6339596	L3MBTL4	insT	IVS1-27973_1-27972insT
18	6341870	L3MBTL4	delA	IVS1-30246delA
18	6342401	L3MBTL4	delA	IVS1-30777delA
18	6342409	L3MBTL4	delA	IVS1-30785delA
18	6342892	L3MBTL4	delA	IVS1-31268delA
18	6343675	L3MBTL4	delAA	IVS1-32051_1-32050delAA
18	6345428	L3MBTL4	delA	IVS1-33804delA
18	6346292	L3MBTL4	insT	IVS1-34669_1-34668insT
18	6346502	L3MBTL4	delA	IVS1-34878delA
18	6347740	L3MBTL4	insT	IVS1-36117_1-36116insT
18	6347869	L3MBTL4	delA	IVS1-36245delA
18	6348166	L3MBTL4	delA	IVS1-36542delA
18	6348687	L3MBTL4	delA	IVS1-37063delA
18	6348694	L3MBTL4	delA	IVS1-37070delA
18	6349355	L3MBTL4	delA	IVS1-37731delA
18	6349724	L3MBTL4	delA	IVS1-38100delA
18	6349734	L3MBTL4	delA	IVS1-38110delA
18	6349932	L3MBTL4	delA	IVS1-38308delA

18	6350551	L3MBTL4	delA	IVS1-38927delA
18	6351202	L3MBTL4	insA	IVS1-39579_1-39578insA
18	6352786	L3MBTL4	delA	IVS1-41162delA
18	6353292	L3MBTL4	delAA	IVS1-41668_1-41667delAA
18	6353795	L3MBTL4	delA	IVS1-42171delA
18	6354887	L3MBTL4	delA	IVS1-43263delA
18	6355018	L3MBTL4	insA	IVS1-43395_1-43394insA
18	6355343	L3MBTL4	delA	IVS1-43719delA
18	6355437	L3MBTL4	delT	IVS1-43813delT
18	6355853	L3MBTL4	delA	IVS1-44229delA
18	6356647	L3MBTL4	delA	IVS1-45023delA
18	6359596	L3MBTL4	delA	IVS1-47972delA
18	6360860	L3MBTL4	insTTTAAA	IVS1-49237_1-49236insTTTAAA
18	6362115	L3MBTL4	delA	IVS1-50491delA
18	6363337	L3MBTL4	insC	IVS1-51714_1-51713insC
18	6363948	L3MBTL4	delA	IVS1-52324delA
18	6367052	L3MBTL4	delG	IVS1-55428delG
18	6370151	L3MBTL4	delAA	IVS1-58527_1-58526delAA
18	6370823	L3MBTL4	delA	IVS1-59199delA
18	6371062	L3MBTL4	delT	IVS1-59438delT
18	6372811	L3MBTL4	delA	IVS1-61187delA
18	6373001	L3MBTL4	delT	IVS1-61377delT
18	6376753	L3MBTL4	delA	IVS1-65129delA
18	6377321	L3MBTL4	delT	IVS1-65697delT
18	6386235	L3MBTL4	delA	IVS1-74611delA
18	6389522	L3MBTL4	delA	IVS1-77898delA
18	6390054	L3MBTL4	delA	IVS1-78430delA
18	6390272	L3MBTL4	delT	IVS1-78648delT
18	6391033	L3MBTL4	delA	IVS1-79409delA
18	6392957	L3MBTL4	delA	IVS1-81333delA
18	6395993	L3MBTL4	delA	IVS1-84369delA
18	6396054	L3MBTL4	delA	IVS1-84430delA
18	6397691	L3MBTL4	delA	IVS1-86067delA
18	6397698	L3MBTL4	delA	IVS1-86074delA
18	6397705	L3MBTL4	delA	IVS1-86081delA
18	6398070	L3MBTL4	delT	IVS1-86446delT
18	6398772	L3MBTL4	delA	IVS1-87148delA
18	6399442	L3MBTL4	delA	IVS1-87818delA
18	6404865	L3MBTL4	delTTT	IVS1-93241_1-93239delTTT
18	6407404	L3MBTL4	delA	IVS1-95780delA
18	6408320	L3MBTL4	delA	IVS1-96696delA
18	6408684	L3MBTL4	delT	IVS1-97060delT
18	6412932	L3MBTL4	delA	IVS1-101308delA
18	6515009	LOC1001304	delT	IVS197+1284delT
18	6516365	LOC1001304	delA	IVS197+2640delA

18	6516517	LOC1001304	delA	IVS197+2792delA
18	6518103	LOC1001304	delT	IVS197+4378delT
18	6518115	LOC1001304	delT	IVS197+4390delT
18	6520455	LOC1001304	delA	IVS197+6730delA
18	6520525	LOC1001304	delT	IVS197+6800delT
18	6521359	LOC1001304	delCTCT	IVS197+7634_197+7637delCTCT
18	6521417	LOC1001304	insT	IVS197+7691_197+7692insT
18	6521622	LOC1001304	delT	IVS197+7897delT
18	6525474	LOC1001304	delA	IVS197+11749delA
18	6526200	LOC1001304	insT	IVS197+12474_197+12475insT
18	6529549	LOC1001304	delA	IVS197+15824delA
18	6530452	LOC1001304	delT	IVS197+16727delT
18	6530585	LOC1001304	delT	IVS197+16860delT
18	6531376	LOC1001304	delAA	IVS197+17651_197+17652delAA
18	6534865	LOC1001304	delA	IVS197+21140delA
18	6535080	LOC1001304	TC	IVS197+21355C>CT
18	6536187	LOC1001304	delT	IVS197+22462delT
18	6538210	LOC1001304	insA	IVS197+24484_197+24485insA
18	6538852	LOC1001304	delA	IVS197+25127delA
18	6538862	LOC1001304	delA	IVS197+25137delA
18	6835870	ARHGAP28	delA	IVS1-1478delA
18	6836874	ARHGAP28	delT	IVS1-474delT
18	6837097	ARHGAP28	delA	IVS1-251delA
18	6838219	ARHGAP28	delA	IVS66+806delA
18	6839775	ARHGAP28	delA	IVS66+2362delA
18	6840922	ARHGAP28	delA	IVS66+3509delA
18	6841208	ARHGAP28	insT;delC	IVS66+3794_66+3795insT;IVS66+3795delC
18	6841601	ARHGAP28	delT	IVS66+4188delT
18	6842192	ARHGAP28	delAT	IVS66+4779_66+4780delAT
18	6842381	ARHGAP28	delA	IVS66+4968delA
18	6842984	ARHGAP28	delA	IVS66+5571delA
18	6844768	ARHGAP28	delA	IVS67-6265delA
18	6846540	ARHGAP28	delT	IVS67-4493delT
18	6847653	ARHGAP28	GT	IVS67-3380G>GT
18	6849107	ARHGAP28	delA	IVS67-1926delA
18	6849590	ARHGAP28	delA	IVS67-1443delA
18	6849830	ARHGAP28	delT	IVS67-1203delT
18	6851680	ARHGAP28	delA	IVS159+555delA
18	6854765	ARHGAP28	delCCCACCTGCAGCT	IVS159+3640_159+3652delCCCACCTGCAGC
18	6856406	ARHGAP28	delT	IVS160-3401delT
18	6856518	ARHGAP28	delT	IVS160-3289delT
18	6858157	ARHGAP28	delT	IVS160-1650delT
18	6858880	ARHGAP28	delA	IVS160-927delA
18	6859171	ARHGAP28	delA	IVS160-636delA
18	6859430	ARHGAP28	delT	IVS160-377delT

18	6868978	ARHGAP28	delA	IVS334+745delA
18	6868984	ARHGAP28	delA	IVS334+751delA
18	6868991	ARHGAP28	delA	IVS334+758delA
18	6869251	ARHGAP28	delT	IVS334+1018delT
18	6871680	ARHGAP28	TC	IVS477+949C>CT
18	6872573	ARHGAP28	delT	IVS478-835delT
18	6873001	ARHGAP28	delA	IVS478-407delA
18	6873382	ARHGAP28	delA	IVS478-26delA
18	6873656	ARHGAP28	delTT	IVS644-27_644-26delTT
18	6878503	ARHGAP28	delAA	IVS813+2296_813+2297delAA
18	6883252	ARHGAP28	delTTT	IVS976+954_976+956delTTT
18	6884387	ARHGAP28	delA	IVS976+2089delA
18	6888008	ARHGAP28	TC	IVS1059+770C>CT
18	6888651	ARHGAP28	delA	IVS1060-1236delA
18	6889275	ARHGAP28	delT	IVS1060-612delT
18	6892554	ARHGAP28	delT	IVS1371+2012delT
18	6893881	ARHGAP28	delTT	IVS1372-953_1372-952delTT
18	6898297	ARHGAP28	delT	IVS1553+1672delT
18	6899021	ARHGAP28	delA	IVS1553+2396delA
18	6899038	ARHGAP28	delA	IVS1553+2413delA
18	6899038	ARHGAP28	delA	IVS1553+2413delA
18	6899067	ARHGAP28	delA	IVS1553+2442delA
18	6900648	ARHGAP28	delA	IVS1553+4023delA
18	6901553	ARHGAP28	delA	IVS1553+4928delA
18	6901555	ARHGAP28	delA	IVS1553+4930delA
18	6903847	ARHGAP28	delA	IVS1554-5112delA
18	6904066	ARHGAP28	delA	IVS1554-4893delA
18	6904423	ARHGAP28	delA	IVS1554-4536delA
18	6905098	ARHGAP28	delA	IVS1554-3861delA
18	6905112	ARHGAP28	delA	IVS1554-3847delA
18	6905239	ARHGAP28	delT	IVS1554-3720delT
18	6907405	ARHGAP28	delAA	IVS1554-1554_1554-1553delAA
18	6907408	ARHGAP28	delA	IVS1554-1551delA
18	6907913	ARHGAP28	delA	IVS1554-1046delA
18	6910848	ARHGAP28	delTTT	IVS1619-1211_1619-1209delTTT
18	6910861	ARHGAP28	delT	IVS1619-1198delT
18	6911398	ARHGAP28	delGCTAATATA	IVS1619-661_1619-653delGCTAATATA
18	6911454	ARHGAP28	insT	IVS1619-606_1619-605insT
18	6911873	ARHGAP28	delA	IVS1619-186delA
18	6914984	ARHGAP28	delA	IVS1713+2831delA
18	6942335	LAMA1	delT	IVS9068-97delT
18	6942429	LAMA1	delTT	IVS9068-191_9068-190delTT
18	6946143	LAMA1	delA	IVS8844+1019delA
18	6946853	LAMA1	delA	IVS8844+309delA
18	6946995	LAMA1	delA	IVS8844+167delA

18	6948634	LAMA1	delCA	IVS8557-79_8557-78delCA
18	6956273	LAMA1	delTT	IVS8094+362_8094+363delTT
18	6957889	LAMA1	GC	IVS7964+587C>CG
18	6960674	LAMA1	delA	IVS7626+911delA
18	6960716	LAMA1	GA	IVS7626+869G>AG
18	6962263	LAMA1	delA	IVS7338-205delA
18	6962270	LAMA1	delA	IVS7338-212delA
18	6962433	LAMA1	delG	IVS7338-375delG
18	6962440	LAMA1	delG	IVS7338-382delG
18	6963072	LAMA1	delTT	IVS7338-1014_7338-1013delTT
18	6970623	LAMA1	delT	IVS6899+1233delT
18	6971226	LAMA1	delG	IVS6899+630delG
18	6972697	LAMA1	delT	IVS6774+359delT
18	6972719	LAMA1	insT	IVS6774+336_6774+337insT
18	6973500	LAMA1	delA	IVS6624-294delA
18	6973951	LAMA1	delT	IVS6624-745delT
18	6975856	LAMA1	delT	IVS6489+80delT
18	6976779	LAMA1	delT	IVS6346-700delT
18	6976788	LAMA1	delT	IVS6346-709delT
18	6979189	LAMA1	delA	IVS6008-812delA
18	6979194	LAMA1	delA	IVS6008-817delA
18	6979206	LAMA1	delA	IVS6008-829delA
18	6979596	LAMA1	delA	IVS6007+924delA
18	6981120	LAMA1	delA	IVS5891-484delA
18	6981689	LAMA1	delA	IVS5890+807delA
18	6984011	LAMA1	delA	IVS5661-778delA
18	6986932	LAMA1	delT	IVS5169-586delT
18	6987493	LAMA1	delT	IVS5169-1147delT
18	6989278	LAMA1	delTT	IVS5169-2932_5169-2931delTT
18	6990035	LAMA1	delT	IVS5168+2525delT
18	6991305	LAMA1	delA	IVS5168+1255delA
18	6991387	LAMA1	delT	IVS5168+1173delT
18	6991408	LAMA1	delG	IVS5168+1152delG
18	6992223	LAMA1	delT	IVS5168+337delT
18	6993439	LAMA1	delA	IVS5008+201delA
18	6994697	LAMA1	delCA	IVS4896+659_4896+660delCA
18	6995573	LAMA1	TA	IVS4807-128A>AT
18	6995581	LAMA1	delA	IVS4807-136delA
18	6996119	LAMA1	delA	IVS4807-674delA
18	6996138	LAMA1	delT	IVS4807-693delT
18	6999034	LAMA1	delT	IVS4663+410delT
18	6999040	LAMA1	delT	IVS4663+404delT
18	7003266	LAMA1	delT	IVS4261-882delT
18	7003268	LAMA1	delG	IVS4261-884delG
18	7003878	LAMA1	delT	IVS4261-1494delT

18	7004368	LAMA1	delT	IVS4261-1984delT
18	7007489	LAMA1	delA	IVS4123-214delA
18	7014575	LAMA1	delG	IVS3127-525delG
18	7014577	LAMA1	delTACAGTGAGCTTTG	IVS3127-527_3127-
18	7014641	LAMA1	TA	IVS3127-591T>AT
18	7014643	LAMA1	AT	IVS3127-593A>AT
18	7015020	LAMA1	delT	IVS3126+701delT
18	7015461	LAMA1	insT	IVS3126+259_3126+260insT
18	7015608	LAMA1	delT	IVS3126+113delT
18	7016065	LAMA1	insG	IVS2990-209_2990-208insG
18	7016077	LAMA1	insG	IVS2990-221_2990-220insG
18	7019697	LAMA1	delT	IVS2702-2314delT
18	7019700	LAMA1	insT	IVS2702-2318_2702-2317insT
18	7024017	LAMA1	delT	IVS2489+362delT
18	7026846	LAMA1	delA	IVS2275-741delA
18	7027341	LAMA1	insT	IVS2275-1237_2275-1236insT
18	7027503	LAMA1	delT	IVS2275-1398delT
18	7027588	LAMA1	delA	IVS2275-1483delA
18	7027958	LAMA1	delA	IVS2275-1853delA
18	7028171	LAMA1	delA	IVS2275-2066delA
18	7028176	LAMA1	delA	IVS2275-2071delA
18	7028719	LAMA1	delC	IVS2275-2614delC
18	7029988	LAMA1	insA	IVS2274+2076_2274+2077insA
18	7030448	LAMA1	delA	IVS2274+1617delA
18	7031541	LAMA1	delA	IVS2274+524delA
18	7031888	LAMA1	delA	IVS2274+177delA
18	7034938	LAMA1	delA	IVS1840-249delA
18	7035441	LAMA1	insT	IVS1839+544_1839+545insT
18	7038791	LAMA1	insCGC	IVS1563+17_1563+18insCGC
18	7039879	LAMA1	delA	IVS1422+196delA
18	7042904	LAMA1	delA	IVS1155+322delA
18	7044161	LAMA1	insA;delC	IVS976+559_976+560insA;IVS976+560delC
18	7044163	LAMA1	delAA	IVS976+558_976+559delAA
18	7044184	LAMA1	delA	IVS976+537delA
18	7045061	LAMA1	CT	IVS859-223C>CT
18	7047043	LAMA1	delT	IVS769-677delT
18	7047832	LAMA1	delA	IVS768+1245delA
18	7053521	LAMA1	delA	IVS346-2586delA
18	7053566	LAMA1	delA	IVS346-2631delA
18	7054085	LAMA1	delT	IVS346-3150delT
18	7054703	LAMA1	delT	IVS346-3768delT
18	7056078	LAMA1	AC	IVS346-5143A>AC
18	7056089	LAMA1	delAA	IVS346-5154_346-5153delAA
18	7056888	LAMA1	delT	IVS346-5953delT
18	7057981	LAMA1	delTT	IVS346-7046_346-7045delTT

18	7064061	LAMA1	TA	IVS346-13126T>AT
18	7569184	PTPRM	delT	IVS73+1295delT
18	7570123	PTPRM	insA	IVS73+2233_73+2234insA
18	7570253	PTPRM	delT	IVS73+2364delT
18	7570952	PTPRM	delTT	IVS73+3063_73+3064delTT
18	7571253	PTPRM	delT	IVS73+3364delT
18	7575036	PTPRM	insA	IVS73+7146_73+7147insA
18	7576484	PTPRM	delT	IVS73+8595delT
18	7577089	PTPRM	delT	IVS73+9200delT
18	7578329	PTPRM	insT	IVS73+10439_73+10440insT
18	7581366	PTPRM	delA	IVS73+13477delA
18	7581653	PTPRM	insT	IVS73+13763_73+13764insT
18	7581669	PTPRM	delTT	IVS73+13780_73+13781delTT
18	7581672	PTPRM	delG	IVS73+13783delG
18	7581973	PTPRM	delT	IVS73+14084delT
18	7585271	PTPRM	delT	IVS73+17382delT
18	7585623	PTPRM	delT	IVS73+17734delT
18	7585626	PTPRM	TA	IVS73+17737T>AT
18	7588587	PTPRM	insT	IVS73+20697_73+20698insT
18	7589026	PTPRM	delT	IVS73+21137delT
18	7589560	PTPRM	delC	IVS73+21671delC
18	7589563	PTPRM	delC	IVS73+21674delC
18	7594494	PTPRM	delA	IVS73+26605delA
18	7594512	PTPRM	delA	IVS73+26623delA
18	7598702	PTPRM	delT	IVS73+30813delT
18	7598803	PTPRM	insT	IVS73+30913_73+30914insT
18	7599172	PTPRM	delA	IVS73+31283delA
18	7602441	PTPRM	delTT	IVS73+34552_73+34553delTT
18	7602658	PTPRM	delT	IVS73+34769delT
18	7606517	PTPRM	delT	IVS73+38628delT
18	7607419	PTPRM	delT	IVS73+39530delT
18	7607421	PTPRM	delAAG	IVS73+39532_73+39534delAAG
18	7607422	PTPRM	-;GG	-;IVS73+39533A>G
18	7607423	PTPRM	-;TT	-;IVS73+39534G>T
18	7607425	PTPRM	delAG	IVS73+39536_73+39537delAG
18	7607451	PTPRM	delCTAG	IVS73+39562_73+39565delCTAG
18	7609542	PTPRM	insT	IVS73+41652_73+41653insT
18	7613684	PTPRM	delA	IVS73+45795delA
18	7615257	PTPRM	delA	IVS73+47368delA
18	7618919	PTPRM	delT	IVS73+51030delT
18	7622253	PTPRM	delT	IVS73+54364delT
18	7624501	PTPRM	delT	IVS73+56612delT
18	7625550	PTPRM	delT	IVS73+57661delT
18	7628626	PTPRM	delT	IVS73+60737delT
18	7631307	PTPRM	delT	IVS73+63418delT

18	7631461	PTPRM	delTT	IVS73+63572_73+63573delTT
18	7634240	PTPRM	delT	IVS73+66351delT
18	7634990	PTPRM	delT	IVS73+67101delT
18	7636995	PTPRM	insC	IVS73+69105_73+69106insC
18	7637181	PTPRM	delAAA	IVS73+69292_73+69294delAAA
18	7642617	PTPRM	delT	IVS73+74728delT
18	7642844	PTPRM	insA	IVS73+74954_73+74955insA
18	7642864	PTPRM	delA	IVS73+74975delA
18	7643330	PTPRM	delT	IVS73+75441delT
18	7643680	PTPRM	delT	IVS73+75791delT
18	7644083	PTPRM	insA	IVS73+76193_73+76194insA
18	7649765	PTPRM	insA	IVS73+81875_73+81876insA
18	7650455	PTPRM	delC	IVS73+82566delC
18	7652925	PTPRM	delA	IVS73+85036delA
18	7659172	PTPRM	delT	IVS73+91283delT
18	7660185	PTPRM	delA	IVS73+92296delA
18	7660975	PTPRM	insA	IVS73+93085_73+93086insA
18	7660988	PTPRM	delA	IVS73+93099delA
18	7661933	PTPRM	delT	IVS73+94044delT
18	7662614	PTPRM	delA	IVS73+94725delA
18	7662873	PTPRM	delA	IVS73+94984delA
18	7667757	PTPRM	delT	IVS73+99868delT
18	7667948	PTPRM	delT	IVS73+100059delT
18	7668074	PTPRM	delT	IVS73+100185delT
18	7669737	PTPRM	delT	IVS73+101848delT
18	7672306	PTPRM	delT	IVS74-101841delT
18	7673726	PTPRM	delA	IVS74-100421delA
18	7674174	PTPRM	delA	IVS74-99973delA
18	7675750	PTPRM	delT	IVS74-98397delT
18	7676231	PTPRM	delA	IVS74-97916delA
18	7676321	PTPRM	delA	IVS74-97826delA
18	7681145	PTPRM	delT	IVS74-93002delT
18	7681580	PTPRM	insA	IVS74-92568_74-92567insA
18	7681593	PTPRM	delAA	IVS74-92554_74-92553delAA
18	7681601	PTPRM	delA	IVS74-92546delA
18	7681612	PTPRM	delT	IVS74-92535delT
18	7683334	PTPRM	delA	IVS74-90813delA
18	7683344	PTPRM	delA	IVS74-90803delA
18	7683684	PTPRM	delT	IVS74-90463delT
18	7686579	PTPRM	delA	IVS74-87568delA
18	7687798	PTPRM	insT	IVS74-86350_74-86349insT
18	7689806	PTPRM	delT	IVS74-84341delT
18	7691025	PTPRM	delT	IVS74-83122delT
18	7691122	PTPRM	delT	IVS74-83025delT
18	7691767	PTPRM	delA	IVS74-82380delA

18	7691776	PTPRM	delA	IVS74-82371delA
18	7691784	PTPRM	delA	IVS74-82363delA
18	7691876	PTPRM	delG	IVS74-82271delG
18	7694427	PTPRM	delT	IVS74-79720delT
18	7694446	PTPRM	insT	IVS74-79702_74-79701insT
18	7699778	PTPRM	delT	IVS74-74369delT
18	7706438	PTPRM	delA	IVS74-67709delA
18	7706632	PTPRM	insA	IVS74-67516_74-67515insA
18	7708736	PTPRM	delT	IVS74-65411delT
18	7710566	PTPRM	delA	IVS74-63581delA
18	7718063	PTPRM	delA	IVS74-56084delA
18	7718883	PTPRM	delA	IVS74-55264delA
18	7719549	PTPRM	delA	IVS74-54598delA
18	7719556	PTPRM	delA	IVS74-54591delA
18	7719563	PTPRM	delA	IVS74-54584delA
18	7720712	PTPRM	AC	IVS74-53435A>AC
18	7722484	PTPRM	delA	IVS74-51663delA
18	7723007	PTPRM	delT	IVS74-51140delT
18	7724083	PTPRM	delTT	IVS74-50064_74-50063delTT
18	7729790	PTPRM	delA	IVS74-44357delA
18	7735588	PTPRM	delA	IVS74-38559delA
18	7735601	PTPRM	delA	IVS74-38546delA
18	7736337	PTPRM	insT	IVS74-37811_74-37810insT
18	7738967	PTPRM	delT	IVS74-35180delT
18	7741583	PTPRM	delT	IVS74-32564delT
18	7746469	PTPRM	delT	IVS74-27678delT
18	7747169	PTPRM	delT	IVS74-26978delT
18	7749331	PTPRM	delA	IVS74-24816delA
18	7749358	PTPRM	delT	IVS74-24789delT
18	7750960	PTPRM	delC	IVS74-23187delC
18	7750963	PTPRM	delC	IVS74-23184delC
18	7752357	PTPRM	delT	IVS74-21790delT
18	7752595	PTPRM	delT	IVS74-21552delT
18	7752604	PTPRM	delT	IVS74-21543delT
18	7752998	PTPRM	delG	IVS74-21149delG
18	7753854	PTPRM	delT	IVS74-20293delT
18	7755244	PTPRM	delAA	IVS74-18903_74-18902delAA
18	7760389	PTPRM	delT	IVS74-13758delT
18	7761574	PTPRM	delT	IVS74-12573delT
18	7761583	PTPRM	delT	IVS74-12564delT
18	7765550	PTPRM	delT	IVS74-8597delT
18	7766892	PTPRM	delT	IVS74-7255delT
18	7767391	PTPRM	delA	IVS74-6756delA
18	7768329	PTPRM	delA	IVS74-5818delA
18	7768340	PTPRM	delA	IVS74-5807delA

18	7769576	PTPRM	delT	IVS74-4571delT
18	7772232	PTPRM	insA	IVS74-1916_74-1915insA
18	7772349	PTPRM	delCT	IVS74-1798_74-1797delCT
18	7772521	PTPRM	insT	IVS74-1627_74-1626insT
18	7773110	PTPRM	delT	IVS74-1037delT
18	7773591	PTPRM	insT	IVS74-557_74-556insT
18	7781811	PTPRM	delT	IVS196+7542delT
18	7782125	PTPRM	delA	IVS196+7856delA
18	7783747	PTPRM	delT	IVS196+9478delT
18	7788967	PTPRM	delT	IVS196+14698delT
18	7790918	PTPRM	delA	IVS196+16649delA
18	7791246	PTPRM	delT	IVS196+16977delT
18	7792373	PTPRM	insC	IVS196+18103_196+18104insC
18	7792966	PTPRM	delT	IVS196+18697delT
18	7793804	PTPRM	CT	IVS196+19535C>CT
18	7795481	PTPRM	delAA	IVS196+21212_196+21213delAA
18	7795796	PTPRM	insT	IVS196+21526_196+21527insT
18	7795830	PTPRM	insT	IVS196+21560_196+21561insT
18	7798537	PTPRM	delAA	IVS196+24268_196+24269delAA

Appendix 1.3. Structural variant candidates within the 18p11 critical region.

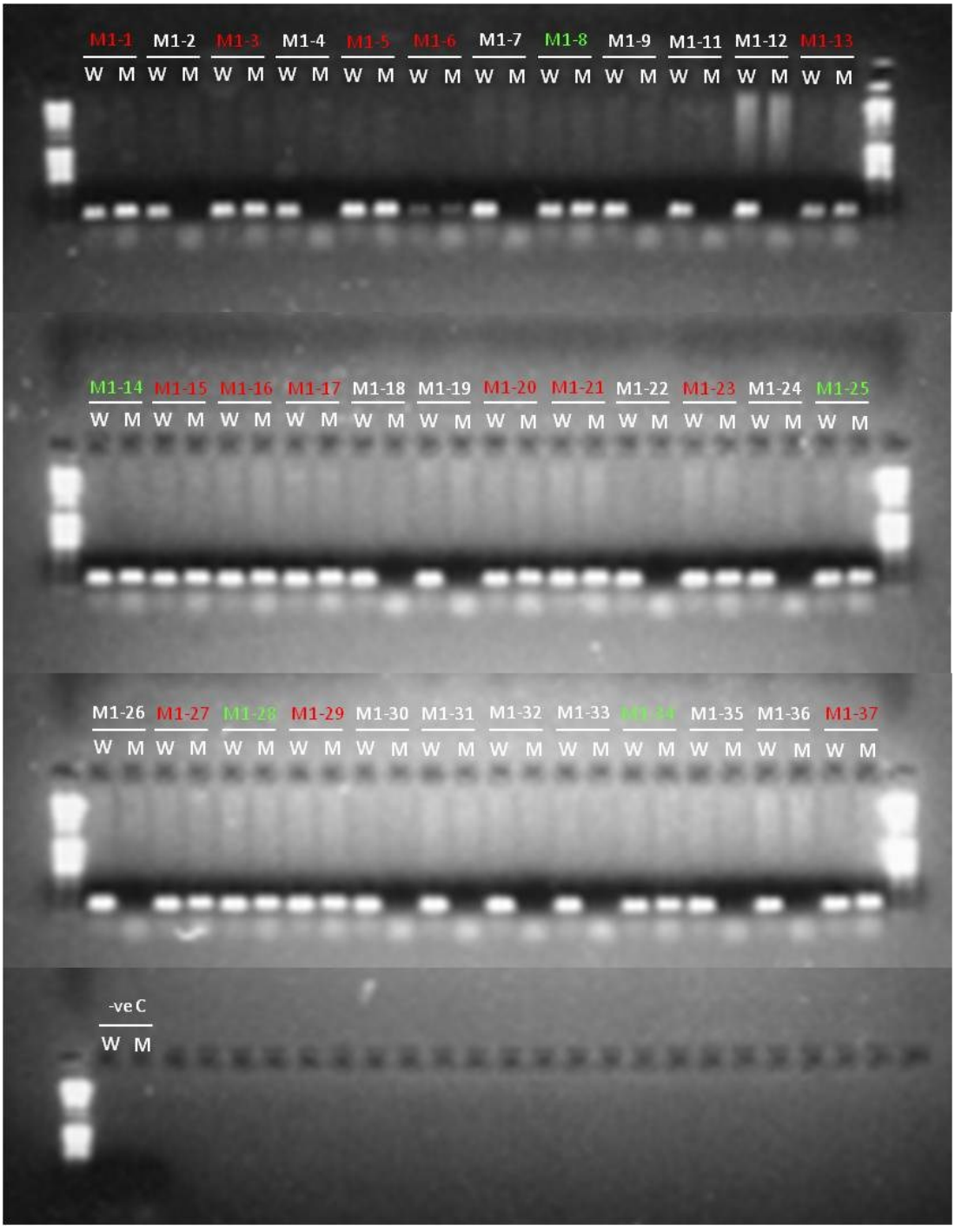
Identifier	Chr Position	Description
SV-Mut1	18:4578525-4579175	Section of reads with what appears to be breakpoints on either end, as well as a deletion within the aligned region
SV-Mut2	18:4652047-4652082	36 bp deletion
SV-Mut3	18:4628036-4628084	49 bp deletion
SV-Mut4	18:4723165-4723203	39 bp deletion
SV-Mut5	18:4760670-4760679	10 bp deletion
SV-Mut6	18:4813790-4814490	700 bp deletion
SV-Mut7	18:4874440-4874840	huge amount of coverage with very clear breakpoint, likely an insertion
SV-Mut8	18:4950225-4950360	repetitive region, misaligned sequence flanking ACACA repeat
SV-Mut9	18:5020998-5021035	38 bp deletion
SV-Mut10	18:5036820-5036965	clear breakpoint
SV-Mut11	18:5085903-5085926	24 bp deletion
SV-Mut12	18:5160860-5161040	small deletion flanked by misaligned sequence
SV-Mut13	18:5162653-5162664	12 bp deletion in a repetitive region
SV-Mut14	18:5324640-5326240	~1600 bp deletion
SV-Mut15	18:5395000-5395280	misaligned reads through exon 20 of EPB41L3
SV-Mut16	18:5406675-5406950	misaligned reads through exon 15 of EPB41L3
SV-Mut17	18:5457080-5457240	~100 bp deletion
SV-Mut18	18:5469445-5469480	small deletion within a repetitive region
SV-Mut19	18:5511760-5511820	60 bp deletion
SV-Mut20	18:5535555-5535645	breakpoint?
SV-Mut21	18:5555410-5555795	clear breakpoint, possible insertion?
SV-Mut22	18:5668280-5668720	highly repetitive region, a lot of misaligned sequence
SV-Mut23	18:5700666-5700692	26 bp deletion
SV-Mut24	18:5911245-5911380	breakpoint?
SV-Mut25	18:5934555-5934598	43 bp deletion
SV-Mut26	18:5959620-5959800	breakpoint?
SV-Mut27	18:5959960-5960120	misaligned sequence
SV-Mut28	18:5972805-5972980	highly repetitive region, a lot of misaligned sequence
SV-Mut29	18:6024480-6024580	breakpoint? Occurs within an intron of L3MBTL4
SV-Mut30	18:6094684-6094745	55 bp deletion. Occurs within an intron of L3MBTL4
SV-Mut31	18:6118155-6118275	repeat expansion?
SV-Mut32	18:6139100-6139300	misalignment that may correspond with a deletion
SV-Mut33	18:6199680-6199860	several reads with misalignment
SV-Mut34	18:6206460-6206580	breakpoint?
SV-Mut35	18:6214120-6214280	clear breakpoint. Occurs within an intron of L3MBTL4. ~570 bp deletion.
SV-Mut36	18:6516480-6516615	possible breakpoint within a highly repetitive region. Occurs within an intron of LOC10013480.

SV-Mut37	18:6541564-6541626	62 bp deletion
SV-Mut38	18:6543950	29 bp insertion. Occurs within an intron of LOC100130480.
SV-Mut39	18:6544170-6544500	region with high coverage, a lot of misalignment
SV-Mut40	18:6569480-6569680	breakpoint?
SV-Mut41	18:6725744-6725768	25 bp deletion
SV-Mut42	18:6729200-6729640	highly repetitive region, a lot of misaligned sequence
SV-Mut43	18:6876480-6876540	~60 bp deletion
SV-Mut44	18:6879450-6879810	potential breakpoints?
SV-Mut45	18:6920240-6920440	highly repetitive region, likely an insertion
SV-Mut46	18:6953164-6953203	31 bp deletion
SV-Mut47	18:6972780-6973000	breakpoint?
SV-Mut48	18:7014574-7014591	18 bp deletion
SV-Mut49	18:7172522	17 bp insertion
SV-Mut50	18:7419637-7419670	34 bp deletion
SV-Mut51	18:7485398-7485414	17 bp deletion
SV-Mut52	18:7497125-7497400	highly repetitive region, a lot of misaligned sequence
SV-Mut53	18:7620330-7620480	region of misalignment
SV-Mut54	18:7686160-7686380	breakpoint?
SV-Mut55	18:7692383-7692416	34 bp deletion

Appendix 1.4. Expressed sequence tag (ESTs) variant candidates within the 18p11 critical region.

Chr	Position	EST ID	Genotype	Mutation Call
18	5288157	CD108131	insACC	IVS1351+2699_1351+2700insACC
18	7118268	AI864567	TA	IVS1-549T>AT
18	7244243	BF695255	TC	IVS1-126524T>CT

Appendix 1.5. Products from allele-specific PCR on all members of Myoclonus Family 1 (MF1), testing for the ACC duplication in the CD108131 EST. The 161 bp product is specific to the wild-type (W) allele and the 163 bp product is specific to the mutant (M) allele. Given their proximity in size, the wild-type and mutant PCR products were run in separate lanes on the agarose gel. Individuals in red font are affected with MD, individuals in green font are carriers of MD disease haplotype, but are unaffected.



Appendix 2.1. Nucleotide sequence of the CD108131 EST, obtained from www.genome.ucsc.edu. The genomic sequence is 838 bp in size, from chr18:5287450-5288289.

cDNA CD108131

```
AGTGAGGTTG ACACTAACAA ATTCACCCTA TTTTGTTAGG AAAAAAAAGA  50
CAAGATGAGG GAAAATGATG CAAATTTATG TAGTTTCCAA CTTAATGGGT  100
TATTTTGTGT AGTTATCTGT AAGTCATTTG GAAGGTGGGA TTCATTGTCA  150
CATAGTAATG CTATTAGTGG TGTTTTGTTC TCAGGATTAG CAAGCCTTAG  200
GGTGCAATTG TAGTAGAGCT GGCACATAGT AAGGATTGTG GTGCAGGGTT  250
GAGTCCTAGG CGGAGGCAGA GAACTGTTAC AGCAAAGAGA ACGGgAAGCT  300
CTCACTCCCT TTGAAGTCTG TCTCCATTTT TGCTTACCCT CTTGTAACTG  350
GAGCTCTTTC CCCCTGGGAA CTATTTATAA TAGGGTTAAT CCTGCATCTA  400
CTTAAGGTAT GTAAAGATAG CATGTTGCCT ACTTGTGTTG AGCATCATAT  450
GCGAGGCACT GGGCTGGACC CTGGAGATGC AAAGGCAGAA GCAGTTTGGT  500
CCCTGGCTGT TAGAGTGAGT GTGCCCCTTC CCAGTCTGGT CTTACATGTT  550
gTTTTTGAAA ATGTTTTTCC ATCACCTGTT ATCCATGATT CCTTAAAGGT  600
TGCCATACTG ATCTCATCTT AGGGGCCCTG GGTGTAATTT ATCTGGATTG  650
TGACAGATTT TAGCAGTTGG GTGTTTTCAT GATCTCTTCA AAAGATGTTC  700
CCACCTACCA ATACTCATTT AAATCTTTTC AGTATAACGA GGAaGTGAGG  750
tTTTAGTGGA AAGACTTTAC AAGACAAACT TtAGATTTGG AaTCCTgGGT  800
TTTGCCctcT CcCAGTAACT GCAAATTtCA GACnAagtTT CTTAAaCTGC  850
Cctttttagc cca
```

Genomic chr18 (reverse strand):

```
tgaggaagag aaatttagga agtttaaatg taaagtaatt tgccggccat  5288340
taacttgatc actaatttag aattgttaga aaatacagag gtcgccatca  5288290
AGTGAGGTTG ACACTAACAA ATTCACCCTA TTTTGTTAGG AAAAAAAAGA  5288240
CAAGATGAGG GAAAATGATG CAAATTTATG TAGTTTCCAA CTTAATGGGT  5288190
TATTTTGTGT AGTTATCTGT AAGTCATTTG GAAGGTGGGA TTCATTGTCA  5288140
CATAGTAATG CTATTAGTGG TGTTTTGTTC TCAGGATTAG CAAGCCTTAG  5288090
GGTGCAATTG TAGTAGAGCT GGCACATAGT AAGGATTGTG GTGCAGGGTT  5288040
GAGTCCTAGG CGGAGGCAGA GAACTGTTAC AGCAAAGAGA ACGGaAAGCT  5287990
CTCACTCCCT TTGAAGTCTG TCTCCATTTT TGCTTACCCT CTTGTAACTG  5287940
GAGCTCTTTC CCCCTGGGAA CTATTTATAA TAGGGTTAAT CCTGCATCTA  5287890
CTTAAGGTAT GTAAAGATAG CATGTTGCCT ACTTGTGTTG AGCATCATAT  5287840
GCGAGGCACT GGGCTGGACC CTGGAGATGC AAAGGCAGAA GCAGTTTGGT  5287790
CCCTGGCTGT TAGAGTGAGT GTGCCCCTTC CCAGTCTGGT CTTACATGTT  5287740
cTTTTTGAAA ATGTTTTTCC ATCACCTGTT ATCCATGATT CCTTAAAGGT  5287690
TGCCATACTG ATCTCATCTT AGGGGCCCTG GGTGTAATTT ATCTGGATTG  5287640
TGACAGATTT TAGCAGTTGG GTGTTTTCAT GATCTCTTCA AAAGATGTTC  5287590
CCACCTACCA ATACTCATTT AAATCTTTTC AGTATAACGA GGAgGTGAGG  5287540
TTTAGTGGAA AGACTTTACA AGACAAACTT AGATTTGGAT CCTGGTTTTG  5287490
CCTCtCAGTA ACTGCAAATT CAGACaAgTT CTTAACTGCC ttcagcctga  5287440
gtttcaccat ctctaaaata ctggtatttc tgcctcacag cctgctgagg  5287390
attaatgtga tggccaggaa aaaacattaa ccccagcatc
```

Appendix 2.2. Open reading frame prediction of the CD108131 EST using the "Translate Tool" in the ExPASy Bioinformatics Resource Portal (http://web.expasy.org/translate/).

5'3' Frame 1

S E V D T N K F T L F C **Stop** E K K D K **Met** R E N D A N L C S F Q L N G L F C V V I C K S F G R W D S L S H S N A I S G V L F S G L A S L R V Q L **Stop** **Stop** S W H I V R I V V Q G **Stop** V L G G G R E L L Q Q R E R E A L T P F E V C L H F C L P S C N W S S F P L G T I Y N R V N P A S T **Stop** G **Met** **Stop** R **Stop** H V A Y L C **Stop** A S Y A R H W A G P W R C K G R S S L V P G C **Stop** S E C A P S Q S G L T C C F **Stop** K C F S I T C Y P **Stop** F L K G C H T D L I L G A L G V I Y L D C D R F **Stop** Q L G V F **Met** I S S K D V P T Y Q Y S F K S F Q Y N E E V R F **Stop** W K D F T R Q T L D L E S W V L P S P S N C K F Q X K F L K L P F L A

5'3' Frame 2

V R L T L T N S P Y F V R K K K T R **Stop** G K **Met** **Met** Q I Y V V S N L **Met** G Y F V **Stop** L S V S H L E G G I H C H I V **Met** L L V V F C S Q D **Stop** Q A L G C N C S R A G T **Stop** **Stop** G L W C R V E S **Stop** A E A E N C Y S K E N G K L S L P L K S V S I F A Y P L V T G A L S P W E L F I I G L I L H L L K V C K D S **Met** L P T C V E H H **Met** R G T G L D P G D A K A E A V W S L A V R V S V P L P S L V L H V V F E N V F P S P V I H D S L K V A I L I S S **Stop** G P W V **Stop** F I W I V T D F S S W V F S **Stop** S L Q K **Met** F P P T N T H L N L F S I T R K **Stop** G F S G K T L Q D K L **Stop** I W N P G F C P L P V T A N F R X S F L N C P F **Stop** P

5'3' Frame 3

Stop G **Stop** H **Stop** Q I H P I L L G K K R Q D E G K **Stop** C K F **Met** **Stop** F P T **Stop** W V I L C S Y L **Stop** V I W K V G F I V T **Stop** **Stop** C Y **Stop** W C F V L R I S K P **Stop** G A I V V E L A H S K D C G A G L S P R R R Q R T V T A K R T G S S H S L **Stop** S L S P F L L T L L **Stop** L E L F P P G N Y L **Stop** **Stop** G **Stop** S C I Y L R Y V K I A C C L L V L S I I C E A L G W T L E **Met** Q R Q K Q F G P W L L E **Stop** V C P F P V W S Y **Met** L F L K **Met** F F H H L L S **Met** I P **Stop** R L P Y **Stop** S H L R G P G C N L S G L **Stop** Q I L A V G C F H D L F K R C S H L P I L I **Stop** I F S V **Stop** R G S E V L V E R L Y K T N F R F G I L G F A L S Q **Stop** L Q I S D X V S **Stop** T A L F S P

3'5' Frame 1

W A K K G S L R N X V **Stop** N L Q L L G E G K T Q D S K S K V C L V K S F H **Stop** N L T S S L Y **Stop** K D L N E Y W **Stop** V G T S F E E I **Met** K T P N C **Stop** N L S Q S R **Stop** I T P R A P K **Met** R S V W Q P L R N H G **Stop** Q V **Met** E K H F Q K Q H V R P D W E G A H S L **Stop** Q P G T K L L L P L H L Q G P A Q C L A Y D A Q H K **Stop** A T C Y L Y I P **Stop** V D A G L T L L **Stop** I V P R G K E L Q L Q E G K Q K W R Q T S K G V R A S R S L C C N S S L P P P R T Q P C T T I L T **Met** C Q L Y Y N C T L R L A N P E N K T P L I A L L C D N E S H L P N D L Q I T T Q N N P L S W K L H K F A S F S L I L S F F S **Stop** Q N R V N L L V S T S

3'5' Frame 2

G L K R A V **Stop** E T X S E I C S Y W E R A K P R I P N L K F V L **Stop** S L S T K T S L P R Y T E K I **Stop** **Met** S I G R W E H L L K R S **Stop** K H P T A K I C H N P D K L H P G P L R **Stop** D Q Y G N L **Stop** G I **Met** D N R **Stop** W K N I F K N N **Met** **Stop** D Q T G K G H T H S N S Q G P N C F C L C I S R V Q P S A S H **Met** **Met** L N T S R Q H A I F T Y L K **Stop** **Met** Q D **Stop** P Y Y K **Stop** F P G G K S S S Y K R V S K N G D R L Q R E **Stop** E L P V L F A V T V L C L R L G L N P A P Q S L L C A S S T T I A P **Stop** G L L I L R T K H H **Stop** **Stop** H Y Y V T **Met** N P T F Q **Met** T Y R **Stop** L H K I T H **Stop** V G N Y I N L H H F P S S C L F F P N K I G **Stop** I C **Stop** C Q P H

3'5' Frame 3

G **Stop** K G Q F K K L X L K F A V T G R G Q N P G F Q I **Stop** S L S C K V F P L K P H F L V I L K R F K **Stop** V L V G G N I F **Stop** R D H E N T Q L L K S V T I Q I N Y T Q G P **Stop** D E I S **Met** A T F K E S W I T G D G K T F S K T T C K T R L G R G T L T L T A R D Q T A S A F A S P G S S P V P R I **Stop** C S T Q V G N **Met** L S L H T L S R C R I N P I I N S S Q G E R A P V T R G **Stop** A K **Met** E T D F K G S E S F P F S L L **Stop** Q F S A S A **Stop** D S T L H H N P Y Y V P A L L Q L H P K A C **Stop** S **Stop** E Q N T T N S I T **Met** **Stop** Q **Stop** I P P S K **Stop** L T D N Y T K **Stop** P I K L E T T **Stop** I C I I F P H L V F F F L T K **Stop** G E F V S V N L T

CURRICULUM VITAE

EDUCATIONAL BACKGROUND

- **Bachelor of Science Honours in Biology, Minor in Psychology** **2010**
  Carleton University, Ottawa, Ontario

- **Student of the Faculty of Science, Major in Pre-Medicine** **2005-2007**
  University of Regina, Regina, Saskatchewan

ACADEMIC HIGHLIGHTS AND ACHIEVEMENTS

- Awarded a poster prize at the 4th annual Brain Health Research Day **2012**
- Awarded 1st place at the annual Department of Biochemistry, Microbiology and Immunology Graduate Poster Day **2012**
- Awarded 3rd place at the annual Department of Biochemistry, Microbiology and Immunology Graduate Seminar Day **2012**
- Recipient of an Ontario Graduate Scholarship **2011/2012**
- Awarded a poster prize at the 3rd annual Brain Health Research Day **2011**
- Attended the International Congress of Human Genetics Conference **2011**
- Recipient of the Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarships Master's Award **2010/2011**
- Awarded the University of Ottawa National Excellence Scholarship **2010-2012**
- Achieved Bachelor of Science with Highest Honours **2010**
- Recipient of the NSERC Undergraduate Student Research Award **2009-2010**
- Awarded the Gerhard Herzberg Scholarship (Carleton University) **2008-2010**
- Awarded the General In-Course Scholarship (Carleton University) **2007-2008**
- Appointed to the Deans' Honour List (Carleton University) **2007-2010**
- Awarded the Centennial Merit Plus Scholarship (University of Regina) **2005**
- Awarded the Dorothy Boyle Memorial Scholarship (University of Regina) **2005**
- Awarded the University of Regina Alumni Association Children of Alumni Entrance Scholarship **2005**
- Awarded the Saskatchewan General Proficiency Award (Government of Saskatchewan) **2005**
- Awarded the Governor General's Academic Medal (Bronze, secondary school level) **2005**

VOLUNTEER EXPERIENCE

- Let's Talk Science Program at the University of Ottawa **2010/2011**
- University of Regina Ambassador, participant in the UR Connected event to welcome new students **2006**