Improving Diagnostic Detection Strategies of Choroideremia

by

Rachel Mah

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Medical Sciences – Medical Genetics
University of Alberta

© Rachel Mah, 2018

Abstract

Choroideremia is an X-linked monogenic inherited retinal disease. It affects males starting in their teenage years with night blindness followed by progressive vision loss starting in the peripherals and ending with total vision loss late in life. It is estimated that 1 in 50,000 individuals are affected by this genetic disorder. Female carriers are usually asymptomatic although some may present with an uneven degenerative fundus. With clinical gene therapy trials underway in locations across the world, the molecular diagnosis of choroideremia is increasingly important. In order for a patient to be enrolled in a gene therapy trial, an unequivocal diagnosis of choroideremia is a prerequisite. In the past, research labs performed immunoblot analysis to demonstrate a lack of Rab Escort Protein-1 (REP-1) in patients phenotypically suggestive of choroideremia. Now, exon sequencing of the causative gene, *CHM*, is a more cost effective and reliable diagnostic tool. However, the sensitivity of exon sequencing is not 100% and occasionally individuals will lack expression of the REP-1 protein, but present no identifiable defect in the coding sequence of the gene. It is important to investigate such patients and examine for further alterations. In this study, we discovered a novel splice mutation, c.1245-521A>G, in two unrelated individuals. In addition, we report, the first inversion mutation in *CHM*, c.-839_49+5528inv. A fourth patient was determined to be a CHM phenocopy; he bore no mutations in the *CHM* gene and was eventually found to be REP-1 positive. He instead underwent whole exome sequencing (WES) with both his parents to determine candidate genes for his ocular choroideremia-like phenotype. In conclusion, novel non-coding mutations were determined to be responsible for choroideremia observed in three individuals. Further work is required on the possible pathogenicity of the candidate genes identified in the fourth patient in order to determine the variant responsible for his unique ocular

phenotype. These findings suggest the importance of sequencing non-coding regions of the *CHM* gene in order to improve diagnostic sensitivity and provide molecular diagnoses to patients looking for enrolment in a gene therapy trial.

Preface

This thesis contains original work by Rachel Mah. No part of this work has been previously published.

Acknowledgments

I would like to begin by thanking my supervisor, Dr. Stacey Hume, for her excellent guidance throughout my project. Whenever I hit a problem or roadblock in my research she was always there with a handful of solutions I could try. I greatly appreciate the amount of resources she allowed me to have access to through the Molecular Diagnostics Laboratory, allowing me to see and conduct experiments that most graduate students do not. Thank you too all the staff of MDL for accommodating my diagnostic needs, abundant questions and student aliquots, specifically Christine Walker and Kathleen Sprysak for their patience.

Thank you to my supervisor, Dr. Ian MacDonald and members of his clinic and lab. I am incredibly grateful for the opportunities you have provided for me both in the lab and outside. Being so directly linked to a clinical setting and volunteering at Foundation Fighting Blindness events to see firsthand where my research is significant is something I never thought I would have the opportunity for in grad school. A massive thank you is extended to Alina Radziwon for her taking me under her wing and teaching me everything I needed to learn in the lab, as well as her insight when results did not make sense. Thank you to Lance Doucette for solving my western blot woes and allowing me to collaborate with him to gain bioinformatic experience.

Thank you to my final committee member, Dr. Elena Posse de Chavez for her expertise and collaboration throughout my project. Thank you to her lab members, Jennifer Ling and Sarah Samuelson for their support and input during my research. As well as Dr. Yves Sauvé and Dr. Heather McDermid for reading my thesis and acting as my examining committee for my defence.

I would also like to thank the Department of Medical Genetics for accepting me into the program and providing support and experience throughout my degree. Thank you to Dr. Michael

Walter for answering any questions or concerns I was facing and for sharing stories of Northern Ontario when I was feeling particularly far away from home. Thank you to our graduate coordinator, Sarah Hughes for keeping my supervisors and myself on track to graduate. Thank you to all the students in the department for making my hours spent here more enjoyable and for providing a listening ear on the harder days. Special thanks to Matthea Sanderson for always being supportive, best of luck with your PhD and Ashley Russell for her honest advice, I know we will remain in touch and you will make an amazing nurse.

Finally I would like to thank my family and friends, without them I would not be where I am today. Thank you to my mother, Sonya Mah, for instilling in me the importance of education and the desire to help others. Thank you for answering my tearful phone calls and providing the best possible advice. Thank you to my father, Cory Mah, for all the support a daughter could need and teaching me how to do adult things while out here on my own. Thank you to my friends back home in Ontario for answering my random phone calls, meeting me at airports and always finding a reason to celebrate whenever I came home. Special thanks to my best friend, Jill Heron, for always being there, never judging and keeping me motivated, no matter what.

Last but not least, thank you to all of the wonderful people I've met during my time in Edmonton. Thank you for making it feel like a second place to call home and showing me all the wonderful parts of the city.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAV	**Adeno-associated virus**
ACMG	**American College of Medical Genetics and Genomics**
BCA	**Bicinchoninic acid assay**
cDNA	**Complementary deoxyribonucleic acid**
CHM	**Choroideremia**
COL18A1	**Collagen type XVII alpha 1 chain**
CRB1	**Crumbs homolog 1**
CRD	**Cone-rod dystrophy**
DMEM	**Dulbecco's Modified Eagle's Medium**
DNA	**Deoxyribonucleic acid**
dNTPs	**Deoxyribonucleotide triphosphate**
EJC	**Exon junction complex**
EORCD	**Early onset rod-cone dystrophy**
ESE	**Exonic splicing enhancer**
ESS	**Exonic splicing silencer**
FBS	**Fetal bovine serum**

FISH	**Fluorescent *in situ* hybridization**
GDF	**GDI-displacement factor**
gDNA	**Genomic DNA**
GDP	**Guanosine diphosphate**
GDI	**GDP-dissociation inhibitor**
GTP	**Guanosine triphosphate**
HGVS	**Human Genome Variation Society**
HID	**Human identification**
HSF	**Human Splice Finder**
iPSCs	**Induced pluripotent stem cells**
IRD	**Inherited retinal disease**
LCA	**Leber Congenital Amaurosis**
MAF	**Minor allele frequency**
mRNA	**Messenger ribonucleic acid**
MLPA	**Multiplex ligation-dependent probe amplification**
NGS	**Next generation sequencing**
NMD	**Nonsense-mediated decay**

NNSplice	**Neural Network Splice**
PBS	**Phosphate-buffered saline**
PCR	**Polymerase chain reaction**
PTCs	**Premature termination codons**
PROM1	**Prominin-1**
PVDF	**Polyvinylidene difluoride**
RabGDI	**Rab GDP-dissociation inhibitor**
Rab-GGTase	**Rab geranylgeranyl transferase**
REP-1	**Rab escort protein-1**
REP-2	**Rab escort protein-2**
RNA	**Ribonucleic acid**
RP	**Retinitis pigmentosa**
RPE	**Retinal pigment epithelium**
SIFT	**Sorting Intolerant From Tolerant**
SNARE	**SNAp receptor**
snRNPs	**Small nuclear ribonucleoproteins**
SSF	**Splice Site Finder**

STR	**Short tandem repeat**
TBS	**Tris-buffered saline**
TBS-T	**Tris-buffered saline plus Tween**
WES	**Whole exome sequencing**
VCF	**Variant Call Format**
WGS	**Whole genome sequencing**

CHAPTER 1 INTRODUCTION

1.1 Inherited Retinal Diseases

Inherited retinal diseases (IRDs) are a group of visual impairment conditions caused by the dysfunction and/or degeneration of the neural retina or retinal pigment epithelium (Carss et al., 2017). All present with a clinical phenotype but many exhibit a high degree of phenotypic, genetic and allelic heterogeneity. More than 250 genes and loci are associated with this group of diseases and can be inherited in an autosomal dominant, autosomal recessive, X-linked, or mitochondrial manner (Carss et al., 2017). IRDs are among the many rare diseases that are increasingly becoming investigated with high-throughput sequencing.

1.2 Choroideremia

Choroideremia (CHM) is an X-linked, monogenic disorder (MacDonald, Sereda, McTaggart, & Mah, 2004). It is characterized by progressive degeneration of the retinal pigment epithelium (RPE), photoreceptors and choroid (Dimopoulos, Radziwon, St Laurent, & MacDonald, 2017). Peripheral degeneration begins in teenage years and progresses until mid-life when patients become completely blind (Seabra, Brown, & Goldstein, 1993). Female carriers are often asymptomatic, although some present with a patchy degeneration of the fundus. Choroideremia is a rare genetic disease, occurring in 1 in 50,000 individuals, accounting for approximately 4% of blindness worldwide (Coussa & Traboulsi, 2012). Current treatments include surgery for complications such as retinal detachment and cataracts, UV blocking sunglasses, low vision services and genetic counselling.

1.3 *CHM* gene

The *CHM* gene encodes for Rab escort protein-1 and pathogenic variants in this gene lead to choroideremia. *CHM* is located on the X chromosome and spans 186,382 bp, while the mRNA is 5,442 bp and comprised of 15 exons (van Bokhoven et al., 1994). Exons 1-14 are all less than 400 bp long and exon 15 is 3,642 bp (van Bokhoven et al., 1994). The relatively small reading frame is one of the reasons this gene is ideal for gene therapy (discussed below). The gene produces a 653 amino acid protein that has a molecular weight of 95 kDa (van Bokhoven et al., 1994). Many different types of mutations have been reported as disease-causing in this gene including: small deletions, nonsense mutations, missense mutations, frameshifts, splice site defects, retrotransposon insertions and copy number variants (Chi, MacDonald, & Hume, 2013; van den Hurk et al., 2003).

1.4 Endocytic Vesicular Trafficking

Movement of different molecules and particles from one cellular component to another is accomplished by intracellular vesicular transport. Cellular compartments are encircled by intracellular membranes, which bud off to form vesicles carrying the molecules that need to be transported elsewhere within the cell. There are 3 steps to vesicular transport; budding, targeting/docking and fusion (Alory & Balch, 2003). Budding begins from a donor membrane to which cargo and targeting molecules are recruited (Bonifacino & Glick, 2004). Targeting or docking is the step where the vesicle becomes attached to its acceptor membrane (Bonifacino & Glick, 2004). Finally, fusion includes SNAp REceptor (SNARE) proteins and other factors, which bridge the vesicular membrane and acceptor membranes together (Bonifacino & Glick, 2004). Although incomplete, the current understanding is that Rab GTPases are required to

recruit tethering factors and SNARE proteins on both the vesicle membrane and acceptor membrane to facilitate the docking step of this process (Bonifacino & Glick, 2004). With more than 60 Rab proteins identified, they can be developmentally and spatially regulated within higher eukaryotes (Alory & Balch, 2003). In order for these Rab proteins to attach to cell membranes, they must be prenylated, a process facilitated by REP-1 or REP-2 (Alory & Balch, 2003). The Rab protein family is comprised of small GTPases that regulate vesicular membrane trafficking, which are involved in many inherited retinal disorders (Hutagalung & Novick, 2011; Roosing, *et al*., 2014) (Figure 1.1). Loss of REP-1 function leads to disruption of proper Rab protein intracellular trafficking and progressive degeneration of ocular structures (Dimopoulos et al., 2017).

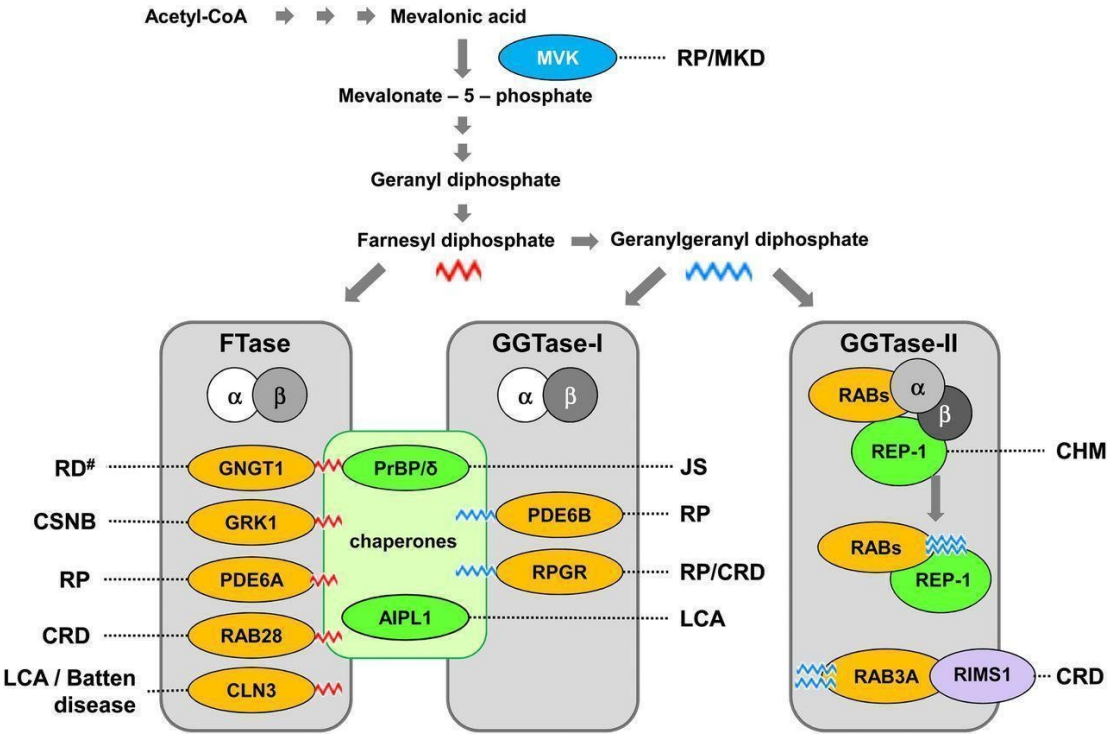


Figure 1.1. **Prenylated proteins in ocular disorders.** Schematic representation of proteins that are prenylated in various inherited retinal disorders. Some ocular disorders listed here include retinitis pigmentosa (RP), retinal detachment (RD), congenital stationary night blindness (CSNB), cone rod dystrophy (CRD), Leber congenital amaurosis (LCA), Joubert syndrome (JS).

Copyright © 2014 Susanne Roosing, Rob W J Collin, Anneke I den Hollander, Frans P M Cremers, Anna M Siemiatkowska, British Medical Journal.

1.5 Rab/REP-1 Protein Pathway

To function as vesicular transport chaperones, Rab proteins must undergo prenylation, a post-translational modification consisting of the addition of geranylgeranyl lipid groups on the carboxyl terminus of the Rab protein (Preising & Ayuso, 2004; Hutagalung & Novick, 2011) (Figure 1.2). This prenylation step allows the Rab proteins to be linked to their target membranes through the added geranylgeranyl moiety (Preising & Ayuso, 2004). Rab proteins are found in either a cytosolic, inactive GDP-bound state or in an active GTP-bound state (Preising & Ayuso, 2004). Rab Escort Protein-1 (REP-1) functions as an accessory protein that delivers the Rab proteins to the prenyltransferase, geranylgeranyltransferase (RabGGTase II) (Alexandrov, Horiuchi, Steele-Mortimer, Seabra, & Zerial, 1994). REP-1 binds newly synthesized Rab proteins, and presents them to the RabGGTase for the prenylation reaction. Once prenylated, REP-1 removes the protein from the catalytic site of the transferase and finally, delivers them to an acceptor membrane (Andres et al., 1993).

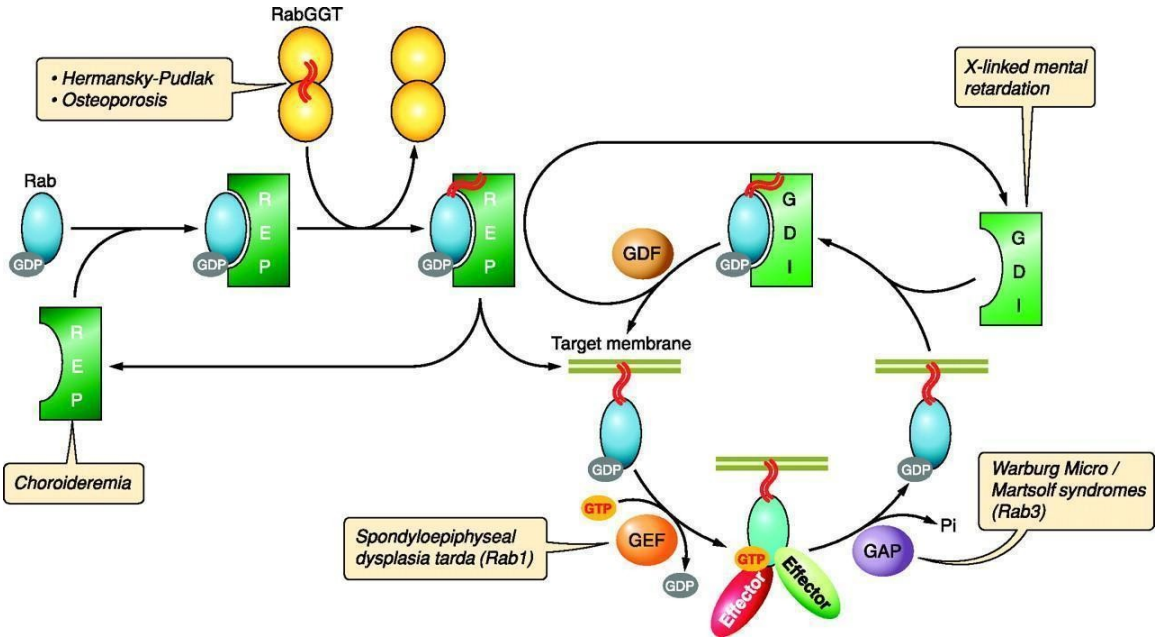


Figure 1.2. Rab cycle. Schematic of the proteins involved in the Rab cycle.

Copyright © 2011, Hutagalung & Novick, The American Physiological Society

1.6 Rab Escort Protein 2

Both mammalian REP-1 and REP-2 proteins are ubiquitously distributed but REP-1 is more highly expressed than REP-2 in the eye (Alory & Balch, 2003). This likely explains why CHM is not a disorder resulting in widespread cellular death in all tissues (Cremers et al., 1994). It has been determined that REP-1 and REP-2 bind with nearly the same affinity to both Rab1a and Rab5a (Anant et al., 1998). This data, along with other similar findings suggests that REP-1 and REP-2 are functionally redundant and aid in the prenylation of many different Rab proteins. Importantly, REP-2 can still allow for low geranylgeranylation of Rab27 in the absence of REP-1 (Seabra et al., 1995). Kohnke, et al. (2013) suggest a prenylation hierarchy exists *in vivo* in which some Rab proteins have a slower prenylation rate and are therefore, more vulnerable to underprenylation. Rab27 and Rab38 are some of the Rab proteins with the slowest rate of GTP hydrolysis (Brooks et al., 2007). Currently, there are Rab proteins which are described as melanosome specific; Rab27a, Rab27b and Rab38 (Hutagalung & Novick, 2011). Since melanosome synthesis in the RPE occurs before birth and no maintenance of melanosomes occurs in the RPE, malfunctioning of the prenylation of these Rab proteins in CHM patients leads to cellular apoptosis (Kohnke et al., 2013). In rescue experiments conducted by the same researchers, REP-1 was able to reverse the underprenylation of these Rab proteins but REP-2 could not, suggesting that REP-2 may not be able to prenylate specific Rab proteins with a high enough efficiency to reverse underprenylation in the absence of REP-1 (Kohnke et al., 2013). The only symptom in choroideremia patients is blindness, therefore, REP-2 is thought to be able to compensate in other tissues. This may be due to the tissue specific composition of Rab proteins or the levels of REP expression, or some combination of both instances and requires further investigation (Kohnke et al., 2013).

1.7 Rab27 and Rab38

A lack of REP-1 protein in choroideremia patients causes only an ocular phenotype. Both REP proteins chaperone Rab proteins to RabGTPases for prenylation, which is essential for their function (Preising & Ayuso, 2004). In choroideremia patient lymphoblasts, there is an accumulation of unprenylated Rab proteins that would normally be preferentially prenylated by REP-1 (Seabra, Ho, & Anant, 1995). REP-1 has been demonstrated to interact specifically with over 60 different Rab proteins (Pereira-Leal, Strom, Godfrey, & Seabra, 2003). Importantly, Seabra and colleagues determined that two particular Rab proteins, Rab27 and Rab38, are prenylated *in vitro* more efficiently by REP-1 than REP-2 and are identified in an unprenylated state in choroideremia patient cells (Seabra et al., 1995). Rab27 is found in many tissues but notably it was demonstrated to be expressed in the retinal pigment epithelium and choriocapillaris, cell layers of the eye that degenerate earliest in choroideremia patients (Seabra et al., 1995). Mutations to the Rab38 protein in mouse models demonstrate a slow *in vivo* prenylation rate and cause ocular hypopigmentation and thinning of the RPE (Brooks et al., 2007; Lopes, Wasmeier, Seabra, & Futter, 2007). It is suggested that Rab38 remains in an unprenylated and non-functional state in choroideremia patients (Kohnke et al., 2013).

1.8 RabGDI

It was previously suggested that a cellular factor would be required to dissociate a Rab protein from REP-1 in order for the Rab protein to be recycled in the prenylation reaction. Rab GDP-dissociation inhibitor (RabGDI) was suggested to be the cellular factor required for this process (Andres et al., 1993). REP-1 and RabGDI are similar in both function and structure.

RabGDI is a member of the Rab GTPase functional cycle (Alory & Balch, 2001). RabGDI forms a complex with prenylated GDP-bound Rab proteins in the cytosol (Alory & Balch, 2001). GDI family members work to mediate both the delivery of Rab proteins to various membranes during vesicular formation as well as retrieve them following vesicular fusion (Alory & Balch, 2001). RabGDI works by forming a stable complex with Rab proteins in an inactive, GDP-bound state (Matsui et al., 1990). This allows Rab proteins to stay in the aqueous cytosolic environment preventing improper membrane binding (Matsui et al., 1990). For the attachment of Rab proteins to their target membrane, activation is required which involves dissociating from RabGDI and exchanging a GDP for a GTP molecule (Dirac-Svejstrup, Sumizawa, & Pfeffer, 1997). A protein, GDI-displacement factor (GDF), accomplishes this by releasing prenylated Rab proteins from RabGDI (Dirac-Svejstrup et al., 1997). Through further investigation, Alexandrov et al. (1994) determined that REP-1 delivers prenylated Rab proteins to their target compartments with the dissociation process occuring in the cytosol as a RabGDI independent event.

1.9 Gene therapy

Technological advances in gene delivery have allowed for rapid evolution of gene therapies. With the development of several safe gene delivery techniques, gene therapy is now realizing its potential as a medicine (Wang & Gao, 2014). Current gene delivery methods are categorized into two classes: DNA (non-viral) vectors and viral vectors. Non-viral techniques involve plasmid DNA entering cells either with or without chemicals to enhance stability and efficiency (Wang & Gao, 2014). Non-viral gene expression cassettes are typically composed of a promoter, the transgene of interest and a termination signal. The cassette is embedded within a

circular, double-stranded plasmid for delivery and this can then be directly injected *in vivo* (Wang & Gao, 2014). Viral vectors use the natural infectious characteristics of certain viruses yet by removing as many viral genes as possible, it allows them to carry the gene of choice. Viral genes are removed for safety purposes but all information that allows the virus to insert itself into the genome of its host are left intact. Viruses are naturally evolved gene delivery shuttles and their surface proteins interact with target cell receptors to trigger endocytosis (Wang & Gao, 2014). The adeno-associated virus (AAV) is the current choice for gene therapy in choroideremia due to its success in treating Leber congenital amaurosis and due to it's ease of delivery as a it is a rub-retinal injection. AAVs fall into a group of small, simple, helper-dependent, non-pathogenic, single-stranded DNA viruses (Wang & Gao, 2014). Some other of using AAV advantages include it's ability to transduce both dividing and non-dividing cells and the discovery that some serotypes can cross the blood-brain barrier (Duque et al., 2009). Although not an obstacle for the *CHM* gene, a major drawback of AAV vectors is that they can only carry a gene that is no longer than approximately 4.5 kb.

1.10 Gene therapy in ophthalmology

Choroideremia became a candidate for gene therapy trials following the success of gene therapy in treating Leber Congenital Amaurosis (Simonelli et al., 2010). Choroideremia is considered a good disease candidate for multiple reasons: it is monogenic, it has a slow progression which extends the treatment window, the gene has a small coding region and a diagnosis is relatively straightforward with a clearly identifiable phenotype and an available clinical genetics test (Tolmachova et al., 2013). Clinical CHM trials using AAV-mediated gene therapy are underway at multiple sites including the University of Alberta (NCT02077361), the

University of Pennsylvania (NCT02341807), the University of Miami (NCT02553135) and the University of Oxford (NCT02407678) (http://curechm.org/research/clinical-trials/?lang=can). AAV carrying *CHM* cDNA has been injected *in vitro* into human patient lymphoblasts and induced pluripotent stem cells (iPSCs) to determine its ability to restore normal REP-1 function (Vasireddy et al., 2013). The safety has also been evaluated in normal-sighted mice and no toxicity has been found (Vasireddy et al., 2013). Clinical trials have moved forward to phase 1 and 2 in human choroideremia patients and initial reports have shown an improved rod and cone function in the treated eye (MacLaren et al., 2014). In order to be included in a gene therapy trial, patients must have a definitive molecular diagnosis of choroideremia.

1.11 Pathogenic Variants in the *CHM* gene

Numerous pathogenic variants have been described in the *CHM* gene, the most frequent being nonsense or frameshift mutations leading to premature truncation (Freund, Sergeev, & MacDonald, 2016; Simunovic et al., 2016). In a large cohort study of 128 affected males, the most common mutations found were nonsense mutations (41%), exon deletions (37%) and splice site mutations (14%) which all were associated with a loss of functional protein (Freund et al., 2016). The remaining missense mutations were predicted to result in severe changes to protein structure and folding with effects similar to that of null mutations (Freund et al., 2016). These missense mutations represented 4% of the cohort population and were further investigated for phenotype correlation. Freund *et al.* determined that given the spectrum of mutations documented in *CHM*, no genotype-phenotype correlation could be made with respect to age of onset of visual symptoms, visual acuity or width of visual fields. Therefore, no preferential treatment for inclusion in a gene replacement therapy trial should be given based on genotype.

1.12 Splicing

The majority of pathogenic variants characterized in the *CHM* gene are nonsense mutations leading to premature stop codons or changes in splice acceptor or donor sequences causing splice defects. Splicing is the process of editing non-coding sequences, called introns, out of the pre-mRNA sequence (da Costa, Menezes, & Romao, 2017). The final product of the splicing process is mature mRNA, containing only the coding regions, called exons (da Costa et al., 2017). This process occurs co-transcriptionally and is catalyzed by spliceosomes (da Costa et al., 2017). A spliceosome is an RNA-protein complex that is comprised of five small nuclear ribonucleoproteins (snRNPs) and over 180 splicing protein factors (da Costa et al., 2017). Two of these snRNPs are required for recognizing a very specific region of the pre-mRNA, the splice site (da Costa et al., 2017). U1 and U2 recognize the 5' splice site and splice branches, forming the pre-spliceosomal A complex (da Costa et al., 2017) (Figure 1.3). This defines the exon and intron boundaries, which are then bound to by U4/U5-U6 snRNPs to form the pre-spliceosomal B complex (Sanford & Caceres, 2004). This complex is then converted to the C complex through a series of RNA-RNA and RNA-protein rearrangements (Sanford & Caceres, 2004). The complex completes the splicing process and results in two exons ligated together and removal of the intervening intron (da Costa et al., 2017). There are still many associated proteins awaiting characterization and the mechanism of splice site identification is not fully understood (da Costa et al., 2017). Ward & Cooper (2010) estimated that up to 50% of disease-causing mutations affect splicing.

Da Costa et al., 2017 also proposed that 95% of coding and non-coding regions of the genome are affected by alternative splicing. Alternate splicing allows specific genes to produce multiple mRNA transcripts from the same locus, leading to protein diversity (da Costa et al.,

2017). This is accomplished through exon skipping, varying use of 5' or 3' splice sites and intron inclusion (da Costa et al., 2017). *CHM* does have a shorter, protein coding alternative transcript of only 110 amino acids but is lacking any experimental confirmation to date about function or location (Alexandrov et al., 1994).

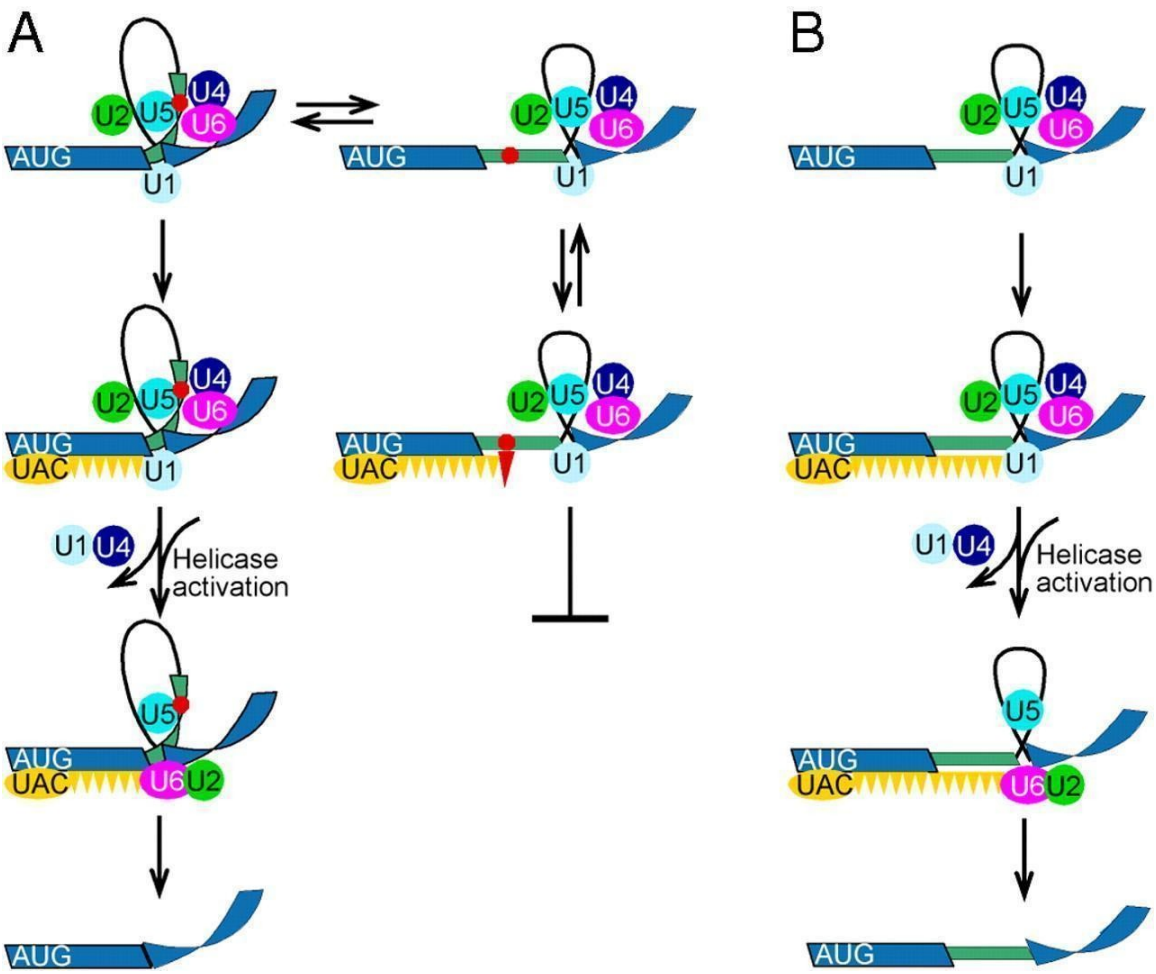


Figure 1.3. Splicing machinery. A schematic of the splicing process. A) Splicing occurring at the genuine 5' splice site (left) and splicing occurring at the latent (comply with consensus splice sites but are not used normally for splicing) 5' splice site (right). B) Splicing at the latent 5' splice site after removal of the in-frame stop codon.

Copyright © 2010, National Academy of Sciences, Kamhi, Raitskin, Sperling, & Sperling

1.13 X-inactivation

Designing an animal model for choroideremia has been a complicated task. Previous mouse models showed that null mutations in *Chm* were embryonically lethal in males and when the mutant allele was inherited from a carrier female, lethal in female embryos as well (van den Hurk et al., 1997). Shi, W. et al. (2004) conclude this lethality is due to defects in trophoblast development and vascularization. To avoid this, conditional knockout models using a Cre/loxP approach was used to spatially and temporally control the gene knockout (Tolmachova et al., 2006). The inability to create male mouse models requires the consideration of X-inactivation in the female heterozygous mice. It is widely accepted that during early female development, most genes on one X chromosome are transcriptionally silenced. However, some X-linked genes have been determined to "escape" X-inactivation and both copies are expressed. Generally, it is assumed that the X-inactivation status of a gene is constant among different females in a particular species (Disteche, 1995). In a study by (Carrel & Willard, 1999) determined that REP-1 showed both monoallelic expression and biallelic expression in different cell lines. This research shows a third expression pattern for some X-linked genes, heterogenous X-inactivation. The *CHM* gene is subject to X-inactivation on some X chromosomes and escapes inactivation to different extents on others; this effect is suggested to be the reason for varied clinical manifestation in female carriers of choroideremia (Carrel & Willard, 1999). X-inactivation of cells can also be random and create a mosaic of *Chm* wildtype and *Chm* knockout cells in mouse models (Tolmachova et al., 2006). Both of these reasons complicate the functional testing of *Chm* knockout cells or mice.

1.14 Rationale and Hypothesis

In the current study, we examined REP-1 absent choroideremia patients who lacked a molecular diagnosis. Patients with a clinical diagnosis of choroideremia but who lacked a molecular diagnosis through the sequencing of coding regions will be verified to lack the REP-1 protein by immunoblot analysis. Screening only the coding regions of the gene could miss non-coding mutations in some REP-1 absent patients. Furthermore, some of these patients could have been mistakenly clinically characterized as having choroideremia but they may not lack the REP-1 protein. These misdiagnosed patients may have a condition phenotypically similar to CHM, but potentially carry a mutation in a different gene. I hypothesize that examination of the RNA in REP-1 deficient patients will reveal splicing defects, while whole exome sequencing on REP-1 expressing patients may identify new genes that cause a similar phenotype. Currently, in our research database, we have 280 choroideremia families of which, only 84% have a coding sequence mutation and 9% have an exonic copy number variation. The aim of this research is to improve the sensitivity of the current diagnostic test for choroideremia within our patient database and discover other genes that when altered, lead to a similar phenotype as CHM.

Test Method	Sensitivity in Affected Males
Coding sequence analysis	84%
Exon CNV analysis	9%
Unknown	6%

(MacDonald, IM research database)

Table 1.1. Current sensitivity of CHM diagnostic testing. Choroideremia diagnostic sensitivity at the start of this study within the MacDonald research patient database.

CHAPTER 2 MATERIALS AND METHODS

2.1 Patient samples

Consent was obtained from all patients through a research ethics board-approved protocol by either Dr. Ian MacDonald or Stephanie Chan (genetic counsellor).

2.2 Cell culture

Human fibroblast cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM), supplemented with 10% Fetal Bovine Serum, 1x PenStrep, 4 mmol L-glutamine, 1x sodium pyruvate and 1x non-essential amino acids. The cells were incubated at 37°C and allowed to grow until they were 80% confluent. The media was removed and the cells were rinsed with 3.0 ml of PBS, and subsequently harvested by incubation in 3.0 ml of trypsin for 7 min. The trypsin was inactivated using 4.0 ml of media. The cells were centrifuged at 0.4 g for 5 min. The cell pellet was then rinsed with PBS at which point the cell samples were ready for DNA, RNA or protein isolation.

2.3 RNA extraction

RNA was extracted using a Macherey-Nagel RNA Isolation NucleoSpin RNA Extraction Kit according to the manufacturer's instructions. This method lyses cells in a buffer containing chaotropic ions, which immediately inactivates RNases and creates favourable binding conditions to the silica membrane. Contaminating DNA bound to the silica membrane is removed with a DNase solution. Two different buffers are used to wash RNA of salts, metabolites and macromolecular cellular components and RNA is eluted with RNase-free water. RNA quantity and quality was then assessed using a NanoDrop spectrophotometer at 260 nm and 280 nm and stored at -80°C.

2.4 cDNA synthesis

Synthesis of cDNA was completed using the Thermo Fisher Scientific cDNA kit #K1621 with Random Hexamer primer according to the manufacturer's instructions. The reverse transcriptase in this kit is RevertAid Reverse Transcriptase which has lower RNase H compared to other reverse transcriptases. The RiboLock RNase Inhibitor protects template RNA from degradation. The samples were stored at -4°C.

2.5 Protein extraction

Human fibroblast cells were cultured as previously described. One quarter of a Thermo Fisher Scientific Pierce Phosphatase Inhibitor tablet was added to 2 ml of ice-cold lysis buffer (0.05M Tris-Base, 0.15M NaCl, 1% TritonX) and this buffer was added to the cell pellet. The tube was vortexed and left on ice for ten min. The solution was then spun down for 10 min at maximum speed at 4°C. The lysate was decanted and kept frozen at -80°C.

2.6 Protein quantification

Protein quantification was completed using the Thermo Fisher Scientific Pierce BCA Protein Assay Kit according to the manufacturer's instructions. The working reagent was mixed in a 50:1 ratio of Reagent A: Reagent B. Lysis buffer was used as a blank measurement and each protein sample was loaded in 100%, 20% and 5% dilutions. 25 ul of each sample was loaded into the plate and 200 ul of working reagent added to each sample. The plate was covered and placed on a shaker for 30 sec, followed by a 30 min incubation at 37°C. Absorbance at 562nm was measured using the Thermo Fisher Scientific Multiskan GO, and protein concentration was determined from a standard curve.

2.7 Western blot

2.7.1 Protein preparation

Protein samples were prepared by mixing 100 uL of lysate with 25 uL of Laemmlli buffer (60 mM Tris-Cl pH 6.8, 2% SDS, 10% glycerol, 5% β-mercaptoethanol, 0.01% bromophenol blue). SDS-PAGE was carried out with a 4% stacking and 10% resolving gel in running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS). Proteins were run alongside a Thermo Fisher Scientific Spectra multicolour broad range protein ladder at 100V until the 25 kDa protein-dye band reached the bottom of the gel. Electrophoresis was conducted using a Bio Rad power pac 300 mini protean 3.

2.7.2 Transfer

Bio Rad polyvinylidene difluoride (PVDF) membrane was activated in methanol and allowed to sit in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol) for 5 min along with the gel, fibre pads and filter paper. The gel was stacked while in buffer within the cassette in this order: fibre pad, two pieces of filter paper, the gel, the PVDF membrane, two more pieces of filter paper and a second fibre pad. The electrophoresis apparatus was assembled according to the manufacturer's instructions and the protein was transferred to the membrane in a cold room at 90mA overnight.

2.7.3 Blocking

The membrane was immersed in TBS-T (Tris-buffered saline, 0.1% Tween 20) with 5% non-fat dry milk powder (TBST-M) for 1 hr on Thermo Fisher Scientific Ocelot rocking platform.

Subsequently, the membrane was cut at the 50 kDa protein-dye band. The top half of the membrane was then incubated in a 1/500 dilution of SC2060 (Santa Cruz Biotechnologies) REP-1 mouse primary antibody in TBST-M and the bottom half of membrane was incubated in a 1/20,000 dilution of β-actin-HRP conjugated antibody in TBST-M. The membranes were rocked with primary antibody at 4°C for 1 hr. Following this incubation, the top half of the membrane was rinsed with TBS-T, followed by two TBS (Abcam recipe 20 mM Tris and 150 mM NaCl, pH 7.6) washes for 5 min each and a final 5 min wash in TBS-T on the rocker. The membrane was then incubated in a 1/10,000 dilution of A31571-donkey-anti-mouse secondary antibody (Thermo Fisher Scientific) in TBST-M in the cold room for 1 hr. Following the final incubation, the same washing steps were carried out as before.

2.7.4 Detection and development

GE Healthcare ECL prime western blotting developing solution was used according to the manufacturer's instructions. 1.4625 ml of Solution A was added to 0.0375 ml of Solution B. The developing solution was added on top of the membrane and allowed to sit in the dark for 5 min. The blot was imaged with Li-COR c-DiGit blot scanner.

2.8 DNA extraction from blood

Extraction of DNA from blood samples was performed using a QIAGEN Puregene Blood Core Kit C #158389 according to the manufacturer's instructions. This kit uses an anionic detergent to lyse cells in the presence of a DNA stabilizer, which limits the activity of DNases. An RNA digesting enzyme is used to remove RNA from the sample through salt precipitation. The DNA

was precipitated with alcohol and dissolved in the hydration solution. Extracted genomic DNA was stored in the refrigerator at 4°C.

2.9 DNA extraction from cell lines

Extraction of DNA from cultured human cells was performed using the QIAGEN DNeasy Blood and Tissue Kit #69504 according to the manufacturer's instructions. This kit lyses samples using proteinase K. Buffering solutions are added to optimize DNA binding to the column during centrifugation while contaminants pass through. DNA bound to the column is washed twice and is eluted using water. The extracted DNA was stored at -20°C.

2.10 Human identification

Human identification (HID) was carried out using the Thermo Fisher Scientific AmpFSTR Identifiler PCR Amplification Kit according to the manufacturer's instructions. This is a short tandem repeat (STR) multiplex assay that amplifies 15 tetranucleotide repeat loci and a sex chromosome specific marker in a single PCR amplification. The PCR conditions are initial denaturation at 95°C for 11 min, followed by 28 cycles of 94°C for 1 min, 59°C for 1 min, 72°C for 1 min and a final extension at 60°C for 60 min. Five different fluorescent dyes are used in the analysis. The PCR products were combined with Hi-Di formamide (Thermo Fisher Scientific) and GeneScan-500 (LIZ) (Thermo Fisher Scientific) standard on a 96-well plate; samples were denatured at 94°C for 5 min and chilled at -20°C for 2 min. Capillary electrophoresis was completed on an Applied Biosystems 3130 (Thermo Fisher Scientific) and sample results were exported to the Applied Biosystems GeneMapper ID 3.2 software for analysis.

2.11 Multiplex ligation-dependent probe amplification (MLPA)

MLPA was completed using a MRC-Holland SALSA PCR kit (P366-A2) according to the manufacturer's instructions. Four wildtype controls were used along with one duplication control, low TE buffer was used as a negative control. DNA samples were diluted to 20 ng/uL. Denaturation was completed in a thermocycler at 98°C for 5 min and held at 25°C. The probe mix consisted of 1.5 uL of SALSA probe mix and 1.5 uL MLPA buffer for each sample. 3.0 uL of probe mix was added to each sample and the probes permitted to anneal to the DNA in the thermocycler (95°C for 1 min then a hold at 60°C for 16-18 hr). Next, 32.0 uL of Ligase-65 mix (3.0 uL Ligase-65 buffer A, 3.0uL Ligase-65 buffer B, 25.0 uL PCR grade water, 1.0 uL Ligase - 65 for each sample) was added to each sample tube at 54°C. Ligation was completed in the thermocycler at 54°C for 15 min, followed by 98°C for 5 min and held at 4°C. The PCR mix was then made (2.0 uL SALSA PCR primer mix, 0.5 uL SALSA polymerase, 7.5 uL PCR grade water) and 10 uL was added to each sample tube at room temperature. PCR (35 x [95°C for 30 sec, 60°C for 30 sec, 72°C for 1 min], 72°C for 20 mins, held at 15°C) was completed in the thermocycler. The PCR products were diluted by adding 1.0 uL to 9.0 uL formamide mix (8.7 uL Hi-Di Formamide, 0.3 uL GS500 LIZ (Thermo Fisher Scientific)) in a 96-well plate. The samples were denatured at 95°C for 5 min, then snap-cooled on ice. Samples were run by capillary electrophoresis on an Applied Biosystems 3130 (Thermo Fisher Scientific) and data analyzed using MRC-Holland Coffalyser software.

2.12 PCR

Polymerase chain reactions were carried out using Thermo Fisher Scientific Phusion Taq polymerase. Each reaction was carried out in 20 uL volumes. Reactions contained 4.0 uL of 5X Phusion HF Buffer, 0.4 uL of 10 mM dNTPs, 1.0 uL of 10 uM forward and reverse primers, 1.0 uL of template DNA and 0.2 uL of Phusion polymerase. The remaining volume was nuclease-free water to 20 uL. Primer names, locations, directions and melting points are found in Table 2.1 (all primers were ordered from integrated DNA technologies (IDT)). PCR conditions were 98°C for 30 sec, [98°C for 10 sec, 60°C for 30 sec, 72°C for 20 sec] x 35, 72°C for 5 min. Annealing temperatures were adjusted based on melting temperatures (T_m) of the primers used in each reaction and extension time was adjusted to meet 10-30 seconds per kilobase of expected read length. The reactions were performed in a Primer 3 thermocycler (Thermo Fisher Scientific). All primers used in experiments and their annealing location are listed in Table 2.1.

2.13 PCR purification

PCR products were purified using the Fermentas Life Sciences GeneJET PCR purification kit according to the manufacturer's instructions. This kit uses a silica-based membrane technology in a spin column to remove primers, dNTPs, enzymes, and other reaction components. Once purified, PCR products were kept at -20°C.

2.14 Sanger Sequencing

PCR products were sequenced at the University of Alberta TAGC core lab facility. Samples were submitted according to dilution guidelines with 3.2 pmol/uL primer. Sequencing reactions were carried out using the Thermo Fisher Scientific Big Dye Seq Kit according to the

manufacturer's instructions. Samples were run on an Applied Biosystems 3130 Genetic Analyzer (Thermo Fisher Scientific) and sequences were analyzed in Biomatters Limited program Geneious. Alamut Visual Interactive Biosoftware was used to predict whether any variants affected splicing.

2.15 Bioinformatics

Trio Whole Exome Sequencing (WES) was completed off-site by DNALink on the patient, mother and father. Based on the inheritance pattern and family history, the data was filtered for compound heterozygous candidates (Figure 2.1). Two total records files were opened, the first for share variants between the patient and his mother and a second for share variants between the patient and his father. One document was filtered to contain only C300-TA (patient) and C300-AL (mother) under the "Hetero" column and other document was filtered to contain only C300-TA (patient) and C300-JA (father). Both documents were filtered under the "1000g MAF" column to remove all blank values and only variants with minor allele frequency less than 0.01. Both documents were filtered to remove blanks and synonymous variants in the column named "HGVS p." Under the "Impact" column, all variants predicted as "LOW" or "MODIFIER" were removed. The remaining variants of each patient-mother and patient-father lists were compared to each other and all non-shared genes were removed. Candidate genes were classified after cross-referencing genes that appeared in both patient-mother and patient-father documents with a list of genes currently known to show expression in the retina and that play a role in eye development. Following HGVS guidelines, variants were investigated through databases and freely available software including: dbSNP (available from: http://www.ncbi.nlm.nih.gov/SNP/, Sherry ST, et al., 2001), SIFT (available from: http://sift.jcvi.org/, Ng PC and Henikoff S, 2002) and PolyPhen-2 (available from: http://genetics.bwh.harvard.edu/pph2/, Adzhubei I, et al., 2013).

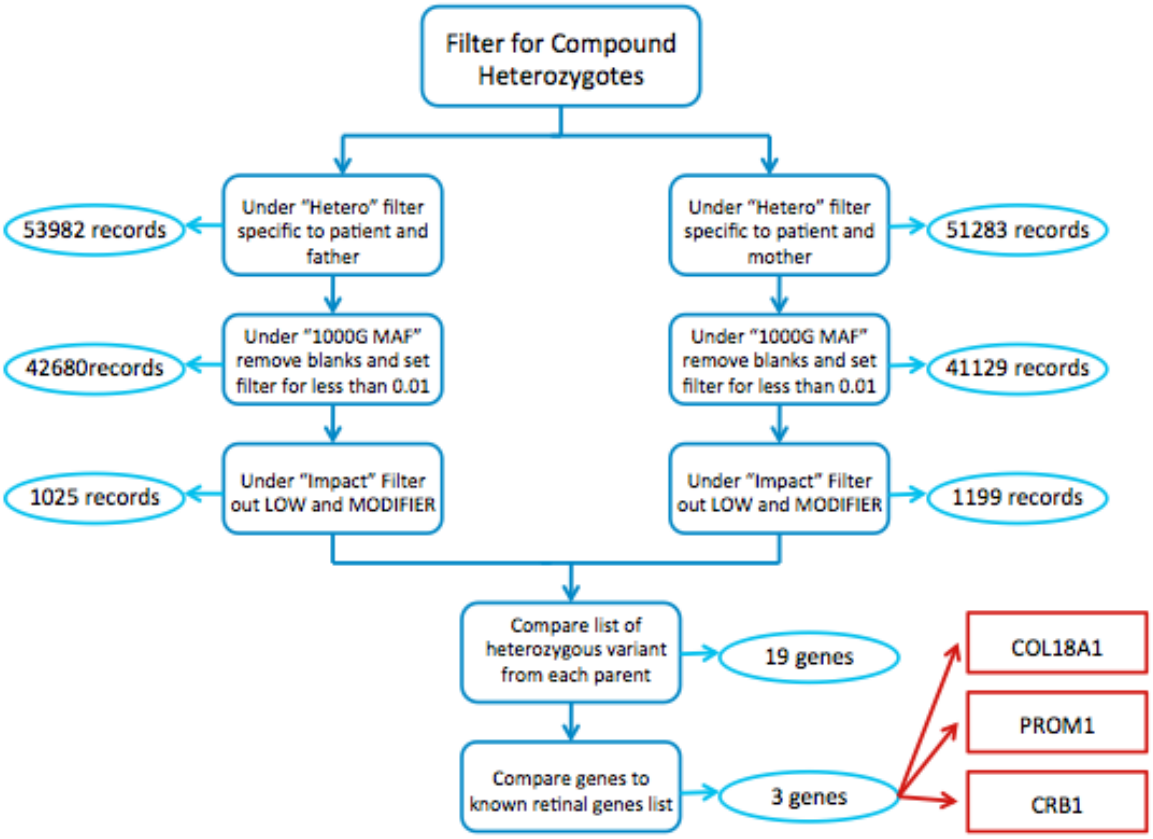


Figure 2.1: Bioinformatic workflow. Bioinformatic workflow of filtering variants for compound heterozygotes.

2.16 Mouse X-inactivation assay design

2.16.1 PCR

All mouse samples were previously genotyped to be heterozygous for the *Chm* wildtype and *Chm* null alleles. RNA was extracted and cDNA was produced as previously described. The *Chm* null allele has a deletion of exon 4. Three primers were designed and purchased from integrated DNA technologies (IDT): a forward primer (5'—ACTGGGCCAGTTTCAGCTTT-3') and two reverse primers fluorescently tagged with 56-FAM for quantification purposes, one in exon 4 (5'-TGCCACATTGAATTTTCAGTCACA-3') and one in exon 5 (5'-TTTTTCTGCAGTGCACCAGC-3'). Product lengths were designed to be similar but not identical so the difference in fluorescent peaks could be seen by capillary electrophoresis. PCR conditions were 98°C for 30 sec, 35 x [98°C for 10 sec, 60°C for 30 sec, 72°C for 10 sec], 72°C for 5 min. The PCR products were run on 1% agarose gel at 120V for 20 min.

2.16.2 Electrophoresis Injection

A master mix containing 0.3 ul GeneScan500 (LIZ) (Thermo Fisher Scientific) and 9.0 ul HiDi (Thermo Fisher Scientific) per reaction was made. PCR products were diluted to 100%, 50% and 25% using water. 9.0 ul of this master mix and 1.0 ul of PCR product dilutions were added to each well on the plate. The plate was gently vortexed, spun down and put on hot plate to denature for 3 min at 95°C. The plate was then cooled on an ice block for 2 min. The samples were then injected on the Applied Biosystems 3130 genetic analyzer (Thermo Fisher Scientific) and the data was exported into the GeneMapper ID software (Thermo Fisher Scientific) for analysis.

Table 2.1: Primer table. Primer nomenclature, target description, direction of binding, sequence and melting temperature.

Name	Target description	Direction	Primer sequence	Tm (°C)
16	CHM exon 1	Forward	CTT TCC TAC TTC AAC CCT CCA G	54.9
17	CHM exon 1	Reverse	ACA GTC TTC CTA AAC TTT GTC C	52.4
50	CHM cDNA	Forward	TAC CAG GAT TCT TGC ATT TCG	52.9
53	CHM cDNA exon 5	Reverse	AGG CAG GAA GGC AGA ATC T	56.2
54	CHM cDNA exon 6	Reverse	CCT GTT CCA CTC GTC CTT CT	56.4
55	CHM cDNA exon 9	Reverse	AGG CAC TGT ACT GAA TGG CG	57.6
56	CHM cDNA exon 13/14	Reverse	TTG TTC ATT TTC TAT CTC CAT TTC AG	51.9
132	CHM 5' promoter region	Reverse	CTT GTG GAA ATG AGA TCA AGT TAG G	53.7
141	CHM 5' promoter region ~1997 bp	Forward	GGC TTC TCA TTA TTA GAG AAG ATT TTG G	54.1
142	CHM 5' promoter region ~1063 bp	Forward	CAG GGA AGG CCC ACT ACT GC	60.4
145	CHM sequencing 5' promoter region ~1200 bp	Forward	CAG CGA CAT CCT GGG AAT AAC	55.6
146	CHM sequencing 5' promoter region ~360 bp	Reverse	TTA GGC TGG AAT GAA AGT CTC C	54.5
150	CHML (REP2) position -50	Forward	CAC CTC ATT TCT TTC ATC AG	48.7
151	CHML (REP2) position 812	Reverse	TTC CCG AAA TGC AAG AAT CC	53.3
183	CHM exon 1 mRNA check	Reverse	CCG TCC CTA TTA CGA TCA CAT CAA AC	57.2
184	CHM 5' UTR mRNA check	Forward	TAA TAG TCA CAT GAC ACG TTT CCC G	56.8
185	CHM exon 1/2 boundary mRNA check	Reverse	TGG ATT CAG GCA AAC CCG T	57.2
186	CHM exon 10 mRNA	Reverse	TGT CCT CCA CGA GGA AAT GCT CA	60.3
187	CHM exon 11 mRNA	Reverse	CAG ATC TAT CTG TAA TCA GCA CTG CCC	58.6
188	CHM exon 12 mRNA	Reverse	ACC CGA ACA GCA AAA GTT CCT GG	60
189	CHM exon 13 mRNA	Reverse	CTT GCT GTT TTA GAA GAT GTG CAA GTC	56.7
191	CHM promoter 524 bp upstream	Forward	GAC ATG GAA ACG AAA AAG ATA TGG	52.4
192	CHM promoter 411 bp upstream	Forward	AG GGC CAA CAC GAA AGT ACA G	57.4
193	CHM 320 bp upstream	Forward	AAT TTC ACT CCG CTC TCC GTG	57.4
279	CHM exon 2	Reverse	TCT CCG GCC ACT TCT TGA AC	57.2
280	CHM exon 3	Reverse	CTG GCC CAG TTT CCT CCA TA	56.8
281	CHM exon 4	Reverse	ACT TGG TCT TGC CAC ACT GG	57.8

Table 2.1: Continued.

Name	Target description	Direction	Primer sequence	Tm (°C)
282	CHM 25kb upstream	Forward	TCA GAA ACA AGG GAG GCC AC	57.4
283	CHM 25kb upstream	Reverse	ACT GTG ACC TGC CTG CAT TT	57.4
284	CHM 50kb upstream	Forward	TCC TGT TGG TCC TCA AGC AC	57.3
285	CHM 50kb upstream	Reverse	AGC AAA CAA CAG CCC TCT CT	56.8
286	CHM 100kb upstream into DACH2	Forward	CAA CAA ACA GTG AGC AGC CC	57
287	CHM 100kb upstream into DACH2	Reverse	CCA TCT ATT AGC CCC CGC AG	57.6
288	CHM 10kb upstream	Forward	CCC CAC ATG CAC ACA CAT TG	57.2
289	CHM 10kb upstream	Reverse	CAA CCC TTG CAC TGG CTT TC	57.1
290	CHM 15kb upstream	Forward	CAC TGT GCT CCA GGG CTT AT	57.2
291	CHM 15kb upstream	Reverse	CCC TAA AAC CCA CAG ACC CC	57.8
292	CHM 20kb upstream	Forward	AGG TCC TCC TTC ACT TCC CT	57.4
293	CHM 20kb upstream	Reverse	TCA TCT CGG CCC AAA AGC TT	57.1
294	CHM 7.5kb upstream	Forward	TCT CTG CAA AGC CAA GGG AG	57.4
295	CHM 7.5kb upstream	Reverse	TTG GCC ATC TTG GAA CCT CC	57.5
296	CHM 8.5kb upstream	Forward	TAT CAG GTG GCC TCC TCC TC	58.1
297	CHM 8.5kb upstream	Reverse	AGC CTC TCT TTC CAG GTC CA	57.9
303	CHM intron 1 5kb	Reverse	ACC AAG GCA GGT AGA TTG CTT GAG	59.1
304	CHM intron 1 5kb	Forward	AAT CCT CCC ACC TCA GAC TCC TG	60
305	CHM intron 1 10kb	Reverse	GCG ATT CTC CTG CCT CAG CTT C	60.2
306	CHM intron 1 10kb	Forward	GGC AAG TGG ATC ACC TGA GGT C	59.8
307	CHM intron 1 15kb	Reverse	GGC ATT CAG TAA ATG TTG CTT CTC TTC C	58.2
308	CHM intron 1 15kb	Forward	GCA GAG CAA GTT GTT TGG TAC AAC C	58.8
315	CHM intron 1	Reverse	ACT CAC GTG CTC CCT CCT AT	57.6

CHAPTER 3 RESULTS

3.1 Patient A

Patient A, who is of Asian ancestry, had a reported family history of vision loss. As choroideremia was suspected (Figure 3.1), immunoblot analysis was performed and confirmed a lack of REP-1 protein (Figure 3.2). To determine the molecular basis for this result, sequencing of coding regions of the *CHM* gene along with copy number analysis using multiplex ligation-dependent probe amplification (MLPA) was performed previously in a diagnostic laboratory (data not shown). These methods did not identify any pathogenic variants. As this patient had agreed (years ago) to have a fibroblast cell line produced from a biopsy, RNA was extracted and cDNA was produced. Our hypothesis was that this patient likely had a deep intronic splice mutation that was not detectable by the above molecular assay. To assess this we planned on amplifying the cDNA derived from the mRNA transcript in sequential steps to reveal any splicing errors. The first amplicon included exons 1-5, the second amplicon included exons 1-6, a third amplicon included exons 1-9 and a fourth amplicon included exons 1-14. All reactions demonstrated a product of the expected length except for the fourth amplicon spanning exons 1-14 (Figure 3.3). This implied that a defect lied between exons 9-14. To narrow down the affected region further, reverse primers were designed along the cDNA transcript in exons 9, 10, 11, 12, 13, 14 and the 3' UTR and PCR analysis determined the presence, absence, or variable size of the *CHM* transcript (Figure 3.4). PCR products were of expected size until exon 10 (Figure 3.5). Although exon 10 did amplify, a closer examination of the size revealed the fragment was 100 bp larger than expected (Figure 3.5). Therefore, this PCR product was purified and Sanger sequenced. Alignment of the resulting abi file in the software Geneious (Biomatters Limited©) revealed a 114 bp insertion that mapped to the adjacent intronic region (Figure 3.7). In order to determine the underlying genomic variant that caused this mis-splicing,

primers were designed to the area surrounding the region. Sanger sequencing of the genomic DNA identified a point mutation in the intronic region (Figure 3.7). According to HGVS nomenclature, this variant is c.1245-521A>G and is predicted by the software Alamut ® (Interactive Biosoftware) which utilizes splicing prediction software such as Human Splice Finder (HSF), GeneSplicer, NNSplice and Splice Site Finder (SSF) to cause a new, more favourable splice acceptor. As a result, 114 bases of intronic sequence are introduced into the mRNA transcript prior to exon 10. This point mutation changes the lysine to cysteine (p.Lys415_Cys416ins22Ter) as well as inserts 21 amino acids into the protein sequence before the introduction of a new stop codon. This mutation was confirmed through Sanger sequencing of genomic DNA derived from patient blood samples kept in the research lab and also through an independent sample banked in the diagnostic laboratory.

This patient's sister was interested in knowing her carrier status for family planning purposes. Through fundus examination, she was determined by Dr. MacDonald to have a female carrier phenotype (Figure 3.9). This phenotype is an irregular, patchy retinal pigment epithelium. Sanger sequencing completed on patient's genomic DNA sample determined that she carried the familial mutation (Figure 3.9).

3.2 Patient B

A second, unrelated individual was first clinically diagnosed with choroideremia in 1997 (Figure 3.1). This patient, who is of Hungarian ancestry, also had a reported family history of vision loss. As choroideremia was suspected, immunoblot analysis was performed by the MacDonald lab and confirmed a lack of REP-1 protein (Figure 3.2). To determine the molecular basis for this result, sequencing of coding regions of the *CHM* gene along with copy number analysis using multiplex ligation-dependent probe amplification (MLPA) was performed (data not shown). These methods also did not find any pathogenic variants. This patient had previously agreed to have a fibroblast cell line produced from a biopsy so that cells would be available for future research. RNA was extracted from these cell lines and cDNA was produced. Again, PCR analysis of cDNA derived from patient cell lines was completed, following the same steps as the previous patient. Similar to patient A, a change in band size was observed at exon 10 (Figure 3.6). The PCR product was purified and Sanger sequenced. Alignment of the resulting abi file in the Geneious software (Biomatters Limited ©) revealed the same 114 bp insertion that mapped to the adjacent intronic region (Figure 3.7). The causative mutation was verified through gDNA sequencing from a banked genomic DNA sample (Figure 3.7). As these two patients were of different ancestry, the finding of the same mutation was unexpected. In order to verify that the samples were not mixed-up (either by myself or the MacDonald laboratory), microsatellite genotyping was completed on DNA derived from cell lines used in the research laboratory and the stock DNA banked in the clinical diagnostic laboratory (Figure 3.8). It was determined that the two patient microsatellite profiles were unique and more importantly, the cell lines matched their respective genomic DNA stored in the clinical lab, thereby ensuring no sample mix-up occurred. Additionally the MacDonald laboratory examined a nearby (approx. 3600 bp) single nucleotide polymorphism (SNP) rs#2223344 which had a minor allele frequency

of 27.3%. This analysis revealed the c.1245-4158G>C was present in the hemizygous form in Patient B while Patient A was hemizygous for the reference allele (G). This SNP difference again verified sample identity and implied that this variant was not due to a common descent. All other choroideremia patient samples lacking a molecular diagnosis in the MacDonald research laboratory were sequenced and determined not to have this splice mutation.

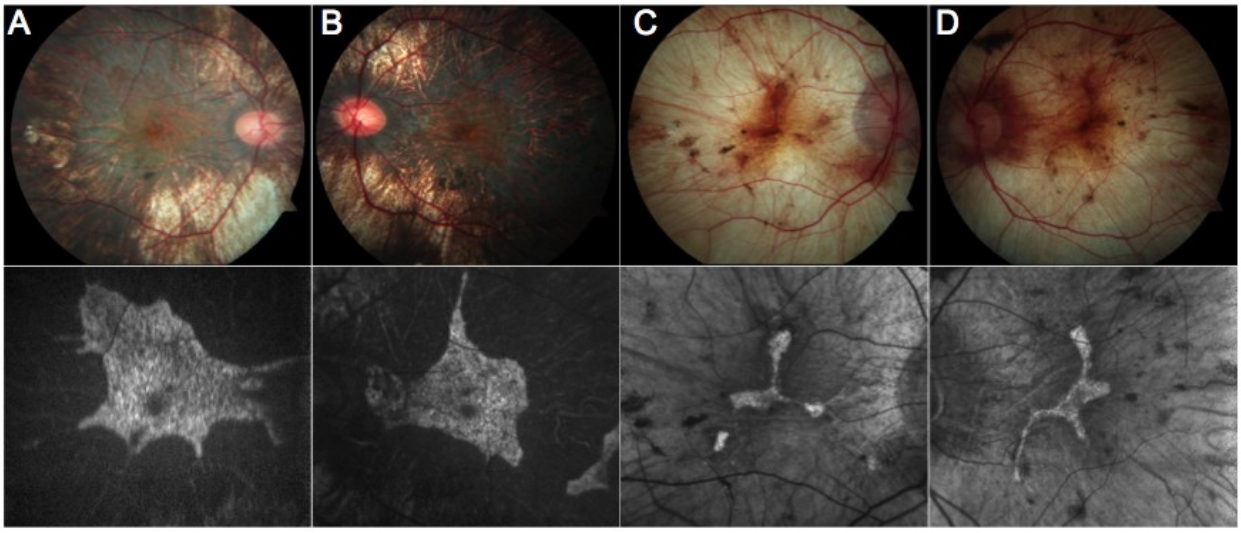


Figure 3.1: Patient A and Patient B photos of back of the eye. Top panels show colour fundus photos of the macula (50 degrees) and the bottom panels are corresponding fundus autofluorescence (30 degrees). A) and B) from patient A and C) and D) from patient B. All fundus photos provided by Dr. Dimopoulous.

Figure 3.2: Patient A and B Immunoblot analysis. Western blot analysis performed on fibroblast protein taken from two affected individuals, showing the absence of REP-1 protein. Lane 1: REP-1 present control; Lane 2: negative control; Lane 3: Patient A; Lane 4: Patient B.

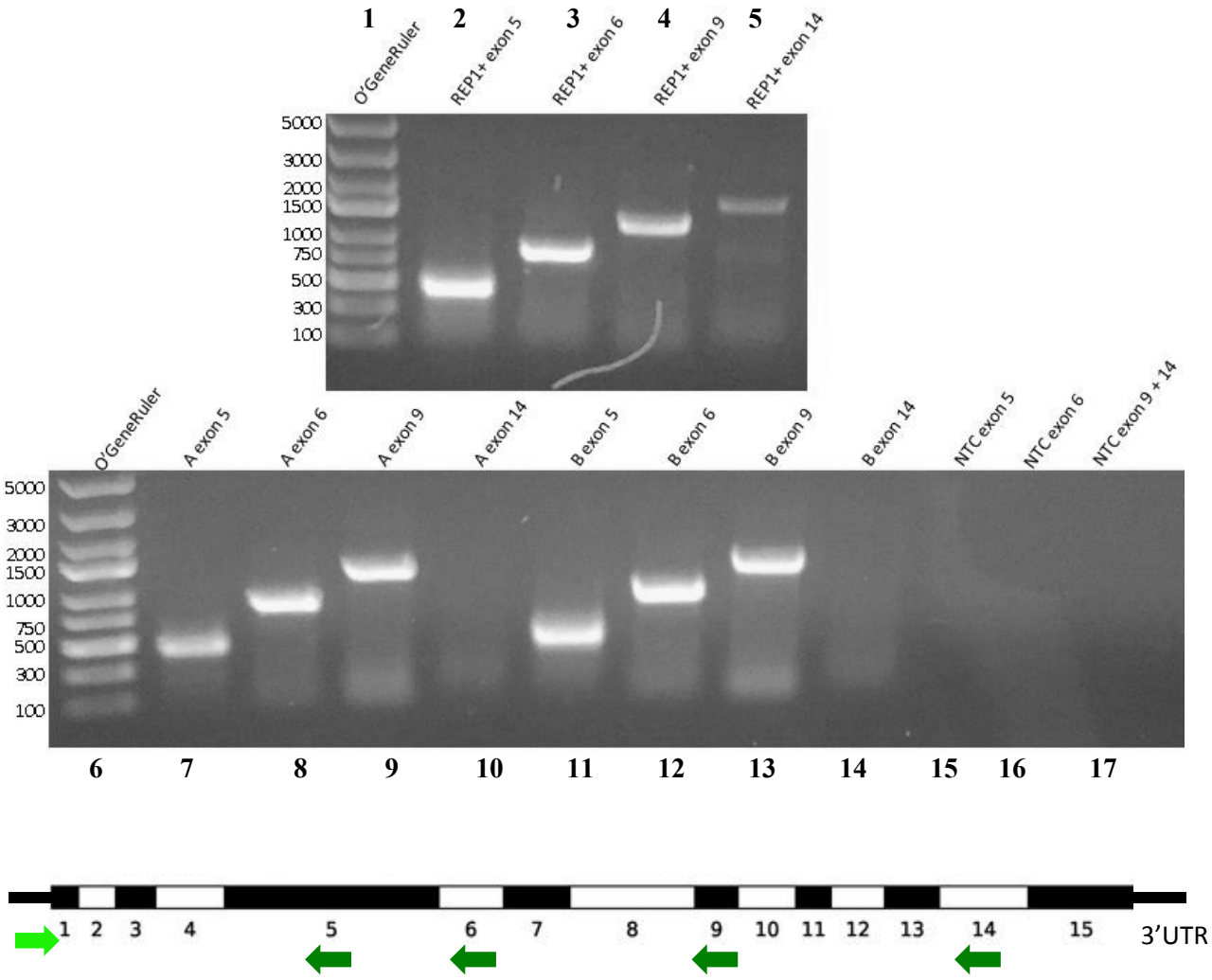


Figure 3.3: Patient A and B exons 5, 6, 9 and 14. Transcript walking of Patient A and B's cDNA. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control exon 1-5; Lane 3: REP-1 present control exon 1-6; Lane 4: REP-1 present control exon 1-9; Lane 5: REP-1 present control exon 1-14; BOTTOM Lane 6: Thermo Fisher Scientific O'GeneRuler Ladder, Lane 7: Patient A exon 1-5; Lane 8: Patient A exon 1-6; Lane 9: Patient A exon 1-9; Lane 10: Patient A exon 1-14; Lane 11: Patient B exon 1-5; Lane 12: Patient B exon 1-6; Lane 13: Patient exon 1-9; Lane 14: Patient B exon 1-14; Lane 15: negative control exon 1-5; Lane 16: negative control exon 1-6; Lane 17: negative control exon 1-9 and 1-14.

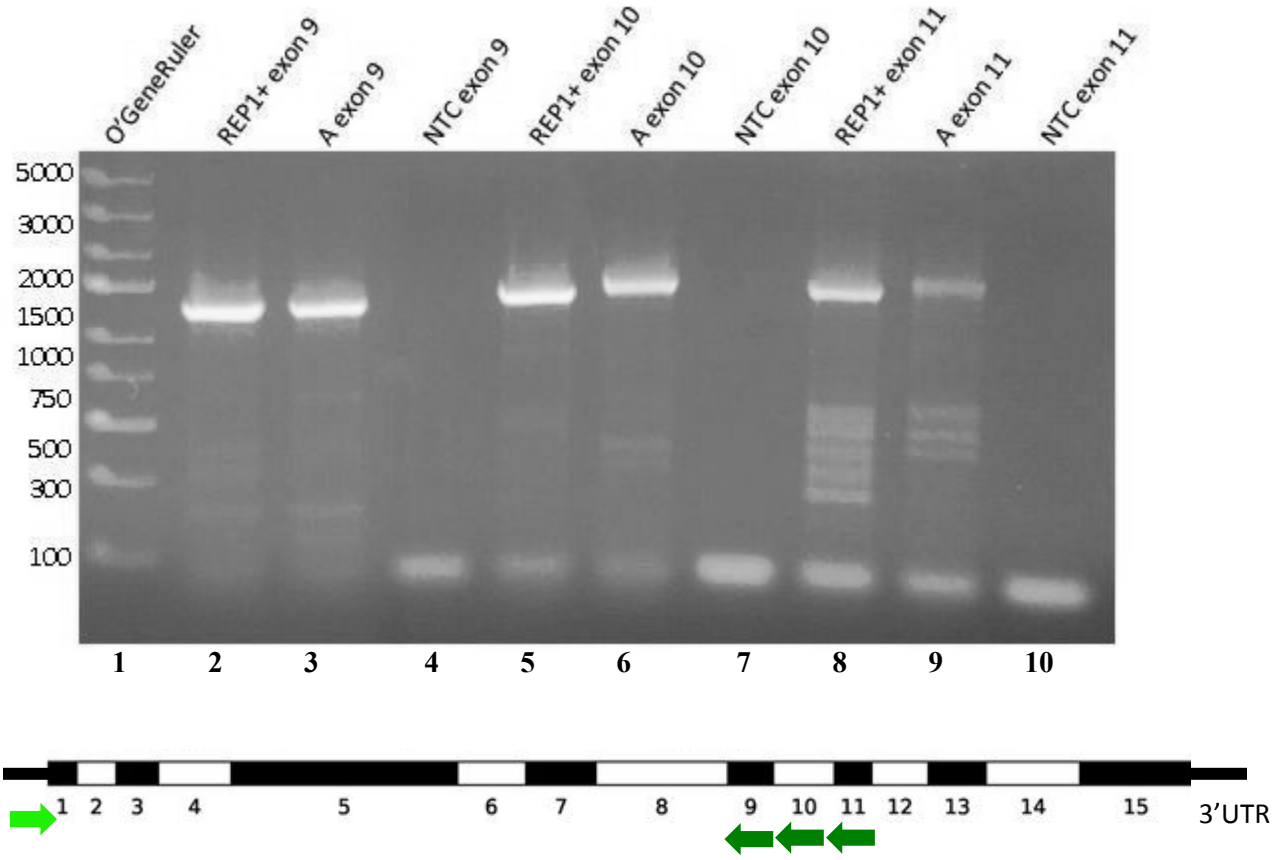


Figure 3.4: Patient A PCR analysis of cDNA exons 9, 10 and 11. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control exon 1-9; Lane 3: Patient A exon 1-9; Lane 4: negative control exon 1-9; Lane 5: REP-1 present control exon 1-10; Lane 6: Patient A exon 11-0; Lane 7: negative control exon 1-10; Lane 8: REP-1 present control exon 1-11; Lane 9: Patient A exon 1-11; Lane 10: negative control exon 1-11.

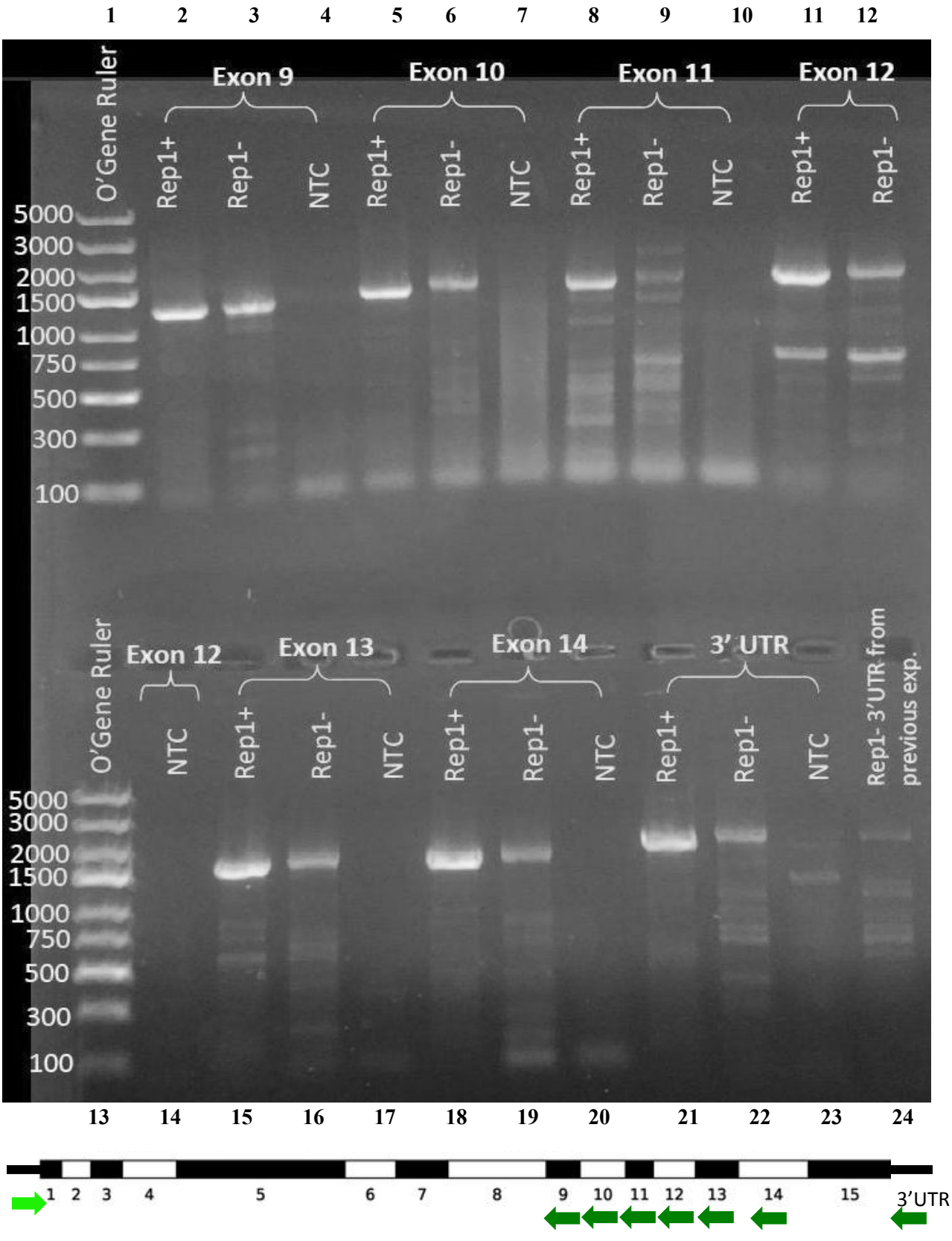

1 2 3 4 5 6 7 8 9 10 11 12
O'Gene Ruler
5000
3000
2000
1500
1000
750
500
300
100
Exon 9
Exon 10
Exon 11
Exon 12
Rep1+
Rep1-
NTC
Exon 12
Exon 13
Exon 14
3' UTR
Rep1- 3'UTR from previous exp.
13 14 15 16 17 18 19 20 21 22 23 24
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
3'UTR

Figure 3.5: Patient A PCR of cDNA exons 9 – 3' UTR. Noticed band height change after exon 9 and continued through the rest of the transcript. TOP Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control exon 1-9; Lane 3: Patient A exon 1-9; Lane 4: negative control exon 1-9; Lane 5: REP-1 present control exon 1-10; Lane 6: Patient A exon 1-10; Lane 7: negative control exon 1-10; Lane 8: REP-1 present control exon 1-11; Lane 9: Patient A exon 1-11; Lane 10: negative control exon 1-11; Lane 11: REP-1 present control exon 1-12; Lane 12: Patient A exon 1-12; BOTTOM Lane 13: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 14: negative control exon 1-12; Lane 15: REP-1 present control exon 1-13; Lane 16: Patient A exon 1-13; Lane 17: negative control exon 1-13; Lane 18: REP-1 present control exon 1-14; Lane 19: Patient A exon 1-14; Lane 20: negative control exon 1-14; Lane 21: REP-1 present control exon 1-3'UTR; Lane 22: Patient A exon 1-3'UTR; Lane 23: negative control exon 1-3'UTR.

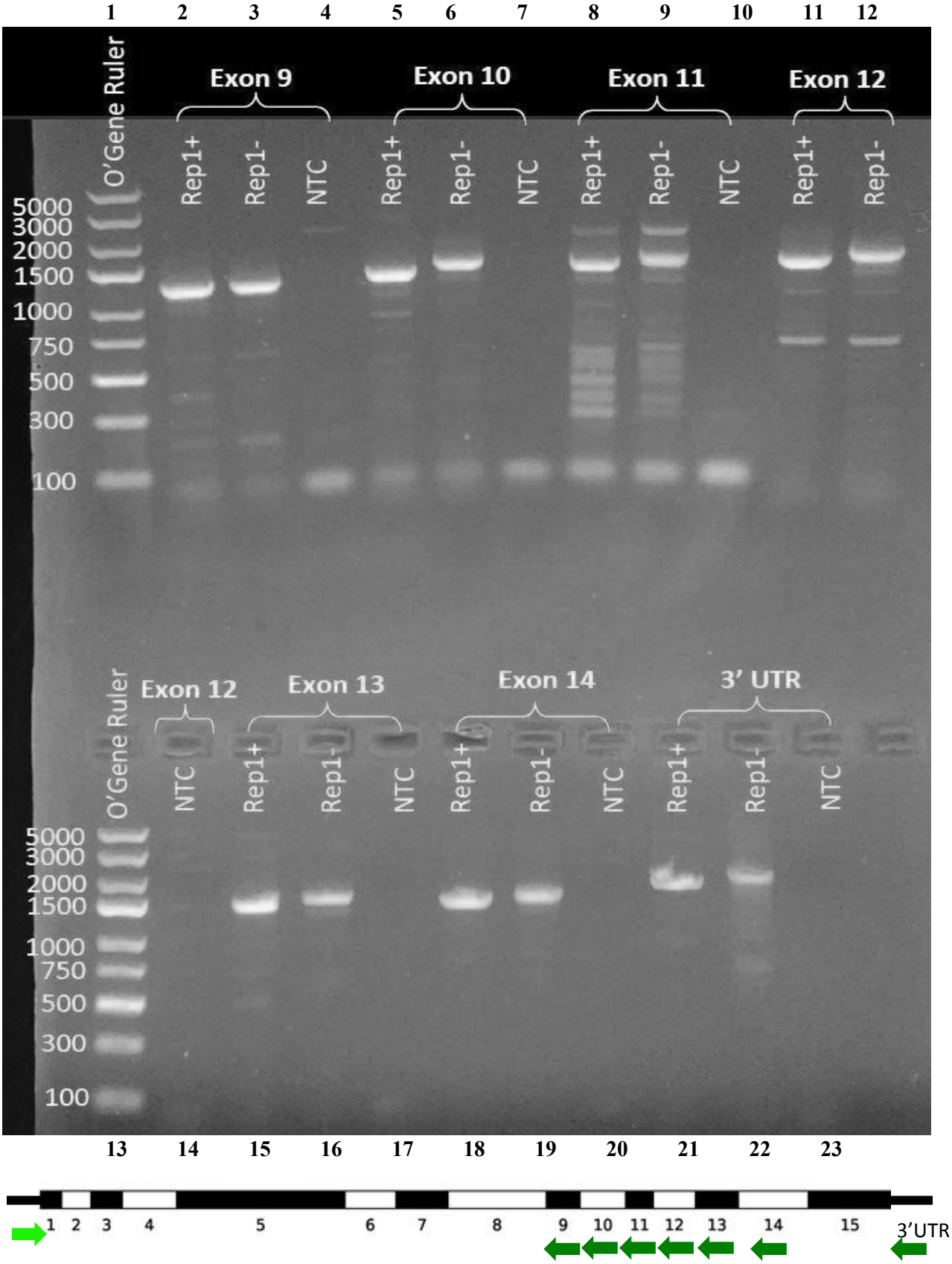

1 2 3 4 5 6 7 8 9 10 11 12
O'Gene Ruler
Exon 9
Exon 10
Exon 11
Exon 12
Rep1+
Rep1-
NTC
5000
3000
2000
1500
1000
750
500
300
100
Exon 12
Exon 13
Exon 14
3' UTR
13 14 15 16 17 18 19 20 21 22 23
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
3'UTR

Figure 3.6: Patient B PCR of cDNA exons 9 – 3' UTR. Noticed band height change after exon 9 and continued through the rest of the transcript. TOP Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control exon 1-9; Lane 3: Patient B exon 1-9; Lane 4: negative control exon 1-9; Lane 5: REP-1 present control exon 1-10; Lane 6: Patient B exon 1-10; Lane 7: negative control exon 1-10; Lane 8: REP-1 present control exon 1-11; Lane 9: Patient B exon 1-11; Lane 10: negative control exon 1-11; Lane 11: REP-1 present control exon 1-12; Lane 12: Patient B exon 1-12; BOTTOM Lane 13: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 14: negative control exon 1-12; Lane 15: REP-1 present control exon 1-13; Lane 16: Patient B exon 1-13; Lane 17: negative control exon 1-13; Lane 18: REP-1 present control exon 1-14; Lane 19: Patient B exon 1-14; Lane 20: negative control exon 1-14; Lane 21: REP-1 present control exon 1-3'UTR; Lane 22: Patient B exon 1-3'UTR; Lane 23: negative control exon 1-3'UTR.

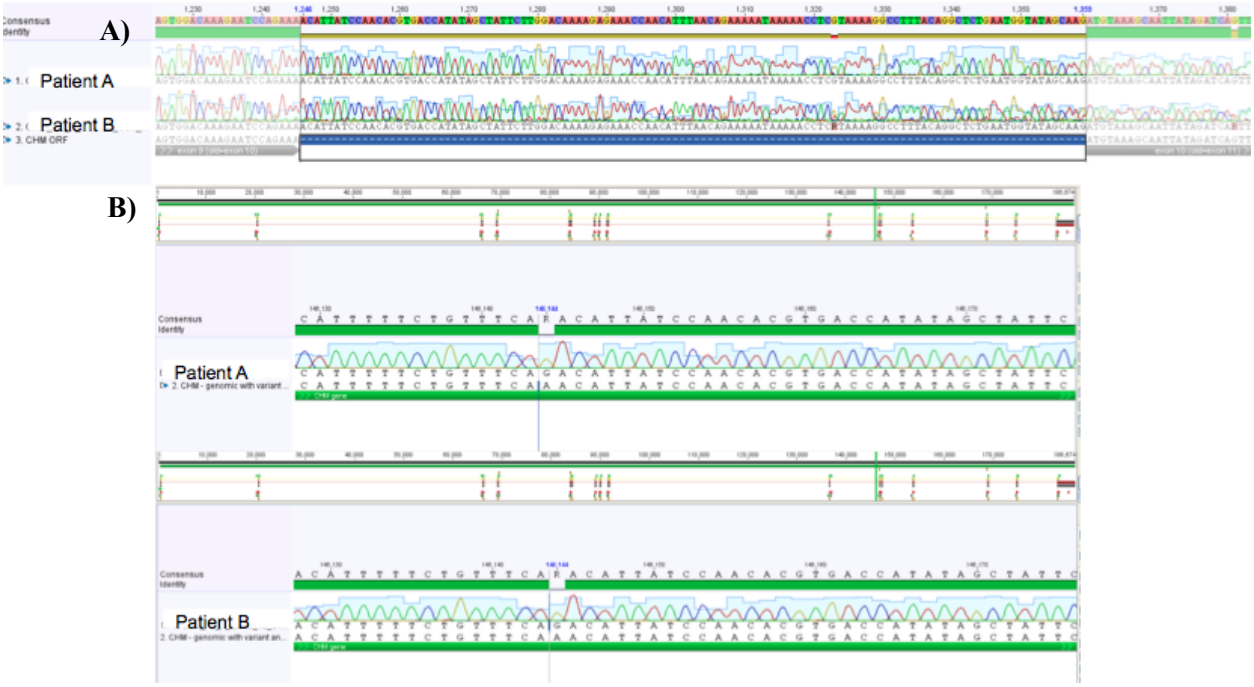
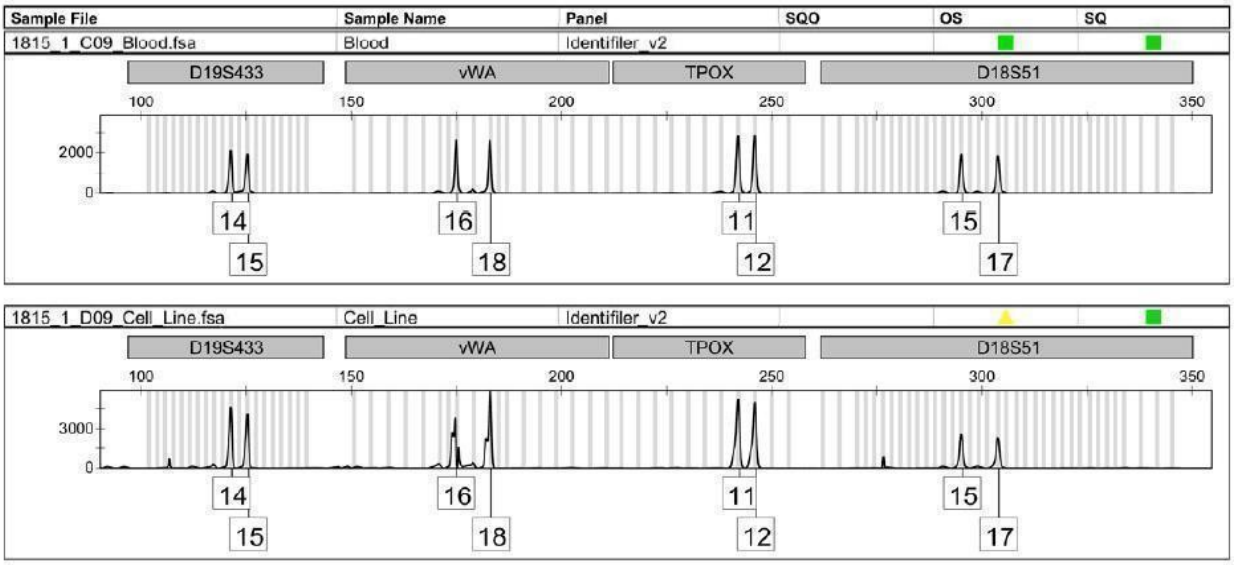


Figure 3.7: Sequencing results for patients A and B. A) Sequencing of Patient A (top) and Patient B (bottom) cDNA showing 114bp of inserted deep intronic sequence. B) Sequencing of Patient A (top) and Patient B (bottom) gDNA showing the single nucleotide change c.1245-521A>G

Applied Biosystems

GeneMapper ID v3.2.1

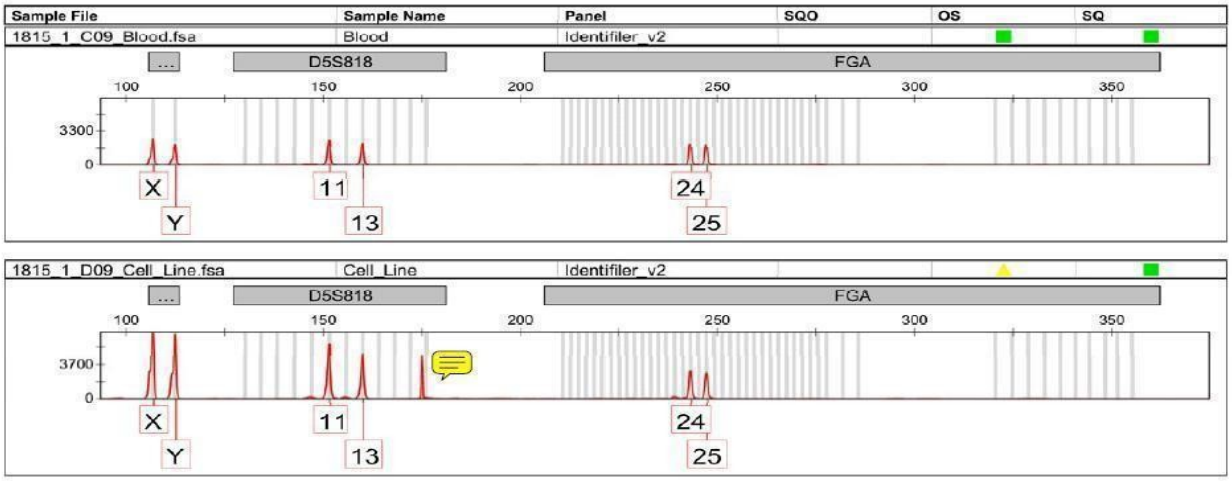
HID_2016_RM



Applied Biosystems

GeneMapper ID v3.2.1

HID_2016_RM



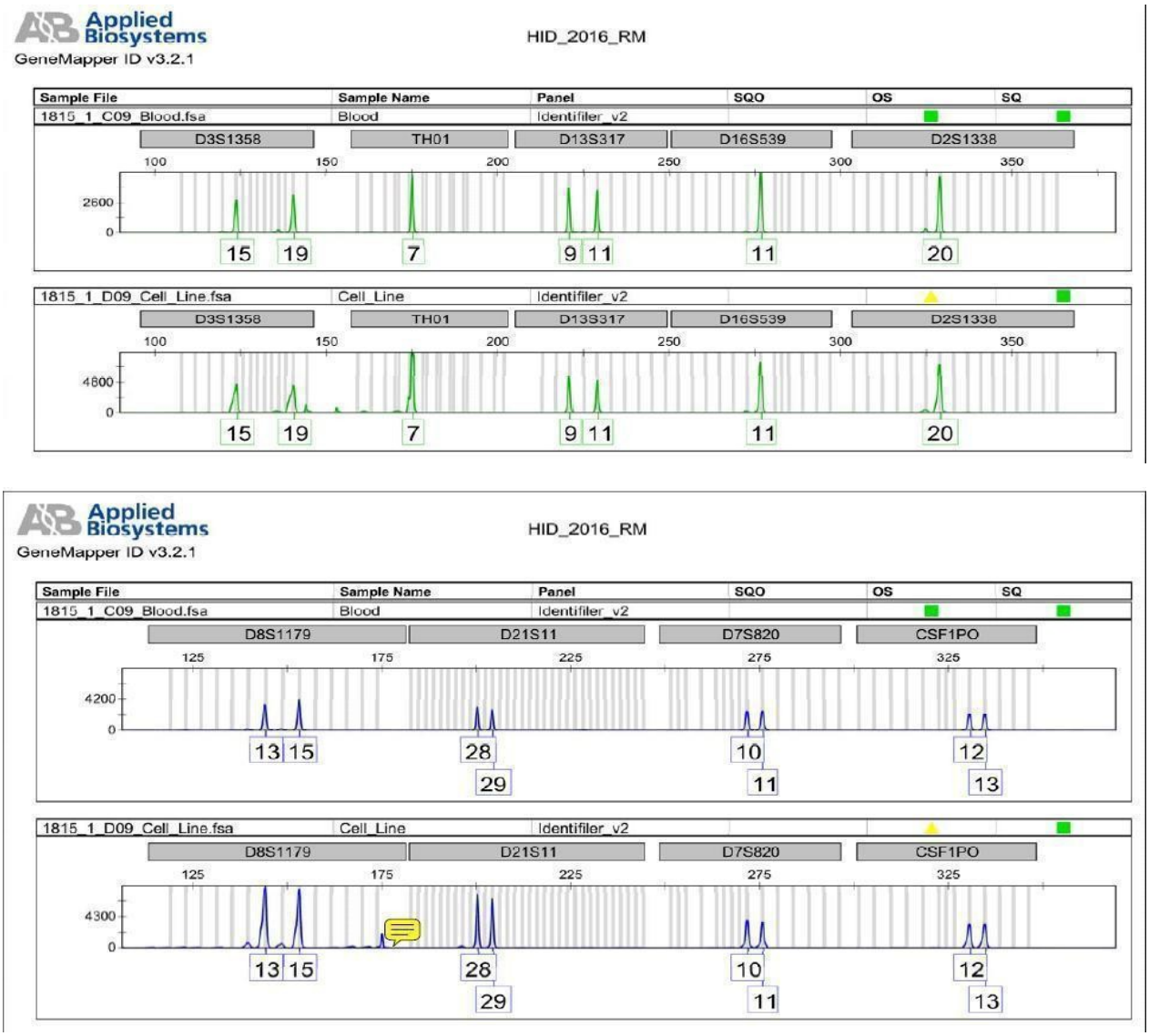
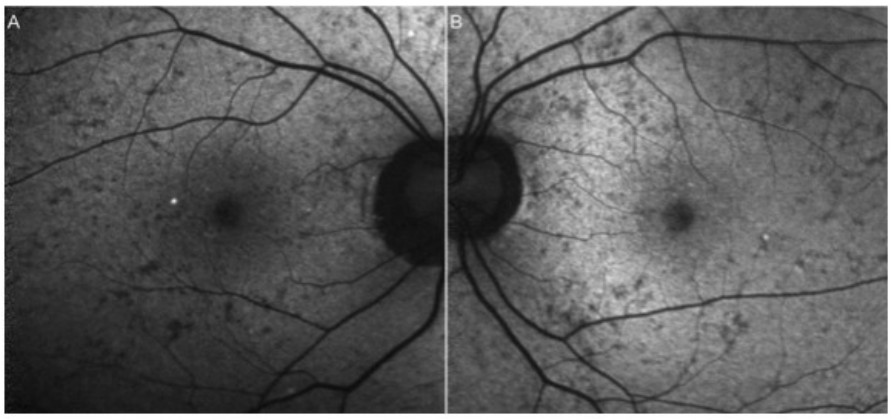


Figure 3.8: Human identification. Human Identification of various genetic markers comparing cell line DNA and stock genomic DNA of Patient A sample to determine no cell mix up occurred. A) yellow, B) red, C) green and D) blue wavelengths. Peaks marked with yellow speech bubble are bleed through fluorescence of another colour wavelength.

A)



B)

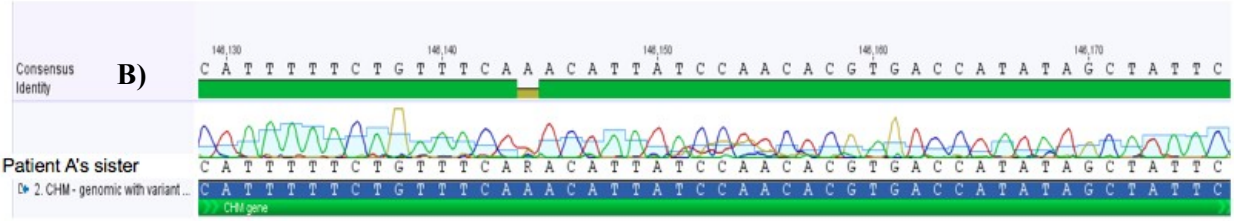


Figure 3.9: Patient A's sister's carrier status. A) Patient A's sister's fundus photos obtained from Dr. Dimopoulous. B) Sanger sequencing of gDNA showing expression of both the normal allele and c.1245-521A>G.

3.3 Patient C

This patient was diagnosed by Dr. Ian MacDonald as likely having choroideremia although he presented with an unusually severe choroideremia-like phenotype for his young age of 18 years. As this patient had also agreed (years ago) to have a fibroblast cell line produced from a biopsy, RNA was extracted and cDNA was produced. A similar transcript walking approach was used as performed for patients A and B. The first amplicon included exon 1 through to the exon 1/exon 2 boundary, the second amplicon included exons 1 through to the 3'UTR. In contrast to the previous patients, Patient C had a PCR product consistent with the full length *CHM* transcript (Figure 3.10). All PCR products representing various exons from the transcript of the *CHM* gene were also of expected length (Figure 3.11, Figure 3.12). As the PCR analysis revealed the entire *CHM* transcript, immunoblot analysis was performed to examine for the REP-1 protein. This analysis confirmed that this patient expressed the REP-1 protein at the expected molecular weight (Figure 3.13). Due to the lack of REP-1 protein expression, the early onset of this patient's symptoms and the patient not having myopia, we concluded that this patient had been misdiagnosed with CHM and began to look for an alternative cause for his disease.

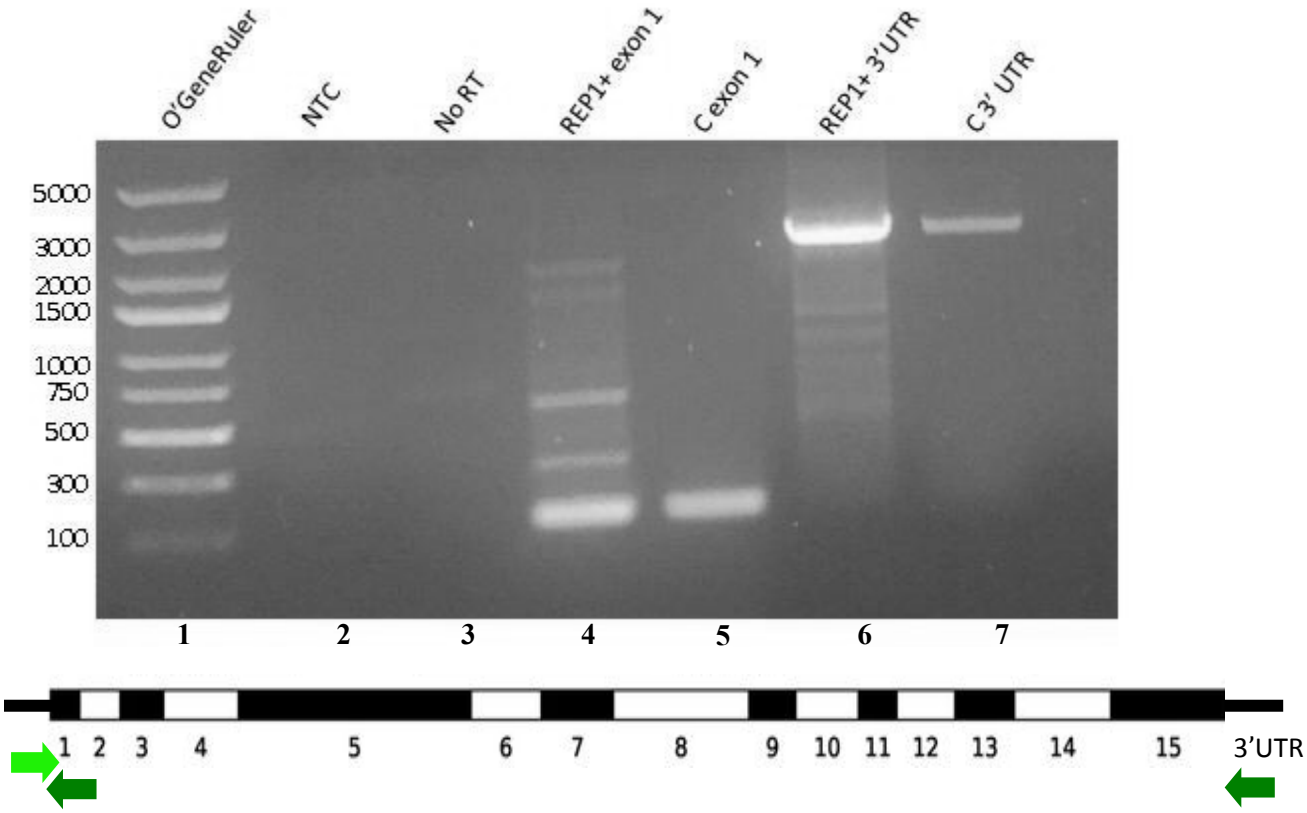


Figure 3.10: Patient C full length PCR analysis of REP-1 transcript. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: negative control exon 1; Lane 3: no reverse transcriptase control; Lane 4: REP-1 present control exon 1; Lane 5: Patient C exon 1; Lane 6: REP-1 present control exon 1 through to the 3' UTR; Lane 7: Patient C exon 1 through to the 3'UTR.

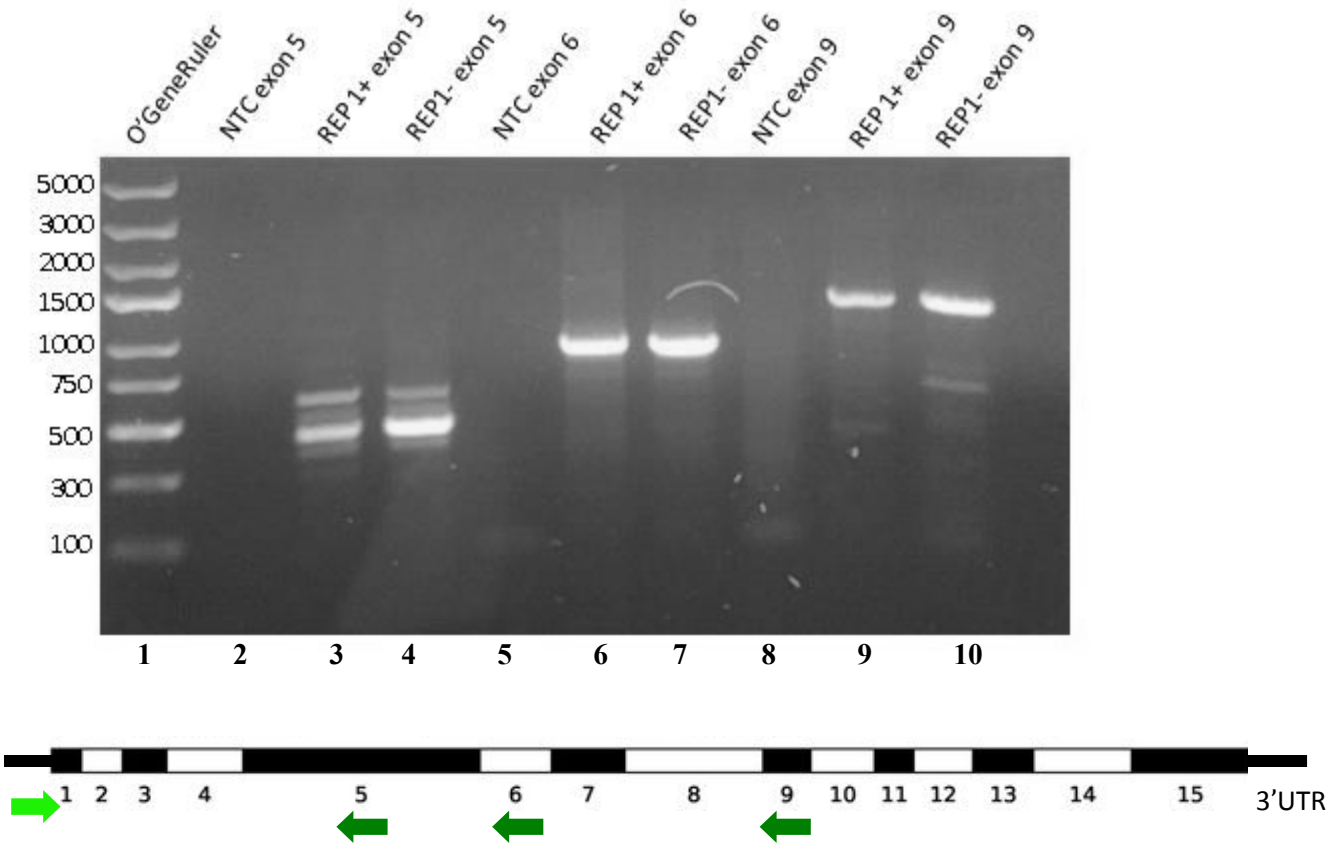


Figure 3.11: Patient C cDNA exons 5, 6 and 9. All appear normal when compared to the REP-1 REP-1 present control. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: negative control exon 1-5; Lane 3: REP-1 present control exon 1-5; Lane 4: Patient C exon 1-5; Lane 5: negative control exon 1-6; Lane 6: REP-1 present control exon 1-6; Lane 7: Patient C exon 1-6; Lane 8: negative control exon 1-9; Lane 9: REP-1 present control exon 1-9; Lane 10: Patient C exon 1-9.

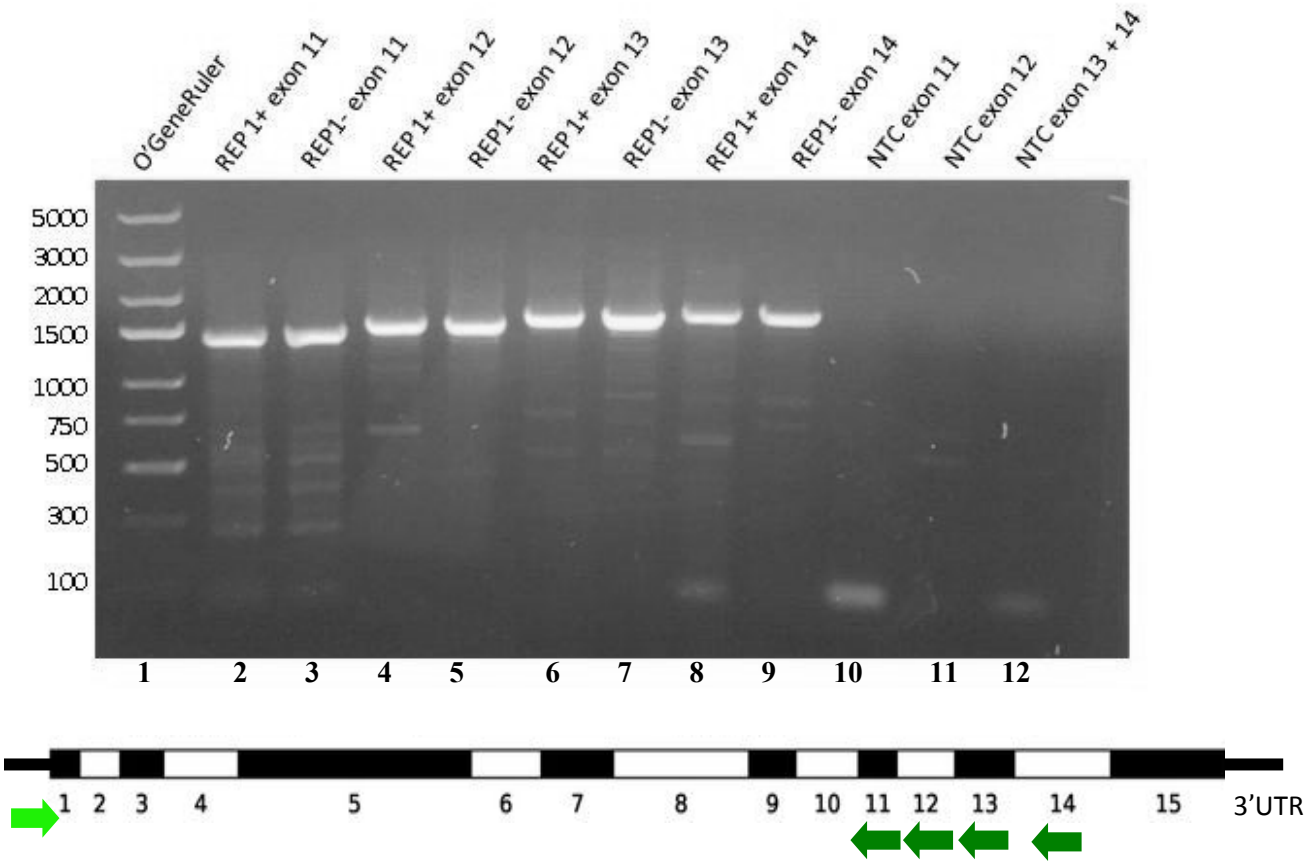


Figure 3.12: Patient C cDNA exons 11, 12, 13 and 14. All appear normal when compared to the REP-1 REP-1 present control. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control exon 1-11; Lane 3: Patient C exon 1-11; Lane 4: REP-1 present control exon 1-12; Lane 5: Patient C exon 1-12; Lane 6: REP-1 present control exon 1-13; Lane 7: Patient C exon 1-13; Lane 8: REP-1 present control exon 1-14; Lane 9: Patient C exon 1-14; Lane 10: negative control exon 1-11; Lane 11: negative control exon 1-12; Lane 12: negative control exon 1-13 and exon 1-14.

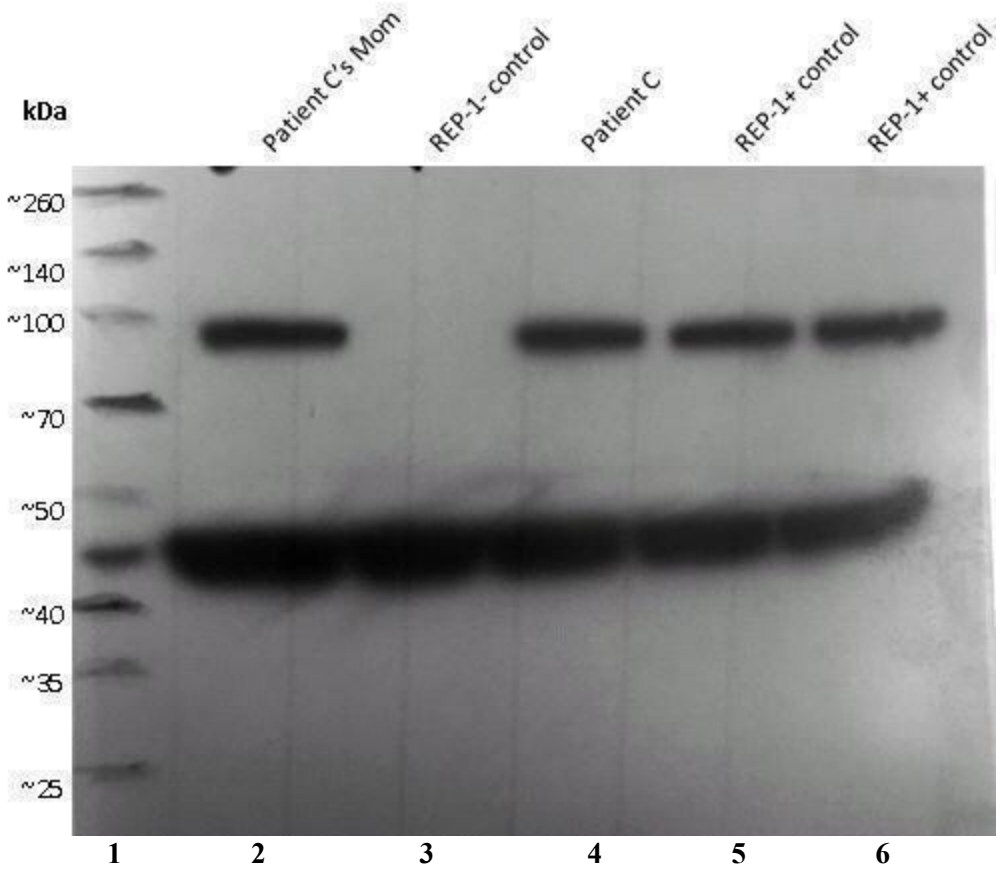


Figure 3.13: Patient C Immunoblot analysis. Patient C expresses REP-1 protein of normal weight. Lane 1: Ladder; Lane 2: Patient C's mother; Lane 3: negative control; Lane 4: Patient C; Lane 5: REP-1 present control; Lane 6: REP-1 present control.

As it was known that novel inherited ocular disorders are being investigated successfully with whole exome sequencing (WES), we chose to send Patient C and his parents to DNALink (San Diego, USA) for WES analysis (Tiwari et al., 2016). We obtained the vcf file in an excel format and then proceeded to filter the variants for possible candidate genes (as outlined in Figure 2.1). Due to the inheritance pattern, an autosomal recessive mode of inheritance seemed most likely, and compound heterozygous mutations were sought. Variants in three candidate genes, *COL18A1*, *PROM1* and *CRB1* were investigated according to American College of Medical Genetics and Genomics (ACMG) guidelines (Richards et al., 2015) (Table 3.1). These variant classification guidelines were decided upon based on expert opinion of a working group of clinical laboratory directors and clinicians. Due to the increased complexity of interpreting variants acquired from new genetic testing technologies, ACMG recommends classifying variants into five standardized categories: "pathogenic", "likely pathogenic", "uncertain significance", "likely benign" and "benign" (Richards et al., 2015). Variants are sorted through information available in population data, bioinformatics predictions, functional studies, familial inheritance and other data that may be applicable. Evidence supporting benign or pathogenic classifications is categorized and a final classification is determined.

The variant annotation software, Alamut ® (Interactive Biosoftware), was used to gather bioinformatic evidence. Variant c.1175T>C in *COL18A1* has a minor allele frequency (MAF) of 0.033. Three out of four software that predict the functional effect of missense changes to proteins predicted this alteration as having no consequence (Align GVGD, SIFT, and MutationTaster), while PolyPhen-2 predicted this variants to be probably damaging (Figures A1 and A2). As three of the four prediction programs suggested benign, I chose to check off "predict neutral" on the guideline forms. Splice prediction using several software algorithms

(SSF, MaxENT, NNSPLICE, GeneSplicer and Human Splice Finder - HSF) showed no changes in splicing (Figure A3). These analyses were summarized according to ACMG guidelines with the variant determined to be of unknown significance (VUS) (Figure A4).

Variant c.4060C>T in the *COL18A1* gene has a recorded MAF of 8% in some ethnicities and is predicted by Align GVGD and SIFT to have no consequence, while MutationTaster and PolyPhen-2 predict this variant to be probably damaging (Figures A5 and A6). Alamut-associated splice prediction programs showed a possibly more favourable splice acceptor (Figure A7). These analyses were summarized according to ACMG guidelines and the variant determined to benign (Figure A8).

Variant c.868A>C in *PROM1* has a minor allele frequency (MAF) of 0.004, predicted as tolerated by AlignGVGD, SIFT and MutationTaster, while PolyPhen-2 predicted this variant to be possibly damaging (Figures A9 and 10). Alamut-associated splice prediction programs showed a loss of splice acceptor site (Figure A11). These analyses were summarized according to ACMG guidelines and the variant was determined to be of unknown significance (VUS) (Figure A12).

Variant c.1559C>T in the *PROM1* gene has no reported MAF and is predicted by all four software to have no consequence (Figures A13 and A14). Alamut-associated splice prediction programs showed no favourable changes to splicing (Figure A15). These analyses were summarized according to ACMG guidelines and the variant determined to be of unknown significance (VUS) (Figure A16).

The variant c.3202A>G in *CRB1* has been previously reported three times in ClinVar in patients with pigmented paravenous chorioretinal atrophy, Leber congenital amaurosis and recessive retinitis pigmentosa. Only one variant was investigated in the gene because the

variants filtered in the other parent were common SNPs but listed in the variant report with incorrect reference numbers. Align GVGD, SIFT and PolyPhen-2 predict this variant as having no consequence, while MutationTaster classifies it as disease causing (Figures A17 and A18). Alamut splice predictor showed no favourable splicing changes (Figure A19). These analyses were summarized according to ACMG guidelines and the variant determined to be of unknown significance (VUS) (Figure A20).

Gene	Variant	MAF	SIFT	Polyphen-2	Align GVGD	MutationTaster	Predicted splicing changes	Classification
COL18A	c.1175T>C	0.033	Tolerated	Probably damaging	No consequence	Polymorphism	None	Unknown significance
COL18A	c.4060C>T	0.08	Tolerated	Probably damaging	No consequence	Disease causing	More favourable splice acceptor	Benign
PROM1	c.868A>C	0.004	Tolerated	Possibly damaging	No consequence	Polymorphism	Loss of splice acceptor site	Unknown significance
PROM1	c.1559C>T	None	Tolerated	Benign	No consequence	Polymorphism	None	Unknown significance
CRB1	c.3202A>G	None	Tolerated	Benign	No consequence	Disease causing	None	Unknown significance

Table 3.1 Summary of variants. Candidate variants and information used in classification of the variants as benign, pathogenic or unknown significance.

3.4 Patient D

This patient sample came to us from Finland. The patient has a reported family history of vision loss. As choroideremia was suspected, the MacDonald laboratory performed immunoblot analysis and he was confirmed to have a lack of REP-1 protein (data not shown). To determine the molecular basis for this result, sequencing of coding regions of the *CHM* gene along with copy number analysis using multiplex ligation-dependent probe amplification (MLPA) was performed (Figure 3.14). These analyses did not find any pathogenic variants. As this patient had also agreed to have a fibroblast cell line produced from a biopsy, RNA was extracted and cDNA was produced. Again, a similar approach was taken to examine for splice alterations – PCR amplicons were "walked" along the exons. The PCR amplicon from exon 1 through to the exon 1/exon 2 boundary was not present and the full-length transcript (exon 1 through to the 3'UTR) appeared weaker and of longer length (Figure 3.15). Exon walking of the transcript similar to patients A and B, revealed no normal PCR products in Patient D (Figure 3.16). Given that all individual exons did not appear to be present and no REP-1 protein was produced, I hypothesized that this patient could have a regulatory mutation such as a point mutation, insertion, deletion or other gross rearrangement in the non-coding region of the gene. Therefore, using gDNA derived from the same patient cell line, I started analysis upstream of the *CHM* translational start codon. The core promoter region in this patient, spanning c.-119 to c.-79 (Radziwon et al., 2017) had just been determined by the MacDonald laboratory to be free of mutations. Therefore there was potential for a novel regulatory element to be altered further upstream. Normal-sized PCR product was observed 300 bp upstream of the transcriptional start; however, PCR product was not amplified at a location 2 kb upstream (Figure 3.17). Looking internal to those two primer sets, a normal amplicon was produced at 500 bp upstream yet no

product was produced 1 kb upstream (Figure 3.18). Suspecting a large deletion, multiple primer pairs that served as short probes were designed to investigate the deletion's 5' boundary (Figure 3.19): primers 286 and 287 which are 100 kb upstream of *CHM* in the neighbouring gene, *DACH2*, primers 284 and 285 which are 50 kb upstream and primers 282 and 283 which are 25 kb upstream. PCR analysis showed the primers produced amplicons of the same size and amount as those from a REP-1 present control (Figure 3.20). This was repeated with primers 288 and 289 at 10 kb upstream, primers 290 and 291 at 15 kb upstream and primers 292 and 293 at 20 kb upstream, which again demonstrated amplification of products identical to the REP-1 present control (Figure 3.21). Two more primer pairs were designed: 294 and 295 were 7.5 kb upstream and primer pair 296 and 297 were 8.5 kb upstream. Again, there was no difference between the REP-1 present control sample and Patient D (Figure 3.22). To investigate whether a deletion including exon 1 through to 7 kb upstream was occurring, a long-range PCR was completed. Failure to produce a band of similar strength to the REP-1 present control at a smaller size minimized the likelihood that the mutation in this patient was a deletion (Figure 3.23). The PCR results thus far could only be explained by one of two hypotheses (Figure 3.24). To investigate whether an inversion took place, pairs of forward primers in the area of interest were used. Several forward primer pairings amplified the region of interest from patient DNA, however as expected in the REP-1 present control no amplification occurred (Figure 3.25). Schematics of primer locations and PCR product results in REP-1 present control and our patient are summarized in Table 3.3. Through further amplifications it was determined that the upstream region of the gene as well as the first exon of the *CHM* gene was involved in an inversion (Figure 3.26). This novel inversion was validated through sequencing and has been named c.-839_49+5528inv (Figure 3.27). A diagnostic multiplex PCR was created in order to

screen future choroideremia patients and was demonstrated to correctly identify this variant in Patient D (Figure 3.28). This assay was shown to be specific to this inversion as it did not create false positives in multiple REP-1 present control samples (Figure 3.29). Although a female carrier was not available to run with this assay, it is expected that a female carrier would produce two identifiable bands representing both the normal and inversion allele.

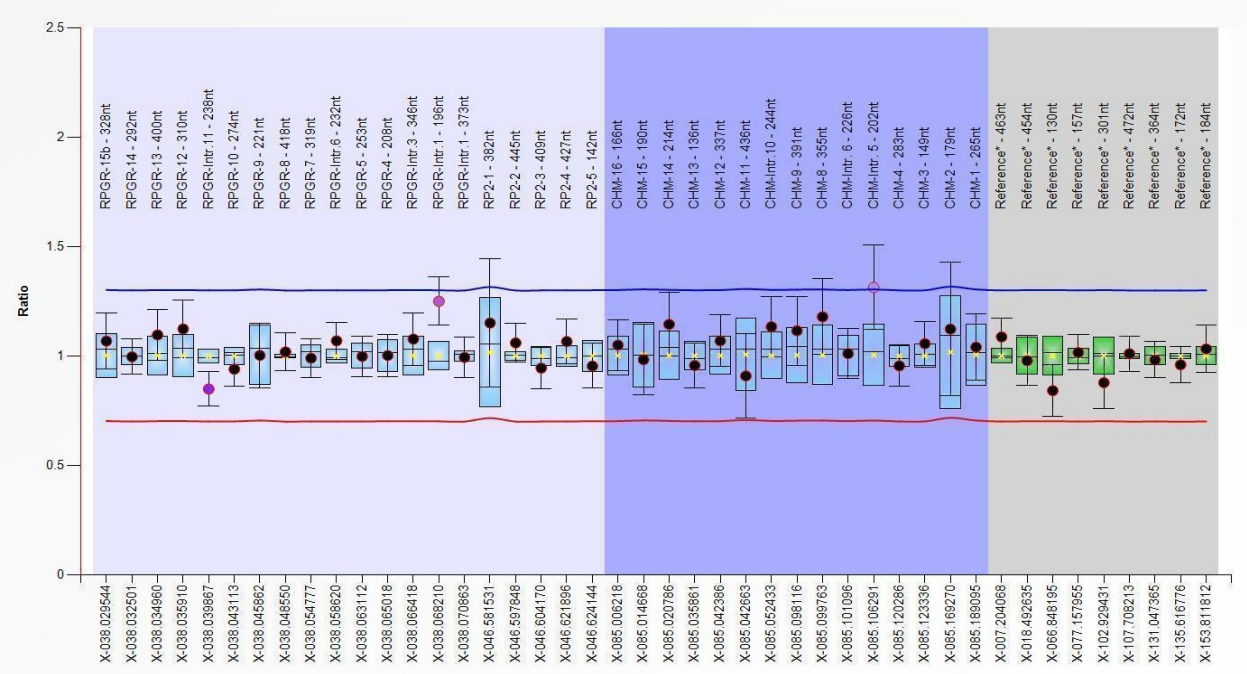


Figure 3.14: Copy number variant analysis in Patient D. Ratio of markers in RP associated genes (light purple), choroideremia coding regions (blue) and reference genes (grey). No fold change in ratio, therefore no deletions or duplications present in the coding region of Patient C.

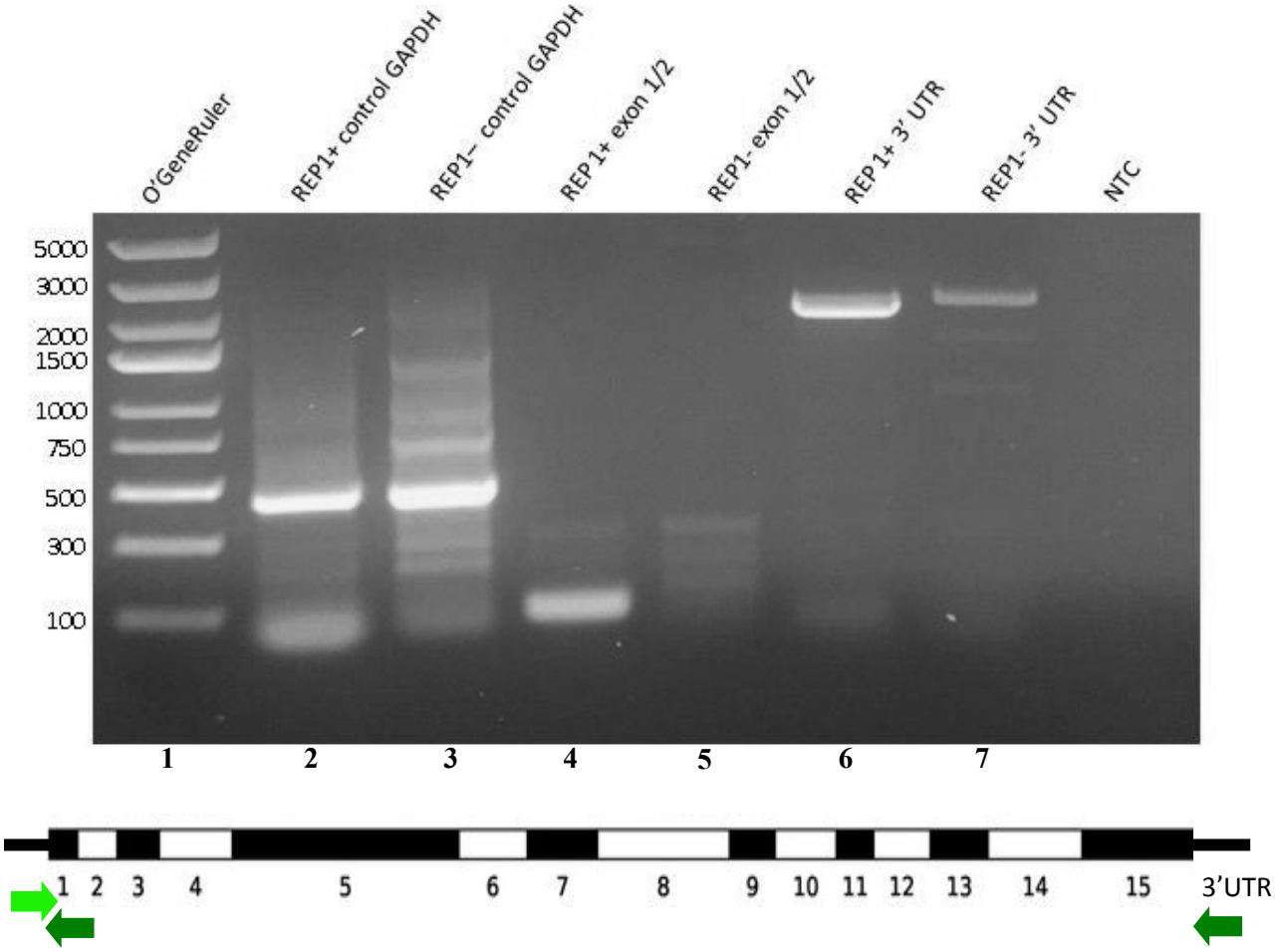


Figure 3.15: Patient D full length PCR analysis of REP-1 transcript. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present GAPDH positive control; Lane 3: Patient D GAPDH positive control; Lane 4: REP-1 present control exon 1; Lane 5: Patient D exon 1; Lane 6: REP-1 present control exon 1-3' UTR; Lane 7: Patient D exon 1-3'UTR.

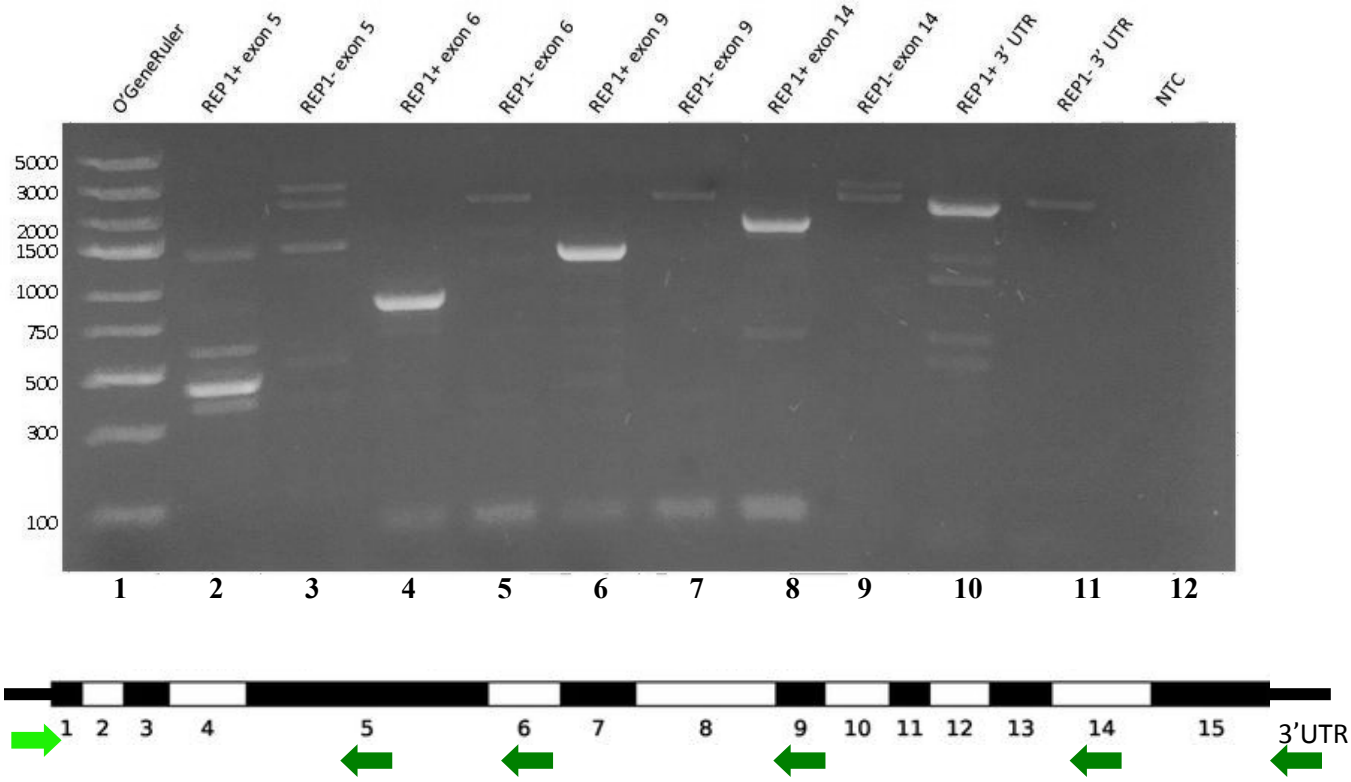


Figure 3.16: Patient D cDNA PCR analysis of exons 5, 6, 9, 14 and 3'UTR. All appear normal when compared to the REP-1 present control. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control exon 1-5; Lane 3: Patient D exon 1-5; Lane 4: REP-1 present control exon 1-6; Lane 5: Patient D exon 1-6; Lane 6: REP-1 present control exon 1-9; Lane 7: Patient D exon 1-9; Lane 8: REP-1 present control exon 1-14; Lane 9: Patient D exon 1-14; Lane 10: negative control exon 1-3'UTR; Lane 11: negative control exon 1-3'UTR; Lane 12: no template control.

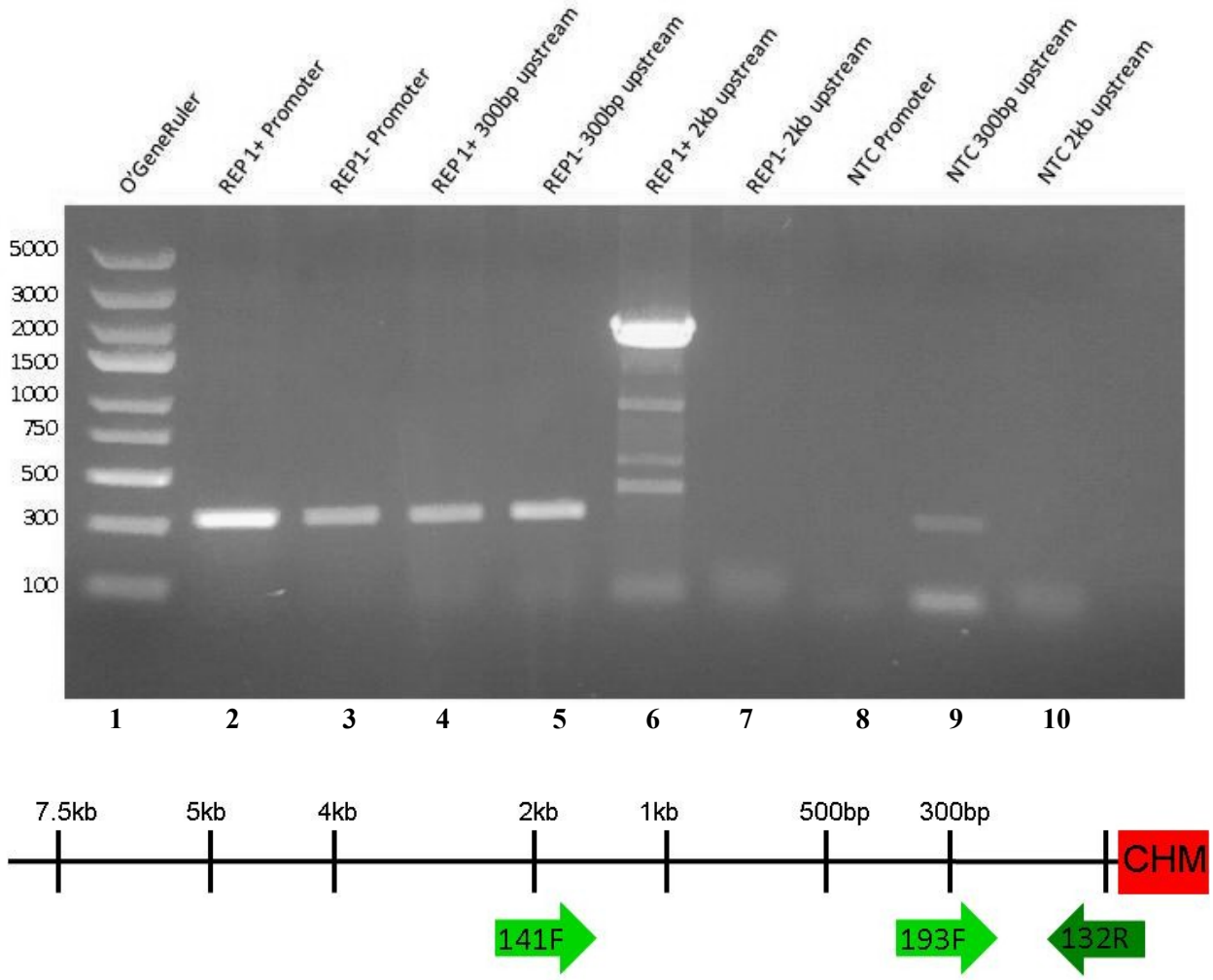


Figure 3.17: 300 bp and 2 kb upstream analysis of *CHM* gene in Patient D. The promoter region and 300 bp upstream appear normal while a complete loss of band is observed at 2 kb upstream. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control promoter region; Lane 3: Patient D promoter region; Lane 4: REP-1 present control 300 bp upstream; Lane 5: Patient D 300 bp upstream; Lane 6: REP-1 present control 2 kb upstream; Lane 7: Patient D 2 kb upstream; Lane 8: negative control promoter region; Lane 9: negative control 300 bp upstream; Lane 10: negative control 2 kb upstream.

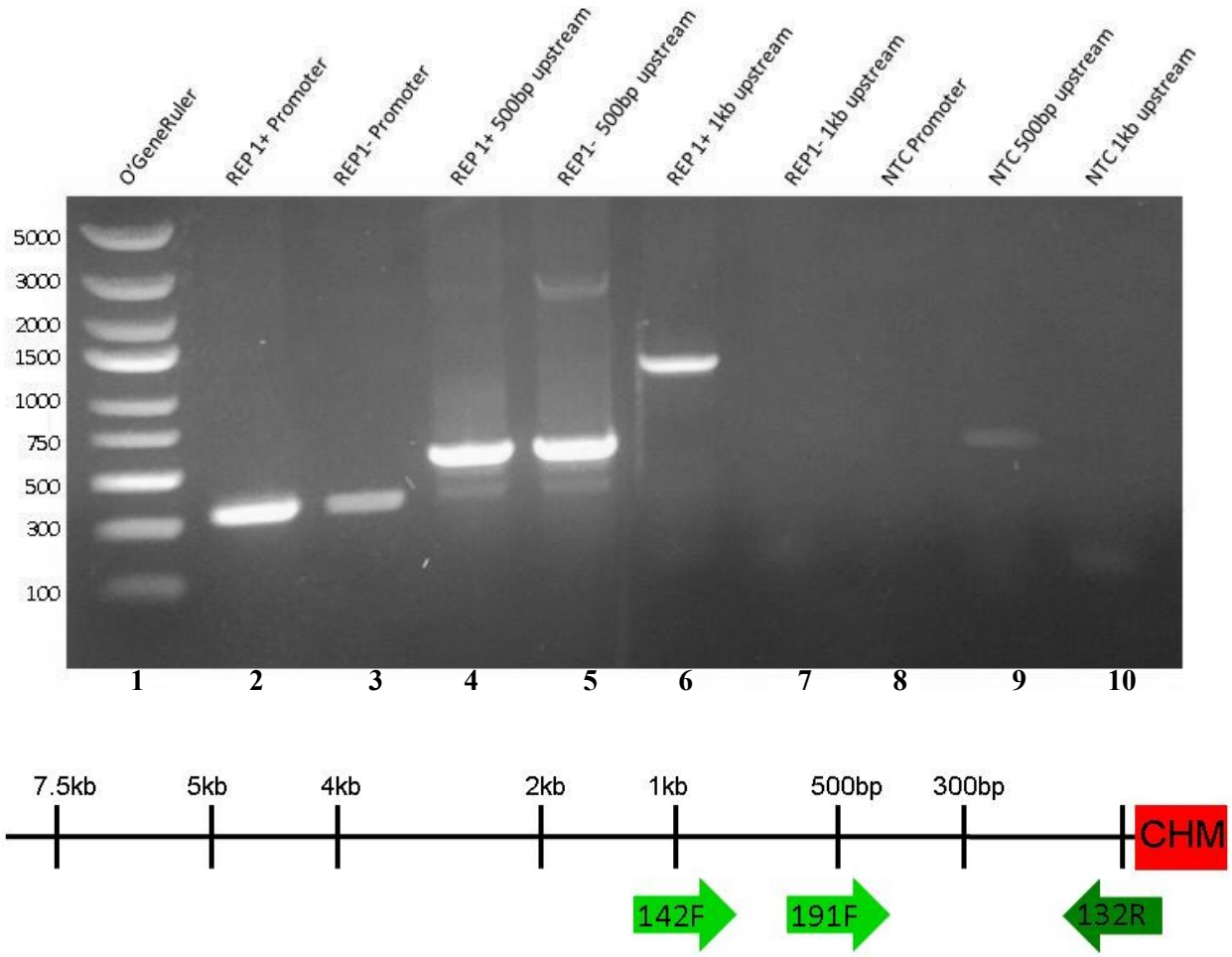


Figure 3.18: 500 bp and 1 kb upstream analysis of *CHM* gene in Patient D. The promoter region and 500 bp upstream appear normal while a complete loss of band is observed at 1 kb upstream. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control promoter region; Lane 3: Patient D promoter region; Lane 4: REP-1 present control 500 bp upstream; Lane 5: Patient D 500 bp upstream; Lane 6: REP-1 present control 1 kb upstream; Lane 7: Patient D 1 kb upstream; Lane 8: negative control promoter region; Lane 9: negative control 500 bp upstream; Lane 10: negative control 1 kb upstream.

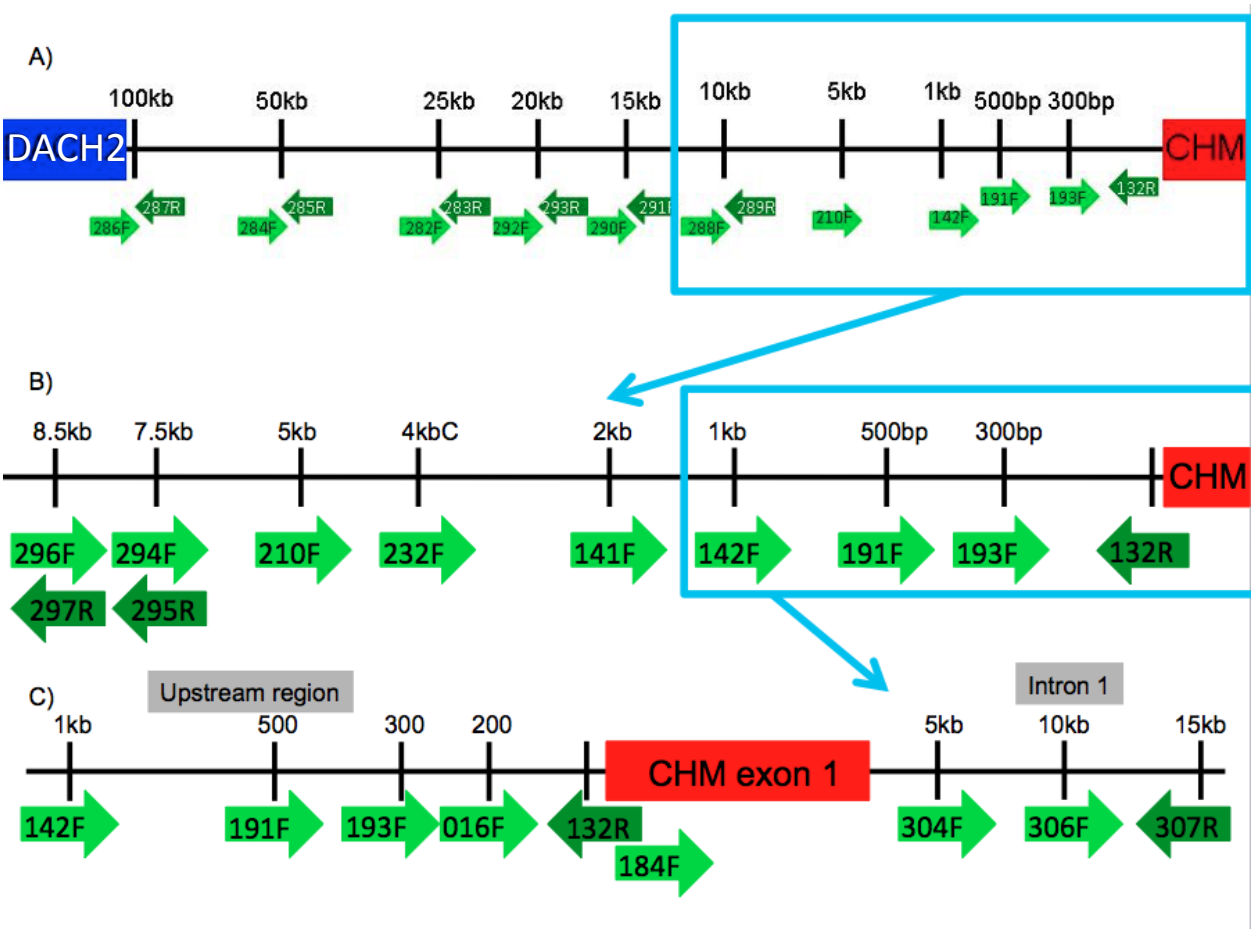


Figure 3.19: Primer locations. Schematic of primer locations used in upstream PCR analysis and inversion investigation of Patient D. A) Locations of primers between *CHM* and upstream neighbouring gene DACH2, B) a closer look at primer locations upstream of *CHM* and C) primer locations upstream of *CHM* and in the first intron.

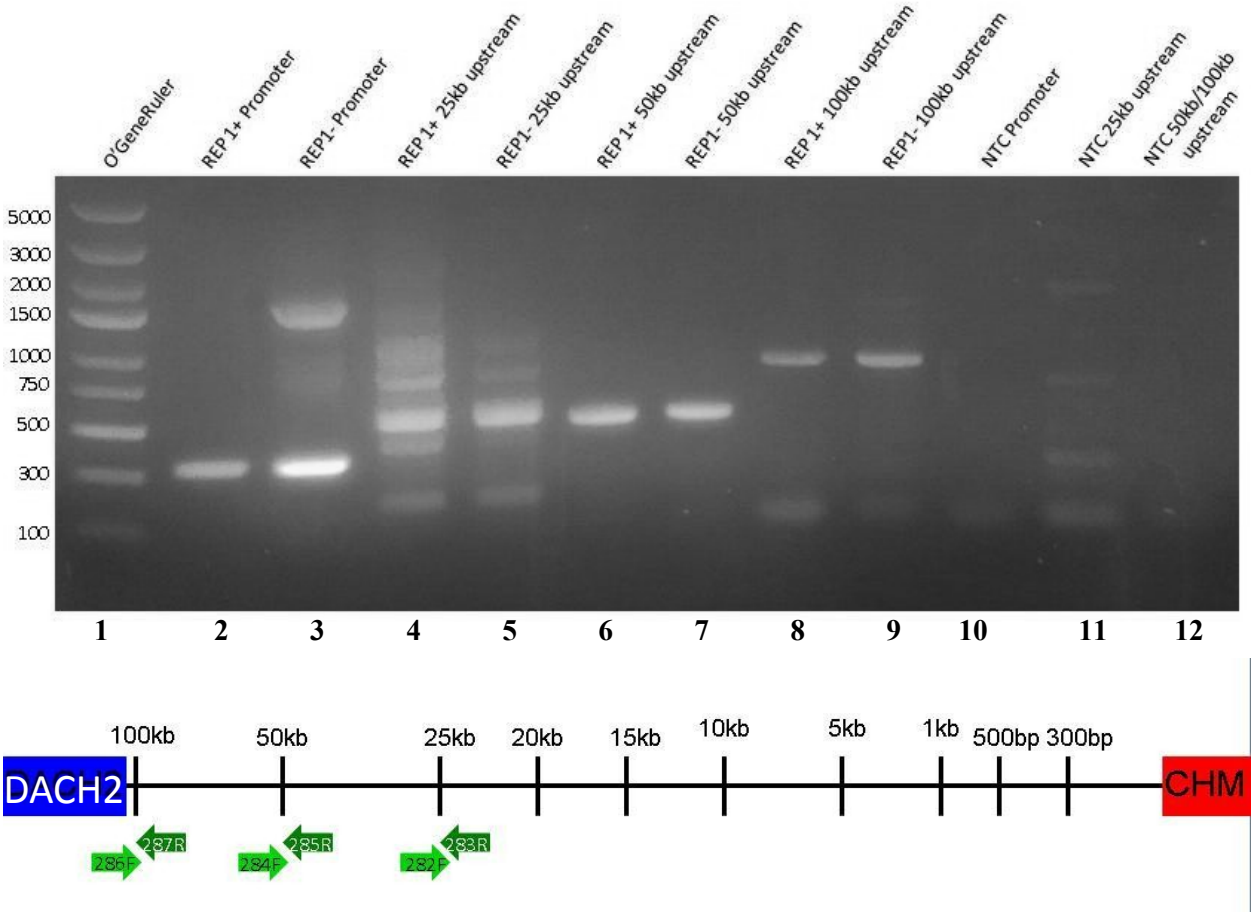


Figure 3.20: 25 kb, 50 kb and 100 kb upstream analysis of *CHM* gene in Patient D. The promoter region, 25 kb, 50 kb and 100 kb all appear normal compared to REP-1 REP-1 present individual. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control promoter region; Lane 3: Patient D promoter region; Lane 4: REP-1 present control 25 kb upstream; Lane 5: Patient D 25 kb upstream; Lane 6: REP-1 present control 50 kb upstream; Lane 7: Patient D 50 kb upstream; Lane 8: REP-1 present control 100 kb upstream; Lane 9: Patient D 100 kb upstream; Lane 10: negative control promoter region; Lane 11: negative control 25 kb upstream; Lane 12: negative control 50 kb/100 kb upstream.

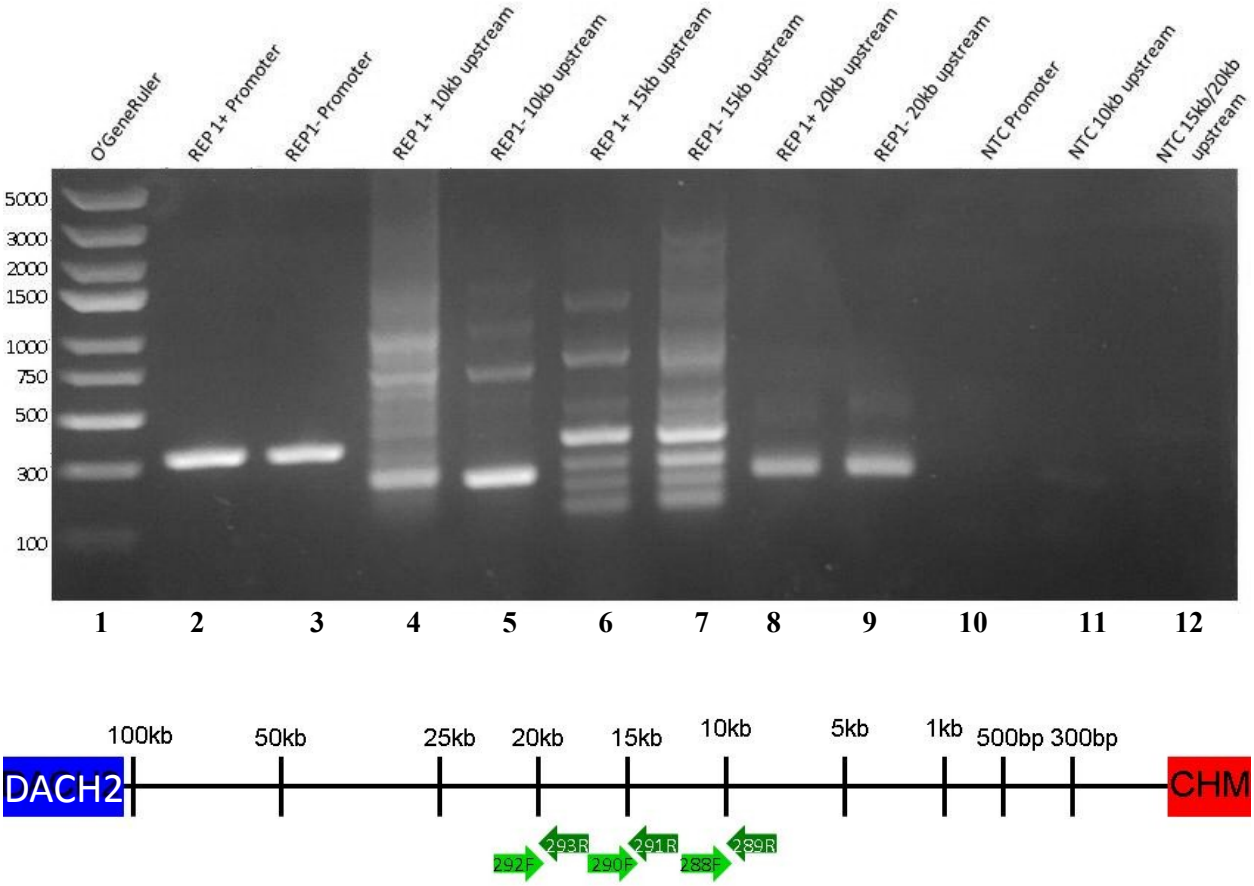


Figure 3.21: 10 kb, 15 kb and 25 kb upstream analysis of *CHM* gene in Patient D. The promoter region, 10 kb, 15 kb and 20 kb all appear normal compared to REP-1 present individual. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control promoter region; Lane 3: Patient D promoter region; Lane 4: REP-1 present control 10 kb upstream; Lane 5: Patient D 10 kb upstream; Lane 6: REP-1 present control 15 kb upstream; Lane 7: Patient D 15 kb upstream; Lane 8: REP-1 present control 20 kb upstream; Lane 9: Patient D 20 kb upstream; Lane 10: negative control promoter region; Lane 11: negative control 10 kb upstream; Lane 12: negative control 15 kb/20 kb upstream.

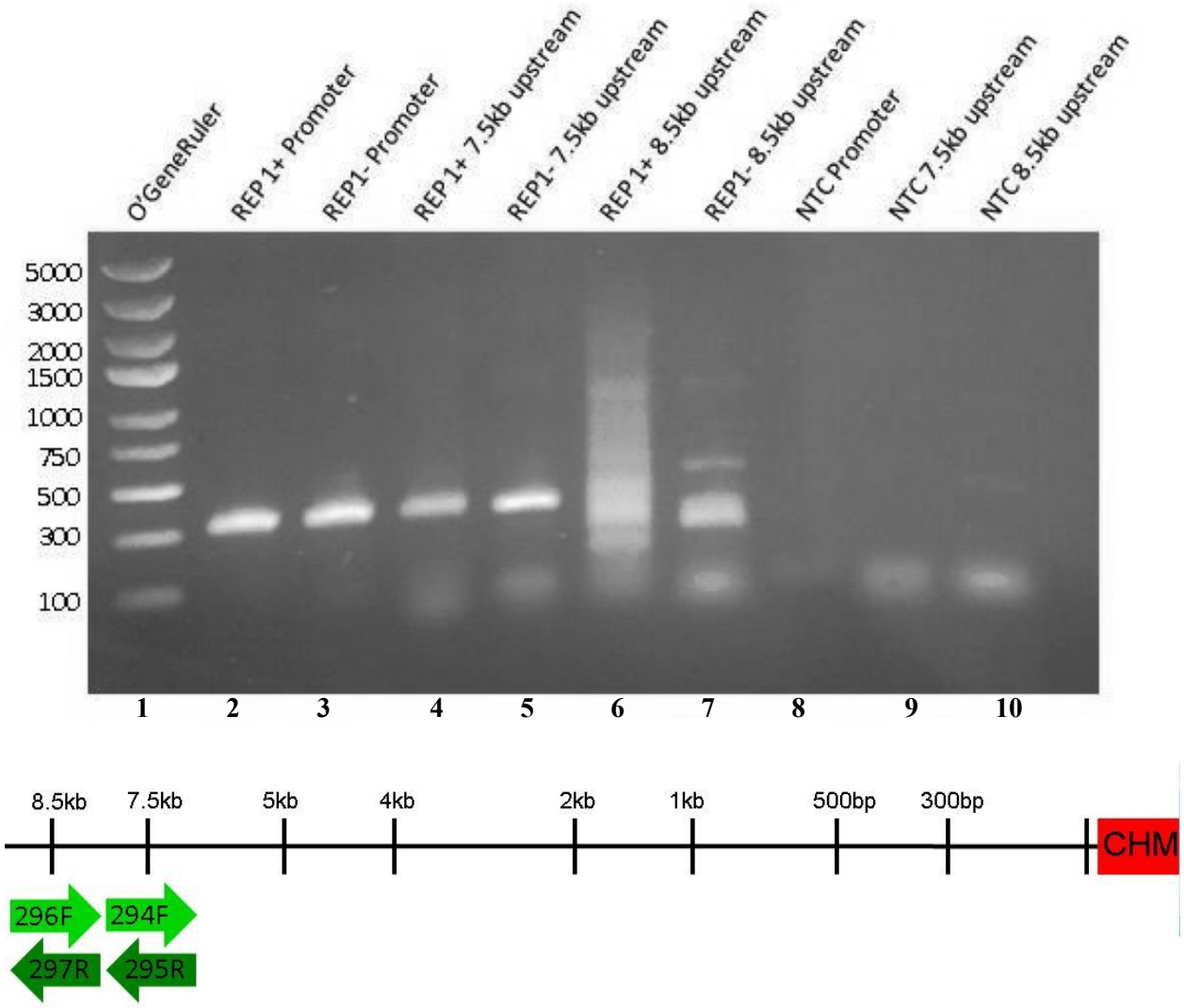


Figure 3.22: 7.5 kb and 8.5 kb upstream analysis of *CHM* gene in Patient D. The promoter region, 7.5 kb and 8.5 kb appear normal compared to REP-1 present individual. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: REP-1 present control promoter region; Lane 3: Patient D promoter region; Lane 4: REP-1 present control 7.5 kb upstream; Lane 5: Patient D 7.5 kb upstream; Lane 6: REP-1 present control 8.5 kb upstream; Lane 7: Patient D 8.5 kb upstream; Lane 8: negative control promoter region; Lane 9: negative control 7.5 kb upstream; Lane 10: negative control 8.5 kb upstream.

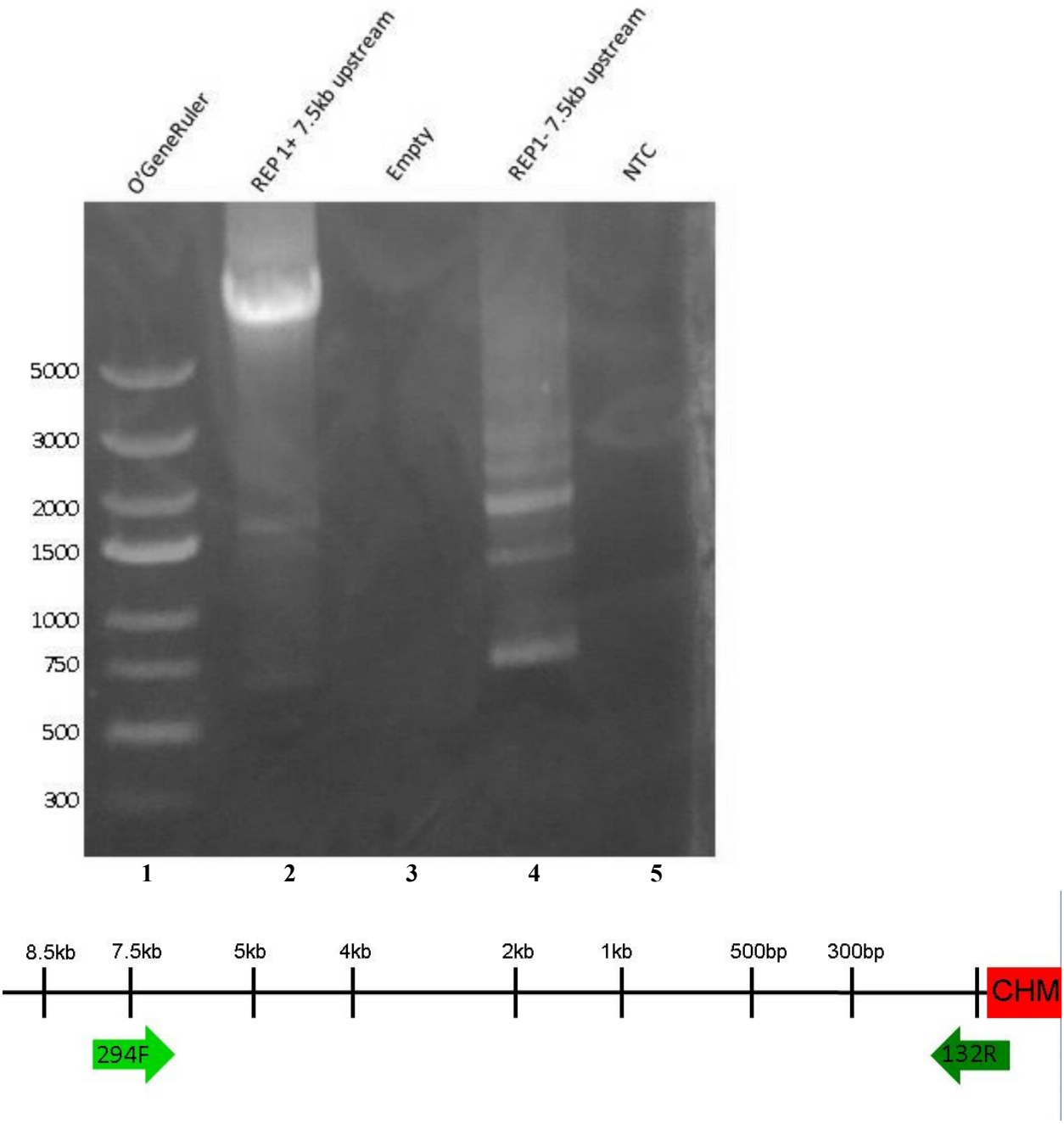


Figure 3.23: Long-range PCR. Long-range PCR to check for a deletion upstream of *CHM* gene in patient D. No product was produced in patient sample, therefore unable to confirm deletion. Lane 1: Thermo Fisher Scientific O'GeneRuler; Lane 2: REP-1 present control 7.5 kb upstream; Lane 3: Empty, nothing loaded; Lane 4: Patient D 7.5 kb upstream; Lane 5: negative control 7.5 kb upstream.

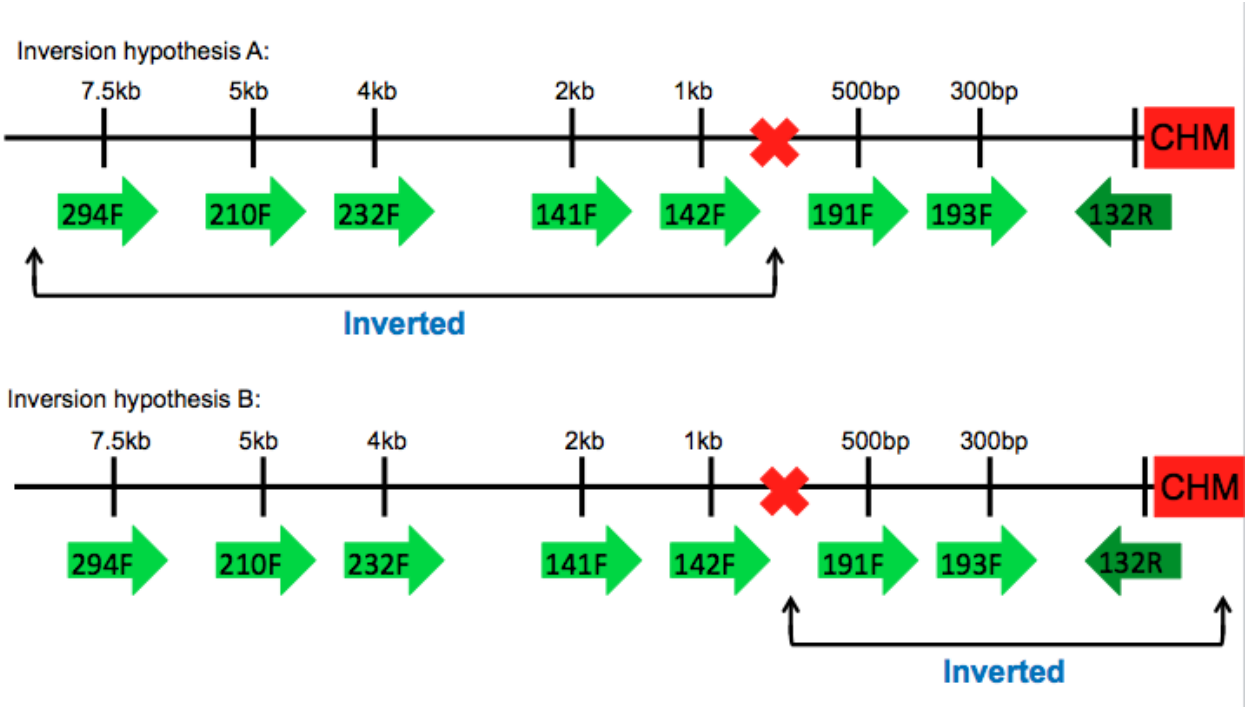


Figure 3.24: Inversion hypotheses. Schematic of two possible inversions occurring in Patient D based on PCR result thus far.

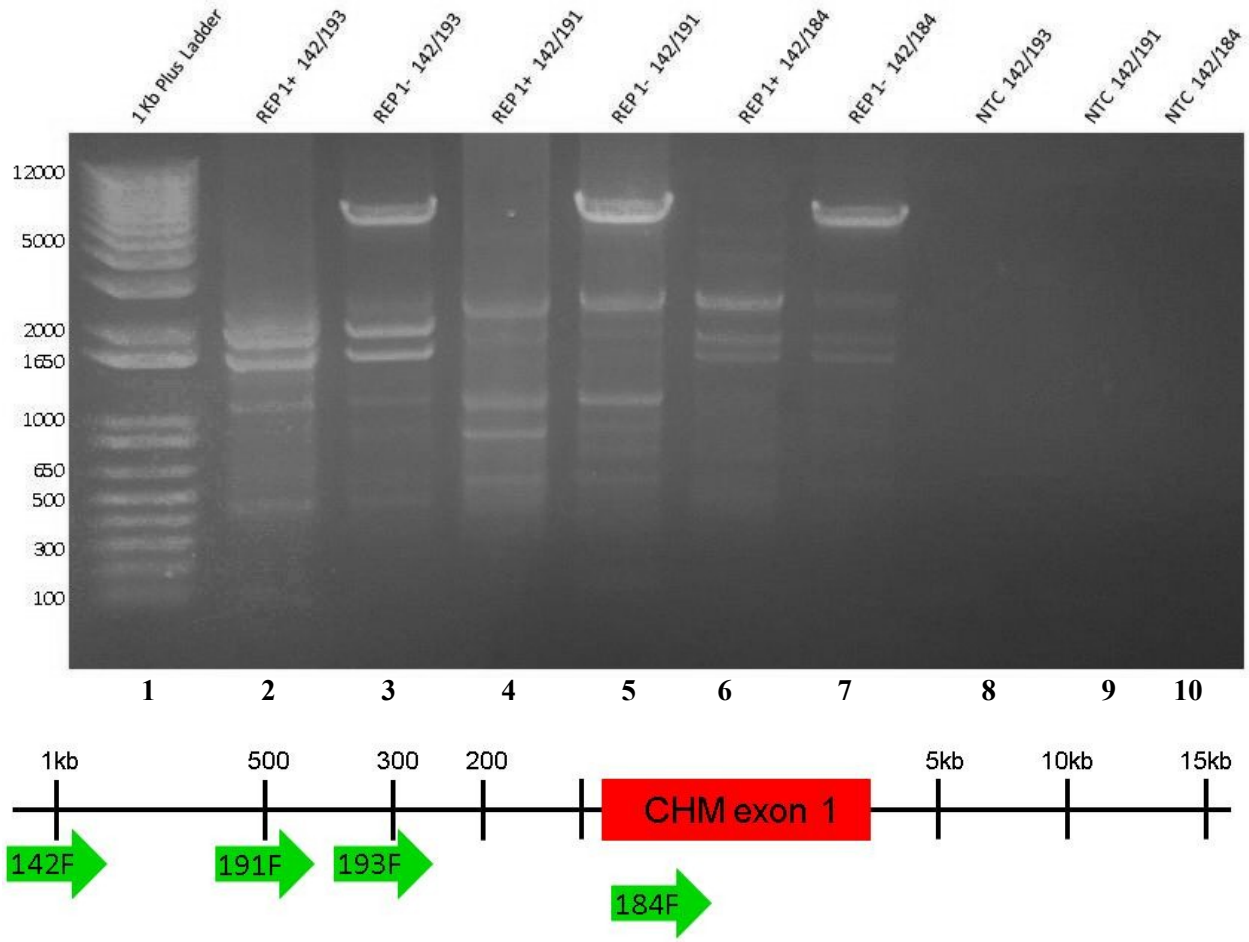


Figure 3.25: PCR using pairings of two forward primers. No amplification in REP-1 present individual but amplification in Patient D, suggests inversion upstream of *CHM* gene. Lane 1: 1 kb Plus Ladder; Lane 2: REP-1 present control with primers 142 and 193; Lane 3: Patient D with primers 142 and 193; Lane 4: REP-1 present control with primers 142 and 191; Lane 5: Patient D with primers 142 and 191; Lane 6: REP-1 present control with primers 142 and 184; Lane 7: Patient D with primers 142 and 184; Lane 8: negative control with primers 142 and 193; Lane 9: negative control with primers 142 and 191; Lane 10: negative control with primers 142 and 184.

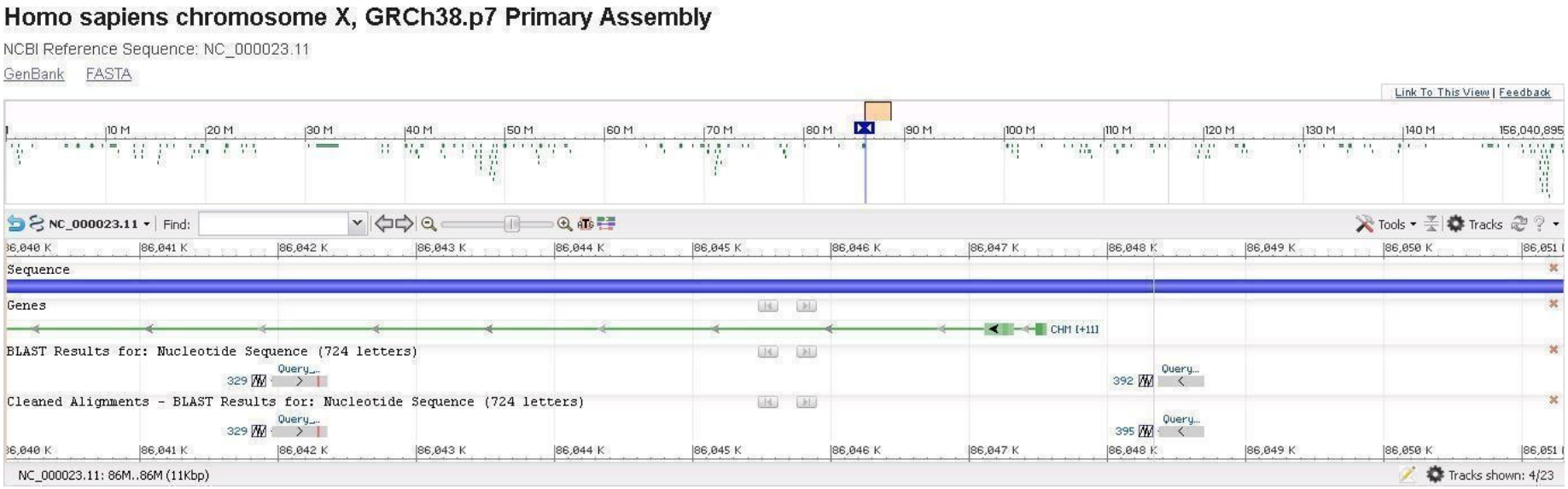


Figure :3.26. Patient D inversion sequence alignment. Sequence alignment of PCR product of inversion c.-839_49+5528inv in Patient D.

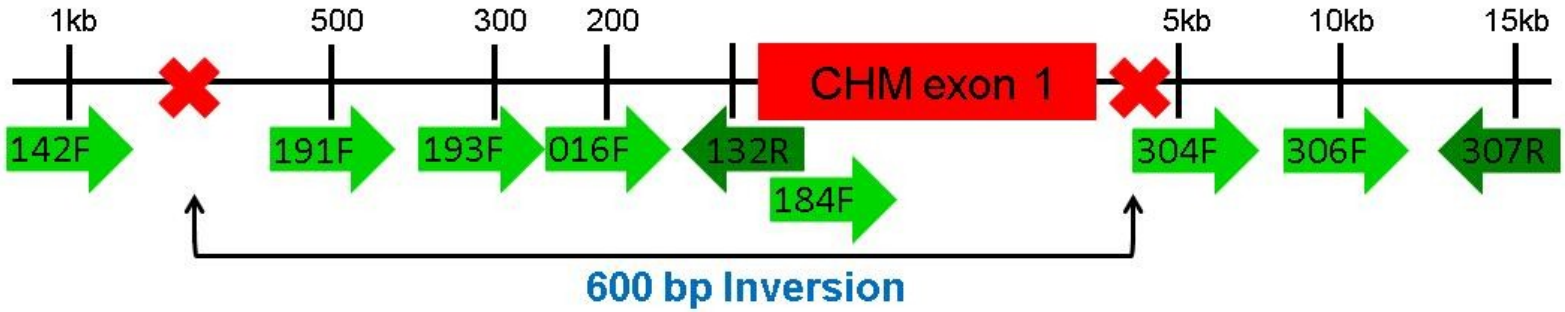


Figure 3.27: Inversion c.-839_49+5528inv. Schematic REP-1 presentation of inversion c.-839_49+5528inv breakpoints and primer locations in *CHM* gDNA.

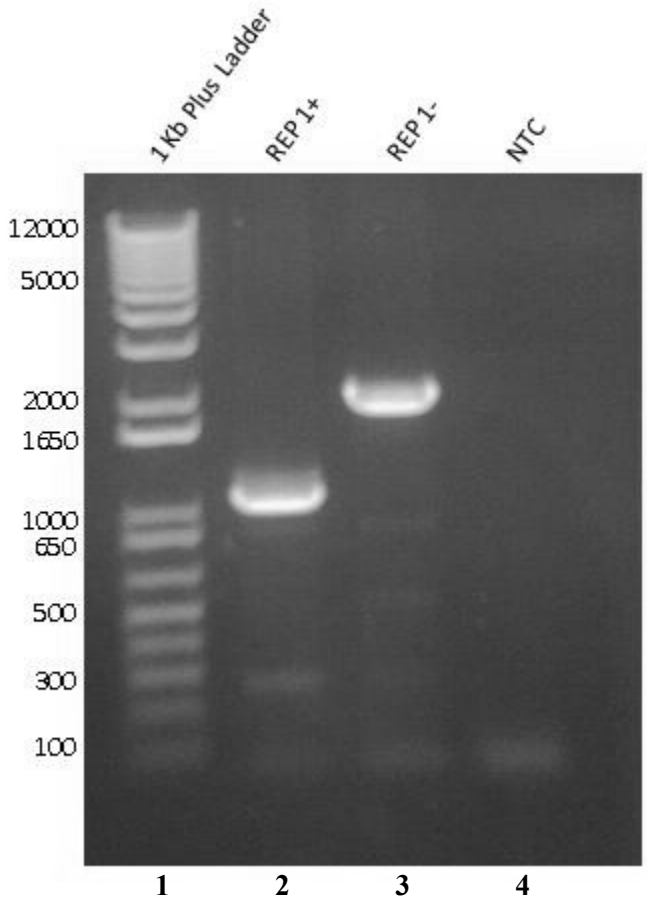


Figure 3.28: Inversion diagnostic multiplex PCR. Diagnostic multiplex PCR on REP-1 present control patient and Patient D. Lane 1: 1 kb Plus Ladder; Lane 2: REP-1 present control with primers 142, 017, 132 and 315; Lane 3: Patient D with primers 142, 017, 132 and 315; Lane 4: negative control.

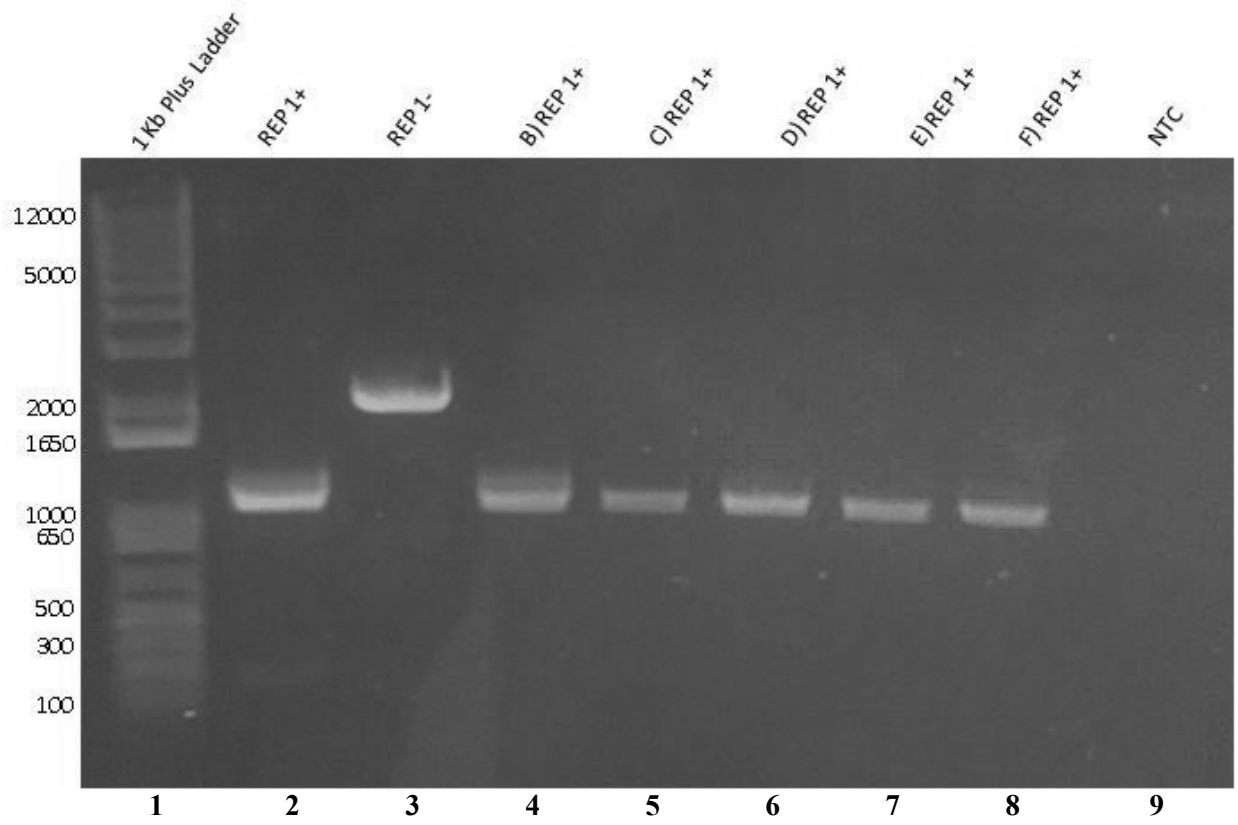
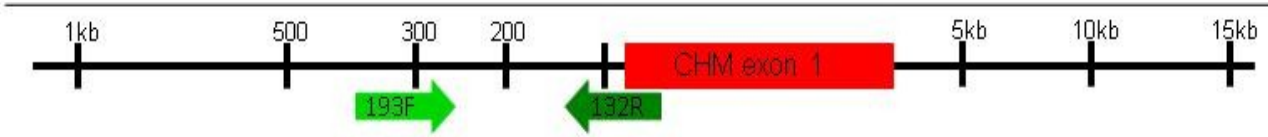
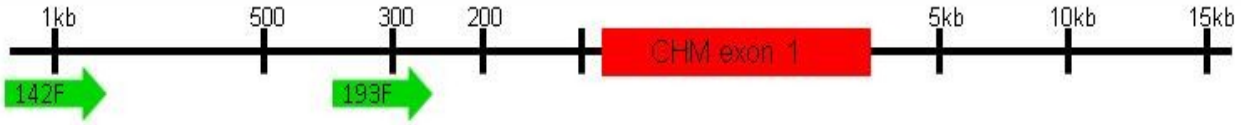
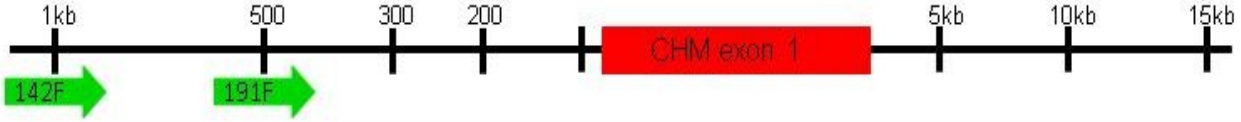
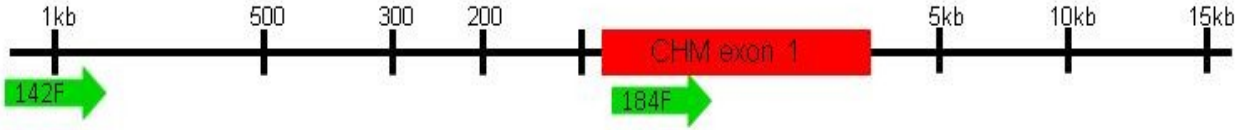


Figure 3.29: Diagnostic multiplex PCR. Diagnostic multiplex PCR with primers 142, 017, 132 and 315 for mutation c.-839_49+5528inv completed on patient D carrying the inversion mutation and six REP-1 present control individuals. Lane 1: 1 kb Plus Ladder; Lane 2: REP-1 present control; Lane 3: Patient D; Lane 4: individual without choroideremia; Lane 5: individual without choroideremia; Lane 6: individual without choroideremia; Lane 7: individual without choroideremia; Lane 8: individual without choroideremia; Lane 9: no template negative control.

Patient	Age	Refractive error	Vision	Family history	Ancestry
Patient A	39	(OD) -4.00 (OS) -4.00	(OD) 20/32 (OS) 20/32	Reported, mother confirmed carrier	Korean
Patient B	30	(OD) -3.50 (OS) -4.50	(OD) 20/30 (OS) 20/40	Reported, sister confirmed carrier	Hungarian
Patient C	16	No glasses	(OD) 20/400 (OS) 20/150	None reported	First Nations
Patient D	67	(OD) +0.75 (OS) +2.50	(OD) 20/32 (OS) 20/63	Reported, mother confirmed carrier	Finnish

Table 3.2 Patient summary. Patient information including phenotypic information when study started: age, refractive error, vision, reported family history and ancestry.

Table 3.3: Inversion primer locations and results. Primer positions and PCR result for Patient D inversion analysis.

Primers Used	REP-1 +	Patient D
	Band	Band
	No band	Band
	No band	Band
	No band	Band
	No band	Band

3.5 X-Inactivation

Due to the inheritance pattern of *Chm*, heterozygous null females are mosaic for REP-1 present and REP-1 absent cells (Tolmachova et al., 2006). It is estimated that 50% of cells are expressing the knockout version of the gene, but the expression of the REP-1 protein from the normal allele could vary due to random X-inactivation. Such a difference in REP-1 expression could lead to variable results when REP-1 function was studied using a biochemical assay (ie. some heterozygous females may have increased function while others may show very little function depending on their random *CHM* inactivation). In order to correlate REP-1 expression due to X-inactivation patterns with the biochemical assay currently being performed in the Posse de Chavez laboratory, a quantitative X-inactivation assay was developed. This assay relies on measuring the amount of fluorescence from primers designed to flank the region of interest in the transcript. The amount of fluorescently labelled PCR product amplified in the region correlates to the level of expression of the specific X chromosome transcript as the primers incorporated exon 4, which is missing in the *Chm* null mouse (Figure 3.50). RNA extracted from ear punch samples from heterozygous mice was converted to cDNA and PCR analysis was completed to ensure proper primer specificity (Figure 3.51). PCR products dilutions of 50% and 20% were injected and electrophoresis was measured with peaks at expected lengths for both the wildtype and Chm^{3lox} alleles (Figure 3.52). The wildtype allele (using the primer located in exon 4) will show a peak at approximately 85 bp and the Chm^{3lox} allele (using the primer located in exon 5) at 95 bp and in the wildtype allele at 220 bp. Two sets of primers were designed to produce fragments of similar length to account for PCR amplification bias. PCR reactions could be combined into one multiplex PCR reaction in the future using the primer set with the reverse primer in exon 5. This would require taking into consideration the PCR amplification bias for

the smaller, knockout fragment (excluding exon 4) amplifying at a higher efficiency than the longer, wildtype allele product (including exon 4).

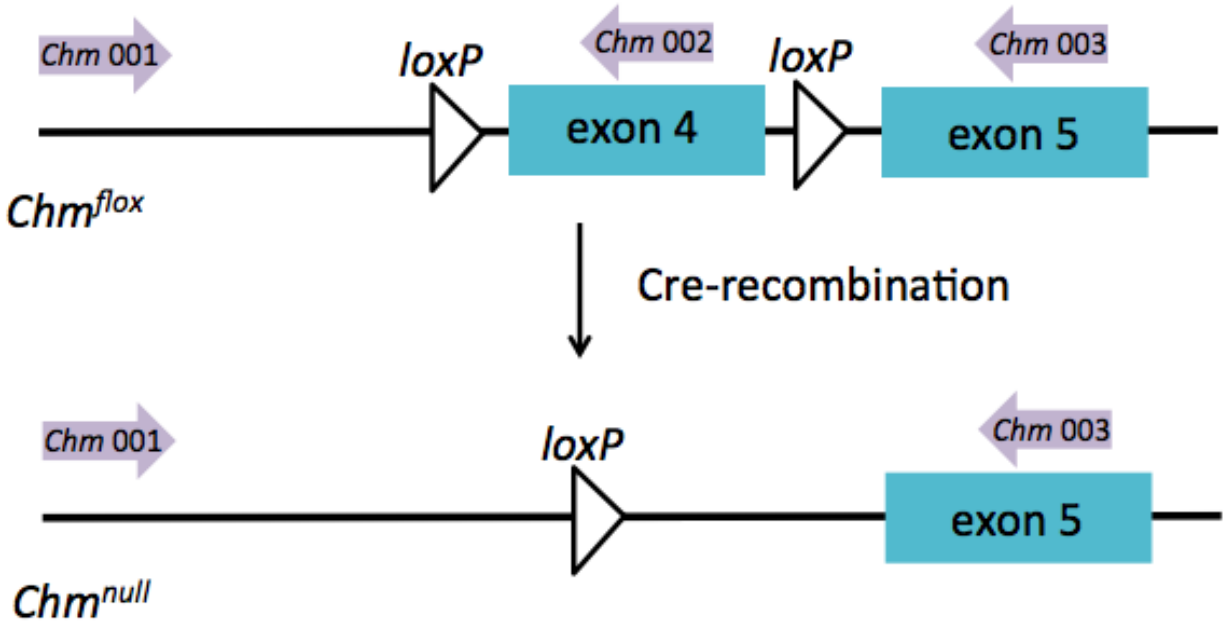


Figure 3.30: Schematic representation of *Chm* transcripts. Location of *loxP* sites in *Chm*^{3lox} mouse transcript and primers used for the X-inactivation assay.

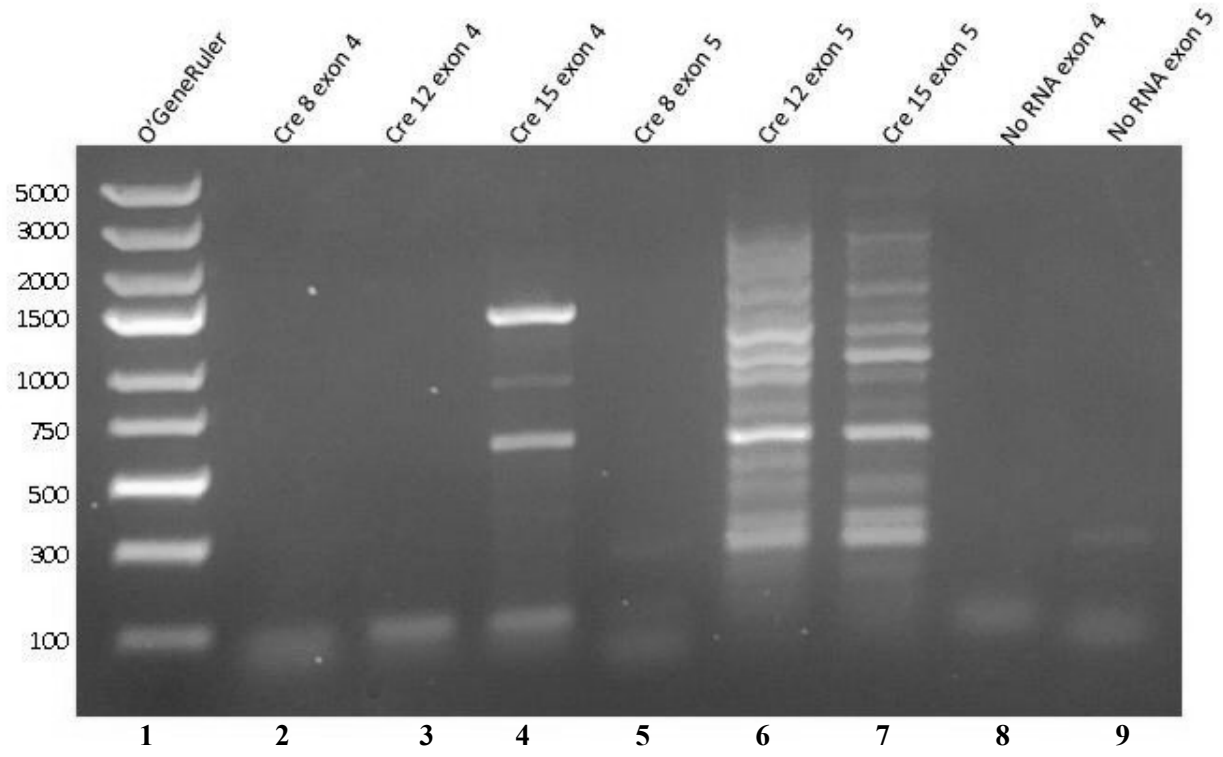
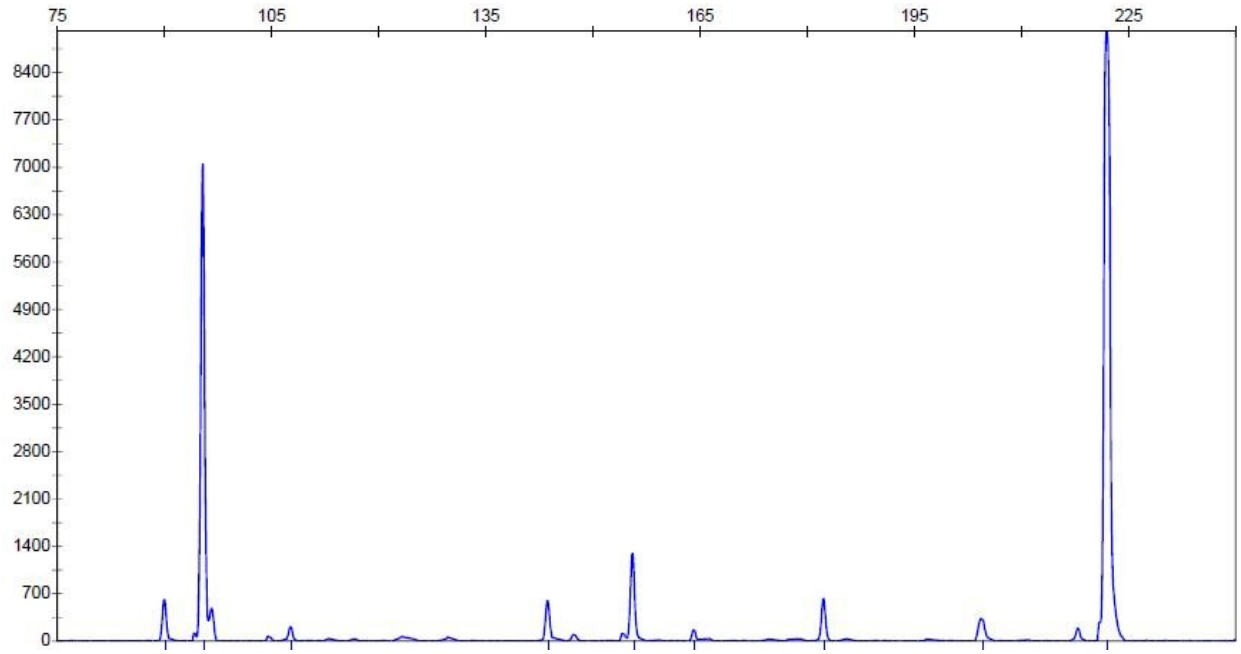


Figure 3.31: X-inactivation mouse assay. PCR products run on 1% agarose gel. Lane 1: Thermo Fisher Scientific O'GeneRuler Ladder; Lane 2: mouse cre 8 exon 1-4; Lane 3: mouse cre 12 exon 1-4; Lane 4: mouse cre 15 exon 1-4; Lane 5: mouse cre 8 exon 1-5; Lane 6: mouse cre 12 exon 1-5; Lane 7: mouse cre 15 exon 1-5; Lane 8: no RNA control exon 1-4; Lane 9: no RNA control exon 1-5.

A)


75
105
135
165
195
225
8400
7700
7000
6300
5600
4900
4200
3500
2800
2100
1400
700
0

B)

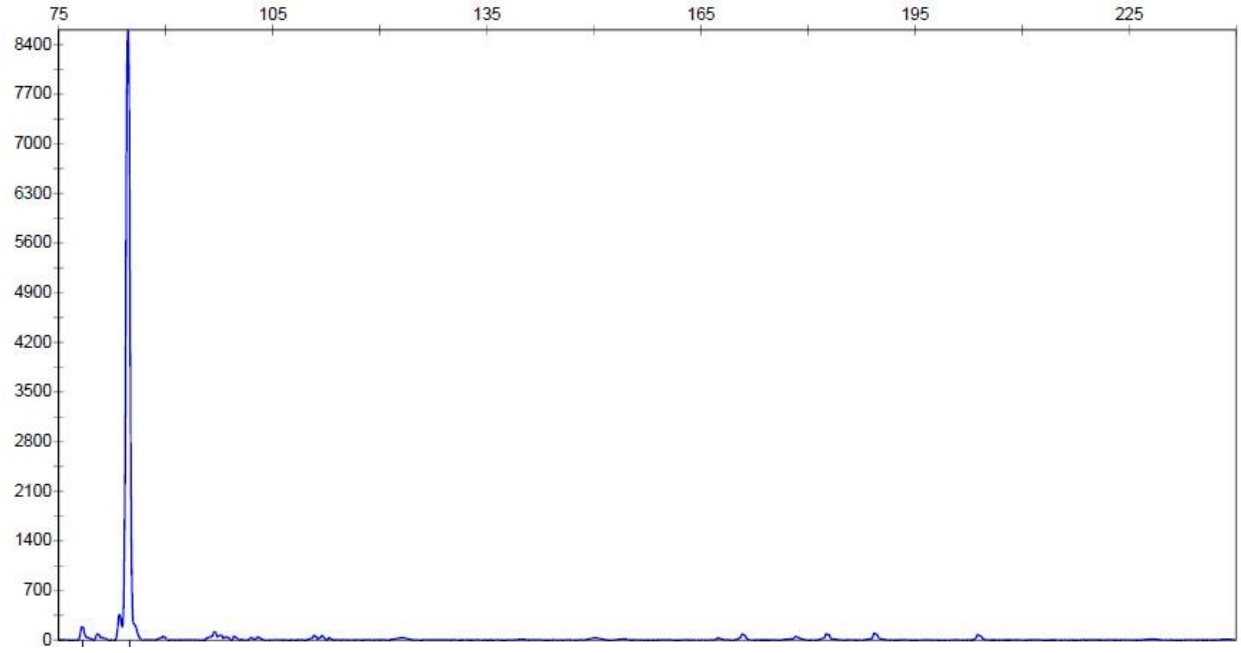

75
105
135
165
195
225
8400
7700
7000
6300
5600
4900
4200
3500
2800
2100
1400
700
0

Figure 3.32: Electrophoresis graphs of PCR product dilutions. PCR dilution electrophoresis graphs showing expected peak location for wildtype and *Chm*^{3lox} alleles. A) Reverse primer in exon 5 picks up transcripts from both wildtype (220 bp) and *Chm*^{3lox} (95 bp) alleles. B) Reverse primer in exon 4 only picks up wildtype transcript (84 bp).

CHAPTER 4 DISCUSSION

4.1 Importance of identifying novel CHM mutations

Due to the phenotypic similarities that choroideremia shares with other ocular disorders, such as retinitis pigmentosa (RP), it is crucial to identify the molecular diagnosis of choroideremia in order for patients to have the hope of potential gene therapy treatment or for family or career planning purposes. Retinitis pigmentosa is much more complicated genetically than monogenic choroideremia, having more than 40 associated genes (Ferrari et al., 2011). Being misdiagnosed with RP, patients will miss out on the opportunity to potentially be included in a gene therapy trial or treatment. Although sequencing the coding region of the *CHM* gene appears to be the standard for the diagnosis of choroideremia, our research shows that sequencing the non-coding regions of the gene may also be required in some patients to confirm the diagnosis. Developing assays that target the non-coding region may noticeably increase the diagnostic sensitivity for choroideremia.

4.2 Novel splice mutation

The identification of the same novel splice mutation c.1245-521A>G in two unrelated individuals from different ancestries was unexpected. A human identification protocol was completed on DNA obtained from cell lines in the research laboratory and DNA obtained directly from patient samples in the molecular diagnostic lab. Analysis of a nearby SNP showed different nucleotides in both patients, suggesting that this mutation may not have arisen on a single genetic background. The confirmation of sample identity and absence of common descent suggests that this area of the gene could potentially be a hotspot for mutations. Upon further investigation of that region and surrounding areas, some inverted repeats were discovered nearby. The repetitive nature of the sequence surrounding this mutation can cause DNA

instability and cause replication machinery to make mistakes, leading to a more favourable mutation site. In addition, the deamination of adenine is a relatively common mutation, accounting for approximately 15% of A>G transitions (Parry, 2007). The deamination of adenine changes adenine to hypoxanthine, which pairs with cytosine and upon replication, binds to guanine, resulting in the A>G transition (Parry, 2007). As only a single SNP was analyzed, further SNPs would need to be assessed to determine if the patients with the same mutation indeed carry different haplotypes. Regardless, the different ethnicities of the patients support the hypothesis that this area is a hotspot for mutations.

Analysis of RNA samples from patients is typically the easiest and most reliable way to detect splicing defects. However, if RNA is not available, or RNA is not expressed in the tissue obtained, such as lymphocytes from peripheral blood, other widely used laboratory techniques are the *in vitro* splicing assays or minigene splicing assays. These splicing assays require two key components: minigene constructs and mammalian crude pre-mRNA nuclear extracts, usually generated from HeLa cells (Movassat, Mueller, & Hertel, 2014). The minigene constructs identify specific splicing regulatory elements and identify features that control intron and exon usage as well as *cis*-acting and *trans*-acting factors (Movassat et al., 2014). The limitations of this assay include a decreased rate of intron removal in comparison to the *in vivo* reaction and the restricted size of RNA than can be used (less than 2,000 bp) (Movassat et al., 2014). However, laboratory testing for all splicing variants is expensive and time consuming, therefore, *in silico* software technologies are more amendable to a diagnostic laboratory.

Current *in silico* software relies on machine learning, whereby known splice sites are used to 'teach' the computer; the user can then input an unknown sequence and the software will predict the likelihood of splicing based on its knowledge of actual splice sites. This method is

economical as it allows molecular geneticists to screen thousands of variants and narrow down candidates for further experimental confirmation. The c.1245-521A>G variant in *CHM* is predicted to introduce a canonical splice acceptor. Alamut maximum entropy splicing prediction software demonstrated that this new splice acceptor is more favourable in terms of entropy. The aberrant splicing was confirmed through sequencing of the resulting RNA transcript. All programs summarized by Alamut splice predictor (SSF, MaxENT, NNSPLICE, GeneSplicer and HSF) yielded the same prediction. Due to aberrant splicing, it is likely that the RNA transcript will undergo nonsense-mediated decay (NMD) and therefore, no REP-1 protein is detectable in this patient.

4.3 Nonsense mediated decay

The nonsense-mediated mRNA decay (NMD) pathway occurs in the cytoplasm and selectively degrades mRNAs containing premature termination codons (PTCs) as well as cellular RNAs that are abundant (Hug, Longman, & Caceres, 2016). Without NMD, the mRNAs containing PTCs would be translated into truncated proteins, which could potentially have a deleterious effect on the organism (Hug et al., 2016). The purpose of the nonsense mediated decay pathway is to degrade transcripts containing PTCs and also to aid in the regulation of normal transcripts, which works to control gene expression (Hug et al., 2016). The NMD process must be able to differentiate a normal termination codon from a premature one (Hug et al., 2016). Being closely linked to pre-mRNA splicing, the presence of an exon junction complex (EJC) downstream of a termination codon is a good indicator it is a PTC rather than normal stop codon (da Costa et al., 2017). The EJC is a multi subunit protein complex deposited 20-24 nucleotides upstream of an exon-exon junction during pre-mRNA splicing and remains there until it is removed by translational machinery (Hug et al., 2016).

NMD also degrades alternative spliced transcripts with frameshifts that lead to the introduction of PTCs (da Costa et al., 2017). Bioinformatic predictions suggest that one third of alternatively spliced human mRNAs leads to introduction of a PTC (da Costa et al., 2017). This makes alternative splicing NMD an important post-translational mechanism for controlling gene expression.

4.4 Improved diagnostic sensitivity

Improving the sensitivity of the diagnosis of choroideremia is important. Diagnostic laboratories generally do not perform immunoblot analysis and instead rely on DNA testing. If a clinical phenotype suggestive of choroideremia cannot be confirmed through a diagnostic test, treatment by gene therapy is not an option. Some patients that phenotypically resemble choroideremia may not have a mutation in *CHM* and may express REP-1 protein; an example in this research would be Patient C. Without the improvements made from this research to the sensitivity of diagnosis, Patient C may have undergone gene replacement therapy when it was not needed. Overall, considering all the pathogenic mutations identified to date, a diagnostic assay for CHM can reach 95% sensitivity. The diagnostic test sensitivity following the discovery of the two novel mutations was determined by taking the number of patients with a molecular diagnosis and dividing it by the total number of patients in the research database (that are clinically suspected to have choroideremia) (Table 4.1). The remaining unknown patients were tested for the novel splice mutation (c.1245-521A>G) and the novel inversion mutation (c.-839_49+5528inv) and determined to not carry either of those mutations. It is entirely possible that some of the remaining "unknown" patients have been misdiagnosed clinically and this

would improve the overall test sensitivity. Hence, the numbers calculated in Table 4.1 likely represent a conservative estimate of the sensitivity of *CHM* gene testing for choroideremia.

Test Method	Sensitivity in Affected Males
Coding sequence analysis	84%
Exon CNV analysis	9%
Promoter sequencing	0.3%
Intronic sequencing for c.1245-521A>G	0.7%
Inversion analysis for c.-839_49+5528inv	0.3%
Unknown	4.4%

Table 4.1: Improved diagnostic sensitivity. This research and now published research categorizing the promoter of *CHM* (Radziwon et al., 2017) has improved the diagnostic sensitivity for choroideremia.

4.5 Patient A's sister's carrier status

Determining the carrier status of Patient A's sister was very important to her for family planning purposes. Carrying this mutation gives her a 50% chance of having an affected male child, as well as a 50% chance of having a carrier daughter. In a study conducted about fragile X syndrome, Wehbe and colleagues investigated young adult females with a history of fragile X syndrome for their opinions on the ideal ages for: learning fragile X is inherited, learning one could be a carrier and offering carrier testing (Wehbe, Spiridigliozzi, Heise, Dawson, & McConkie-Rosell, 2009). No participants in this study expressed regret about the timing of disclosure of their genetic risk and their opinions for appropriate ages were linked to their own experiences and their carrier status knowledge (Wehbe et al., 2009). Often it is recommended that carrier testing should be deferred until the child is of an appropriate age to provide informed consent. There are varying guidelines around the world of what age genetic and carrier testing is acceptable for early onset diseases such as Tay-Sachs disease and cystic fibrosis (Nazareth, Lazarin, & Goldberg, 2015). In the case of our patient, it is usually recommended that she wait to have her female child undergo genetic testing until the child is an appropriate age to provide consent, as female carriers of choroideremia are usually asymptomatic. If a female carrier of choroideremia knew her child was male, genetic testing could be offered as early as requested (i.e. even prenatally), as the decision to terminate a pregnancy remains her choice until approximately 22 6/7 weeks. If testing is prolonged until after the birth of the child, it is hoped that they could become eligible for gene therapy. The goal of gene therapy for choroideremia is to prevent any further progression of the disease; therefore, being able to administer the therapy before any vision loss occurred would be ideal.

4.6 Patient C

Patient C, although phenotypically resembling choroideremia, is likely a phenocopy as he does not carry a mutation in the *CHM* gene and was found to express the REP-1 protein by immunoblot analysis. Since neither parent is affected and they report no family history of vision problems, a compound heterozygous inheritance pattern seems most likely. The patient and both parents were sent for whole exome sequencing and through iterative filtering, it was determined that collagen alpha-1 (*COL18A1*), prominin-1 (*PROM1*) or crumbs-1 (*CRB1*) genes were good candidates for investigation. Mutations in collagen alpha-1 *COL18A1* gene are known to cause Knobloch syndrome (Haghighi et al., 2014). Knobloch syndrome presents with vitroretinal degradation, recurrent retinal detachment, retinitis pigmentosa-like features, lens subluxation, congenital high myopia and macular abnormalities (Haghighi et al., 2014). Although our patient presents with retinitis pigmentosa-like features, the chorioretinal degeneration is a very rare manifestation in Knobloch syndrome. In addition, our patient does not present with any other syndromic characteristics of Knobloch. The variant c.1175T>C in *COL18A1* is a missense variant located in exon 1 which changes a leucine to a proline. There are no other reported pathogenic variants in this functional domain of the protein. The variant c.4060T>C in *COL18A1* is a missense variant in exon 33, which converts a glycine to a serine. There are reported likely benign missense and pathogenic frameshift mutations in this functional domain of the protein.

Prominin-1 *(PROM1)* is expressed *in vivo* in the retina but its function remains mostly unknown. Previously, a single nucleotide deletion has been reported to cause an autosomal recessive retinal disorder (Jaszai, Fargeas, Florek, Huttner, & Corbeil, 2007). The affected individual presented with symptoms similar to choroideremia, such as, night blindness, and

progressive peripheral vision loss (Jaszai et al., 2007). For this reason, it is possible that the two variants, one inherited from each parent, could combine and cause phenotypic symptoms similar to those in our patient. The variant c.868A>C in *PROM1* is a missense variant in exon 10, which converts serine to arginine. This functional domain of this variant also contains previously reported variants of unknown significance. The variant c.1559C>T is a missense variant in exon 15, which changes threonine to methionine. There is a previously reported pathogenic frameshift variant as well as a reported missense variant of unknown significance located in the same functional domain as variant c.1559C>T.

As for crumbs-1 (*CRB1)*, this gene is known to cause many different ophthalmic phenotypes, which have a wide variety of clinical features with no genotype correlation (Quinn, Pellissier, & Wijnholds, 2017). *CRB1* is strongly linked to Leber Congenital Amaurosis (LCA) and early-onset retinitis pigmentosa (RP) as well as early-onset rod-cone dystrophy (EORCD) and cone-rod dystrophy (CRD) (Ehrenberg et al., 2013). The variant in our patient, c.3202A>G, a missense mutation in exon 9 changing threonine to alanine, has been previously reported three times in ClinVar to be pathogenic causing LCA, pigmented paravenous chorioretinal atrophy and recessive retinitis pigmentosa. Although I did not identify another obvious variant in *CRB1*, it could be in the non-coding region, or incorrectly filtered out by my filters (i.e. it is falsely predicted to be a low or moderate likelihood of pathogenicity, or a *de novo* change not present in one parent). To follow up on this it would be important to remove all variant filters and examine for any changes in the *CRB1* gene in this patient. Therefore, it is possible that the above variant in this gene in our patient could be causing another ophthalmic phenotype that has not yet been seen.

It is also possible that these candidate variants are not the causative pathogenic variants in this individual. Other inheritance patterns may also require investigation such as homozygous autosomal recessive mutations, one inherited recessive mutation and one acquired *de novo* mutation, a *de novo* autosomal dominant mutation or an X-linked mutation. In addition, because exome sequencing was undertaken, it is possible there is a causative mutation in the non-coding regions of a gene. It is also possible that the causative gene for this disorder is not in an already known retina-expressed gene and in a gene with unknown expression. In order to investigate homozygous recessive mutations, the first step in Figure 2.1 could be modified to filter only the homozygous, and filtering would continue as described. To investigate an X-linked mutation, the filters would be adjusted to contain only patient and mother in step two of Figure 2.1. The filtering would require adding a step to filter only for variants on the X chromosome and continue filtering as described.

Definitively assigning the genetic cause of disease requires further investigation of candidate genes. With autosomal recessive disorders, examining segregation in the family members will not be helpful. Therefore, after narrowing down suspect genes, investigating whether protein is translated from the mutated transcripts would be an appropriate starting point.

4.7 Identifying structural rearrangements

DNA recombination is an important evolutionary process and it is essential for DNA repair and replication. Occasionally when a gene pairs alongside its homologous chromosome during meiosis or it's complementary strand during DNA replication or repair, the sequence may loop, break and be end-joined. This can cause the sequence to become inverted (Heyer & Kanaar, 2004). The result is a balanced inversion, whereby frequently, little to no genetic material is lost and the direction of the sequence is reversed.

Traditional exon sequencing may miss these structural rearrangements, as demonstrated by our inversion patient. Structural rearrangements present a particularly difficult diagnostic challenge. Karyotyping and fluorescent *in situ* hybridization (FISH) have low sensitivity and furthermore, they require the initial knowledge of the inversion for the assay design. Microarrays can be used to narrow down breakpoints but they do not detect balanced rearrangements. Next generation high-throughput DNA sequencing strategies have allowed the detection of balanced chromosomal changes (Talkowski et al., 2011). Paired-end sequencing uses short reads (up to 500 bp) from both ends of millions of DNA fragments to detect and characterize rearrangements (Campbell et al., 2008). In this strategy, small fragments have the advantage of greater sensitivity for small intrachromosomal rearrangements and straightforward PCR confirmation (Campbell et al., 2008). If slightly larger fragments are used instead of small fragments less than 500 bp, it becomes more difficult to sequence and annotate the breakpoints. Another strategy is large-insert jumping libraries, which shear genomic DNA into a targeted insert size, those fragments are circularized and then fragmented with the circularization junctions retained (Hanscom & Talkowski, 2014). This derives short genomic fragments in which the ends are separated by the size of the circle, which enables massively parallel sequencing of the ends of the fragments and the size of the circle to provide information about the assembly of the genome (Hanscom & Talkowski, 2014). Although our RNA analysis suggested a regulatory mutation or potential upstream variant due to lack of transcript expression, diagnostic laboratories do not have routine access to patient RNA. Therefore, these other new sequencing techniques are becoming important diagnostic tools. The main disadvantage of most of these strategies is that knowledge of where the inversion is occurring is important for diagnosis. Balanced inversions are especially difficult as all the expected genetic material is present and without finding the

inversion breakpoints, the genotype can appear normal on most assays. My diagnostic assay will only detect this particular inversion, but there are likely other inversions in the *CHM* gene that will need detection and can be identified by adapting this strategy.

4.8 Novel inversion mutation

The novel inversion mutation, c.-839_49+5528inv, was identified in the fourth patient in this study. Unknown inversions present a particularly difficult diagnostic challenge as none of the genetic material is missing. The typical short-range PCR-based assay will yield a false negative result as primer pairs will still bind and amplify regions of the gene unless they fortuitously span the inversion breakpoints. I have designed a diagnostic multiplex PCR to accomplish identification of this particular inversion in future choroideremia patients. Southern blots are typically used to detect inversion mutations, such as in haemophilia A families (Deutz-Terlouw et al., 1995). DNA samples are digested with specific restriction enzymes, the fragments are size separated by gel electrophoresis and following DNA transfer to a membrane a probe is hybridized to the region in question (Deutz-Terlouw et al., 1995). Southern blot analysis is time consuming, labour intensive and expensive; therefore scientists have turned to next generation sequencing (NGS) methods to detect inversions. Long-range PCR protocols can be used to rule out deletions, but in order to be effective in diagnosing inversions; there must be knowledge of the inversion location. This is the first inversion mutation characterized in the choroideremia gene. The assay I designed will be much cheaper than a Southern blot (approximately $5 vs $300) and faster (1 day vs 3-4 days). This makes it an ideal assay for detecting a known inversion in a diagnostic laboratory.

4.9 Genotype vs. phenotype

Many questions still remain in regards to a genotype-phenotype correlation within choroideremia. It is suggested that because this is a monogenic disorder, all patients with a mutation in this gene, resulting in lack of protein should have a similar phenotype but this is not the case. All patients present with different severities and different ages of onset and the individuals in this study are no different. The phenotypic variability could be due to a number of factors: environmental, epigenetic or stochastic. Depending on the patient's environment, job, or lifestyle this could affect the severity of their symptoms of disease. It is possible that other genes or proteins are involved in the mechanism of the disease (i.e. the other proteins involved in membrane trafficking). Stochastic events taking place throughout development and aging, such as unequal cellular divisions, may also lead to the variation observed in patients. One of our novel mutations involves the region upstream of the gene and the newly characterized promoter region being inverted, this patient's phenotype is not markedly different than any other choroideremia patient, suggesting that the mutation upstream of the gene does not change the phenotypic outcome.

4.10 Clinical trials

Current clinical trials performed on patients with a complete loss of protein is considered advantageous in terms of determining appropriate gene dosage (i.e. expression) levels. In clinical trials for Leber Congenital Amaurosis (LCA), a misfolding mutation in RPE65 cannot be treated through gene therapy because the misfolding, mislocalization and aggregation of mutant protein causes cytotoxic effects (Li et al., 2014). A similar situation could occur with

choroideremia gene therapy in patients with misfolding proteins. The importance of having a molecular diagnosis and understanding of the patient's mutation is increasingly important with the idea of treatment. The aim of gene therapy would be to successfully prevent further death of photoreceptor cells and therefore, treating a patient with very few healthy photoreceptor cells left would be hard to consider as a success. The ideal candidate for gene therapy is one who is just in the early stage of visual deterioration of their visual acuity. Patients with many remaining viable photoreceptor cells are most likely to be successful candidates, although this usually means they are younger in age, before deterioration has occurred. With knowledge of family history and genetic diagnosis, these individuals would be more easily selected. No preference should be given to one particular type of genetic mutation over another, as no genotype phenotype correlation has been established (Freund et al., 2016).

4.11 X-inactivation

The fluorescence graphs produced were of expected size from the PCR products. This assay allows an easily measurement of X-inactivation in the *Chm* mouse model. The scheduling of model mice fell behind and the assay has not yet been completed on the mice in this study. More work may need to be done when the mice are ready if the Posse de Chavez laboratory wishes to use only one set of primers. Trials will need to be run on carrier female mice to determine what the PCR amplification bias is and that difference can then be accounted for in the resulting fluorescent graphs. Other laboratories working with this mouse knockout model of choroideremia could use the assay. This assay is expected to permit comparison of fluctuations in biochemical responses with the degree of X-inactivation present (i.e. provides a sensitive measure of REP-1 expression).

4.12 Future directions

There are still many questions when it comes to choroideremia and gene therapy. With next generation sequencing becoming more affordable and accessible it may become the standard to sequence the entire genetic region of the *CHM* gene to screen for any mutations that may be found in the expansive non-coding region of the gene, which takes up 97% of *CHM*. Our research demonstrates how important it is to also consider mutations in non-coding regions that can be disease causing. There are approximately 14 patients (4.5%) remaining in our research database with phenotypic symptoms of choroideremia and who lack protein expression through immunoblot analysis but for whom mutations in the coding region have not been identified. We improved the diagnostic sensitivity of testing for choroideremia by 4% in our research population of choroideremia families by assessing the non-coding regions of the *CHM* gene.

Many questions remain in terms of gene therapy in choroideremia patients. Notably, the percentage of normal REP-1 function that is required to stop the progression of visual degeneration and the number of viral particles required for injection. There may also be other possibilities for treatment. Other proteins in the pathway could be targeted to improve vesicular transport in the eye and ultimately improve vision. Other interesting avenues to look into would be REP-2 and whether it can be modified to also compensate for REP-1 in the layers of the eye, in which REP-2 is not normally expressed. The Rab proteins specifically affected by loss of REP-1 may also lead to a possible treatment option. This may compensate for the lack of REP-1 and restore or prevent future loss of vision.

Understanding the genotype of Patient A and B and the introduction of a splice acceptor could also allow for patient specific therapies. Exon skipping may be applicable to this

particular mutation and allow for therapy by agents that induce skipping of the errant sequence from being incorporated into mRNA. Designing an antisense oligonucleotide that binds to the new splice acceptor site would silence it and allow splicing to occur normally (Nakamura, 2017). This therapy is currently being implemented in muscular dystrophy patients (Nakamura, 2017). Another possibility would be editing out the area around this single base mutation using a CRISPR/Cas9 method. Creating guide RNA to direct Cas9 to cut the DNA on either side of the point mutation and then allowing the cell to repair the cut in the DNA through homologous end joining, the mutation would be successfully edited out and no aberrant splicing would occur (Hess, Tycko, Yao, & Bassik, 2017).

Patient C requires more investigation. It is possible that the mode of inheritance is not compound heterozygous and could be another inheritance pattern, producing additional candidate genes for analysis. In order to investigate our candidate genes further, a first step would be to sequence the mutations in order to verify their presence. Next, protein analysis by western blot would assess protein expression and would provide insight into expression changes, or the production of an aberrant protein. This would give some indication of whether the mutation is disease causing or not. Functional assays could be designed in cultured cell lines including the mutated transcript to determine if the mutation is affecting cell viability.

Reference List

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]*, 0 7, Unit7.20. http://doi.org/10.1002/0471142905.hg0720s76

Alexandrov, K., Horiuchi, H., Steele-Mortimer, O., Seabra, M. C., & Zerial, M. (1994). Rab escort protein-1 is a multifunctional protein that accompanies newly prenylated rab proteins to their target membranes. *EMBO J, 13*(22), 5262-5273.

Alory, C., & Balch, W. E. (2001). Organization of the Rab-GDI/CHM superfamily: the functional basis for choroideremia disease. *Traffic, 2*(8), 532-543.

Alory, C., & Balch, W. E. (2003). Molecular evolution of the Rab-escort-protein/guanine-nucleotide-dissociation-inhibitor superfamily. *Mol Biol Cell, 14*(9), 3857-3867. doi:10.1091/E03-04-0227

Anant, J. S., Desnoyers, L., Machius, M., Demeler, B., Hansen, J. C., Westover, K. D., . . . Seabra, M. C. (1998). Mechanism of Rab geranylgeranylation: formation of the catalytic ternary complex. *Biochemistry, 37*(36), 12559-12568. doi:10.1021/bi980881a

Andres, D. A., Seabra, M. C., Brown, M. S., Armstrong, S. A., Smeland, T. E., Cremers, F. P., & Goldstein, J. L. (1993). cDNA cloning of component A of Rab geranylgeranyl transferase and demonstration of its role as a Rab escort protein. *Cell, 73*(6), 1091-1099.

Bonifacino, J. S., & Glick, B. S. (2004). The mechanisms of vesicle budding and fusion. *Cell, 116*(2), 153-166.

Brooks, B. P., Larson, D. M., Chan, C. C., Kjellstrom, S., Smith, R. S., Crawford, M. A., . . . Pavan, W. J. (2007). Analysis of ocular hypopigmentation in Rab38cht/cht mice. *Invest Ophthalmol Vis Sci, 48*(9), 3905-3913. doi:10.1167/iovs.06-1464

Campbell, P. J., Stephens, P. J., Pleasance, E. D., O'Meara, S., Li, H., Santarius, T., . . . Futreal, P. A. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet, 40*(6), 722-729. doi:10.1038/ng.128

Carrel, L., & Willard, H. F. (1999). Heterogeneous gene expression from the inactive X chromosome: an X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. *Proc Natl Acad Sci U S A, 96*(13), 7364-7369.

Carss, K. J., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S., . . . Raymond, F. L. (2017). Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet, 100*(1), 75-90. doi:10.1016/j.ajhg.2016.12.003

Chi, J. Y., MacDonald, I. M., & Hume, S. (2013). Copy number variant analysis in CHM to detect duplications underlying choroideremia. *Ophthalmic Genet, 34*(4), 229-233. doi:10.3109/13816810.2012.752016

Collins, M., & Thrasher, A. (2015). Gene therapy: progress and predictions. *Proc Biol Sci, 282*(1821), 20143003. doi:10.1098/rspb.2014.3003

Coussa, R. G., & Traboulsi, E. I. (2012). Choroideremia: a review of general findings and pathogenesis. *Ophthalmic Genet, 33*(2), 57-65. doi:10.3109/13816810.2011.620056

Cremers, F. P., Armstrong, S. A., Seabra, M. C., Brown, M. S., & Goldstein, J. L. (1994). REP-2, a Rab escort protein encoded by the choroideremia-like gene. *J Biol Chem, 269*(3), 2111-2117.

da Costa, P. J., Menezes, J., & Romao, L. (2017). The role of alternative splicing coupled to nonsense-mediated mRNA decay in human disease. *Int J Biochem Cell Biol*. doi:10.1016/j.biocel.2017.07.013

Deutz-Terlouw, P. P., Losekoot, M., Olmer, R., Pieneman, W. C., de Vries-v d Weerd, S., Briet, E., & Bakker, E. (1995). Inversions in the factor VIII gene: improvement of carrier detection and prenatal diagnosis in Dutch haemophilia A families. *J Med Genet, 32*(4), 296-300.

Dimopoulos, I. S., Radziwon, A., St Laurent, C. D., & MacDonald, I. M. (2017). Choroideremia. *Curr Opin Ophthalmol, 28*(5), 410-415. doi:10.1097/ICU.0000000000000392

Dirac-Svejstrup, A. B., Sumizawa, T., & Pfeffer, S. R. (1997). Identification of a GDI displacement factor that releases endosomal Rab GTPases from Rab-GDI. *EMBO J, 16*(3), 465-472. doi:10.1093/emboj/16.3.465

Disteche, C. M. (1995). Escape from X inactivation in human and mouse. *Trends Genet, 11*(1), 17-22.

Duque, S., Joussemet, B., Riviere, C., Marais, T., Dubreil, L., Douar, A. M., . . . Barkats, M. (2009). Intravenous administration of self-complementary AAV9 enables transgene delivery to adult motor neurons. *Mol Ther, 17*(7), 1187-1196. doi:10.1038/mt.2009.71

Ehrenberg, M., Pierce, E. A., Cox, G. F., & Fulton, A. B. (2013). CRB1: one gene, many phenotypes. *Semin Ophthalmol, 28*(5-6), 397-405. doi:10.3109/08820538.2013.825277

Ferrari, S., Di Iorio, E., Barbaro, V., Ponzin, D., Sorrentino, F. S., & Parmeggiani, F. (2011). Retinitis pigmentosa: genes and disease mechanisms. *Curr Genomics, 12*(4), 238-249. doi:10.2174/138920211795860107

Freund, P. R., Sergeev, Y. V., & MacDonald, I. M. (2016). Analysis of a large choroideremia dataset does not suggest a preference for inclusion of certain genotypes in future trials of gene therapy. *Mol Genet Genomic Med, 4*(3), 344-358. doi:10.1002/mgg3.208

Haghighi, A., Tiwari, A., Piri, N., Nurnberg, G., Saleh-Gohari, N., Haghighi, A., . . . Berger, W. (2014). Homozygosity mapping and whole exome sequencing reveal a novel homozygous COL18A1 mutation causing Knobloch syndrome. *PLoS One, 9*(11), e112747. doi:10.1371/journal.pone.0112747

Hanscom, C., & Talkowski, M. (2014). Design of large-insert jumping libraries for structural variant detection using Illumina sequencing. *Curr Protoc Hum Genet, 80*, 7 22 21-29. doi:10.1002/0471142905.hg0722s80

Hess, G. T., Tycko, J., Yao, D., & Bassik, M. C. (2017). Methods and Applications of CRISPR-Mediated Base Editing in Eukaryotic Genomes. *Mol Cell, 68*(1), 26-43. doi:10.1016/j.molcel.2017.09.029

Heyer, W. D., & Kanaar, R. (2004). Recombination mechanisms; fortieth anniversary meeting of the Holliday model. *Mol Cell, 16*(1), 1-9. doi:10.1016/j.molcel.2004.09.025

Hug, N., Longman, D., & Caceres, J. F. (2016). Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res, 44*(4), 1483-1495. doi:10.1093/nar/gkw010

Hutagalung, A. H., & Novick, P. J. (2011). Role of Rab GTPases in membrane traffic and cell physiology. *Physiol Rev, 91*(1), 119-149. doi:10.1152/physrev.00059.2009
Hutagalung, A. H., & Novick, P. J. (2011). Role of Rab GTPases in membrane traffic and cell physiology. *Physiol Rev, 91*(1), 119-149. doi:10.1152/physrev.00059.2009

Jaszai, J., Fargeas, C. A., Florek, M., Huttner, W. B., & Corbeil, D. (2007). Focus on molecules: prominin-1 (CD133). *Exp Eye Res, 85*(5), 585-586. doi:10.1016/j.exer.2006.03.022

Jaszai, J., Fargeas, C. A., Florek, M., Huttner, W. B., & Corbeil, D. (2007). Focus on molecules: prominin-1 (CD133). *Exp Eye Res, 85*(5), 585-586. doi:10.1016/j.exer.2006.03.022

Jian, X., Boerwinkle, E., & Liu, X. (2014). In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med, 16*(7), 497-503. doi:10.1038/gim.2013.176

Kamhi, E., Raitskin, O., Sperling, R., & Sperling, J. (2010). A potential role for initiator-tRNA in pre-mRNA splicing regulation. *Proc Natl Acad Sci U S A, 107*(25), 11319-11324. doi:10.1073/pnas.0911561107

Kohnke, M., Delon, C., Hastie, M. L., Nguyen, U. T., Wu, Y. W., Waldmann, H., . . . Alexandrov, K. (2013). Rab GTPase prenylation hierarchy and its potential role in choroideremia disease. *PLoS One, 8*(12), e81758. doi:10.1371/journal.pone.0081758

Li, S., Izumi, T., Hu, J., Jin, H. H., Siddiqui, A. A., Jacobson, S. G., . . . Jin, M. (2014). Rescue of enzymatic function for disease-associated RPE65 proteins containing various missense

mutations in non-active sites. *J Biol Chem, 289*(27), 18943-18956. doi:10.1074/jbc.M114.552117

Lopes, V. S., Wasmeier, C., Seabra, M. C., & Futter, C. E. (2007). Melanosome maturation defect in Rab38-deficient retinal pigment epithelium results in instability of immature melanosomes during transient melanogenesis. *Mol Biol Cell, 18*(10), 3914-3927. doi:10.1091/mbc.E07-03-0268

MacDonald, I. M., Sereda, C., McTaggart, K., & Mah, D. (2004). Choroideremia gene testing. *Expert Rev Mol Diagn, 4*(4), 478-484. doi:10.1586/14737159.4.4.478

MacLaren, R. E., Groppe, M., Barnard, A. R., Cottriall, C. L., Tolmachova, T., Seymour, L., . . . Seabra, M. C. (2014). Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial. *Lancet, 383*(9923), 1129-1137. doi:10.1016/S0140-6736(13)62117-0

Matsui, Y., Kikuchi, A., Araki, S., Hata, Y., Kondo, J., Teranishi, Y., & Takai, Y. (1990). Molecular cloning and characterization of a novel type of regulatory protein (GDI) for smg p25A, a ras p21-like GTP-binding protein. *Mol Cell Biol, 10*(8), 4116-4122.

Movassat, M., Mueller, W. F., & Hertel, K. J. (2014). In vitro assay of pre-mRNA splicing in mammalian nuclear extract. *Methods Mol Biol, 1126*, 151-160. doi:10.1007/978-1-62703-980-2_11

Nakamura, A. (2017). Moving towards successful exon-skipping therapy for Duchenne muscular dystrophy. *J Hum Genet, 62*(10), 871-876. doi:10.1038/jhg.2017.57

Nazareth, S. B., Lazarin, G. A., & Goldberg, J. D. (2015). Changing trends in carrier screening for genetic disease in the United States. *Prenat Diagn, 35*(10), 931-935. doi:10.1002/pd.4647

Ng PC, Henikoff S. Accounting for Human Polymorphisms Predicted to Affect Protein Function. *Genome Res.* 2002 Mar;12(3):436-46.

Parry, T. E. (2007). On the mutagenic action of adenine. *Leuk Res. 31*(12), 1621-1624.

Pereira-Leal, J. B., Strom, M., Godfrey, R. F., & Seabra, M. C. (2003). Structural determinants of Rab and Rab Escort Protein interaction: Rab family motifs define a conserved binding surface. *Biochem Biophys Res Commun, 301*(1), 92-97.

Preising, M., & Ayuso, C. (2004). Rab escort protein 1 (REP-1) in intracellular traffic: a functional and pathophysiological overview. *Ophthalmic Genet, 25*(2), 101-110. doi:10.1080/13816810490514333

Quinn, P. M., Pellissier, L. P., & Wijnholds, J. (2017). The CRB1 Complex: Following the Trail of Crumbs to a Feasible Gene Therapy Strategy. *Front Neurosci, 11*, 175. doi:10.3389/fnins.2017.00175

Radziwon, A., Arno, G., D, K. W., McDonagh, E. M., Baple, E. L., Webb-Jones, K., . . . MacDonald, I. M. (2017). Single-base substitutions in the CHM promoter as a cause of choroideremia. *Hum Mutat, 38*(6), 704-715. doi:10.1002/humu.23212

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics

and the Association for Molecular Pathology. *Genet Med, 17*(5), 405-424. doi:10.1038/gim.2015.30

Roosing, S., Collin, R. W., den Hollander, A. I., Cremers, F. P., & Siemiatkowska, A. M. (2014). Prenylation defects in inherited retinal diseases. *J Med Genet, 51*(3), 143-151. doi:10.1136/jmedgenet-2013-102138

Sanford, J. R., & Caceres, J. F. (2004). Pre-mRNA splicing: life at the centre of the central dogma. *J Cell Sci, 117*(Pt 26), 6261-6263. doi:10.1242/jcs.01513

Seabra, M. C. (1996). New insights into the pathogenesis of choroideremia: a tale of two REPs. *Ophthalmic Genet, 17*(2), 43-46.

Seabra, M. C., Brown, M. S., & Goldstein, J. L. (1993). Retinal degeneration in choroideremia: deficiency of rab geranylgeranyl transferase. *Science, 259*(5093), 377-381.

Seabra, M. C., Ho, Y. K., & Anant, J. S. (1995). Deficient geranylgeranylation of Ram/Rab27 in choroideremia. *J Biol Chem, 270*(41), 24420-24427.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan 1;29(1):308-11.

Shi, W., et al. (2004). Choroideremia gene product affects trophoblast development and vascularization in mouse extra-embryonic tissues. *Dev Biol* 272(1), 53-65.

Simonelli, F., Maguire, A. M., Testa, F., Pierce, E. A., Mingozzi, F., Bennicelli, J. L., . . . Auricchio, A. (2010). Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. *Mol Ther, 18*(3), 643-650. doi:10.1038/mt.2009.277

Simunovic, M. P., Jolly, J. K., Xue, K., Edwards, T. L., Groppe, M., Downes, S. M., & MacLaren, R. E. (2016). The Spectrum of CHM Gene Mutations in Choroideremia and Their Relationship to Clinical Phenotype. *Invest Ophthalmol Vis Sci, 57*(14), 6033-6039. doi:10.1167/iovs.16-20230

Talkowski, M. E., Ernst, C., Heilbut, A., Chiang, C., Hanscom, C., Lindgren, A., . . . Gusella, J. F. (2011). Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am J Hum Genet, 88*(4), 469-481. doi:10.1016/j.ajhg.2011.03.013

Tiwari, A., Lemke, J., Altmueller, J., Thiele, H., Glaus, E., Fleischhauer, J., . . . Berger, W. (2016). Identification of Novel and Recurrent Disease-Causing Mutations in Retinal Dystrophies Using Whole Exome Sequencing (WES): Benefits and Limitations. *PLoS One, 11*(7), e0158692. doi:10.1371/journal.pone.0158692

Tolmachova, T., Anders, R., Abrink, M., Bugeon, L., Dallman, M. J., Futter, C. E., . . . Seabra, M. C. (2006). Independent degeneration of photoreceptors and retinal pigment epithelium in conditional knockout mouse models of choroideremia. *J Clin Invest, 116*(2), 386-394. doi:10.1172/JCI26617

Tolmachova, T., Anders, R., Abrink, M., Bugeon, L., Dallman, M. J., Futter, C. E., . . . Seabra, M. C. (2006). Independent degeneration of photoreceptors and retinal pigment epithelium in conditional knockout mouse models of choroideremia. *J Clin Invest, 116*(2), 386-394. doi:10.1172/JCI26617

Tolmachova, T., Tolmachov, O. E., Barnard, A. R., de Silva, S. R., Lipinski, D. M., Walker, N. J., . . . Seabra, M. C. (2013). Functional expression of Rab escort protein 1 following

AAV2-mediated gene delivery in the retina of choroideremia mice and human cells ex vivo. *J Mol Med (Berl), 91*(7), 825-837. doi:10.1007/s00109-013-1006-4

van Bokhoven, H., van den Hurk, J. A., Bogerd, L., Philippe, C., Gilgenkrantz, S., de Jong, P., . . . Cremers, F. P. (1994). Cloning and characterization of the human choroideremia gene. *Hum Mol Genet, 3*(7), 1041-1046.

van den Hurk, J. A., Hendriks, W., van de Pol, D. J., Oerlemans, F., Jaissle, G., Ruther, K., . . . Cremers, F. P. (1997). Mouse choroideremia gene mutation causes photoreceptor cell degeneration and is not transmitted through the female germline. *Hum Mol Genet, 6*(6), 851-858.

van den Hurk, J. A., van de Pol, D. J., Wissinger, B., van Driel, M. A., Hoefsloot, L. H., de Wijs, I. J., . . . Cremers, F. P. (2003). Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon. *Hum Genet, 113*(3), 268-275. doi:10.1007/s00439-003-0970-0

Vasireddy, V., Mills, J. A., Gaddameedi, R., Basner-Tschakarjan, E., Kohnke, M., Black, A. D., . . . Bennett, J. (2013). AAV-mediated gene therapy for choroideremia: preclinical studies in personalized models. *PLoS One, 8*(5), e61396. doi:10.1371/journal.pone.0061396

Wang, D., & Gao, G. (2014). State-of-the-art human gene therapy: part I. Gene delivery technologies. *Discov Med, 18*(97), 67-77.

Ward, A. J., & Cooper, T. A. (2010). The pathobiology of splicing. *J Pathol, 220*(2), 152-163. doi:10.1002/path.2649

Wehbe, R. M., Spiridigliozzi, G. A., Heise, E. M., Dawson, D. V., & McConkie-Rosell, A. (2009). When to tell and test for genetic carrier status: perspectives of adolescents and young adults from fragile X families. *Am J Med Genet A, 149A*(6), 1190-1199. doi:10.1002/ajmg.a.32840

APPENDIX

Classification of candidate variants in Patient C

To examine and classify the candidate variants in Patient C following filtering, variants were investigated in Alamut ® (Interactive Biosoftware). Summary window of variant c.1175T>C in *COL18A1* including MAF, predictions for pathogenesis from Align GVGD, SIFT, MutationTaster and PolyPhen-2 shown in Figure A1. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species are shown in Figure A2. Splice prediction using several software algorithms (SSF, MaxENT, NNSPLICE, GeneSplicer and Human Splice Finder - HSF) are shown in Figure A3. These analyses were summarized according to ACMG guidelines and a variant classification form was completed, determining this variant to be of unknown significance (VUS) (Figure A4).

Summary window of variant c.4060C>T in the *COL18A1* gene including MAF, predictions for pathogenesis from Align GVGD, SIFT, MutationTaster and PolyPhen-2 shown in Figure A5. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species are shown in Figure A6. Splice prediction using several software algorithms (SSF, MaxENT, NNSPLICE, GeneSplicer and Human Splice Finder - HSF) are shown in Figure A7. These analyses were summarized according to ACMG guidelines and a variant classification form was completed, determining this variant to be benign due to the high MAF in some ethnicities (Figure A8).

Summary window of variant c.868A>C in *PROM1* including MAF, predictions for pathogenesis from Align GVGD, SIFT, MutationTaster and PolyPhen-2 shown in Figure A9. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species are shown in Figure A10. Splice prediction using several software algorithms (SSF, MaxENT, NNSPLICE, GeneSplicer and Human Splice Finder - HSF) are

shown in Figure A11. These analyses were summarized according to ACMG guidelines and a variant classification form was completed, determining this variant to be of unknown significance (VUS) (Figure A12).

Summary window of variant c.1559C>T in the *PROM1* including MAF, predictions for pathogenesis from Align GVGD, SIFT, MutationTaster and PolyPhen-2 shown in Figure A13. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species are shown in Figure A14. Splice prediction using several software algorithms (SSF, MaxENT, NNSPLICE, GeneSplicer and Human Splice Finder - HSF) are shown in Figure A15. These analyses were summarized according to ACMG guidelines and a variant classification form was completed, determining this variant to be of unknown significance (VUS) (Figure A16).

Summary window of variant c.3202A>G in *CRB1* including MAF, predictions for pathogenesis from Align GVGD, SIFT, MutationTaster and PolyPhen-2 shown in Figure A17. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species are shown in Figure A18. Splice prediction using several software algorithms (SSF, MaxENT, NNSPLICE, GeneSplicer and Human Splice Finder - HSF) are shown in Figure A19. These analyses were summarized according to ACMG guidelines and a variant classification form was completed, determining this variant to be of unknown significance (VUS) (Figure A20).

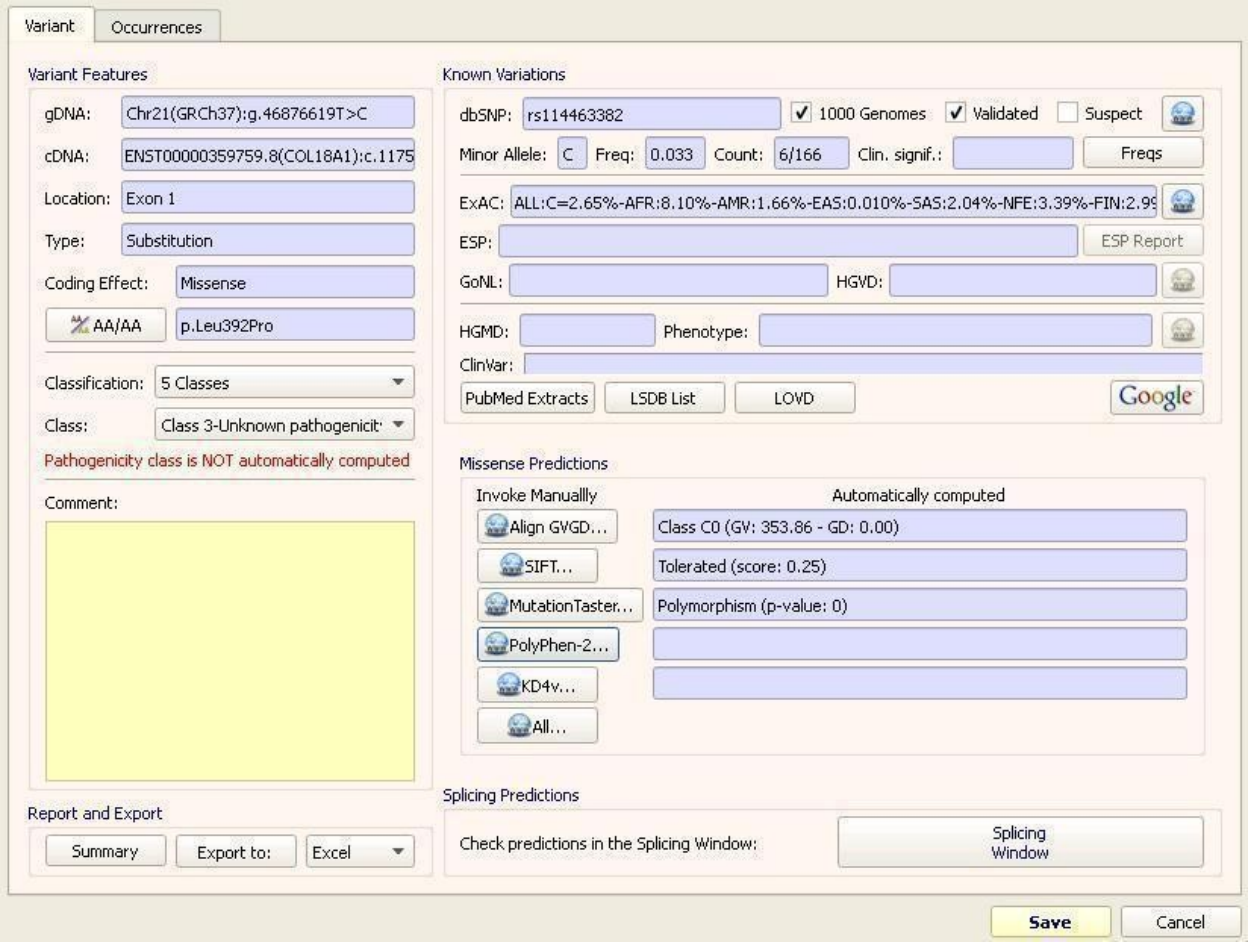


Figure A1: *COL18A1* variant c.1175T>C Alamut features. *COL18A1* variant c.1175T>C features analysis from Alamut software with reported MAF and pathogenic predictions from Align GVGD, SIFT, MutationTaster.

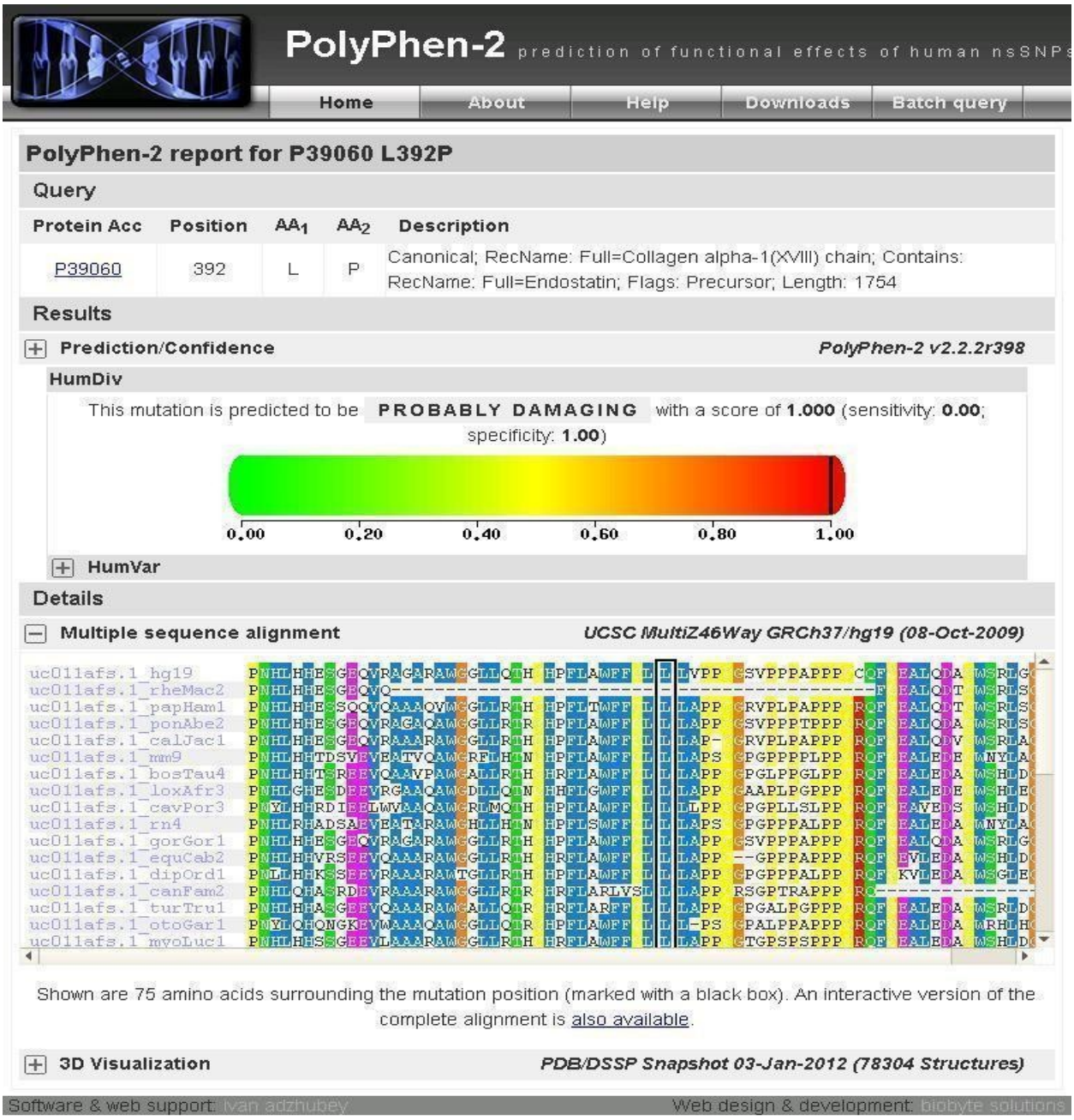
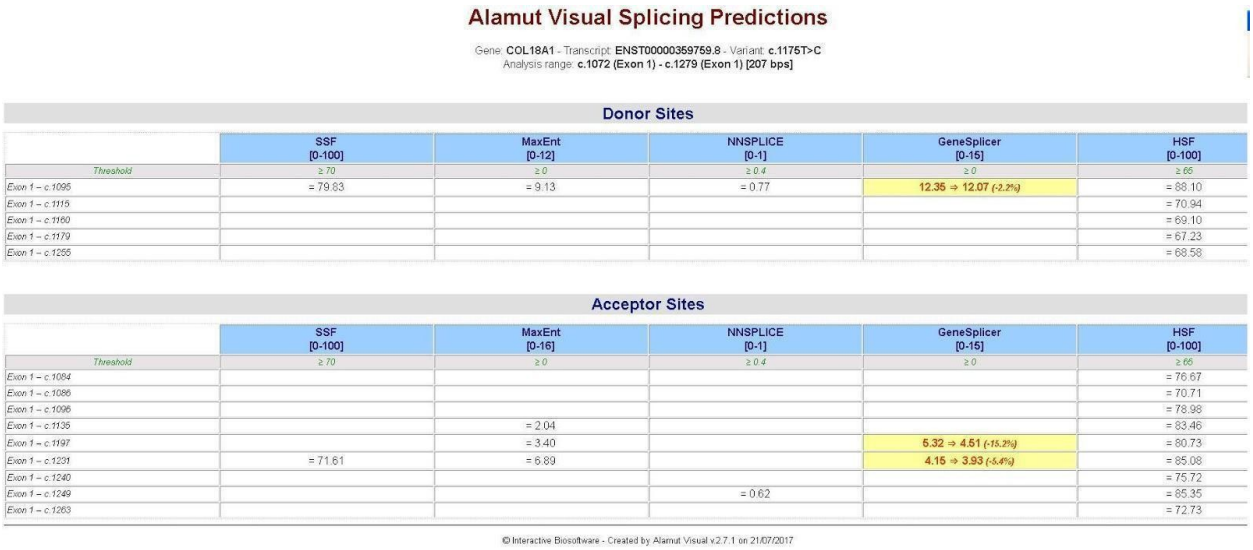


Figure A2: Polyphen-2 summary of variant c.1175T>C in *COL18A1*. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species.



Alamut Visual Splicing Predictions

Gene: COL18A1 - Transcript: ENST00000359759.8 - Variant: c.1175T>C
Analysis range: c.1072 (Exon 1) - c.1279 (Exon 1) [207 bps]

Donor Sites

	SSF [0-100]	MaxEnt [0-12]	NNSPLICE [0-1]	GeneSplicer [0-15]	HSF [0-100]
Threshold	≥ 70	≥ 0	≥ 0.4	≥ 0	≥ 66
Exon 1 – c.1095	= 79.83	= 9.13	= 0.77	12.35 ⇒ 12.07 (-2.2%)	= 88.10
Exon 1 – c.1115					= 70.94
Exon 1 – c.1160					= 69.10
Exon 1 – c.1179					= 67.23
Exon 1 – c.1255					= 68.58

Acceptor Sites

	SSF [0-100]	MaxEnt [0-16]	NNSPLICE [0-1]	GeneSplicer [0-15]	HSF [0-100]
Threshold	≥ 70	≥ 0	≥ 0.4	≥ 0	≥ 66
Exon 1 – c.1084					= 76.67
Exon 1 – c.1086					= 70.71
Exon 1 – c.1096					= 78.98
Exon 1 – c.1135		= 2.04			= 83.46
Exon 1 – c.1197		= 3.40		5.32 ⇒ 4.51 (-15.2%)	= 80.73
Exon 1 – c.1231	= 71.61	= 6.89		4.15 ⇒ 3.93 (-5.4%)	= 85.08
Exon 1 – c.1240					= 75.72
Exon 1 – c.1249			= 0.62		= 85.35
Exon 1 – c.1263					= 72.73

© Interactive Biosoftware - Created by Alamut Visual v.2.7.1 on 21/07/2017

Figure A3: Variant c.1175T>C in *COL18A1* splice predictions. Alamut splice predictor for variant c.1175T>C in *COL18A1,* any large changes in entropy suggests changes in splicing.

Variant Review

Gene: COL18A1 Transcript: ENST00000359759.8

HGVS nomenclature cDNA: c.1175T>C protein: p.Leu392Pro

Historic nomenclature:

Classification: Uncertain Significance Patient ID: Patient C

☐ First Review Previously Reviewed: Change in Classification? n/a

Disease Specific Databases

additional databases can be added in "Other Information" if applicable

ClinVar	n/a	Times Reported:
HGMD	n/a	Times Reported:
LSDB 1	enter DB (if applicable)	
result 1	n/a	Times Reported:
LSDB 2	enter DB (if applicable)	
result 2	n/a	Times Reported:

Alamut version 2.7.1

GRCh

dbSNP rs #114463382

MAF 0.015 Population Aggregated

☑ ExAc ☐ ESP ☐ 1000 Genomes

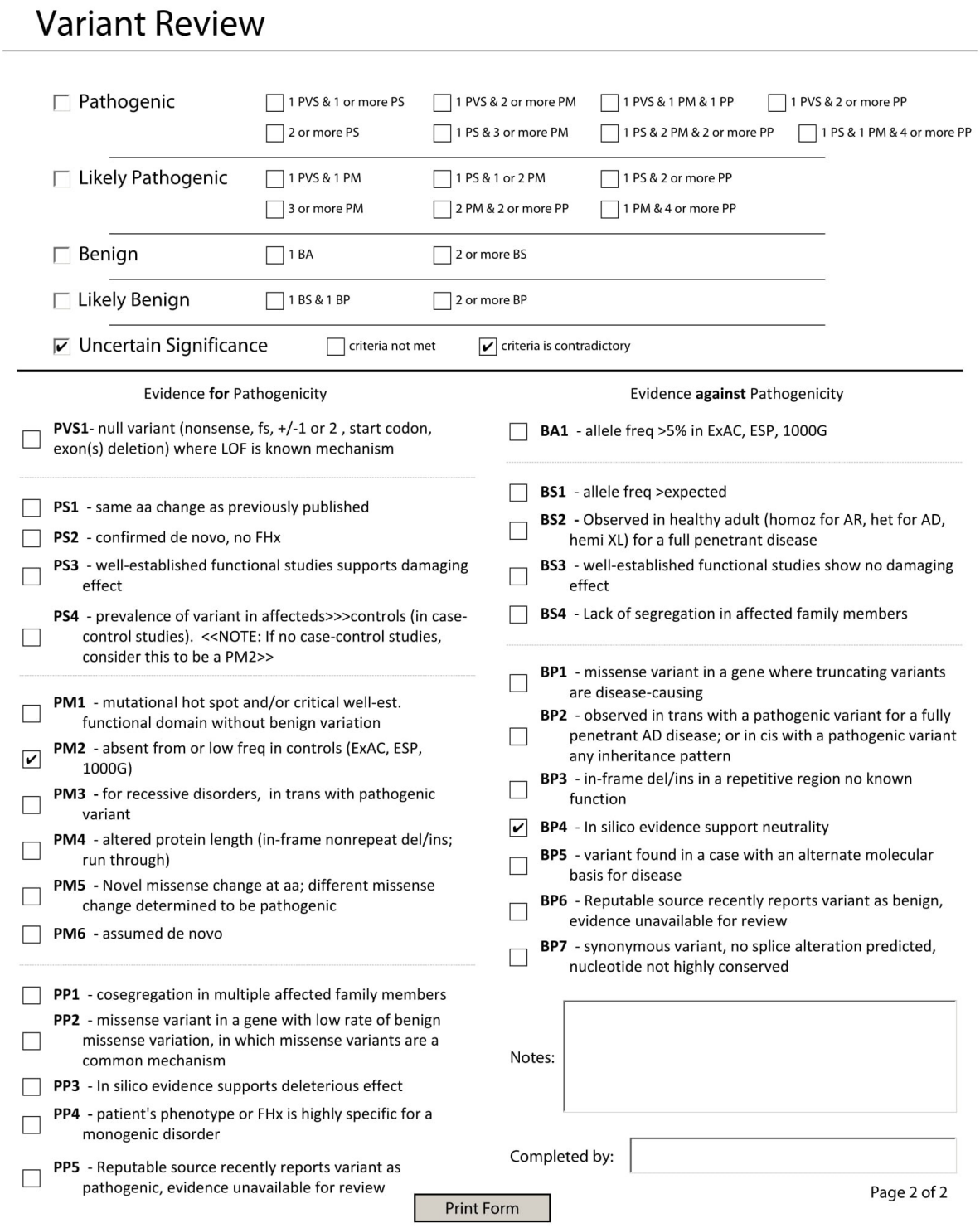
Align GVGD	Class: CO	(Less Likely C0 --> C65 Most Likely)		
SIFT	Change : Tolerated		Score: 0.25	
Mutation Taster	Change : polymorphism		p-value: 0	
Polyphen-2	HumVar: Probably Damagin		Sensitivity: 0.00	Specificity: 1.00

Splicing Predictions ☑ No splice change predicted ☐ Splice change predicted

which programs? (if change predicted)

Other Information

(i.e Google Search Literature Search, other databases)



Variant Review

Pathogenic
1 PVS & 1 or more PS | 1 PVS & 2 or more PM | 1 PVS & 1 PM & 1 PP | 1 PVS & 2 or more PP
2 or more PS | 1 PS & 3 or more PM | 1 PS & 2 PM & 2 or more PP | 1 PS & 1 PM & 4 or more PP

Likely Pathogenic
1 PVS & 1 PM | 1 PS & 1 or 2 PM | 1 PS & 2 or more PP
3 or more PM | 2 PM & 2 or more PP | 1 PM & 4 or more PP

Benign
1 BA | 2 or more BS

Likely Benign
1 BS & 1 BP | 2 or more BP

☑ Uncertain Significance
☐ criteria not met | ☑ criteria is contradictory

Evidence **for** Pathogenicity

☐ **PVS1**- null variant (nonsense, fs, +/-1 or 2 , start codon, exon(s) deletion) where LOF is known mechanism

☐ **PS1** - same aa change as previously published
☐ **PS2** - confirmed de novo, no FHx
☐ **PS3** - well-established functional studies supports damaging effect
☐ **PS4** - prevalence of variant in affecteds>>>controls (in case-control studies). <<NOTE: If no case-control studies, consider this to be a PM2>>

☐ **PM1** - mutational hot spot and/or critical well-est. functional domain without benign variation
☑ **PM2** - absent from or low freq in controls (ExAC, ESP, 1000G)
☐ **PM3** - for recessive disorders, in trans with pathogenic variant
☐ **PM4** - altered protein length (in-frame nonrepeat del/ins; run through)
☐ **PM5** - Novel missense change at aa; different missense change determined to be pathogenic
☐ **PM6** - assumed de novo

☐ **PP1** - cosegregation in multiple affected family members
☐ **PP2** - missense variant in a gene with low rate of benign missense variation, in which missense variants are a common mechanism
☐ **PP3** - In silico evidence supports deleterious effect
☐ **PP4** - patient's phenotype or FHx is highly specific for a monogenic disorder
☐ **PP5** - Reputable source recently reports variant as pathogenic, evidence unavailable for review

Evidence **against** Pathogenicity

☐ **BA1** - allele freq >5% in ExAC, ESP, 1000G

☐ **BS1** - allele freq >expected
☐ **BS2** - Observed in healthy adult (homoz for AR, het for AD, hemi XL) for a full penetrant disease
☐ **BS3** - well-established functional studies show no damaging effect
☐ **BS4** - Lack of segregation in affected family members

☐ **BP1** - missense variant in a gene where truncating variants are disease-causing
☐ **BP2** - observed in trans with a pathogenic variant for a fully penetrant AD disease; or in cis with a pathogenic variant any inheritance pattern
☐ **BP3** - in-frame del/ins in a repetitive region no known function
☑ **BP4** - In silico evidence support neutrality
☐ **BP5** - variant found in a case with an alternate molecular basis for disease
☐ **BP6** - Reputable source recently reports variant as benign, evidence unavailable for review
☐ **BP7** - synonymous variant, no splice alteration predicted, nucleotide not highly conserved

Notes:

Completed by:

Page 2 of 2

Print Form

Figure A4: ACMG variant summary classification form for c.1175T>C in *COL18A1*. Resulting classification of this variant is unknown significance.

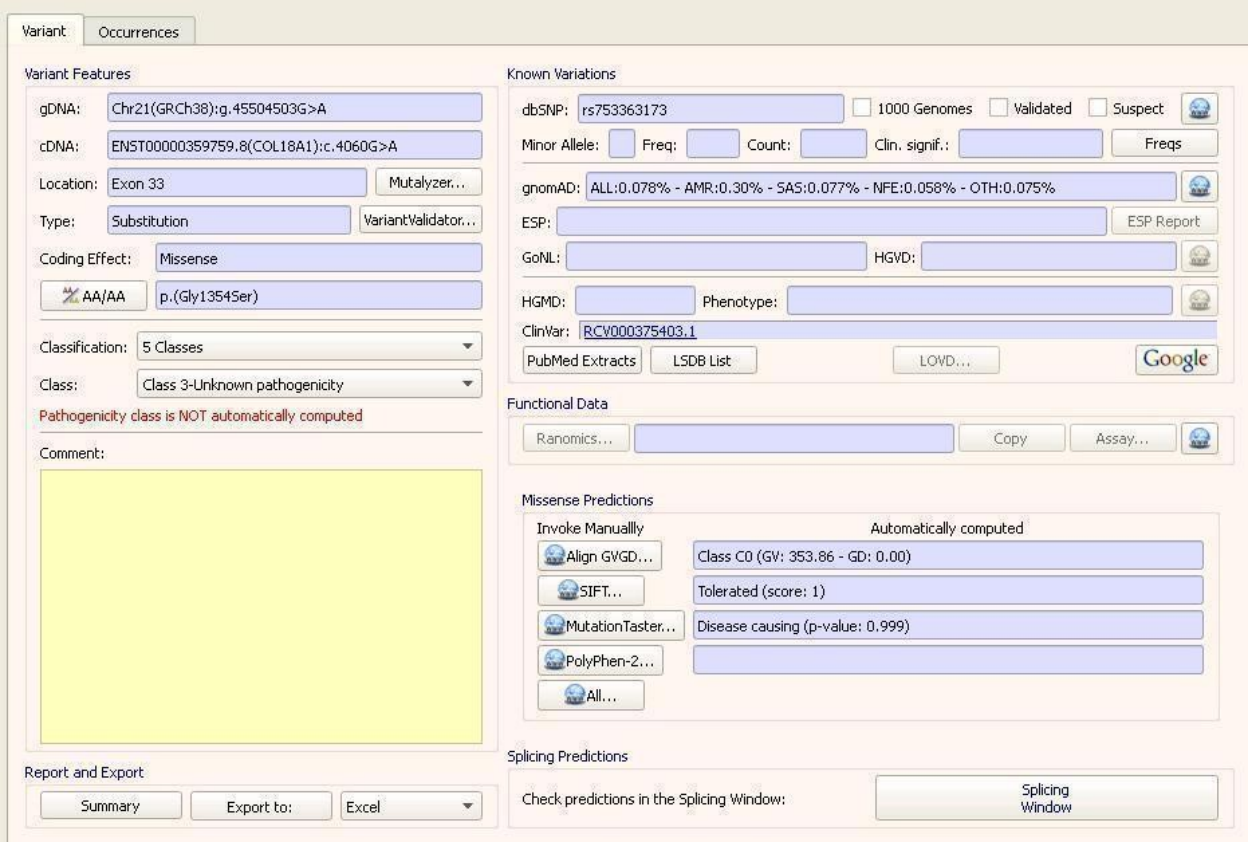


Figure A5: *COL18A1* variant c.4060T>C Alamut features. Features analysis from Alamut software for variant c.4060T>C in *COL18A1* with reported MAF and pathogenic predictions from Align GVGD, SIFT, MutationTaster.

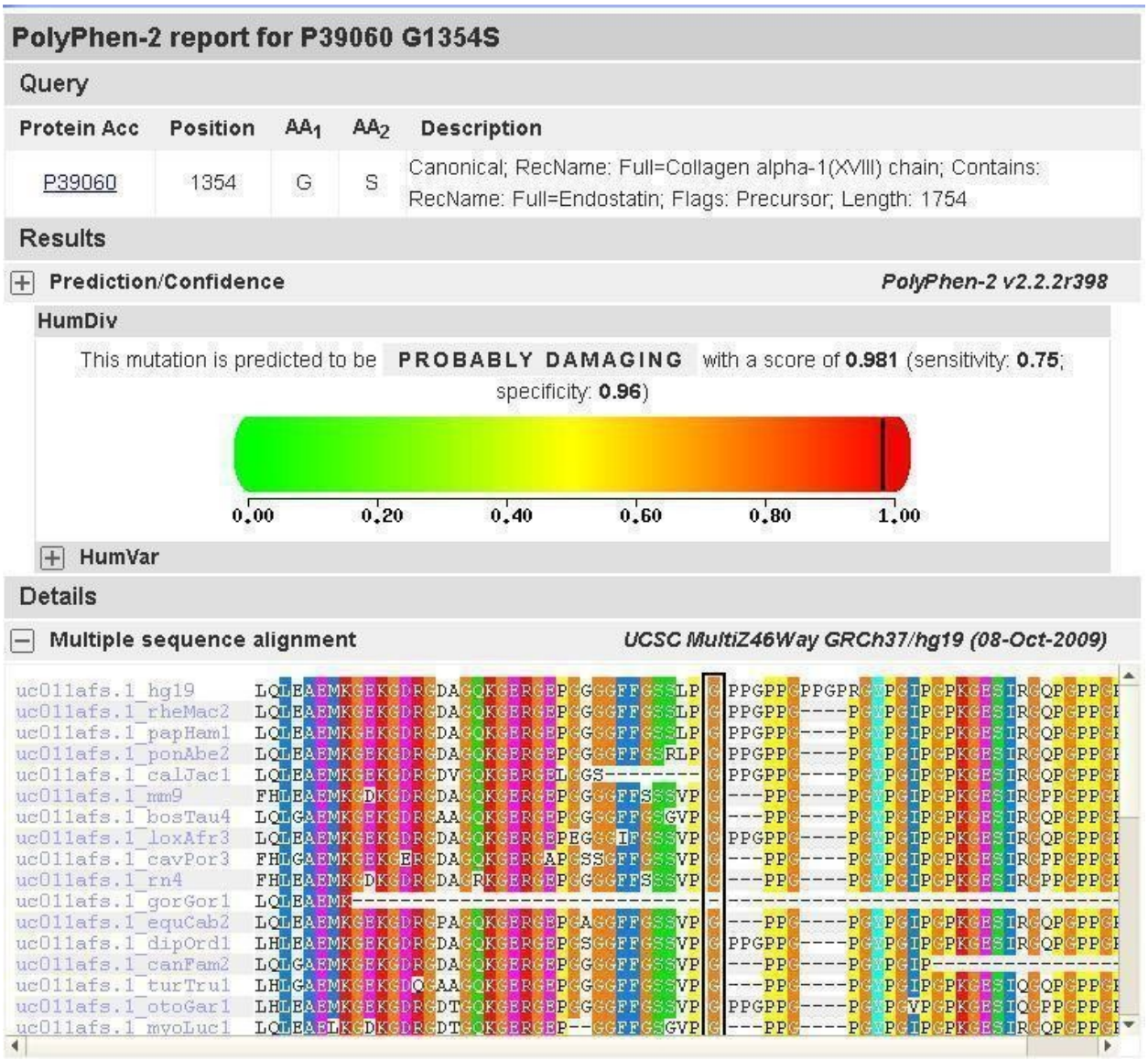


Figure A6: Polyphen-2 summary of variant c.4060T>C in *COL18A1*. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species.

Donor Sites

	SSF [0-100]	MaxEnt [0-12]	NNSPLICE [0-1]	GeneSplicer [0-24]	HSF [0-100]
Threshold	*≥ 70*	*≥ 0*	*≥ 0.4*	*≥ 0*	*≥ 65*
Intron 32 – c.3973-24					= 69.93
Intron 32 – c.3973-17					= 67.14
Exon 33 – c.3991	= 75.17	= 4.44		4.63 ⇒ 4.66 *(+0.6%)*	= 82.83
Exon 33 – c.4006					= 67.65
Exon 33 – c.4093					= 67.47
Exon 33 – c.4113 N	= 78.95	= 9.09	= 0.99	9.23 ⇒ 8.99 *(-2.6%)*	= 84.67
Intron 33 – c.4113+40					= 77.62
Intron 33 – c.4113+46					= 69.10

Natural Splice Site

Acceptor Sites

	SSF [0-100]	MaxEnt [0-16]	NNSPLICE [0-1]	GeneSplicer [0-21]	HSF [0-100]
Threshold	*≥ 70*	*≥ 0*	*≥ 0.4*	*≥ 0*	*≥ 65*
Intron 32 – c.3973-15					= 69.60
Exon 33 – c.3973 N	= 87.96	= 9.57	= 0.74	= 10.45	= 93.10
Exon 33 – c.3979					= 65.36
Exon 33 – c.3982					= 65.84
Exon 33 – c.3992					= 66.98
Exon 33 – c.4001					= 78.70
Exon 33 – c.4006					= 75.68
Exon 33 – c.4010					= 68.22
Exon 33 – c.4017					= 67.18
Exon 33 – c.4053	= 86.11	= 3.84	0.80 ⇒ 0.68 *(-14.3%)*		= 87.40
Exon 33 – c.4062					— ⇒ 80.76
Exon 33 – c.4079	= 76.36	8.25 ⇒ 6.76 *(-18.0%)*	0.75 ⇒ 0.71 *(-4.9%)*	5.13 ⇒ 4.73 *(-7.9%)*	= 88.98
Exon 33 – c.4088		= 1.10			= 86.91
Intron 33 – c.4113+2	= 73.17	= 0.74			= 81.73
Intron 33 – c.4113+6					= 70.69
Intron 33 – c.4113+12					= 78.00
Intron 33 – c.4113+21					= 85.19
Intron 33 – c.4113+25					= 81.73

Figure A7: c.4060T>C in *COL18A1* splice predictions. Alamut splice predictor for variant c.4060T>C in *COL18A1,* any large changes in entropy suggests changes in splicing..

Variant Review

Gene: COL18A1 Transcript: ENST00000359759.8

HGVS nomenclature cDNA: c.4060G>A protein: p.Gly1354Ser

Historic nomenclature:

Classification: Benign Patient ID: Patient C

☐ First Review Previously Reviewed: Change in Classification? n/a

Disease Specific Databases

additional databases can be added in "Other Information" if applicable

ClinVar	n/a	Times Reported:
HGMD	n/a	Times Reported:
LSDB 1	enter DB (if applicable)	
result 1	n/a	Times Reported:
LSDB 2	enter DB (if applicable)	
result 2	n/a	Times Reported:

Alamut version 2.7.1

GRCh

dbSNP rs #753363173

MAF 0.078% Population

☐ ExAc ☐ ESP ☐ 1000 Genomes

Align GVGD Class: C0 (Less Likely C0 --> C65 Most Likely)

SIFT Change : Tolerated Score: 1

Mutation Taster Change : disease causing p-value: 0.999

Polyphen-2 HumVar: Probably Damagi Sensitivity: 0.75 Specificity: 0.96

Splicing Predictions ☐ No splice change predicted ☐ Splice change predicted

MaxEnt, NNSplice, GeneSplicer, HSF

Other Information

(i.e Google Search Literature Search, other databases)

Page 1 of 2

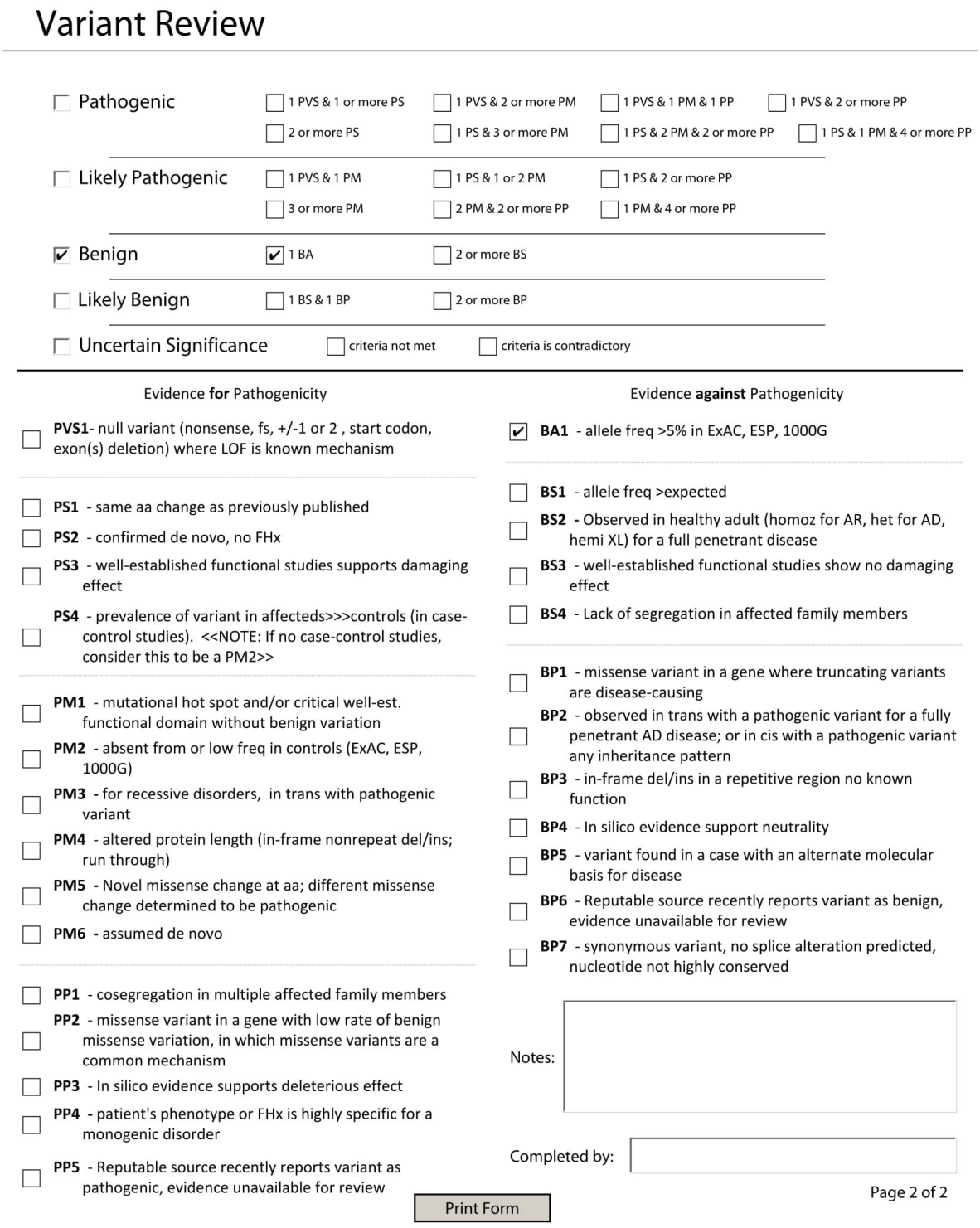


Figure A8: ACMG variant summary classification form for c.4060T>C in *COL18A1*. Resulting classification of this variant is benign.

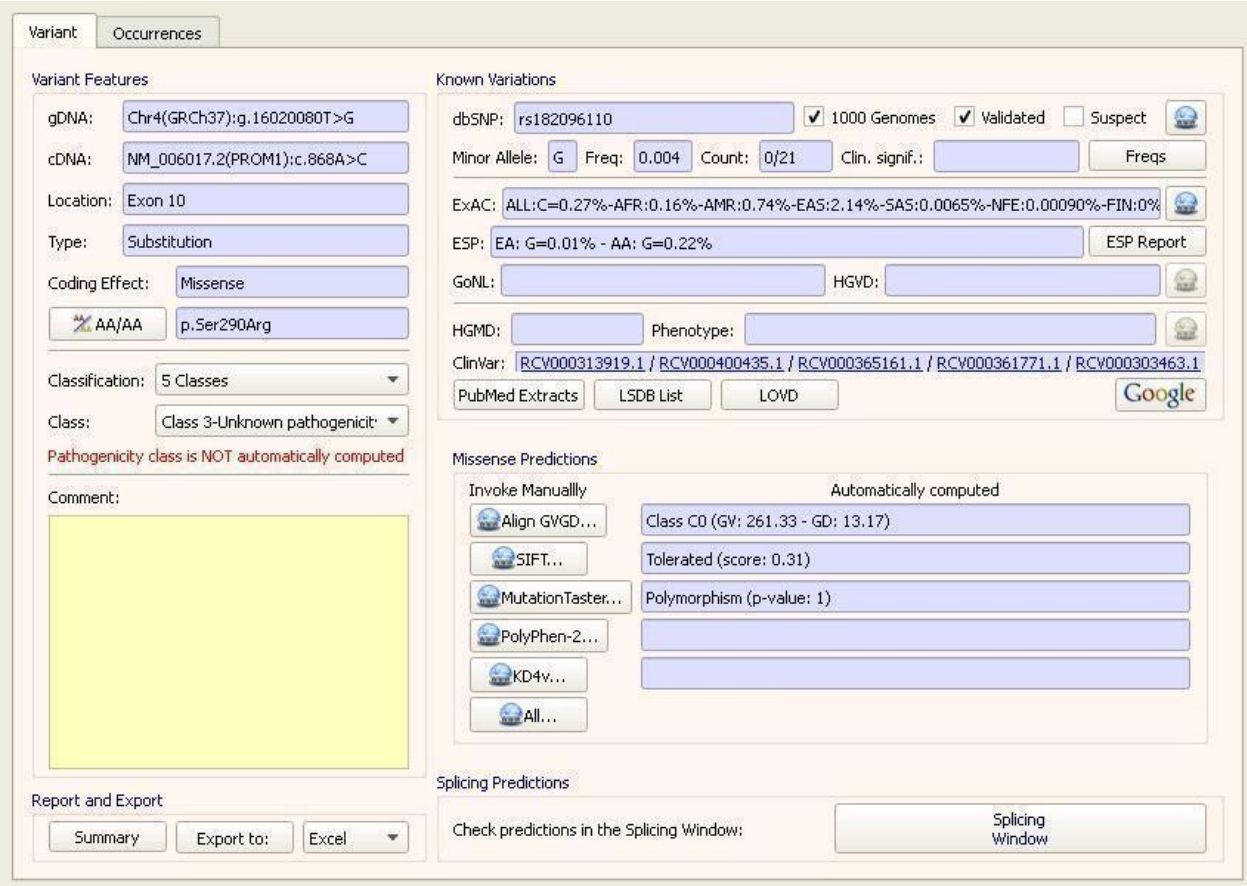


Figure A9: *PROM1* variant c.868A>C Alamut features. *PROM1* variant c.868A>C features analysis from Alamut software with reported MAF and pathogenic predictions from Align GVGD, SIFT, MutationTaster.

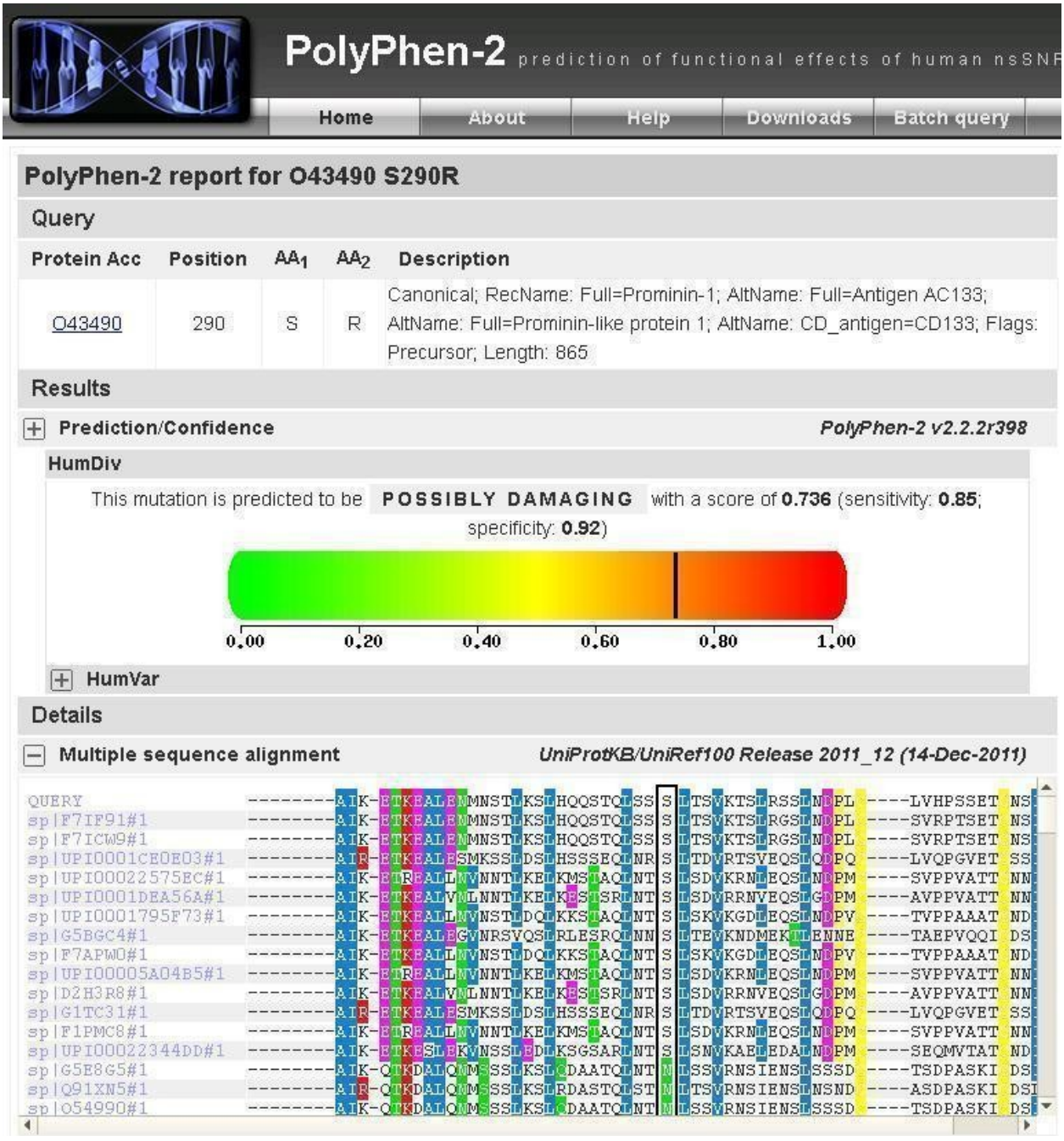


Figure A10: Polyphen-2 summary of variant c.868A>C in *PROM1*. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species.

Gene: **PROM1** - Transcript: **NM_006017.2** - Variant: **c.868A>C**
Analysis range: **c.785-16 (Intron 9) - c.968 (Exon 10) [199 bps]**

Donor Sites

	SSF [0-100]	MaxEnt [0-12]	NNSPLICE [0-1]	GeneSplicer [0-15]	HSF [0-100]
Threshold	≥ 70	≥ 0	≥ 0.4	≥ 0	≥ 65
Exon 10 – c.806					= 66.67
Exon 10 – c.860					= 65.32
Exon 10 – c.868					66.75 ⇒ 66.31 *(-0.7%)*
Exon 10 – c.879					= 67.52
Exon 10 – c.896					= 73.97
Exon 10 – c.924					= 71.14
Exon 10 – c.937					= 73.76

Acceptor Sites

	SSF [0-100]	MaxEnt [0-16]	NNSPLICE [0-1]	GeneSplicer [0-15]	HSF [0-100]
Threshold	≥ 70	≥ 0	≥ 0.4	≥ 0	≥ 65
Exon 10 – c.785 N	= 77.86			= 2.72	= 78.84
Exon 10 – c.793					= 74.16
Exon 10 – c.796					= 66.21
Exon 10 – c.803					= 67.01
Exon 10 – c.805					= 67.95
Exon 10 – c.814					= 68.65
Exon 10 – c.826					= 72.99
Exon 10 – c.835					= 74.67
Exon 10 – c.837					= 69.85
Exon 10 – c.852					= 68.71
Exon 10 – c.859					= 76.43
Exon 10 – c.864					= 71.93
Exon 10 – c.867					78.32 ⇒ 78.25 *(-0.1%)*
Exon 10 – c.870					80.44 ⇒ —
Exon 10 – c.879					79.53 ⇒ 81.60 *(+2.6%)*
Exon 10 – c.891					= 67.15
Exon 10 – c.939		= 0.04			= 74.09

Figure A11: *PROM1* variant c.868A>C splicing prediction. Alamut splice predictor for variant c.868A>C in *PROM1,* any large changes in entropy suggests changes in splicing.

Variant Review

Gene: PROM1 Transcript: NM_006017.2

HGVS nomenclature cDNA: c.868A>C protein: p.Ser290Arg

Historic nomenclature:

Classification: Uncertain Significance Patient ID: Patient C

☐ First Review Previously Reviewed: Change in Classification? n/a

Disease Specific Databases

additional databases can be added in "Other Information" if applicable

ClinVar	n/a	Times Reported:
HGMD	n/a	Times Reported:
LSDB 1	enter DB (if applicable)	
result 1	n/a	Times Reported:
LSDB 2	enter DB (if applicable)	
result 2	n/a	Times Reported:

Alamut version 2.7.1

GRCh

dbSNP rs #182096110

MAF 0.25% Population Aggregated

☐ ExAc ☐ ESP ☐ 1000 Genomes

Align GVGD	Class: C0	(Less Likely C0 --> C65 Most Likely)		
SIFT	Change : Tolerated		Score: 0.31	
Mutation Taster	Change : polymorphism		p-value: 1	
Polyphen-2	HumVar: Probably Damagin		Sensitivity: 0.85	Specificity: 0.92

Splicing Predictions ☐ No splice change predicted ☐ Splice change predicted

which programs? (if change predicted)

Other Information

(i.e Google Search Literature Search, other databases)

Page 1 of 2

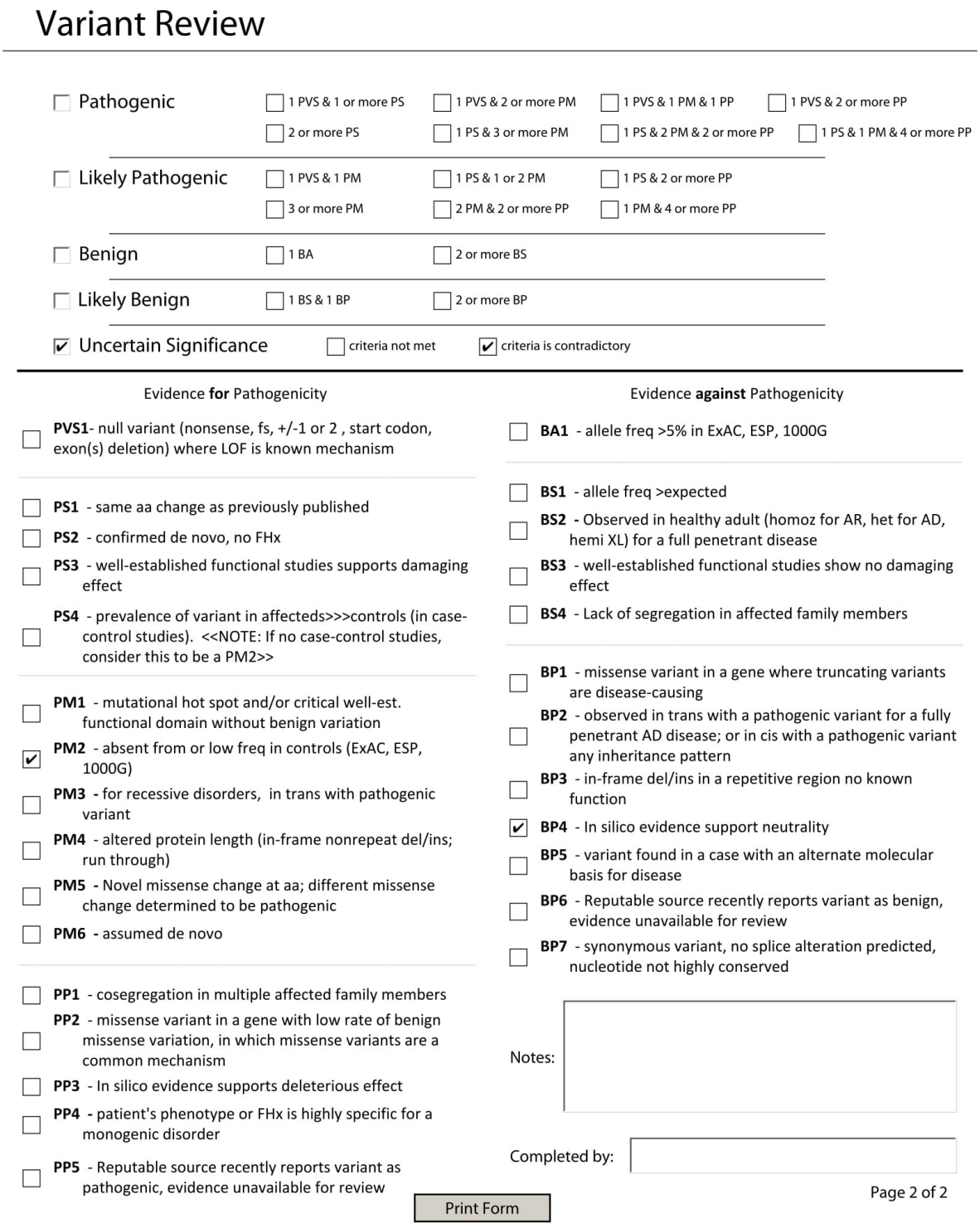


Figure A12: ACMG variant summary classification form for c.868A>C in *PROM1*. Resulting classification of this variant is unknown significance.

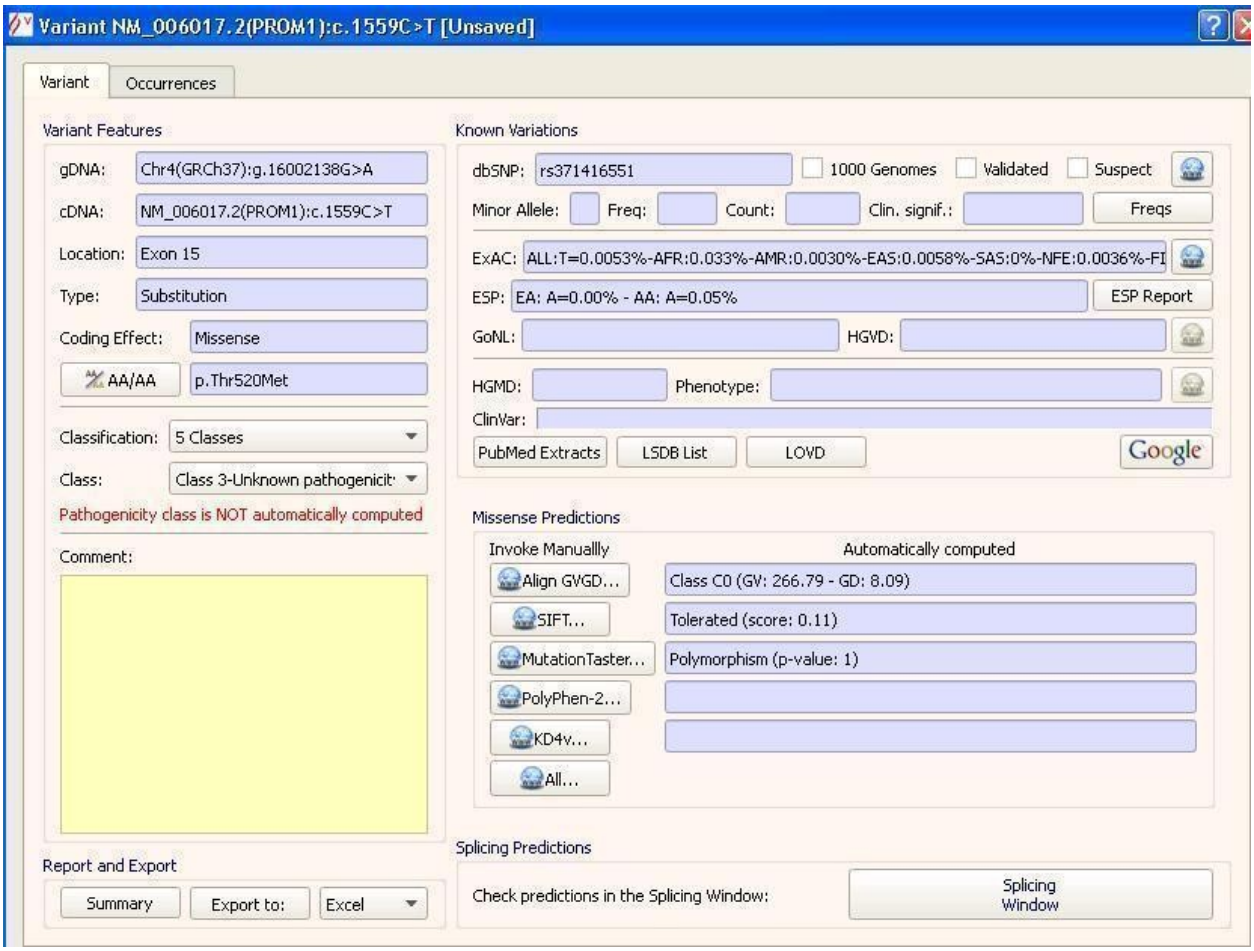


Figure A13: *PROM1* variant c.1559C>T Alamut features. *PROM1* variant c.1559C>T features analysis from Alamut software with reported MAF and pathogenic predictions from Align GVGD, SIFT, MutationTaster.

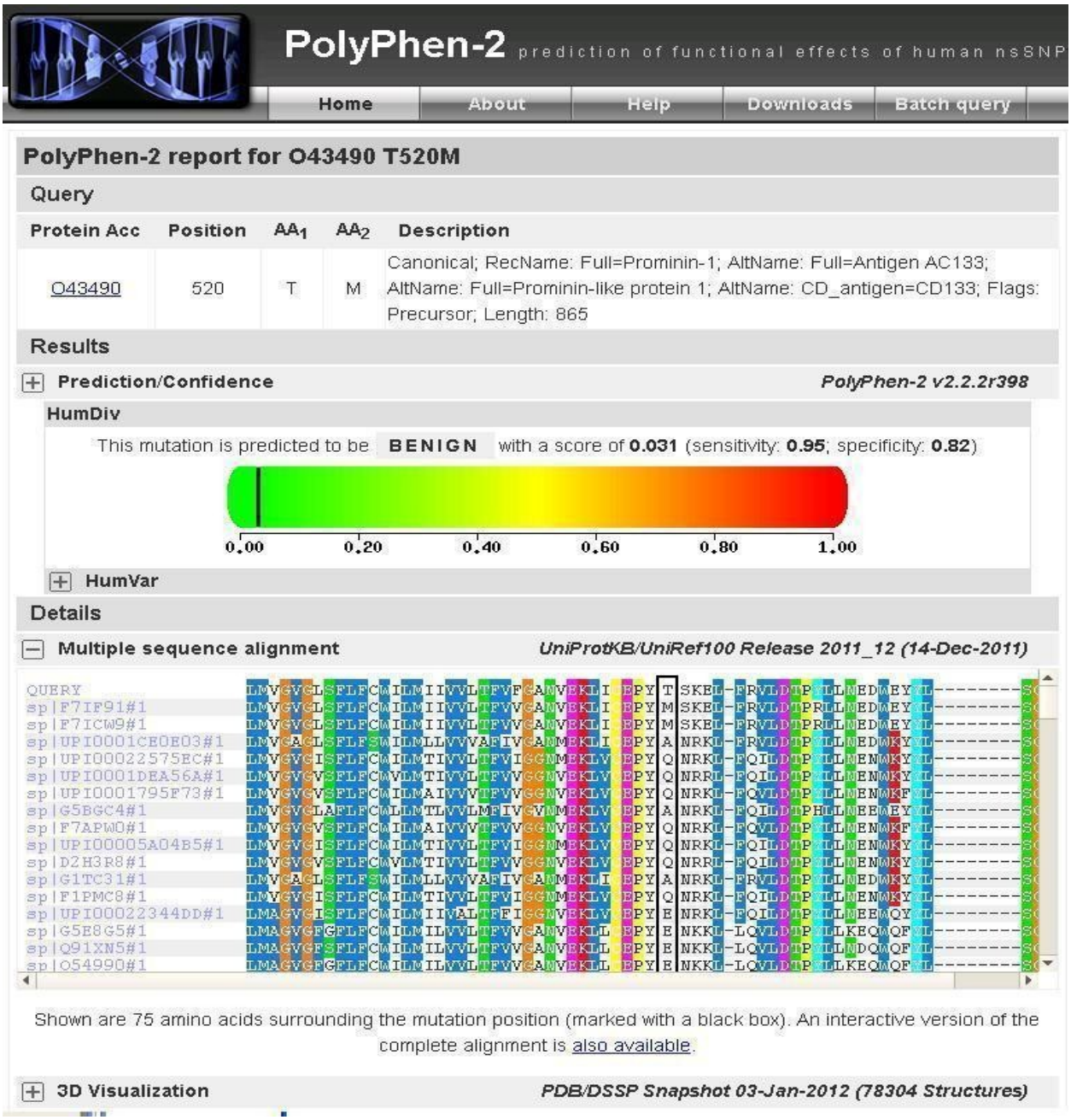
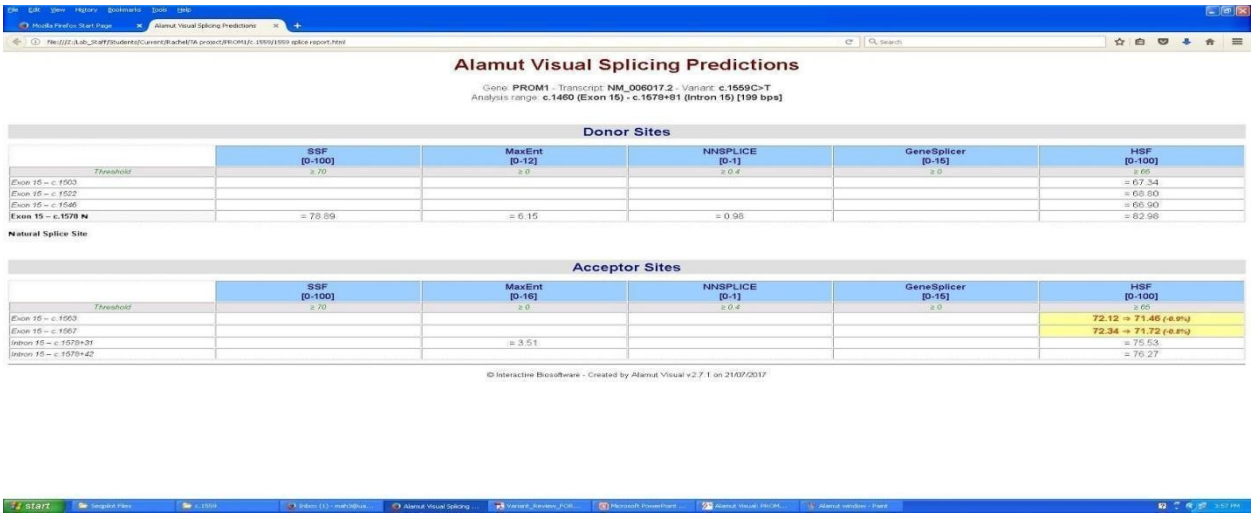


Figure A14: Polyphen-2 summary of variant c.1559C>T in *PROM1*. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species.



Alamut Visual Splicing Predictions

Gene: PROM1 - Transcript: NM_006017.2 - Variant: c.1559C>T
Analysis range: c.1460 (Exon 15) - c.1578+81 (Intron 15) [199 bps]

Donor Sites

	SSF [0-100]	MaxEnt [0-12]	NNSPLICE [0-1]	GeneSplicer [0-15]	HSF [0-100]
Threshold	≥ 70	≥ 0	≥ 0.4	≥ 0	≥ 66
Exon 15 – c.1503					= 67.34
Exon 15 – c.1522					= 68.80
Exon 15 – c.1546					= 66.90
Exon 15 – c.1578 N	= 78.89	= 6.15	= 0.98		= 82.98

Natural Splice Site

Acceptor Sites

	SSF [0-100]	MaxEnt [0-16]	NNSPLICE [0-1]	GeneSplicer [0-15]	HSF [0-100]
Threshold	≥ 70	≥ 0	≥ 0.4	≥ 0	≥ 66
Exon 15 – c.1563					72.12 ⇒ 71.46 (-0.9%)
Exon 15 – c.1567					72.34 ⇒ 71.72 (-0.8%)
Intron 15 – c.1578+31		= 3.51			= 75.53
Intron 15 – c.1578+42					= 76.27

© Interactive Biosoftware - Created by Alamut Visual v2.7.1 on 21/07/2017

Figure A15: *PROM1* variant c.1559C>T splicing prediction. Alamut splice predictor for variant c.1559C>T in *PROM1,* any large changes in entropy suggests changes in splicing.

Variant Review

Gene: PROM1 Transcript: NM_006017.2

HGVS nomenclature cDNA: c.1559C>T protein: p.Thr520Met

Historic nomenclature:

Classification: Uncertain Significance Patient ID: Patient C

☐ First Review Previously Reviewed: Change in Classification? n/a

Disease Specific Databases

additional databases can be added in "Other Information" if applicable

ClinVar	n/a	Times Reported:
HGMD	n/a	Times Reported:
LSDB 1	enter DB (if applicable)	
result 1	n/a	Times Reported:
LSDB 2	enter DB (if applicable)	
result 2	n/a	Times Reported:

Alamut version 2.7.1

GRCh

dbSNP

rs #371416551

MAF 0.0058% Population Aggregated

☐ ExAc ☐ ESP ☐ 1000 Genomes

Align GVGD	Class: CO	(Less Likely C0 --> C65 Most Likely)	
SIFT	Change: Tolerated	Score: 0.11	
Mutation Taster	Change: polymorphism	p-value: 1.0	
Polyphen-2	HumVar: Benign	Sensitivity: 0.95	Specificity: 0.82

Splicing Predictions ☑ No splice change predicted ☐ Splice change predicted

which programs? (if change predicted)

Other Information

(i.e Google Search Literature Search, other databases)

Page 1 of 2

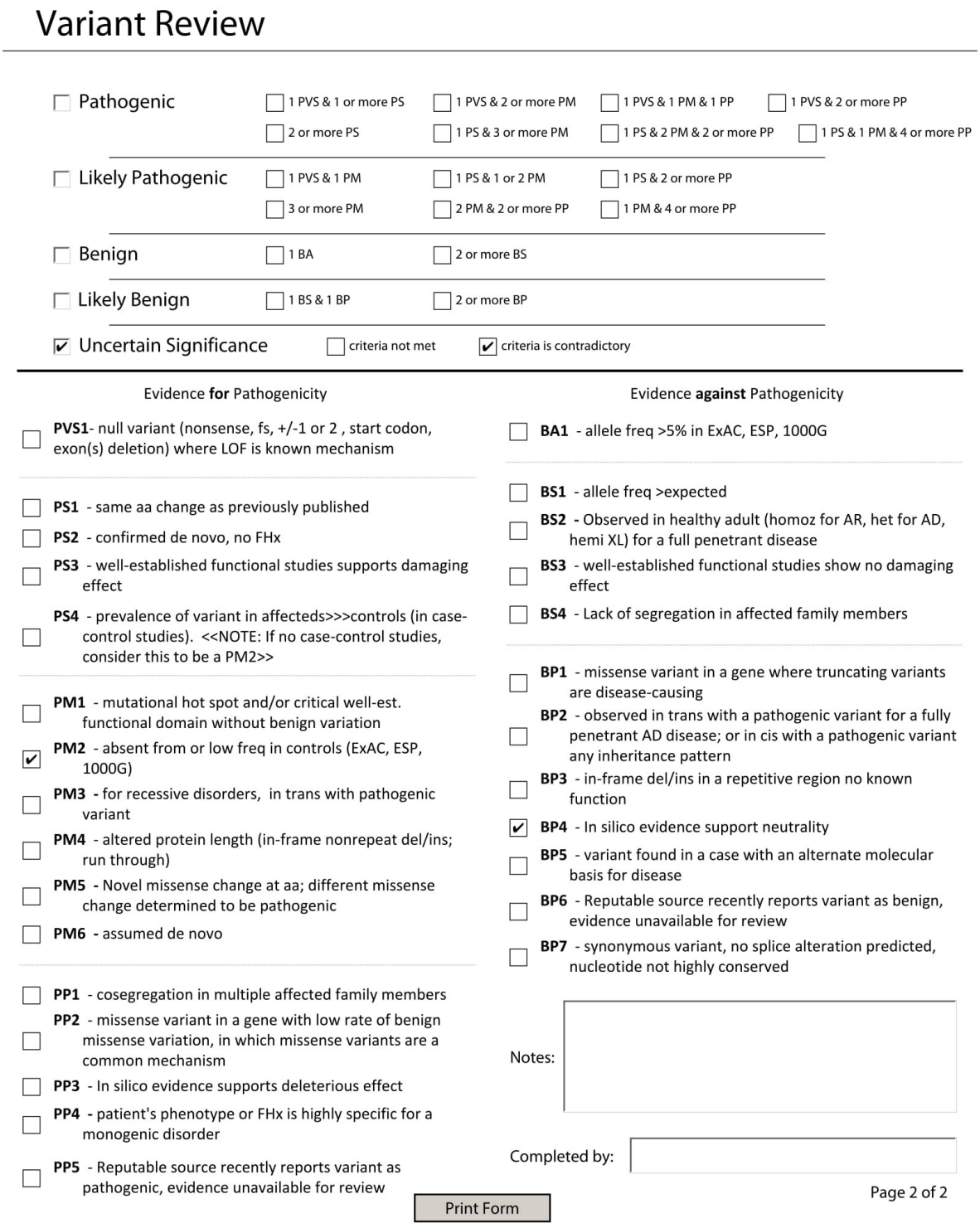


Figure A16: ACMG variant summary classification form for c.1559C>T in *PROM1*. Resulting classification of this variant is unknown significance.

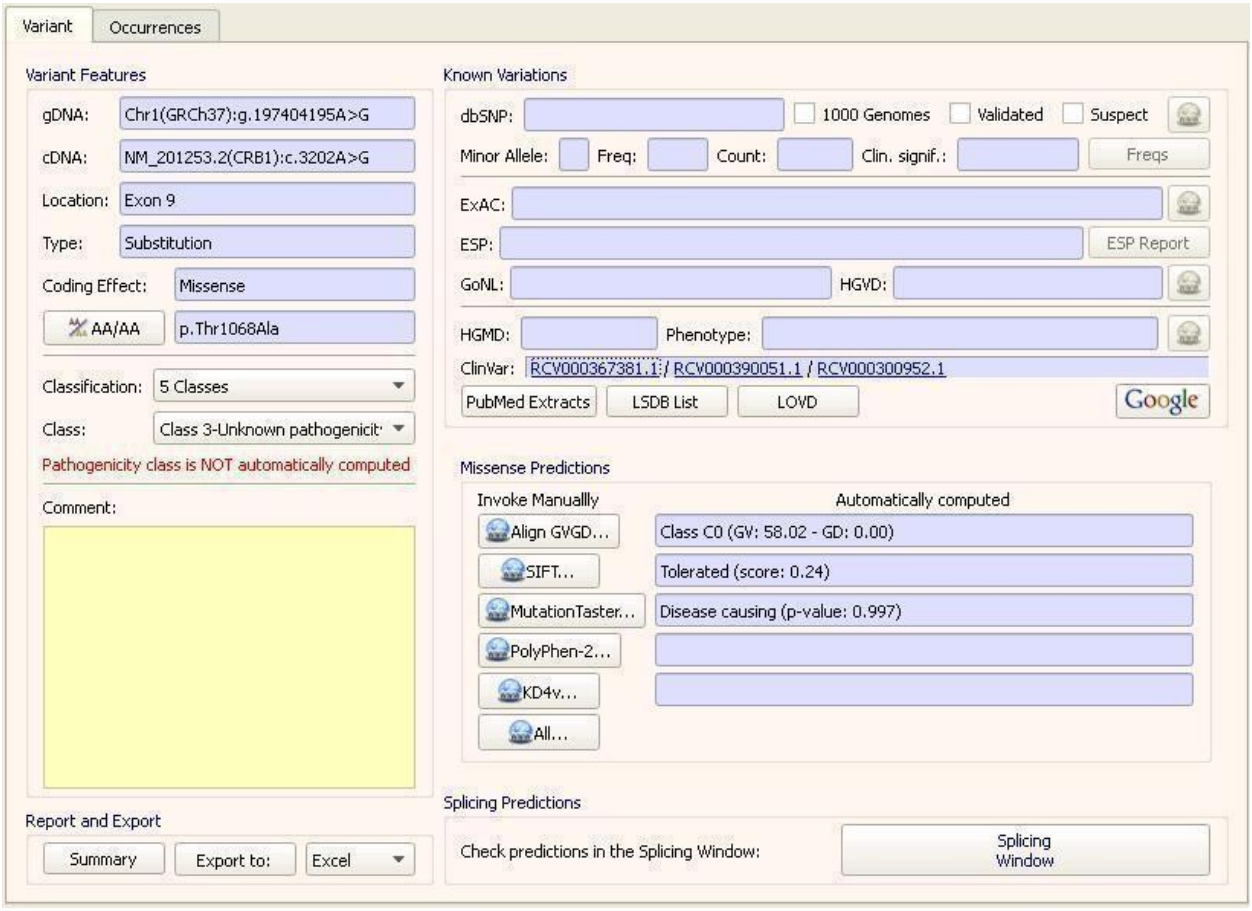


Figure A17: CRB1 variant c.3202A>G Alamut features. *CRB1* variant c.3202A>G features analysis from Alamut software with reported MAF and pathogenic predictions from Align GVGD, SIFT, MutationTaster.

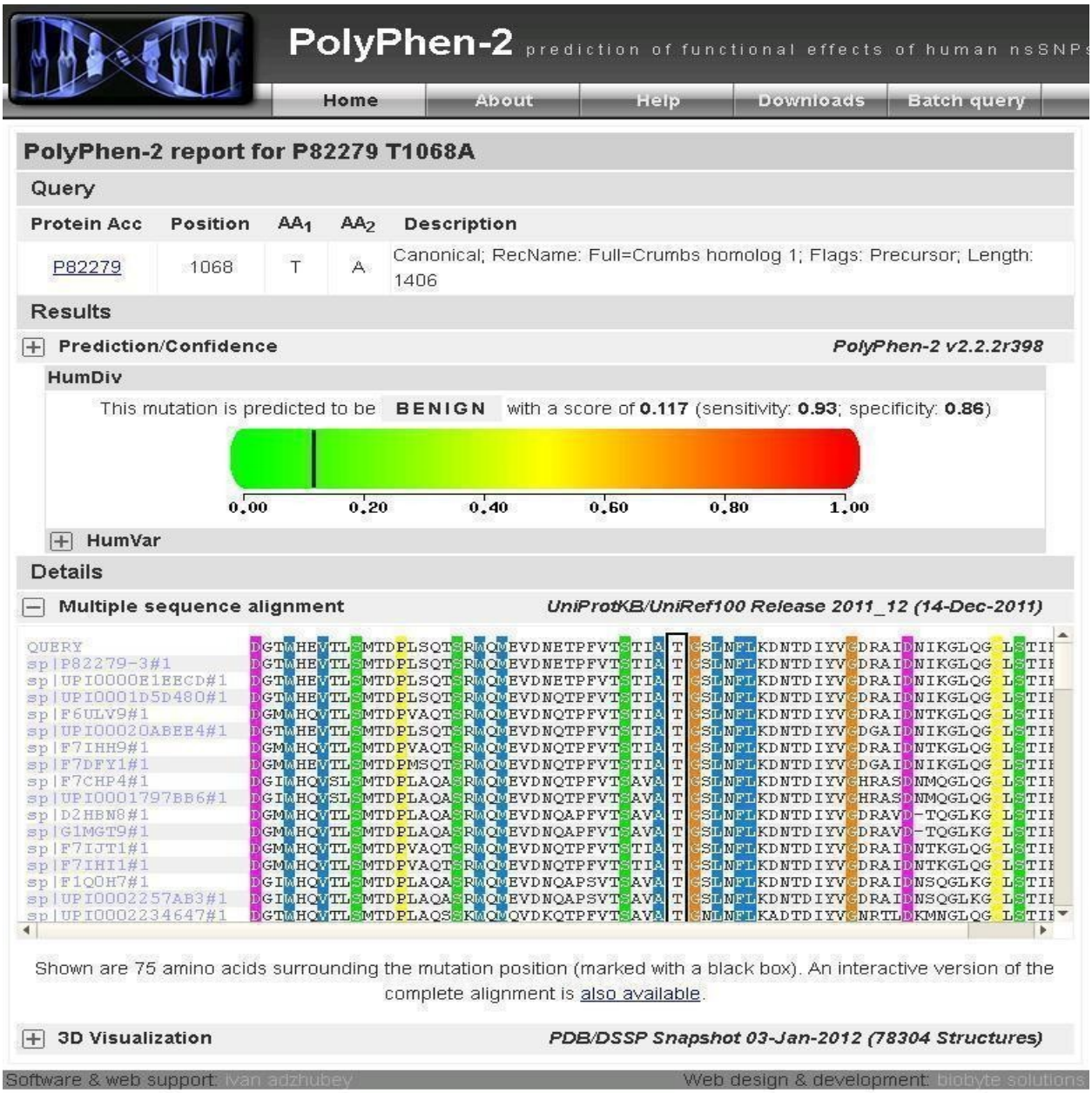


Figure A18: Polyphen-2 summary of variant c.3202A>G in *CRB1*. Results from Polyphen-2 prediction software using HumDiv and the variant's conservation through different species.

Gene: CRB1 - Transcript: NM_201253.2 - Variant: c.3202A>G
Analysis range: c.3099 (Exon 9) - c.3306 (Exon 9) [207 bps]

Donor Sites

	SSF [0-100]	MaxEnt [0-12]	NNSPLICE [0-1]	GeneSplicer [0-15]	HSF [0-100]
Threshold	*≥ 70*	*≥ 0*	*≥ 0.4*	*≥ 0*	*≥ 66*
Exon 9 – c.3101	= 70.01				
Exon 9 – c.3108					= 72.31
Exon 9 – c.3149		= 1.45		— ⇒ 0.49	= 76.15
Exon 9 – c.3152	= 73.67				
Exon 9 – c.3183					= 67.05
Exon 9 – c.3243					= 75.32
Exon 9 – c.3284					= 72.21

Acceptor Sites

	SSF [0-100]	MaxEnt [0-16]	NNSPLICE [0-1]	GeneSplicer [0-15]	HSF [0-100]
Threshold	*≥ 70*	*≥ 0*	*≥ 0.4*	*≥ 0*	*≥ 66*
Exon 9 – c.3128	= 84.37	= 6.96	= 0.66	1.63 ⇒ 1.74 (+6.5%)	= 87.43
Exon 9 – c.3142	= 73.06	= 2.93			= 83.76
Exon 9 – c.3150		= 3.20			= 88.00
Exon 9 – c.3192	= 81.17	= 2.78			= 85.07
Exon 9 – c.3210					73.73 ⇒ 73.80 (+0.1%)
Exon 9 – c.3226	= 81.40	= 2.74			= 79.56
Exon 9 – c.3236					= 76.41
Exon 9 – c.3251	= 72.44				= 74.38
Exon 9 – c.3255					= 76.75
Exon 9 – c.3257					= 65.62
Exon 9 – c.3274					= 72.01
Exon 9 – c.3284					= 72.54
Exon 9 – c.3294					= 66.92
Exon 9 – c.3302					= 70.91

Figure A19: *CRB1* variant c.3202A>G splicing prediction. Alamut splice predictor for variant c.3202A>G in *CRB,* any large changes in entropy suggests changes in splicing.

Variant Review

Gene: CRB1 Transcript: NM_201253.2

HGVS nomenclature cDNA: c.3202A>G protein: p.Thr1068Ala

Historic nomenclature:

Classification: Uncertain Significance Patient ID: Patient C

☐ First Review Previously Reviewed: Change in Classification? n/a

Disease Specific Databases

additional databases can be added in "Other Information" if applicable

ClinVar	Unclassified	Times Reported:	3
HGMD	n/a	Times Reported:	
LSDB 1	enter DB (if applicable)		
result 1	n/a	Times Reported:	
LSDB 2	enter DB (if applicable)		
result 2	n/a	Times Reported:	

Alamut version 2.7.1

GRCh

dbSNP rs #886045787

MAF 0.000006 Population

☐ ExAc ☐ ESP ☐ 1000 Genomes

Align GVGD	Class: CO	(Less Likely C0 --> C65 Most Likely)		
SIFT	Change : Tolerated		Score: 0.24	
Mutation Taster	Change : disease causing		p-value: 0.997	
Polyphen-2	HumVar: Benign		Sensitivity: 0.93	Specificity: 0.86

Splicing Predictions ☐ No splice change predicted ☐ Splice change predicted

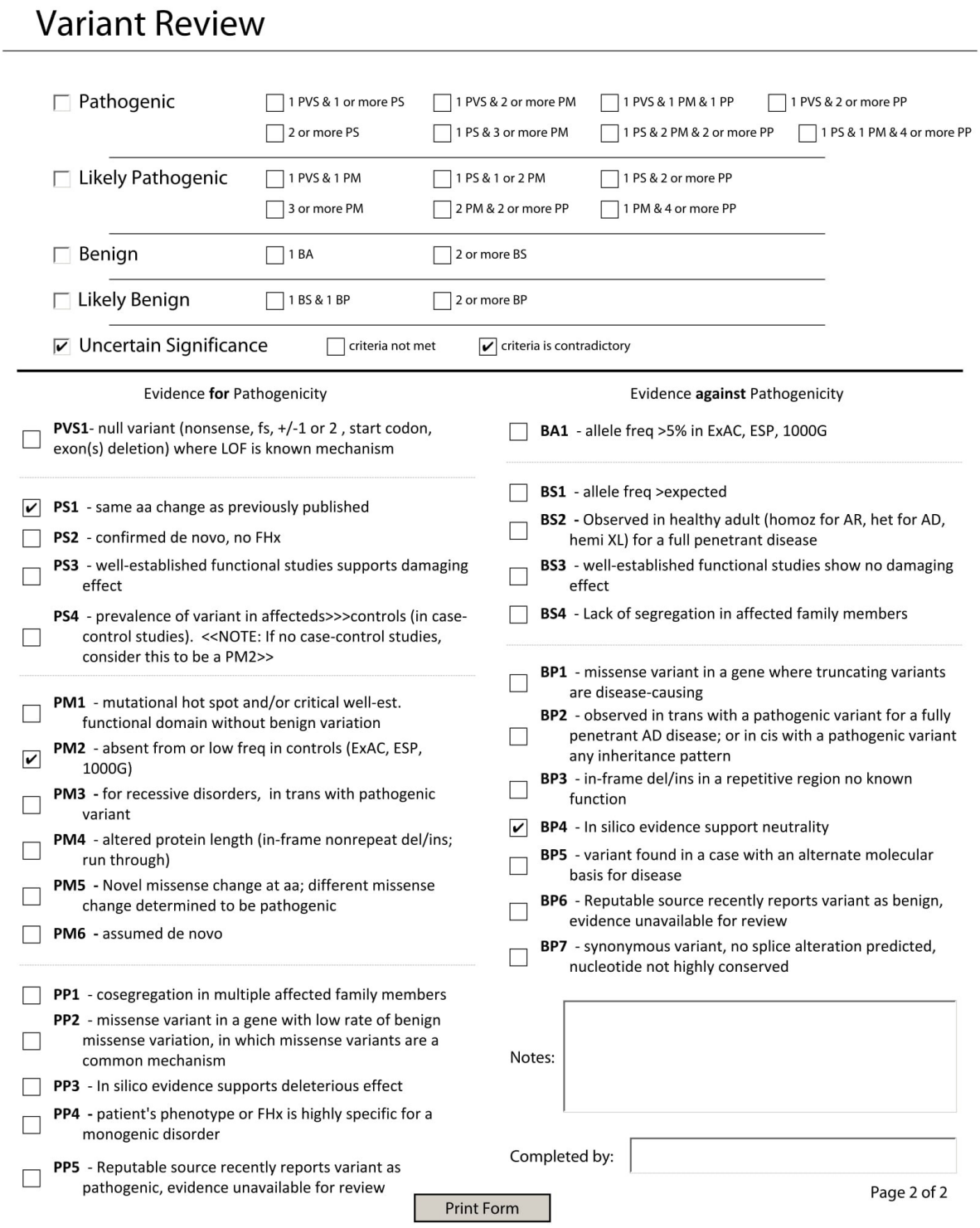
which programs? (if change predicted)

Other Information

(i.e Google Search Literature Search, other databases)

ClinVar reported 3 times; Pigmented paravenous chorioretinal atrophy, Leber congenital amaurosis and retinitis pigmentosa, recessive.

Page 1 of 2



Variant Review

Pathogenic
1 PVS & 1 or more PS | 1 PVS & 2 or more PM | 1 PVS & 1 PM & 1 PP | 1 PVS & 2 or more PP
2 or more PS | 1 PS & 3 or more PM | 1 PS & 2 PM & 2 or more PP | 1 PS & 1 PM & 4 or more PP

Likely Pathogenic
1 PVS & 1 PM | 1 PS & 1 or 2 PM | 1 PS & 2 or more PP
3 or more PM | 2 PM & 2 or more PP | 1 PM & 4 or more PP

Benign
1 BA | 2 or more BS

Likely Benign
1 BS & 1 BP | 2 or more BP

☑ Uncertain Significance
☐ criteria not met | ☑ criteria is contradictory

Evidence **for** Pathogenicity

- ☐ **PVS1**- null variant (nonsense, fs, +/-1 or 2 , start codon, exon(s) deletion) where LOF is known mechanism
- ☑ **PS1** - same aa change as previously published
- ☐ **PS2** - confirmed de novo, no FHx
- ☐ **PS3** - well-established functional studies supports damaging effect
- ☐ **PS4** - prevalence of variant in affecteds>>>controls (in case-control studies). <<NOTE: If no case-control studies, consider this to be a PM2>>
- ☐ **PM1** - mutational hot spot and/or critical well-est. functional domain without benign variation
- ☑ **PM2** - absent from or low freq in controls (ExAC, ESP, 1000G)
- ☐ **PM3** - for recessive disorders, in trans with pathogenic variant
- ☐ **PM4** - altered protein length (in-frame nonrepeat del/ins; run through)
- ☐ **PM5** - Novel missense change at aa; different missense change determined to be pathogenic
- ☐ **PM6** - assumed de novo
- ☐ **PP1** - cosegregation in multiple affected family members
- ☐ **PP2** - missense variant in a gene with low rate of benign missense variation, in which missense variants are a common mechanism
- ☐ **PP3** - In silico evidence supports deleterious effect
- ☐ **PP4** - patient's phenotype or FHx is highly specific for a monogenic disorder
- ☐ **PP5** - Reputable source recently reports variant as pathogenic, evidence unavailable for review

Evidence **against** Pathogenicity

- ☐ **BA1** - allele freq >5% in ExAC, ESP, 1000G
- ☐ **BS1** - allele freq >expected
- ☐ **BS2** - Observed in healthy adult (homoz for AR, het for AD, hemi XL) for a full penetrant disease
- ☐ **BS3** - well-established functional studies show no damaging effect
- ☐ **BS4** - Lack of segregation in affected family members
- ☐ **BP1** - missense variant in a gene where truncating variants are disease-causing
- ☐ **BP2** - observed in trans with a pathogenic variant for a fully penetrant AD disease; or in cis with a pathogenic variant any inheritance pattern
- ☐ **BP3** - in-frame del/ins in a repetitive region no known function
- ☑ **BP4** - In silico evidence support neutrality
- ☐ **BP5** - variant found in a case with an alternate molecular basis for disease
- ☐ **BP6** - Reputable source recently reports variant as benign, evidence unavailable for review
- ☐ **BP7** - synonymous variant, no splice alteration predicted, nucleotide not highly conserved

Notes:

Completed by:

Page 2 of 2

Print Form

Figure A20: ACMG variant summary classification form for c.3202A>G in *CRB1*. Resulting classification of this variant is unknown significance.