*Your file Votre référence*
*ISBN:* 978-0-494-80901-3
*Our file Notre référence*
*ISBN:* 978-0-494-80901-3

# A molecular genetics approach to gene discovery of Mendelian diseases on the island of Newfoundland

By



A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in the Faculty of Medicine, Discipline of Genetics, Memorial University, St. John's, Newfoundland and Labrador, Canada.

2011

## Abstract

### Background

Newfoundland is an island off Canada's east coast that has a unique population for gene discovery. Newfoundland out-port communities were founded by English Protestant or Irish Catholic fisherman, and were subject to isolation due to geographical distance between communities and religious segregation. As such, many genetic isolates formed and have since aided in several gene discoveries.

### Objective

The main objective of this thesis was to identify novel disease-causing genes involved in arrhythmogenic right ventricular cardiomyopathy/dysplasia type 5 (ARVC/D), deafness and breast cancer, by studying Newfoundland families.

### Results

ARVC is an arrhythmic disorder characterized by fat-fibro replacement of the myocardium. The causal gene for one subtype of ARVC, *ARVD5*, was identified by studying 15 Newfoundland ARVC families. Haplotype analysis initially revealed a 2.36 Mb critical region that all affected individuals shared and after screening positional candidate genes, a missense variant, *transmembranre protein 43* (*TMEM43*) c.1073C>T, was determined to be the causal variant.

A large, extended Newfoundland family with non-syndromic hearing loss was suspected to have an X-linked mode of inheritance. Haplotypes spanning the entire X chromosome revealed that only a single region was shared among all affected individuals, a 13.3 Mb region on Xp. One key individual, whose parents were both affected and related, had a

0.96 Mb region of homozygosity within the *dystrophin* (*DMD*) locus, however, segregation of the affected haplotype on the maternal side was not confirmed. Screening cochlea expressed positional candidate genes in the 13.3 Mb region revealed no deleterious variants.

Screening a cohort of 96 Newfoundland breast cancer probands, with a family history of breast cancer, for *BRCA1* and *BRCA2* (*breast cancer susceptibility genes 1 and 2*) deleterious variants (phase 1) identified 15 truncation mutations solving only 15.6% (15/96) of the cohort. In addition, targeted screening of the 15 Newfoundland *BRCA1* and *BRCA2* mutations in 57 newly recruited probands (phase 2) only resulted in one positive screen; one proband screened positive for the *BRCA2* c.6714delACAA mutation. The two *BRCA2* c.6714delACAA probands (one from phase 1 and the other from phase 2) appeared to share a 6.02 Mb haplotype on chromosome 13 between markers *D13S1242* and *D13S220* – suggesting it may be a founder mutation. Furthermore, screening all probands for a low penetrant allele in the *CHK2* gene identified three mutation carriers.

**Conclusions**

The population of Newfoundland provides opportunity for novel gene discovery particularly in autosomal dominant and X-linked disorders. By studying 15 Newfoundland ARVC families we have identified the cause of ARVD5 as a missense mutation in a novel gene, *TMEM43*. This discovery, through mutation screening, now aids in disease diagnosis and identifies at-risk individuals, which allows the appropriate life-saving precautions to be taken, including the use of an implantable cardioverter defibrillator.

A critical region on the X chromosome has been identified that segregates with non-syndromic deafness in a large Newfoundland family. Despite great efforts, the causal gene has not yet been identified and thus there still remains an opportunity for a novel gene discovery.

The Newfoundland population also provides an opportunity to identify a novel gene(s) in breast cancer. Studying the unsolved families (approximately 84% of the *BRCA1* and *BRCA2* screened cohort) and determining which ones originate from the same fishing communities may represent clusters of related families that could be used to search for new genes.

## Acknowledgements

I would like to thank:

**Brad Merner:** for knowing me better than I know myself, never losing faith in my abilities and for keeping me motivated through all the ups and downs. Ultimately, I thank you for your unconditional love and for being my best friend.

**Abby Jayne Merner:** for the pure joy and fulfillment you have given me. You and daddy are my life and together we are one solid unit. Deciding to have a baby in the middle of my graduate studies was an easy decison, but challenging at the same time. Through this accomplishment I would like you to realize that any goal is obtainable if you put your mind to it. Women in science, or in any field for that matter, are essential. Do not let anything hold you back.

**My parents (Jane and Neville Hackett):** for providing me with an amazing childhood. Your love and acceptance provided me with a stable household that has influenced who I am today. Also, I thank you for being such amazing grandparents and helping me with Abby while writing my thesis.

**My parents-in-law:** for accepting me into your family with open arms and helping take care of Abby as I finished my doctoral program.

**The Young Lab:** for being great, supportive co-workers that truly became amazing friends. I will never forget our accomplishments, friendly competitions (Dante, Sandra and Lance), and fun times together. I loved going to work and seeing you everyday. Specific thanks to Ingrid Pardoe, Dante Galutira, Jim Houston and Carol Negrjin for assisiting me no matter what the cause - your skills and generosity were greatly appreciated; Annika Haywood and Kathy Hodgkinson for being valuable members of the A team - I have learned so much from you; Sandra Cooke, Lance Doucette, Vanessa French, Nelly Abdelfatah, Nancy Whalen, Yoella Teplitsky and Krista Mahoney for being excellent peers and making learning fun.

**My friends:** for helping me in any way possible, from a fun night out to helping with Abby.

**Supervisory committee:** for your encouraging feedback and support, and for exposing me to other scientific opinions and influences.

**Terry-Lynn Young:** for being an amazing supervisor and mentor. It was you who truly made me the scientist I am today. I thank you for your support, generosity, and faith in

my abilities, as well as for providing me with the opportunities of a life-time. I will treasure this experience and our relationship forever.

**Funding sources:** Genome Canada (Atlantic Medical Genetics and Genomics Initiative), The Canadian Breast Cancer Foundation, and the Memorial University Graduate Studies Dean's Fellowship.

**American Journal of Human Genetics:** for publishing "Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the *TMEM43* gene" in the April 2008, issue 82, volume 4, pages 809-21.

*To Abby Jayne*

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations and Symbols

| | |
|---|---|
| AAPC | Attenuated adenomatous polyposis coli |
| *Abra1* | *Abraxas* |
| *ACTN2* | *Actinin alpha-2* |
| *AFG3L2* | *ATPase family gene 3-like 2* |
| ALS | Amyotrophic lateral sclerosis |
| ARSACS | Autosomal recessive spastic ataxia of Charlevoix-Saguenay |
| ARVC/D | Arrhythmogenic right ventricular cardiomyopathy/dysplasia |
| AT | Ataxia-telangiectasia |
| ATM | Ataxia-telangiectasia mutated protein |
| *ATM* | *Ataxia-telangiectasia mutated gene* |
| BARD1 | Breast cancer associated risk domain 1 |
| BBS | Bardet-Biedl syndrome |
| *BRCA* | *Breast cancer susceptibilty gene* |
| BRCT | BRCA1 C-terminal |
| BRIP1 | BRCA1-interacting protein 1 |
| *CDH23* | *Cadherin 23* |
| cDNA | Complementary deoxyribonucleic acid |
| *CHK2* | *Check point kinase 3* |
| CLS | Coffin-Lowry syndrome |
| cM | Centimorgan |
| CMD1C | Dilated cardiomyopathy 1C |
| CNVs | Copy number variants |
| COX | Cytochrome c oxidase |
| dB | Decibels |
| *DFN* | X-linked deafness loci (old designation) |
| *DFNA* | Autosomal dominant non-syndromic deafness loci |
| *DFNB* | Autosomal recessive non-syndromic deafness loci |
| *DFNX* | X-linked deafness loci (new designation) |
| *DMD* | *Duchenne muscular dystrophy* |
| DMF | Dimethyl-formamide |
| DNA | Deoxyribonucleic acid |
| *DSC2* | *Desmocollin*-2 |
| *DSG2* | *Desmoglein*-2 |
| *DSP* | *Desmoplakin* |

| | |
|---|---|
| EST | Expressed sequence tag |
| ET | Essential tremor |
| FA | Fanconi anemia |
| *FTHL17* | *Ferritin heavy polypeptide like-15* |
| *GJB2* | *Gap junction protein beta 2* gene |
| $H_2O$ | Water |
| HIC | Human Investigation Committee |
| HNPCC1 | Hereditary non-polyposis colorectal cancer syndrome |
| HSAN | Hereditary sensory and autonomic neuropathy |
| Hz | Hertz |
| Indels | Small insertions and deletions |
| *JUP* | *Plakoglobin* |
| LCRs | Low-copy repeats |
| LOD | Logarithm of odds |
| Mb | Megabases |
| MEN1 | Multiple endocrine neoplasia type 1 |
| MLPA | Multiplex ligation-dependent probe amplification |
| *NBS1* | *Nijmegen breakage syndrome 1* |
| NGS | Next generation sequencing |
| OCCR | Ovarian cancer cluster region |
| ORF | Open reading frame |
| *OTOF* | *Otoferlin* |
| *PALB2* | *Partner and Localizer of BRCA2* |
| *PCDH15* | *Protocadherin 15* |
| PCR | Polymerase chain reaction |
| *PKP-2* | *Plakophilin-2* |
| *POU3F4* | *POU-domain class 3 transcription factor 4* |
| *POU4F3* | *POU-domain class 4 transcription factor 3* |
| PPREs | Peroxisome proliferator response elements |
| PRPP | Phosphoribosyl pyrophosphate |
| *PRPS1* | PRPP synthetase 1 |
| *PTEN* | *Phosphatase and tensin homolog* |
| RING | Really Interesting New Gene |
| RPAC | Research Proposals Approval Committee |
| *RYR2* | *Ryanodine receptor 2* |
| SCA | Spinocerebellar ataxia |
| SCD | Sudden cardiac death |
| SNPs | Single nucleotide polymorphisms |

| | |
|---|---|
| SSCP | Single-strand conformation polymorphism |
| ssDNA | Single stranded DNA |
| *TAB3* | *TAK1 binding protein 3* |
| TD | Touchdown |
| TGFβ-3 | *Transforming growth factor β3* |
| *TMEM43* | *Transmembrane protein 43* |
| *TMPRSS3* | *Transmembrane protease serine 3* |
| UTRs | Untranslated regions |
| VT | Ventricular arrhythmias |
| *WFSI* | *Wolframin* |
| WHO | World Health Organization |
| Z | LOD score |
| | |

# List of Appendices

# Chapter 1: General introduction

## 1.1 Aims of this study

The ultimate goal of this thesis was to identify disease genes that segregated in Newfoundland families; three hereditary diseases including arrhythmogenic right ventricular cardiomyopathy/dysplasia (ARVC/D), deafness and breast cancer were studied. This general introduction discusses the modern scope of genetics, familial gene discovery approaches, and the benefits of using founder populations for gene discovery. Each chapter that follows is dedicated to one of the three studied diseases where the clinical and genetic characteristics are described more specifically.

## 1.2 Why make gene discovery efforts?

Identifying disease genes improves disease management at many levels. Immediate impacts include the improvement of disease diagnosis and risk assessment. A simple genetic test can provide a diagnosis of disease, which can be beneficial especially for diseases where early diagnosis is critical and can improve prognosis, such as in the case of cardiomyopathies (1). The risk of disease onset can also be better assessed by identifying mutation carriers, who, once identified, can be closely monitored for early detection, or maybe even prevention in the future (2). Hamilton discusses the different preventative measures, including surveillance and prophylactic surgeries that positive *BRCA* (breast cancer susceptibility gene) mutation carriers, who are at a higher risk of developing breast cancer, have to face (2).

The complete molecular mechanisms that underlie human disease are poorly understood, but new insights can be derived from identifying disease genes (3). Such discoveries help elucidate the cellular pathways responsible for disease and may provide new knowledge for drug therapies (3-5). A recent review about calcium deregulation in amyotrophic lateral sclerosis (ALS) and novel targets for neuroprotective therapies is such an example (6). Nearly a decade has passed since the first draft of the human genome was sequenced (7-9). This ground-breaking research has provided a blueprint for the study of transcriptomics (transcriptional regulation), proteomics (biological function of gene products) and interactomics (the interaction of proteins), all of which are crucial in order to better understand disease pathogenicity (5).

## 1.3 Genes – definition and number

A question that has been posed since the human genome sequence became publically available is "how many genes are in the human genome?" According to the most recent assembly of the human genome (GRCh37 – February 2009) and Genebuild by Ensembl (database version 56.37c – released May 2010) there are 21,701 protein coding genes, 8,483 RNA genes and 12,599 pseudogenes that have been annotated with a total of 148,792 transcripts. The answer to the above question, however, has been suggested to largely depend on how one defines the word "gene" (10). Gerstein et al. published a comprehensive review discussing the evolution of the definition of a gene, which indicated that in the late 1800s and early 1900s the term "gene" was originally defined as a distinct unit of heredity (11). Later, in the 1940s, it was termed as a blueprint for a

protein, but as the years passed (between the 1960s and 1980s) and RNA genes were recognized, it was defined as a transcript that gives rise to a functional product (11). More recently, a gene has been described as a DNA segment that contributes to a particular function or phenotype which, in the absence of known function, can be characterized by sequence, transcription or homology (gene annotation) (11). The genome is however a complex entity. Some genes overlap each other either on the same or different DNA strand; some share regulatory or coding regions (12), while others are completely independent but are within an intron of an overlapping gene (13). Other genes undergo alternate splicing and, for example, have a number of different first exons and promoters (14), and/or have tissue-specific isoforms (14, 15). More uncommon are bicistronic genes, where two genes in *cis* are transcribed together from a single transcription site but are later cleaved and processed as different proteins (16). There is also recent evidence of two different RNA products being cleaved then joined together as one transcript, which is termed trans-splicing (17). Defining and categorizing such transcripts as multiple genes or a single gene is debatable. Furthermore, considering *cis*-acting regulatory elements, should they be included in the boundary of a gene? What if such elements are remote and not even on the same chromosome? Both structure and function need to be considered when generating a general definition of a "gene" (10), and when screening candidate disease genes one has to take all this information into consideration as well.

## 1.4 The mutome – common versus rare disease variants

The mutome is defined as the spectrum of mutations that underlie or are associated with disease (18). Over 100,000 germline mutations that either cause or are associated with a disease have been identified in 3700 different genes. Each year approximately 10,000 new gene mutations and 300 new "inherited disease genes" are being identified (10, 18). Since 2005, when the International Human HapMap Project reported that the human genome is structured into general haplotypes that are divided into blocks and, within those blocks, genetic variants are in linkage disequilibrium (19), association studies became a popular method to study complex diseases (4). Common, low-risk variants that confer only a small risk of disease were aimed to be identified (20, 21). There is a debate on how far these studies will take us in understanding the risks and causes of disease (22-24), and whether multiple common or rare variants actually contribute to complex diseases, such as autism (25, 26). What is certain though, is that in order to have a better understanding of disease it is still very important to identify variants that cause monogenic traits. Furthermore, a single disease variant that is rare, infers a high risk for disease, and segregates within families can be readily identified by traditional approaches.

## 1.5 Gene discovery efforts of monogenic traits – a systematic approach using families

Family studies have been used effectively to identify disease genes over the last 20 years (4). The resulting discoveries have mainly revealed deleterious variants that affect the protein sequence directly and increase disease risk substantially. According to Online

Mendelian Inheritance in Man (OMIM) there are approximately 2800 monogenic (Mendelian) diseases with a known molecular basis (27). However, the molecular basis remains unknown for approximately 1800 described Mendelian diseases and another 2000 diseases with a suspected Mendelian inheritance (27). If large, informative, properly ascertained disease families are available for study then traditional gene discovery approaches can be applied to identify the causal genes for the aforementioned unsolved Mendelian diseases.

### 1.5.1 Systematic approach – Part 1 (ascertainment of families)

There is an ideal, systematic approach to gene discovery when studying families. It first requires the disease of interest to be selected and its clinical features (also known as the disease phenotype) to be well defined, preferably with precise diagnostic criteria. Families with the disease are then ascertained and the family members are clinically assessed by health professionals. Correctly determining the affection status of all family members is crucial for the success of the gene discovery effort. An accurately constructed pedigree will then give insight into the mode of disease inheritance and the appropriate genetic model that should be used to identify the disease gene. The general modes of inheritance that will be discussed in this thesis include autosomal dominant and recessive, requiring one and two mutations (one inherited from each parent), respectively, in a gene located on one of the 22 autosomes, and sex-linked, involving mutations inherited on either the X or the Y chromosome. X-linked inheritance will be discussed in more detail in Chapter 3. Properly diagnosing a disease, assigning the correct affection

status and determining its mode of inheritance is not always an easy task. Complications pertaining to this are discussed below.

### *Variable expression*

The clinical features of many diseases can be expressed variably, which complicates defining the disease phenotype and identifying affected individuals. Essential tremor (ET) [MIM #190300], a common neurological disorder, is such a disease (28). ET is characterized by postural tremor (when a patient voluntarily attempts to maintain a steady posture) that subsequently worsens with movement and affects mainly the upper limbs and, less commonly, the head, voice, face, tongue, trunk and lower limbs. The clinical variability of the phenotype however makes its diagnosis difficult. In fact, there are no validated diagnostic tests for ET and the lack of consensus on clinical diagnostic criteria has resulted in 30-50% of patients previously diagnosed with ET to be misdiagnosed (28). Recognizing this complication, ET is generally diagnosed as either "definite," "probable," or "possible." Despite being a hereditary disorder in 50-70% of patients, studying ET families has not yet led to the identification of a causative gene (29), which can be attributed to misdiagnosis. If an individual from an ET family is diagnosed as "possibly" affected, but is truly unaffected, classifying that individual as affected for the genetic analysis will hinder any gene discovery efforts.

### *Penetrance*

Penetrance of a phenotype, given a particular genotype, is defined as the probability that a person who has the genotype will manifest the phenotype. If an allele is highly (100%) penetrant, as is the c.338T>C mutation in *connexin 43* [MIM #121014], which causes autosomal dominant, congenital, oculodentodigital dysplasia [MIM #164200] (30), the trait will always be apparent in an individual carrying the allele.

Penetrance is said to be reduced or incomplete when some individuals fail to express the trait, even though they carry a mutant allele. Regarding autosomal dominant inherited traits, if a mutation carrier is asymptomatic but has an affected child the mode of inheritance may be mistaken as recessive. There are examples of hereditary breast cancer families with high risk *breast cancer susceptibility gene 1* or *2* (*BRCA1* [MIM #113705] or *BRCA2* [MIM #600185]) mutations where carriers never present the disease phenotype (31, 32). By 70 years of age, breast cancer penetrance for *BRCA1* and *BRCA2* mutation carriers is estimated to range from 14% to 87% (33-36).

Penetrance can also be age-related. The penetrance of highly penetrant *BRCA1* and *BRCA2* alleles are a good example of this as well (31, 32). It has been shown that different *BRCA1* mutations result in different median ages of onset (32). For example, c.185delAG in exon 2 has a median age of onset of 55 years, compared to variants with a higher penetrance, c.4184delTCAA in exon 11 and an exon 13 duplication, where the median ages of onset are 47 and 41, respectively (32).

Low penetrance alleles also exist, which are alleles that have a low associated risk of developing the disease. For example, 12 *Ataxia-telangiectasia mutated* (*ATM*) [MIM #607585] mutations were identified in breast cancer patients after studying 443 familial breast cancer pedigrees. Women with these variants were determined to have an estimated relative risk of 2.37 for developing breast cancer (37).

### 1.5.2 Systematic approach – Part 2 (genetic analysis of families)

#### *Exclusion of known genes or loci*

Before the search for a new gene commences, known disease genes and/or loci are generally excluded as the cause of disease. This is common when studying genetically heterogeneous diseases and will save time, effort and money if the family can be solved using this approach. However, when a disease has many known causatives genes, such as deafness (38), it is sometimes more effective to skip the exclusion process or screen for only the most commonly mutated genes. Autosomal dominant cerebellar ataxias are a clinically and genetically heterogeneous group of neurodegenerative disorders characterized by imbalance, progressive gait and limb ataxia, and dysarthria (39, 40). The first gene, *SCA1* (responsible for spinocerebellar ataxia type 1) was identified in 1993 (41). At present, at least 25 distinct genetic forms of SCA are known, and many are caused by tri-nucleotide repeat expansion mutations within the coding region of the corresponding genes (42). In 2006, *SCA28* was a novel locus identified on chromosome 18p11.22-q11.2 by studying a four generation Italian family. Before the search for the novel locus commenced, genetic analysis excluded the presence of repeat expansions in

the SCA genes associated with those types of mutations, and linkage analysis was performed to exclude the remaining known loci (40).

### *Novel gene discovery approaches in families - Candidate gene and genome scan approach*

Two research approaches have been commonly used to identify monogenic disease genes when studying disease families, the candidate gene approach and the genomic screening approach.

### *(1) Candidate gene approach*

For the candidate gene approach, genes are selected based on their known or predicted biological function and/or on their relation to the disease. For example, hereditary sensory and autonomic neuropathy type 4 (HSAN4) [MIM #256800] is a very rare autosomal recessive disorder that involves both pain insensitivity and autonomic defects, which is described as a congenital insensitivity to pain with anhidrosis (inability to perspire) (43, 44). In 1996 Indo et al. screened 3 candidate genes (*NTRK1*, *NGF* and *p75* neurotrophin receptor) in three unrelated HSAN4 patients, all of whom had consanguineous parents, based on similar phenotypes that were reported in the knock-out mouse models for these genes (45). No mutations in *NGF* or *p75* neurotrophin receptor were identified, however three different *NTRK1* [MIM #191315] variants (a deletion, splice mutation, and missense mutation) were detected which suggested *NTRK1* was the pathogenic HSAN4 gene (45). Since that time over 40 different HSAN4 *NTRK1*

mutations have been reported in several different ethnic groups (Human Gene Mutation Database: http://www.hgmd.cf.ac.uk/ac/gene.php?gene=NTRK1).

### *(2a) Genome-wide or chromosome scan approach*

When families that have gene discovery potential are identified a genome-wide scan or chromosome scan (for sex-linked disorders) has been the traditional approach used. Well-ascertained, informative families are required for this approach (see section 1.3.1). Genetic markers spanning each chromosome are genotyped in available individuals. A total of 300 to 400 evenly spaced microsatellite markers that cover the whole genome were most commonly typed. Microsatellites are simple sequence repeats of mono-, di-, tri- or tetra-nucleotides that represent 2% of the genome. However, more recently, thousands of single nucleotide polymorphisms (SNPs) are being typed on oligonucleotide chips, which are devices that have thousands of short hybridized DNA sequences, each containing a specific SNP that enables the discrimination between alternative bases at the SNP site. This technology allows many SNPs to be analyzed in parallel (46, 47).

The objective of the genotyping scan is to detect genetic linkage between the disease mutation that is segregating in a family and a genotyped marker. Genetic linkage is a term that describes the tendency of alleles to be inherited together. Alleles that are on the same chromosome and physically close to one another tend to stay together during meiosis, and are thus genetically linked. These linked alleles segregate together in blocks known as haplotypes. How closely linked two alleles are, is measured by genetic

recombination or crossover. Chromosomes undergo crossovers during meiosis, which is the process of exchanging DNA (deoxyribonucleic acid), usually, between matching regions on homologous chromosomes. A recombination fraction measures the likeliness of a crossover between two markers, which also correlates with genetic distance. The closer two markers are on the same chromosome, the less likely a crossover event will occur between them, thus the markers are said to be linked (48). If recombinants are never seen then the recombination fraction is 0 and the alleles are tightly linked; if recombinants are seen, the fraction will have a value greater than 0 but less than 0.5 (the recombination fraction is 0.5 when there is no linkage) (48). Therefore, a pathogenic variant that segregates in a family has a distinct haplotype of closely linked markers on the ancestral disease chromosome. At least one of the markers genotyped in the scan is hoped to be linked to the disease variant.

In order to interpret the data from the genotyping scan a parametric or nonparametric linkage analysis can be used. The parametric linkage analysis is a model-based approach, which includes clearly defining the affection status and pattern of inheritance, predicting the allele frequency of the disease variant (is the disease common or rare?), and considering the penetrance of the disease variant (since that is crucial in determining the probability that a clinically unaffected individual is truly unaffected, a true non-carrier, or a non-penetrant carrier). The frequency of phenocopies (same phenotype but different genotype) is also important to note, since they are common in some diseases, for example Huntington's disease (49). The nonparametric linkage analysis is a model-free approach

which is appropriate for more complex traits where the exact genetic model is hard to define. The basic concept behind a nonparametric linkage analysis is to identify markers where the same identical-by-decent (IBD) allele is shared between affected individuals - between all sib-pairs in a pedigree, for example (50).

Parametric linkage has been a very successful and common approach for mapping Mendelian disease loci. Two likelihoods are considered, the first being the likelihood of observing a phenotype in a family due to genetic linkage between loci, where the recombination fraction between the loci is some value of theta ($\theta$). The second is the likelihood of observing a phenotype in a family assuming there is no linkage between loci thus the recombination fraction is 0.5. The ratio of these two likelihoods gives the odds of linkage compared to no linkage. The logarithm of the odds ratio is the LOD score (Z). LOD scores are a function of $\theta$, thus they are calculated for a range of $\theta$ values for each marker. The $\theta$ value with the highest LOD score is the most likely recombination fraction. A two-point linkage analysis is the simplest calculation and involves independently calculating LOD scores for each typed marker and the disease locus. The threshold for accepting linkage is a LOD score of 3, which corresponds to an odds ratio of 1000:1 for linkage ($\log(1000) = 3$). Similarly, linkage can be rejected when the LOD score is less than or equal to –2. Regarding the SCA example that was mentioned above, after all known loci were excluded a genome wide scan was carried out on the Italian family and linkage was identified with markers on chromosome 18. A maximum two-point LOD score of 4.20 ($\theta=0$) was observed for marker *D18S53* (40). A multi-point

linkage analysis can also be carried out which takes into consideration the framework of genotyped markers and analyzes more than two loci simultaneously. This approach is usually performed after a two-point linkage analysis; markers are grouped into linkage groups and multiple markers in a particular chromosomal region are evaluated together for linkage. Multi-point linkage analysis can increase power since the analysis uses haplotype information from several markers.

### *(2b) Fine mapping of the disease-linked region*

Once linkage has been detected, markers in that chromosomal region are typed to define the disease locus and its boundaries. The minimal region that is shared by all affected individuals is characterized by generating the disease-associated haplotype and detecting key recombinations. Microsatellites have traditionally been used for fine mapping but with the availability of the new SNP chips, fine mapping data is obtained from the original analysis because the SNP chips are so dense (~100,000 SNPs). Analyzing as many individuals as possible can help narrow the shared region since increasing the number of meioses increases the chance of detecting key genetic recombinations.

Fine mapping of the *SCA28* disease-linked region around *D18S53*, in the aforementioned Italian family, was carried out using eleven polymorphic markers (40). Multipoint LOD scores were calculated and reached a maximal value of 4.77 at marker *D18S453*. A common disease haplotype was shared by all the affected individuals and key recombinants refined the disease region to 7.9 megabases (Mb) between markers

*D18S1418* and *D18S1104*, which ultimately defined the new SCA locus, *SCA28* [MIM #610246] (40).

### *(2c) Assessing the disease-linked locus*

Due to the extensive effort of the Human Genome Project (7, 8) one can now obtain the DNA sequence of the disease region by downloading it from the world-wide-web using databases such as the University of California Santa Cruz (UCSC) Genome browser (51, 52). Before this luxury existed a contig of genomic clones had to be established across the candidate region, which is a group of genomic clones that overlap with each other and together represent the original sequence of the chromosome segment. This was the technique used by the Human Genome Project to establish the framework for sequencing the human genome (7, 8).

Genes within the disease-linked region are called positional candidates and can be identified using genome browser databases as well (52). This is possible because another goal of the Human Genome Project was to construct the human gene map (7, 8), thus the disease-linked region and all annotated positional candidates can be graphically displayed on these sites making it extremely easy to identify candidate genes. The 7.9 Mb interval segregating with SCA in the Italian family contained over 30 genes (Build 36 version 2) (40). It is important to note that the list of human genes and isoforms is incomplete, so one should not solely rely on these databases.

### *(2d) Positional candidate gene screening*

Sanger sequencing, which involves dideoxy nucleotide chain-terminating inhibitors of DNA polymerase to determine the nucleotide sequence, is a popular method for screening positional candidate genes. This technique is very good at detecting point mutations, and small insertions and deletions (indels) (53, 54). When sequencing positional candidates, it is common to prioritize gene screening based on the expression pattern, function and homology of the corresponding proteins. In an effort to identify the *SCA28* mutant gene, Di Bella et al. recently screened 12 genes within the 7.9-megabase critical region based on nervous system expression (55). They identified *AFG3L2* heterozygous missense mutations that were likely disease-causing in five unrelated SCA families (including the Italian family that was used to identify the locus) (55).

It is important to note that when using Sanger sequencing, the coding exons and intron/exon boundaries are the regions most commonly screened. As such, often variants in introns, untranslated regions (UTRs) and promoters will go undetected. It is also important to mention here that due to the demand for fast, high volume, inexpensive and reliable sequencing methods, new sequencing technologies now exist, which have the ability to massively parallel sequence chromosomal regions that have been capture-targeted (from whole genome to disease loci) (56). These novel technologies will help detect non-coding sequencing variants in the future.

More recently large genomic deletions and duplications have been recognized as disease causing. The mechanism that causes many of these rearrangements is nonallelic homologous recombination or unequal crossovers of chromosomes between low-copy repeats (LCRs) (also known as segmental duplications) (57). A single exon, a series of exons, a whole gene, or even large chunks of DNA ranging from tens of thousands to millions of base pairs can be deleted or duplicated (these variants are known as copy number variants or CNVs), all of which have a Mendelian pattern of inheritance. One relatively new technique that enables the identification of exonic duplications and deletions is multiplex ligation-dependent probe amplification (MLPA) (58). The detection of such rearrangements has solved several families that were previously linked to known loci but were found to be mutation-negative by Sanger sequencing. Large deletions/duplications have been found in genes that cause hereditary colon (59, 60) and breast cancer (61, 62). In fact, at least 81 different *BRCA1* genomic rearrangements and 17 *BRCA2* genomic rearrangements have been reported to date (63), and studies carried out mainly in Caucasian populations have shown that 4–28% of *BRCA* mutations may be genomic rearrangements (64). There have also been at least 20 different human diseases described that are caused by deletions of *cis*-acting regulatory elements (65). Half of the mutations in the *POU-domain class 3 transcription factor 4* (*POU3F4*) [MIM #300039] gene that cause deafness are deletions of a regulatory element (66). In most cases, deletions of regulatory elements are not as common as deletions in the open reading frame but that could be due to screening biases (65, 67), hence the proportion of

pathogenic rearrangements that play a role in genetic disorders is likely underestimated (67).

### *(2e) Mutation validation of variants detected in positional candidate genes*

Generally, a small number of affected family members and unaffected controls are screened for mutation detection. After the mutation screen, variants that are exclusively found in affected individuals are highlighted. The next step in the exclusion process is determining which of these variants segregate on the disease haplotype, followed by the screening of the segregating variants in population controls to determine allele frequency. Rare variants are more likely to be pathogenic, and are thus further analyzed. Performing additional bioinformatic analyses can help to predict the effect a variant has on RNA splicing and protein function. If a variant is in a coding region it is more likely to be pathogenic if an amino acid is highly conserved in orthologs, thus conservation analysis is also useful (68-70). Another helpful clue is if the variant is in a gene whose protein product has a similar function or is in the same signaling pathway as the encoded proteins of other causal genes whose mutations cause the same disease. Ultimately, the most ideal validation methods include identifying additional families that have a mutation(s) in the same gene, and performing functional studies in order to show pathogenicity.

*AFG3L2*, which encodes a mitochondrial metalloprotease (ATPase family gene 3-like 2) was presumed to be the *SCA28* mutant gene after mutation screening and detecting different heterozygous *SCA28* mutations in five different *SCA28* families. For validation,

the missense mutations were determined to be located in highly conservative amino acids, and functional studies demonstrated that expressing the mutant human AFG3L2 protein in yeast resulted in a defective activity of cytochrome c oxidase (COX) and impairment of cell respiration (55). Di Bella et al. also noted that AFG3L2 is highly and predominantly expressed in cerebellar Purkinje cells, which is consistent with the pathological phenotype (55).

## 1.6 Gene discovery in isolated populations

Gene discovery efforts have been very successful in isolated populations, which are groups of people who have been restricted from intermarriage with other groups due to differences in culture, religion, language, geography or other factors (71). Generally these populations descend from a limited number of common ancestors; this can also define the group as a founder population. These populations are more homogeneous than out bred populations thus the majority of individuals with a disease will often carry the same ancestral mutation, which can benefit gene discovery efforts (71).

### 1.6.1 Example of a founder population and global medical benefits

There are several examples of founder populations around the world that have been used to study genetic diseases. The French-Canadian population is one prominent example. It is a founder population that exceeds 7 million people, 6 million of whom are of French decent (72, 73). The Canadian province of Quebec was founded by French immigrants who settled along the St. Lawrence River in the early 1600s. In 1608, immigrants first

settled in what is today Quebec City. Immigration continued along the St. Lawrence River until 1759, when English settlers took over the land and halted French immigration. A total of 8500 French immigrants had settled permanently; however, by the late 1600s natural expansion had exceeded immigration (72, 74), such that by 1759, 76,000 French Canadians lived in settlements along the river (73). Both language and religion restricted marriages with the newly arrived English speaking Protestants (French Canadians are almost exclusively Roman Catholic). In fact, in order to preserve their culture, French-Canadians were encouraged to have large families, known in Quebec as "the revenge of the cradles." This led to several founder effects that were specific to different regions of the province since most settlements along the river were geographically isolated (72). Despite inter-regional mixing, urbanization and decreased birth rates in the 20th century, the historical founder effects remain a key component of medical genetics in Quebec (72, 73).

Many genetic diseases have been recognized in the Quebec population (73), for example autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) [MIM #270550]. ARSACS was reported initially in the province (75), and the causative gene was identified by studying the Quebec population (76, 77). When the genetics of ARSACS was first studied, 300 patients were identified, all of whom originated from the Charlevoix-Saguenay region of northeastern Quebec. A genome-wide scan was performed on 12 families and excess shared homozygosity was observed at 13q11. Additional analysis on 19 families suggested two different ARSACS haplotypes existed

in this region (77), and as a result two French Canadian founder mutations in the *SACS* gene [MIM #604490], c.6594delT and c.5254C>T, have now been identified (76). Since then, mutations in *SACS* have been recognized as causing the same condition in other populations around the world. *SACS* mutations have been found in Tunisian, Italian, Spanish, Japanese, Turkish and Dutch patients (78-85). This example demonstrates how Quebec gene discovery efforts permit the molecular diagnosis for patients outside the founder population, and illustrates the global benefits of studying isolated populations.

### 1.6.2 Newfoundland as a genetic isolate

Newfoundland, part of the Canadian province Newfoundland and Labrador, is an island off the country's east coast. It has many distinct genetic isolates, particularly along its coastline in out-port communities that were first established as fishing communities (86). Newfoundland is recognized as being first discovered in 1497 by an Italian navigator, John Cabot, and seasonal immigrants first settled in the early 1600s. Permanent immigration did not occur, however, until the mid 1700s, which mainly included English Protestants from south-west England and Irish Catholics from south-east Ireland (86). There were about 20,000 initial settlers and the population grew through natural expansion to 200,000 by the late 1800s. Today there are approximately 500,000 residents, 98% of whom are of English or Irish descent, and 60% of whom live in communities of less than 2500 residents (87-89). Geographical distance between communities and religious segregation are the two main factors that have contributed to the development of several Newfoundland isolates (87).

The founder effect has been reported for many diseases in the Newfoundland population. One of the first was in 1998, when affected individuals of four large families with familial multiple endocrine neoplasia type 1 (MEN1) [MIM #131100] from the Burin peninsula/Fortune Bay area were determined to share a founder nonsense mutation in the *MEN1* gene (90). In 1999, five AAPC (attenuated adenomatous polyposis coli) [MIM #175100] families were reported to have the same ancestral *APC* [MIM #611731] splice mutation that resulted in the omission of exon 4 (91). There have also been 12 large Newfoundland HNPCC (hereditary non-polyposis colorectal cancer syndrome) [MIM #120435] families reported to carry the same *MSH2* [MIM #609309] founder mutation (92, 93), three large Newfoundland families with a novel and clinically variable spastic ataxia and supranuclear gaze palsy [MIM #108600] that carry the same disease haplotype on chromosome 12p13 (94, 95), and four families with diffuse gastric cancer [MIM #137215] from the south-east coast of Newfoundland that carry the same ancestral *CDH1* [MIM #192090] gene mutation (96). A *factor VIII* [MIM #306700] variant that causes a mild form of hemophilia A that is highly prevalent in the Newfoundland population has also been determined to be a founder mutation (97).

The unique population of Newfoundland has aided in multiple gene discoveries as well. Newfoundland families have been used to identify genes for hereditary hearing loss [MIM #606201] (98), HNPCC (99, 100), hereditary sensory and autonomic neuropathy type 2 [MIM #605232] (15, 101), and interleukin 1 receptor antagonist deficiency (102). The Newfoundland population, perhaps, most greatly influenced gene discovery efforts

for Bardet-Biedl syndrome (BBS) [MIM #209900], for which Newfoundland families aided in the discovery of four of the 14 known *BBS* loci or genes. Five years after the discovery of the first *BBS* locus [MIM #209901] (103), the disease interval was refined by studying a group of *BBS1*-linked Newfoundland families that shared a rare ancestral disease haplotype (104), which in turn aided in the 2002 gene discovery by Mykytyn et al. (105). Similarly, by studying a Newfoundland BBS kindred in 1998 (106), the *BBS3* locus [MIM # 209900] was refined four years after its initial discovery (107). The causative gene was later discovered in 2004 (108). Furthermore, homozygosity mapping of another large consanguineous Newfoundland family identified the *BBS5* locus [MIM #603650] on chromosome 2q31 (109) and led to the identification of the causative gene, *DKFZp762I194* in 2004 (110). Finally, a genome scan that was performed on several BBS Newfoundland families that had been excluded from *BBS1-5* identified linkage to a marker on chromosome 20, *D20S189*, and ultimately *MKKS* [MIM #604896] as the causative *BBS6* gene (111). BBS, despite being a great example of how the Newfoundland population is a genetic gold mine, highlights the significant genetic diversity of a single disease within the island's population.

The main objective of this thesis was to take advantage of Newfoundland's unique population in order to identify novel disease causing genes involved in ARVC, deafness and breast cancer, and to gain new insights on disease management and pathogenicity.

# Co-authorship Statement

### (1) Design and identification of research proposal

Nancy Merner designed the research proposal of all the genetic research sections with the guidance of Dr. Terry-Lynn Young.

### (2) Practical aspects of the research

**Chapter 2:** Fine mapping to reduce the disease-associated region was the work of Vanessa French (M.Sc. candidate), Dante Galutira (research assistant) and Ingrid Pardoe (technical lab manager). Nancy Merner was the team leader of the positional gene screening phase and was responsible for developing and ordering primer sets for all exons of each candidate gene, performing trial PCRs to determine optimal amplification conditions for each primer set, designing the mutation screening panel and screening the panel for all candidate genes, analyzing sequence traces for variants, designing a database to record all detected variants, and training fellow team members (Annika Haywood, post-doctoral fellow, and Vanessa French) to perform the above duties as well.

**Chapter 3** Nancy Merner was responsible for all of the research in this chapter. Jim Houston (research assistant) provided technical help by screening four positional candidate genes.

**Chapter 4:** Phase one was the postdoctoral work of Dr. Terry-Lynn Young under the supervision of Dr. Mary Claire King at the University of Washington. Phase two was carried out solely by Nancy Merner.

**(3) Data analysis**

**Chapter 2:** Once all candidate genes were screened and all variants recorded, it was solely Nancy Merner's responsibility to determine which variants were detected only in affected individuals on the screening panel and not in any controls, carry out segregation analysis to determine which of those variants truly segregated on the affected haplotype, determine the allele frequencies of all segregating variants, and screen all 295 subjects born at *a priori* 50% risk of ARVC for the rare variants. Once the mutation was identified, both Nancy Merner and Annika Haywood performed bioinformatic analysis to predict the mutation effect. Annika Haywood, Fahad Chowdury (undergraduate summer student) and Dante Galutira predicted potential protein localization, function and structure. Kathy Hodgkinson (Ph.D. candidate) was responsible for the clinical assessment.

**Chapter 3:** Nancy Merner performed all the analysis, with the exception that Jim Houston performed the initial analysis on the four genes that he screened.

**Chapter 4:** Nancy Merner was responsible for the analysis of the *BRCA1*, *BRCA2* and *CHK2* targeted mutation screen, protein truncation test and muliplex ligation dependent probe amplification on the newly recruited probands.

**(4) Manuscript preparation**

**Chapter 2:** This work was published in the American Journal of Human Genetics as "Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the *TMEM43* gene" in the April

2008, issue 82, volume 4, pages 809-21. Chapter 2 of this thesis is a more elaborate version of the *TMEM43* ARVC gene discovery and was authored by Nancy Merner.

**Chapter 3:** Authored by Nancy Merner.

**Chapter 4:** Authored by Nancy Merner.

# Chapter 2: Arrhythmogenic right ventricular cardiomyopathy type 5 (*ARVD5*) is a lethal arrhythmic disorder caused by an amino acid substitution in the *TMEM43* gene

## 2.1 Introduction

### 2.1.1 What is ARVC?

ARVC is one of four phenotypic groups of hereditary cardiomyopathies, the other three being hypertrophic cardiomyopathy, dilated cardiomyopathy and restrictive cardiomyopathy; a cardiomyopathy is a heart muscle disorder that is caused by structural and functional abnormalities (112). ARVC is characterized by progressive fibro-fatty replacement of the right ventricular myocardium. As a result, individuals with ARVC are at risk for life-threatening ventricular arrhythmias (VT) (a resting heart rate faster than 100 beats/minute), for which classic symptoms include palpitations, presyncope (light-headedness), syncope (a temporary loss of consciousness), and dizziness. VT can ultimately result in sudden cardiac death (SCD) which makes ARVC one of the major genetic causes of SCD in young people.

The first systematic description of ARVC was written in 1982 (113), however, a 1977 report by Fontaine is generally considered as well for the recognition of ARVC (114). There are even earlier reports of potential ARVC cases where patients had fatty infiltration of the right ventricle (115, 116). Through the years, ARVC has undergone various name changes. In Fontaine's report, the disease was recognized as a pre-excitation syndrome (114). It was later described as a dysplastic disorder (abnormal

growth/development of cells) (117), which then evolved to include an arrhythmic factor and resulted in a name change to arrhythmogenic right ventricular dysplasia (ARVD) (118). It is currently recognized as a cardiomyopathy, thus known as ARVC (119).

### 2.1.2 Natural history of ARVC

The natural history of a disease is defined as the clinical presentation from birth to death. The natural history of ARVC can generally be divided into four phases (120). The first phase, the concealed phase, is normally asymptomatic but there can be subtle right ventricle structural changes with or without mild arrhythmias. Sudden death can occur in this stage and it can be the first manifestation of the disease. The second phase is known as the overt (obvious) arrhythmic phase. Symptomatic right ventricle arrhythmias can occur, possibly leading to sudden cardiac death. There are also obvious right ventricle functional and structural abnormalities. The predominant symptoms in this stage are ventricular arrhythmias – including palpitations and sustained ventricular tachycardia. The third phase, the global right ventricle dysfunctional phase, generally results in the progression and extension of muscle disease with relatively preserved left ventricular function. Impaired contractions and isolated right heart failure can occur in this phase. The final phase involves bi-ventricular pump failure and pronounced left ventricular failure as well. At this stage ARVC mimics bi-ventricular dilated cardiomyopathy (DCM) leading to congestive heart failure and related complications (121).

### 2.1.3 Diagnosis of ARVC

Initial reports on ARVC concluded that it was a rare, severe disorder and all patients had a dilated right ventricle and sustained VT (113), thus milder cases or more severe cases that caused death before 'presenting' the disease were overlooked or misdiagnosed. Since then, additional features have been recognized, for example the importance of histopathology for diagnosis in order to detect the fibro-fatty replacement of the myocardium, and the frequent involvement of both ventricles (122-125). In 1994, a set of International Task Force diagnostic criteria for ARVC was created that included six different classes with major and minor criteria (126). For diagnosis, an individual has to fulfill either two major, one major plus two minor, or four minor criteria (126). These criteria have since been modified (127) in order to recognize as affected, for example, less severely affected family members of known ARVC families, who did not fulfill the Task Force criteria (127). Current limitations of the Task Force criteria still include not recognizing deceased relatives as affected, even when an autopsy suggests ARVC, which is detrimental when an extended family history is available and SCD is a primary presenting feature. Also, the level of clinical testing needed for a Task Force clinical diagnosis is only offered at tertiary centers (hospitals that offer a full complement of services), thus some cases are unrecognized and perhaps misdiagnosed only because the proper testing is not available.

### 2.1.4 How common is ARVC?

ARVC has been reported worldwide (128-131), however, its exact prevalence and incidence have not been determined (120). Italy has been the population of focus for many ARVC studies (132-135), and after conducting postmortem studies on 60 individuals under 35 years of age who had died suddenly in north-eastern Italy from 1979 to 1986, ARVC was estimated to account for 20% of deaths in young Italian people (123). A group in the USA studied sudden unexpected nontraumatic deaths between 1960 to 1989 in a young adult population (aged 20 to 40 years old) from Minnesota, and suggested that 17% of SCDs in those young individuals were a result of ARVC (136). Another study in France performed a retrospective analysis of 1700 forensic autopsies following unexpected sudden cardiac death that occurred between 1981 and 1997, and revealed a group of 50 cases where death occurred during surgery/anesthesia administration. Patients were young with no history of cardiac disease, however the authors suggested that 18 of the 50 cases (36%) were ARVC-related (137).

### 2.1.5 The cause of ARVC

ARVC is recognized as a heritable disorder with, typically, an autosomal dominant pattern of inheritance. It is genetically heterogeneous with 12 known autosomal dominant loci (*ARVD1-12*). One recessive syndromic form of ARVC (Naxos disease) [MIM #601214] exists (Table 2.1). In fact, the gene for Naxos disease was the first ARVC gene to be cloned.

### 2.1.6 Autosomal recessive syndromic ARVC

Mutations in *plakoglobin* (gamma catenin) (*JUP*) [MIM #173325] cause a recessive syndromic form of ARVC called Naxos disease (138). This syndrome involves a triad of wooly hair, a skin disorder known as palmoplantar keratoderma, and ARVC, and was first clinically reported in 1986 (139). Naxos disease was mapped to chromosome 17, in 1998, by studying nine Greek families with 21 affected individuals (140), and in 2000, McKoy et al. identified a homozygous 2 bp deletion in the *C*-terminus of plakoglobin. Plakoglobin functions in cell assembly in both the desmosome and adherens but it also has a role in gene regulation in the WNT (wingless type) signaling pathway. *Plakoglobin* mutations, therefore, possibly disturb proper cell-cell adhesion and/or a critical signaling pathway (121).

### 2.1.7 Autosomal dominant ARVC

#### (i) *ARVD1* [MIM #107970]

The *ARVD1* locus, located at 14q23-q24, was mapped in 1994 by studying two large Italian families with ARVC (133). A maximum two-point LOD score of 6.04 ($\theta$=0) was obtained at marker *D14S42* after a genome-wide scan (133). In 2003, two additional families (one Italian and the other German) were determined through linkage exclusion to have distinct haplotypes that segregated with ARVC at the *ARVD1* locus. Maximum LOD scores (at $\theta$=0) of 4.41 with marker *D14S254* and 4.06 with marker *D14S983* were obtained for the Italian family. However, for the German family, a maximum LOD score of 1.51 ($\theta$=0) was obtained for marker *D14S59*, which the author suggested was due to

the small family size (141). In 2005, mutation screening of the regulatory 5′ UTR of *TGFβ-3* in one of the Italian families described above, identified a variant (c.1-36G>A) that segregated with the disease (141, 142). Screening 30 additional ARVC probands for *TGFβ*-3 variants identified another variant (c.1723C>T) in a single proband. Neither nucleotide change was found in 300 control subjects (142). *TGFβ*-3 mutations have not been reported for the other *ARVD1*-linked families.

### (ii) *ARVD2* [MIM #600996]

The second ARVD locus was mapped to 1q42-q43 by studying a north-eastern Italian family (134). A LOD score of 4.02 ($\theta$=0), assuming a 95% penetrance, was obtained using a CA-repeat polymorphism within the gene *actinin alpha-2* (*ACTN2*) [MIM #102573]. The family also showed significantly positive LOD scores for markers flanking the *ACTN2* gene (134). In 2001, the locus was refined and the causative gene identified after studying four Italian families, including the family used to originally map *ARVD2*, which had been clinically updated in a later report (143, 144). After screening *ryanodine receptor 2* (*RYR2*) [MIM #180902], which encodes a calcium channel, four missense variants p.R176Q, p.L433P, p.N2386I and p.T2504M were identified, all of which were located in highly conserved regions of the protein and critical for the regulation of the channel.

### (iii) *ARVD3* [MIM #602086]

*ARVD3* was identified by studying three unrelated families with ARVC from Italy, Slovenia, and Belgium (132). ARVC in all three families linked to the chromosome region 14q12-q22 with a cumulative two-point LOD score of 3.26 for *D14S252* ($\theta$=0) and a multipoint maximal cumulative LOD score of 4.7 between loci *D14S252* and *D14S257* (132). This locus was provisionally named *ARVD2* in the original manuscript however it has since been called *ARVD3*. The manuscript describing the discovery of the current *ARVD2* locus (134) had been accepted but not published in the journal Human Molecular Genetics when the *ARVD3* discovery manuscript was first submitted to Genomics (132). The causative gene remains unknown.

### (iv) *ARVD4* [MIM #602087]

The discovery of *ARVD4* involved a linkage analysis on three families (135). Two of those families were Italian, and one was American with European ancestry, and was previously described in 1993 (145). All three ARVC families were characterized by localized involvement of the left ventricle, and were excluded from previously linked loci. Therefore, Rampazzo et al. suspected further genetic heterogeneity and performed a two-point linkage analysis to identify the fourth ARVC locus (135). The disease appeared to be transmitted with three polymorphic markers (*D2S152*, *D2S103*, and *D2S389*) on 2q32.1-q32.3. A maximum LOD score of 3.46 ($\theta$=0) for the marker *D2S152* was obtained (135). The causative gene remains to be identified.

### (v) *ARVD5* [MIM #604400]

The *ARVD5* locus was mapped in an extended seven generation family from the genetically isolated population of Newfoundland (146). This family was first identified in the 1980s (147). There were 200 individuals in this large family, including 10 living affected individuals. There were also 17 individuals who died suddenly, four of whom had an autopsy and all had fat-fibro replacement. A two-point linkage analysis was performed, assuming an autosomal dominant inheritance and a penetrance of 20%, 60%, 80% and 95% for individuals under the age of 15, between 15 and 35 years of age, between 35 and 55 years of age, and over 55 years of age, respectively. Only one locus with a LOD score over 1.5 was identified. Marker *D3S3613* on 3p25 had a peak 2-point LOD score of 6.91 with zero recombination. Haplotype analysis identified a shared region of 9.3 cM between markers *D3S3610* and *D3S3659*. The *ARVD5* gene identification will be addressed in this chapter.

### (vi) *ARVD6* [MIM #604401]

The *ARVD6* locus was identified in a large Caucasian North American family with a highly penetrant, early-onset ARVC (148). The first five known ARVD loci were initially excluded through linkage studies, thus a genome scan identified a novel locus on chromosome 10p14-p12. A maximum 2-point LOD score of 3.92 ($\theta$=0) was obtained with marker *D10S1664*. A 10.6 cM shared haplotype was generated between markers *D10S547* and *D10S1653* (148). A second family with *ARVD6* was identified in 2006 and was used to narrow the critical region to 2.9 Mb (149). This family was of South African

descent. The exonic regions of two positional candidate genes that were involved in cell-cell adhesion (*ITG8* and *FRMD4A*) were screened but no causative mutations were detected (149). Thus, the ARVD6 gene has yet to be identified.

### (vii) *ARVD7* [MIM #609160]

*ARVD7* is also known as myofibrillar myopathy (MFM) with ARVC. This locus was identified in a Swedish family whose affected individuals suffered from a myopathy (morphological changes in the skeletal muscle) and ARVC (150). Linkage analysis showed a maximum 2-point LOD score of 2.76 ($\theta$=0) with marker *D10S1752*, and a multi-point peak LOD score of 3.06 between markers *D10S605* and *D10S215*. This disease interval is located at 10q22.3 (150). After re-examination in 2008, the critical interval was refined to 4.37 Mb between *D10S1645* and *D10S1786* (151). Screening 17 positional candidate genes, including *ZASP* [MIM #605906], which was previously associated with myofibrillar myopathy (152), revealed no pathogenic variants (151). Interestingly, also linked to this critical region is dilated cardiomyopathy 1C (CMD1C) whose causative gene is *LDB3* [MIM #605906] (153). Kuhl et al. suggested that ARVD7 and CMD1C may be allelic disorders (151), however the causal gene for ARVD7 remains to be identifed.

### (viii) *ARVD8* [MIM #607450]

In 2002, Rampazzo et al. identified an Italian family that was unlinked to the other known ARVC loci (154). After performing a genome-wide scan the disease appeared to be

linked to 6p24. Several markers in that region showed positive linkage and the maximum 2-point LOD score was 4.32 ($\theta$=0) for marker *D6S309*. All affected individuals shared a common haplotype between markers *D6S1574* and *D6S1721* that encompassed approximately 6 Mb (154). Within the *ARVD8* locus was *desmoplakin* (*DSP*) [MIM #125647], a desmosomal gene, for which *C*-terminal mutations had been previously reported to cause a recessively inherited triad syndrome involving dilated left ventricular cardiomyopathy, wooly hair, and a skin condition known as generalized striate keratoderma (154, 155). A missense mutation (p.S299R) in exon 7 was identified that segregated with ARVC in the Italian family (154); it is in the *N*-terminal and it modifies the putative protein kinase C (PKC) phosphorylation site that interacts with plakoglobin.

### (ix) *ARVD9* [MIM #609040]

*Plakophilin-2* (*PKP-2*) [MIM #602861], another desmosomal gene, is the causal gene in *ARVD9* (156). Gerull et al. selected *PKP-2* as the candidate gene for ARVC because Pkp2-null mice died at embryonic day 10.75 due to a defect in cardiac morphogenesis (157). After screening 120 western Europe probands with ARVC, 25 different mutations were identified solving 32 cases (156). Another study has reported 14 *PKP-2* mutations in 24 of 56 Dutch probands, 11 of which were novel (158). The mutations found in both studies mainly resulted in truncated or aberrant proteins as a result of deletion/insertion, nonsense or splice site mutations. Many of the missense mutations changed highly conserved amino acids to amino acids of different charges (156, 158). *PKP-2* mutations are thought to cause of up to 40-70% of familial ARVC cases (158, 159).

### (x) *ARVD10* [MIM #610193] and *ARVD11* [MIM #610476]

After *PKP-2* was identified as the *ARVD9* gene in 2004, it was suggested that ARVC may be a disease of the desmosomes. *PKP-2* was the third known desmosomal gene associated with the disease, along with *DSP* (*ARVD8*) and *JUP* (Naxos disease) (156). Several groups took a candidate gene approach and screened *desmoglein*-2 (*DSG2*) [MIM #125671] and *desmocollin*-2 (*DSC2*) [MIM #125645], two desmosomal cadherin proteins. Mutations in both genes are now associated with *ARVD10* and *ARVD11* respectively (160).

The *DSG2* gene was first analyzed in 54 Italian ARVC probands who screened negative for mutations in the *TGFβ3*, *DSP* and *PKP-2* genes. *DSG2* mutations were detected in eight probands, including five missense mutations, two insertion-deletions, one nonsense mutation, and one splice site mutation. Seven probands were heterozygous and one proband was compound heterozygous (161). Neither mutation was detected in 560 control chromosomes. Several other studies have found *DSG2* mutations in ARVC probands as well (162, 163).

Two manuscripts suggesting that mutations in *DSC2* cause ARVC were accepted for publication on the same day (September 6, 2006). One study screened 77 unrelated ARVC probands and identified two different *DSC2* heterozygous mutations, a deletion and an insertion, in four probands from four unrelated families. Both mutations resulted in frameshifts and premature truncation of the desmocollin-2 protein (164). The second

study screened 88 unrelated patients with ARVC for *DSC2* mutations and identified a heterozygous splice acceptor site mutation in intron 5. This resulted in the use of a cryptic splice acceptor site and created a downstream premature termination codon. This mutation was not detected in 500 control chromosomes (165).

### (xi) *ARVD12* [MIM #611528]

Mutations in plakoglobin (*JUP*) have been, more recently, determined to also cause a dominantly inherited non-syndromic form of ARVC, *ARVD12* (166). A novel dominant mutation in *plakoglobin* was identified in a German family, which is located in the *N*-terminus of the protein and inserts an extra serine residue after serine 39 (166). This *N*-terminal mutation is not associated with other abnormalities like the *C*-terminal mutation resulting in the syndrome Naxos disease. This is similar to mutation effects in desmoplakin, where mutations in the *N*-terminus are dominantly inherited and do not affect the skin.

## 2.1.8 ARVC – a disease of the desmosomes

There are 12 known autosomal dominant *ARVD* loci and seven known genes, five of which code for desmosomal proteins. Thus, ARVC has been considered a disease of the desmosomes, which indicates the importance of cell adhesion molecules in ARVC pathogenesis (167). Desmosomes as well as gap junctions and adherens junctions are three types of cell-cell connections called intercalated discs. Cardiac cells rely on intercalated discs for both electrical and mechanical purposes (168). Desmosomes are

responsible for cell-cell adhesion and help to resist forces between cells. Proteins from three separate families make the desmosome: the cadherins (four desmogleins and three desmocollins), the armadillo proteins (plakophilin and plakoglobin) and the plakins (desmoplakin) (Figure 2.1) (168, 169).

### (i) Desmosomal cadherins

The cadherins are the transmembrane component of the desmosome and the extracellular domains interact with the neighboring cell for adhesion (Figure 2.1). In addition to adhesion purposes another role of the desmosomal cadherins is to regulate morphogenesis (168). The two cadherins that have been associated with ARVC are desmoglein-2 (DSG-2) and desmocollin-2 (DSG2). They are both calcium-binding transmembrane glycoprotein that have a extracellular region with cadherin domains (which calcium binds to stabilize), a short transmembrane domain, and cytoplasmic region (170). The majority of DSG-2 mutations that are associated with ARVC are extracellular, suggesting that cell-cell adhesion has an important role in cardiomyopathies (161).

Interestingly, one of the non-desmosomal ARVC genes encodes RYR-2, a calcium channel. It would be worth studying the relationship between this channel and the desmosomal cadherins, considering these proteins are stabilized by calcium binding. It is known, however, that in myocardial cells RYR-2 is associated with the prolyl *cis-trans* isomerase FKBP1B (also known as FKBP-12.6), which is a protein that plays a role in excitation-contraction coupling in cardiac muscle by controlling the opening of the RYR-

2 channels and regulating cytosolic calcium concentrations (51, 52). Releasing calcium from a cell terminates muscle contractions and prevents diastolic depolarization, which can cause ventricular arrhythmias (121).

### (ii) Armadillo proteins

The desmosomal armadillo proteins (plakoglobin and plakophilin) connect the desmosomal cadherins and desmoplakin (Figure 2.1). Plakoglobin functions in cell assembly in both the desmosome and adheren junctions, but is also expressed in the WNT signaling pathway (170). It has been shown, in a myocyte cell line, that when plakoglobin is freed from the desmosomal complex it translocates to the nucleus where it competes and opposes the action of β-catenin and down-regulates the canonical *WNT*/β-catenin signaling pathway (171). Suppression of the canonical *WNT*/β-catenin signaling up-regulates the expression of adipogenic and fibrogenic genes and causes fat droplets to accumulate in the cells, thus, re-generating the ARVC phenotype in myocytes (171).

Plakophilin-2 is another armadillo protein associated with ARVC (158, 159). Plakophilins are widely expressed in many epithelial tissues, but plakophilin-2 is the only member of this family expressed in cardiomyocytes (157). It was recently determined that mutant plakophilin-2 leads to the disruption of desmosome assembly and specific mutations have different disruptive effects (159). Plakophilin-2 does not localize to the plasma membrane or have the ability to recruit desmoplakin to the desmosomal complex when the *carboxyl*-terminus is mutated (159).

### (iii) Desmosomal plakin – Desmoplakin

Desmoplakin (DSP) is exclusively found in desmosomes but is ubiquitously expressed. It is the key protein to the inner desmosome. It is responsible for desmosomal assembly, stability, regulation and modeling of tissues during embryogenesis (168). The *N*-terminal domain of desmoplakin binds to the armadillo proteins, and the *C*-terminus of desmoplakin is responsible for anchoring desmin (an intermediate filament in the sacromere) to the cell surface (Figure 2.1) (168, 169). A DSP mutant (R2834H) mouse model was established and over-expression of that protein led to cardiac bi-ventricular apoptosis, fibrosis, enlargement and dysfunction, whereas wild-type DSP mice had no adverse effects (172).

## 2.1.9 ARVC variable expression and penetrance

Specific studies attempting to characterize genotype-phenotype correlations have been carried out in ARVC cohorts and the phenotype has been shown to be variable. One study provided a clinical evaluation of ARVC families harboring mutations in *PKP-2* (173). The clinical expression of *PKP-2* mutations in affected individuals varied tremendously, even amongst first-degree relatives, ranging from a complete lack of symptoms to a severe disease phenotype. Evaluation of the genotyped family members showed that strict adherence to the International Task Force diagnostic criteria would have omitted several affected individuals. The age of disease expression appeared to vary quite widely in these *PKP-2* families as well. The age at diagnosis was between 8 to 45 years (median of 33 years) (173).

Another study investigated the genotype-phenotype correlation for *DSG2* mutations and ARVC clinical expression (163). The mean age at diagnosis for the probands was 32.6 (a range from 14–59 years), indicating that ARVC is a progressive disorder so younger asymptomatic individuals may develop the disease later in life. Dividing gene positive individuals into age groups (20–40 and >40 years of age) showed higher penetrance in the older group (50 and 75% respectively, for Task Force criteria) (163).

### 2.1.10 Aims of this work

The most recent ARVC gene discoveries have taken a candidate gene approach by screening desmosomal genes for mutations. The main objective of this work was to identify the *ARVD5* disease gene. The *ARVD5* locus was previously mapped in a large Newfoundland family (146). Therefore, positional cloning was used to discover the causative gene.

## 2.2 Materials and Methods

### 2.2.1 Study population

Fifteen families were referred to either the Newfoundland Provincial Medical Genetics Program or the Newfoundland Labrador genetics cardiomyopathy clinic because of a family history of cardiomyopathy and sudden death (Figure 2.2, 2.3, 2.4, and 2.5). The first and largest family (Family 64) (Figure 2.2) was described as having ARVC in the late 1980s (147). Correctly diagnosing the individuals in these families using the International Task Force criteria (126) was difficult. Each family had an extended pedigree, and SCD was generally the primary disease presenting feature that appeared to be inherited in an autosomal dominant pattern; autopsy reports were not available to confirm ARVC as the cause of death in most of the deceased individuals of earlier generations. As well, these families originate from central Newfoundland where a tertiary medical center is not accessible. As such, the 'standard' testing required to diagnosis ARVC based on the International Task Force criteria was not available which affected the ability to define affected individuals.

In this study, a subset of disease features was used to define affection status (Figure 2.6) (174), and only well-ascertained individuals born at *a priori* 50% risk were investigated. These individuals were divided into three groups according to their presentation of disease; primary affection status (clinically affected), secondary affection status and clinically unaffected (defined in Figure 2.6). A total of 295 individuals across the 15 families had blood drawn for the genetic analysis (Figure 2.6). Originally, genomic DNA

of participants was extracted from lymphocytes and stored in the DNA diagnostic lab of Eastern Health. Recently, blood samples were sent to the research laboratory of Dr. Terry-Lynn Young where genomic DNA was extracted and cell lines established. Informed consent was obtained in compliance with the Human Investigation Committee requirements of the Eastern Health Corporation of St. John's, Newfoundland, Canada (study number 00-176).

### 2.2.2 Defining the *ARVD5* locus

Family 64 (Figure 2.2) was the family used to map the *ARVD5* locus to chromosome 3p (146). Only one marker with a LOD score over 1.5 was identified, *D3S3613* (a peak 2-point LOD score of 6.91 ($\theta$=0)). Ahmad et al. built a thirteen marker haplotype (Table 2.2) using 10 living affected individuals and two deceased individuals, whose haplotypes were inferred from their children (146). A telomeric recombination between *D3S3610* and *D3S1585* in individuals V:19, VI:19 and VI:21, along with a centromeric recombination between *D3S1293* and *D3S3659* in individuals IV:18 and IV:32, identified a 9.93 Mb disease region between markers *D3S3610-D3S3659* (Table 2.2) (146).

The *ARVD5* haplotype was recapitulated in the Young lab (175) by first genotyping genomic DNA from individuals in Family 64 using 18 microsatellite markers (including markers in the original marker set) (Table 2.2) and manually reconstructing haplotypes (Figure 2.7A). Similarly, genomic DNA from all available family members of the 14 additional families was genotyped and the same disease-associated haplotype was

determined to also segregate with ARVC in those families. The shared haplotype between families was presumed to be ancestral (Figure 2.7A). In the original mapping paper, *D3S3610* was the telomeric boundary marker (146). In the Young lab, the ancestral *ARVD5* allele at that marker was called '246' (Figure 2.7A). This boundary was confirmed in families 76 and 453 (Figure 2.3 and 2.4) where all individuals with the *ARVD5* haplotype had the allele '256' instead of '246' (Figure 2.7A). Another recombinant haplotype was identified in family 840 (Figure 2.4) where all individuals with the *ARVD5* haplotype had allele '351' instead of '347' at marker *D3S1516* (Figure 2.7A). However, it was cautiously decided to only narrow the region if recombinations were seen in two or more families. On the centromeric side of the *ARVD5* haplotype, a recombination was observed between markers *D3S3595* and *D3S3613* in all clinically affected individuals in families 69 and 273 (Figure 2.3, 2.4 and 2.7A). That key recombination reduced the critical region to 2.36 Mb from 9.93 Mb between markers *D3S3610* and *DS33613* (Figure 2.7A) (175).

### 2.2.3 Screening positional candidate *ARVD5* genes

The contig of the 2.36 Mb critical region was determined using the UCSC Genome Browser homepage (http://genome.ucsc.edu/index.html?org=Human), and the March 2006 assembly (Build 36.1). In search for the *ARVD5* gene, all annotated genes in the critical region from Refseq were noted (Figure 2.7B and Table 2.3) and screened.

### (i) Customized primer design and work up

Primer sets of all coding and non-coding exons, and intron-exon boundaries of all positional candidate genes for *ARVD5* were designed using Primer 3 (176). In order to determine optimal amplification conditions of these primer sets, trial PCRs were run on the newly designed primers. A typical trial for one primer set involved two different PCR cocktails, one with betaine and the other without, because amplification of GC rich regions can be enhanced by using betaine (Appendix 1). A set of three different DNA samples and one $H_2O$ (water) control was used for each trial reaction. Amplification was first carried out using touchdown (TD) 54 cycling conditions (Appendix 2).

In order to determine, for each primer set, which of the two cocktails (betaine or no betaine) under the TD54 cycling conditions amplified the DNA best, reactions were electrophoresed on a 1% agarose gel made with 1X TBE (Tris/Borate/EDTA) buffer, and 5 µl of ethidium bromide (from a stock solution of 10 mg/ml) per 100 ml of gel solution for a final concentration of 0.5 µg/ml. Generally, 5 µl of PCR product and 1 µl of bromophenol blue/xylene-cyanol dye were added to each well. A 100 bp ladder from Invitrogen was used for band sizing. The Kodak MI software was used to visualize the banding pattern on a gel.

If no product was observed using these conditions, then the annealing temperature was lowered and/or the final $MgCl_2$ concentration in the PCR cocktail was increased to make primer binding less specific/stringent. In contrast, if multiple bands were seen after the

first trial then the annealing temperature was increased and/or the final $MgCl_2$ concentration in the PCR cocktail was decreased to make primer binding more specific/stringent. If more or less $MgCl_2$ was added to the PCR mix then the $H_2O$ volume was adjusted according so that the final volume of the 1X PCR mix was still 25 µl. If all else failed, an additional primer set was ordered. The conditions for all the primer sets used during this project are listed in Appendix 3.

The PCR products were purified using 50% sephacryl (Amersham Biosciences) and MultiScreen HTS filter plates (Millipore Corporation). Purified PCR products were cycle sequenced in both the forward and reverse directions with the use of a BigDye Terminator V3.1 cycle sequencing kit on an automated ABI 3700 DNA analyzer. The sequencing reaction had a final volume of 20 µl and 1/16 of the recommended volume of sequencing mix was used. A single reaction contained 0.5 µl of sequencing mix, 2 µl of 5X Sequencing Buffer, 1 µl of purified PCR product, 2 µl of either the forward or reverse primer (1.6 µM) and 14.5 µl of $H_2O$. The cycle sequencing PCR had 25 cycles. Each cycle involved a 96°C – 10 second denaturing period, 50°C – 5 second annealing period, and a 60°C – 4 minute extension on an ABI PCR GeneAmp 9700 thermocycler.

After cycle sequencing, 5 µl of 125 mM EDTA and 65 µl 95% ethanol were added to each sample. The samples were then precipitated over night, in darkness, preferably at 4°C. After precipitation, the samples were centrifuged at 3000 g for 30 minutes to pellet the DNA. The plate was then inverted and placed in the centrifuge for approximately 20

seconds at 200 rpm to remove the supernatant. Then the samples were washed by adding 150 µl of 70% ethanol and centrifuging the plate at 3000 g for 15 min. Another quick inverted spin was carried out to remove the excess ethanol. The samples were then left for 15 minutes to air dry. At that time, 15 µl of dimethyl-formamide (DMF) was added to each sample. The samples were then denatured for 2 minutes at 95°C and ready to be sequenced. Sequencing electropherograms were inspected manually (ABI Sequencing Analysis version 5.2) and analyzed with Mutation Surveyor software (Transition Technologies) (Appendix 4).

### (ii) Gene/Mutation screening

A mutation screening panel was established that comprised seven genomic DNA samples from four clinically affected subjects, from three families (64, 453, and 840), three unrelated spouses (controls) and one $H_2O$ control. The four affected individuals chosen to be on this panel had primary affection status (Figure 2.6) and the disease haplotype. Based on the haplotypes that were built using microsatellite markers during the re-construction of the *ARVD5* haplotype, some affected individuals appeared to be homozygous for distal portions of the haplotype; in addition, clinically unaffected individuals had distal portions as well. The affected individuals that were chosen to be on the mutation screening panel had only one copy of the haplotype. The three controls were unrelated to the ARVC families, did not have the disease haplotype and showed no clinical signs. These eight samples occupied one column on a 96 well plate (8 rows X 12 columns), therefore, 12 different primer sets (generally representing one exon) were

amplified on one plate. This panel was screened for all primer sets (Appendix 3) under the optimal conditions that were determined from the above PCR trials.

Priority screening of the positional candidate genes was based on gene function, expression, and information given to the Young laboratory from Dr. Ludwig Thierfelder in Germany, who previously attempted positional gene screening. The first gene to be screened was *fibulin-2* (*FBLN2*). *FBLN2* was an interesting candidate because previous screening detected several potential pathogenic variants (data not published, Thierfelder personal communication), and furthermore, the gene product is involved in organ development, particularly, the differentiation of heart, skeletal and neuronal structures. The second gene to be screened was *WNT7A*. The WNT gene family consists of structurally related genes that encode secreted signaling proteins. These proteins are involved in regulating cell fate and patterning during embryogenesis. It was recently determined that suppression of the WNT/β-catenin pathway by plakoglobin resulted in the ARVC phenotype (171), thus *WNT7A* was a good candidate. The remaining positional candidates were generally chosen based on their size, with the smaller genes screened first.

As each positional candidate gene was sequenced all variants were recorded in an excel database (Appendix 5). This database was created before the mutation screening commenced, at which time, only the genotypes of each microsatellite marker in the *ARVD5* haplotype for all seven individuals chosen to be on the mutation screening panel

were included. Each marker in the haplotype was designated to a row in the database (based on their Mb position on chromosome 3), such that the previously determined haplotypes (175) were visible vertically in columns. As each new sequencing variant was identified, we were able to continue building the haplotypes by placing the variants in the database (based on its chromosome position) and displaying the corresponding genotypes for each individual on the screening panel (based on the most likely haplotypes). If a variant was detected in all affected individuals on the screening panel, that variant was assumed to be on the *ARVD5* haplotype during this stage of the analysis.

### 2.2.4 Segregation Analysis

Variants that were found exclusively in clinically affected subjects on the mutation screening panel were of interest. In order to verify if these variants did in fact reside on the *ARVD5* haplotype segregation analysis in family 1139 was carried out (Figure 2.5).

### 2.2.5 *ARVD5* Allele Frequencies

The allele frequencies of the *ARVD5* sequencing variants were determined using Newfoundland population-based controls, which were obtained through random digit phone dialing, as part of a large colorectal cancer study (177). Sequencing variants that were determined to be rare (<1% of the control alleles screened) were screened in all affected individuals for which DNA had been collected to determine which rare variant(s) were shared amongst all clinically affected individuals.

### 2.2.6 Bioinformatic Analysis

Conservation of the TMEM43 protein across species was determined using ClustalW and Weblogo (170, 178, 179), and the effects of amino acid substitutions on protein function were also predicted (180-184). Potential protein localization, function, structure and post-translational modification sites were predicted using the online tools via the ExPASy web site (170).

## 2.3 Results

### 2.3.1 Positional candidate genes and mutation screening

The 2.36 Mb critical region contained 20 annotated genes (Figure 2.7B and Table 2.3). After all 20 genes were screened a total of 240 variants were identified (Appendix 5). Below the genes are described in the order they appear on the ARVC haplotype (Figure 2.7B).

***IQSEC1* (*IQ motif and Sec7 domain 1 protein*)**

*IQSEC1* encodes a protein of unknown function (51). Five non-coding variants were found in *IQSEC1*, three within introns and two within the 3′ UTR. Four of these variants were detected in controls on the screening panel and one, c.65+73C>G, was detected in only one affected individual on the panel. None of these five variants were found exclusively in all affected individuals on the screening panel so they were excluded as the causal *ARVD5* variant (Appendix 5).

***NUP210* (*Nucleoporin 210 precursor*)**

*NUP210* encodes a membrane-spanning glycoprotein known as a nucleoporin protein. The nucleoporin family of proteins is the main component of the nuclear pore complex, the structure that acts like a gateway and regulates the flow of macromolecules between the nucleus and the cytoplasm (51). Thirty three variants were found in *NUP210*. Twenty were non-coding (five were within the 3′ UTR and 15 were intronic) and 13 were coding. Twenty-eight of these variants were found in controls on the screening panel and five, c.5664+365A>T, c.4582C>T, c.2964+143C>T, c.305-17T>C and c.166+130G>A,

were found in only one affected individual. Again, none of these variants were found exclusively in all affected individuals on the screening panel so they were excluded as the causal *ARVD5* variant (Appendix 5).

***HDAC11 (histone deacetylase 11)***

*HDAC11* encodes a class IV histone deacetylase. This protein localizes to the nucleus and may be involved in regulating the expression of interleukin 10 (51). There were five variants detected in *HDAC11*, all of which were non-coding, intronic variants. Of these variants, three were found in controls on the screening panel, and one, c.490 -97 C>T, was found in only one affected individual. Thus, these four variants were excluded (Appendix 5). There was one variant of interest, c.369+18_369+19insG that was found exclusively in all the affected individuals. This variant was not excluded despite one affected individual on the screening panel being a homozygote (Appendix 5).

***FBLN2 (fibulin 2)***

*FBLN2* encodes an extracellular matrix protein that belongs to the fibulin family. This protein is known to bind various extracellular ligands and calcium, and it is suspected to play a role during organ development. More specifically, it may be involved in the differentiation of heart, skeletal and neuronal structures (51). Twenty six variants were detected in *FBLN2*. Thirteen of these variants were non-coding (six were intronic and seven were within the 3′ UTR), seven were coding and six were found in intragenic ESTs (expressed sequence tags) that were screened (Appendix 5). All of these variants were

excluded from further analysis because they were detected in controls on the screening panel (Appendix 5).

***WNT7A* (*wingless-type MMTV integration site family member 7A*)**

*WNT7A* is part of the WNT gene family, a group of structurally related genes that encode secreted signaling proteins. These proteins are involved in regulating cell fate and patterning during embryogenesis (51). Seven variants were detected in *WNT7A*, three non-coding (two intronic and one 5′ UTR variant), two were coding and two were found in intragenic ESTs that were screened. Five of the variants were found in controls and two (one EST variant and the 5′ UTR variant) were only found in one affected individual. All *WNT7A* variants were excluded from further analysis (Appendix 5).

***TPRXL* (*tetrapeptide repeat homeobox-like protein*)**

*TPRXL* is a homeobox gene since it has a conserved DNA sequence called the homeobox that encodes a DNA-binding domain – the homeodomain. Many homeobox gene products are thought to be involved in early embryonic development, however the exact function of TPRXL is unknown (51). Two non-coding intronic variants were detected in *TPRXL*, which were found in controls on the screening panel thus excluded (Appendix 5).

***CHCHD4* (*coiled-coil-helix-coiled-coil-helix domain containing protein 4*)**

*CHCHD4* is a component of human mitochondria and belongs to a protein family that has 6 highly conserved cysteine residues within a particular motif in the C terminus (51). Four non-coding variants were detected in *CHCHD4*, including two intronic variants, one

3′ UTR variant and one 5′ UTR variant. All were observed in controls so were excluded from further analysis (Appendix 5).

***TMEM43* (transmembrane protein 43)**

*TMEM43* encodes an inner nuclear membrane protein that is also known as LUMA (185). There were fifteen *TMEM43* variants detected. Thirteen of these were non-coding variants (six intronic and seven within the 3′ UTR) and two coding variants. All variants could be excluded except one, c.1073C>T, which was found exclusively in all affected individuals on the screening panel (Appendix 5).

***XPC* (*xeroderma pigmentosum, complementation group C*)**

*XPC* encodes a protein that is involved in the nucleotide excision repair pathway (51). Twenty four variants were detected in *XPC*. Sixteen were non-coding (five were within the 3′ UTR, 10 were intronic and one was within the 5′ UTR) and eight were coding. Twenty-three of these variants were found in controls on the screening panel and were excluded, however one variant, c.2823+684G>C, was found exclusively in all affected individuals on the screening panel (Appendix 5).

***LSM3***

*LSM3* encodes a Sm-like protein, which has sequence homology with the Sm protein family. LSM3 contains the Sm sequence motif, like all Sm-like proteins, and may be involved in splicing (51). There were no variants detected in *LSM3*.

***SLC6A6* (*solute carrier family 6 member 6*)**

*SLC6A6* encodes a neurotransmitter, more specifically, it belongs to the sodium:neurotransmitter symporter family (51). There were thirty two variants found in *SLC6A6*, all of which were non-coding. Twelve were 5′ UTR variants, nine were intronic and 11 were 3′ UTR variants. Twenty-nine variants were excluded from further analysis, three of which were found in only one affected individual, and 26 were found in the controls on the panel (Appendix 5). There were three *SLC6A6* variants, c.1-27420G>A, c.599+370A>G and c.733-1226A>G, however that were found exclusively in all affected individuals on the screening panel (Appendix 5).

***GRIP2* (*glutamate receptor interacting protein 2*)**

GRIP2 plays an important role in neuronal cells by acting as a scaffold for the assembly of multiprotein signaling complexs and mediating the trafficking of its binding partners (51). There were fourteen variants detected in *GRIP2*. Eight were non-coding intronic variants and six were coding variants. All fourteen variants were excluded from further analysis, six were found in only one affected individual and the remaining eight were found in controls (Appendix 5).

***C3orf19***

C3orf19 is a hypothetical protein, also known as LOC51244. Twelve variants were detected in *C3orf19*, all of which were non-coding. Eleven were intronic and one was within the 3′ UTR. All variants were excluded since they were all detected in controls on the screening panel (Appendix 5).

***C3orf20***

C3orf20 is another hypothetical protein in the *ARVD5* critical region that is also referred to as LOC84077. There were fifteen variants detected in *C3orf20*. Eight variants were non-coding (three were within the 5′ UTR and five were intronic) and seven were coding. All variants were excluded. Ten were found in control samples, one was found in only one affected individual and two variants were found in two of the four affected individual on the panel (Appendix 5).

***FGD5 (FYVE, RhoGEF and PH domain containing 5)***

FGD5 is a protein of unknown function but it may activate members of the Ras-like family of Rho- and Rac proteins, and/or play a role in regulating the actin cytoskeleton and cell shape (51). Ten variants were detected in *FGD5*, seven of which were non-coding intronic variants and three were coding. Six of these variants, c.934G>A, c.2186+22G>A, c.2187-82G>A, c.2220G>T, c.2613+50C>T and c.3085-74G>A, were found exclusively in all the affected individuals (Appendix 5). Only four were excluded, three of which were seen in controls and one was seen in only two of the affected individuals on the panel (Appendix 5).

***NR2C2 (nuclear receptor subfamily 2, group C, member 2)***

NR2C2 is a member of the nuclear hormone receptor family, and acts as a ligand-activated transcription factor (51). Nine variants were detected in *NR2C2*. All variants, but one, c.1685G>A, were non-coding. Three of these variants, c.855+70G>A, c.1848+365T>A, and c.1848+2965_1848+2966insGATA, were of interest because they

were exclusively seen in all affected individuals on the panel (Appendix 5). The remaining six variants were excluded. One was detected in only one affected individual and five were seen in the controls (Appendix 5).

***MRPS25* (*mitochondrial ribosomal protein S25*)**

*MRPS25* is a nuclear gene that encodes a mitochondrial ribosomal protein, which aids in protein synthesis within the mitochondrion (51). Six variants were detected in *MRPS25*, all of which were non-coding. One variant was within intron 2 and the remaining five were in the 3′ UTR. Four of the six variants could be excluded; one variant was seen in only one affected individual and the other three were seen in controls on the panel (Appendix 5). One variant, c.522+1059G>A, was observed in all affected individuals exclusively (Appendix 5).

***ZFYVE20* (zinc finger *FYVE domain-containing protein 20*)**

ZFYVE20 is a Rab4/Rab5 effector protein that acts in early endocytic membrane fusion and membrane trafficking of recycling endosomes (51). Seven variants were detected in *ZFYVE20.* One variant, c.1734C>G, was coding and six variants were non-coding. The non-coding variants consisted of one intronic variant and five 3′ UTR variants. All variants were excluded because they were detected in controls (Appendix 5).

***CAPN7* (*calpain 7*)**

CAPN7 is a member of the calpain family of proteins, which are a well-conserved family of calcium-dependent, cysteine proteases. The exact function of CAPN7 is however unknown (51). Six variants were detected in *CAPN7*, all of which were non-coding. One

variant was detected in the 5′ UTR and the rest were intronic (Appendix 5). Two of these variants, c.1289+68C>T and c.1430-28T>C, were of interest since they were detected exclusively in all the affected individuals on the screening panel (Appendix 5). The remaining four were excluded (Appendix 5).

***SH3BP5 (SH3-domain binding protein 5)***

*SH3BP5* encodes a SH3-domain binding protein of unknown function that was discovered due to its association with bruton tyrosine kinase, a cytoplasmic tyrosine kinase that is crucial for the maturation of B-lineage cells (186). Eight variants were detected in *SH3BP5*. Two variants were coding and six were non-coding. The non-coding variants included three intronic variants, two 3′ UTR variants and one 5′ UTR variant. All variants were excluded from further analysis because they were detected in control individuals on the screening panel (Appendix 5).

Of the 240 variants reported, in the 20 positional genes, a total of 18 variants were found exclusively in clinically affected individuals on the mutation screening panel and were used in further investigation (Appendix 5).

## 2.3.2 Segregation analysis

In order to determine which of the 18 sequencing variants, that were found exclusively in affected individuals from the screening panel, truly segregated on the *ARVD5* haplotype, six individuals from Family 1139 (Global ID: 708, 709, 710, 714, 716 and 718) were

genotyped for all 18 variants and haplotypes were created (Table 2.4, Figure 2.5 and 2.8). Two clinically affected individuals 710 and 709 (mother and daughter) who were determined to have the *ARVD5* haplotype using microsatellite markers were chosen for this analysis, along with the 21 year old daughter of 709, 714, who was also previously determined to carry the *ARVD5* haplotype, however, had a recombination between *D3S1516* and *D3S3608* (Figure 2.8). Individual 714 has yet to show clinical signs, thus, her recombination was not used to reduce the *ARVD5* disease region. Individual 708, who also showed no clinical signs, was another recombinant chosen for the segregation analysis (Table 2.4 and Figure 2.8). During the earlier microsatellite analysis a recombination between *D3S3595* and *D3S3613* was detected in this individual. This recombination is between the same markers as the recombination that reduced the critical region in families 69 and 273. However, clinically affected individuals in families 69 and 273 have the telomeric portion of the *ARVD5* haplotype (Figure 2.7A and 2.9), whereas individual 708 has the most distal centromeric portion (Figure 2.8 and Table 2.4). Two other individuals, 716 and 718, who did not have any clinical signs or the *ARVD5* haplotype were also chosen (Table 2.4 and Figure 2.8).

Segregation analysis in Family 1139 determined that five of the 18 variants, *XPC* c.2823+684G>C, *FGD5* c.2186+22G>A, *FGD5* c.2187-82G>A, *FGD5* c.2220G>T, and *FGD5* c.2613+50C>T did not reside on the *ARVD5* ancestral haplotype (yellow) (Table 2.4 and Figure 2.8). Of the remaining 13 variants that segregated on the *ARVD5* haplotype, two variants, *HDAC11* c.369+18_369+19insG and *SLC6A6* c.1-27420G>A,

were excluded as being the *ARVD5* mutation because they were observed on at least one other haplotype that segregated through Family 1139 (Table 2.4). In addition to the yellow *ARVD5* haplotype, *HDAC11* c.369+18_369+19insG also segregated on the red, grey and orange haplotype, and *SLC6A6* c.1-27420G>A segregated on the blue haplotype (Table 2.4). Eleven *ARVD5* variants remained of interest (Table 2.4 – highlighted grey).

### 2.3.3 Identification of the causal variant

Allele frequencies of the 11 *ARVD5* variants were determined. Six variants were considered common with allele frequencies ranging from 9 to 55% (Table 2.4). The remaining five, however, were rare alleles (<1% of the alleles screened) with frequencies between 0 and 0.60% (Table 2.4). The first rare allele to be screened in all clinically affected individuals from all 15 families was *FGD5* c.934G>A. Not all clinically affected individuals screened positive for this variant. The affected members of families 69 and 273 were wild-type for that allele (Figure 2.9). Family 69 had three individuals (two female, one male) with primary clinical affection status and two females with secondary affection status, all of whom, screened negative for that variation, and Family 273 had two individuals with primary affection status that screened negative as well (Figure 2.9). Those two families previously defined the new centromeric boundary by identifying a recombination between markers *D3S3595* and *D3S3613* (Figure 2.7A). The exact crossover point is unknown, however, *FDG5* lies within those two markers (Figure 2.7B) thus suggesting that the recombination occurred more telomeric than *FGD5* (closer to *D3S3595*). Three other rare variants, *MRPS25* c.522+1059G>A, *CAPN7* c.1289+68C>T,

and *CAPN7* c.1430-28T>C (Table 2.4), were located between these markers as well, and were more centromeric than *FGD5* (Figure 2.7B). The affected family members of families 69 and 273 screened wild-type for those alleles as well (Figure 2.9).

The only rare variant that was shared by all clinically affected subjects across all 15 families was *TMEM43* c.1073C>T (p.S358L) (Figure 2.10). In fact, *TMEM43* c.1073C>T was the only rare variant retained on key recombinant *ARVD5* haplotypes identified in clinically affected individuals from families 69 and 273 (Figure 2.9). This suggested that *TMEM43* is *ARVD5*.

### 2.3.4 *TMEM43* mutation screening in *ARVD5* linked families

All available subjects born at *a priori* 50% risk (n=295) across the 15 ARVC families were sequenced for the presence of the c.1073C>T *TMEM43* variant (Figure 2.6). All clinically affected individuals with primary affection status (n=83/83, 40 males, 43 females) were mutation carriers. All individuals with secondary affection status (n=23/23, 8 males, 15 females) were also mutation carriers. Twenty percent of clinically unaffected individuals (n=38/189, 10 males, 28 females) were mutation carriers (Figure 2.6). The clinically unaffected mutation carriers were at a median age of 22 and 33 years for males and females, respectively. The 151 subjects with no clinical signs who did not have the *TMEM43* variant were considered unaffected (Figure. 2.6).

### 2.3.5 Additional evidence supporting *TMEM43* as the cause of *ARVD5*

After screening all available spouses (n=47) and population controls (n=161) for the *TMEM43* variant, no mutation carriers were detected (416 mutation negative chromosomes) (Table 2.4). Also, clinically unaffected adults (from *ARVD5* families) who shared distal sections of the *ARVD5* haplotype that lacked the *TMEM43* mutation were identified (Figure 2.11). There were several cases where individuals, with no clinical signs of ARVC, had the centromeric portion of the disease haplotype, including four of the five rare variants, *FGD5* c.934G>A, *MRPS25* c.522+1059G>A, *CAPN7* c.1289+68C>T, and *CAPN7* c.1430-28T>C. These individuals included individual 708 (also known as 1139.016), a female from Family 1139 who is in her late 40s with no clinical symptoms to date (Table 2.4 and Figure 2.5 and 2.8), and three other individuals from Family 64 with no clinical signs, individual 0064.1011 (a female in her late 60s) (Figure 2.2), and individuals 0064.0029 and 0064.0030 (females in their late 40s who are not in the reduced pedigree in Figure 2.2). There was also one case where an individual from Family 964 (0964.0004 - a male in his late 60s with no clinical signs) was determined, through fine-mapping, to have the telomeric portion of the disease haplotype, but had a recombination and therefore lacked *TMEM43* c.1073C>T and the rest of the haplotype (Figure 2.5 and 2.11).

### 2.3.6 ARVC linked to the *ARVD5* locus is caused by a missense mutation in *TMEM43*

*TMEM43* (Genbank Accession number NM_024334) has 12 exons, and encodes a 400 amino protein known as transmembrane protein 43 and/or LUMA (185, 187). The protein is 98% similar to the mouse protein and is well conserved across all eukaryotic and prokaryotic species (Figure 2.12). In 2001, TMEM43 was first recognized as an inner nuclear membrane protein (185), a result that was later confirmed in another independent proteomics study (188). Interestingly, the first characterization of the TMEM43 protein was recently published, linking it to proteins known to cause cardiac disease (187). Bioinformatic analysis of TMEM43 predicts it to be a membrane protein with several potential post-translation modification sites (Figure 2.13). However, unlike the transmembrane proteins of the desmosome (desmocollin and desmoglein), TMEM43 does not have a cadherin domain. Furthermore, protein sequence alignments with desmocollin and desmoglein show less than 10% identity and less than 12% similarity. The mutation, p.S358L, occurs within the third predicted transmembrane domain and is highly conserved in mammalian, avian, amphibian, and insect orthologs (Figure 2.12 and 2.13). Interestingly, a leucine residue at this position is found in the bacterium *Rhizobium loti* but it is not found in any multicellular organisms (Figure 2.12). The p.S358L mutation is also predicted through bioinformatic analysis to have a deleterious effect on TMEM43 structure and function.

## 2.4 Discussion

A founder effect within the population of Newfoundland enabled the *ARVD5* gene discovery. By studying 15 ARVC Newfoundland families and using a positional mapping approach, *TMEM43* was identified as the causal gene for *ARVD5*. The *ARVD5* locus was initially mapped in one of the 15 families, Family 64, which at the time, extended seven generations and contained 200 individuals (146). Ten living affected individuals and 17 deceased individuals (who had all died suddenly) were noted on the pedigree. After a two-point linkage analysis, marker *D3S3613* on 3p25 (LOD score 6.91 ($\theta$=0)) was determined to be the only marker with a LOD score above 1.5, and fine mapping identified a 9.3 cM disease haplotype between markers *D3S3610* and *D3S3659*. Despite that the pedigree of Family 64 has currently been extended from 200 to 1200 individuals, the disease region would not have been reduced to 2.36 Mb if the additional ARVC families, especially families 69 and 273, were not available for study. Affected individuals within families 69 and 273 all had a recombination between markers *D3S3595* and *D3S3613*, which reduced the centromeric side of the *ARVD5* haplotype. Perhaps individuals in families 69 and 273 are more closely related and are descendents of an affected recombinant individual from years ago. The telomeric boundary that was detected in the original mapping paper was confirmed in families 76 and 453, where all individuals had the same recombinant allele at *D3S3610*. Perhaps, again, the individuals from families 76 and 453 are descendents of another recombinant affected individual.

The 2.36 Mb region contained 20 positional candidate genes. Deciding which methodology that should be used to screen the positional candidate genes, how to prioritize each candidate's screening, and which individuals were to be placed on the mutation screening panel were very important for efficiency and reliability. With 357 amplicons required to amplify the 275 exons of the 20 positional candidate genes, Sanger sequencing was used to identify all point mutations and small insertions/deletions with certainty. The screening began by selecting the best candidates based on expression, function and prior screening knowledge. However, in the end all positional candidates were screened. Many clinically affected individuals with the *ARVD5* ancestral haplotype were identified during the reconstruction of the haplotype (175), however, only a select few of these individuals needed to be on the mutation screening panel since they all shared the same ancestral haplotype. To be placed on the screening panel, the affected individual had to ultimately have a primary affection status. Carefully selecting the controls to be placed on the mutation screening panel was critical as well. Since ARVC is difficult to diagnose, controls were chosen if they were not blood related to the 15 families, did not have clinical signs of ARVC, and were previously determined not to have the *ARVD5* haplotype. Screening four clinically affected and three unaffected controls, along with a $H_2O$ control, enabled 12 amplicons per sequencing plate, making the screening process as efficient as possible.

After mutation screening, a critical process of elimination was carried out to identify the disease variant. There were initially 240 variants detected, however only 18 variants

were found exclusively in clinically affected individuals on the mutation screening panel and were further investigated. Segregation analysis determined that 11 of those variants segregated on the *ARVD5* haplotype, but after calculating the Newfoundland population allele frequencies only five rare variants remained of interest. Four of those variants happened to be located between markers *D3S3595* and *D3S3613*, where the critical recombination was noted in families 69 and 273 that reduced the disease region to 2.36 Mb. These four variants were determined not to be present in any affected individual in these two key recombinant families, identifying the *TMEM43* variant, c.1073C>T, as the only rare *ARVD5* variant that was shared by all affected individuals within the 15 ARVC Newfoundland families.

*TMEM43* is the third non-desmosomal ARVC gene to be identified. The TMEM43 protein is evolutionarily conserved and the presumed amino acid substitution (p.S358L) is predicted to be deleterious. This mutation was not detected in spouses or population controls. Bengtsson and Otto specifically determined that TMEM43 interacts with emerin and participates in controlling its distribution along the nuclear membrane (187). Emerin [MIM #300384] is a lamina-associated LEM-domain protein that causes an X-linked form of Emery-Dreifuss muscular dystrophy (EDMD) [MIM #310300] (189). EDMD is a laminopathy, which are a group of rare genetic disorders caused by mutations in genes encoding proteins of the nuclear lamina and are generally due to mutations in A-type lamins (190). Bengtsson and Otto also determined that TMEM43 binds both A- and B-type lamins and depends on A-type lamins for its localization. Thus, they suggested

that TMEM43 is an integral nuclear membrane protein that structurally and functionally organizes inner nuclear membrane protein complexes and has the potential to cause pathological changes to the nuclear envelope (187).

Recently, signaling pathways have been implicated in ARVC pathogenesis as well (171, 191). For example, plakoglobin, when freed from desmosomal complexes, translocates to the nucleus where it competes and opposes the action of β-catenin and down regulates the canonical *WNT*/β-catenin signaling pathway (171). Suppression of the canonical *WNT*/β-catenin signaling up-regulates two adipogenic transcription factors, C/EBP-α [MIM #116897] and PPARγ [MIM #601487] (171). A genome wide scan for peroxisome proliferator response elements (PPREs) identified 1085 potential target genes of PPARγ, including *TMEM43* (192). If TMEM43 is a part of an adipogenic pathway regulated by PPARγ, then perhaps, dysregulation of this pathway may explain the fibrofatty replacement of the myocardium in ARVC patients.

After the third desmosomal gene, *PKP-2*, was identified as a causative ARVC gene in 2004, it was suggested that ARVC may be a disease of the desmosomes (156). This initiated ARVC gene discovery efforts using a candidate gene approach. Several groups screened *DSG2* and *DSC2*, two genes that encode desmosomal cadherin proteins, in ARVC probands and identified the *ARVD10* and *ARVD11* loci, respectively (161-165). Another group screened plakoglobin (*JUP*), the recessively mutated Naxos disease gene, in a German family with autosomal dominant ARVC and identified the first autosomal

dominant ARVC *JUP* mutation (166). Before this *ARVD5* gene discovery, there were 12 known autosomal dominant *ARVD* loci and seven known genes, five of which code for desmosomal proteins; noting the importance of cell adhesion molecules in ARVC pathogenesis (167). Despite the success of the recent candidate gene approaches, the work presented in this chapter exemplifies how that approach will not solve every story.

The two previously determined non-desmosomal ARVC genes are *ryanodine receptor 2* (*RYR2*) and *transforming growth factor 3* (*TGFβ-3*) (141, 143). Multiple missense variants in *RYR2* have been reported as pathogenic (143). *RYR2* encodes a membrane calcium channel, and associates with the prolyl cis-trans isomerase FKBP1B, a protein that plays a role in excitation-contraction coupling in cardiac muscle by controlling the opening of the RYR-2 channels and regulating cytosolic calcium concentrations (51, 52). Regarding ARVC pathogenesis, if *RYR2* is mutated, ventricular arrhythmias may arise if calcium is no longer released from a cell, as this affects the termination of muscle contractions and prevents diastolic depolarization (121). Interesting as well, desmosomal cadherins are stabilized by calcium binding, so a relationship between RYR-2 and the desmosomal cadherins is worth studying. *TGFβ-3* as an ARVC gene suggests that there is a disruption of some type of signaling pathway in ARVC pathogenesis (121). The TGFβ-3 protein is a cytokine with regulatory roles in tissue repair and remodeling. It plays a major role in cardiac morphogenesis, and knock-out studies in mice determined it specifically plays a role in mesenchyme differentiation and the development of fibrous septum of the atrium and fibrous skeleton of the heart tissue (193). *In vitro* expression

assays with ARVC-mutant constructs showed twofold increase of TGFβ-3 expression compared to wild-type constructs (142), which may explain the excess fibrosis observed in ARVC cases. However, it is important to note that no mutations have been reported in the other families previously linked to *ARVD1* mutations (133, 141), which questions the gene's validity.

Strictly using the International Task Force criteria (126) to diagnose ARVC in this study was difficult because of the large historic nature of the pedigrees (SCD being the primary disease feature) and the lack of availability of tertiary testing centers. Therefore, a subset of disease features that defined affection status was established in this study (174). Mutation status for the defined three groups was as follows: primary affection status (clinically affected) (83/83), secondary affection status (23/23) and clinically unaffected (38/189). All individuals that had either primary affection status or secondary affection status were mutation carriers. This highlights the variability of the expressed phenotype. Also, 20% of people that had no clinical signs were mutation carriers; the clinically unaffected mutations carriers were determined to be young (median age of 22 years for males and 33 years for females). This can be explained by age-related penetrance. The results of a penetrance study performed by Kathy Hodgkinson showed that ARVC was fully penetrant in males by the age of 63 and in females by the age of 76 (174).

In regard to future perspectives, the size of the ARVC cohort in this study provides an opportunity to compare and define ARVC-specific clinical features in affected and

unaffected subjects. One hundred and forty-four individuals with the *TMEM43* mutation are available for study, as well as 151 unaffected controls. Previous studies that assessed the ARVC phenotype in patients with mutations in *plakophilin-2* (194), *desmoglein* (163) and *plakoglobin* (195) typically evaluated clinical features in only mutation carriers, with the assumption that all cardiac signs present in affected subjects are due to the underlying genetic defect. However, it may be that some clinical features are due to the genetic background of either the family or the source population, which would be interesting to study using this cohort. Furthermore, due to the serious repercussions of ARVC, this gene discovery has enabled a clinical diagnostic test to be designed and offered to at-risk family members, which will help with early diagnosis and the proper medical precautions that need to be taken in order to save lives - implantation of an implantable cardioverter-defibrillator. Finally, recent functional studies suggest that the responsible mechanism for the pathogenesis of ARVC is the suppression of canonical *WNT* signaling by nuclear plakoglobin, which enhances expression of adipogenic factors and leads to the differentiation of a subset of cardiac progenitor cells to adipocytes (171). ARVC is the first disease to be recognized as disrupting the differentiation of cardiac progenitor cells (196), and it will be interesting in the future to see what therapeutic developments unfold in order to reverse or prevent the ARVC phenotype.

As Bengtsson and Otto suggested (187), *TMEM43* has the potential to cause disease. Functionally studying the effects of the TMEM43 mutation (p.S358L) can validate its pathogenic effect and will enable a greater understanding of ARVC pathogenesis.

Detecting additional *TMEM43* mutations in ARVC probands from other populations would validate this study, as well as, provide evidence for the fact that *TMEM43* has a world-wide impact on ARVC. Interestingly, after analyzing the personal genome of a patient with a family history of vascular disease and sudden death, and querying disease-specific mutation databases, Ashely et al. reported rare variants in three genes associated with SCD, including *TMEM43* (197). Also, most recently, two *TMEM43* sequence variants, including c.1073C>T, were identified in a Danish ARVC cohort. Segregation analysis indicated that the Danish *TMEM43* c.1073C>T variant segregated with ARVC in the affected family (198). It would be interesting to determine through haplotype analysis if this is a founder mutation or a recurrent mutation from different populations. In the same report, evaluation of the expression of the desmosomal protein plakoglobin in *TMEM43* mutation carriers indicated reduced levels of the plakoglobin protein, which suggests that the *ARVD5* mutated protein TMEM43 may share a final common pathway with desmosome-associated ARVC (198). It will be interesting to determine precisely how these gene products link to a common disease mechanism.

**Table 2.1: The known ARVC loci and genes.**

| Locus | Inheritance Pattern | Syndromic | Chromosome location | Gene | Reference | |
|---|---|---|---|---|---|---|
| | | | | | Loci | Gene |
| Naxos | AR | Yes | 17q21 | *Plakoglobin (PKGB)* | Coonar et al., 1998 | McKoy et al., 2000 |
| ARVD1 | AD | No | 14q23-q24 | *Transforming growth factor β3 (TGFβ-3)* | Rampazzo et al., 1994 | Beffagna et al., 2005 |
| ARVD2 | AD | No | 1q41.2-q43 | *Cardiac ryanodine receptor (RYR-2)* | Rampazzo et al., 1995 | Tiso et al., 2001 |
| ARVD3 | AD | No | 14q12-q22 | - | Severini et al., 1996 | - |
| ARVD4 | AD | No | 2q32.1-32.3 | - | Rampazzo et al., 1997 | - |
| ARVD5 | AD | No | 3p25 | - | Ahmad et al., 1998 | - |
| ARVD6 | AD | No | 10p14-p12 | - | Li et al., 2000 | - |
| ARVD7 | AD | Yes | 10q22.3 | - | Melberg et al., 1999 | - |
| ARVD8 | AD | No | 6p24 | *Desmoplakin (DSP)* | Rampazzo et al., 2002 | Rampazzo et al., 2002 |
| ARVD9 | AD | No | 12p11 | *Plakophilin-2 (PKP-2)* | Gerull et al., 2004 | Gerull et al., 2004 |
| ARVD10 | AD | No | 18q12.1-12.2 | *Desmoglein-2 (DSG2)* | Pilichou et al., 2006 | Pilichou et al., 2006 |
| ARVD11 | AD | No | 18q12.1 | *Desmocollin-2 (DSC2)* | Syrris et al., 2006 | Syrris et al., 2006 |
| ARVD12 | AD | No | 17q21 | *Plakoglobin (PKGB)* | Asimaki et al., 2007 | Asimaki et al., 2007 |

**Table 2.2: Markers in the *ARVD5* haplotype.**

| Position along chr. 3 (Mb) Build 36.1 | *ARVD5* haplotype markers | |
|---|---|---|
| | Ahmad et al. markers | Young lab markers |
| 11.52 | *D3S1263* | - |
| 12.07 | *D3S 1259* | *D3S 1259* |
| 12.98 | *D3S 3610* | *D3S 3610* |
| 13.15 | - | *D3S 2403* |
| 13.63 | - | *D3S 1516* |
| 13.68 | - | *D3S 3608* |
| 13.85 | - | *D3S 2385* |
| 13.9 | - | *D3S 3602* |
| 13.92 | *D3S 1585* | *D3S 1585* |
| 14.34 | - | *D3S 1554* |
| - | *D3S 1255** | - |
| 14.62 | *D3S 3595*** | *D3S 3595* |
| 15.34 | *D3S 3613* | *D3S 3613* |
| 16.5 | *D3S 3473* | *D3S 3473* |
| 16.87 | *D3S 2338* | *D3S 2338* |
| 17.93 | - | *D3S 4547* |
| 19.07 | - | *D3S 3510* |
| 21.9 | *D3S 1293* | *D3S 1293* |
| 21.92 | - | *D3S 3038* |
| 22.91 | *D3S 3659* | *D3S 3659* |
| 23.91 | *D3S 3700* | - |
| 27.96 | *D3S 1266* | - |
| Total number | 13 | 18 |

* *D3S1255* was not on the Genethon map and was placed in the haplotype since it co-segregated with the haplotype. ***D3S3595* was positioned between *D3S3473* and *D3S2338* in Ahmad et al.

**Table 2.3: The 20 positional candidate genes for *ARVD5*.**

| Genes | Accession Number | MIM number | Strand | Genomic Position | | Exons |
|---|---|---|---|---|---|---|
| | | | | Start | End | |
| *IQSEC1* | NM_014869 | 610166 | - | 13003536 | 12917079 | 13 |
| *NUP210* | NM_024923 | 607703 | - | 13436809 | 13332737 | 40 |
| *HDAC11* | NM_024827 | 607226 | + | 13496824 | 13521834 | 10 |
| *FBLN2* | NM_001004019 | 135821 | + | 13565625 | 13654922 | 18 |
| *WNT7A* | NM_004625 | 601570 | - | 13896619 | 13835083 | 4 |
| *TPRXL* | AK092426 | 611167 | + | 13953902 | 14082480 | 3 |
| *CHCHD4* | NM_144636 | 611077 | - | 14141323 | 14128584 | 4 |
| *TMEM43* | NM_024334 | na | + | 14141546 | 14160180 | 12 |
| *XPC* | NM_004628 | 278720 | - | 14195143 | 14161651 | 16 |
| *LSM3* | NM_014463 | 607283 | + | 14195341 | 14214840 | 4 |
| *SLC6A6* | NM_003043 | 186854 | + | 14419110 | 14503973 | 15 |
| *GRIP2* | NM_001080423 | na | - | 14558592 | 14510177 | 25 |
| *C3orf19* | NM_016474 | na | + | 14668278 | 14689167 | 11 |
| *C3orf20* | NM_032137 | na | + | 14691658 | 14789544 | 17 |
| *FGD5* | NM_152536 | na | + | 14835810 | 14950899 | 20 |
| *NR2C2* | NM_003298 | 601426 | + | 14964240 | 15065782 | 15 |
| *MRPS25* | NM_022497 | na | - | 15081820 | 15065024 | 4 |
| *ZFYVE20* | NM_022340 | 609511 | - | 15115659 | 15086584 | 14 |
| *CAPN7* | NM_014296 | 606400 | + | 15222737 | 15269426 | 21 |
| *SH3BP5* | NM_004844 | 605612 | - | 15349108 | 15271250 | 9 |
| | | | | | Total | 275 |

**Table 2.4: Sequencing variants identified exclusively in clinically affected individuals and their segregation in Family 1139.** Eleven of the 18 variants, which were determined to be on the *ARVD5* ancestral haplotype (yellow), were not seen on any other haplotype segregating in Family 1139 and were subsequently sequenced in population controls (grey). The allele frequency of variants previously reported in the NCBI dbSNP was used when available. Rare variants (<1% of the alleles screened) are marked with an *.

| Gene name | Variants | Segregated on *ARVD5* haplotype | Classification | Population control screening | Allele frequency | Global ID | 710 | | 709 | | 714 | | 718 | | 708 | | 716 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Family # | 1139.1000 | | 1139.0000 | | 1139.A000 | | 1139.1008 | | 1139.0016 | | 1139.0004 | |
| | | | | | | Clinical affection status | primary | | primary | | unaffected | | unaffected | | unaffected | | unaffected | |
| | | | | | | Microsatellite *ARVD5* haplotype | carrier | | carrier | | recombinant carrier | | no | | recombinant carrier | | no | |
| | | | | | | Chromosome# | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| HDAC11 | c.369+18_369+19insG | yes | Non-coding | nd | nd | | ins | | ins | | ins | ins | | | | | | |
| TMEM43 | c 1073C>T | yes | Missense (S>L) | 0/322 | 0%* | | T | | T | | T | C | | | | | | |
| XPC | c.2823+684G>C | no | Non-coding | nd | nd | | G | | G | | G | C | | | | | | |
| SLC6A6 | c.1-27420G>A | yes | Non-coding | nd | nd | | A | | A | | A | G | | | | | | |
| SLC6A6 | c 599+370A>G | yes | Non-coding | dbSNP | 46% | | G | | G | | G | A | | | | | | |
| SLC6A6 | c 733-1226A>G | yes | Non-coding | 50/90 | 55 60% | | G | | G | | G | A | | | | | | |
| FGD5 | c.934G>A | yes | Missense (V>M) | 2/318 | 0.6%* | | A | | A | | A | G | | | | | | |
| FGD5 | c.2186+22G>A | no | Non-coding | nd | nd | | G | | G | | G | G | | | | | | |
| FGD5 | c.2187-82G>A | no | Non-coding | nd | nd | | G | | G | | G | G | | | | | | |
| FGD5 | c.2220G>T | no | Synonymous (L>L) | nd | nd | | G | | G | | G | G | | | | | | |
| FGD5 | c.2613+50C>T | no | Non-coding | nd | nd | | C | | C | | C | C | | | | | | |
| FGD5 | c 3085-74G>A | yes | Non-coding | 15/166 | 9 00% | | A | | A | | A | G | | | | | | |
| NR2C2 | c.855+70G>A | yes | Non-coding | 8/88 | 9 10% | | A | | A | | A | G | | | | | | |
| NR2C2 | c 1848+365T>A | yes | Non-coding | 16/90 | 17 80% | | A | | A | | A | A | | | | | | |
| NR2C2 | c.1848+2965_1848+2966insGATA | yes | Non-coding | 23/126 | 18 30% | | ins | | ins | | ins | wt | | | | | | |
| MRPS25 | c.522+1059G>A | yes | Non-coding | 0/138 | 0%* | | A | | A | | A | G | | | | | | |
| CAPN7 | c 1289+68C>T | yes | Non-coding | 1/162 | 0.01%* | | T | | T | | T | C | | | | | | |
| CAPN7 | c 1430-28T>C | yes | Non-coding | 0/160 | 0%* | | C | | C | | C | T | | | | | | |

**nd means 'no data'**

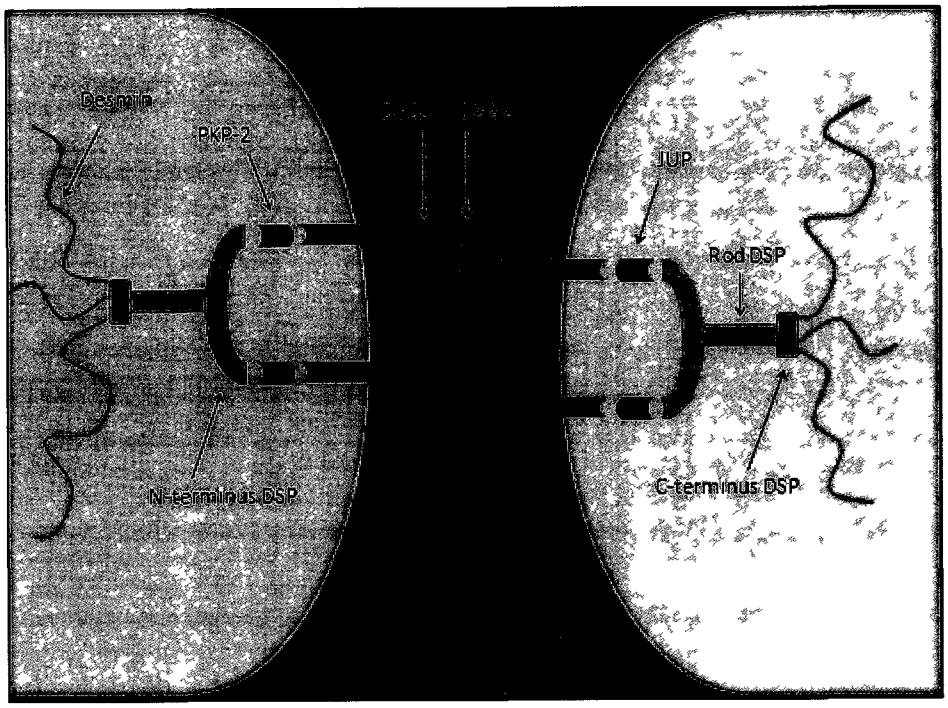



**Figure 2.1: Schematic of the desmosomal structure.**

## Family 64

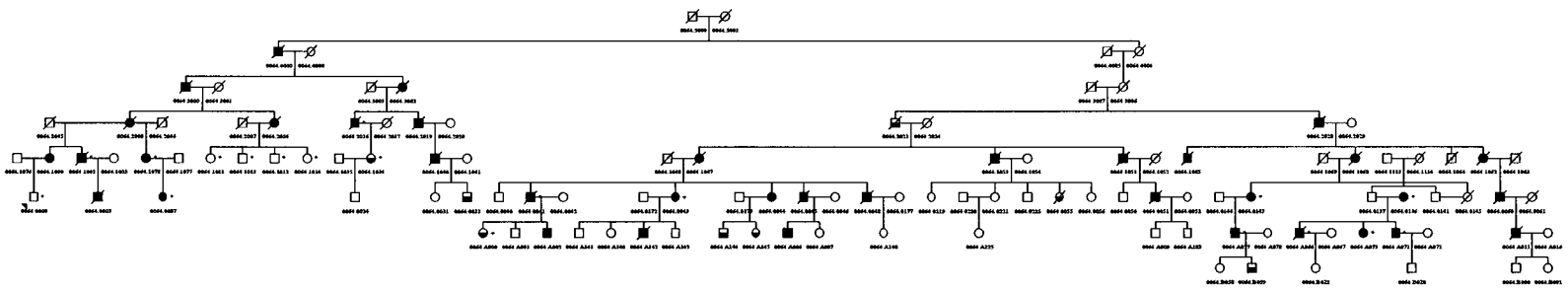


**Figure 2.2: Core Pedigree of Family 64, the original *ARVD5* family.** Abbreviated pedigree structure used in the study. ○=clinically unaffected female; □=clinically unaffected male; ●=1° clinical affection status; ◒=2° clinical affection status; and ⊙=obligate carrier.

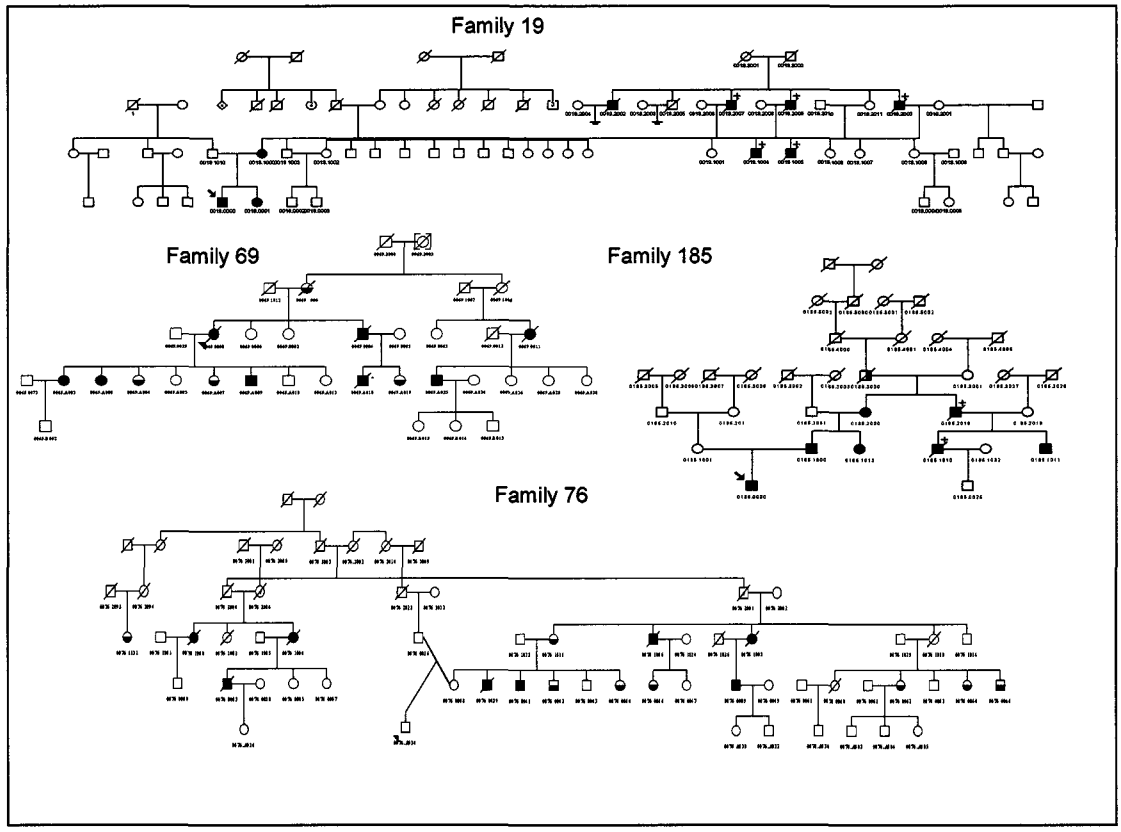



**Figure 2.3: Pedigrees of families 19, 69, 76 and 185.** The symbols in each pedigree represent: ○=clinically unaffected female; □=clinically unaffected male; ●=1° clinical affection status; ◒=2° clinical affection status; and ⊙=obligate carrier.

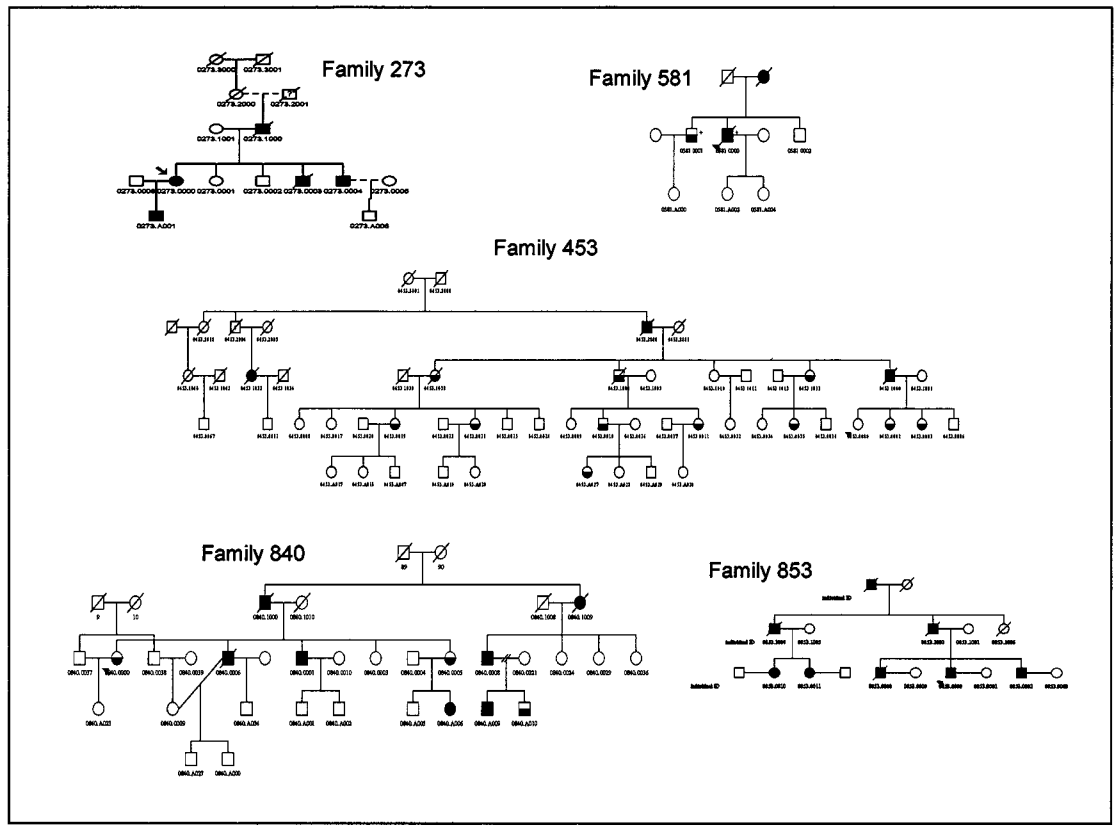



**Figure 2.4: Pedigrees of families 273, 453, 581, 840 and 853.** The symbols in each pedigree represent: ○=clinically unaffected female; □=clinically unaffected male; ●=1° clinical affection status; ◒=2° clinical affection status; and ⊙=obligate carrier.

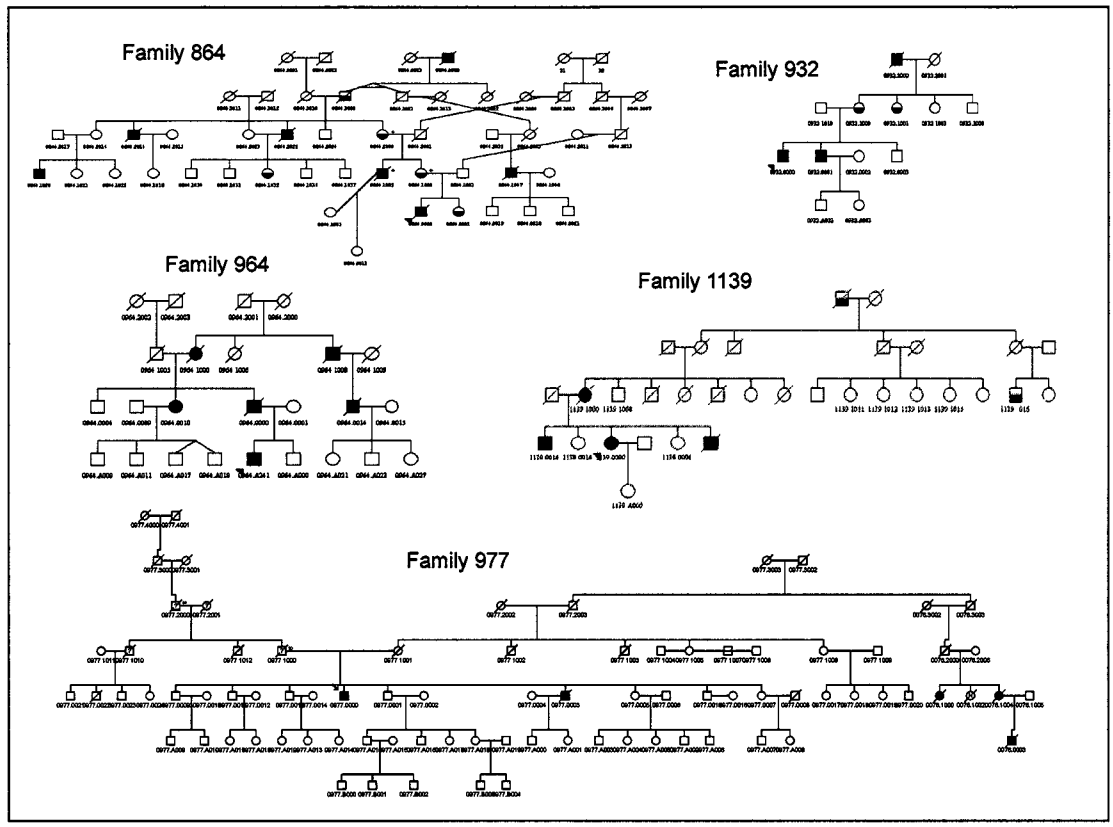



**Figure 2.5: Pedigrees of families 864, 932, 964, 977 and 1139.** The symbols in each pedigree represent: ○=clinically unaffected female; □=clinically unaffected male; ●=1° clinical affection status; ◒=2° clinical affection status; and ⊙=obligate carrier.

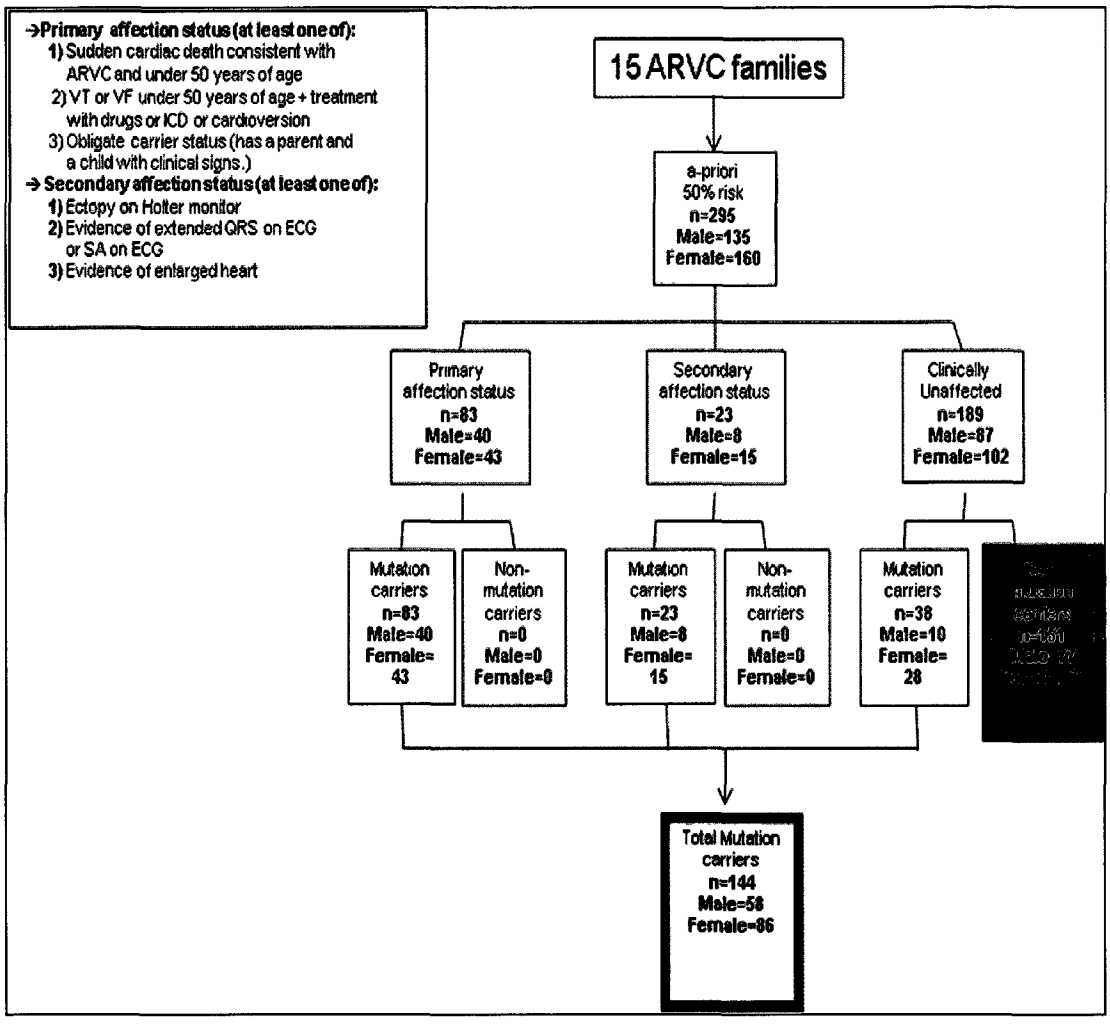
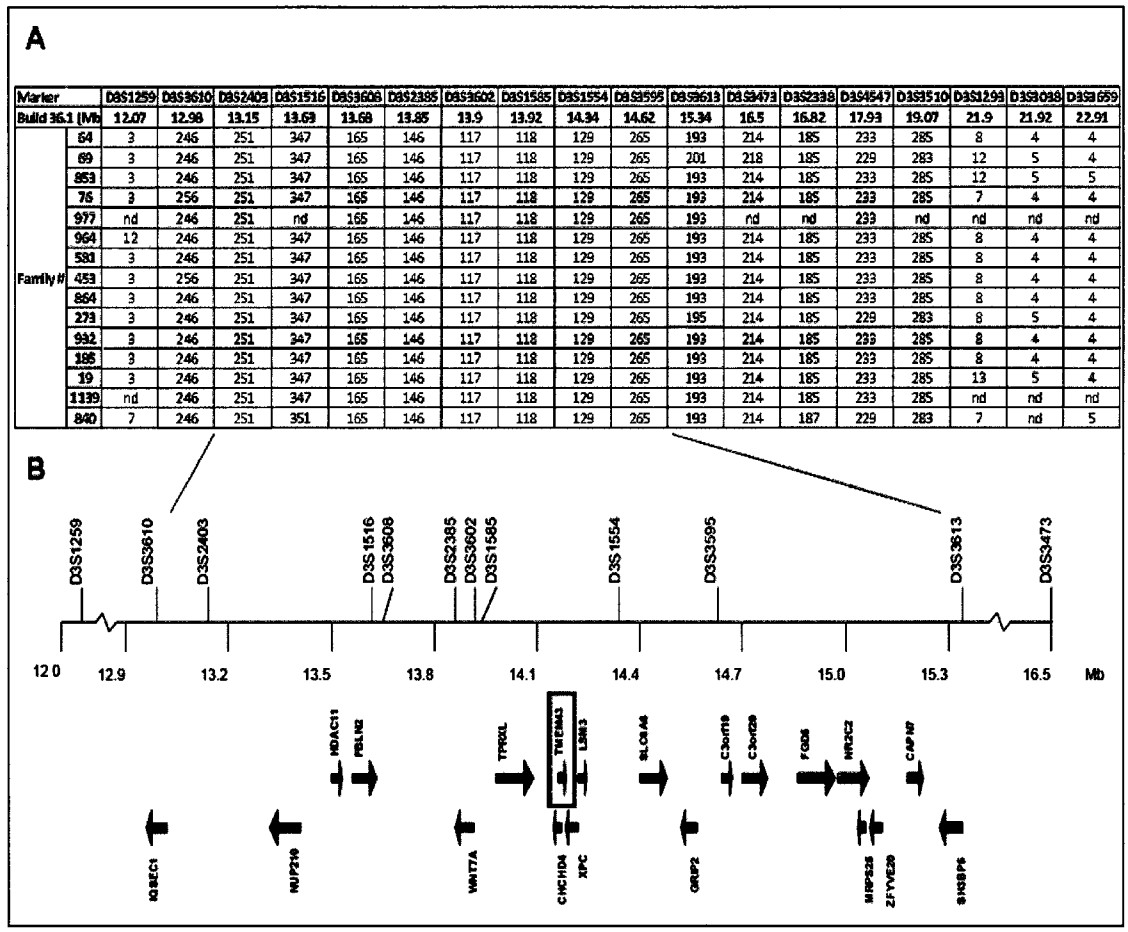



**Figure 2.6: Affection status and mutation screening of all individuals born at a-priori 50% risk with DNA collected.**




| Marker | | D3S1259 | D3S3610 | D3S2403 | D3S1516 | D3S3608 | D3S2385 | D3S3602 | D3S1585 | D3S1554 | D3S3595 | D3S3613 | D3S3473 | D3S2338 | D3S4547 | D3S3510 | D3S1293 | D3S3038 | D3S3659 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Build 36.1 (Mb | | 12.07 | 12.98 | 13.15 | 13.69 | 13.68 | 13.85 | 13.9 | 13.92 | 14.34 | 14.62 | 15.34 | 16.5 | 16.82 | 17.93 | 19.07 | 21.9 | 21.92 | 22.91 |
| Family # | 64 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 8 | 4 | 4 |
| | 69 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 201 | 218 | 185 | 229 | 283 | 12 | 5 | 4 |
| | 853 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 12 | 5 | 5 |
| | 76 | 3 | 256 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 7 | 4 | 4 |
| | 977 | nd | 246 | 251 | nd | 165 | 146 | 117 | 118 | 129 | 265 | 193 | nd | nd | 233 | nd | nd | nd | nd |
| | 964 | 12 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 8 | 4 | 4 |
| | 581 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 8 | 4 | 4 |
| | 453 | 3 | 256 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 8 | 4 | 4 |
| | 864 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 8 | 4 | 4 |
| | 273 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 195 | 214 | 185 | 229 | 283 | 8 | 5 | 4 |
| | 932 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 8 | 4 | 4 |
| | 186 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 8 | 4 | 4 |
| | 19 | 3 | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | 13 | 5 | 4 |
| | 1139 | nd | 246 | 251 | 347 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 185 | 233 | 285 | nd | nd | nd |
| | 840 | 7 | 246 | 251 | 351 | 165 | 146 | 117 | 118 | 129 | 265 | 193 | 214 | 187 | 229 | 283 | 7 | nd | 5 |

**Figure 2.7: Physical map of the *ARVD5* critical region.** (A) Summary recombinant *ARVD5* haplotypes (yellow) identified across the 15 ARVC families using microsatellite markers (alleles are either numbered (1-9) or given in base pairs). (B) The physical map of the *ARVD5* critical region. Physical distances were captured from the March 2006 freeze of the UCSC Genome Browser. Arrows show the direction of transcription of each annotated gene.

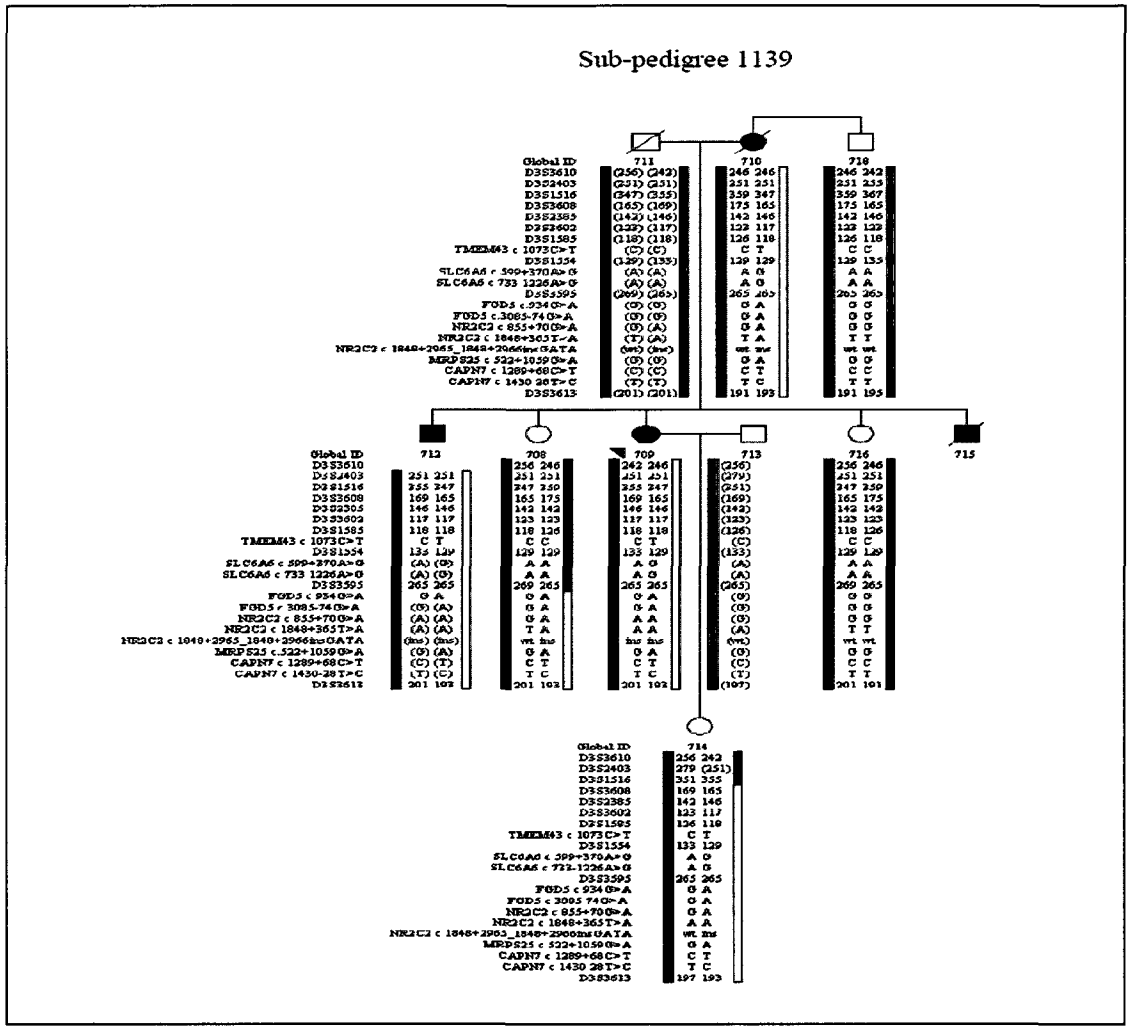



**Figure 2.8: Segregation analysis in sub-pedigree 1139.** Of the 18 variants found exclusively in clinically affected subjects on the mutation screening panel, only 11 were found to reside exclusively on the *ARVD5* ancestral haplotype (yellow). Note that a clinically unaffected subject (Global ID 708 – also referred to as 1139.0016) inherited a recombinant *ARVD5* haplotype from her clinically affected mother (Global ID 710). Alleles in brackets have been inferred. Abbreviated pedigree structure used in the study: ○=clinically unaffected female; □=clinically unaffected male; ●=1° clinical affection status; ◒=2° clinical affection status; and ⊙=obligate carrier.

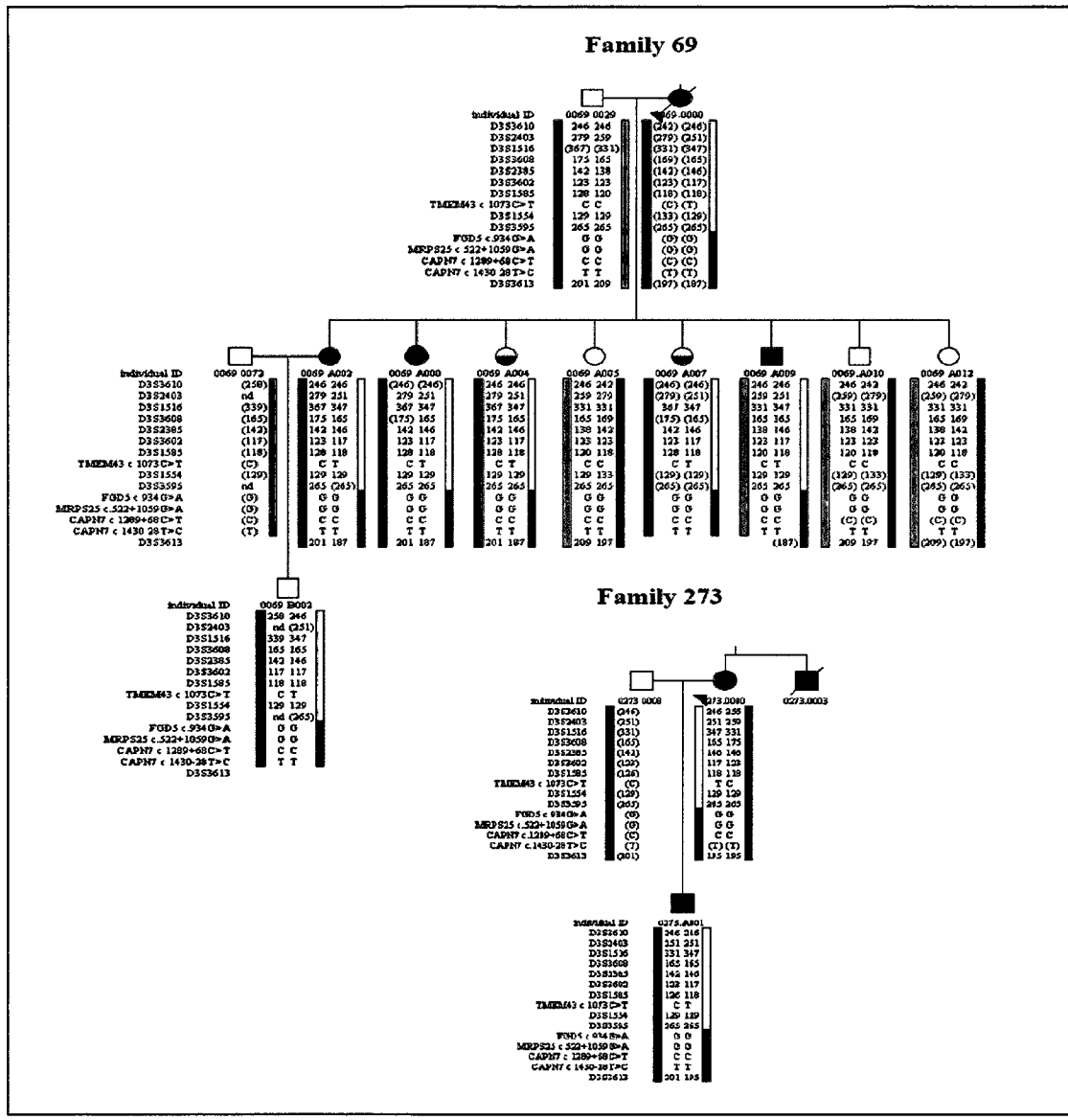



**Figure 2.9: Segregation analysis in families 69 and 273 (partial pedigrees).** Clinically affected subjects only have one of the five rare variants due to a historical recombination event on the *ARVD5* haplotype (yellow). Alleles in brackets have been inferred. Abbreviated pedigree structure used in the study: ○=clinically unaffected female; □=clinically unaffected male; ●=1° clinical affection status; ◒=2° clinical affection status; and ⊙=obligate carrier.

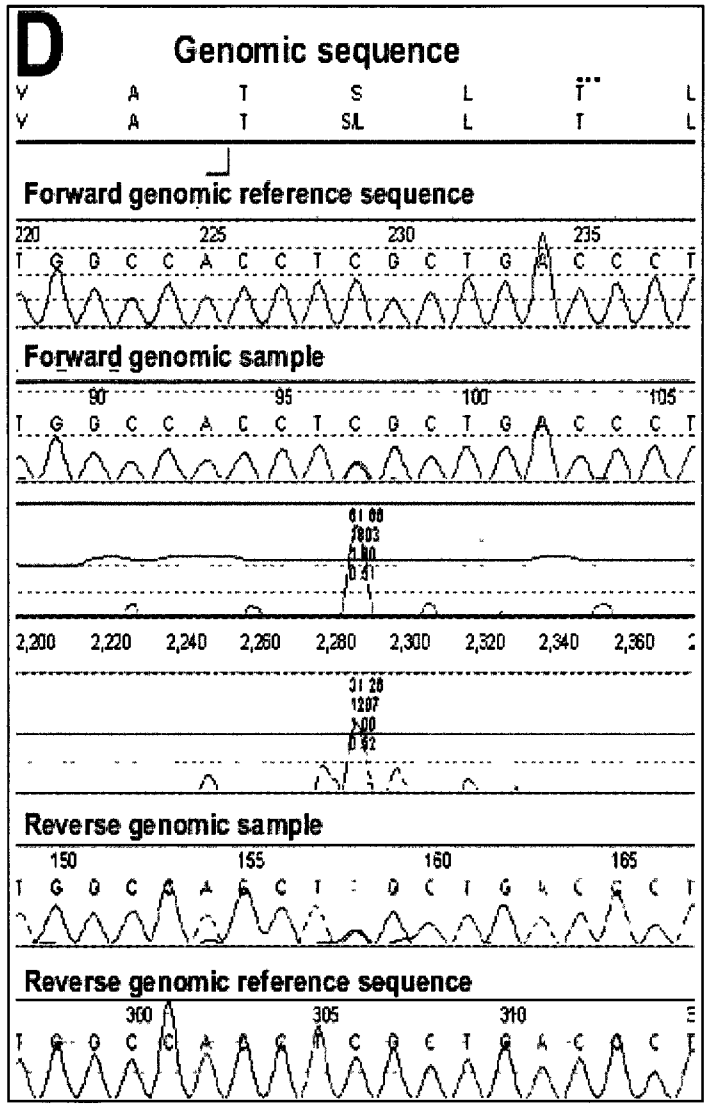



**Figure 2.10: *TMEM43* mutation sequence trace.** Forward and reverse sequencing traces showing the *TMEM43* c.1073 C>T mutation of an affected individual's genomic DNA. The amino acid translation (top) shows the S358L amino acid substitution.

| Markers | 3p25 location (bp) | Affected haplotype | Unaffected recombinants | |
|---|---|---|---|---|
| | | | R1139.0016 | R0964.0004 |
| D3S2403 | 13147397 - 13147709 | 251 | 251 | 251 |
| D3S1516 | 13628628 - 13629103 | 347 | 359 | 347 |
| D3S3608 | 13670236 - 13679541 | 165 | 175 | 165 |
| D3S2385 | 13853945 - 13854287 | 146 | 142 | 146 |
| D3S3602 | 13900968 - 13901215 | 113 | 123 | 113 |
| D3S1585 | 13916682 - 13916861 | 118 | 126 | 118 |
| **TMEM43** | **14158166** | **T** | **C** | **C** |
| D3S1554 | 14342778 - 14343106 | 129 | 129 | 133 |
| D3S3595 | 14617332 - 14617642 | 265 | 265 | 265 |
| D3S3613 | 15271250 - 15357905 | 193 | 193 | 187 |

**Figure 2.11: Clinically unaffected individuals that lack the *TMEM43* mutation but have distal portions of the *ARVD* haplotype.** R1139.0016 has the centromeric portion where 4 of the 5 rare alleles lie, and R0964.0004 has the telomeric portion.

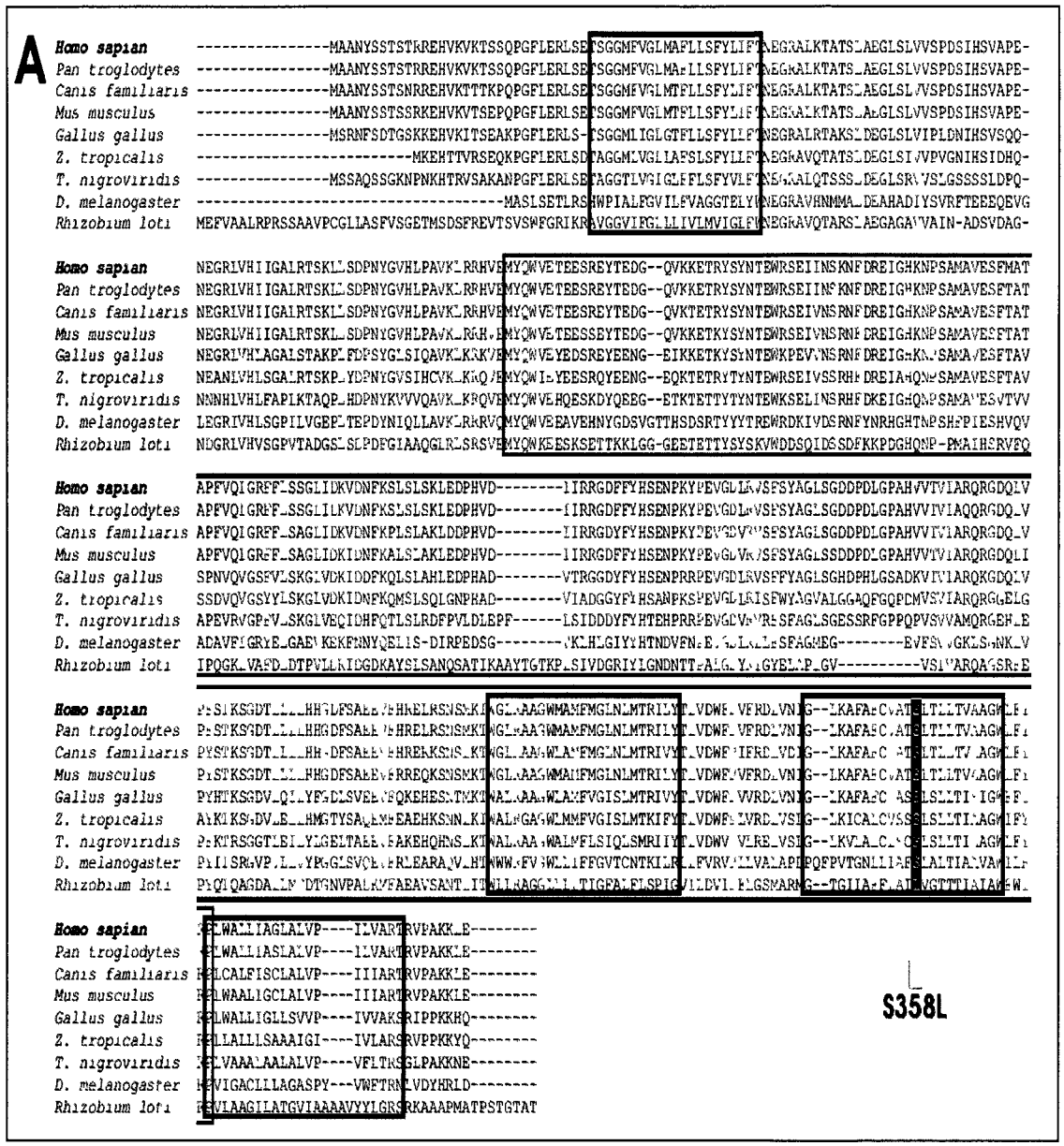



**Figure 2.12: Multiple alignment of the *TMEM43* gene across eight eukaryotic and prokaryotic species.** (A) Clustal W align (70) was used to align orthologues from *Homo sapiens* (NP_077310), *Pan trogLODytes* (XP_516299), *Canis familiaris* (XP_541751), *Mus musculus* (NP_083042), *Gallus gallus* (XP_414378), *Zenopus tropicalis* (UP10004D5297), *Tetraodon nigroviridis* (Q4RXL8), *Drosophila melanogaster* (NP_64162) and *Rhizobium loti* (Q98HF3) (179). The blue box outlines the DUF1625 domain and the black boxes outline predicted transmembrane domains. Completely conserved residues are red, strongly similar residues are green and weakly similar residues are blue. The S358L mutation is marked by a yellow arrow.

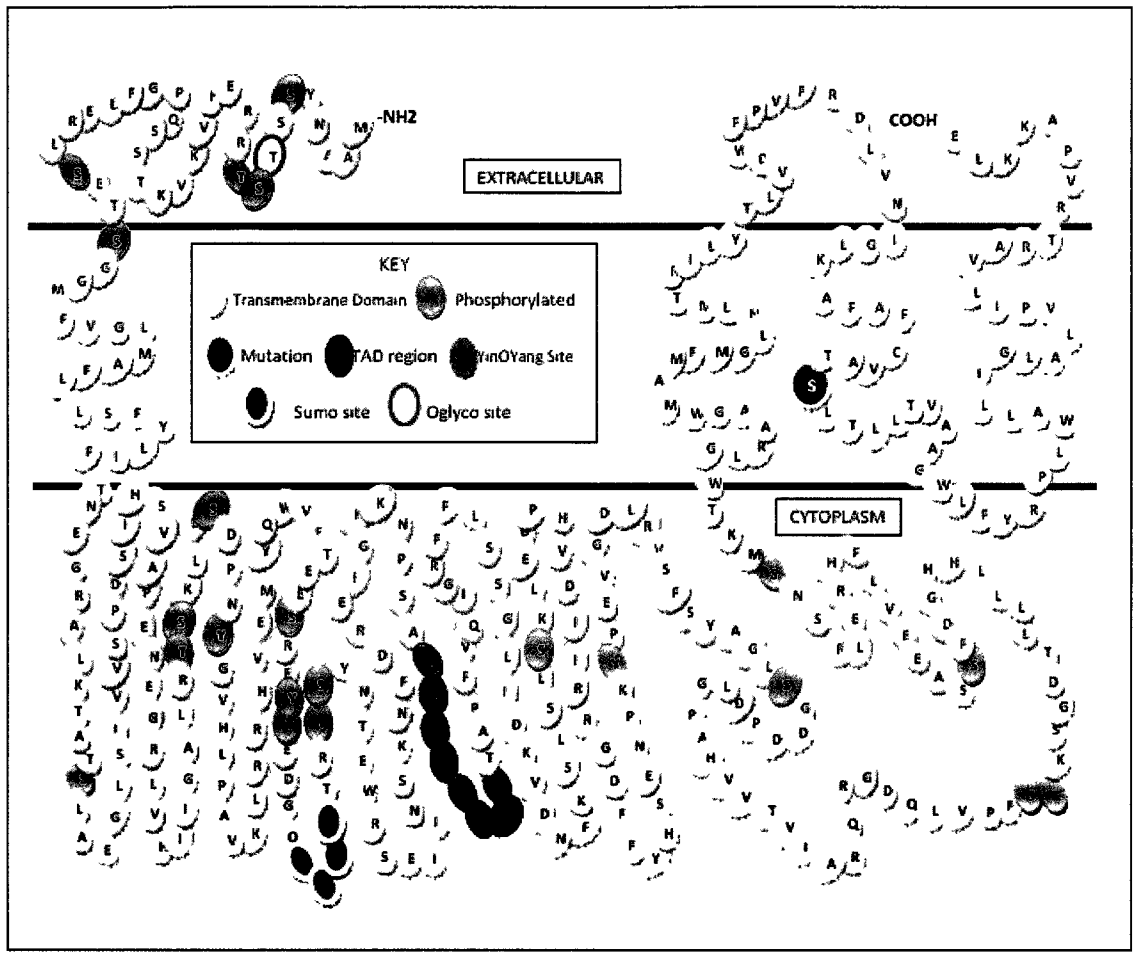



**Figure 2.13: Predicted topography of the TMEM43 protein.** Indicated are transmembrane domains (beige), phosphorylation sites (green), a transactivation domain (red), YingOYang sites (orange), a SUMO attachment site (purple), and an O-glycosylation site (blue open). The extracellular and cytoplasmic regions may be reversed: there is evidence supporting either orientation.

# Chapter 3: Non-syndromic sensorineural hearing loss in a large extended Newfoundland family maps to Xp21

## 3.1 Introduction

### 3.1.1 Sound and measurement of hearing

Sound is perceived through the detection of sound waves, which are pressure waves that propagate through the air. All sound waves have particular frequencies; frequency is defined as the measurement of the number of times a repeated event occurs per unit of time. The unit of frequency measurement is the hertz (Hz), which detects the number of cycles per second. Hearing is measured in decibels (dB) across all frequencies. The normal threshold for hearing is 0 dB, which is when normal young adults perceive a tone burst 50% of the time at a certain frequency. Hearing is considered normal if an individual's thresholds are within 15 dB of normal threshold. An audiogram is a graph that depicts the ability to hear sounds at different frequencies (Figure 3.1), and it is usually plotted as hearing loss in dB as a function of frequency (199).

### 3.1.2 Deafness - a common disorder

Deafness affects 6-8% of the population in developed nations and is the most prevalent sensorineural disorder (200, 201). It is also the most common birth defect (200, 201). Approximately, 1 in 1000 newborns is profoundly deaf, another 1 in 300 has a congenital hearing loss to a lesser degree, and an additional 1 in 1000 become profoundly hearing impaired before adulthood (200-202).

### 3.1.3 Classification of hearing loss

There are several ways to classify hearing loss. Firstly, upon analysis of an audiogram, the severity (and the corresponding frequency) of the hearing loss can be determined. The World Health Organization (WHO) divides the severity of hearing impairment into several grades: mild or slight (26-40 dB), moderate (41-60 dB), severe (61-80 dB) and profound (81 dB or greater); and the frequency of hearing loss is designated as low (<500 Hz), middle (501-2000 Hz) and high (>2000 Hz) (199). Secondly, one can classify hearing loss based on age of onset. If hearing loss develops before the ability to speak it is known as pre-lingual, which may or may not be congenital. If it occurs after language development it is known as post-lingual. A third way to classify hearing loss is based on its stability over time. If the hearing loss worsens over time it is known as progressive and if it remains stable over time it is known as non-progressive, however hearing loss can fluctuate over time as well, meaning that after successive testing hearing loss can appear better or worse (199).

Hearing loss can also be classified based on etiology. It can be due to environmental factors, genetic defects, or a combination of the two (mixed). It is estimated that at least 50% of pre-lingual hearing loss is genetic, 25% is environmental (infections, trauma, etc.) and the remaining 25% is unknown, but possibly and probably genetic (200, 201). Focusing on the physiological malformation, hearing loss can be defined as conductive (due to external ear anomalies or abnormalities of the ossicles in the middle ear), sensorineural (due to inner ear malfunctions including defected hair cells or cochlea),

central (due to defects in the VIIIth nerve, brain stem or cerebral cortex), and mixed (a combination of conductive and sensorineural hearing loss) (200, 201).

### 3.1.4 Inherited hearing loss

Inherited hearing loss, in most cases described, is monogenic (200). It can be inherited alone with no other manifestations, which is known as non-syndromic, or it can be inherited as a syndrome with additional manifestations. The inheritance pattern can be autosomal recessive, autosomal dominant, X-linked, Y-linked or mitochondrial. In a 2009 review by Hilgert et al., it was estimated that 70% of inherited deafness is non-syndromic, with 80% being recessive (DFNB), 15-20% being dominant (DFNA), ~1% being X-linked (DFNX) and at least 1% being Y-linked or mitochondrial. Non-syndromic hearing loss is generally sensorineural. Autosomal dominant deafness is generally post-lingual and progressive, and recessive deafness is generally pre-lingual (70-85%) (201).

### 3.1.5 Heterogeneity of deafness

#### (i) Genetic heterogeneity

Deafness is very genetically heterogeneous with approximately 1% of all human genes involved in the hearing process; proteins comprising ion channels, the extracellular matrix, the cytoskeleton, and transcription factors all interact for proper auditory function (203). According to the Hereditary Hearing Loss Homepage (http://webh01.ua.ac.be/hhh/), there have been over 120 non-syndromic deafness loci

identified for autosomal dominant (*DFNA*) and autosomal recessive (*DFNB*) deafness combined (38). The number identifying the locus, for example *DFNA1*, represents the chronological order in which the loci were discovered (38). Of the identified loci, over 45 causative genes have been identified (200). The *gap junction protein beta 2* gene (*GJB2*) [MIM #121011], which encodes the protein connexin 26, is the most common cause of recessive deafness in most populations and in some populations is the cause of 50% of deafness cases (204). There are over 220 *GJB2* mutations that have been reported (The Connexin-Deafness Homepage: http://davinci.crg.es/deafness/index.php) (200), and the most commonly reported mutation in Caucasians with European ancestry is c.35delG, which may explain up to 70% of all cases (205).

### (ii) Clinical Heterogeneity

Clinical heterogeneity is defined as the development of different clinical phenotypes from different mutations within the same gene. Different mutations in many non-syndromic deafness genes cause syndromic forms of deafness as well, and a good example of this is Usher syndrome [MIM #276900], which is associated with deafness and blindness. This syndrome normally presents with pre-lingual sensorineural hearing loss, with or without vestibular function, and a later onset of retinitis pigmentosa (206). There are a number of different Usher syndrome phenotypes that differ based on vestibular function and retinitis pigmentosa age of onset (207, 208). Many non-syndromic deafness loci and Usher loci have the same causative gene (209), for example, *DFNB12* [MIM #601386] and *Ush1D*

[MIM #601067] are caused by *cadherin 23* (*CDH23*) [MIM #605516] mutations (210, 211).

### 3.1.6 Hereditary deafness gene discovery

#### (i) Mapping hearing loss loci

Due to the genetic heterogeneity of deafness, gene discovery efforts have been most successful in isolated populations by studying large consanguineous families. Approximately 80% of non-syndromic deafness cases are recessive, and because of the typical small pedigree sizes that are ascertained in most first-world urban areas, linkage has been more difficult in those diverse populations (201). Also, since there are over 100 known loci for non-syndromic deafness the likeliness that two random, hereditary deafness families from first-world urban areas share the same causative gene is low. Therefore grouping multiple deafness families for linkage analysis, which is a traditional method for gene identification and was used for the identification of *BRCA1*, is generally not very useful (201, 212). Considering that deafness is a common disorder and very genetically heterogeneous as well as easily influenced by environmental factors, phenocopies can be present even within one hereditary deafness family. This can also hinder gene discovery efforts, therefore it is best if the clinical characteristics are well defined.

In populations such as Tunisia and Pakistan, consanguinity is common and preferred (213, 214). Religious and cultural beliefs, as well as social and economic considerations

all influence attitudes toward consanguinity. For example, in certain populations consanguineous marriages are believed to strengthen family ties and reduce financial problems (215). This belief system results in married couples who share many of the same ancestral loci, and consequently recessive conditions such as deafness have a higher prevalence. Autozygosity mapping, which is the mapping of recessive alleles identical by decent in consanguineous families, has been performed in these populations to find deafness loci. For example, *DFNB8* [MIM #601072] and *DFNB10* [MIM #605316] were independently identified through linkage analysis in two large consanguineous Pakistan families (216, 217). These two loci are now known to overlap and have the same causative gene, *transmembrane protease serine 3* (*TMPRSS3*) [MIM #605511] (218).

### (ii) Identifying causal genes within hearing loss loci

Once deafness loci have been mapped, there are several ways to prioritize the selection of positional candidate genes for screening. One method of selection is based on gene expression. Immense effort has been made to identify genes selectively expressed in the cochlea and to determine whether these genes map to known deafness loci (219). A mouse cochlea cDNA library was established and it has aided in the gene discovery of several deafness genes. This approach was used to identify the gene *otoferlin* (*OTOF*) [MIM #603681] within *DFNB9* [MIM #601071] (220).

The use of animal models, particularly the mouse model, has also been a successful way to identify deafness genes. When an unidentified deafness gene has been mapped to a

locus and a mouse model with a mutated gene in the corresponding mouse locus has been previously identified, that gene would be a key candidate for screening. This approach was used to identify *MYO15A* [MIM #602666] as the *DFNB3* [MIM #600316] gene (221-224). In addition to aiding in gene discovery, animal models for hereditary deafness also enable the determination of the phenotype's anatomical, biochemical, and cellular features, which otherwise could not be studied *in vivo*. The shaker-2 mouse model was used for morphology studies of mutant-*MYO15A* deafness, which showed abnormally short stereocilia bundles (225).

### 3.1.7 Hearing loss in the population of Newfoundland

Hearing impairment appears to be prevalent in the Newfoundland population, particularly on the south coast of the province, and three large south coast families have previously been studied to identify causal deafness genes. Positional cloning in a Newfoundland (south coast) family with autosomal dominant low frequency hearing loss identified *wolframin* (*WFSI*) [MIM #606201], the gene previously known to cause Wolfram syndrome [MIM #222300], as the causative *DFNA38* [MIM #600965] deafness gene (98). *WFS1* is the causative gene for low-frequency hearing loss at *DFNA6/14/38*, which were separately identified loci that were first thought to be non-overlapping loci at 4p16.3 (98, 226-228). There are now at least 12 additional families with autosomal dominant low frequency hearing loss that have *WFS1* causal variants, all of which are missense mutations (228-233). In 2004, a second south coast family was studied whose affected individuals presented with an autosomal recessive, congenital, non-syndromic hearing

loss (234). Using a candidate gene approach, *TMPRSS3* was determined to be the causative gene. Two mutations segregated in the family, one in exon 4, c.207delC, and another in exon 8, c.782+3delGAG. Affected family members were either homozygous for the exon 4 mutation or compound heterozygous for the exon 4 and 8 mutations (234). Thirdly, a family with autosomal recessive, congenital, profound, non-syndromic hearing loss was recently determined to have a homozygous mutation in *PCDH15* (235). After performing a genome wide linkage scan, this family was determined to be linked to the previously mapped *DFNB23* locus. In 2003, Ahmed et al. took a candidate gene approach and screened 400 families with recessive pre-lingual hearing loss for mutations in the *protocadherin 15*, *PCDH15*, gene (236). Two different mutations were identified, c.400C>T (p.R134G) and c.785G>A (p.G262D), solving two of the 400 families. Therefore, once linkage to the same locus was determined in the Newfoundland family, *PCDH15* was sequenced and a missense mutation, c.1583T>A, in exon 13, was identified.

Due to the success in identifying hearing loss genes in Newfoundland, additional probands with hereditary hearing loss were ascertained by the Young laboratory. Once ascertained, probands from each family were first screened for the known Newfoundland mutations (*WFS1* c.2146G>A, *TMPRSS3* c.207delC, *TMPRSS3* c.782+3delGAG, and *PCDH15* c.1583T>A) and for variants in *GJB2* (connexin 26) and *GJB6* (connexin 30). The connexin genes were added to the screening panel because variants in those genes are major contributors to non-syndromic deafness (204). After the initial screen, only six of

the 47 probands participating in the study were solved. All six probands were determined to have variants in *GJB2* or *GJB6* (Young, unpublished data). The remaining 41 families are considered to have a potential for gene discovery. One extended family was the only family to have an apparent X-linked mode of inheritance (no male-to-male transmission and variable expression in females). This family has major potential for gene discovery and will be the focus of this chapter.

### 3.1.8 X-linked non-syndromic deafness

X-linked non-syndromic deafness is relatively rare, estimated to contribute to approximately 1% of all non-syndromic deafness cases (66). There is however an excess of of males among the deaf (237) and X-linked deafness can explain this excess; it is estimated that approximately 5% of pre-lingual male deafness is X-linked (237-239). X-linked disorders are caused by mutations in genes on the X chromosome, one of the two sex chromosomes. Females are XX and males are XY. Assuming no sex anomalies, a female will transmit one of her X chromosomes to her children and a male will transmit an X or a Y (50% chance of either case). Therefore, the father's sex chromosome determines the sex of a child and an X-linked trait is not transmitted from male to male. A father with X-linked hereditary hearing loss will not pass the mutation to his son(s) but will pass it to all of his daughters. A female who is a carrier has a 50% chance of transmitting the deafness-causing mutation with each pregnancy. Sons who inherit the mutation from their mother will be deaf, and mothers who have a son with X-linked hearing loss are considered obligate heterozygotes (unless one can prove that the mutation

is *de novo*) (240). Daughters who inherit the mutation either from their mother or father are carriers and are likely to have normal hearing or a less severe hearing loss (241).

### 3.1.9 Variable expression in X-linked diseases

Variable expression of X-linked traits in heterozygous females is common and may be due to the specific mutation itself or the mechanism of X chromosome inactivation, which operates in females to balance gene dosage in males and females (241). X inactivation is the phenomenon in female mammals by which one X chromosome, either the maternally or paternally derived X, is randomly inactivated in embryonic cells. Once a female diploid cell has an X chromosome inactivated, the same X is inactivated in all descendent cells (240). Dobyns et al., 2006, recently suggested to refer to X-linked traits as solely 'X-linked' instead of the traditional terms 'dominant' and 'recessive' because of the frequent occurrence of variable expression across many phenotypes in female carriers. An X-linked trait was once considered dominantly inherited when the daughter of an affected father was also affected, and was considered recessive when the disease was passed to an affected son from an unaffected mother (in this case the mother was considered an obligate carrier) (241).

### 3.1.10 X-linked non-syndromic deafness loci

To date, eight non-syndromic deafness loci have been designated as X-linked (*DFN1-8*) (Table 3.1) (38). However, *DFN1* is no longer considered a non-syndromic X-linked deafness locus. The large Norwegian family originally linked to *DFN1* (242) has since

been shown to have visual disability, dystonia, fractures and mental deficiency, in addition to hearing loss (243). *DFN5* and *DFN7* have been withdrawn from the list and the *DFN8* locus has been reserved but the map location not released. Therefore, only four mapped *DFN* loci exist, *DFN2*, *DFN3*, *DFN4* and *DFN6*. The causative genes for *DFN2* (244) and *DFN3* (245) have been identified. Recently, it has been suggested to designate non-syndromic X-linked loci as DFNX (66), therefore the current designation in OMIM is *DFNX1* (*DFN2*), *DFNX2* (*DFN3*), *DFNX3* (*DFN4*) and *DFNX4* (*DFN6*) (Table 3.1) (160).

X-linked deafness was first reported in the 20$^{th}$ century with at least seven different families described between 1930 and 1970 (239). In the late 1980s, one of the first-ever linkage studies was performed on X-linked, non-syndromic, deafness families from which one large locus on Xq13-q21 was identified (246-248). Cytogenetically detectable deletions were observed across that linked region in some male patients, however there were clear audiological differences between the patients with deletions and the patients without (248-250). This suggested that maybe two different deafness loci were in that chromosomal region, hence they were given distinct designations of *DFNX1 (DFN2)* and *DFNX2 (DFN3)*.

### (i) *DFNX1* [MIM #304500]

*DFNX1* was officially mapped in 1996 using a British-American family with a congenital, profound, sensorineural, X-linked hearing loss, originally described by

Reardon et al. (248, 251). This 4 generation family had seven affected males and 14 female carriers. Eight of those females had mild-moderate hearing loss upon audiological testing, the remainder either had a normal audiogram or their hearing status was unknown. After linkage analysis hearing loss in that family mapped between *DXS990* and *DXS1001* (Xq21.32-Xq24, a 26.8 Mb region) with a maximum two-point LOD score of 2.91 ($0=\theta$) at dinucleotide repeats at *COL4A5* and *DXS1106*. Two additional families have since been linked to the same locus (252, 253). Manolis et al. described a large American family with a profound, sensorineural, X-linked hearing loss, however the deafness was post-lingual (252). Cui et al. described a Chinese family with congenital, profound sensorineural hearing loss (253). The female carriers in these additional families had a mild/moderate hearing loss as well. A typical affected male had a symmetrical hearing loss around 100 dB and females who have a hearing defect detected had a milder loss between 10 to 60 dB over all frequencies (252, 253). The *DFNX1* gene has been recently reported (244). A second Chinese family was determined to be linked to the *DFNX1* locus, and mutation screening identified a missense mutation in the *PRPS1* gene [MIM #311850], which encodes phosphoribosyl pyrophosphate (PRPP) synthetase 1. Screening the three previously identified *DFNX1*-linked families revealed missense mutations as well, all of which were determined to be loss of function (244).

### (ii) *DFNX2* [MIM #304400]

Linkage analysis for the *DFNX2* locus was first performed in a large Dutch kindred with X-linked, progressive, mixed deafness with perilymphatic gusher during stapes surgery

(246). The gene for *DFNX2* was localized to a 500 kb region on Xq21 and was identified as *POU3F4* [MIM #300039], using five unrelated families that mapped to *DFNX2* (245). *DFNX2* is the most common cause of X-linked non-syndromic deafness, representing 50% of all cases, and at least 50 families have been reported (66, 245, 254). The phenotype is described as a profound, sensorineural deafness with or without a conductive component associated with a unique developmental abnormality of the ear (example: perilymphatic gusher after stapedectomy, and a dilated internal auditory meatus). It is also known as deafness with stapes fixation syndrome, perilymphatic gusher-deafness syndrome or Nance deafness (160, 255). In affected males the hearing loss is congenital and rapidly progresses to severe hearing loss of all tones in the first decade. Female carriers may show a slight hearing loss with or without the perilymphatic gusher (255, 256).

*POU3F4* (also referred to as *BRAIN4*) encodes a POU-domain transcription factor and is part of a transcription factor family with a homologous DNA-binding (POU) domain. In addition to *POU3F4*, mutations of another POU domain transcription factor, *POU4F3* [MIM #602460], also cause a deafness phenotype (*DFNA15* [MIM #602459]). Mutations in *POU3F4* that have been implicated in hearing loss include point mutations and small deletions, partial or complete deletions of the *POU3F4* gene, and deletions and duplications of DNA proximal to but not including *POU3F4*, suggesting the presence of an important regulatory element that affects *POU3F4* function (249, 254, 256-260).

Interestingly, 50% of *POU3F4* deafness mutations are caused by deletions of the regulatory element (66).

### (iii) *DFNX3* [MIM #300030]

In 1991 Reardon et al. reported one family with X-linked congenital, profound, sensorineural hearing loss that did not link to Xq13-21.2 (248). Adult female carriers in this family had a mild-moderate high frequency hearing loss. It was later determined that the deafness in that family linked to a novel locus, Xp21.2 (261). The linked region included 8.1 Mb with 24 annotated genes, including *DMD*, the gene responsible for *Duchenne* muscular dystrophy (51, 52). The family members showed no clinical signs of muscular dystrophy (261) .

A second *DFNX3* family was identified in 1998 (262). Affected males were deaf at birth and female carriers had a stable moderate hearing loss that only affected the high frequencies (> 1000 Hz). It was initially thought that recombination events reduced the linked region to within the *DMD* gene; a 5′ crossover between *DXS1219* and intron 44 of *DMD*, and a 3′ crossover between *DXS1214* and *DXS985*, defined the boundaries. *DXS985* is not within the *DMD* locus, in fact it is approximately 560kb downstream and encompasses two additional genes, *FTHL17* (*ferritin heavy polypeptide like-15*) [MIM #300308] and *TAB3* (TAK1 binding protein 3) [MIM #300480] (51, 52). Also, the boundaries were determined using an 'unaffected' female, who could be a non-penetrant carrier, and an 'affected' female, who could be a phenocopy, especially due to the

variable expression in affected female phenotypes (262). It is possible that *DMD* is the causative gene, and there is a mouse model, mdx, with a stop codon in *dmd* reported to have auditory problems (263) but that data could not be validated (264).

### (iv) *DFNX4* [MIM #300066]

Del Castillo et al. linked hearing loss in a single Spanish family, with bilateral, sensorineural and progressive hearing loss to 12 Mb region on Xp22 (265). The age of onset in males was approximately five years, and the hearing loss initially affected only the high frequencies. The hearing loss later progressed to affect all frequencies and became severe-profound. Carrier females have a moderate hearing loss in the high frequency and the onset is generally in the fourth decade. No other *DFNX4* families have been reported, but expansion of the Spanish family later refined the disease linked region to 8.4 Mb between markers *DXS8022* and *DXS7105* (66).

## 3.1.11 Aims of this work

This chapter concentrated on the disease gene identification of the only identified hearing loss family with an apparent X-linked mode of inheritance in the Young laboratory. Haplotype analysis of the X chromosome was first carried out to identify chromosomal regions that were exclusively shared by affected individuals, followed by the screening of positional candidate genes.

## 3.2 Materials and Methods

### 3.2.1 Recruitment

Family 2024 (Figure 3.2) was recruited through the Provincial Medical Genetics Program of Eastern Health, as part of a Newfoundland population-based, hereditary deafness study. This project was approved by the Human Investigation Committee of Memorial University (#01.186) and the Research Proposal Approval Committee (RPAC) of Eastern Health, St. John's, NL, Canada. All recruited individuals that consented to the study had a blood sample taken for DNA isolation and cell-line establishment. Family 2024 is an extended 6 generation family (Figure 3.2) that was first recognized by Dr. Claire Neville Smith in the early 1970s as two separate families (families 24 and 25). The probands for family 24 and 25 were 2024.3019 and 2024.3000, respectively (Figure 3.2).

### 3.2.2 Clinical assessment

An individual was considered affected if he/she was reported to have hearing loss. Due to the historic nature of this pedigree most individuals were designated as having hearing loss by either Dr. Smith herself, or from a family history taken by Dr. Smith in the early 1970s; however, limited medical records were included in the archived family files. Deafness was only confirmed through medical records (audiology report or doctor's report) in 15 individuals. There were a total of 50 apparently affected individuals in the family and DNA was collected for seven of them (all of whom had the proper medical records available). An individual with known audiology had pure-tone audiometric evaluations to determine the hearing loss severity across all frequencies. CT (computed

tomography) scans were performed on two affected individuals, one male and one female, to search for inner ear abnormalities. Affected individuals were further grouped based on age of onset. Three groups were established; (1) age of onset under 10 years of age, (2) age of onset over the age of 10, and (3) age of onset unknown (Figure 3.2).

### 3.2.3 A scan of the X chromosome

A genotype scan of the X chromosome was performed on five individuals, four affected and one unaffected. The affected individuals included a father (2024.1000), his daughter (2024.0000), and his two grandchildren (2024.A001 and 2024.A002). The unaffected individual was the father of the two grandchildren (2024.0002) who married into the family (spouse 2024.0000) (Figure 3.2). Microsatellite markers from Panel 28 of the ABI PRISM Linkage Mapping Set v2.5-MD10 kit were used for the scan. Panel 28 consisted of 18 microsatellite markers spaced approximately 10 cM across the X chromosome (Appendix 6). Markers were labeled with one of the four fluorescent dyes; 6-FAM (blue), VIC (green), NED (yellow) or PET (Red). A 1X PCR mix had a final volume of 7.5 µl, consisting of 0.75 µl of 10X PCR Buffer, 0.75 µl of dNTPs (250 mM), 0.75 µl of $MgCl_2$ (25 mM), 3.49 µl of $H_2O$, 0.06 µl of Taq DNA polymerase (5 Units/µl), 0.5 µl of primer mix, and 1.2 µl of 25 ng/ul genomic DNA. Amplification was carried out by an initial 12 minute 95°C hold, followed by 10 cycles of a 15 second 94°C denaturing period, a 15 second 55°C annealing period, and a 30 second 72°C extension period, followed by 20 cycles of a 15 second 89°C denaturing period, a 15 second 55°C annealing period, and

a 30 second 72°C extension period, which was ultimately followed by a 10 minute 72°C final extension. The ABI PCR GeneAmp 9700 thermocycler was used for all reactions.

After amplification, PCR products were pooled for genotyping. PCR samples from one individual were pooled together if the amplicons from each primer set were labeled with a different dye or if the amplicons were different sizes (non-overlapping) but labeled with the same dye set. A total of 2-4 amplicons were pooled together using 1 µl of each PCR sample, 0.4 µl of GeneScan™ 1200 LIZ® size standard, and 9 µl of DMF (ABI PRISM Linkage Mapping Set Version 2.5 User Guide and Panel Guide). Samples were electrophoresed on the ABI 3100 or the 3130, and the data was analyzed using GeneMapper version 4.

### 3.2.4 Building X chromosome haplotypes

Haplotypes (combinations of alleles that are transmitted together) of the entire X chromosome were built manually using all typed markers, in order to determine which regions were shared amongst the affected individuals. When building haplotypes, phase, defined as knowing from which parent a child inherited the alleles, and recombinants (defined in Chapter 1) need to be determined; studying families with three or more generations makes this task easier. In regard to the individuals genotyped in Family 2024, phase was simply determined for the females (2024.0000 and 2024.A0002) because the paternal X chromosome does not recombine (there is only one copy) and their father's

DNA was available for analysis. Since phase was known for 2024.0000, the maternal crossovers in her children, 2024.A001 and 2024.A002, were correctly determined as well.

### 3.2.5 Fine Mapping

Fine mapping was performed in regions where the genotypes of 2024.0000 were uninformative in order to determine if her children (2024.A001 and 2024:A002) were recombinant or non-recombinant. Additional markers were selected from the UCSC Genome Browser homepage (http://genome.ucsc.edu/index.html?org=Human). A total of 49 additional markers were typed (Appendix 7).

The Invitrogen™ *Taq* DNA Polymerase kit (Cat. #: 10342020) was used to amplify all fine mapping markers (Appendix 1). Trial PCRs were run to determine the best amplification conditions, similar to the trials run in Chapter 2, section 2.2.3(i). However, the recommended annealing temperature for most fine mapping markers was 50°C, thus amplification was first carried out using a standard PCR cycling program with 50°C as the annealing temperature. All steps (denaturation, annealing and extension) were 30 seconds long and there were a total of 30 cycles.

After amplification, the electrophoresis procedure and imaging of the gel followed the same protocol as carried out in Chapter 2, section 2.2.3(i). All markers amplified without betaine under the standard cycling program (annealing temperature 50°C). Agarose gel

electrophoresis was no longer performed once conditions were optimized. Samples were run on the ABI 3100 or the 3130 and analyzed using GeneMapper version 4.

As additional affected individuals within the large multi-generational family were recruited (total=3), they were typed for the markers in the shared region(s) to determine if they had the same 'affected' haplotype. Recombinants of these new individuals were used to narrow the region.

### 3.2.6 Identifying and screening genes within the candidate region

The UCSC Genome Browser homepage (URL: http://genome.ucsc.edu/index.html?org=Human), and the March 2006 assembly (NCBI build 36.1) were used to identify the genomic contig of the candidate region and positional candidate genes. All genes in Refseq were noted (Table 3.2 and 3.3).

Positional candidate genes were prioritized for screening based on expression, function, previously associations with hearing loss, and overlap with known *DFNX* loci. The Morton cochlear cDNA library was used to determine which genes in the candidate region were expressed in the cochlea (URL: http://www.brighamandwomens.org/bwh_hearing/human-cochlear-ests.aspx) (219). Only the genes that were listed in that library were screened in this study. For all expressed genes, primer sets for all coding exons and intron-exon boundaries were designed using Primer 3 (176). The same trial methodology used in Chapter 2, section

2.2.3(i), was applied to determine the optimal amplification conditions for each primer set. See Appendix 8 for the primer sequences and amplification conditions.

### 3.2.7 Creation of a mutation screening panel

A mutation screening panel was established that comprised seven genomic DNA samples from four affected subjects and three unaffected controls, and one $H_2O$ blank. The affected individuals that were chosen included two males (2024.A006 and 2024.1000) and two females (2024.0000 and 2024.A008). The logic behind using seven DNA samples and one $H_2O$ control was the same as in Chapter 2 gene screening (section 2.2.3(ii)). This panel was screened for all primer sets under the optimal conditions that were determined above (Appendix 8). The same sequencing protocol was followed as in Chapter 2.2.3(ii).

### 3.2.8 Detection of genomic rearrangements in *DMD*

The *dystrophin* (*DMD*) gene was a positional candidate in this study, and was also the positional candidate that Pfister et al. in 1998 thought caused deafness at *DFNX3* (262). *DMD* is known to have many exonic duplications and deletions, which can be detected by MLPA (58, 266) by using MLPA probe mixes P034 and P035. MLPA was used in this study to screen for such mutations. Mixes P034 and P035 contain probes for each of the 79 exons in the longest DMD isoform (accession number X14298). In addition, a probe is present for the alternative exon 1 of isoform Dp427c.

Dp427c is not the only isoform of *DMD* with a non-overlapping first exon. Within the refined deafness region Dp40, Dp71 and Dp116 have alternate first exons. Dp40 and Dp71 actually have the same alternate first exon, thus there were two additional exons that did not overlap with the exons of the longest DMD isoform (accession number X14298) that also needed to be screened for duplications and deletions. MLPA probes were designed for these exons manually following the detailed instructions of 'Designing synthetic MLPA probes' version 8 available at www.mlpa.com. See Appendix 9 for specific details on the manually synthesized probes.

The MLPA step-by-step protocol for DNA detection/quantification is available at www.mlpa.com. Below is the protocol for a 1X reaction.

**DNA denaturation and hybridisation of the SALSA MLPA probes**

A DNA sample (50-200 ng of DNA) was diluted with TE to 5 µl and added to a 96 well PCR plate. It was then heated for 5 minutes at 98°C in the thermocycler, which was cooled to 25°C before opening. A mixture of 1.5 µl of SALSA probe mix (black cap) and 1.5 µl of MLPA buffer (yellow cap) was then added to each well while the PCR plate was still in the thermocycler. The mixture was thoroughly mixed by gently re-suspending with a pipette then incubated for 1 minute at 95°C, followed by a 16 hour incubation period at 60°C.

**Ligation reaction**

After the 16 hour incubation, the samples still remained in the thermocycler and the temperature was reduced to 54°C. While at 54°C, 32 µl of Ligase-65 mix was added to each sample and mixed well. That mixture was incubated for 15 minutes at 54°C then heated for 5 minutes at 98°C.

Note: The Ligase-65 mix was made less than 1 hour before use and was stored on ice. To make the Ligase-65 mix, 3 µl of Ligase-65 buffer A (transparent cap), 3 µl of Ligase-65 buffer B (white cap), and 25 µl of $H_2O$ were initially mixed together. Then 1 µl of Ligase-65 (brown cap) was added and the complete mixture was mixed one final time.

**PCR**

In a new PCR plate, 4 µl of SALSA PCR buffer (red cap), 26 µl of $H_2O$ and 10 µl of MLPA ligation reaction were mixed together. This was placed in the thermocycler at 60°C and 10 µl of Polymerase mix was added, which immediately followed the start of the PCR reaction.

Note: The Polymerase mix was made less than 1 hour before use and was stored on ice. To make the Polymerase mix 2 µl of SALSA PCR-primers (purple cap), 2 µl of SALSA Enzyme Dilution buffer (blue cap) and 5.5 µl of $H_2O$ were mixed together. Then 0.5 µl of SALSA Polymerase (orange cap) was added and the whole mixture was mixed again thoroughly.

The PCR had 35 cycles, each with a 30 second 95°C denaturing period, a 30 second 60°C annealing period, and a 60 second 72°C extension period. After the 35 cycles, the samples were then held at 72°C for 20 minutes.

**Separation of amplification products by electrophoresis**

Following the amplification, 1-3 µl of the PCR reaction, 0.3 µl of the internal size standard and 9 µl DMF were mixed together in an individual well of an ABI 96 well plate. The mixture was incubated for 2 minutes at 94°C then held at 4°C for 5 minutes. The sample was then placed on the ABI-Prism 3100 Genetic Analyzer. For specific settings see 'Settings electrophoresis instruments' of the MLPA webpage (58).

## 3.3 Results

### 3.3.1 The hearing loss phenotype in Family 2024 is variable

There were a total of 50 apparently affected individuals (25 males, 25 females) in Family 2024, however medical records were available for only 15 of those individuals to confirm and correctly diagnose the hearing loss (Figure 3.2). Based on the medical records obtained, the hearing loss that segregates in this family can be described as a non-syndromic, bilateral, progressive, sloping hearing loss that is moderate to severe in severity (Figure 3.3), with no radiographic abnormality of the temporal bone.

The severity seemed to differ in males and females. Of the 15 affected individuals with medical reports (10 males, 5 females), the severity of hearing loss was noted for 12 of them (7 males, 5 females) (Figure 3.2 and 3.4). There were two males under the age of 10 with their audiology tested. Their hearing loss, mild-moderate (2024.A001) and moderate-severe (2024.B000) was not as severe as the hearing loss reported from males tested over 40 years of age (2024.1043, 2024.1046, and 2024.1000) (Figure 3.2 and 3.4). Females had a variable and more moderate hearing loss; after the age of 40 there were three females whose hearing loss severity was known; they were classified as mild-moderate (2024.A008), moderate (2024.1024) and moderate-severe (2024.0000) (Figure 3.2 and 3.4).

The age of onset differed between males and females as well. There were 22 cases where the age of onset was not reported, however there were 12 individuals (10 males, 2

females) that were reported to have hearing loss under the age of 10, and 16 individuals (1 male, 15 females) that were first recognized as having hearing loss over the age of 10 (Figure 3.2). Of the 11 total men with a known age of onset, 10 had their hearing loss first reported under the age of 10. The one male that was reported over the age of 10 (2024.1046) was actually 11 years old when he was first recognized to have a hearing problem. Of the 17 women with a known age of onset, 15 (88%) had their hearing loss first reported over the age of 10. Of the female cases where a specific age of onset was reported, the age ranged from as early as three years to 74 years of age, with the majority reported from 20-40 years of age.

The universal newborn hearing screening, implemented in the early 2000s, was only performed on one affected male (2024.A001) who passed the screening but was subsequently diagnosed at the age of two. It is now questioned whether the result of his newborn screening was a false negative, as the first audiology report for 2024.A001 had a "cookie bite" appearance (Figure 3.3). Follow up reports on 2024.A001 also showed hearing loss progression over time at the higher frequencies.

### 3.3.2 Investigation of the mode of inheritance for this extended pedigree

Historically this family was considered to have a sex-linked mode of inheritance. The male-female ratio for individuals reported to have hearing loss is, however, 50/50. Despite the fact that the number of affected males is not larger than females there is a significant difference in the male and female phenotype in regard to age of onset and

severity. The hearing loss that is seen in females is generally less severe, reported later in life, and variable. This supports an X-linked mode of inheritance. Of the 25 affected males, only six had children. These six affected men had a total of 15 children (5 males, 10 females). None of the five sons were affected therefore there was no evidence of male-male transmission. This also supports an X-linked mode of inheritance. Four of the 10 daughters were affected, which demonstrates male-female transmission making mitochondrial inheritance unlikely. If the mode of inheritance is X-linked then the six females that were not deaf are carriers and do not have signs of hearing loss because of penetrance issues or X-inactivation.

There were 22 matrimonial unions that resulted in at least one affected child, eleven of which were unlikely consanguineous, two were possibly consanguineous (due to common pedigree surnames), five were definitely consanguineous and four were unknown (Figure 3.2). The likeliness that all eleven of the unions that seem to be non-consanguineous are actually consanguineous is low, thus an autosomal recessive form of deafness caused by an IBD mutation is unlikely. Furthermore, if this familial deafness is inherited recessively, the likeliness that all married-in spouses in all the non-consanguinous unions were carriers of a recessive deafness allele is also unlikely. However, it is important to note that due to the genetic heterogeneity of deafness, it is a possibility that some branches of the pedigree may be caused by different genetic forms of deafness. Lastly, the female, 2024.3000, had children with two different men, one union was consanguineous and the other unlikely consanguineous. Both relations resulted in

children with hearing loss (Figure 3.2). This is more evidence that a recessive mode of inheritance is unlikely.

### 3.3.3 X chromosome haplotype analysis

Haplotypes spanning the whole X chromosome were created using all genotyped markers (Figure 3.5). The paternally inherited chromosomes of females 2024.0000 and 2024.A002 were confirmed by haplotyping their fathers, 2024.1000 and 2024.0002, respectively. Phase was therefore simply determined for the females because the paternal chromosome does not recombine. Since phase was known for 2024.0000, the maternal crossovers in her children 2024.A001 and 2024.A002 were correctly determined as well. Individual 2024.A001 had three crossover events over the whole X chromosome and 2024.A002 had only one (Figure 3.5).

The four affected individuals shared two common regions on the X chromosome (Figure 3.5). One was a 27.6 Mb region on Xp with crossovers occurring in 2024.A001 between *DXS8051* and *DXS987* at the telomeric end, and between *DXS1049* and *DXS8090* at the centromeric end (Figure 3.5). They also shared the genomic region around *DXS1106* on Xq. The crossovers occurred between *DXS8089* and *DXS1106* in 2024.A002 and between *DXS1106* and *DXS8055* in the 2024.A001. Therefore two candidate disease regions were shared, one between *DXS8051* and *DXS8090* and the other between *DXS8089* and *DXS8055* (Figure 3.5).

The Xq region was excluded after haplotyping three additional affected individuals from Family 2024 (Figure 3.2 and 3.6). One of the three additional affected relatives, 2024.A006, did not share the same allele (122) at *DXS1106*, nor did he share either allele of the two surrounding markers, *DXS8089* and *DXS8055*, in fact he inherited a completely different haplotype (Figure 3.6). Another affected individual, 2024.A008 did not share the same allele at *DXS1106* either. Her parents were not typed so phase could not be determined but she was heterozygous for *DXS1106* and did not have the 122 allele. Her haplotypes were inferred but it is possible that she shared the proximal half of the potential disease-associated yellow haplotype. The third affected relative, 2024.0007, shared the whole Xq region with affected individuals (Figure 3.6).

The region on Xp however could not be excluded after haplotyping the additional affected family members; all affected individuals shared a portion of the Xp disease-associated yellow haplotype (Figure 3.2 and 3.7). A crossover in 2024.A006 between markers *DXS999* and *DXS7592* narrowed the telomeric region, and a crossover in 2024.A008 between markers *DXS997* and *DXS1219* narrowed the centromeric region. Therefore the boundary markers for the shared region are *DXS999* and *DXS1219*, narrowing the region to 13.3 Mb from 27.3 Mb (Figure 3.7). This interval overlaps the two previously discovered Xp *DFNX* loci, *DFNX3* and *DFNX4*. *DFNX3* was mapped to a 1.5 Mb interval that is completely within the Xp region that is shared by affected individuals in Family 2024, and shares the same centromeric boundary marker *DXS1219*. The overlap with *DFNX4* is 3.5 Mb (Figure 3.8).

Interestingly, individual 2024.0000 is homozygous for alleles on the Xp haplotype between and including the markers *DXS1214* and *DXS997* (Figure 3.5 and 3.7). It should be noted that the parents of 2024.0000 are related (first cousins once removed) and her mother, 2024.1006, is affected as well (Figure 3.2, 3.5 and 3.7). Therefore they might share the same segregating ancestral genetic mutation. The DNA of 2024.1006 was not available for genotyping, however the X chromosome that 2024.0000 inherited from her mother was deciphered because the DNA of her father was available, the male X chromosome does not recombine, and phase could be 100% determined. 2024.0000 had a 50% chance of inheriting the X-linked deafness allele from her mother and a 100% chance of inheriting it from her father. It is possible that 2024.0000 inherited two copies of the ancestral disease haplotype. Of all the affected women with medical records, 2024.0000 had the most severe hearing loss, which supports the fact that X-inactivation would not affect her phenotype (if she is homozygous for the disease variant). If that is the case, the small region of homozygosity is 0.94 Mb between markers *DXS1234* and *DXS1219*. This only overlaps the *DFNX3* locus (Figure 3.8).

### 3.3.4 Screening of positional candidate genes

Within the 13.3 Mb Xp critical region, there were 48 positional candidate genes (Table 3.2 and 3.3), 13 of which are expressed in the cochlea, according to the Morton cDNA library (Table 3.2). The first 18 genes listed in Table 3.2 overlap with *DFNX4* and the last three (*TAB3*, *FTHL17* and *DMD*) with *DFNX3* (Table 3.2 and Figure 3.8). The 0.94 Mb region of homozygosity in 2024.0000 only has one annotated gene, *DMD* (Figure 3.8

and Table 3.3). All 13 cochlea expressed genes were sequenced. The genes that overlapped with the two known deafness loci, were previously associated with hearing loss, and had functions related to the hearing process were prioritized for screening. The *Duchenne muscular dystrophy* (*DMD*) gene was the first gene to be screened. This gene was the only gene within the region of homozygosity of 2024.0000, is expressed in the cochlea, and was the key candidate for the *DFNX3* locus. The largest isoform of *DMD* (X14298) has 79 exons and exons 45-79 were within the disease interval (Table 3.3). Sequencing those exons in addition to other exons from different isoforms that did not overlap with the exons of X14298, for example, exon 1 of Dp71 (Table 3.3) revealed no sequencing variants. No genomic rearrangements were detected either.

The protein kinase *RSK2*, along with *SMS* and *PHEX* were screened next (Table-3.2). These three genes overlap with the *DFNX4* locus, are all expressed in the cochlea and have been associated with hearing loss either in human or the mouse. *RSK2* is the causal gene for Coffin-Lowry syndrome (CLS) [MIM #303600], a rare form of X-linked mental retardation characterized by skeletal malformations, growth retardation, paroxysmal movement disorders, cognitive impairment and a hearing deficit in affected males and some carrier females (267, 268). Both *SMS* and *PHEX* have partial deletions in the Gyro mouse, which is the mouse model for X-linked hypophosphatemia [MIM #307800], a disease characterized by growth retardation, bone disease, hypophosphatemia, renal defects, and also neurologic abnormalities, including deafness, hyperactivity, circling behavior, and inner ear abnormalities (269, 270). Only one variant was detected, *SMS*

c.170+30C>T (Table 3.4). This variant was found exclusively in all affected individuals on the mutation screening panel but according to dbSNP has a frequency of over 40% in the African population.

Overall, after all 13 genes were screened, a total of 11 variants were detected (Table 3.4). None of these could be deemed pathogenic as they were either not detected in all affected individuals or found in control individuals.

## 3.4 Discussion

The hearing loss that segregates in Family 2024 has a pattern that models X-linked inheritance. Firstly, there was no evidence of male to male transmission. Six affected males, all together, had 15 children. Five of those 15 children were males, and all five sons were not affected. Secondly, there was a significant difference in the male and female phenotype in regard to severity and age of onset. After 40 years of age, all affected males with known audiology had a severe hearing loss, whereas the hearing loss of the affected women was more moderate. There was also a range of severity for the hearing loss of females over 40 years of age, which ranged from mild-moderate to moderate-severe, representing a variable hearing loss. The age of onset differed between sexes as well; 91% of males versus only 12% of females, with a known age of onset, were reported to have their hearing loss before the age of 10.

It is important to discuss that incomplete medical records are not ideal for gene discovery efforts. The X-linked pattern of inheritance in Family 2024 was determined using all 50 'affected' individuals (35 individuals without, and 15 with the proper medical records). The affection status for all the affected individuals without the proper medical records is, in fact, uncertain – some individuals may truly be unaffected. Also, considering that in the 1970s the affection status of some individuals was dependent upon a relative's recollection of the family history, some unaffected individuals may actually be truly affected as well. Both scenarios could change the visible mode of inheritance of the pedigree. Furthermore, without the proper medical records, the exact deafness phenotype

(severity/frequencies affected) for each affected individual cannot be determined. Considering the genetic heterogeneity of deafness and the range of possible hearing loss phenotypes, some affected individuals of the large pedigree of Family 2024 may have different genetic or environmental causes. It also has to be considered, however, that having the proper medical records will not always identify true phenocopies. Consider, for example, the variable phenotype of females within an X-linked deafness family - the medical records for all the affected women will vary in severity and age of onset. The deafness phenotype could be, and is presumed to be, due to the same genetic cause, however one or more individuals could be a phenocopy. All the affected individuals that were used for the genetic analysis in this study had medical records available; however, it is possible that one of those individuals is still a phenocopy. The presence of phenocopies in hereditary deafness families is a common phenomenon and can affect gene discovery efforts. However, this does not that mean that most hereditary deafness gene discovery efforts are not worth a try. Some efforts will fail while others will succeed. The clinical information that was available for members of Family 2024 was incomplete but with the presenting evidence for X-linked inheritance this genetic study had merit.

The hearing loss in Family 2024 appears to segregates with a region on the small arm of the X chromosome. A 13.3 Mb region was determined to be shared by all affected individuals (n=7) that took part in the genetic analysis. In addition, a small region of homozygosity in 2024.0000 may potentially narrow the region to 0.94 Mb. The father

and mother of 2024.0000 were first cousins once removed. Family 2024 was created by linking two separately recruited families, family 24 and 25. The father (2024.1000) of 2024.0000 came from Family 24 (his mother was the proband), and the mother (2024.1006) came from Family 25 (her maternal grandmother was the proband). The only individuals with DNA available were from Family 24 and a 13.3 Mb region was shown to segregate with the disease through that part of the pedigree. Haplotyping individuals in Family 25 (particularly 2024.1006 and 2024.2000) could confirm the segregation of the same ancestral haplotype through this side of the family and clearly demonstrate that a recombination between *DXS1234* and *DXS1214* reduces the region to 0.94 Mb.

The 13.3 Mb region overlaps with two known loci for non-syndromic, X-linked deafness. Both *DFNX3* and *DFNX4* have been mapped to the p arm of the X chromosome. Two families have been linked to *DFNX3* and one to *DFNX4* (261, 265). The causative genes for these loci are unknown but the causative gene of one of them may be the same as Family 2024. It is also possible that this newly discovered disease interval represents a new deafness locus (*DFNX5*?) because deafness is heterogeneous.

*DFNX4* was mapped in 1996 and a clinical follow up of the family later refined the region to an 8.4 Mb region (66, 265). The region has over fifty positional candidate genes (51, 52) and there is no published data stating whether any of them have been screened. Due to small family size and limited number of families, this disease-linked interval could not

be refined to a desirable size for screening positional candidates, thus the gene discovery efforts have been put on hold (265). If Family 2024 is a *DFNX4* family then the region would be reduced to 3.4 Mb. Perhaps collaborating with the Moreno group in Spain and screening the positional candidates in this region of overlap is more feasible.

The mapped region for *DFNX3* was, however, previously refined to a small size (approximately 1.5 Mb), and the 0.94 Mb region of homozygosity in 2024.0000 overlaps with this locus. This region of homozygosity is entirely within the *DMD* locus, unlike *DFNX3*, which includes two other positional candidate genes, thus this would support the authors beliefs that *DMD* is the causative gene (262). Pfister et al. only screened for genomic deletions/duplications in exons of the large isoform of *DMD* and no variations were detected (262). When deciding which methods would be undertaken to screen *DMD* in Family 2024, many issues had to be considered due to the complexity of the *DMD* gene. All coding exons of all *DMD* isoforms in the disease linked region were sequenced and analyzed for genomic rearrangements. Since truncation mutations generally cause *Duchenne* muscular dystrophy, one possibility was that a missense mutation in an exon of one of the larger isoforms might cause non-syndromic deafness. Another possibility was that a mutation in the first exon of either, Dp71 or Dp116, two smaller *DMD* isoforms that are expressed in the nervous system, and whose first exons do not overlap with any other *DMD* exon, causes deafness. However, no pathogenic variant was detected through sequencing or MLPA.

Affected males in Family 2024 experience a non-syndromic, bilateral, progressive, sloping hearing loss that is very severe. The earliest age of onset was reported as 2 years. This phenotype seems most similar to that seen in the *DFNX4* family, where the hearing loss is detected first at school age, whereas the hearing loss in the two *DFNX3* families is profound at birth. The exact age of onset may not be known for previously described *DFNX4* family members or individuals in Family 2024 but they both have a bilateral, progressive, sloping hearing loss with similar severity. The earliest hearing test of an affected male in Family 2024 was carried out on a 2 year old and the audiogram resulted in a 'cookie bite' configuration. Del Castillo et al. showed the early stage of *DFNX4* hearing loss through an audiogram of a 9 year old male whose hearing loss was sloped and more severely affected the high frequencies (265, 271). This was more similar to the follow up audiograms of the boy in Family 2024 who was first tested at 2 years of age.

Affected females in Family 2024 showed variable expression in regard to both age of onset and severity. In general, there was a much milder hearing loss with a delayed onset compared to males. This is typical of X-linked disorders, and all previously reported non-syndromic X-linked deafness loci have female carriers that are affected in the same manner (245, 251, 261, 265). This can be explained by X-inactivation. Even though female cells have two copies of the X chromosome, for gene dosage purposes, one is inactivated during fetal development. Each tissue in the female body is a mosaic, a compilation of two cell types that differ based on which X chromosome is expressed. The degree of mosaicism differs from tissue to tissue and person to person. The

phenotype of X-linked diseases in females is thus determined by both cell types in the affected tissue, for example the cochlea. If an affected female allele is heterozygous for the disease allele, the levels of the two different synthesized proteins (one mutated and one wild-type) in the affected tissue are responsible for the phenotypic diversity in females (240).

Sequencing the 13 positional candidate genes that are expressed in the cochlea (219, 272) did not revealed a causative varaint. In fact very few variants were found in general, which is not surprising since after the euchromatic sequence of the human X chromosome was determined, analysis of the sequence revealed that it actually had a low heterozygosity level compared to the autosomes, approximately 57% less (273). Ross et al. also determined that the X chromosome had many interspersed repeats, in particularly LINEs (273). It is already known that Xq21, the location of *POU3F4* (*DFNX2*), is subject to many rearrangements and that the deletion of an important regulatory element upstream of the gene is the cause of fifty percent of *DFNX2* deafness (66). It is important to note that large genomic rearrangements of exons or *cis*-acting regulatory elements may be a common cause of X-linked deafness in general. For example, rearrangements happen frequently at the *DMD* locus as well, specifically, it has a recombination frequency of approximately 9-14%, which is sixfold higher than other chromosomal regions of similar size (261, 274-276). Perhaps a duplication or deletion of the promoter or another *cis*-acting regulatory element of the *DMD* isoforms Dp71 or Dp116 causes deafness at *DFNX3*, similar to the situation with *POU3F4*. It is also important to note

that not all genes or isoforms are annotated. Perhaps there is an unidentified cochlea-specific isoform of one of the positional candidate genes with an independent promoter, or maybe a novel intragenic gene exists that is completely independent of the overlapping gene. Finally, the remaining positional candidates that happened not to be in the Morton human fetal cochlear library could actually be involved in deafness and just not expressed at the time of development that the library was created.

Typically, the more refined a critical region, the easier it is to find a disease gene. The key to refining a disease region is finding crucial recombinants. That is more likely to happen by studying a high number of meioses, in other words, by having a high participation rate. The recruitment efforts for this study were good but, for numerous reasons, many individuals in the family did not participate. For example, out-migration is common for the Newfoundland population and many members of Family 2024 sib-ships currently live out of the province. Also, telephone recruitment efforts may not be the best approach for a hereditary deafness study. Family 2024 had 50 apparently affected individuals but DNA was collected from only seven affected individuals. Better enrollment might narrow the disease region. However, next generation sequencing (NGS) has the ability to target-capture and sequence desired regions of the human genome in order to identify causative disease variants using a small sample size (56, 277); so maybe the recruitment problem can be overcome. One future possibility would be to target-capture the Xp region and sequence that genomic region in all (or most) the affected individuals that participated in the genetic study, and some controls. If the Xp

region is the true disease locus and the causative variant is a sequencing variant then this approach would theoretically detect the variant no matter if it was coding or non-coding. Another possibility would be to exome sequence the affected individuals (277, 278). One would have to hope that the disease variant is coding and covered with the exome sequencing kit, but this approach would be a way to study different possible modes of inheritance because the whole exome will be sequenced. The X chromosome could be analyzed first, particularly the genes in the Xp region, but the data would be available to look for autosomal dominant and recessive variants as well.

X-linked diseases have had a major impact on medical genetics. Firstly, they are well studied because the mode of inheritance is recognized relatively easily. Secondly, there have been a large number of diseases associated with the X chromosome (because of male hemizygosity and the phenotypic consequence of carrying a single mutation); just over 8% of recognized Mendelian diseases (342 of 4140) are X-linked (160). This is despite the fact that the X chromosome has a small number (n=1098) of annotated genes, representing 4% of all the human genes (273). Progress in gene identification of many human diseases has been influenced by studying the X chromosome. To date, there are 205 X-linked phenotypes for which the molecular basis is known (160). In fact, the first gene associated with non-syndromic deafness was the X-linked gene *POU3F4*, the causal gene of the *DFNX2* locus (66, 245). Identifying the deafness mutation in Family 2024 has the potential to help additional families with X-linked deafness around the world.

Gene discoveries help elucidate the pathogenesis of deafness, which involves approximately 1% of all human genes. Proteins that comprise ion channels, the extracellular matrix, the cytoskeleton, and transcription factors all regulate proper auditory function. Each new finding helps piece together how the ear actually works. For example, in 2005 digenic inheritance involving *PCDH15* and *CDH23* mutations was reported to cause non-syndromic deafness in mice and humans (279). Through functional analysis it was then determined that *CDH23* and *PCDH15* interact to form tip-links at the end of hair cell stereocilia where mechanoelectrical transduction occurs. If the interaction of these proteins is disrupted by either a homozygous mutation in either gene or digenic inheritance, then deafness occurs (280). Gene discovery followed by functional work is the key to identifying deafness genes, determining their specific function, and ultimately linking the functionally diverse group of proteins together. This may even lead to new therapeutic approaches for the hearing impaired in addition to the use of hearing aids and cochlear implants.

**Table 3.1: Non-syndromic X-linked deafness loci.**

| Locus | | Location | Gene | Hearing loss type | Severity | Tones | Onset | Progression | References |
|---|---|---|---|---|---|---|---|---|---|
| previous designation | new designation | | | | | | | | |
| DFN1 | excluded | Xq21 | *TIMM8A* | syndromic | - | - | - | - | Tranebjaerg et al (1995) - Jin et al (1996) |
| DFN2 | DFNX1 | Xq21.33-Xq22.3 | *PRPS1* | Sensorineural | Severe-profound | All | Variable | Variable | Tyson et al ( 1996) - Liu et al (2010) |
| DFN3 | DFNX2 | Xq21 | *POU3F4* | Mixed | Severe | All | Congenital | Rapid Progression | De Kok et al (1995) |
| DFN4 | DFNX3 | Xp21-Xp11 | - | Sensorineural | Severe-profound | All | Congenital | Progressive | Lalwani et al (1994) |
| DFN5 | - | withdrawn | - | - | | | | | - |
| DFN6 | DFNX4 | Xp22 | - | Sensorineural | Severe-profound | High --> All | First decade | Progressive | del Castillo et al (1996) |
| DFN7 | - | withdrawn | - | - | - | - | - | - | - |
| DFN8 | - | reserved | - | - | - | - | - | - | - |

## Table 3.2: The 48 genes within the 13.3 Mb region on Xp

| Gene Number | Gene Name | Gene expression in the human cochlear (Morton cDNA library) | Accession Number | Strand | Start | End | Exons |
|---|---|---|---|---|---|---|---|
| 1 | PPEF1 | No | NM_006240 | + | 18618966 | 18755953 | 19 |
| | | | NM_152224 | + | 18618966 | 18755953 | 19 |
| | | | NM_152226 | + | 18618966 | 18755953 | 18 |
| 2 | PHKA2 | Yes | NM_000292 | - | 18820337 | 18912401 | 33 |
| 3 | GPR64 | Yes | NM_001079859 | - | 18917346 | 19050598 | 28 |
| | | | NM_005756 | - | 19050598 | 19050598 | 29 |
| | | | NM_001079860 | - | 18917353 | 19050598 | 27 |
| | | | NM_001079858 | - | 18917353 | 19050598 | 29 |
| 4 | PDHA1 | Yes | NM_000284 | + | 19271932 | 19289723 | 11 |
| 5 | MAP3K15 | No | NM_001001671 | - | 19410315 | 19410315 | 30 |
| 6 | SH3KBP1 | Yes | NM_031892 | - | 19463441 | 19815640 | 18 |
| | | | NM_001024666 | - | 19463751 | 19727790 | 17 |
| 7 | CXorf23 | No | NM_198279 | - | 19844041 | 19898303 | 11 |
| 8 | MAP7D2 | No | NM_152780 | - | 19935150 | 20044937 | 16 |
| 9 | E1F1AX | No | NM_001412 | - | 20052558 | 20069887 | 7 |
| 10 | RSK2 (RPS6KA3) | Yes | NM_004586 | - | 20077951 | 20194671 | 22 |
| 11 | CNKSR2 | No | NM_014927 | + | 21302901 | 21580700 | 22 |
| 12 | KLHL34 (FLJ34960) | Yes | NM_153270 | - | 21583530 | 21586369 | 1 |
| 13 | SMPX | No | NM_014332 | - | 21634013 | 21686151 | 5 |
| 14 | MBTPS2 | No | NM_015884 | + | 21767675 | 21810794 | 11 |
| 15 | YY2 | No | NM_206923 | + | 21784524 | 21785642 | 1 |
| 16 | SMS | Yes | NM_004595 | + | 21868763 | 21922875 | 11 |
| 17 | PHEX | Yes | NM_000444 | + | 21960842 | 22176397 | 22 |
| 18 | ZNF645 | No | NM_152577 | + | 22200981 | 22202495 | 1 |
| 19 | DDX53 | No | NM_182699 | + | 22928008 | 22930125 | 1 |
| 20 | PTCHD1 | No | NM_173495 | + | 23262906 | 23324839 | 3 |
| 21 | PRDX4 | No | NM_006406 | + | 23595566 | 23614434 | 7 |
| 22 | ACOT9 | No | NM_001037171 | - | 23631699 | 23671328 | 16 |
| | | | NM_001033583 | - | 23631699 | 23671328 | 15 |
| 23 | SAT1 | Yes | NM_002970 | + | 23711225 | 23714246 | 6 |
| 24 | APOO | No | NM_024122 | - | 23761398 | 23835955 | 9 |
| 25 | CXorf58 | No | NM_152761 | + | 23836044 | 23867545 | 9 |
| 26 | KLHL15 | No | NM_030624 | - | 23915884 | 23955224 | 4 |
| 27 | EIF2S3 | Yes | NM_001415 | + | 23982986 | 24006846 | 12 |
| 28 | ZFX | No | NM_003410 | + | 24079729 | 24142548 | 9 |
| 29 | PDK3 | No | NM_005391 | + | 24393475 | 24462463 | 11 |
| 30 | PCYT1B | No | NM_004845 | - | 24486126 | 24575274 | 8 |
| 31 | POLA1 | No | NM_016937 | + | 24621977 | 24925021 | 37 |
| 32 | LOC644820 | No | NM_001089876 | - | 24716384 | 24717149 | 1 |
| 33 | ARX | No | NM_139058 | - | 24931732 | 24943986 | 5 |
| 34 | MAGEB18 | No | NM_173699 | + | 26066384 | 26068773 | 3 |
| 35 | MAGEB6 | No | NM_173523 | + | 26120478 | 26123683 | 2 |
| 36 | MAGEB10 | No | NM_182506 | + | 27736046 | 27750683 | 3 |
| 37 | WDR42B | No | NM_001017930 | - | 27906031 | 27909487 | 1 |
| 38 | IL1RAPL1 | Yes | NM_014271 | + | 28515480 | 29884749 | 11 |
| 39 | MAGEB2 | No | NM_002364 | + | 30143601 | 30148125 | 2 |
| 40 | MAGEB3 | No | NM_002365 | + | 30158474 | 30165528 | 5 |
| 41 | MAGEB4 | No | NM_002367 | + | 30170090 | 30172227 | 1 |
| 42 | MAGEB1 | No | NM_002363 | + | 30171769 | 30180075 | 4 |
| | | | NM_177415 | + | 30171769 | 30180075 | 3 |
| | | | NM_177404 | + | 30175192 | 30180076 | 2 |
| 43 | NR0B1 | No | NM_000475 | - | 30232506 | 30237413 | 2 |
| 44 | CXorf21 | No | NM_025159 | - | 30486862 | 30505835 | 3 |
| 45 | GK | No | NM_000167 | + | 30581461 | 30658646 | 19 |
| | | | NM_203391 | + | 30581461 | 30658646 | 20 |
| 46 | TAB3 (MAP3K7IP3) | Yes | NM_152787 | - | 30755480 | 30817432 | 11 |
| 47 | FTHL17 | No | NM_031894 | - | 30999279 | 31000091 | 1 |
| 48 | DMD | Yes | See Table 5.3 | | | | |

**Table 3.3: *DMD* isoforms and their corresponding exons within the disease interval.**

| Isoform | | Accession number | Exons within the disease interval (the number in each cell corresponds to the exon number of each specific isoform and each column represents the same exonic sequence - overlapping exons) |
|---|---|---|---|
| Number | Name | | |
| 1 | Dp40 | NM_004019 | 1 2 3 4 5 6 7 8 9* |
| 2 | Dp71 | NM_004015 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 |
| | Dp71a | NM_004017 | 1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 |
| | Dp71ab | NM_004018 | 1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 18* |
| | Dp71b | NM_004016 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 18* |
| 3 | Dp116 | NM_004014 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 |
| 4 | Dp140 | NM_004013 | 8* 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 |
| | Dp140ab | NM_004022 | 8* 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34* |
| | Dp140b | NM_004021 | 8* 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35* |
| | Dp140bc | NM_004023 | 8* 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27* 23 29 30 31* |
| | Dp140c | NM_004020 | 8* 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27* 23 29 30 31 32 |
| 5 | Dp260-1 | NM_004011 | 17 18 19 20* 21* 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 |
| | Dp260-2 | NM_004012 | 17 18 19 20* 21* 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 |
| 6 | Dp427p1 | NM_004009 | 45 46 47 48* 49* 51 52 53 54 55 56 57 58 59 60 51 62 63 64 65 66 67 68 69 70 71 72 73 74 75 75 77 78 79 80 |
| | Dp427p2 | NM_004007 | 44 45 46 47* 48* 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 |
| | Dp427m | NM_004006 | 45 46 47 48* 49* 50 51 52 53 54 55 56 57 58 59 50 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 |
| | Dp427c | NM_001109 | 45 46 47 48* 49* 50 51 52 53 54 55 56 57 58 59 50 61 62 63 64 65* 66 67 68 69 70 71 72 73 74 75 76 77 78 79 |
| 7 | DMD | X14258 | 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 50 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 |

Only translated exons in the disease region are shown. When an exon was missing from a particular isoform it is grayed out. Exon 1 in Dp40, Dp71 and Dp116 are non-overlapping. Dp140 8* is the 8th exon of Dp140 and translation starts 71bp into this exon. Dp427c 65* is the 65th exon in this isoform and is 3 nucleotides shorter than the same exon in other isoforms. Dp140bc and Dp140c 27* is shorter than the other exons of the same sequence in other isoforms. Dp40 9* is 4bp longer than the same exon in other isoforms therefore a stop codon is introduced. Dp71ab, Dp71b 18* and Dp140ab 34*, Dp140b 35*, Dp140bc 31* lack the normally present second last exon therefore the open reading frame is different in the last exon. Dp260-1, Dp260-2 20* and Dp427p1, Dp427m, Dp427c 48* and Dp427p2 47* are 3bp shorter than the overlapping X14298 exon. Dp260-1, Dp260-2 21* and Dp427p1, Dp427m, Dp427c 49* and Dp427p2 48* is 3bp larger than the overlapping X14298 exon.

**Table 3.4: The 11 variants detected after sequencing the 13 positional candidate genes that are expressed in the cochlea.**

| Gene | Variants | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| PHKA2 | c.2138 -63 delA | c.3708+725 delA | - | - |
| GPR64 | c.510-21G>A | c.1414-120 C>T | c.1557+69 T>G | c.2340_2341 insA |
| PDHA1 | - | - | - | - |
| SH3KBP1 | - | - | - | - |
| RSK2 | - | - | - | - |
| KLHL34 | - | - | - | - |
| SMS | c.170+30 C>T | - | - | - |
| PHEX | - | - | - | - |
| SAT1 | c.66+59 A>G | - | - | - |
| EIF2S3 | c.99C>T (H33H) | c.374 A>G (K125R) | - | - |
| IL1RAPL1 | c.1-19 G>A | c.703+85 C>A | - | - |
| TAB3 | - | - | - | - |
| DMD | - | - | - | - |

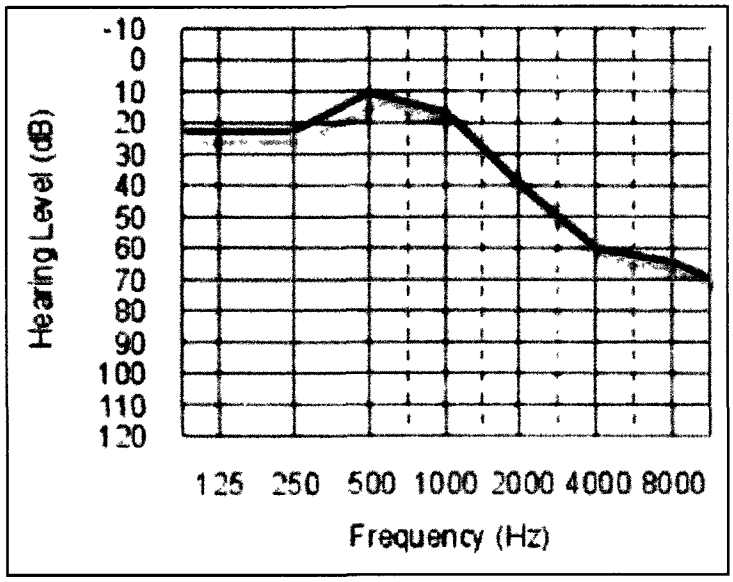



**Figure 3.1: The audiogram.**

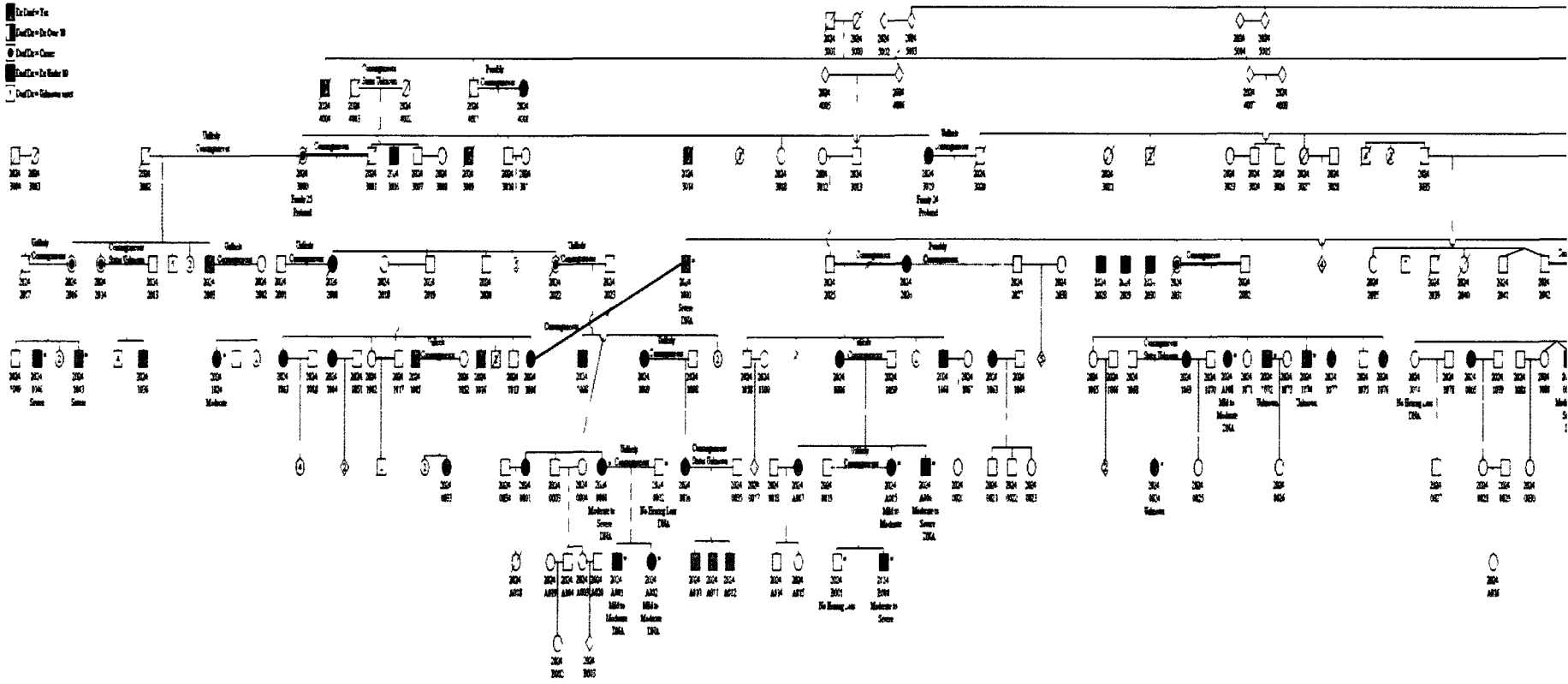
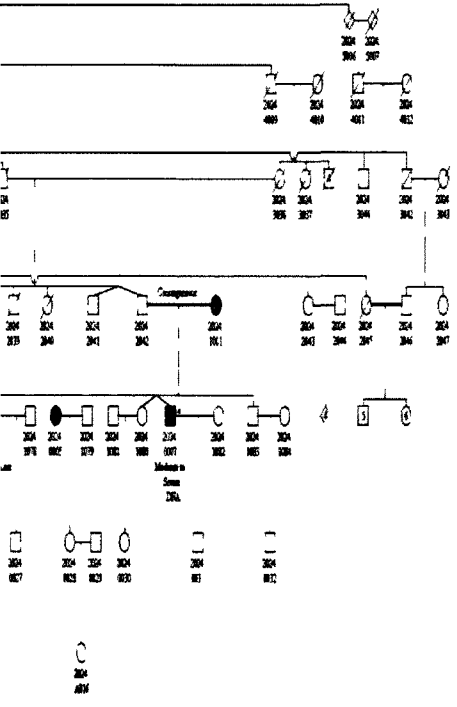


**Figure 3.2: Pedigree of Family 2024.** The different symbols represent hearing loss onset. An asterisk indicates that medical records were available. of each individual with medical records is written under their symbol. If the severity was unknown it is also indicated. DNA is written under each in blood sample was collected and genetic analysis was performed.



ere available. The hearing loss severity
under each individual for whom a

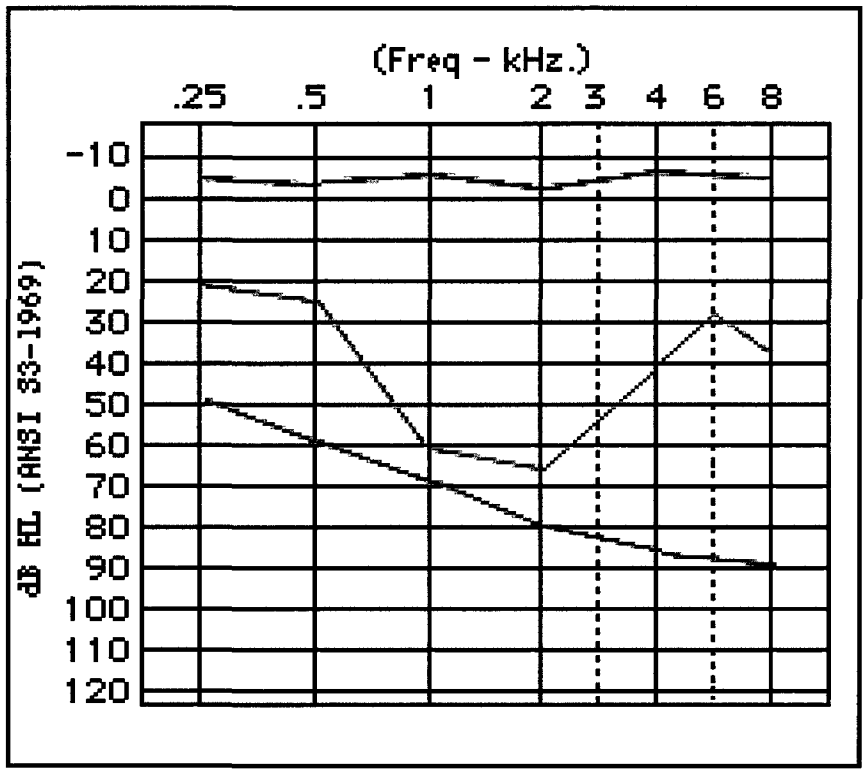
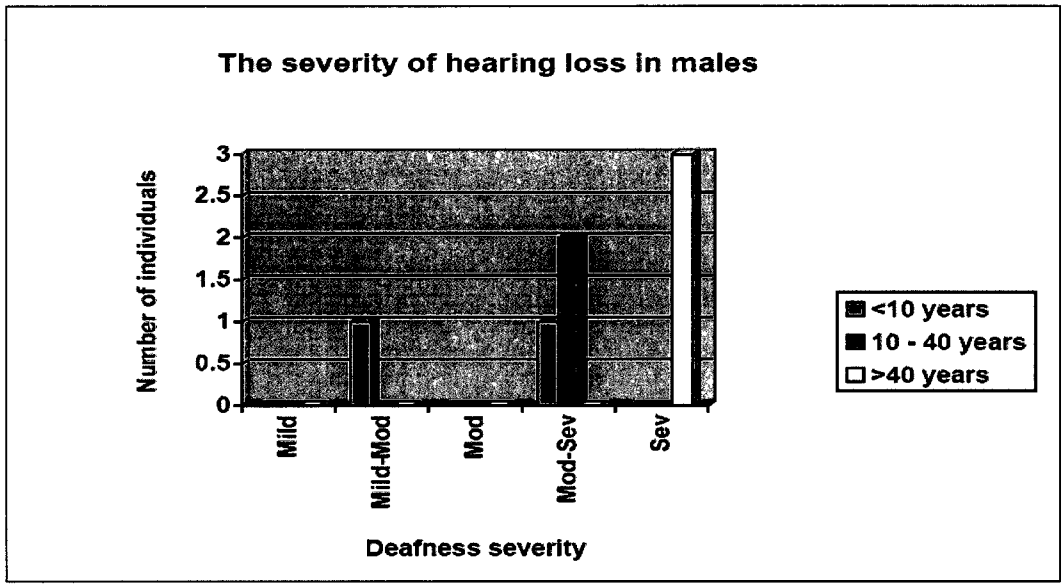



**Figure 3.3: Audiology reports of affected individuals in Family 2024.** The blue line represents a normal hearing individual, the green line represents an affected male – 2024.A001 tested at 2 years of age, the red line represents an affected male adult – 2024.A006 tested at 22 years of age.

**A**




**B**

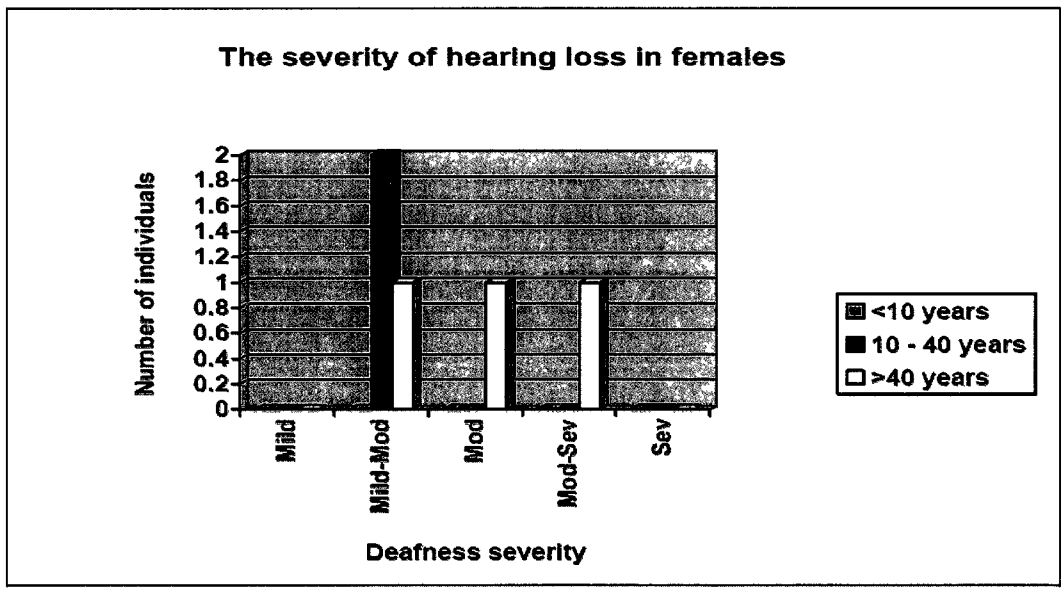
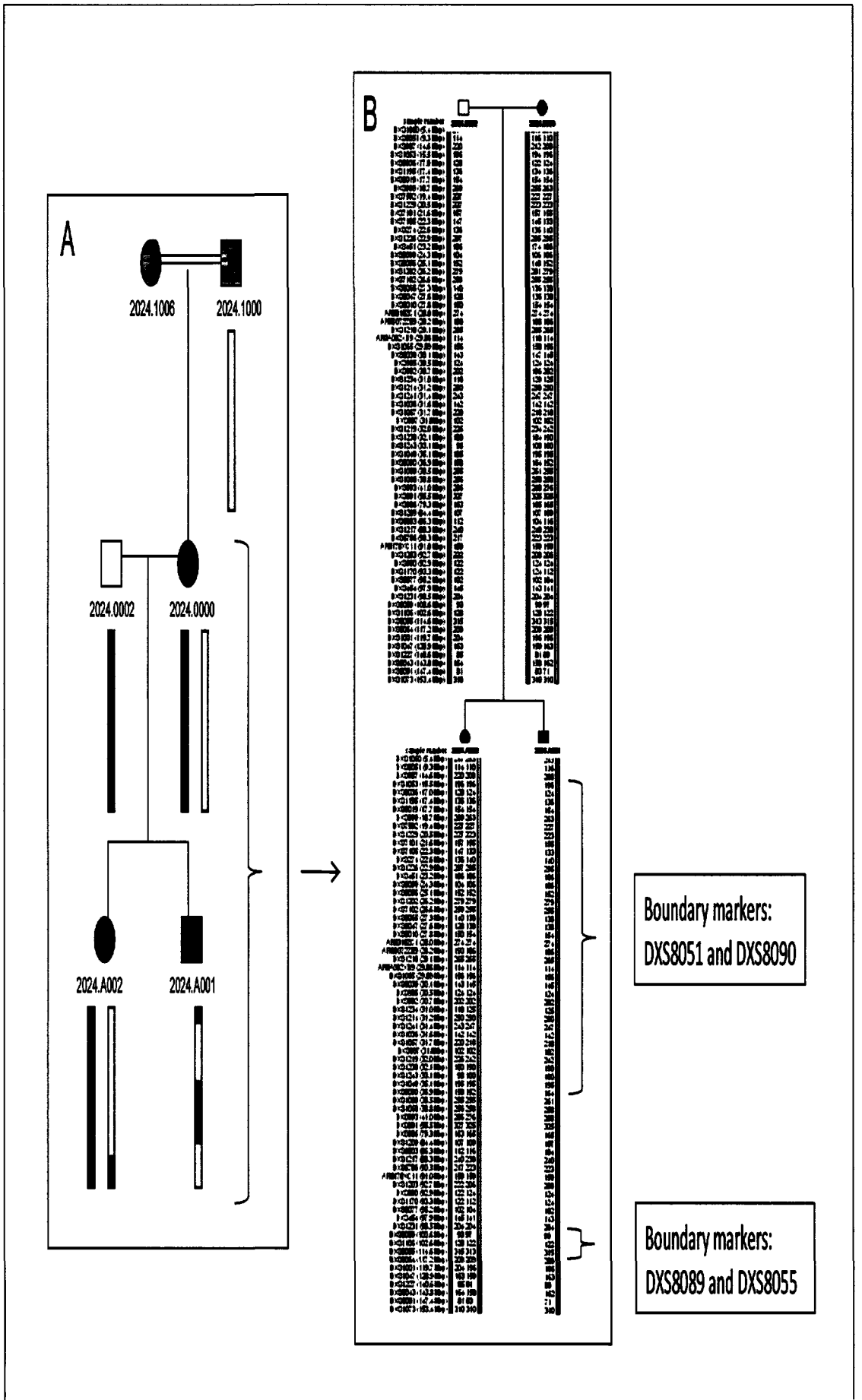



**Figure 3.4: The severity of hearing loss of affected individuals within various age groups.** Each individual was classified into an age group according to when the severity was reported.


A
2024.1006
2024.1000
2024.0002
2024.0000
2024.A002
2024.A001
B
Boundary markers:
DXS8051 and DXS8090
Boundary markers:
DXS8089 and DXS8055

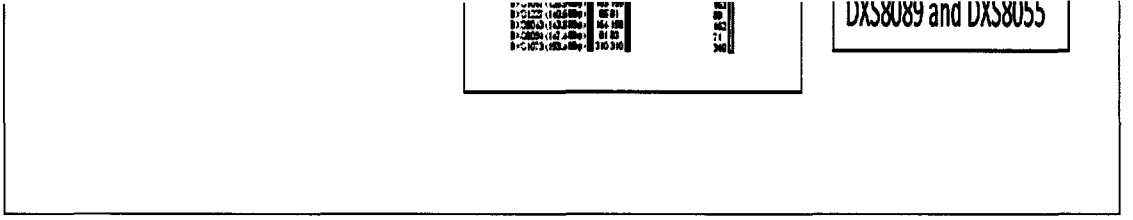
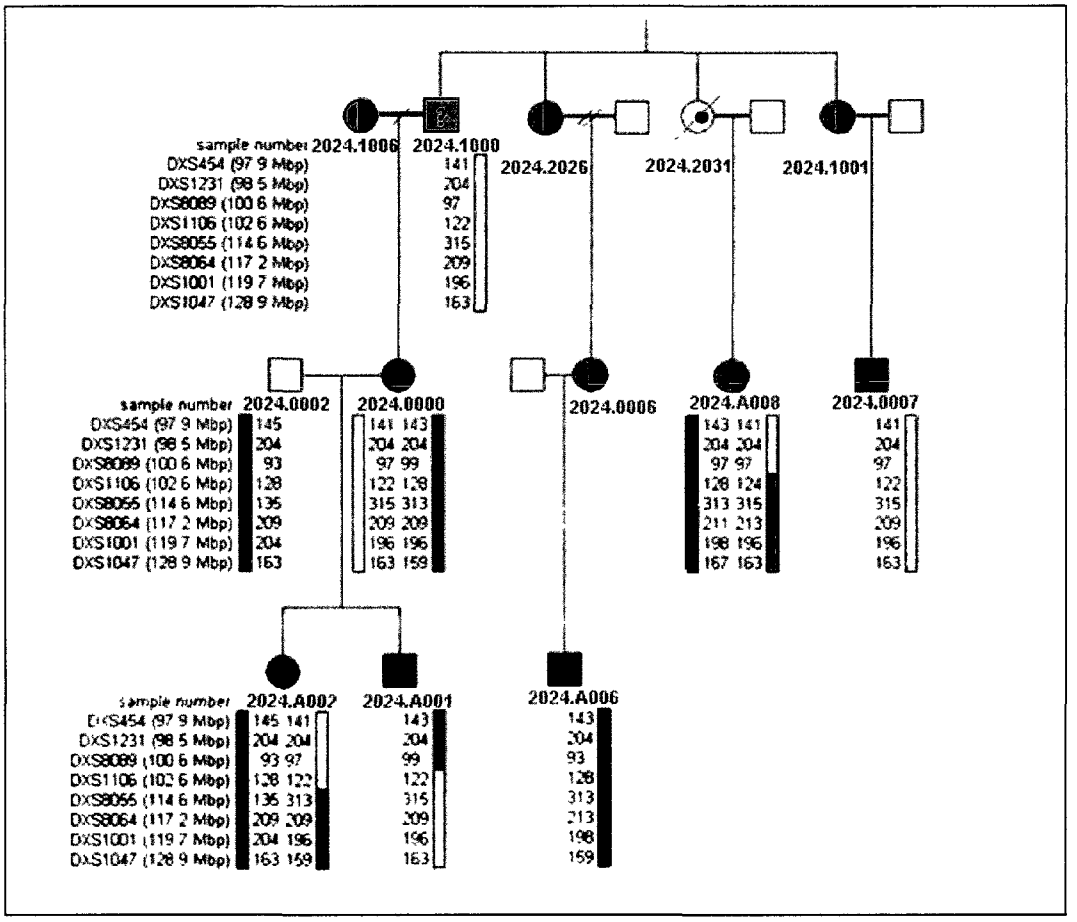



**Figure 3.5: Haplotypes spanning the entire X chromosome reveal the regions shared by affected individuals.** (A) Haplotypes were determined for four affected individuals in this three generation sub-pedigree of Family 2024. (B) Haplotypes using all genotyped markers are shown for the affected individuals in the youngest 2 generations. The two regions shared by all affected individuals are indicated with brackets.

A




B

| Markers | Mbps | 2024.1000 | 2024.0000 | 2024.A002 | 2024.A001 | 2024.A006 | 2024.A008 | 2024.0007 |
|---|---|---|---|---|---|---|---|---|
| | | affected | affected | affected | affected | affected | affected | affected |
| DXS454 | 97.9 | 141 | 141 | 141 | [illegible] | 143 | 141 | 141 |
| DXS1231 | 98.5 | 204 | 204 | 204 | [illegible] | 204 | 204 | 204 |
| DXS8089 | 100.6 | 97 | 97 | 97 | [illegible] | 93 | 97 | 97 |
| DXS1106 | 102.6 | 122 | 122 | 122 | 122 | 128 | 124 | 122 |
| DXS8055 | 114.6 | 315 | 315 | [illegible] | 315 | 313 | 315 | 315 |
| DXS8064 | 117.2 | 209 | 209 | [illegible] | 209 | 213 | 213 | 209 |
| DXS1001 | 119.7 | 196 | 196 | [illegible] | 196 | 198 | 196 | 196 |
| DXS1047 | 128.9 | 163 | 163 | [illegible] | 163 | 159 | 163 | 163 |

**Figure 3.6: Segregation analysis for the exclusion of the Xq region between markers *DXS8089* and *DXS8055*.** (A) A sub-pedigree of Family 2024 with the additional affected individuals and their haplotypes. (B) Xp haplotypes lined up for comparison.

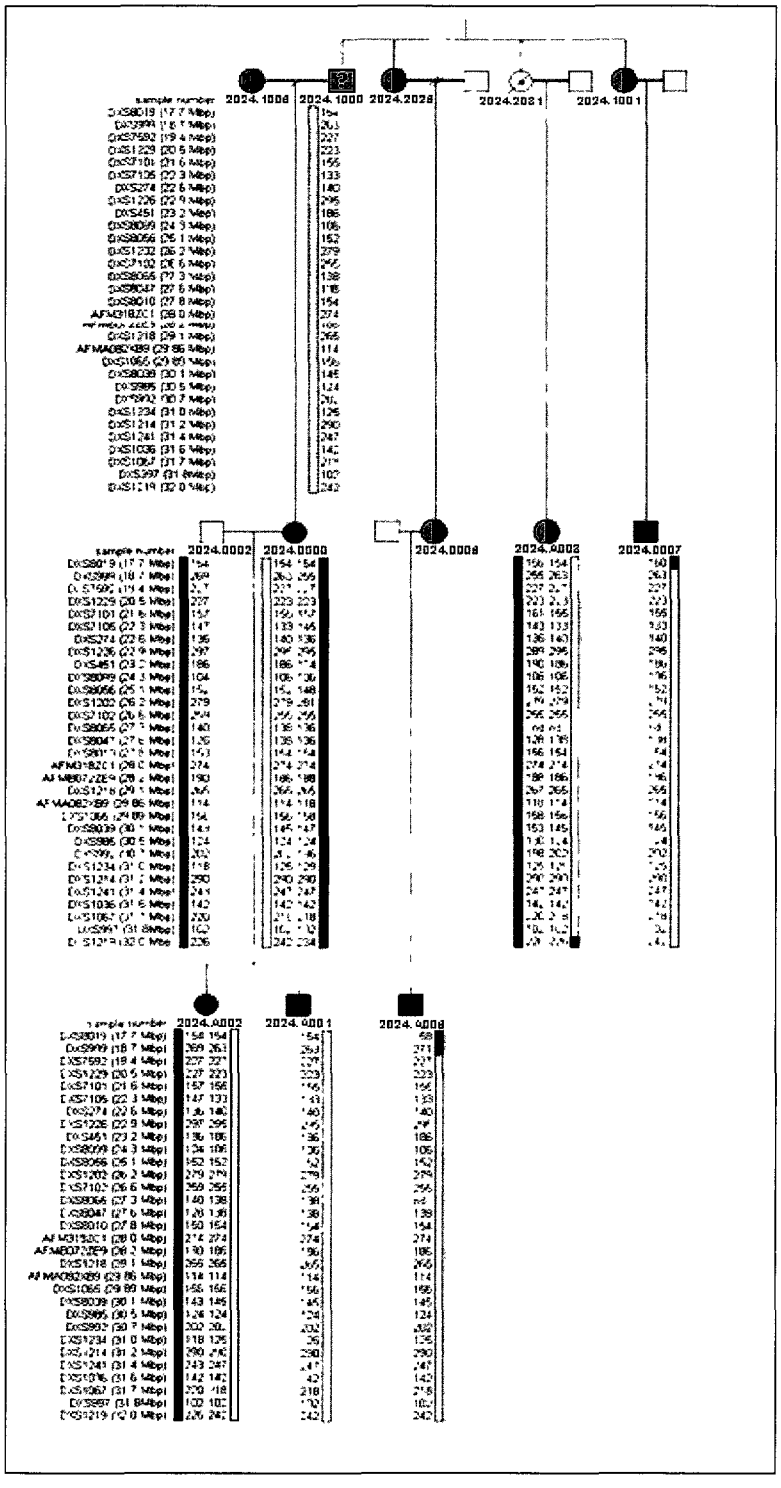


**Figure 3.7: The 13.3 Mb region between markers *DXS999* and *DXS1219* on Xp that is shared by all affected individuals.**

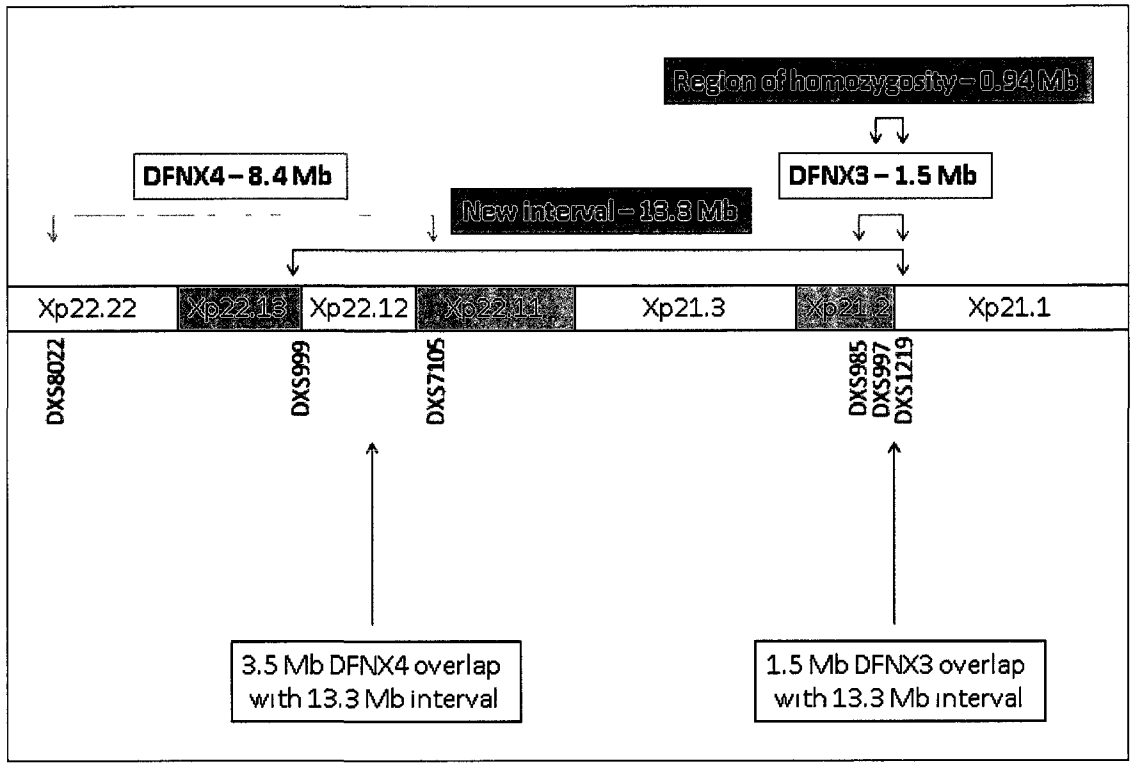



**Figure 3.8: The regions of overlap of *DFNX* Xp loci.** The 13.3 Mb region overlaps 3.5 Mb with the *DFNX4* loci and 1.5 Mb with the *DFNX3* loci. The region of homozygosity in 2024.0000 may potentially reduce the disease interval to 0.94 Mb, which only overlaps with *DFNX3*.

# Chapter 4: Newfoundland - a potential population for novel gene discovery for hereditary breast cancer

## 4.1 Introduction

### 4.1.1 General breast cancer statistics

One in nine Canadian women will develop breast cancer at some point during their life (11.1% life-time risk). In 2010, breast cancer was the most commonly diagnosed cancer among Canadian women and the second leading cause of death (281). Most cases of breast cancer are sporadic, however, approximately 15% of cases have some type of familial aggregation and another 5% have an obvious family history (282).

### 4.1.2 Breast cancer genetics

Hereditary breast cancer is both locus and allelic heterogeneous. More than 10 genes have been associated with an increase risk of breast cancer. These genes have been identified through traditional linkage approaches, candidate gene association studies and genome-wide association studies, and can be classified based on penetrance (283-286). *BRCA1* and *BRCA2* are two highly penetrant breast cancer susceptibly genes that were discovered in the mid-1990s (212, 287), and can confer up to a 20-fold increased risk of breast cancer, depending on the specific mutation (288). Attempts to identify additional highly penetrant breast cancer genes were unsuccessful for approximately 15 years, however, *RAD51C* was just recently acknowledged as the third highly penetrant breast cancer gene by studying German families with hereditary breast and ovarian cancer (289).

Before the *RAD51C* gene discovery, during the seemingly fruitless efforts towards the discovery of additional high penetrant breast cancer genes, the involvement of BRCA1 and BRCA2 in DNA repair became more evident and candidate gene association studies that concentrated on other DNA repair genes identified several moderate-risk breast cancer genes (*CHK2*, *ATM*, *BRIP1*, *NBS1*, *RAD50* and *PALB2*) (37, 290-294). The breast cancer variants in the aforementioned moderate-risk breast cancer genes are rare and a single mutation in either of these genes is enough to increase breast cancer risk (286); in addition, it was recently determined that common alleles in these genes are unlikely to increase breast cancer risk, either individually or in combination with each other (295). Furthermore, genome-wide association studies have identified common low penetrant variants in at least 12 susceptibilty loci that have been associated with breast cancer risk (296-303). These variants indicate a statistically significant risk for breast cancer risk but occur at a high frequency in the general population and, thus, are frequently found in controls (284). The exact causal variants and biological mechanisms underlying most of these common allele associations are unknown; further investigation is required. Finally, mutations in the genes that encode the tumor protein 53 (p53) and the phosphatase and tensin homolog (PTEN) are associated with rare cancer syndromes that confer a high risk of breast cancer (304-307).

### 4.1.3 High penetrant genes

#### (i) *Breast cancer susceptibility gene 1* (*BRCA1*) [MIM #113705]

In 1990 breast cancer susceptibility was first linked to a locus on chromosome 17q in 26 high risk hereditary breast cancer families (308). These families shared characteristic features commonly associated with familial breast cancer: (i) younger age at diagnosis, (ii) frequent bilateral disease, (iii) the presence of ovarian cancer and (iv) frequent occurrence of the disease amongst men. Early-onset families obtained an accumulative LOD score of 5.98 for linkage of breast cancer susceptibility to marker *D17S74*, whereas negative LOD scores were obtained in families with late-onset disease (308). *BRCA1* was first identified in 1994, by detecting mutations in five of eight kindreds whose breast cancer was previously linked to the *BRCA1* locus (212). The five mutations consisted of loss of function alleles, including (i) an 11-base pair deletion, (ii) a 1-base pair insertion, (iii) a stop codon, (iv) a missense substitution, and (v) an inferred regulatory mutation (212). *BRCA1* is located on chromosome 17q21 and codes for a 7.5 kb mRNA transcript and an 1863 amino acid protein. It has 24 exons, 22 of which are coding. Exon 11 represents 60% of the entire gene sequence (309).

The BRCA1 protein is a tumor suppressor that plays critical roles in DNA repair, cell cycle checkpoint control, and maintenance of genomic stability (310-312). The *C*-terminus of the protein has two BRCT (BRCA1 C-terminal) repeats that bind pSPxF phospho-amino acid motifs, and the *N*-terminus has a RING (Really Interesting New Gene) domain that binds to BARD1 (breast cancer associated risk domain 1) [MIM #601593], which forms an E3 ubiqutin ligase (312-315). There are at least three distinct complexes that BRCA1 forms: BRCA1-A, BRCA1-B, and BRCA1-C. Different adaptor

proteins bind the pSPxF motifs to form each unique complex. The adaptor proteins are Abraxas (Abra1) for the BRCA1-A complex, BRIP1 (BRCA1-interacting protein 1) for the BRCA1-B complex, and CTIP (another BRCA1 interacting protein) for the BRCA1-C complex (316-318).

### (ii) *Breast cancer susceptibility gene 2* (*BRCA2*) [MIM #600185]

The *BRCA2* locus was identified in 1994 through a genome wide scan analysis performed on 15 high-risk breast cancer families who were previously unlinked to the *BRCA1* locus (319). The critical region was a 6 cM interval on chromosome 13q12-13 (319). A year later, *BRCA2* was cloned using yeast artificial chromosome and P1 artificial chromosome contigs to identify trapped exons within the region, by focusing on a 600-kb interval centered around *D13S171* where the gene was thought to be located. Six different germ-line mutations in the *BRCA2* gene were identified (287). *BRCA2* codes for a 10.5 kb transcript. It has 27 exons, and exon 11 represents 50% of the coding sequence (309).

The BRCA2 protein is involved in DNA double-strand break repair in homologous recombination. In contrast to BRCA1, BRCA2 is not involved in non-homologous recombinations. Thus, BRCA2 has some specific function in the regulation of homologous recombination, which may ultimately maintain genomic integrity and suppress tumor development in cells (320). In 2004, biallelic mutations in *BRCA2* were reported to cause Fanconi anemia (FA) [MIM #227650], which is an autosomal recessive disorder characterized by aplastic anemia, cancer susceptibility and cellular sensitivity to

DNA-crosslinking agents. It is also associated with cardiac, renal, and limb malformations, as well as pigmentation changes (321, 322). There are 12 different sub-types of FA that are divided into complementation groups (FANC-A, B, C, D1, D2, E, F, G, I, J, L, and M) based on the genetic cause. BRCA2 causes FANC-D1. Eight FA proteins (FANC-A, -B, -C, -E, -F, -G, -L and -M) and three non-FA proteins (FAAP100, FAAP24 and HES1) form the FA nuclear core complex (286, 322), which is required for monoubiquitination of the FANC-D2/FANC-I dimer upon DNA damage (322). Monoubiquitinated FANC-D2 then translocates to the site of damaged DNA where BRCA1, BRCA2 (FANC-D1), and RAD51 are located and allows the repair mechanism to commence (286). Of the 12 FA sub-types, mutations in *BRCA2* cause a unique form of the disease, which involves an early-onset of acute leukemia. Wagner et al. identified seven affected FA children from five kindreds with biallelic mutations in the *BRCA2* gene, six of whom had leukemia (321). Leukemia occurred at a median of 2.2 years of age in these *BRCA2* patients, in contrast to a median onset of 13.4 years in all other Fanconi anemia patients. Breast cancer was also noted in four of the five kindreds (321).

### 4.1.4 *BRCA1* and *BRCA2* contribution to familial breast cancer and penetrance

Estimates of the proportion of all hereditary breast cancer cases attributable to *BRCA1* and *BRCA2* vary, and range from $\leq$ 20% (323) to 30-50% (324). By the age of 70 years, penetrance estimates for germ-line *BRCA1* and *BRCA2* mutation carriers range from 14-87% for breast cancer and 10-68% for ovarian cancer (33-36). The wide penetrance range may be explained by mutation-specific penetrance because there is significant

evidence that age at diagnosis varies by mutation (325). However, penetrance is hard to estimate because of the low frequency of specific mutated alleles, and the often uncertain nature of family ascertainment.

### 4.1.5 Ovarian Cancer attributed to *BRCA1* and *BRCA2*

*BRCA1* and *BRCA2* confer a high risk to both breast and ovarian cancer (326). The majority of families with multiple cases of breast and ovarian cancer have inherited mutations in *BRCA1* and *BRCA2* (327, 328). Specifically, *BRCA1* and *BRCA2* are responsible for approximately half of all families containing two or more ovarian cancer cases (326). When *BRCA2* was first identified, it was noted that it did not confer as much of an elevated risk of ovarian cancer compared to *BRCA1* (287). The cumulative life-time risks of ovarian cancer associated with these genes are estimated to be 40 to 53% for *BRCA1* mutation carriers, and 20 to 30% in *BRCA2* mutation carriers (328, 329). *BRCA2* families that have a high proportion of ovarian cancers, relative to the frequency of breast cancer, tend to have mutations located within exon 11. This region is known as the "ovarian cancer cluster region" or OCCR and is between nucleotides 3035 and 6629 (330).

### 4.1.6 *BRCA1* and *BRCA2* and risks of other cancers

Mutations in *BRCA1* and *BRCA2* are mainly associated with breast and ovary cancer, however other organs can be affected (331). There is actually an elevated risk for all cancers in *BRCA1* and *BRCA2* mutation carriers, and the organs at risk differ between

families. Specific examples of additional organs at risk include the stomach, pancreas, prostate, and colon, with the stomach and pancreas having the highest increased risk. For any particular cancer the increased risk ranges from about 20 to 60%. The normal functioning BRCA1 and BRCA2 proteins must be involved in the protection against a variety of cancers, thus the pathways involved must overlap (331).

### 4.1.7 *BRCA1* and *BRCA2* disease-predisposing alleles

#### (i) Simple sequencing variants

According to the Breast Cancer Information Core (BIC) database, sequencing variants have been detected in every exon of *BRCA1*, and in all but exon 1 of *BRCA2*. There have been a total of 1643 distinct *BRCA1* variants reported, and 841 (51%) of them are in exon 11. There have been a total of 1856 distinct variants reported in *BRCA2*, and 787 (42%) of these are located in exon 11 (332).

It was originally thought that frameshift or nonsense mutations that truncate the BRCA1 and BRCA2 proteins were the most common breast cancer predisposing alleles. When looking at BIC and noting the number of entries of the top 20 recorded variants for each variant type (frame shift, nonsense and missense), there are a total of 5805 frameshift and 1309 nonsense mutation entries in *BRCA1*, and a total of 3436 frameshifts and 996 nonsense mutation entries reported in *BRCA2* (332). However, missense variants are the second most frequently reported variant in the database for *BRCA1* (with a total of 3559 entries), and the most frequently reported in *BRCA2* (with 6168 entries). Many *BRCA1*

and *BRCA2* missense variants are known as unclassified variants (UCVs) since their affect on protein function and disease pathogenesis is unknown; determining if these variants are neutral or disease-causing is very important (333-335). There are different ways to help classify such variants including functional assays (335), segregation analysis and observed co-occurrence with known pathogenic mutations (334).

### (ii) Genomic rearrangements leading to copy number variants in *BRCA1*

It was not until three years after the discovery of *BRCA1* and the identification of hundreds of point mutations and small insertions/deletions (336, 337) that the first *BRCA1* rearrangement was reported (338). This rearrangement was discovered in a large American family with a multipoint LOD score of 3.62 for two markers that flanked the *BRCA1* locus; originally, no mutations were detected after screening the coding sequence, thus further analysis revealed a deletion of exon 17 (338). This three year delay was attributed to primarily relying on PCR-based techniques for mutation detection, which does not readily detect copy number variations. In fact, rearrangements contributing to the *BRCA1* mutation spectrum only became apparent after sequencing the whole gene and determining that a very high density of Alu sequences, which could possibly act as hotspots for unequal homologous recombination, were located within the gene region (339).

Over 30 different *BRCA1* germ-line rearrangements have been reported, the majority are deletions but also included are duplications, triplications, and a combination of both

deletion and insertion events (64). These rearrangements are scattered throughout the gene, and every exon (except the terminal exon 24) has been involved in at least one rearrangement. Some rearrangements involve only one exon while others involve a series of exons, the largest of which includes a 22 exon rearrangement. The rearranged regions range in size from 510 nucleotides (deletion of exon 22) to 160,880 nucleotides (deletion of exons 1 to 22). *BRCA1* rearrangements larger than 10 Mb have never been reported - perhaps they are lethal (64). *BRCA1* rearrangements account for approximately 10 to 20% of all *BRCA1* mutations in the general population (64). For example, it has been estimated that 8% of all French *BRCA1* mutations and 12% of all German *BRCA1* mutations are due to genomic rearrangements (340, 341).

### (iii) Genomic rearrangements leading to copy number variants in *BRCA2*

*BRCA2* genomic rearrangements appear to be less common than *BRCA1* genomic rearrangements, and estimating their frequency in the overall *BRCA2* mutation spectrum has been difficult. The first rearrangement was reported in 1996 when an insertion of an Alu sequence into exon 22 was noted (342). In 1998, a 5 kb deletion of the 3′ end of exon 3 and most of intron 3 was detected, which led to the skipping of exon 3 in the mRNA (343), and in 2001, a 6.2 kb deletion that resulted in the loss of exons 12 and 13, and an insertion of approximately 60 adenine residues was reported (344). Since then, only a few other *BRCA2* rearrangements have been reported. For instance, a deletion of exons 1 to 13 in a high risk US family (345), and an exon 20 deletion in a Danish family (346). However, overall these rearrangements do not seem to be common.

### 4.1.8 *BRCA1 and BRCA2* founder mutations

Founder mutations in *BRCA1* and *BRCA2* have been intensely studied. An example of a founder mutation is the *BRCA2* c.999del5 mutation in the Icelandic population; interestingly no other *BRCA2* mutation has been reported in this population. The c.999del5 mutation is approximately 20 times more prevalent in Iceland than the estimated allele frequency of *BRCA2* in the general worldwide Caucasian population (347, 348), and it proved to be the cause of 76% (16/21) of female breast cancer in high-risk breast cancer families (349). In 632 Icelandic breast cancer cases unselected for a family history, the *BRCA2* c.999del5 mutation was detected in 7.7% of female breast cancer diagnosed at any age, and 24% of those diagnosed at age 40 years or younger (348).

Another example is the three founder mutations that have been observed in Ashkenazi Jewish breast and ovarian cancer patients. In the Ashkenazi Jewish population, the *BRCA2* c.6174delT mutation has a frequency of 0.9-1.5%, the *BRCA1* c.185delAG mutation has a frequency of 0.8 to 1.1%, and the *BRCA1* c.5382insC has a frequency of 0.13-0.3%. The population prevalence for these three mutations combined is 2-2.5%, which is approximately 10-50 times higher than the allele frequency in the general world population (350-352). For breast cancer cases unselected for family history, these mutations account for 30% of those diagnosed at less than 40 years of age (353).

There are at least six *BRCA1* rearrangements that are founder mutations (64). In a study of 805 Dutch breast cancer families two mutations, the deletion of exon 13 and the deletion of exon 22, were shown to represent 23% of all *BRCA1* mutations identified (354). The Rotterdam Family Cancer Clinic determined that 20 of the families that carried the exon 13 deletion all originated from a small, isolated, southwestern region of the Netherlands and likely shared a common ancestor (355). In such founder populations estimates for the frequency of rearrangements contributing to the overall number of *BRCA1* mutations is higher than in other populations. All *BRCA1* rearrangements account for 27% of Dutch *BRCA1* mutations (354).

### 4.1.9 Non-*BRCA1/2* families

In regard to families in which a hereditary component is suspected but no mutation in *BRCA1* or *BRCA2* is detected, negative results may be due to many factors, which include (i) selecting the wrong individual in a *BRCA1/2* family for screening (perhaps a phenocopy was chosen), (ii) regulatory element *BRCA1*/2 mutations, or (iii) epigenetic phenomena that inhibits *BRCA1*/2 gene expression. Another possibility (iv) is the presence of mutations in other genes (333).

### 4.1.10 Rare low to moderate penetrant breast cancer genes

Rare mutations in less penetrant breast cancer susceptibility genes that function in DNA repair have been associated with at least doubling the risk of breast cancer (286).

### (i) *Check point kinase 2* (*CHK2* or *CHK2*) [MIM #604373]

CHK2 is a cell-cycle checkpoint kinase that is in the same signaling pathway as BRCA1 and p53, and is activated in response to DNA damage. It was the first low penetrant breast cancer gene to be discovered. The mutation, c.1100delC, is a low penetrant breast cancer predisposition allele that was identified by studying a single breast cancer family (290). A genome-wide linkage search in a large non-BRCA1/2 family revealed the highest LOD score (Z=1.2) at a region on chromosome 22q, between *D22S1150* and *D22S928*. A haplotype was created that showed partial segregation with breast cancer. The *CHK2* gene was a positional candidate and considered a plausible candidate gene because of its known function. After screening, *CHK2* c.1100delC, a truncation mutation that abolishes its kinase activity was detected. This variant was present in 5.1% of individuals with breast cancer from 718 non-*BRCA1/2* families and 13.5% of individuals from families with male breast cancer. It, however, has a frequency of 1.1% in healthy individuals. It is estimated that *CHK2* c.1100delC results in an approximately two-fold increase of breast cancer risk in women and a ten-fold increase of risk in men (290). This variant does not increase the risk of breast cancer in *BRCA1* or *BRCA2* mutation carriers. CHK2 activates BRCA1 therefore mutations in *BRCA1* override the elevated breast cancer risk from *CHK2* mutations (290). In 2006, another *CHK2* variant associated with a similar breast cancer risk was identified (356). A novel 5.6 kb genomic deletion of the *CHK2* gene was discovered in two Czechoslovakian families. This deletion was found in 8 of 631 (1.3%) patients with breast cancer and in zero of 367 healthy controls in the

Czech and Slovak Republics (356). Mutations in *CHK2* are estimated to account for approximately 1% of all breast cancers (357).

### (ii) *Ataxia-telangiectasia mutated gene* (*ATM*) [MIM #607585]

The ATM protein is a member of the phosphatidylinositol 3-kinase protein family that phosphorylates its substrates upon response to DNA damage. Biallelic mutations in the *ATM* gene cause ataxia-telangiectasia (AT) [MIM #208900], a rare, childhood neurological disorder that causes brain degeneration in areas that controls motor movements and speech (358). Approximately 20% of people diagnosed with AT also develop cancer. Heterozygous mutations in *ATM* have been thought to increase breast cancer risk, and recently 12 mutations in affected breast cancer individuals have been identified after studying 443 familial breast cancer pedigrees. Heterozygous carriers have an estimated relative risk of 2.37 (37).

### (iii) *BRCA1-interacting protein 1* (*BRIP1*) [MIM #605882]

In 2001, *BRIP1* mutations were identified in two of 65 patients with early-onset breast cancer, 35 of whom had a strong family history of breast and/or ovarian cancer and lacked mutations in either the *BRCA1* or *BRCA2* genes. The mutations were not found in 200 matched controls and as such, it was suggested that mutated *BRIP1* might cause hereditary breast cancer (359). BRIP1 encodes a helicase that interacts with the BRCT domain of BRCA1 and has BRCA1-dependent DNA repair and checkpoint functions

(359). It is also an FA protein (FANCJ), thus biallelic mutations in *BRIP1* have been determined to cause Fanconi anemia (360).

In 2006, Seal et al. screened *BRIP1* in breast cancer patients that screened negative for *BRCA1* and *BRCA2* (292). The full coding sequence and intron-exon boundaries of *BRIP1* were screened in 1,212 women with familial breast cancer and 2,081 controls. Five different truncating mutations were detected in nine of the 1,212 affected women. The relative risk for breast cancer associated with truncating mutations in *BRIP1* was determined to be 2. Due to the low relative risk, like all low penetrant alleles, there was limited evidence of linkage in the *BRIP1*-positive pedigrees. It was estimated that *BRIP1* mutations attribute to 0.20% of all breast cancers (292).

### (iv) *Nijmegen breakage syndrome 1* (*NBS1*) [MIM #602667]

The NBS1 protein plays a crucial role in the detection and repair of chromosome breaks. The *NBS1* gene was first discovered as the causative gene for Nijmegen breakage syndrome [MIM #251260] and hence bears the name of the disease (361). A homozygous mutation, a deletion of five nucleotides in exon 6 (c.657del5), was the most common variant identified (361).

The same *NBS1* mutation also had a significant, but moderate, age-related risk for breast cancer. In populations with a high c.657del5 carrier frequency this mutation is thought to contribute substantially to the overall incidence of breast cancer, particularly in younger

age groups (294, 362). Gorski et al. studied three groups of patients from Poland: (i) 150 consecutive breast cancer patients who were diagnosed under the age of 50 and had histological confirmed breast cancer, two of which (1.3%) had the *NBS1* mutation, (ii) 80 breast cancer patients with a family history of breast cancer in their first-degree relatives, three of which (3.7%) had the mutation, and (iii) 530 randomly and consecutively doctor-selected control individuals without the diagnosis of breast cancer, three of which carried the mutation, giving it a 0.6% frequency in the general population (362). Steffen et al. studied 562 non-selected breast cancer patients from Central Poland and found 11 (1.96%) 657del5 mutation carriers. They estimated the risk of c.657del5 mutation carriers for breast cancer onset according to age; less than 40 years: 8.36; less than 50 years: 4.27; less than or equal to 50 years: 2.40, and for all ages: 3.13 (294).

More recently, a missense mutation in the *NBS1* gene, p.I171V, has been associated with a nine-fold increased risk of breast cancer in Polish patients (363, 364). In a group of 270 women with breast cancer, seven cases of mutated *NBS1* were revealed, five of which carried the mutation p.I171V in exon 5. The rate of p.I171V mutation in the group of breast cancer patients was significantly higher than in the controls (364).

### (v) *RAD50* [MIM #604040]

RAD50 is a part of the Mre11 protein complex, which is important for recombination, repair, and genomic stability. *RAD50* was screened for mutations in 151 northern Finnish breast cancer families as part of a process to determine if any Mre11 complex genes

(*Mre11A* [MIM #600814], *RAD50* and *NBS1*) had a predisposition to breast cancer (365). A truncation variant, c.687delT, was detected in two of these families. Due to the detection of this variant in unaffected individuals (ranging from 35-81 years of age) and because it had a prevalence of 0.6% in the general population the authors suggested the allele had a low penetrance (365). A later case-control study in the Finnish population revealed that the same truncating *RAD50* variant increased breast cancer risk by four-fold (366). Eight out of 317 consecutive, newly diagnosed northern Finnish breast cancer patients carried the mutation (366).

### (vi) *Partner and Localizer of BRCA2* (*PALB2*) [MIM #610355]

PALB2 promotes BRCA2 localization and stability, which enables it to properly function in recombination repair (367). It was recently determined that PALB2 is also an FA protein and biallelic mutations in *PALB2* cause FA (368). The FA phenotype caused by malformations in *PALB2* is similar to the sub-type caused by biallelic mutations in *BRCA2* (368). Thus Rahman et al. investigated whether monoallelic *PALB2* mutations confer susceptibility to breast cancer by sequencing 923 breast cancer patients who had a family history of the disease and were *BRCA1* or *BRCA2* negative, and 1,084 controls (293). Monoallelic truncating *PALB2* mutations were identified in 10 patients and in no controls. It was also determined that such mutations confer a 2.3-fold higher risk of breast cancer (293). A highly penetrant *PALB2* allele (c.3113G>A (W1038X)) has been detected in the Australian population, solving a total of 13 breast cancer families (369). The mutation segregated with breast cancer in these families and was not found in 764

population-based unaffected controls. The corresponding cumulative risk estimates were 49% at age 50 and 91% at age 70 (369).

### 4.1.11 Moderately penetrant genes with founder mutations

Founder mutations have also been observed in the known moderately penetrant genes. For example, the truncating *RAD50* variant c.687delT is thought to be a founder mutation in the Finnish population. The variant has a relatively high frequency in the Finnish population but is not observed in other Nordic cohorts such as Sweden, Norway and Iceland. The same haplotype was observed in 14 carriers, suggesting that c.687delT originated in Finland from a common ancestor (366). The *PALB2* mutation c.1592delT has also been reported to be a founder breast cancer Finnish mutation. This mutation has a significantly elevated frequency in the Finnish population and is estimated to represent 1% of all breast cancer cases, which is remarkable considering that *BRCA1* and *BRCA2* together account for about 1.8% of all Finnish breast cancer cases (370).

### 4.1.12 *RAD51C* [MIM #602774]- a third highly penetrant breast cancer gene

The candidate gene approach has been proven successful, particularly for identifying rare alleles that increase breast cancer risk moderately. In general, the genes were chosen for mutation screening based on their involvement in DNA repair, similar to *BRCA1* and *BRCA2*. Biallelic mutations in two of the moderate-risk genes, *BRIP1* and *PALB2*, along with *BRCA2* cause FA. Most recently, *RAD51C* was chosen for candidate gene screening because a biallelic missense mutation in the *RAD51C* gene was recently determined to be

the cause of a Fanconi anemia–like phenotype in a consangiunous Pakistan family (371). The screening of index cases from 1,100 German families with gynecological malignancies for *RAD51C* variations identified six monoallelic pathogenic mutations (two frameshift-causing insertions, two splice-site mutations and two nonfunctional missense mutations) that confer an increased risk for breast and ovarian cancer (289). The mutations were detected in only breast/ovarian cancer families, and were not in 620 families with breast cancer only or in 2,912 healthy German controls. All *RAD51C* mutations segregated perfectly with the disease, unlike the alleles of the identified moderate-risk genes. The authors suggest the identification of *RAD51C* as cancer susceptibility gene supports the 'common disease, rare allele' hypothesis (289). A follow up report by Akbari et al. sequenced an additional 454 familial breast/ovarian cancer cases (of various ethnic groups) that previously screened negative for *BRCA1/BRCA2* screening and found no deleterious mutations. They suggested that *RAD51C* mutations may not be as common as the initial report suggested (372).

### 4.1.13 Common low penetrant variants with increased risk of breast cancer

Genome-wide association studies have identified common low penetrant variants in at least 12 susceptibilty loci that have been associated with breast cancer risk (296-303). The exact causal variants and biological mechanisms underlying most of these associations are unknown; further investigation is required. *FGFR2*, however, a fibroblast growth factor receptor, which is known to be over expressed in sporadic breast cancer cases (373), has two SNPs in intron 2 (rs2981582 and rs1219648) that, in two

independent multicenter genome-wide association studies, have been associated with an increased risk of breast cancer with odds ratios of 1.23-1.24 in heterozygotes (296, 297). These SNPs are, however, present in 47-48% of healthy control cohorts, which makes it difficult to use these SNPs as predictive measures of disease onset (284). This association has also recently been confirmed in an Ashkenazi Jewish three phase association study when comparing *BRCA1/BRCA2*-negative high risk breast cancer cases to healthy controls (374). Interestingly, the two intronic *FGFR2* SNPs seem to alter the binding affinity of particular transcription factors which increases *FGFR2* expression (375).

### 4.1.14 Syndromic breast cancer

Mutations in the tumor protein gene *p53* [MIM #191170] and in the *phosphatase and tensin homolog* (*PTEN*) gene [MIM #601728] confer a high risk of breast cancer and are associated with rare cancer syndromes. The transcription factor p53 responds to diverse cellular stresses to regulate genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism (304). Germ-line mutations in the *p53* gene cause the cancer predisposition syndrome Li-Fraumeni syndrome [MIM #151623]. Li-Fraumeni syndrome is inherited in an autosomal dominant fashion and is characterized by early onset tumors including sarcomas, breast cancer, leukemia, brain tumors, and adrenocortical carcinoma (305). PTEN is thought to play a role in preparation for DNA replication during the cell cycle (306). Germ-line mutations in the *PTEN* tumor-suppressor gene causes Cowden syndrome [MIM #158350]. It is a cancer predisposition

syndrome associated with an increased risk of breast, thyroid, and endometrial cancers, and benign tumors (307).

### 4.1.15 Identifying novel breast cancer mutations and genes

Approximately 50% of hereditary breast cancer cases remain unsolved. There may be other mutations in each of the known genes that confer a risk of breast cancer, however, it will only be a matter of time before they are identified by studying other populations, classifying detected missense variants (UCVs) as pathogenic, and evaluating other mutation types such as genomic rearrangements and regulatory malfunctions (286).

Despite that many researchers believe in polygenic breast cancer susceptibility and strive to pursue association studies that will identify low-moderate risk alleles (283), the recent identification of the highly penetrant *RAD51C* mutations supports the 'common disease, rare allele' hypothesis (289), which indicates that other very rare and highly penetrant genes may exist. Attempts to identify new highly penetrant hereditary breast cancer genes have, however, been relatively unsuccessful. There have been several genome-wide linkage scans performed on multiple non-*BRCA1/2* breast cancer families (376-379). Huusko *et al.* studied non-*BRCA1/2* Finnish families and found linkage at marker *D2SS2262* (2q32) (378), and Bergman *et al.* studied non-*BRCA1/2* Swedish families and reported linkage at 10q23.32-q25.2, 12q14-q21 and 19p13.3-q12 (376). Another study was performed by the Breast Cancer Linkage Consortium using 149 non-*BRCA1/2* families, and despite the large number of families used, no significant linkage was found

(379). Gonzalez-Neira *et al.* more recently identified five suggestive loci related to hereditary breast cancer with moderate linkage values at 2p22.3, 4p14q12, 7q21.11–7q21.3, 11q13.5–11q14.3 and 14q21.1–14q21.3 (377). No causative genes have been identified in the areas of suggestive linkage. Regarding future studies, it has been suggested to use families from a homogeneous population to reduce genetic variation (377). It is possible that other homogeneous populations have founder mutations in known or yet undiscovered breast cancer genes; and the new, next-generation sequencing technology may ease their identification. Hereditary breast cancer has not been studied in the Newfoundland population.

### 4.1.16 Aims of this study

The goal of this study was to evaluate, for the first time, familial breast cancer cases in Newfoundland in order to find cases attributed to *BRCA1*, *BRCA2* and *CHK2*, and to identify families suitable for novel gene discovery in breast cancer.

## 4.2 Materials and Methods

### 4.2.1 Recruitment

This project was approved by the Human Investigation Committee, Faculty of Medicine, Memorial University (Reference #05.97). Probands were ascertained through the Newfoundland Provincial Medical Genetics Program since 1991. After giving consent, a blood sample was taken from the proband and used to extract DNA and, in some cases, establish a cell-line. The proband's family history was also recorded and the family was categorized as high-risk (four or more cases in the family) or moderate-risk (two to three cases in the family). A total of 153 probands were recruited into the study, 75 of whom were from high-risk families and 78 from moderate-risk families (Table 4.1).

The project was divided into two phases (Figure 4.1). Phase 1 was completed in Seattle, Washington by Dr. Terry-Lynn Young, and the results of this phase are presented in Appendix 10. In Phase 1, the first 96 recruited probands (45 high-risk families and 51 moderate-risk families) had full gene screening for the detection of point mutations in both *BRCA1* and *BRCA2* (Table 4.1 and Figure 4.1). Sixty-three of those families had only breast cancer cases (23 high-risk, 40 moderate-risk) and 33 families had cases of both breast and ovarian cancer (22 high-risk, 11 moderate-risk). Nine of the 96 families had at least one male breast cancer case. The majority of the probands recruited in this phase were in their forties (41/96) (Table 4.1).

Phase 2 was carried out at Memorial University as a part of my graduate studies. In phase 2, 57 of the more recently recruited probands (30 from high-risk families and 27 from moderate-risk families) had targeted mutation screening based on the mutations found in phase 1 (Table 4.1 and Figure 4.1). Thirty-six of those families had breast cancer cases only (17 high-risk and 19 moderate-risk), and 21 families had cases of both breast and ovarian cancer (13 high-risk, 8 moderate-risk). Five of the 57 families had at least one male breast cancer case.

### 4.2.2 *BRCA1* and *BRCA2* conventional gene screening

In phase 1, the first 96 recruited probands had conventional full gene screening for *BRCA1* and *BRCA2*. The protein truncation test (PTT) was first performed on each proband for the detection of truncation mutations in *BRCA1* exon 11, and *BRCA2* exons 10 and 11. The remaining open reading frame of *BRCA1* and *BRCA2* was screened by either single-strand conformation polymorphism (SSCP) and/or by direct sequencing.

In phase 2, 57 newly recruited probands had targeted mutation screening by direct sequencing for the 15 truncation mutations found in phase 1. These probands were recruited after phase 1 from 2002 to 2006. PTT was also performed to screen for additional truncation mutations in *BRCA1* exon 11, and *BRCA2* exons 10 and 11 since they represent over 60% of each gene's coding region (309).

### (i) Protein truncation test (PTT)

In brief, PTT involves an *in vitro* transcription mechanism and protein synthesis that enables the detection of truncated proteins. Specially designed PTT primers are used to amplify DNA and the resulting amplified DNA products are then used for *in vitro* transcription and translation to generate the protein that would be synthesized *in vivo*. These protein products are then analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Normal protein products show a particular pattern on the gel whereas truncated proteins from mutated genes are shorter, run faster and have a different pattern from that of a normal protein (380, 381).

PTT was performed by first amplifying exon 11 of *BRCA1*, and exons 10 and 11 of *BRCA2* from genomic DNA. A standard 1X PCR reaction cocktail without betaine was used for all primer sets (Appendix 1). All amplifications consisted of 35 cycles. Each cycle had a 94°C 30 second denaturing period, a 30 second annealing period at primer specific temperatures, and a 72°C extension period at primer specific times. The primer sets and amplification conditions that were used had been previously determined in the laboratory of Dr. Mary-Claire King at the University of Washington (Appendix 11).

The Promega TNT® T7 Quick kit was used to couple the transcription and translation reaction. A single reaction included, 10 µl (one fourth the recommended amount) of TNT Quick Master mix, 1 µl of [$^{35}$S]methionine, and 1.5 µl of the PCR-generated DNA template, for a total volume of 12.5 µl. The reaction mixture was incubated at 30°C for

90 minutes. After incubation, 2 µl of the translated sample was added to 20 µl of 6X loading buffer (the remaining translated sample was stored at -80°C). The 22 µl mixture was denatured at 99°C for 2 to 5 minutes then the whole volume was run out on a 12% SDS-polyacrylamide gel until the bromophenol blue dye reached the bottom of the gel. After the electrophoresis, the gel was fixed for 30 minutes in a solution of 50% methanol, 10% glacial acetic acid, and 40% $H_2O$. After fixation, the gel was soaked in an aqueous solution of 7% acetic acid, 7% methanol, and 10% glycerol for 5 minutes to prevent gel cracking. The gel was then dried on Whatman® 3MM filter paper at 80°C for 30 to 90 minutes under a vacuum seal. It was then exposed on Kodak X-OMAT® AR film up to 6 days, depending on the desired band intensity.

### (ii) Single-strand conformation polymorphism (SSCP)

SSCP takes advantage of the properties of single stranded DNA (ssDNA) during electrophoresis. After denaturation, ssDNA undergoes 3-dimensional folding and, based on its DNA sequence, may assume a distinct conformation. The conformational differences between two ssDNA strands with different sequences can cause them to migrate differently on an electrophoresis gel, even though the number of nucleotides (and molecular weight) is the same. Thus, a single nucleotide change can be detected using this approach. However, not all sequence changes are detected, which decreases its sensitivity (382).

All PCR products that were screened by SSCP were amplified using primer sets and amplification conditions that were also determined in the King laboratory (Appendix 12 and 13). A $^{32}P$ dCTP-labeled PCR was carried out. A 1X master mix included: 2.5 µl of 10X PCR Buffer, 2.5 µl of dNTPs (2 mM) (with 5% of the normal concentration of dCTP), 0.75 µl of $MgCl_2$ (50 mM), 0.2 µl of Taq DNA polymerase, 15.85 µl of $H_2O$, 1 µl of the forward primer, 1 µl of the reverse primer, and 0.2 µl of $^{32}P$ dCTP. Touchdown PCRs were generally used to amplify the desired product (Appendix 2).

The SSCP analyses were performed using 4µl of $^{32}P$ dCTP-labeled PCR product and the same amount of denaturing solution (95% formamide, 20 mM EDTA (pH 8.0), and 0.5 mg/ml bromophenol blue) for each sample. Samples were heated at 98°C for 10 minutes and immediately chilled on ice. The DNA strands were separated by electrophoresis in a non-denaturing polyacrylamide gel (10% acrylamide) at 100 V for 16 hours at 25°C.

### (iii) Automated Cycle Sequencing

The sequencing primers sets were the same as the primer sets used for SSCP (Appendix 12 and 13). A standard 1X PCR reaction cocktail without betaine was used for all primer sets (Appendix 1). Touchdown PCRs were used to amplify the desired product (Appendix 2). The desired PCR products were purified and prepared for sequencing following the procedure described in Chapter 2, section 2.2.3(i).

### 4.2.3 Founder mutation identification

Probands that shared the same disease-causing mutation were genotyped for several microsatellite markers surrounding the desired locus in order to determine if they shared a common haplotype, and thus a common ancestor. Six markers flanking the *BRCA2* locus were genotyped; *D13S1242*, *D13S1299*, *D13S1287*, *D13S260*, *D13S171*, and *D13S220* (Appendix 14). A standard PCR cocktail without betaine was used (Appendix 1). The PCR products were amplified using a general cycling program with a 50°C annealing temperature (all steps were 30 seconds in duration). Genotyping was performed by fluorescent-based PCR (ABI 3100, dye set DS-33).

### 4.2.4 Screening for *BRCA1* and *BRCA2* copy number variants (genomic duplications and deletions)

#### (i) Amplification across the breakpoints of known variants

Targeted mutation screening can be performed on known copy number variants when PCR-based applications have been developed to amplify across the specific breakpoints. Six *BRCA1* variants were selected for targeted screening: (i) a deletion of exon 3 (383), (ii) a duplication of exon 13 (384), (iii) a deletion of exons 14-20 (385), (iv) a deletion of exon 20 (386), (v) a deletion of exons 20-22 (354), and (vi) a deletion of exon 22 (387). For information on each primer set and amplification conditions see Appendices 1, 2 and 15. Only probands recruited in phase 2 were screened using this method.

### (ii) Multiplex ligation-dependent probe amplification (MLPA)

Copy number variants in *BRCA1* and *BRCA2* can also be detected by MLPA. MLPA probes mixes P002 and P045 have been used for the detection of copy number variants in *BRCA1* and *BRCA2*, respectively. Only probands recruited in phase 2 were screened using this method. The MLPA step-by-step protocol for DNA detection/quantification is available at www.mlpa.com. See Chapter 3, section 3.2.7 for a description of the standard protocol used for a 1X MLPA reaction.

## 4.2.5 *CHK2* c.1100delC screening

### (i) Restriction test

All 153 probands were screened for the *CHK2* variant c.1100delC. Since pseudogenes are present in the *CHK2* locus, a nested PCR was created to increase specificity and amplify the desired *CHK2* fragment. The external and internal primer sets and amplification conditions can be seen in Appendix 16. The initial PCR using the external primer set involved a standard PCR cocktail without betaine (Appendix 1). The PCR products were amplified using a TD 60 cycling program. A $H_2O$ control was used in this reaction. The nested PCR with the internal primer set also used a standard PCR cocktail without betaine, but instead of genomic DNA as the template, the PCR product was used (Appendix 1). This PCR was carried out using the TD 50 cycling program. The PCR product of the $H_2O$ control was carried over to the nested PCR as well to ensure there were no contamination issues.

A BsrI restriction site is disrupted by the c.1100delC mutation. The undigested amplicon for this experiment was 124 bp. After digestion the wild-type allele was cut two times to give three bands (74 bp, 31 bp, and 19 bp), and the mutant allele was cut only one time to give two different bands (93 bp, and 31 bp). The digest reaction mixture contained 5 µl of the nested PCR product, 2 µl of 10X Buffer 3, 0.5 µl of BsrI (5U/µl) and 12.5 µl of $H_2O$. The reaction was carried out at 65°C overnight. A 2% agarose gel was used for the electrophoresis.

**(ii) MLPA**

The *CHK2* variant could also be detected by using the *BRCA2* MLPA probe mix P045. A probe in this kit was designed such that amplification would occur only when a DNA sample contained the *CHK2* c.1100delC mutation. Only probands recruited in phase 2 were screened using this method.

## 4.3 Results

### 4.3.1 Phase 1 – *BRCA1* and *BRCA2* full gene screening

See Appendix 10.

### 4.3.2 Phase 2 – Targeted mutation screening in newly recruited probands

Fifty-seven probands that were recruited after phase 1 had targeted mutation screening for the 15 Newfoundland *BRCA1* and *BRCA2* mutations that were detected during phase 1 (Appendix 10 and Figure 4.1).

#### (i) Positive screen for *BRCA2* c.6714delACAA – founder mutation

Only one of the 57 probands screened positive for a known mutation. That proband had the *BRCA2* mutation c.6714delACAA (Figure 4.2). *BRCA2* c.6714delACAA was first detected (phase 1) in the proband of Family 199. The proband was female and was diagnosed with breast cancer at the age of 59. There were three total cases of breast cancer in that family, all of which were diagnosed over the age of 50 in females. There were two other cancer cases in this family, one lung and one skin (lip) cancer (Appendix 10, Table A10.1). The newly recruited proband that screened positive for *BRCA2* c.6714delACAA was from Family 1191. The proband was female and diagnosed with breast cancer at 37 years of age. She also had a secondary ovarian cancer. The sister of the proband was also diagnosed with breast cancer under the age of 50. The proband had two brothers diagnosed with cancer, one brother had throat cancer and the other had

pancreatic cancer. Both the extended maternal and paternal sides of the pedigree had several cancer cases, so it is unknown which side the mutation segregated from. These two families were not known to be related but after genotyping markers that flanked the *BRCA2* locus, a common haplotype was constructed. Haplotypes were generated without knowing phase so further recruitment is needed to confirm this finding. However, it appears that the two probands share a 6.02 Mb haplotype on chromosome 13 between markers *D13S1242* and *D13S220*. This data suggests that *BRCA2* c.6714delACAA may be a founder mutation (Figure 4.2).

After performing PTT on the 57 newly recruited probands no additional Newfoundland truncation mutations were detected in *BRCA1* exon 11 and *BRCA2* exons 10 and 11 (Figure 4.3).

### 4.3.3 Screening for *BRCA1* and *BRCA2* copy number variants

No copy number variants were detected in the phase 2 probands. When amplifying across the breakpoints of the six common variants in *BRCA1*, the band representing the mutant allele of the positive controls amplified each time (Figure 4.4), but was not observed in any tested probands. MLPA was the second method used to screen for copy number variants, which screened all exons of *BRCA1* and *BRCA2*. The quality of the results using this method was variable. Several attempts were made to validate the results however this was not 100% successful. In the unsuccessful cases, the electropherogram peak heights, which would normally quantify the copy number of the corresponding exon,

were too low to make reliable calls. Of the 57 probands screened by MLPA, 43 and 39 were successfully screened for *BRCA1* and *BRCA2*, respectively. No variants were detected in any of those probands, and the positive controls worked well each time (Figure 4.5).

### 4.3.4 *CHK2* c.1100delC targeted screening

All 153 probands from phase 1 and 2 were screened for the *CHK2* c.1100delC mutation (Figure 4.1). This mutation was found in 2% of the probands studied (3/153) (Table 4.2 and Figure 4.6). Two of the probands were from phase 1 and one was from phase 2. All three probands were female and their primary cancer was in fact breast cancer. All were 52 years of age at disease onset. Two of the three probands had a second primary cancer, one had multiple myeloma and the other had ovarian cancer. There were a total of seven breast cancer cases in all three families, and all were diagnosed over fifty years of age. Only one family showed instances of ovarian cancer. All three families had other cancers associated with the disease but there were no male breast cancer cases observed (Table 4.2).

## 4.4 Discussion

The purpose of this study was to evaluate familial breast cancer cases in Newfoundland in order to find cases attributed to deleterious variants in *BRCA1*, *BRCA2* and *CHK2*, and to identify families suitable for novel gene discovery in breast cancer. Fifteen *BRCA1* and *BRCA2* truncation mutations that contribute to both high- and moderate-risk cases of breast cancer in the province were identified in 15 different probands during phase 1, explaining 15.6% (15/96) of the hereditary breast cancer families that had full gene screening. *BRCA1* mutations explained 8.3% of those families and *BRCA2* explained 7.3%. The average age of diagnosis was 41.8 and 47.9 years for a *BRCA1* and *BRCA2* proband respectively.

The number of breast cancer families with *BRCA1* and *BRCA2* mutations varies in different studied populations. In addition to this, the spectrum of mutations in each population can differ as well (324, 388), and the proportion of *BRCA1* and *BRCA2* families does not correlate with the number of different mutations in a population (388). This Newfoundland cohort had a spectrum of 15 *BRCA1* and *BRCA2* mutations that explained 15.6% of the probands studied. In comparison, Italy, for example, has a fairly high percentage (23%) of its breast cancer families attributed to *BRCA1* and *BRCA2* mutations but the mutation spectrum is very wide with hundreds of different mutations in both genes. Moreover, there have only been a few founder mutations reported in the Italian population, each restricted to a particular isolated geographical area (324). Even more interesting, Russia has an extremely high proportion (79%) of its breast/ovarian

cancer families attributed to *BRCA1* mutations (388) and has a low mutation spectrum, with the most common mutation (*BRCA1* c.5382insC) accounting for 94% of *BRCA1* and *BRCA2* mutations (389).

Since genetic isolates have been established in the Newfoundland population, and founder mutations have been determined to cause other monogenic disorders in Newfoundland (87, 91, 97, 109, 390), it was decided to take a targeted mutation screening approach after phase 1 and screen all newly recruited probands for the *BRCA1/2* Newfoundland mutations to determine if there were any founder mutations in the population (phase 2). Only one of the 57 new probands screened positive. *BRCA2* c.6714delACAA, was detected in a second family and an ancestral haplotype appeared to be shared by the probands. Therefore, it is likely a founder mutation. Unlike this Newfoundland cohort, some populations around the world, like Russia, have a high percentage of *BRCA1* and *BRCA2* involvement due to a limited number of mutations. Thus, targeted screening is very beneficial (389). The unsolved probands from phase 2 therefore need screening of the remaining unscreened exons of *BRCA1* and *BRCA2* in order to exclude these genes as pathogenic, and identify families for novel gene discovery. Despite the low detection rate of known Newfoundland *BRCA1/2* mutations in this second cohort, there is, however, no harm in first screening the exons with known mutations, followed by the screening of the remaining exons. This approach may solve only a small number of families but overall will save lab funds if there is a positive screen - the main purpose of targeted mutation screening.

Newfoundland's ancestry can be traced back to England and Ireland (86, 87), and 13 of the 15 Newfoundland *BRCA1* and *BRCA2* mutations that have been reported in BIC (391) all have a western European ancestry (eight specifically mention Britain or Ireland). Of the two mutations not reported in BIC (391), *BRCA2* c.697delAA has been reported in the Dutch population (392), and c.4642delAAGA, to our knowledge, has not been reported before. This mutation may have occurred more recently, perhaps since the family migrated to Newfoundland. Interesting, the *BRCA1* mutation, c.4446C>T, is a well-known founder mutation of the French Canadians (71, 332). This Newfoundland family is actually a small subset of a larger family with mainland Canadian ancestry therefore this mutation may be the same founder mutation as observed in French Canadians. Haplotyping would confirm this hypothesis.

In addition to the 16 *BRCA1* and *BRCA2* solved families, three other families have screened positive for the *CHK2* c.1100delC mutation. This allele is estimated to increase breast cancer risk by two-fold and account for 1% of all breast cancers (357). It also accounts for approximately 2% of the 153 probands in this study with a family history of breast/ovarian cancer. Despite that *CHK2* c.1100delC mutation carriers generally have an early age of breast cancer onset (357), the three probands that screened positive for the *CHK2* variant in this study were all diagnosed at the age of 52 years. All three probands also had first degree relatives with breast cancer, all of which were diagnosed over 50 years of age. Segregation of the *CHK2* c.1100delC mutation within these three families

was not performed because the DNA of additional family members was not available for screening.

Of the 96 families that had full gene screening, 84.4% were *BRCA1* and *BRCA2* negative (mystery families). However, point mutations in these families may have gone undetected since the sensitivity of the detection methods (PTT and SSCP) was not 100%. If DNA from multiple affected family members is available, linkage exclusion can be performed to rule out *BRCA1* and *BRCA2* involvement. Otherwise, genomic rearrangements have to be considered, as well as, deleterious variants in promoter regions, UTRs and intronic DNA. Many of these families have distinct characteristics associated with an increased likelihood of being a *BRCA1* or *BRCA2* family; such as, an early age of onset, bilateral breast cancer, multiple cases of breast cancer, one or more family member with two primary cancers, etc. (336, 393-395). It is likely that the breast cancer in some of these families is caused by a *BRCA1* or *BRCA2* variant that has not yet been detected for the various reasons listed above. However, the question also arises, could low-moderate penetrant genes or novel high-penetrant genes (*BRCA3*) be causing breast cancer in these families?

Previous attempts to identify new highly-penetrant hereditary breast cancer genes have been relatively unsuccessful; several genome-wide linkage scans were performed on multiple non-*BRCA1/2* breast cancer families to no avail (376-379). More recently, identifying common low-penetrant alleles through genome-wide association studies have

become more of a focus (296-303). The failed attempts of the linkage studies may be explained by genetic heterogeneity and it has been suggested to use families from a homogeneous population to reduce genetic variation and increase linkage power (377). However, the successful introduction of NGS technologies and its ability to use a small number of samples per disease family to identify rare variants that are shared amongst affected individuals (277) appears to be the next most reliable approach. Affected individuals in breast cancer families will share many sequencing variants, depending on how many individuals are sequenced. Exome sequencing many breast cancer families, and comparing the independently generated lists of variants that are shared by all affected individuals within each family may identify new breast cancer genes. This approach would even be more successful in isolated populations where breast cancer families from the same genetic isolates have a higher probability of sharing the same ancestral disease variant. This makes Newfoundland a good study population. Many mystery families have been identified in this chapter that can be exome sequenced. Extended family members of the mystery families are being recruited and the founders of each mystery families are being traced back to their original communities on the island. Mystery families that originate from the same fishing communities may represent clusters of related families that potentially share the same disease variants that can be identified using NGS.

**Table 4.1: Proband categorization into study phase, high- or moderate-risk family, and age of onset.**

| Screening Approach | Probands Grouped by age | High-risk families (4 or more cases) | Moderate-risk families (2 to 3 cases) | Total |
|---|---|---|---|---|
| Phase 1 (Conventional full gene screening) | 20-29 | 1 | 4 | 5 |
| | 30-39 | 9 | 7 | 16 |
| | 40-49 | 20 | 21 | 41 |
| | 50-59 | 10 | 15 | 25 |
| | 60+ | 5 | 4 | 9 |
| | **All ages** | **45** | **51** | **96** |
| Phase 2 (Targeted mutation screening) | 20-29 | 0 | 2 | 2 |
| | 30-39 | 2 | 8 | 10 |
| | 40-49 | 8 | 7 | 15 |
| | 50-59 | 9 | 6 | 15 |
| | 60+ | 11 | 4 | 15 |
| | **All ages** | **30** | **27** | **57** |
| **Total** | **All ages** | **75** | **78** | **153** |

**Table 4.2: *CHK2* c.1100delC mutation positive probands and family characteristics.**

| | | | | Proband | | | | Cancer cases in family | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene | Exon | Mutation | Putative Effect | Sex | Age of Onset | Site of Primary | Other Primary | Br <50 | Br 50+ | M Br | Ov-FT | Total | Other Cancers |
| CHK2 | 10 | c.1100delC | Stop | F | 52 | Breast | Multiple Myeloma | 0 | 2 | 0 | 0 | 2 | Sto, Tes, Thr |
| | 10 | c.1100delC | Stop | F | 52 | Breast | None | 0 | 3 | 0 | 0 | 3 | Leu, Sk, Sto |
| | 10 | c.1100delC | Stop | F | 52 | Breast | Ovanan | 0 | 2 | 0 | 3 | 5 | Cx |

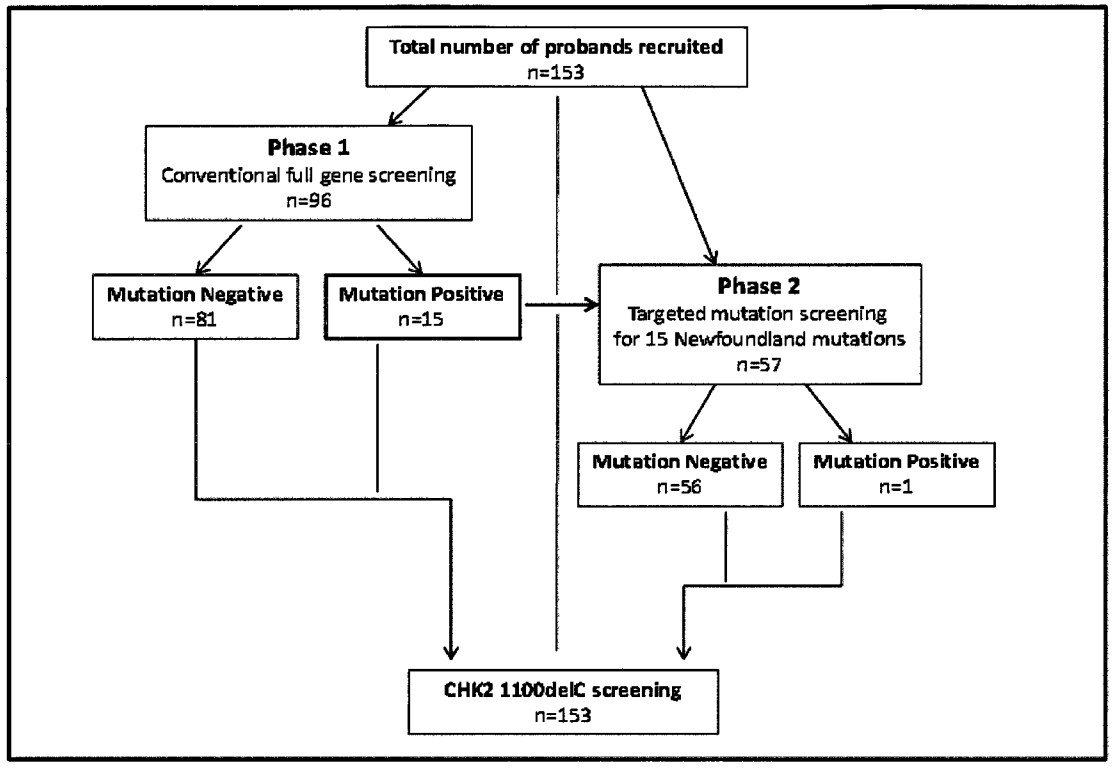



**Figure 4.1: Breast cancer study design and workflow.**

A

| Markers | Build 36.1 (Mb) | Shared Haplotype | | | |
|---|---|---|---|---|---|
| | | Family 199 | | Family 1191 | |
| D13S1242 | 28.04 | 1 | 3 | 1 | 3 |
| D13S1299 | 29.49 | 2 | 5 | 2 | 5 |
| D13S1287 | 30.61 | 4 | 4 | 6 | 4 |
| D13S260 | 31.33 | 5 | 3 | 3 | 3 |
| BRCA2 | 31.79 - 31.87 | wt | 6714del4 | wt | 6714del4 |
| D13S171 | 32.15 | 1 | 2 | 5 | 2 |
| D13S220 | 34.07 | 7 | 5 | 6 | 5 |

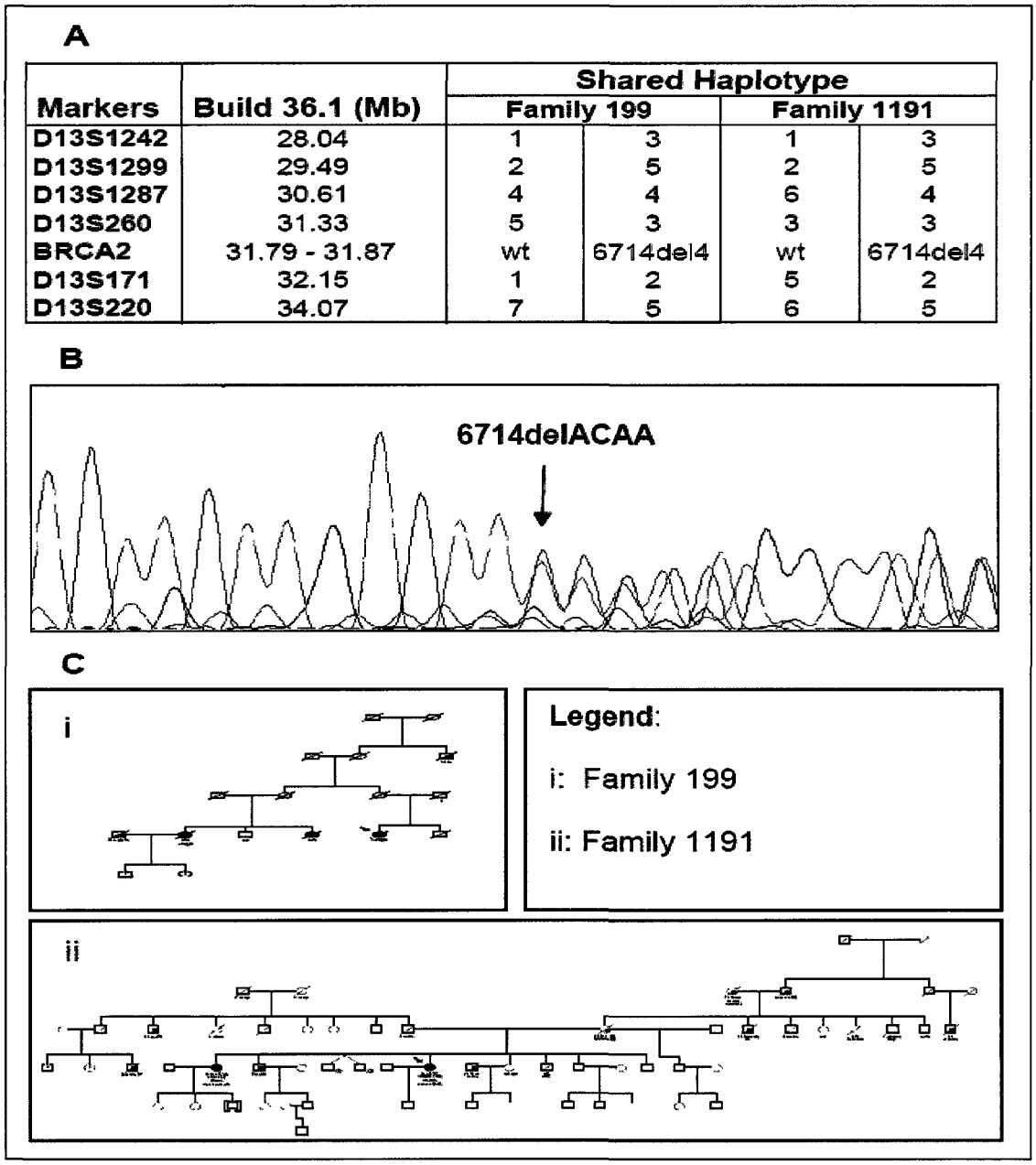



**Figure 4.2: The *BRCA2* 6714delACAA founder mutation.** (A) The shared haplotype on chromosome 13 (yellow) between the probands of families 199 and 1191 who have the *BRCA2* c.6714delACAA mutation. (B) A direct sequencing trace of the *BRCA2* mutation c.6714delACAA. (C) Pedigrees of the two families that share the ancestral mutation.

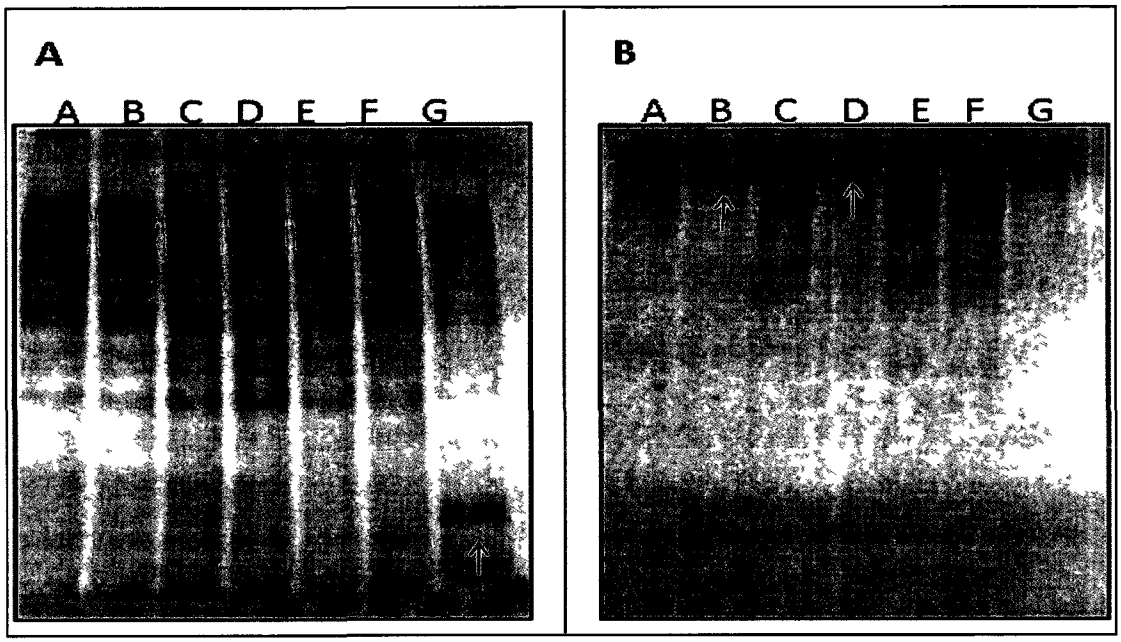



**Figure 4.3: The Protein Truncation Test.** (A) *BRCA1* exon 11, base pairs 1921-3383. Lanes A-F are wild-type (wt) with a normal peptide pattern. Lane G is a positive control for mutation c.2190delA. (B) *BRCA2* exon 11, base pairs 3625-7070. Lanes A, C, E, F and G are wildtype (wt) with a normal peptide pattern. Lane B is a positive control for mutation c.6293 C>G and Lane D is a positive control for mutation c.6714delACAA. Truncated protein is labeled with a black arrow.

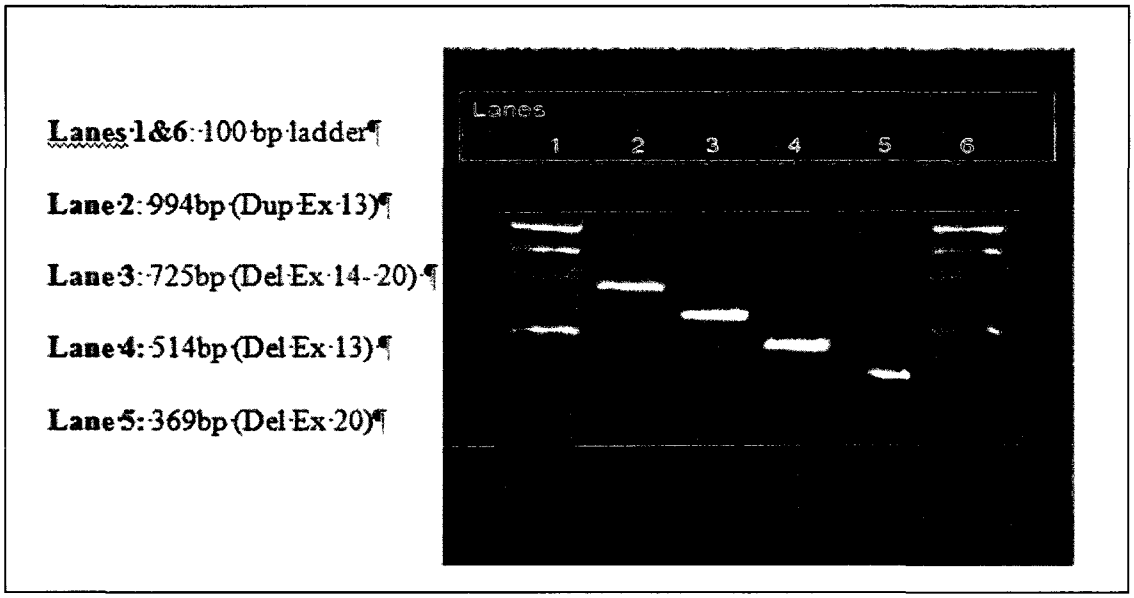
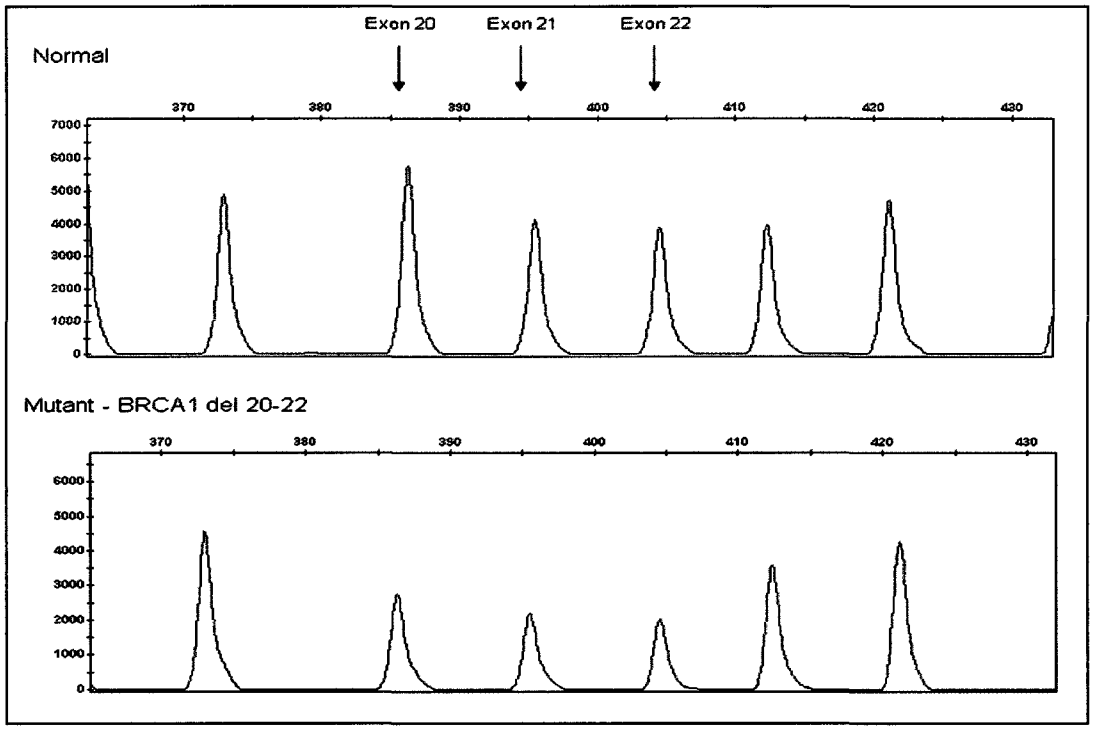



**Figure 4.4: Positive controls for four of the *BRCA1* genomic rearrangements screened by PCR and gel electrophoresis.**

A




B

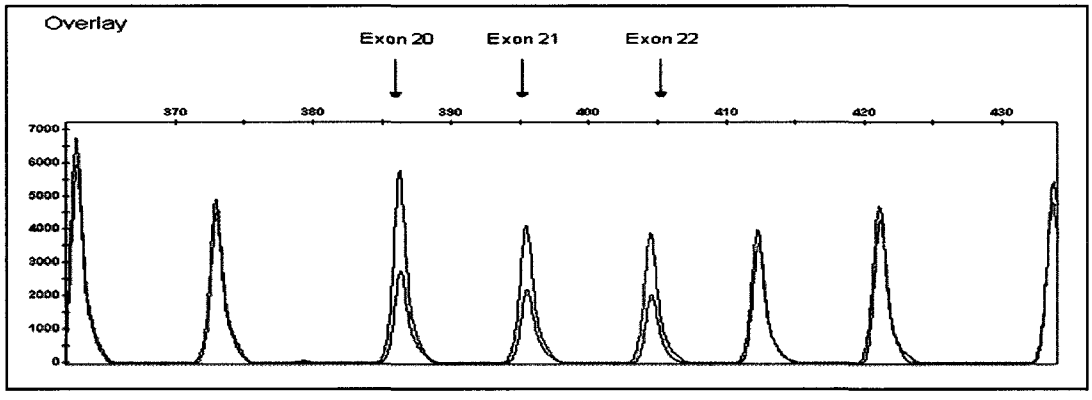



**Figure 4.5: MLPA detection of the *BRCA1* deletion of exon 20-22.** (A) Two electropherograms of a wt and a mutant sample. Notice the peak heights relative to each other. (B) Overlay of the wt and mutant sample. Notice the peak heights of the mutant sample is half that of the wt where there are copy number variants. The surrounding peak heights, where there is no variation, are the same.

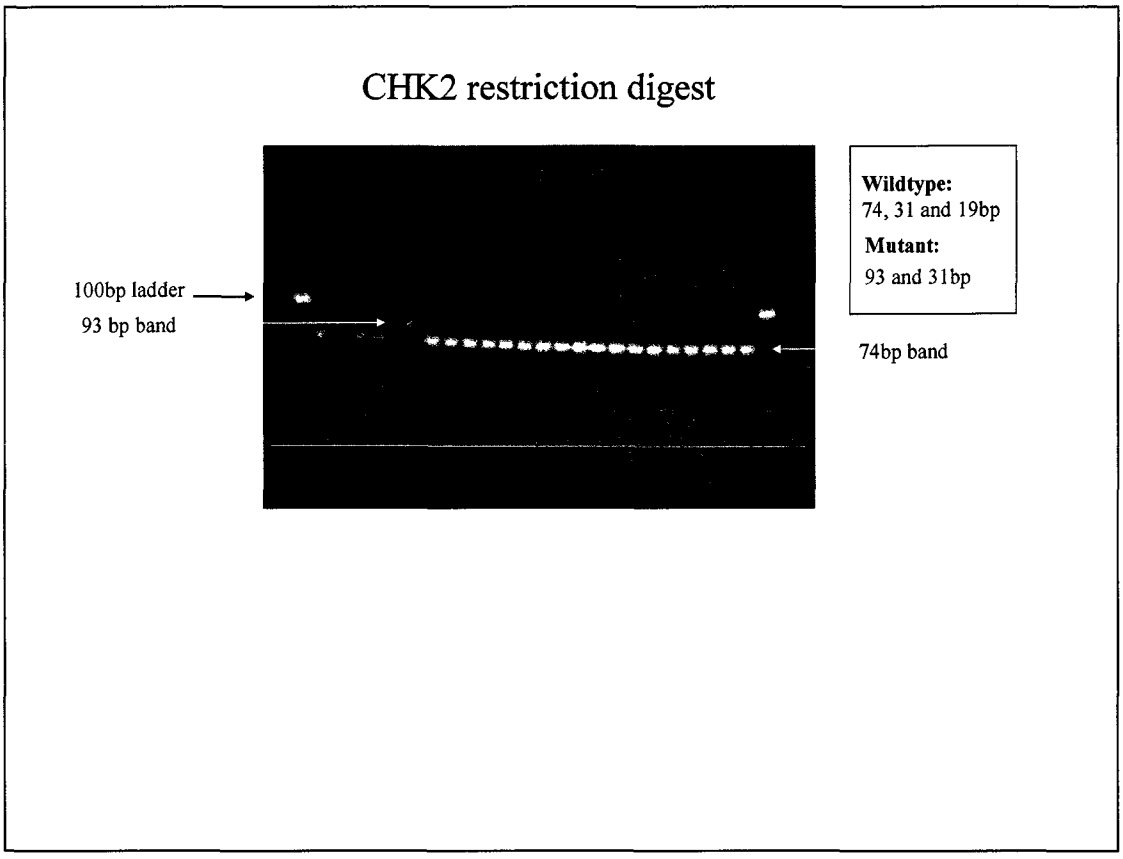



**Figure 4.6: Restriction enzyme test for *CHK2* c.1100delC.** A 2% agarose gel was run with a100 bp ladder in first and last lanes (100 bp band only one visible). The positive control for *CHK2* c.1100delC is in lane 7 (white arrow pointing to 93 bp band), the rest are negative for the mutation (white arrow pointing to 74 bp band).

# Chapter 5: General discussion

Gene discovery efforts have been very successful in isolated populations (71), and in this regard, the Newfoundland population has proven to be a gold mine for gene discovery. In the mid 1700s approximately 20,000 permanent immigrants, whose livelihood was the fishing industry, settled along the coastline of Newfoundland establishing several out-port fishing communities. The settlers mainly included English Protestants from south-west England and Irish Catholics from south-east Ireland (86). These out-port communities led to the development of several Newfoundland genetic isolates because of geographical distance between the communities and religious segregation (87). The Newfoundland population grew steadily through natural expansion up to 200,000 by the late 1800s, and today there are approximately 500,000 residents (98% of whom have English or Irish descent, and 60% of whom live in communities of less than 2500 residents) (87-89). Thus, it is not surprising that the founder effect has been identified in a number of studies on hereditary diseases in Newfoundland (90-97), and several of these studies have lead to past gene discovery successes (15, 98, 101, 102, 109, 111, 396). This thesis is another example of the opportunities, breakthroughs and challenges of gene discovery efforts in this population particularly regarding autosomal dominant and X-linked disorders.

The most significant achievement of this work was the discovery of *TMEM43* as the causal gene for autosomal dominant *ARVD5*. This success-story was made possible through the hard work and dedication of a team of individuals including clinicians, genetic counselors and researchers. Strictly using the International Task Force criteria

(126) to diagnose ARVC in this study was difficult because the historic nature of the Newfoundland ARVC pedigrees (clinical details were not available for many individuals in earlier generations and SCD was the primary disease feature), and presently there is a lack of availability of clinical tertiary care centers in the areas of the province where most affected individuals live. Therefore a subset of disease features was used to define affection status, which overcame diagnostic difficulties that may have hindered previous gene discovery efforts. Crucial for the gene discovery was the recruitment of the additional Newfoundland ARVC families that shared the same ancestral haplotype. Identifying key recombinations in two of these families reduced the critical region to a reasonable size for positional cloning and ultimately led to the identification of *TMEM43* as the causative gene. Furthermore, with the overall goal in gene discovery being the improvement of disease management, it is an honour to report that this gene discovery has enabled a clinical diagnostic test to be designed and offered to at-risk family members. Due to the serious repercussions of ARVC, this gene discovery will aid in early diagnosis and the implementation of the proper preventative measures that will save lives, including the use of implantable cardioverter-defibrillators. As such, there are also potential impacts on public health policy considering that the *TMEM43* founder mutation raises issues regarding the best way to provide genetic health care (mutation testing, genetic counseling, and follow-up specialist interventions).

The X-linked deafness family studied in Chapter 3 represented another large and extended Newfoundland disease pedigree that appeared to have the potential to localize a

disease locus. However, similar to the large ARVC pedigrees, due to the historic nature of this pedigree the proper medical records were not always available to diagnose the disease, especially in the earlier generations. Medical records were only available for 15 of the 50 reported affected individuals and therefore, disease status was mainly determined by a clinical history as recollected by relatives. Large pedigree size and the large number of apparently affected individuals can theoretically facilitate the determination of the mode of deafness inheritance, but not having the proper medical records to confirm disease status in all individuals can be misleading. Taking all modes of Mendelian inheritance into consideration, based on the family history and medical records, the most likely mode of inheritance of hearing loss in Family 2024 was X-linked due to the variable expression and less severe phenotype in females, and no apparent male-male transmission. Newfoundland represents a population of out-migration which makes it difficult to collect DNA from all desired individuals. This affected the recruitment of members of Family 2024, thus DNA was collected from only seven affected individuals. Haplotypes spanning the entire X chromosome revealed only one region that was shared by all affected individuals that participated in the genetics study - a 13.3 Mb region on Xp. Of interest was a 0.96 Mb region of homozygosity in a woman with two related, affected parents. That region of homozygosity could have potentially reduced the disease region to the *DMD* locus; a higher participation rate could have confirmed this. Overall, no pathogenic sequencing variant was detected after sequencing positional candidates in the 13.3 Mb region, unlike the detection of *TMEM43* c.1073C>T (p.S358L) in our ARVC cohort. However, in this case only the coding regions of cochlea

expressed genes were sequenced, whereas in the ARVC study all coding and non-coding exons of all positional candidates were screened. Thus, the variant could be in a gene not yet screened, in a non-coding region of a gene that was screened, within an unidentified isoform or gene within the critical region, or it may even be a copy number variant. NGS will probably be the next best approach to solve this family. A target-capture of the 13.3 Mb disease region can be performed and the region sequenced, which will identify both coding and non-coding variants, and potentially even rearrangements. Or a cheaper option would be to perform an exome capture and sequence basically all coding exons in the genome. Using this approach one would hope that the variant is coding (or in an intronic region that is close to an exon) but it would be beneficial if the mode of inheritance is incorrect because coding regions of autosomal genes will be sequenced as well.

Lastly, the Newfoundland population provides an opportunity to identify a novel breast cancer gene(s). Conventional, full-gene screening in 96 Newfoundland breast cancer probands with a family history of breast cancer identified 15 *BRCA1* and *BRCA2* truncation mutations, solving 15.6% of the cohort (15/96). Taking into consideration that an additional two of the 96 probands later screened positive for *CHK2* c.1100delC and that some of the remaining unsolved "mystery" families (79/96) may in fact be *BRCA1* or *BRCA2* families with yet-unidentified mutations, there remains an opportunity for novel gene discovery. Previous attempts to identify new highly penetrant hereditary breast cancer genes in non-*BRCA1/2* families have been relatively unsuccessful (376-379),

however NGS may ease the difficulty of linkage. Smaller families can be sequenced and comparing variant lists between families may identify a novel gene. Even more beneficial would be to study families from the same genetic isolate. A common founder mutation may be uncovered. The remaining Newfoundland mystery families provide an opportunity to find a highly sought after, novel breast cancer gene. Extended family members are being recruited from the most informative mystery families, and the founders of these families are being traced back to their original communities on the island. Mystery families that originate from the same fishing communities may represent clusters of related families that could be used to identify new genes through NGS. Nevertheless, this unsolved breast cancer cohort can also participate in association studies to identify low-moderate penetrance breast cancer alleles as well.

Overall, it is beneficial to study isolated populations like Newfoundland when making gene discovery efforts. It takes luck and hard work to find the ultimate, a disease gene. I was very fortunate to find one!

# Bibliography

1. Hodgkinson K, Pullman D. Duty to warn and genetic disease. Can J Cardiovasc Nurs 2010: 20: 12-15.

2. Hamilton R. Genetics: breast cancer as an exemplar. Nurs Clin North Am 2009: 44: 327-338.

3. Butcher SP. Target discovery and validation in the post-genomic era. Neurochem Res 2003: 28: 367-371.

4. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med 2009: 360: 1759-1768.

5. Kandpal R, Saviola B, Felton J. The era of 'omics unlimited. Biotechniques 2009: 46: 351-352, 354-355.

6. Grosskreutz J, Van Den Bosch L, Keller BU. Calcium dysregulation in amyotrophic lateral sclerosis. Cell Calcium 2010: 47: 165-174.

7. Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. Nature 2001: 409: 860-921.

8. Venter JC, Adams MD, Myers EW et al. The sequence of the human genome. Science 2001: 291: 1304-1351.

9. International-Human-Genome-Sequencing-Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004: 431: 931-945.

10. Cooper DN, Chen JM, Ball EV et al. Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. Hum Mutat 2010: 31: 631-655.

11. Gerstein MB, Bruce C, Rozowsky JS et al. What is a gene, post-ENCODE? History and updated definition. Genome Res 2007: 17: 669-681.

12. Yang MQ, Elnitski LL. Diversity of core promoter elements comprising human bidirectional promoters. BMC Genomics 2008: 9 Suppl 2: S3.

13. Vuoristo JT, Berrettini WH, Ala-Kokko L. C18orf2, a novel, highly conserved intronless gene within intron 5 of the GNAL gene on chromosome 18p11. Cytogenet Cell Genet 2001: 93: 19-22.

14. Austin RC, Howard PL, D'Souza VN et al. Cloning and characterization of alternatively spliced isoforms of Dp71. Hum Mol Genet 1995: 4: 1475-1483.

15. Shekarabi M, Girard N, Riviere JB et al. Mutations in the nervous system--specific HSN2 exon of WNK1 cause hereditary sensory neuropathy type II. J Clin Invest 2008: 118: 2496-2505.

16. Rajpar MH, Koch MJ, Davies RM et al. Mutation of the signal peptide region of the bicistronic gene DSPP affects translocation to the endoplasmic reticulum and results in defective dentine biomineralization. Hum Mol Genet 2002: 11: 2559-2565.

17. Horiuchi T, Aigaki T. Alternative trans-splicing: a novel mode of pre-mRNA processing. Biol Cell 2006: 98: 135-140.

18. Chen JM, Ferec C, Cooper DN. Revealing the human mutome. Clin Genet 2010.

19. International-Hapmap-Consortium. A haplotype map of the human genome. Nature 2005: 437: 1299-1320.

20. Wang WY, Barratt BJ, Clayton DG et al. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 2005: 6: 109-118.

21. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008: 118: 1590-1605.

22. Kraft P, Hunter DJ. Genetic risk prediction--are we there yet? N Engl J Med 2009: 360: 1701-1703.

23. Goldstein DB. Common genetic variation and human traits. N Engl J Med 2009: 360: 1696-1698.

24. Hirschhorn JN. Genomewide association studies--illuminating biologic pathways. N Engl J Med 2009: 360: 1699-1701.

25. Schork NJ, Murray SS, Frazer KA et al. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 2009: 19: 212-219.

26. Rudan I. New technologies provide insights into genetic basis of psychiatric disorders and explain their co-morbidity. Psychiatr Danub 2010: 22: 190-192.

27. OMIM. OMIM, Online Mendelian Inheritance in Man. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 2010.

28. Louis ED. Essential tremors: a family of neurodegenerative disorders? Arch Neurol 2009: 66: 1202-1208.

29. Deuschl G, Elble R. Essential tremor--neurodegenerative or nondegenerative disease towards a working definition of ET. Mov Disord 2009: 24: 2033-2041.

30. Musa FU, Ratajczak P, Sahu J et al. Ocular manifestations in oculodentodigital dysplasia resulting from a heterozygous missense mutation (L113P) in GJA1 (connexin 43). Eye 2009: 23: 549-555.

31. Fackenthal JD, Olopade OI. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. Nat Rev Cancer 2007: 7: 937-948.

32. Al-Mulla F, Bland JM, Serratt D et al. Age-dependent penetrance of different germline mutations in the BRCA1 gene. J Clin Pathol 2009: 62: 350-356.

33. Antoniou AC, Gayther SA, Stratton JF et al. Risk models for familial ovarian and breast cancer. Genet Epidemiol 2000: 18: 173-190.

34. Chen S, Iversen ES, Friebel T et al. Characterization of BRCA1 and BRCA2 mutations in a large United States sample. J Clin Oncol 2006: 24: 863-871.

35. Ford D, Easton DF, Bishop DT et al. Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Lancet 1994: 343: 692-695.

36. Risch HA, McLaughlin JR, Cole DE et al. Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer. Am J Hum Genet 2001: 68: 700-710.

37. Renwick A, Thompson D, Seal S et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet 2006: 38: 873-875.

38. Van Camp GaS, RJH. Hereditary Hearing Loss Homepage. 2009.

39. Harding AE. The clinical features and classification of the late onset autosomal dominant cerebellar ataxias. A study of 11 families, including descendants of the 'the Drew family of Walworth'. Brain 1982: 105: 1-28.

40. Cagnoli C, Mariotti C, Taroni F et al. SCA28, a novel form of autosomal dominant cerebellar ataxia on chromosome 18p11.22-q11.2. Brain 2006: 129: 235-242.

41. Orr HT, Chung MY, Banfi S et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 1993: 4: 221-226.

42. Schols L, Bauer P, Schmidt T et al. Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. Lancet Neurol 2004: 3: 291-304.

43. Swanson AG. Congenital insensitivity to pain with anhydrosis. A unique syndrome in two male siblings. Arch Neurol 1963: 8: 299-306.

44. Swanson AG, Buchan GC, Alvord EC, Jr. Anatomic Changes in Congenital Insensitivity to Pain. Absence of Small Primary Sensory Neurons in Ganglia, Roots, and Lissauer's Tract. Arch Neurol 1965: 12: 12-18.

45. Indo Y, Tsuruta M, Hayashida Y et al. Mutations in the TRKA/NGF receptor gene in patients with congenital insensitivity to pain with anhidrosis. Nat Genet 1996: 13: 485-488.

46. Bell PA, Chaturvedi S, Gelfand CA et al. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques 2002: Suppl: 70-72, 74, 76-77.

47. Hua J, Craig DW, Brun M et al. SNiPer-HD: improved genotype calling accuracy by an expectation-maximization algorithm for high-density SNP arrays. Bioinformatics 2007: 23: 57-63.

48. Ott J. Analysis of Human Genetic Linkage. Baltimore, Maryland: The John Hopkins University Press, 1991.

49. Wild EJ, Mudanohwo EE, Sweeney MG et al. Huntington's disease phenocopies are clinically and genetically heterogeneous. Mov Disord 2008: 23: 716-720.

50. Kruglyak L, Daly MJ, Reeve-Daly MP et al. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996: 58: 1347-1363.

51. Kent WJ, Sugnet CW, Furey TS et al. The human genome browser at UCSC. Genome Res 2002: 12: 996-1006.

52. Karolchik D, Baertsch R, Diekhans M et al. The UCSC Genome Browser Database. Nucleic Acids Res 2003: 31: 51-54.

53. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977: 74: 5463-5467.

54. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. 1977. Biotechnology 1992: 24: 104-108.

55. Di Bella D, Lazzaro F, Brusco A et al. Mutations in the mitochondrial protease gene AFG3L2 cause dominant hereditary ataxia SCA28. Nat Genet 2010: 42: 313-321.

56. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010: 11: 31-46.

57. Stankiewicz P, Lupski JR. The genomic basis of disease, mechanisms and assays for genomic disorders. Genome Dyn 2006: 1: 1-16.

58. MRC-Holland. MLPA- Homepage. MRC-Holland, 2005.

59. Su LK, Steinbach G, Sawyer JC et al. Genomic rearrangements of the APC tumor-suppressor gene in familial adenomatous polyposis. Hum Genet 2000: 106: 101-107.

60. Spaepen M, Vankeirsbilck B, Van Opstal S et al. Germline mutations of the hMLH1 and hMSH2 mismatch repair genes in Belgian hereditary nonpolyposis colon cancer (HNPCC) patients. Fam Cancer 2006: 5: 179-189.

61. Gutierrez-Enriquez S, de la Hoya M, Martinez-Bouzas C et al. Screening for large rearrangements of the BRCA2 gene in Spanish families with breast/ovarian cancer. Breast Cancer Res Treat 2007: 103: 103-107.

62. Vasickova P, Machackova E, Lukesova M et al. High occurrence of BRCA1 intragenic rearrangements in hereditary breast and ovarian cancer syndrome in the Czech Republic. BMC Med Genet 2007: 8: 32.

63. Sluiter MD, van Rensburg EJ. Large genomic rearrangements of the BRCA1 and BRCA2 genes: review of the literature and report of a novel BRCA1 mutation. Breast Cancer Res Treat 2010.

64. Mazoyer S. Genomic rearrangements in the BRCA1 and BRCA2 genes. Hum Mutat 2005: 25: 415-422.

65. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 2005: 76: 8-32.

66. Petersen MB, Wang Q, Willems PJ. Sex-linked deafness. Clin Genet 2008: 73: 14-23.

67. De Lellis L, Curia MC, Veschi S et al. Methods for routine diagnosis of genomic rearrangements: multiplex PCR-based methods and future perspectives. Expert Rev Mol Diagn 2008: 8: 41-52.

68. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003: 31: 3812-3814.

69. Mi H, Lazareva-Ulitsky B, Loo R et al. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res 2005: 33: D284-288.

70. Larkin MA, Blackshields G, Brown NP et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007: 23: 2947-2948.

71. Neuhausen SL. Founder populations and their uses for breast cancer genetics. Breast Cancer Res 2000: 2: 77-81.

72. Heyer E, Tremblay M. Variability of the genetic contribution of Quebec population founders associated to some deleterious genes. Am J Hum Genet 1995: 56: 970-978.

73. Laberge AM, Michaud J, Richter A et al. Population history and its impact on medical genetics in Quebec. Clin Genet 2005: 68: 287-301.

74. De Braekeleer M, Dao TN. Hereditary disorders in the French Canadian population of Quebec. I. In search of founders. Hum Biol 1994: 66: 205-223.

75. De Braekeleer M, Giasson F, Mathieu J et al. Genetic epidemiology of autosomal recessive spastic ataxia of Charlevoix-Saguenay in northeastern Quebec. Genet Epidemiol 1993: 10: 17-25.

76. Engert JC, Berube P, Mercier J et al. ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF. Nat Genet 2000: 24: 120-125.

77. Richter A, Rioux JD, Bouchard JP et al. Location score and haplotype analyses of the locus for autosomal recessive spastic ataxia of Charlevoix-Saguenay, in chromosome region 13q11. Am J Hum Genet 1999: 64: 768-775.

78. El Euch-Fayache G, Lalani I, Amouri R et al. Phenotypic features and genetic findings in sacsin-related autosomal recessive ataxia in Tunisia. Arch Neurol 2003: 60: 982-988.

79. Criscuolo C, Sacca F, De Michele G et al. Novel mutation of SACS gene in a Spanish family with autosomal recessive spastic ataxia. Mov Disord 2005: 20: 1358-1361.

80. Criscuolo C, Banfi S, Orio M et al. A novel mutation in SACS gene in a family from southern Italy. Neurology 2004: 62: 100-102.

81. Grieco GS, Malandrini A, Comanducci G et al. Novel SACS mutations in autosomal recessive spastic ataxia of Charlevoix-Saguenay type. Neurology 2004: 62: 103-106.

82. Hara K, Onodera O, Endo M et al. Sacsin-related autosomal recessive ataxia without prominent retinal myelinated fibers in Japan. Mov Disord 2005: 20: 380-382.

83. Ogawa T, Takiyama Y, Sakoe K et al. Identification of a SACS gene missense mutation in ARSACS. Neurology 2004: 62: 107-109.

84. Richter AM, Ozgul RK, Poisson VC et al. Private SACS mutations in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) families from Turkey. Neurogenetics 2004: 5: 165-170.

85. Vermeer S, Meijer RP, Pijl BJ et al. ARSACS in the Dutch population: a frequent cause of early-onset cerebellar ataxia. Neurogenetics 2008: 9: 207-214.

86. Mannion J, ed. The peopling of Newfoundland: Essays in Historical Geography. St. John's: Memorial University, 1977.

87. Rahman P, Jones A, Curtis J et al. The Newfoundland population: a unique resource for genetic investigation of complex diseases. Human Molecular Genetics 2003: 12: R167-172.

88. Bear J, Nemec T, Kennedy J. Inbreeding in Outport Newfoundland. American Journal of Medical Genetics 1988: 29: 649-660.

89. Bear J, Nemec T, Kennedy J et al. Persistent genetic isolation in outport Newfoundland. American Journal of Medical Genetics 1987: 27: 807-830.

90. Olufemi SE, Green JS, Manickam P et al. Common ancestral mutation in the MEN1 gene is likely responsible for the prolactinoma variant of MEN1 (MEN1Burin) in four kindreds from Newfoundland. Hum Mutat 1998: 11: 264-269.

91. Spirio L, Green J, Robertson J et al. The identical 5' splice-site acceptor mutation in five attenuated APC families from Newfoundland demonstrates a founder effect. Hum Genet 1999: 105: 388-398.

92. Kopciuk KA, Choi YH, Parkhomenko E et al. Penetrance of HNPCC-related cancers in a retrolective cohort of 12 large Newfoundland families carrying a MSH2 founder mutation: an evaluation using modified segregation models. Hered Cancer Clin Pract 2009: 7: 16.

93. Froggatt N, Green J, Brassett C et al. A common MSH2 mutation in English and North american HNPCC families: origin, phenotypic expression and sex specific differences in colorectal cancer. Journal of Medical Genetics 1999: 36: 97-102.

94. Grewal KK, Stefanelli MG, Meijer IA et al. A founder effect in three large Newfoundland families with a novel clinically variable spastic ataxia and supranuclear gaze palsy. Am J Med Genet A 2004: 131: 249-254.

95. Meijer IA, Hand CK, Grewal KK et al. A locus for autosomal dominant hereditary spastic ataxia, SAX1, maps to chromosome 12p13. Am J Hum Genet 2002: 70: 763-769.

96. Kaurah P, MacMillan A, Boyd N et al. Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. JAMA 2007: 297: 2360-2372.

97. Xie YG, Zheng H, Leggo J et al. A founder factor VIII mutation, valine 2016 to alanine, in a population with an extraordinarily high prevalence of mild hemophilia A. Thromb Haemost 2002: 87: 178-179.

98. Young T-L, Ives E, Lynch E et al. Non-syndromic progressive hearing loss DFNA38 is caused by heterozygous missense mutation in the Wolfram syndrome gene WFS1. Hum Mol Genet 2001: 10: 2509-2514.

99. Peltomaki P, Aaltonen LA, Sistonen P et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science 1993: 260: 810-812.

100. Leach FS, Nicolaides NC, Papadopoulos N et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 1993: 75: 1215-1225.

101. Lafreniere R, MacDonald M, Dube M et al. Identification of a novel gene (HSN2) causing hereditary sensory and autonomic neuropathy type II through the Study of Canadian Genetic Isolates. Am J Hum Genet 2002 Dec; 2004: 74: 1064-1073.

102. Aksentijevich I, Masters SL, Ferguson PJ et al. An autoinflammatory disease with deficiency of the interleukin-1-receptor antagonist. N Engl J Med 2009: 360: 2426-2437.

103. Leppert M, Baird L, Anderson KL et al. Bardet-Biedl syndrome is linked to DNA markers on chromosome 11q and is genetically heterogeneous. Nat Genet 1994: 7: 108-112.

104. Young TL, Woods MO, Parfrey PS et al. A founder effect in the newfoundland population reduces the Bardet-Biedl syndrome I (BBS1) interval to 1 cM. Am J Hum Genet 1999: 65: 1680-1687.

105. Mykytyn K, Nishimura DY, Searby CC et al. Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. Nat Genet 2002: 31: 435-438.

106. Young TL, Woods MO, Parfrey PS et al. Canadian Bardet-Biedl syndrome family reduces the critical region of BBS3 (3p) and presents with a variable phenotype. Am J Med Genet 1998: 78: 461-467.

107. Sheffield VC, Carmi R, Kwitek-Black A et al. Identification of a Bardet-Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. Hum Mol Genet 1994: 3: 1331-1335.

108. Fan Y, Esmail MA, Ansley SJ et al. Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet-Biedl syndrome. Nat Genet 2004: 36: 989-993.

109. Young T, Penney L, Woods M et al. A fifth locus for Bardet Biedl syndrome maps to chromosome 2q31. American Journal of Human Genetics 1999: 64: 900-904.

110. Li JB, Gerdes JM, Haycraft CJ et al. Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell 2004: 117: 541-552.

111. Katsanis N, Beales PL, Woods MO et al. Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. Nat Genet 2000: 26: 67-70.

112. Richard P, Fressart V, Charron P et al. [Genetics of inherited cardiomyopathies.]. Pathol Biol (Paris) 2009.

113. Marcus F, Fontaine G, Guiraudon G. Right ventricular dysplasia: a report of 24 adult cases. Circulation 1982: 65: 384-398.

114. Fontaine G, Guiraudon G, Frank R et al. Stimulation studies and epicardial mapping in VT: Study of mechanisms and selection for surgery. In: Kulbertus H, ed. Reentrant Arrhythmias. Lancaster PA: MTP publishers, 1977: 334-350.

115. Dala Volta S, Fameli O, Maschio G. [The clinical and hemodynamic syndrome of auricularisation of the right ventricle (apropos of 4 personal cases)]. Arch Mal Coeur Vaiss 1965: 58: 1129-1143.

116. Green J, Korovetz M, Shanklin D et al. Sudden Unexpected Death in Three Generations. Arch Intern Med 1969: 124: 359-363.

117. Frank R, Fontaine G, Vedel J. Electrocardiologie de quatre cas de dysplasie ventriculaire droite arythmogene. Arch Mal Coeur Vaiss 1978: 71: 963-972.

118. Fontaine G, Guiraudon G, Frank R et al. [Arrhythmogenic right ventricular dysplasia and Uhl's disease]. Arch Mal Coeur Vaiss 1982: 75: 361-371.

119. WHO/ISFC. Report of the 1995 World Health Organisation/International Society and Federation of Cardiology task force on the definition and classification of cardiomyopathies. Circulation 1996: 93: 841- 843.

120. Corrado D, Fontaine G, Marcus F et al. Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy: Need for an International Registry [Miscellaneous Article]. Circulation 2000: 101: e101-e106.

121. Dokuparti MV, Pamuru PR, Thakkar B et al. Etiopathogenesis of arrhythmogenic right ventricular cardiomyopathy. J Hum Genet 2005: 50: 375-381.

122. Nava A, Scognamiglio R, Thiene G et al. A polymorphic form of familial arrhythmogenic right ventricular dysplasia. American Journal of Cardiology 1987: 59: 1405-1409.

123. Thiene G, Nava A, Corrado D et al. Right ventricular cardiomyopathy and sudden cardiac death in young people. New England Journal of Medicine 1988: 318: 129-133.

124. Shoji T, Kaneko M, Onodera K et al. Arrhythmogenic right ventricular dysplasia with massive involvement of the left ventricle. Canadian Journal of Cardiology 1991: 7: 303-307.

125. Matsuo S, Sato Y, Nakae I et al. Left ventricular involvement in arrhythmogenic right ventricular cardiomyopathy demonstrated by multidetector-row computed tomography. International Journal of Cardiology 2007: 115: e129-e131.

126. McKenna WJ, Thiene G, Nava A et al. Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial disease of the European Society of Cardiology and the Scientific Council on Cardiomyopathies of the International Society and Federation of Cardiology. British Heart Journal 1994: 71: 215-218.

127. Hamid MS, Norman M, Quraishi A et al. Prospective Evaluation of Relatives for Familal Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia Reveals a Need to Broaden Diagnostic Criteria. Journal of the American College of Cardiology 2002: 40: 1445-1450.

128. Cho Y, Park T, Yang DH et al. Arrhythmogenic right ventricular cardiomyopathy and sudden cardiac death in young Koreans. Circ-J 2003: 67: 925-928.

129. Fung W, Sanderson J. Clinical Profile of ARVC in Chinese patients. Int J Cardiol 2001: 81: 9-18.

130. Perzanowski C, Crespo G, Yazdanfar S. Images in Cardiology: familial ventricular tachycardia with mild ventricular dysfunction: a 15 year follow-up of two African American brothers with arrhythmogenic right ventricular cardiomyopathy. Heart 2000: 84: 658.

131. Kaartinen M, Helio T, Lehtonen A et al. Characterization of familial and sporadic arrhythmogenic right ventricular cardiomyopathy in Finland. Ann Med ; 2007: 39: 312-318.

132. Severini GM, Krajinovic M, Pinamonti B et al. A new locus for arrhythmogenic right ventricular dysplasia on the long arm of chromosome 14. Genomics 1996: 31: 193-200.

133. Rampazzo A, Nava A, Danieli GA et al. The gene for arrhythmogenic right ventricular cardiomyopathy maps to chromosome 14q23-q24. Human Molecular Genetics 1994: 3: 959-962.

134. Rampazzo A, Nava A, Erne P et al. A new locus for arrhythmogenic right ventricular cardiomyopathy (ARVD2) maps to chromosome 1q42-q43. Human Molecular Genetics 1995: 4: 2151-2154.

135. Rampazzo A, Nava A, Miorin M et al. ARVD4, a new locus for arrhythmogenic right ventricular cardiomyopathy maps to chromosome 2 long arm. Genomics 1997: 45: 259-263.

136. Shen W, Edwards W, Hammill S et al. Sudden unexpected non-traumatic death in 54 young adults: a 30 year population study. American Journal of Cardiology 1995: 76: 148-152.

137. Tabib A, Loire R, Mirast A et al. Unsuspected cardiac lesions associated with sudden unexpected perioperative death. European Journal of Anaesthesiology 2000: 17: 230-235.

138. McKoy G, Protonotarios N, Crosby A et al. Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). Lancet 2000: 355: 2119-2124.

139. Protonotarios N, Tsatsopoulou A, Patsourakos P et al. Cardiac abnormalities in familial palmoplantar keratosis. Br Heart J 1986: 56: 321-326.

140. Coonar AS, Protonotarios N, Tsatsopoulou A et al. Gene for arrhythmogenic right ventricular cardiomyopathy with diffuse nonepidermolytic palmoplantar keratoderma and woolly hair (Naxos disease) maps to 17q21. Circulation 1998: 97: 2049-2058.

141. Rampazzo A, Beffagna G, Nava A et al. Arrhythmogenic right ventricular cardiomyopathy type 1 (ARVD1): confirmation of locus assignment and mutation screening of four candidate genes. European Journal of Human Genetics 2003: 11: 69-76.

142. Beffagna G, Occhi G, Nava A et al. Regulatory mutations in transforming growth factor-[beta]3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. Cardiovascular Research 2005: 65: 366-373.

143. Tiso N, Stephan DA, Nava A et al. Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). Human Molecular Genetics 2001: 10: 189-194.

144. Bauce B, Nava A, Rampazzo A et al. Familial effort polymorphic ventricular arrhythmias in arrhythmogenic right ventricular cardiomyopathy map to chromosome 1q42-43. The American Journal of Cardiology 2000: 85: 573-579.

145. Kirsch L, Weinstock D, Magid M et al. Treatment of presumed arrhythmogenic right ventricular dysplasia in an adolescent. Chest 1993: 104: 298-300.

146. Ahmad F, Li D, Karibe A et al. Localisation of a gene responsible for arrhythmogenic right ventricular dysplasia to chromosome 3p23. Circulation 1998: 98: 2791-2795.

147. Marshall W, Furey M, Larsen B et al. Right ventricular cardiomyopathy and sudden death in young people. New England Journal of Medicine 1988: 319: 174-175.

148. Li D, Ahmad F, Gardner M et al. The locus of a novel gene responsible for arrhythmogenic right ventricular dysplasia characterised by early onset and high penetrance maps to chromosome 10p12-p14. American Journal of Human Genetics 2000: 66: 148-156.

149. Matolweni LO, Bardien S, Rebello G et al. Arrhythmogenic right ventricular cardiomyopathy type 6 (ARVC6): support for the locus assignment, narrowing of the critical region and mutation screening of three candidate genes. BMC Med Genet 2006: 7: 29.

150. Melberg A, A O, Blomstrom-Lundqvist C et al. Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy linked to chromosome 10q. Annals of Neurology 1999: 46: 684-692.

151. Kuhl A, Melberg A, Meinl E et al. Myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy 7: corroboration and narrowing of the critical region on 10q22.3. Eur J Hum Genet 2008: 16: 367-373.

152. Selcen D, Engel A. Mutations in ZASP define a novel form of muscular dystrophy in humans. Ann Neurol 2005: 57: 269-276.

153. Vatta M, Mohapatra B, Jimenez S et al. Mutations in Cypher/ZASP in patients with dilated cardiomyopathy and left ventricular non-compaction. J Am Coll Cardiol 2003: 42: 2014-2027.

154. Rampazzo A, Nava A, Malacrida S et al. Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet 2002: 71: 1200-1206.

155. Norgett EE, Hatsell SJ, Carvajal-Huerta L et al. Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma. Hum Mol Genet 2000: 9: 2761-2766.

156. Gerull B, Heuser A, Wichter T et al. Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. Nature Genetics 2004: 36:: 1162-1164.

157. Grossmann K, Grund C, Huelsken J et al. Requirement of plakophilin 2 for heart morphogenesis and cardiac junction formation. J Cell Biol 2004: 167: 149-160.

158. van Tintelen JP, Entius MM, Bhuiyan ZA et al. Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circulation 2006: 113: 1650-1658.

159. Hall C, Li S, Li H et al. Arrhythmogenic Right Ventricular Cardiomyopathy Plakophilin-2 Mutations Disrupt Desmosome Assembly and Stability. Cell Commun Adhes 2009: 1-13.

160. OMIM. OMIM, Online Mendelian Inheritance in Man. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 2009.

161. Pilichou K, Nava A, Basso C et al. Mutations in Desmoglein-2 Gene Are Associated With Arrhythmogenic Right Ventricular Cardiomyopathy 10.1161/CIRCULATIONAHA.105.583674. Circulation 2006: 113: 1171-1179.

162. Awad MM, Dalal D, Cho E et al. DSG2 Mutations Contribute to Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy. Am J Hum Genet 2006: 79: 136-142.

163. Syrris P, Ward D, Asimaki A et al. Desmoglein-2 mutations in arrhythmogenic right ventricular cardiomyopathy: a genotype-phenotype characterization of familial disease 10.1093/eurheartj/ehl380. Eur Heart J 2007: 28: 581-588.

164. Syrris P, Ward D, Evans A et al. Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. Am J Hum Genet 2006: 79: 978-984.

165. Heuser A, Plovie ER, Ellinor PT et al. Mutant Desmocollin-2 Causes Arrhythmogenic Right Ventricular Cardiomyopathy. Am J Hum Genet 2006: 79: 1081-1088.

166. Asimaki A, Syrris P, Wichter T et al. A novel dominant mutation in plakoglobin causes arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet 2007: 81: 964-973.

167. Delmar M, McKenna WJ. The Cardiac Desmosome and Arrhythmogenic Cardiomyopathies: From Gene to Disease. Circ Res 2010: 107: 700-714.

168. Sen-Chowdhry S, Syrris P, McKenna W. Genetics of right ventricular cardiomyopathy. J Cardiovasc Electrophysiol 2005: 16: 927-935.

169. Stokes DL. Desmosomes from a structural perspective. Curr Opin Cell Biol 2007: 19: 565-571.

170. Gasteiger E, Gattiker A, Hoogland C et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 2003: 31: 3784-3788.

171. Garcia-Gras E, Lombardi R, Giocondo MJ et al. Suppression of canonical Wnt/beta-catenin signaling by nuclear plakoglobin recapitulates phenotype of arrhythmogenic right ventricular cardiomyopathy. J Clin Invest 2006: 116: 2012-2021.

172. Yang Z, Bowles NE, Scherer SE et al. Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circ Res 2006: 99: 646-655.

173. Syrris P, Ward D, Asimaki A et al. Clinical expression of plakophilin-2 mutations in familial arrhythmogenic right ventricular cardiomyopathy. Circulation 2006: 113: 356-364.

174. Hodgkinson K. The clinical and genetic epidemiology of ARVC in Newfoundland. St. John's: Memorial University, 2009.

175. French V. Fine-mapping and mutation identification for ARVC (Arrhythmogenic Right Ventricular Cardiomyopathy) at the ARVD5 locus in the Newfoundland population., Vol. M.Sc. Thesis. St. John's: Memorial University of Newfoundland, 2008.

176. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000: 132: 365-386.

177. Woods MO, Hyde A, Curtis F et al. High frequency of hereditary colorectal cancer in Newfoundland likely involves novel genes. Clinical Cancer Research 2005: 11: 6853-6861.

178. Crooks GE, Hon G, Chandonia JM et al. WebLogo: a sequence logo generator. Genome Res 2004: 14: 1188-1190.

179. Combet C, Blanchet C, Geourjon C et al. NPS@: network protein sequence analysis. Trends Biochem Sci 2000: 25: 147-150.

180. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002: 30: 3894-3900.

181. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 2002: 12: 436-446.

182. Thomas PD, Kejariwal A, Campbell MJ et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res 2003: 31: 334-341.

183. Yue P, Melamud E, Moult J. SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics 2006: 7: 166.

184. Ferrer-Costa C, Gelpi JL, Zamakola L et al. PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinformatics 2005: 21: 3176-3178.

185. Dreger M, Bengtsson L, Schoneberg T et al. Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane. Proc Natl Acad Sci U S A 2001: 98: 11943-11948.

186. Matsushita M, Yamadori T, Kato S et al. Identification and characterization of a novel SH3-domain binding protein, Sab, which preferentially associates with Bruton's tyrosine kinase (BtK). Biochem Biophys Res Commun 1998: 245: 337-343.

187. Bengtsson L, Otto H. LUMA interacts with emerin and influences its distribution at the inner nuclear membrane. J Cell Sci 2008: 121: 536-548.

188. Schirmer EC, Florens L, Guan T et al. Nuclear membrane proteins with potential disease links found by subtractive proteomics. Science 2003: 301: 1380-1382.

189. Bione S, Maestrini E, Rivella S et al. Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. Nat Genet 1994: 8: 323-327.

190. Bonne G, Di Barletta MR, Varnous S et al. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. Nat Genet 1999: 21: 285-288.

191. MacRae CA, Birchmeier W, Thierfelder L. Arrhythmogenic right ventricular cardiomyopathy: moving toward mechanism. J Clin Invest 2006: 116: 1825-1828.

192. Lemay DG, Hwang DH. Genome-wide identification of peroxisome proliferator response elements using integrated computational genomics. J Lipid Res 2006: 47: 1583-1587.

193. Molin DG, Bartram U, Van der Heiden K et al. Expression patterns of Tgfbeta1-3 associate with myocardialisation of the outflow tract and the development of the epicardium and the fibrous heart skeleton. Dev Dyn 2003: 227: 431-444.

194. Syrris P, Ward D, Asimaki A et al. Clinical Expression of Plakophilin-2 Mutations in Familial Arrhythmogenic Right Ventricular Cardiomyopathy 10.1161/CIRCULATIONAHA.105.561654. Circulation 2006: 113: 356-364.

195. Protonotarios N, Tsatsopoulou A, Anastasakis A et al. Genotype-phenotype assessment in autosomal recessive arrhythmogenic right ventricular cardiomyopathy (Naxos disease) caused by a deletion in plakoglobin. Journal of the American College of Cardiology 2001: 38: 1477-1484.

196. Lombardi R, Marian AJ. Arrhythmogenic right ventricular cardiomyopathy is a disease of cardiac stem cells. Curr Opin Cardiol 2010.

197. Ashley EA, Butte AJ, Wheeler MT et al. Clinical assessment incorporating a personal genome. Lancet 2010: 375: 1525-1535.

198. Christensen AH, Andersen CB, Tybjaerg-Hansen A et al. Mutation Analysis and Evaluation of the Cardiac Localization of TMEM43 in Arrhythmogenic Right Ventricular Cardiomyopathy. Clin Genet 2011.

199. Smith RJ, Camp GV. Deafness and Hereditary Hearing Loss Overview In: GeneReviews at GeneTests: Medical Genetics Information Resource (database online), Vol. 2008, 2007.

200. Hilgert N, Smith RJ, Van Camp G. Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics? Mutat Res 2009: 681: 189-196.

201. Schrijver I. Hereditary non-syndromic sensorineural hearing loss: transforming silence to sound. J Mol Diagn 2004: 6: 275-284.

202. Petersen MB, Willems PJ. Non-syndromic, autosomal-recessive deafness. Clin Genet 2006: 69: 371-392.

203. Friedman TB, Griffith AJ. Human nonsyndromic sensorineural deafness. Annu Rev Genomics Hum Genet 2003: 4: 341-402.

204. Kenneson A, Van Naarden Braun K, Boyle C. GJB2 (connexin 26) variants and nonsyndromic sensorineural hearing loss: a HuGE review. Genet Med 2002: 4: 258-274.

205. Snoeckx RL, Huygen PL, Feldmann D et al. GJB2 mutations and degree of hearing loss: a multicenter study. Am J Hum Genet 2005: 77: 945-957.

206. Lindenov H. The Etiology of Deaf-mutism with Special Reference to Heredity. Copenhagen: E Munksgaard (pub) 1945.

207. Tamayo ML, Bernal JE, Tamayo GE et al. Usher syndrome: results of a screening program in Colombia. Clin Genet 1991: 40: 304-311.

208. Seeliger MW, Fischer MD, Pfister M. [Usher syndrome: clinical features, diagnostic options, and therapeutic prospects]. Ophthalmologe 2009: 106: 505-511.

209. King RA, Rotter JI, Motulsky AG. The Genetic Basis of Common Diseases. Oxford: Oxford University Press, 2002.

210. Bolz H, von Brederlow B, Ramirez A et al. Mutation of CDH23, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D. Nat Genet 2001: 27: 108-112.

211. Bork JM, Peters LM, Riazuddin S et al. Usher syndrome 1D and nonsyndromic autosomal recessive deafness DFNB12 are caused by allelic mutations of the novel cadherin-like gene CDH23. Am J Hum Genet 2001: 68: 26-37.

212. Miki Y, Swensen J, Shattuck-Eidens D et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994: 266: 66-71.

213. Hussain R, Bittles AH. The prevalence and demographic characteristics of consanguineous marriages in Pakistan. J Biosoc Sci 1998: 30: 261-275.

214. Ben Arab S, Masmoudi S, Beltaief N et al. Consanguinity and endogamy in Northern Tunisia and its impact on non-syndromic deafness. Genet Epidemiol 2004: 27: 74-79.

215. Bittles AH. Consanguineous marriage and childhood health. Dev Med Child Neurol 2003: 45: 571-576.

216. Veske A, Oehlmann R, Younus F et al. Autosomal recessive non-syndromic deafness locus (DFNB8) maps on chromosome 21q22 in a large consanguineous kindred from Pakistan. Hum Mol Genet 1996: 5: 165-168.

217. Bonne-Tamir B, DeStefano AL, Briggs CE et al. Linkage of congenital recessive deafness (gene DFNB10) to chromosome 21q22.3. Am J Hum Genet 1996: 58: 1254-1259.

218. Scott HS, Kudoh J, Wattenhofer M et al. Insertion of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. Nat Genet 2001: 27: 59-63.

219. Robertson NG, Khetarpal U, Gutierrez-Espeleta GA et al. Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening. Genomics 1994: 23: 42-50.

220. Yasunaga S, Grati M, Cohen-Salmon M et al. A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes DFNB9, a nonsyndromic form of deafness. Nat Genet 1999: 21: 363-369.

221. Friedman TB, Liang Y, Weber JL et al. A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. Nat Genet 1995: 9: 86-91.

222. Liang Y, Wang A, Probst FJ et al. Genetic mapping refines DFNB3 to 17p11.2, suggests multiple alleles of DFNB3, and supports homology to the mouse model shaker-2. Am J Hum Genet 1998: 62: 904-915.

223. Probst FJ, Fridell RA, Raphael Y et al. Correction of deafness in shaker-2 mice by an unconventional myosin in a BAC transgene. Science 1998: 280: 1444-1447.

224. Wang A, Liang Y, Fridell RA et al. Association of unconventional myosin MYO15 mutations with human nonsyndromic deafness DFNB3. Science 1998: 280: 1447-1451.

225. Belyantseva IA, Boger ET, Naz S et al. Myosin-XVa is required for tip localization of whirlin and differential elongation of hair-cell stereocilia. Nat Cell Biol 2005: 7: 148-156.

226. Lesperance MM, Hall JW, 3rd, Bess FH et al. A gene for autosomal dominant nonsyndromic hereditary hearing impairment maps to 4p16.3. Hum Mol Genet 1995: 4: 1967-1972.

227. Van Camp G, Kunst H, Flothmann K et al. A gene for autosomal dominant hearing impairment (DFNA14) maps to a region on chromosome 4p16.3 that does not overlap the DFNA6 locus. J Med Genet 1999: 36: 532-536.

228. Bespalova IN, Van Camp G, J.H. Bom S et al. Mutations in the Wolfram syndrome 1 gene (WFS1) are a common cause of low frequency sensorineural hearing loss. Hum Mol Genet 2001: 10: 2501-2508.

229. Tsai HT, Wang YP, Chung SF et al. A novel mutation in the WFS1 gene identified in a Taiwanese family with low-frequency hearing impairment. BMC Med Genet 2007: 8: 26.

230. Cryns K, Sivakumaran T, Van den Ouweland J et al. Mutational spectrum of the WFS1 gene in Wolfram syndrome, nonsyndromic hearing impairment, diabetes mellitus, and psychiatric disease. Hum Mutat 2003: 22: 275-287.

231. Komatsu K, Nakamura N, Ghadami M et al. Confirmation of genetic homogeneity of nonsyndromic low-frequency sensorineural hearing loss by linkage analysis and a DFNA6/14 mutation in a Japanese family. J Hum Genet 2002: 47: 395-399.

232. Fukuoka H, Kanda Y, Ohta S et al. Mutations in the WFS1 gene are a frequent cause of autosomal dominant nonsyndromic low-frequency hearing loss in Japanese. J Hum Genet 2007: 52: 510-515.

233. Hildebrand MS, Sorensen JL, Jensen M et al. Autoimmune disease in a DFNA6/14/38 family carrying a novel missense mutation in WFS1. Am J Med Genet A 2008: 146A: 2258-2265.

234. Ahmed ZM, Li XC, Powell SD et al. Characterization of a new full length TMPRSS3 isoform and identification of mutant alleles responsible for nonsyndromic recessive deafness in Newfoundland and Pakistan. BMC Med Genet 2004: 5: 24.

235. Doucette L, Merner ND, Cooke S et al. Profound, prelingual nonsyndromic deafness maps to chromosome 10q21 and is caused by a novel missense mutation in the Usher syndrome type IF gene PCDH15. Eur J Hum Genet 2008.

236. Ahmed ZM, Riazuddin S, Ahmad J et al. PCDH15 is expressed in the neurosensory epithelium of the eye and ear and mutant alleles are responsible for both USH1F and DFNB23. Hum Mol Genet 2003: 12: 3215-3223.

237. Cremers CW, van Rijn PM, Huygen PL. The sex-ratio in childhood deafness, an analysis of the male predominance. Int J Pediatr Otorhinolaryngol 1994: 30: 105-110.

238. Fraser GR. Sex-linked recessive congenital deafness and the excess of males in profound childhood deafness. Ann Hum Genet 1965: 29: 171-196.

239. Reardon W. Sex linked deafness: Wilde revisited. J Med Genet 1990: 27: 376-379.

240. Migeon BR. X inactivation, female mosaicism, and sex differences in renal diseases. J Am Soc Nephrol 2008: 19: 2052-2059.

241. Dobyns WB. The pattern of inheritance of X-linked traits is not dominant or recessive, just X-linked. Acta Paediatr Suppl 2006: 95: 11-15.

242. Mohr J, Mageroy K. Sex-linked deafness of a possibly new type. Acta Genet Stat Med 1960: 10: 54-62.

243. Tranebjaerg L, Schwartz C, Eriksen H et al. A new X linked recessive deafness syndrome with blindness, dystonia, fractures, and mental deficiency is linked to Xq22. J Med Genet 1995: 32: 257-263.

244. Liu X, Han D, Li J et al. Loss-of-function mutations in the PRPS1 gene cause a type of nonsyndromic X-linked sensorineural deafness, DFN2. Am J Hum Genet 2010: 86: 65-71.

245. de Kok YJ, van der Maarel SM, Bitner-Glindzicz M et al. Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. Science 1995: 267: 685-688.

246. Brunner HG, van Bennekom A, Lambermon EM et al. The gene for X-linked progressive mixed deafness with perilymphatic gusher during stapes surgery (DFN3) is linked to PGK. Hum Genet 1988: 80: 337-340.

247. Wallis C, Ballo R, Wallis G et al. X-linked mixed deafness with stapes fixation in a Mauritian kindred: linkage to Xq probe pDP34. Genomics 1988: 3: 299-301.

248. Reardon W, Middleton-Price HR, Sandkuijl L et al. A multipedigree linkage study of X-linked deafness: linkage to Xq13-q21 and evidence for genetic heterogeneity. Genomics 1991: 11: 885-894.

249. Bach I, Brunner HG, Beighton P et al. Microdeletions in patients with gusher-associated, X-linked mixed deafness (DFN3). Am J Hum Genet 1992: 51: 38-44.

250. Reardon W, Middleton-Price HR, Malcolm S et al. Clinical and genetic heterogeneity in X-linked deafness. Br J Audiol 1992: 26: 109-114.

251. Tyson J, Bellman S, Newton V et al. Mapping of DFN2 to Xq22. Hum Mol Genet 1996: 5: 2055-2060.

252. Manolis EN, Eavey RD, Sangwatanaroj S et al. Hereditary postlingual sensorineural hearing loss mapping to chromosome Xq21. Am J Otol 1999: 20: 621-626.

253. Cui B, Zhang H, Lu Y et al. Refinement of the locus for non-syndromic sensorineural deafness (DFN2). J Genet 2004: 83: 35-38.

254. Lee HK, Lee SH, Lee KY et al. Novel POU3F4 mutations and clinical features of DFN3 patients with cochlear implants. Clin Genet 2009: 75: 572-575.

255. Bitner-Glindzicz M, Turnpenny P, Hoglund P et al. Further mutations in Brain 4 (POU3F4) clarify the phenotype in the X-linked deafness, DFN3. Hum Mol Genet 1995: 4: 1467-1469.

256. de Kok YJ, Vossenaar ER, Cremers CW et al. Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. Hum Mol Genet 1996: 5: 1229-1235.

257. de Kok YJ, Merkx GF, van der Maarel SM et al. A duplication/paracentric inversion associated with familial X-linked deafness (DFN3) suggests the presence of a regulatory element more than 400 kb upstream of the POU3F4 gene. Hum Mol Genet 1995: 4: 2145-2150.

258. Hagiwara H, Tamagawa Y, Kitamura K et al. A new mutation in the POU3F4 gene in a Japanese family with X-linked mixed deafness (DFN3). Laryngoscope 1998: 108: 1544-1547.

259. Friedman RA, Bykhovskaya Y, Tu G et al. Molecular analysis of the POU3F4 gene in patients with clinical and radiographic evidence of X-linked mixed deafness with perilymphatic gusher. Ann Otol Rhinol Laryngol 1997: 106: 320-325.

260. Vore AP, Chang EH, Hoppe JE et al. Deletion of and novel missense mutation in POU3F4 in 2 families segregating X-linked nonsyndromic deafness. Arch Otolaryngol Head Neck Surg 2005: 131: 1057-1063.

261. Lalwani AK, Brister JR, Fex J et al. A new nonsyndromic X-linked sensorineural hearing impairment linked to Xp21.2. Am J Hum Genet 1994: 55: 685-694.

262. Pfister MH, Apaydin F, Turan O et al. A second family with nonsyndromic sensorineural hearing loss linked to Xp21.2: refinement of the DFN4 locus within DMD. Genomics 1998: 53: 377-382.

263. Raynor EM, Mulroy MJ. Sensorineural hearing loss in the mdx mouse: a model of Duchenne muscular dystrophy. Laryngoscope 1997: 107: 1053-1056.

264. Pillers DA, Duncan NM, Dwinnell SJ et al. Normal cochlear function in mdx and mdx(Cv3) Duchenne muscular dystrophy mouse models. Laryngoscope 1999: 109: 1310-1312.

265. del Castillo I, Villamar M, Sarduy M et al. A novel locus for non-syndromic sensorineural deafness (DFN6) maps to chromosome Xp22. Hum Mol Genet 1996: 5: 1383-1387.

266. Schouten JP, McElgunn CJ, Waaijer R et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 2002: 30: e57.

267. Kesler SR, Simensen RJ, Voeller K et al. Altered neurodevelopment associated with mutations of RSK2: a morphometric MRI study of Coffin-Lowry syndrome. Neurogenetics 2007: 8: 143-147.

268. Hanauer A, Young ID. Coffin-Lowry syndrome: clinical and molecular features. J Med Genet 2002: 39: 705-713.

269. Lyon MF, Scriver CR, Baker LR et al. The Gy mutation: another cause of X-linked hypophosphatemia in mouse. Proc Natl Acad Sci U S A 1986: 83: 4899-4903.

270. Sabbagh Y, Jones AO, Tenenhouse HS. PHEXdb, a locus-specific database for mutations causing X-linked hypophosphatemia. Hum Mutat 2000: 16: 1-6.

271. del Castillo I, Rodriguez M, Cruz Tapia M et al. X-linked non-syndromic sensorineural deafness: the DFN6 locus. Adv Otorhinolaryngol 2000: 56: 200-202.

272. Morton CC. Gene discovery in the auditory system using a tissue specific approach. Am J Med Genet A 2004: 130A: 26-28.

273. Ross MT, Grafham DV, Coffey AJ et al. The DNA sequence of the human X chromosome. Nature 2005: 434: 325-337.

274. Abbs S, Roberts RG, Mathew CG et al. Accurate assessment of intragenic recombination frequency within the Duchenne muscular dystrophy gene. Genomics 1990: 7: 602-606.

275. Oudet C, Heilig R, Hanauer A et al. Nonradioactive assay for new microsatellite polymorphisms at the 5' end of the dystrophin gene, and estimation of intragenic recombination. Am J Hum Genet 1991: 49: 311-319.

276. Oudet C, Hanauer A, Clemens P et al. Two hot spots of recombination in the DMD gene correlate with the deletion prone regions. Hum Mol Genet 1992: 1: 599-603.

277. Ng SB, Buckingham KJ, Lee C et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010: 42: 30-35.

278. Ng SB, Bigham AW, Buckingham KJ et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010: 42: 790-793.

279. Zheng QY, Yan D, Ouyang XM et al. Digenic inheritance of deafness caused by mutations in genes encoding cadherin 23 and protocadherin 15 in mice and humans. Hum Mol Genet 2005: 14: 103-111.

280. Kazmierczak P, Sakaguchi H, Tokita J et al. Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells. Nature 2007: 449: 87-91.

281. Canadian Cancer Statistics 2010. Special Topic: End-of-Life Care: Produced by: Canadian Cancer Society, Statistics Canada, Provincial/Territorial Cancer Registries and Public Health Agency of Canada

282. Garcia MJ, Benitez J. The Fanconi anaemia/BRCA pathway and cancer susceptibility. Searching for new therapeutic targets. Clin Transl Oncol 2008: 10: 78-84.

283. Hollestelle A, Wasielewski M, Martens JW et al. Discovering moderate-risk breast cancer susceptibility genes. Curr Opin Genet Dev 2010: 20: 268-276.

284. Ripperger T, Gadzicki D, Meindl A et al. Breast cancer susceptibility: current knowledge and implications for genetic counselling. Eur J Hum Genet 2009: 17: 722-731.

285. Cazier JB, Tomlinson I. General lessons from large-scale studies to identify human cancer predisposition genes. J Pathol 2010: 220: 255-262.

286. Walsh T, King MC. Ten genes for inherited breast cancer. Cancer Cell 2007: 11: 103-105.

287. Wooster R, Bignell G, Lancaster J et al. Identification of the breast cancer susceptibility gene BRCA2. Nature 1995: 378: 789-792.

288. Antoniou A, Pharoah PD, Narod S et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003: 72: 1117-1130.

289. Meindl A, Hellebrand H, Wiek C et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet 2010: 42: 410-414.

290. Meijers-Heijboer H, van den Ouweland A, Klijn J et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 2002: 31: 55-59.

291. Vahteristo P, Bartkova J, Eerola H et al. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. Am J Hum Genet 2002: 71: 432-438.

292. Seal S, Thompson D, Renwick A et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet 2006: 38: 1239-1241.

293. Rahman N, Seal S, Thompson D et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet 2007: 39: 165-167.

294. Steffen J, Nowakowska D, Niwinska A et al. Germline mutations 657del5 of the NBS1 gene contribute significantly to the incidence of breast cancer in Central Poland. Int J Cancer 2006: 119: 472-475.

295. Baynes C, Healey CS, Pooley KA et al. Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk. Breast Cancer Res 2007: 9: R27.

296. Hunter DJ, Kraft P, Jacobs KB et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007: 39: 870-874.

297. Easton DF, Pooley KA, Dunning AM et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007: 447: 1087-1093.

298. Cox A, Dunning AM, Garcia-Closas M et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 2007: 39: 352-358.

299. Stacey SN, Manolescu A, Sulem P et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007: 39: 865-869.

300. Ahmed S, Thomas G, Ghoussaini M et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 2009: 41: 585-590.

301. Milne RL, Benitez J, Nevanlinna H et al. Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. J Natl Cancer Inst 2009: 101: 1012-1018.

302. Thomas G, Jacobs KB, Kraft P et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009: 41: 579-584.

303. Zheng W, Cai Q, Signorello LB et al. Evaluation of 11 breast cancer susceptibility loci in African-American women. Cancer Epidemiol Biomarkers Prev 2009: 18: 2761-2764.

304. Bourdon JC. p53 and its isoforms in cancer. Br J Cancer 2007: 97: 277-282.

305. Gonzalez K, Fong C, Buzin C et al. p53 Testing for Li-Fraumeni and Li-Fraumeni-like syndromes. Curr Protoc Hum Genet 2008: Chapter 10: Unit 10 10.

306. Li L, Ernsting BR, Wishart MJ et al. A family of putative tumor suppressors is structurally and functionally conserved in humans and yeast. J Biol Chem 1997: 272: 29403-29406.

307. Gustafson S, Zbuk KM, Scacheri C et al. Cowden syndrome. Semin Oncol 2007: 34: 428-434.

308. Hall JM, Lee MK, Newman B et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 1990: 250: 1684-1689.

309. Garvin AM, Attenhofer-Haner M, Scott RJ. BRCA1 and BRCA2 mutation analysis in 86 early onset breast/ovarian cancer patients. J Med Genet 1997: 34: 990-995.

310. Narod SA, Foulkes WD. BRCA1 and BRCA2: 1994 and beyond. Nat Rev Cancer 2004: 4: 665-676.

311. Futreal PA, Liu Q, Shattuck-Eidens D et al. BRCA1 mutations in primary breast and ovarian carcinomas. Science 1994: 266: 120-122.

312. Wu LC, Wang ZW, Tsan JT et al. Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nat Genet 1996: 14: 430-440.

313. Ruffner H, Joazeiro CA, Hemmati D et al. Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. Proc Natl Acad Sci U S A 2001: 98: 5134-5139.

314. Baer R, Ludwig T. The BRCA1/BARD1 heterodimer, a tumor suppressor complex with ubiquitin E3 ligase activity. Curr Opin Genet Dev 2002: 12: 86-91.

315. Hashizume R, Fukuda M, Maeda I et al. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. J Biol Chem 2001: 276: 14537-14540.

316. Wang B, Matsuoka S, Ballif BA et al. Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. Science 2007: 316: 1194-1198.

317. Yu X, Chen J. DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. Mol Cell Biol 2004: 24: 9478-9486.

318. Greenberg RA, Sobhian B, Pathania S et al. Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. Genes Dev 2006: 20: 34-46.

319. Wooster R, Neuhausen SL, Mangion J et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 1994: 265: 2088-2090.

320. Xia F, Taghian DG, DeFrank JS et al. Deficiency of human BRCA2 leads to impaired homologous recombination but maintains normal nonhomologous end joining. Proc Natl Acad Sci U S A 2001: 98: 8644-8649.

321. Wagner JE, Tolar J, Levran O et al. Germline mutations in BRCA2: shared genetic susceptibility to breast cancer, early onset leukemia, and Fanconi anemia. Blood 2004: 103: 3226-3229.

322. Ali AM, Singh TR, Meetei AR. FANCM-FAAP24 and FANCJ: FA proteins that metabolize DNA. Mutat Res 2009: 668: 20-26.

323. Hartikainen JM, Kataja V, Pirskanen M et al. Screening for BRCA1 and BRCA2 mutations in Eastern Finnish breast/ovarian cancer families. Clin Genet 2007: 72: 311-320.

324. Ferla R, Calo V, Cascio S et al. Founder mutations in BRCA1 and BRCA2 genes. Ann Oncol 2007: 18 Suppl 6: vi93-98.

325. Neuhausen SL, Godwin AK, Gershoni-Baruch R et al. Haplotype and phenotype analysis of nine recurrent BRCA2 mutations in 111 families: results of an international study. Am J Hum Genet 1998: 62: 1381-1388.

326. Ramus SJ, Gayther SA. The contribution of BRCA1 and BRCA2 to ovarian cancer. Mol Oncol 2009: 3: 138-150.

327. Ramus SJ, Harrington PA, Pye C et al. Contribution of BRCA1 and BRCA2 mutations to inherited ovarian cancer. Hum Mutat 2007: 28: 1207-1215.

328. Ford D, Easton DF, Stratton M et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am J Hum Genet 1998: 62: 676-689.

329. Antoniou AC, Pharoah PD, McMullan G et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer 2002: 86: 76-83.

330. Gayther SA, Mangion J, Russell P et al. Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene. Nat Genet 1997: 15: 103-105.

331. Friedenson B. BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. MedGenMed 2005: 7: 60.

332. BIC. Breast Cancer Information Core of the National Human Genome Reserach Institute. 2009.

333. Bellosillo B, Tusquets I. Pitfalls and caveats in BRCA sequencing. Ultrastruct Pathol 2006: 30: 229-235.

334. Osorio A, Milne RL, Honrado E et al. Classification of missense variants of unknown significance in BRCA1 based on clinical and tumor information. Hum Mutat 2007: 28: 477-485.

335. Carvalho MA, Couch FJ, Monteiro AN. Functional assays for BRCA1 and BRCA2. Int J Biochem Cell Biol 2007: 39: 298-310.

336. Shattuck-Eidens D, Oliphant A, McClure M et al. BRCA1 sequence analysis in women at high risk for susceptibility mutations. Risk factor analysis and implications for genetic testing. Jama 1997: 278: 1242-1250.

337. Couch FJ, Weber BL. Mutations and polymorphisms in the familial early-onset breast cancer (BRCA1) gene. Breast Cancer Information Core. Hum Mutat 1996: 8: 8-18.

338. Puget N, Torchard D, Serova-Sinilnikova OM et al. A 1-kb Alu-mediated germ-line deletion removing BRCA1 exon 17. Cancer Res 1997: 57: 828-831.

339. Smith TM, Lee MK, Szabo CI et al. Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. Genome Res 1996: 6: 1029-1049.

340. Gad S, Caux-Moncoutier V, Pages-Berhouet S et al. Significant contribution of large BRCA1 gene rearrangements in 120 French breast and ovarian cancer families. Oncogene 2002: 21: 6841-6847.

341. Hofmann W, Gorgens H, John A et al. Screening for large rearrangements of the BRCA1 gene in German breast or ovarian cancer families using semi-quantitative multiplex PCR method. Hum Mutat 2003: 22: 103-104.

342. Miki Y, Katagiri T, Kasumi F et al. Mutation analysis in the BRCA2 gene in primary breast cancers. Nat Genet 1996: 13: 245-247.

343. Nordling M, Karlsson P, Wahlstrom J et al. A large deletion disrupts the exon 3 transcription activation domain of the BRCA2 gene in a breast/ovarian cancer family. Cancer Res 1998: 58: 1372-1375.

344. Wang T, Lerer I, Gueta Z et al. A deletion/insertion mutation in the BRCA2 gene in a breast cancer family: a possible role of the Alu-polyA tail in the evolution of the deletion. Genes Chromosomes Cancer 2001: 31: 91-95.

345. Palma MD, Domchek SM, Stopfer J et al. The relative contribution of point mutations and genomic rearrangements in BRCA1 and BRCA2 in high-risk breast cancer families. Cancer Res 2008: 68: 7006-7014.

346. Hansen TO, Jonson L, Albrechtsen A et al. Large BRCA1 and BRCA2 genomic rearrangements in Danish high risk breast-ovarian cancer families. Breast Cancer Res Treat 2009: 115: 315-323.

347. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet 1991: 48: 232-242.

348. Thorlacius S, Sigurdsson S, Bjarnadottir H et al. Study of a single BRCA2 mutation with high carrier frequency in a small population. Am J Hum Genet 1997: 60: 1079-1084.

349. Thorlacius S, Olafsdottir G, Tryggvadottir L et al. A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. Nat Genet 1996: 13: 117-119.

350. Oddoux C, Struewing JP, Clayton CM et al. The carrier frequency of the BRCA2 6174delT mutation among Ashkenazi Jewish individuals is approximately 1%. Nat Genet 1996: 14: 188-190.

351. Roa BB, Boyd AA, Volcik K et al. Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. Nat Genet 1996: 14: 185-187.

352. Struewing JP, Abeliovich D, Peretz T et al. The carrier frequency of the BRCA1 185delAG mutation is approximately 1 percent in Ashkenazi Jewish individuals. Nat Genet 1995: 11: 198-200.

353. Abeliovich D, Kaduri L, Lerer I et al. The founder mutations 185delAG and 5382insC in BRCA1 and 6174delT in BRCA2 appear in 60% of ovarian cancer and 30% of early-onset breast cancer patients among Ashkenazi women. Am J Hum Genet 1997: 60: 505-514.

354. Hogervorst FB, Nederlof PM, Gille JJ et al. Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. Cancer Res 2003: 63: 1449-1453.

355. Verhoog LC, van den Ouweland AM, Berns E et al. Large regional differences in the frequency of distinct BRCA1/BRCA2 mutations in 517 Dutch breast and/or ovarian cancer families. Eur J Cancer 2001: 37: 2082-2090.

356. Walsh T, Casadei S, Coats KH et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. Jama 2006: 295: 1379-1388.

357. Nevanlinna H, Bartek J. The CHEK2 gene and inherited breast cancer susceptibility. Oncogene 2006: 25: 5912-5919.

358. Savitsky K, Bar-Shira A, Gilad S et al. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 1995: 268: 1749-1753.

359. Cantor SB, Bell DW, Ganesan S et al. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell 2001: 105: 149-160.

360. Levran O, Attwooll C, Henry RT et al. The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. Nat Genet 2005: 37: 931-933.

361. Varon R, Vissinga C, Platzer M et al. Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. Cell 1998: 93: 467-476.

362. Gorski B, Debniak T, Masojc B et al. Germline 657del5 mutation in the NBS1 gene in breast cancer patients. Int J Cancer 2003: 106: 379-381.

363. Bogdanova N, Schurmann P, Waltes R et al. NBS1 variant I171V and breast cancer risk. Breast Cancer Res Treat 2007.

364. Roznowski K, Januszkiewicz-Lewandowska D, Mosor M et al. I171V germline mutation in the NBS1 gene significantly increases risk of breast cancer. Breast Cancer Res Treat 2008: 110: 343-348.

365. Heikkinen K, Karppinen SM, Soini Y et al. Mutation screening of Mre11 complex genes: indication of RAD50 involvement in breast and ovarian cancer susceptibility. J Med Genet 2003: 40: e131.

366. Heikkinen K, Rapakko K, Karppinen SM et al. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis 2006: 27: 1593-1599.

367. Xia B, Sheng Q, Nakanishi K et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell 2006: 22: 719-729.

368. Xia B, Dorsman JC, Ameziane N et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nat Genet 2007: 39: 159-161.

369. Southey MC, Teo ZL, Dowty JG et al. A PALB2 mutation associated with high risk of breast cancer. Breast Cancer Res 2010: 12: R109.

370. Erkko H, Xia B, Nikkila J et al. A recurrent mutation in PALB2 in Finnish cancer families. Nature 2007: 446: 316-319.

371. Vaz F, Hanenberg H, Schuster B et al. Mutation of the RAD51C gene in a Fanconi anemia-like disorder. Nat Genet 2010: 42: 406-409.

372. Akbari MR, Tonin P, Foulkes WD et al. RAD51C germline mutations in breast and ovarian cancer patients. Breast Cancer Res 2010: 12: 404.

373. Heiskanen M, Kononen J, Barlund M et al. CGH, cDNA and tissue microarray analyses implicate FGFR2 amplification in a small subset of breast tumors. Anal Cell Pathol 2001: 22: 229-234.

374. Gold B, Kirchhoff T, Stefanov S et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci U S A 2008: 105: 4340-4345.

375. Meyer KB, Maia AT, O'Reilly M et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. PLoS Biol 2008: 6: e108.

376. Bergman A, Karlsson P, Berggren J et al. Genome-wide linkage scan for breast cancer susceptibility loci in Swedish hereditary non-BRCA1/2 families: suggestive linkage to 10q23.32-q25.3. Genes Chromosomes Cancer 2007: 46: 302-309.

377. Gonzalez-Neira A, Rosa-Rosa JM, Osorio A et al. Genomewide high-density SNP linkage analysis of non-BRCA1/2 breast cancer families identifies various candidate regions and has greater power than microsatellite studies. BMC Genomics 2007: 8: 299.

378. Huusko P, Juo SH, Gillanders E et al. Genome-wide scanning for linkage in Finnish breast cancer families. Eur J Hum Genet 2004: 12: 98-104.

379. Smith P, McGuffog L, Easton DF et al. A genome wide linkage search for breast cancer susceptibility genes. Genes Chromosomes Cancer 2006: 45: 646-655.

380. Roest PA, Roberts RG, Sugino S et al. Protein truncation test (PTT) for rapid detection of translation-terminating mutations. Hum Mol Genet 1993: 2: 1719-1721.

381. Den Dunnen JT, Van Ommen GJ. The protein truncation test: A review. Hum Mutat 1999: 14: 95-102.

382. Orita M, Iwahana H, Kanazawa H et al. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci U S A 1989: 86: 2766-2770.

383. Payne SR, Newman B, King MC. Complex germline rearrangement of BRCA1 associated with breast and ovarian cancer. Genes Chromosomes Cancer 2000: 29: 58-62.

384. Puget N, Sinilnikova OM, Stoppa-Lyonnet D et al. An Alu-mediated 6-kb duplication in the BRCA1 gene: a new founder mutation? Am J Hum Genet 1999: 64: 300-302.

385. Ward BD, Hendrickson BC, Judkins T et al. A multi-exonic BRCA1 deletion identified in multiple families through single nucleotide polymorphism haplotype pair analysis and gene amplification with widely dispersed primer sets. J Mol Diagn 2005: 7: 139-142.

386. Carson N, Gilpin C, Hunter A et al. An in frane deletion of BRCA1 exon 20 in a family with early onset breast cancer and ovarian cancer. Am J Hum Genet 1999: 65: 1610.

387. Petrij-Bosch A, Peelen T, van Vliet M et al. BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. Nat Genet 1997: 17: 341-345.

388. Szabo CI, King MC. Population genetics of BRCA1 and BRCA2. Am J Hum Genet 1997: 60: 1013-1020.

389. Sokolenko AP, Mitiushkina NV, Buslov KG et al. High frequency of BRCA1 5382insC mutation in Russian breast cancer patients. Eur J Cancer 2006: 42: 1380-1384.

390. Olufemi S, Green J, Manickam P et al. common ancestral mutation in the MEN-1 gene is likely responsible for the prolactinoma variant of MEN1 (MEN1 burin) in four kindreds from Newfoundland. Human Mutation 1998: 11: 264-269.

391. BIC. Breast Cancer Information Core of the National Human Genome Reserach Institute. 2007.

392. Devilee P, Hogervorst F. Summary of Published and Unpublished BRCA2 Mutations in the Netherlands and Belgium. 2002.

393. Couch FJ, DeShano ML, Blackwood MA et al. BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. N Engl J Med 1997: 336: 1409-1415.

394. Frank TS, Manley SA, Olopade OI et al. Sequence analysis of BRCA1 and BRCA2: correlation of mutations with family history and ovarian cancer risk. J Clin Oncol 1998: 16: 2417-2425.

395. Spiegelman D, Colditz GA, Hunter D et al. Validation of the Gail et al. model for predicting individual breast cancer risk. J Natl Cancer Inst 1994: 86: 600-607.

396. Eichers ER, Green JS, Stockton DW et al. Newfoundland rod-cone dystrophy, an early-onset retinal dystrophy, is caused by splice-junction mutations in RLBP1. Am J Hum Genet 2002: 70: 955-964.

# Appendix 1

## Standard 1X PCR cocktail

The Invitrogen™ *Taq* DNA Polymerase kit (Cat. #: 10342020) was used for the amplification.

A standard 1X PCR reaction cocktail without betaine contained:

2.5 µl of 10X PCR Buffer

0.75 µl of $MgCl_2$ (50 mM)

2.5 µl of dNTPs (2 mM)

0.2 µl of Taq DNA polymerase (5 Units/µl)

16.05 µl of $H_2O$

1 µl of each 10 µM forward and reverse primer

1 µl of 25-100 ng/µl DNA template.

---

Total = 25 µl

A standard 1X PCR reaction cocktail with betaine was the same as above except 5 µl of 3.75 M betaine was used and only 11.05 µl of $H_2O$.

# Appendix 2

## Touchdown (TD) PCR

Touchdown PCR is a way to increase specificity. The stringency of primer hybridization (annealing) is very high at temperatures above the Tm, thus at these temperatures spurious products are not favored to anneal and the desired product is predominantly amplified. Therefore, the annealing temperature is first set above the expected Tm and lowered with each additional cycle to a temperature below the Tm.

An example of a typical touchdown cycle is TD 54. There are 5 initial touchdown cycles. The first cycle has an annealing temperature 8°C above (62°C) the desired annealing temperature (54°C). Then in each additional touchdown cycle the annealing temperature decreases by two degrees until it reaches 54°C. The PCR then continues for an additional 30 cycles at annealing temperature 54°C in order to get the desired amount of product. All steps (denaturation, annealing and extension) are 30 seconds long.

# Appendix 3

## *ARVD5* positional genes – primers and amplification conditions

| IQSEC1 (NM_014869) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_014869-Ex1F | acttcagggccacattcctc | 518 | TD58 | yes (0 75) | 1 5 |
| | NM_014869-Ex1R | CTCACCCAAACCCACAGTCT | | | | |
| 2 | NM_014869-Ex2F | ccatggggtgtgagaagaac | 588 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex2R | agctacaggtggagctggaa | | | | |
| 3 | NM_014869-Ex3aF | tgaccaactttgggaccttc | 827 | TD56 | yes (0 75) | 1 5 |
| | NM_014869-Ex3aR | CGTGTCAGCCTTGTCCTCTT | | | | |
| | NM_014869-Ex3bF | CATCGATGAGGAGGAGCTGT | 828 | TD56 | yes (0 75) | 2 5 |
| | NM_014869-Ex3bR | ttgtgtggcagcttcctatg | | | | |
| 4 | NM_014869-Ex4#2F | cagcctgggttcttgagttc | 465 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex4#2R | acatgtcccagcaagtaggg | | | | |
| 5 | NM_014869-Ex5F | ctccccagaagactggaggt | 400 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex5R | atgcagtctttggtccatcc | | | | |
| 6 | NM_014869-Ex6F | gagtgaggagggcaagtgag | 500 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex6R | gtcatctgtgctccccaaag | | | | |
| 7 | NM_014869-Ex7F | tgggtctagtggaaggttcg | 345 | TD54 | no | 1 5 |
| | NM_014869-Ex7R | tggaagagaccacacggagt | | | | |
| 8 | NM_014869-Ex8F | tcccaagggtaggtgtaggg | 243 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex8R | aactcaagcagccagcactt | | | | |
| 9 | NM_014869-Ex9F | ctgtgagggaaggcattttt | 400 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex9R | tgtgaaagacaccaggcaaa | | | | |
| 10 | NM_014869-Ex10F | tctgaccagaggtcccacat | 370 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex10R | tgaggctcagctgactgatg | | | | |
| 11 | NM_014869-Ex11F | gcacaactcagccctaccat | 388 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex11R | agggaaggccagaaaagact | | | | |
| 12 | NM_014869-Ex12F | ggacagtgctagctccaagc | 475 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex12R | tttgacgtttgggtcctgtt | | | | |
| 13 | NM_014869-Ex13F | cctcatgtccgcgtctaact | 250 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex13R | ctctgtgcagtaggggaagg | | | | |
| 14 | NM_014869-Ex14aF | ctgtgctcttgtcgcttctg | 572 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex14aR | GGTCCCATGCTGTTTCTTTC | | | | |
| | NM_014869-Ex14bF | TCCAGAGAGAGCAAATACAGCA | 679 | TD54 | yes (0 75) | 1 5 |
| | NM_014869-Ex14bR | GTTCCAGCAGTTCCTTACGC | | | | |
| | NM_014869-Ex14cF | GAAAAGATTCAGGCCGTTTG | 826 | TD56 | yes (0 75) | 1 5 |
| | NM_014869-Ex14cR | tgcagtgtgtgagggaaaga | | | | |

| NUP210 (NM_024923) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_024923-Ex1-F | caagccttctccctcagctc | 624 | TD56 | yes (1 2) | 1 5 |
| | NM_024923 Ex1-R | CGCCATGACATGAGCAGT | | | | |
| 2 | NM_024923-Ex2-F | cctcagcttgtggaggtctt | 467 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex2-R | tcctgccactcaacaagaga | | | | |
| 3 | NM_024923-Ex3-F | cttggcttctgagggtgaag | 393 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex3-R | ctgaccagagctgccctcta | | | | |
| 4 | NM_024923-Ex4-F | tggtgatgtggagtgtggtt | 281 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex4-R | ctggtgaggagctcacaggt | | | | |
| 5 | NM_024923-Ex5-F | ggacacccctatttcagca | 397 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex5-R | acgggacagcatctcacttc | | | | |
| 6 | NM_024923-Ex6-F | gaaagtgaggatttgggaatga | 398 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex6-R | tgctacctgatggagcactt | | | | |
| 7 | NM_024923-Ex7-F | gcgtcagctgtactgagagc | 394 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex7-R | gttgaggggaaacggctatt | | | | |
| 8 | NM_024923-Ex8-F | cctaagccaggggagttttc | 339 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex8-R | tggtcatagcaagtggcaga | | | | |
| 9 | NM_024923-Ex9-F | cttctcgtgtggatgctgag | 375 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex9-R | agcactgtttcccacagacc | | | | |
| 10 | NM_024923-Ex10-F | tagcttcgtgcccaaattct | 353 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex10-R | atgcaagtagcccaaacagg | | | | |
| 11 | NM_024923-Ex11-F | ggcaccttcctctccactct | 500 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex11-R | gcaccctgccttattctcaa | | | | |
| 12 | NM_024923-Ex12-F | gagtgtgtgcatgggtgaac | 388 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex12-R | agtgaagtcgaggcttgctc | | | | |
| 13 | NM_024923-Ex13-F | gcacacagcagcacttagga | 384 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex13-R | aagccctctctttgggaatc | | | | |
| 14 | NM_024923-Ex14-F | aaagtgatttcagggtgacca | 248 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex14-R | aatttgcatctgggaaggag | | | | |
| 15 | NM_024923-Ex15-F | gctgcagatgacgagaatga | 487 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex15-R | aggagtggctgtctctcagc | | | | |
| 16 | NM_024923-Ex16-F | aaaactgagtcagagggagctg | 390 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex16-R | gaagtgcagcgtggagctat | | | | |
| 17 | NM_024923-Ex17-F | tttctggcatatcccattcc | 422 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex17-R | actaagctcccacaggcact | | | | |
| 18 | NM_024923-Ex18-F | cgaaactgcaaatcagaagga | 292 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex18-R | gactcaggacactgtgataccc | | | | |
| 19 | NM_024923-Ex19&20-F | agacctcaaagggtctgtgg | 572 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex19&20-R | tccgagtgtgagtcaagtgc | | | | |
| 20 | NM_024923-Ex19&20-F | agacctcaaagggtctgtgg | 572 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex19&20-R | tccgagtgtgagtcaagtgc | | | | |
| 21 | NM_024923-Ex21-F | gcactgctcacaaacctgag | 397 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex21-R | agggactctggatcctccac | | | | |
| 22 | NM_024923-Ex22&23-F | tgggtgactctcctgattcc | 600 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex22&23-R | tctctacagtagcttcctggactg | | | | |
| 23 | NM_024923-Ex22&23-F | tgggtgactctcctgattcc | 600 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex22&23-R | tctctacagtagcttcctggactg | | | | |
| 24 | NM_024923-Ex24&25-F | gatgtccatgtcccaggttt | 642 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex24&25-R | gggtggtcctcagaggtaga | | | | |
| 25 | NM_024923-Ex24&25-F | gatgtccatgtcccaggttt | 642 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex24&25-R | gggtggtcctcagaggtaga | | | | |
| 26 | NM_024923-Ex26-F | atggtgctggtcatcctga | 340 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex26-R | gaagcaggccctagataccc | | | | |
| 27 | NM_024923-Ex27-F | aacctgtgccatgctcttct | 473 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex27-R | gcacctactcaccccgtcta | | | | |
| 28 | NM_024923-Ex28-F | cagctctggggttagcagac | 500 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex28-R | atcacacccggatgtttacc | | | | |
| 29 | NM_024923-Ex29-F | aggtcagcccttcacatcat | 374 | TD54 | no | 1 5 |
| | NM_024923-Ex29-R | gcacagcaggagaaaacctc | | | | |
| 30 | NM_024923-Ex30-F | gcatgaggagccacagactc | 465 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex30-R | atcttcggctctgatgacca | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 31 | NM_024923-Ex31-F | gcattgttcctgaggtggat | 429 | TD54 | no | 1.5 |
| | NM_024923-Ex31-R | atgcccacatctcttggaaa | | | | |
| 32 | NM_024923-Ex32-F | gctcagcagaggacatggat | 500 | TD54 | no | 1.5 |
| | NM_024923-Ex32-R | cttgcagaaaacaggcacag | | | | |
| 33 | NM_024923-Ex33-F | gtctcctgaggctgctgagt | 383 | TD54 | no | 1 5 |
| | NM_024923-Ex33-R | tcagacggcttttccctaaa | | | | |
| 34 | NM_024923-Ex34-F | caaggcgtttaggctttgag | 393 | TD54 | no | 1 5 |
| | NM_024923-Ex34-R | ccctctctggagttggtgtg | | | | |
| 35 | NM_024923-Ex35-F | ctggtggaatttggtgtggt | 391 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex35-R | accctctgtgaccagcagtc | | | | |
| 36 | NM_024923-Ex36-F | gacttctcacctgggctttg | 477 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex36-R | taaagagagctggcccagga | | | | |
| 37 | NM_024923-Ex37-F | cagagatggccatgcagag | 399 | TD54 | no | 1 5 |
| | NM_024923-Ex37-R | ctgatagccaggaccactcc | | | | |
| 38 | NM_024923-Ex38&39-F | cctctgagaccaccatcctt | 489 | TD54 | no | 1 5 |
| | NM_024923-Ex38&39-R | gggcctgttgagacttaggg | | | | |
| 39 | NM_024923-Ex38&39-F | cctctgagaccaccatcctt | 489 | TD54 | no | 1.5 |
| | NM_024923-Ex38&39-R | gggcctgttgagacttaggg | | | | |
| 40 | NM_024923-Ex40a-F | tgagagtgtcctgggtgagg | 499 | TD54 | yes (0 75) | 1 5 |
| | NM_024923-Ex40a-R | CAGCACATTGTCCAGAAACC | | | | |
| | NM_024923-Ex40b-F | TGCCGTCTCTTCACACAGAG | 599 | TD54 | yes (0.75) | 1 5 |
| | NM_024923-Ex40b-R | GCCTAAGCCAAGTCCATCAG | | | | |
| | NM_024923-Ex40c-F | GAGGCTGGAGAACACAGGAG | 538 | TD54 | no | 1.5 |
| | NM_024923-Ex40c-R | CAGAATCAGGAGCTGGGAAG | | | | |
| | NM_024923-Ex40d-F | CACTTTCATGGCAGCTCATC | 567 | TD54 | no | 1 5 |
| | NM_024923-Ex40d-R | ccaggccccagagtatcat | | | | |

| HDAC11 (NM_024827) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_024827-Ex1-F | ccagaggacgcggctaag | 240 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex1-R | acccctcaccagcaaatct | | | | |
| 2 | NM_024827-Ex2-F | aaaggcagccagtttaagca | 426 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex2-R | cgagggctgaaactgaagat | | | | |
| 3 | NM_024827-Ex3-F | aagccagcagtgcctaccta | 386 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex3-R | ccacagactcccatgttcct | | | | |
| 4 | NM_024827-Ex4-F | gtgtgcttgaccagggagtt | 400 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex4-R | ggagaaagggtcattgttgc | | | | |
| 5 | NM_024827-Ex5-F | ccagagctggctagaagcag | 284 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex5-R | ggtgacttgggaaaaccaga | | | | |
| 6 | NM_024827-Ex6-F | gaggcacaggttcctgagag | 308 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex6-R | cagcataggccccttgtcta | | | | |
| 7 | NM_024827-Ex7-F | gattacccacaagcattaggc | 298 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex7-R | aagacagaagccgtgagagc | | | | |
| 8 | NM_024827-Ex8-F | caaatggggagtttcctgag | 320 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex8-R | aggatcacctcacagggatg | | | | |
| 9 | NM_024827-Ex9-F | gctgagcagtgctgaatctg | 393 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex9-R | cagtccatctcctccctgac | | | | |
| 10 | NM_024827-Ex10a-F | agcaggacttcctgacacca | 584 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex10a-R | GGACCTAGCCTGTCCTCTCC | | | | |
| | NM_024827-Ex10b-F | TTCTAACCTCATGGGGTGGT | 683 | TD54 | yes (0 75) | 1 5 |
| | NM_024827-Ex10b-R | ttgcactgaacaggcaagac | | | | |

| Fibulin2 (NM_001004019) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | Exon1-NM_001004019-F | gtcagcgagtctgggctct | 463 | TD50 | yes (0 75) | 1 5 |
| | Exon1-NM_001004019-R | GTGCCCGCATCTGTACACT | | | | |
| 2 | Exon2a-NM_001004019-F | tgaggcagtgcttagtgtgc | 577 | TD50 | yes (0 75) | 2 |
| | Exon2a-NM_001004019-R | AGCTGTCAGCCACCACTACA | | | | |
| | Exon2b-NM_001004019-F | GCCGGTCAGTCCTATTTTGT | 575 | TD50 | yes (0 75) | 2 |
| | Exon2b-NM_001004019-R | TCCTCGGTCACTCTCCTAGC | | | | |
| | Ex2C#2-NM_001004019-F | GAGGTGGGAGTCAGCCACT | 484 | TD54 | yes (0 75) | 1 75 |
| | Ex2C#2-NM_001004019-R | GCAGTGATGTGGACAGGATG | | | | |
| | Ex2D-NM_001004019-F | cTGGACCCAGTCTTCCTATCC | 538 | TD54 | yes (0 75) | 1 5 |
| | Ex2D-NM_001004019-R | atgcactctcacagcctcct | | | | |
| 3 | Exon3-NM_001004019-F | tgcccagtaaaggagacctg | 264 | TD54 | yes (0 75) | 2 |
| | Exon3-NM_001004019-R | aagcagggttctggagaagg | | | | |
| 4 | Exon4-NM_001004019-F | agagcttccctcctctgagc | 324 | TD54 | yes (0 75) | 1 5 |
| | Exon4-NM_001004019-R | ccccaaggacagcctaactt | | | | |
| 5 | Exon5-NM_001004019-F | tttatcagtccctccccctga | 358 | TD50 | yes (0 75) | 2 5 |
| | Exon5-NM_001004019-R | tctcaagccacaggtgatga | | | | |
| 6 | Exon6-NM_001004019-F | cccaaagctggttgtctcag | 399 | TD50 | yes (0 75) | 2 |
| | Exon6-NM_001004019-R | ccaacttgtggggtaacagg | | | | |
| 7 | Exon7-NM_001004019-F | tcagtgccctgtgtgagaag | 300 | TD50 | yes (0 75) | 2 |
| | Exon7-NM_001004019-R | ccagacagctcaccacctg | | | | |
| 8 | Exon8-NM_001004019-F | taggttagagccagggatgg | 296 | TD54 | yes (0 75) | 2 |
| | Exon8-NM_001004019-R | tcctgcacaaactcagcaag | | | | |
| 9 | Exon9-NM_001004019-F | tctcctgtcactggatgctc | 294 | 50 | yes (0 75) | 1 5 |
| | Exon9-NM_001004019-R | cacagtgtggtgggcttg | | | | |
| 10 | Exon10-NM_001004019-F | agctctctccctcctggtg | 397 | TD56 | no | 1 5 |
| | Exon10-NM_001004019-R | ccagcaggaaacccaactaa | | | | |
| 11 | Ex11#2-NM_001004019-F | agcttcccagggagatcttt | 381 | TD54 | yes (0 75) | 1 5 |
| | Ex11#2-NM_001004019-R | gagaggaatggagcagcaag | | | | |
| 12 | Exon12&13-NM_001004019-F | ctggccctgcaactctgac | 590 | TD50 | yes (0 75) | 2 |
| | Exon12&13-NM_001004019-R | gaggtggagtgcacagagg | | | | |
| 13 | Exon12&13-NM_001004019-F | ctggccctgcaactctgac | 590 | TD50 | yes (0 75) | 2 |
| | Exon12&13-NM_001004019-R | gaggtggagtgcacagagg | | | | |
| 14 | Ex14#2-NM_001004019-F | gcctctccctgtctttcctt | 374 | TD54 | yes (0 75) | 2 5 |
| | Ex14#2-NM_001004019-R | aaggtgtgagggtgaaggtg | | | | |
| 15 | Exon15-NM_001004019-F | tgtcagacagcatttggatg | 299 | TD56 | no | 1 5 |
| | Exon15-NM_001004019-R | tgctaaaacctttctgtgtcca | | | | |
| 16 | Exon16-NM_001004019-F | gtgacagggacaggctgact | 284 | TD50 | yes (0 75) | 2 |
| | Exon16-NM_001004019-R | gagcttttctgtgcctgctt | | | | |
| 17 | Ex17#2-NM_001004019-F | ggcatgacatgaggtgattg | 482 | TD54 | yes (0 75) | 1 5 |
| | Ex17#2-NM_001004019-R | tttggtcacctagcccacag | | | | |
| 18 | Ex18a#2-NM_001004019-F | ggccattagggagagctg | 691 | TD54 | yes (0 75) | 1 5 |
| | Ex18a#2-NM_001004019-R | TGCAAGCTTCCACTTGGTG | | | | |
| | Exon18b-NM_001004019-F | ACGTGGAGATGAAGCTCTGG | 647 | TD56 | yes (0 75) | 4 5 |
| | Exon18b-NM_001004019-R | gccatagagccctgtgacat | | | | |
| EST1 | BC111426-F | caaaatcctctcccaaatgc | 787 | TD56 | yes (0 75) | 1 5 |
| | BC111426-R | ggcatgaagcacattcacac | | | | |
| EST2 | W70042_FBLN-est2-F | ATATTTCCTGCCGAGCTGAG | 599 | TD54 | yes (0 75) | 1 5 |
| | W70042_FBLN-est2-R | GCCAGGTCTTCCTCACACAT | | | | |
| EST4 | DA856838-FBLN-Est4-F | GGGAGGGAGTGGAGACAGAT | 815 | TD54 | yes (0 75) | 1 5 |
| | DA856838-FBLN-Est4-F | ctgtaagaccctgggggaat | | | | |
| EST5 | DB458158-FBLN-est5-ex1-F | ctggagggagtgttctccac | 247 | TD54 | yes (0 75) | 1 5 |
| | DB458158-FBLN-est5-ex1-R | tggtgctgtctgatggttgt | | | | |
| EST6 | BF368851-FBLN2-Est6-F | ccagccgagactgttctgat | 342 | TD54 | yes (0 75) | 1 5 |
| | BF368851-FBLN2-Est6-R | ctggaccccagctttcact | | | | |

| WNT7A (NM_004625) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_004625-Ex1a#2 | gccctgcaatctgcaagtta | 491 | TD57 (40 cycles) | yes (1 5) | 1 5 |
| | NM_004625-Ex1a#2 | gcggggcaatcaacatag | | | | |
| | NM_004625-Ex1b | cgtctcgcacacttgcac | 527 | TD54 | yes (0 75) | 1 5 |
| | NM_004625-Ex1b | tccccctgcctatatctcct | | | | |
| 2 | Exon2-NM_004625-F | cagtgctggacacctacgaa | 384 | TD56 | yes (0 75) | 2 |
| | Exon2-NM_004625-R | agaggggcacctgagggtatt | | | | |
| 3 | Exon3-NM_004625-F | cgggtactcagtcgagcttt | 498 | TD56 | yes (0 75) | 2 |
| | Exon3-NM_004625-R | ctgtggctcctcaacagaca | | | | |
| | NM_004625-Ex4a#1-F | aagccagacccagataggtg | 387 | TD54 | yes (0 75) | 1 5 |
| | NM_004625-Ex4a#1-R | ggtcctcctcgcagtagttg | | | | |
| | NM_004625-Ex4a#2-F | gaccacactgccacagtttc | 455 | TD54 | yes (0 75) | 1 5 |
| | NM_004625-Ex4a#2-R | tcctcccagcaatctgactt | | | | |
| | Exon4b-NM_004625-F | gcggctgtgacctcatgt | 593 | TD56 | yes (0 75) | 2 |
| | Exon4b-NM_004625-R | tctacacggctccttgtcct | | | | |
| EST1 | AA405144-Exon1-F | cctctgtgagcctcaattttc | 399 | TD56 | yes (0 75) | 3 |
| | AA405144-Exon1-R | tgttttgtttactggatattcttga | | | | |
| | AA405144-Exon2-F | ccatcccagcatcttcctt | 246 | TD54 | no | 1 5 |
| | AA405144-Exon2-R | ctggcctagaggatgtgtgc | | | | |
| EST2 | AA412464-Exon2-F | ctgggcttctgcctctgtag | 399 | TD54 | no | 1 5 |
| | AA412464-Exon2-R | tgaactgaggcttgcaaaaa | | | | |
| EST3 | AA826310WNT-est3-F | cccttcatgtttgagggtgt | 686 | TD54 | yes (0 75) | 2 5 |
| | AA826310WNT-est3-R | tggatgcactgacttcaagc | | | | |

| TPRXL (AK092426) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | AK092426-Ex1-F | ctggtgtgttcaggaagctg | 279 | TD54 | yes (0 75) | 1 5 |
| | AK092426-Ex1-R | gtaaagccactcaccccaag | | | | |
| 2 | AK092426-Ex2-F | ttcctcagcctttgtttttg | 497 | TD56 | yes (0 75) | 2 5 |
| | AK092426-Ex2-R | cccgccagacatacatatcc | | | | |
| 3 | AK092426-Ex3-ext-F | ggcatcttttccctttgtaatg | 2173 | TD54 - ext 1 00 | yes (0 75) | 1 5 |
| | AK092426-Ex3-ext-R | gaagctggaaggcacagttc | | | | |
| | AK092426-Ex3-int-F | gtgcagagccaggagggcgtttg | only for sequencing | | | |
| | AK092426-Ex3-int-R | cctaggattgagcctggacctg | | | | |

| CHCHD4 (NM_144636) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | BC033775-Ex1-F | gaaatgtagtttccggctgag | 397 | TD54 | yes (0 75) | 1 5 |
| | BC033775-Ex1-R | ATCCGCAAAACAGGAGTAGC | | | | |
| 2 | NM_144636-Ex2-F | tcctgaggcttcagacatcc | 380 | TD54 | yes (0 75) | 1 5 |
| | NM_144636-Ex2-R | caactcccccatacagaatca | | | | |
| 3 | BC033775-Ex2-F | ttcaaacttggtttgtgtgg | 400 | TD54 | yes (0 75) | 1 5 |
| | BC033775-Ex2-R | ccagggaagtacatcagaacaa | | | | |
| 4 | BC033775-Ex3-F | gctaaatgggtcccggtatt | 1595 | TD58 | yes (0 75) | 1 5 |
| | BC033775-Ex3-R | ctgtgtgaaggcatctgctc | | | | |
| | BC033775-Ex3-internal-F | ggccttgtcatcaccttccaag | only for sequencing | | | |
| | BC033775-Ex3-internal-R | tcagggcaatctgagaattc | | | | |

| TMEM43 (NM_024334) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_024334-Ex1-F | caatgtcccggaccgtatag | 320 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex1-R | ggcgaaatggacctagagga | | | | |
| 2 | NM_024334-Ex2-F | agttttcattctgttactgtttctttt | 282 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex2-R | ggcccttgattaccaaatcc | | | | |
| 3 | NM_024334-Ex3-F | aactgtacggtggggagatg | 375 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex3-R | atcactcccatgtgtgacca | | | | |
| 4 | NM_024334-Ex4-F | aagaacctgggacagggagt | 472 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex4-R | ctcctggagccactcttcac | | | | |
| 5 | NM_024334-Ex5&6-F | tgatctggtagccctgaggt | 596 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex5&6-R | cacgaggcaggattaactcaa | | | | |
| 6 | NM_024334-Ex5&6-F | tgatctggtagccctgaggt | 596 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex5&6-R | cacgaggcaggattaactcaa | | | | |
| 7 | NM_024334-Ex7-F | cctgggctaatctggacttg | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex7-R | ctgatcctgtgcctttagcc | | | | |
| 8 | NM_024334-Ex8&9-F | cgtggacgagacagagtcag | 700 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex8&9-R | cgctcctgacattgaccaag | | | | |
| 9 | NM_024334-Ex8&9-F | cgtggacgagacagagtcag | 700 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex8&9-R | cgctcctgacattgaccaag | | | | |
| 10 | NM_024334-Ex10-F | gggtttctgtgctcacttcc | 330 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex10-R | tgcctcattcactggctatg | | | | |
| 11 | NM_024334-Ex11-F | tgttcagaaatggccaacag | 394 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex11-R | ctcatcccaaggctatggag | | | | |
| 12 | NM_024334-Ex12a-F | cccatcctcatctagggaca | 559 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex12a-R | GGAAACAGCAGGAGAAGCTG | | | | |
| | NM_024334-Ex12b-F | TGGTGTTCACCAGCTCATGT | 581 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex12b-R | TTCCTCTTGGGGTAGGAAAG | | | | |
| | NM_024334-Ex12c-F | TCCTGAGGAGAAAAGCTGGA | 581 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex12c-R | CGTGGGCATTGTACAACCAG | | | | |
| | NM_024334-Ex12d-F | CGATTAAGAGAAAAGGTTGGAA | 592 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex12d-R | GAGATTTGATGAAATTGCTCATGTA | | | | |
| | NM_024334-Ex12e-F | TTGTGCCTGCTGGGAGTAAT | 561 | TD54 | yes (0 75) | 1 5 |
| | NM_024334-Ex12e-R | atcctatggctgaattctttaca | | | | |

| XPC (NM_004628) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_004628-Ex1-F | gtgactaggcctccaacgaa | 457 | TD54 | no | 1 5 |
| | NM_004628-Ex1-R | agctacgcaggagcttggat | | | | |
| 2 | NM_004628-Ex2-F | ggggtttggagacaggtcat | 399 | TD54 | no | 1 5 |
| | NM_004628-Ex2-R | tggaccccagtgacaagtaa | | | | |
| 3 | NM_004628-Ex3-F | tgaggctcagagcttactaatca | 398 | TD54 | no | 1 5 |
| | NM_004628-Ex3-R | tccctagtacaaccttatctgtgg | | | | |
| 4 | NM_004628-Ex4-F | ttcccagcagaaccttgatt | 300 | TD54 | no | 1 5 |
| | NM_004628-Ex4-R | tcccctacaagtttctccaaa | | | | |
| 5 | NM_004628-Ex5-F | gccttgtgtagggaaacagg | 297 | TD54 | no | 1 5 |
| | NM_004628-Ex5-R | agagcagcaaagccagaaat | | | | |
| 6 | NM_004628-Ex6-F | tttcttatatgtagaaatggcaacac | 390 | TD54 | no | 1 5 |
| | NM_004628-Ex6-R | tagttaccgcctcagggaag | | | | |
| 7 | NM_004628-Ex7-F | aaagtctgagctgggtctgc | 391 | TD54 | no | 1 5 |
| | NM_004628-Ex7-R | tgtcggtaacacacctggaa | | | | |
| 8 | NM_004628-Ex8-F | cctgtcttgaacaagcaccat | 296 | TD54 | no | 1 5 |
| | NM_004628-Ex8-R | acccacactccgtgaatacc | | | | |
| 9 | NM_004628-Ex9a-F | ggggacatcttgatgtattgg | 595 | TD54 | no | 1 5 |
| | NM_004628-Ex9a-R | GGGTCCTGGAGGCACTCT | | | | |
| | NM_004628-Ex9b-F | CCTCTGATGAGGATTCCGAAC | 595 | TD54 | no | 1 5 |
| | NM_004628-Ex9b-R | gctgggcatatataaggtgctc | | | | |
| 10 | NM_004628-Ex10-F | gctccaccatctgttgtcag | 375 | TD54 | no | 1 5 |
| | NM_004628-Ex10-R | tgctgtccagtcagatgagc | | | | |
| 11 | NM_004628-Ex11-F | gcctagcacagcttctctgg | 397 | TD54 | no | 1 5 |
| | NM_004628-Ex11-R | cctccttgaatcctgctcaa | | | | |
| 12 | NM_004628-Ex12&13-F | ttctgagggttcaccaggta | 600 | TD54 | no | 1 5 |
| | NM_004628-Ex12&13-R | aggcctcaactcccagcag | | | | |
| 13 | NM_004628-Ex12&13-F | ttctgagggttcaccaggta | 600 | TD54 | no | 1 5 |
| | NM_004628-Ex12&13-R | aggcctcaactcccagcag | | | | |
| 14 | NM_004628-Ex14-F | tcaggacccctagtgctcag | 368 | TD54 | no | 1 5 |
| | NM_004628-Ex14-R | agcctgctgtattcagtgctc | | | | |
| 15 | NM_004628-Ex15-F | tctgtctttacattcacagtttcca | 294 | TD54 | no | 1 5 |
| | NM_004628-Ex15-R | cacaaagctatccctgacttga | | | | |
| 16 | NM_004628-Ex16a-F | ggaacttgctgcctcttcat | 683 | TD54 | yes (0 75) | 1 5 |
| | NM_004628-Ex16a-R | ATGCACCACCATCCAGAAAT | | | | |
| | NM_004628-Ex16b-F | CAGCCCTTGTCAGATTCACC | 699 | TD54 | yes (0 75) | 1 5 |
| | NM_004628-Ex16b-R | ccctcaggaagaccactcaa | | | | |

| LSM3 (NM_014463) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | Exon1-NM_014463-F | ggaagcaggaaagagcacaa | 221 | TD54 | no | 1 5 |
| | Exon1-NM_014463-R | ggccatttttcctgagtctg | | | | |
| 2 | Exon2-NM_014463-F | cccatggtcacacagcatag | 280 | TD54 | no | 1 5 |
| | Exon2-NM_014463-R | acccttttcactctggacagg | | | | |
| 3 | Exon3-NM_014463-F | aaatgccctagttgtaattgagc | 292 | TD54 | no | 1 5 |
| | Exon3-NM_014463-R | tcaaaacatgacaggcttcc | | | | |
| 4 | Exon4-NM_014463-F | tttgtgtcatgatgaatccattt | 490 | TD56 | yes (0 75) | 2 |
| | Exon4-NM_014463-R | caaggaaacaccaatttaaagaaaa | | | | |

| SLC6A6 (NM_003043) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_003043-Ex1 | cgtgtttgcgtgtatgtgtg | 589 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex1 | GCCGGGGAAACAATAACAG | | | | |
| 2 | NM_003043-Ex2 | gctgggagtctgcatatgatt | 480 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex2 | accaagcaggcacagagtg | | | | |
| 3 | NM_003043-Ex3 | caaagcattggacagacagg | 492 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex3 | ggaaagcagggaaagagagg | | | | |
| 4 | NM_003043-Ex4 | gtccaagagtttggccttga | 351 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex4 | cctggcagacaataggtgct | | | | |
| 5 | NM_003043-Ex5-BC111489 | ataccatggccttccagttg | 793 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex5-BC111489 | caagtctggggatggttgtt | | | | |
| 6 | NM_003043-Ex6 | ctgttaatgtccgagccttg | 295 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex6 | ctaattgcaggcgggaagt | | | | |
| 7 | NM_003043-Ex7 | ctcccctctgcctctctga | 298 | TD54 | no | 1 5 |
| | NM_003043-Ex7 | cctctacatcctgccctgtc | | | | |
| 8 | NM_003043-Ex8&9 | TTGAAAGGCACCTTGGATTC | 591 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex8&9 | TTCATGCCTGCAGTACTTGG | | | | |
| 9 | NM_003043-Ex8&9 | TTGAAAGGCACCTTGGATTC | 591 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex8&9 | TTCATGCCTGCAGTACTTGG | | | | |
| 10 | NM_003043-Ex10 | ccctgctgtatttgggtctg | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex10 | cccaggtggaataaggagtct | | | | |
| 11 | NM_003043-Ex11 | atctcacaggccaattctgg | 385 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex11 | ggagatgaagcccctacctg | | | | |
| 12 | NM_003043-Ex12 | tgccagagctctttgtatatga | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex12 | agggagggcctactgttcaa | | | | |
| 13 | NM_003043-Ex13 | gtcccagagaccccactcat | 294 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex13 | ttataagcccacgctgaaca | | | | |
| 14 | NM_003043-Ex14 | gagagacccttccctcaggt | 399 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex14 | ctggatgggtgagggaacta | | | | |
| 15 | NM_003043-Ex15a | acagcacatgagccaattca | 839 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex15a | ATCTCCTCACAGCCCTCCTT | | | | |
| | NM_003043-Ex15b | CCACTTGAATTGATCTTCTTGC | 761 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex15b | TCCTTTGGTCCACTCACCTC | | | | |
| | NM_003043-Ex15c | CATTCCAGGCAGAGAAGGAG | 810 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex15c | TGTGAAAATTCTGCGGTCTG | | | | |
| | NM_003043-Ex15d | GATCAAGGGCCTTATGTGGA | 850 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex15d | GAGGGCTCTGATTGGAGACA | | | | |
| | NM_003043-Ex15e | GCCACAGTATTTTGGGTTGG | 818 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex15e | GGTCCAGGAATTCTGTGAGG | | | | |
| | NM_003043-Ex15f | GCAGGGAATGGTGAGTGTCT | 843 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex15f | CGATTTTTGCACACAGTGGT | | | | |
| | NM_003043-Ex15g | TGCTCTGGGTAGCCAGTCTAA | 761 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex15g | CCTGCAAAATACAGGGAATCA | | | | |
| | NM_003043-Ex15h | GTCTTGCCTCCTTCCAGAAA | 456 | TD54 | yes (0 75) | 1 5 |
| | NM_003043-Ex15h | gaggagcgtcctgggatt | | | | |
| mRNA1 | BC038790-Ex1a | AGACTCATGCACAGCCTCCT | 581 | TD54 | yes (0 75) | 1 5 |
| | BC038790-Ex1a | GCTCAGACTTCAGAGCTGACAA | | | | |
| mRNA1 | BC038790-Ex1b | GTCAAAGCACCTGGGAAATG | 816 | TD54 | yes (0 75) | 1 5 |
| | BC038790-Ex1b | TCACCTTGGGGGATAAAATG | | | | |
| mRNA1 | BC038790-Ex1c | CTTCGTGACCTGGAGGAGTC | 828 | TD54 | yes (0 75) | 1 5 |
| | BC038790-Ex1c | CAACAAGGTCAGCAAACTGG | | | | |
| mRNA2 | AK023516-Ex1a | CTAAGCAGGCCCTGAAAGC | 700 | TD54 | yes (0 75) | 1 5 |
| | AK023516-Ex1a | GCTTCAGTTCTCATGGATGG | | | | |
| mRNA2 | AK023516-Ex1b | CCACTGTAAACCACCGTCCT | 846 | TD54 | yes (0 75) | 1 5 |
| | AK023516-Ex1b | GCCTTCGTGGAACTGACATT | | | | |
| mRNA2 | AK023516-Ex1c | GGGATTTGAACCCTTGCTTA | 750 | TD54 | yes (0 75) | 1 5 |
| | AK023516-Ex1c | GAGTTCTGGAGCCAGGTTTG | | | | |

| GRIP2 (NM_001080423) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | NM_001080423-Ex1-F | aaggggcagatctaactccaa | 366 | TD54 | no | 1 5 |
| | NM_001080423-Ex1-R | agcagtgatgccaaggtctc | | | | |
| 2 | NM_001080423-Ex2-F | gagccagccttgacaatagc | 327 | TD54 | no | 1 5 |
| | NM_001080423-Ex2-R | caggtctcagccatccagtc | | | | |
| 3 | NM_001080423-Ex3-F | atcttacaggccacgtccac | 374 | TD54 | no | 1 5 |
| | NM_001080423-Ex3-R | gtagggcactctgctgcatt | | | | |
| 4 | NM_001080423-Ex4-F | gccactgacttgctctgtga | 471 | TD54 | no | 1 5 |
| | NM_001080423-Ex4-R | acaccaagttgatgggtgct | | | | |
| 5 | NM_001080423-Ex5-F | tgaccctggacaactgtctg | 477 | TD54 | yes (0 75) | 1 5 |
| | NM_001080423-Ex5-R | tactgcggttgtaggcacag | | | | |
| 6 | NM_001080423-Ex6-F | ctgatcacccagtcctcctg | 299 | TD54 | no | 1 5 |
| | NM_001080423-Ex6-R | ttcaaagtcagaactgagatacaca | | | | |
| 7 | NM_001080423-Ex7-F | ggacaaggtcctcaaacagg | 300 | TD54 | no | 1 5 |
| | NM_001080423-Ex7-R | gcctcgatctcttcatctgg | | | | |
| 8 | NM_001080423-Ex8-F | cattccctgcttcttcatcc | 375 | TD54 | no | 1 5 |
| | NM_001080423-Ex8-R | gactggcccaaggtcataag | | | | |
| 9 | NM_001080423-Ex9&10-F | aaatgggtataacagtttcttcca | 700 | TD54 | no | 1 5 |
| | NM_001080423-Ex9&10-R | tgctgctcttaataatttgagctt | | | | |
| 10 | NM_001080423-Ex9&10-F | aaatgggtataacagtttcttcca | 700 | TD54 | no | 1 5 |
| | NM_001080423-Ex9&10-R | tgctgctcttaataatttgagctt | | | | |
| 11 | NM_001080423-Ex11-F | cgggacatgtgagcagaata | 468 | TD54 | no | 1 5 |
| | NM_001080423-Ex11-R | actcactgcagcctcaacct | | | | |
| 12 | NM_001080423-Ex12-F | aactgatgtgaaggggaag | 383 | TD54 | yes (0 75) | 1 5 |
| | NM_001080423-Ex12-R | tggagctctgcccaaaatac | | | | |
| 13 | NM_001080423-Ex13-F | ggaatcaacgatcccattgt | 466 | TD54 | yes (0 75) | 1 5 |
| | NM_001080423-Ex13-R | cttgtgaagctagggcttgg | | | | |
| 14 | NM_001080423-Ex14-F | gtgtgacccaggttcagtcc | 400 | TD54 | no | 1 5 |
| | NM_001080423-Ex14-R | aaggcttcccttcccactat | | | | |
| 15 | NM_001080423-Ex15-F | gcaggggatgagattagaa | 299 | TD54 | no | 1 5 |
| | NM_001080423-Ex15-R | atggaaatgctggtctgagc | | | | |
| 16 | NM_001080423-Ex16&17-F | agtgggaccaggatgtgagt | 584 | TD54 | no | 1 5 |
| | NM_001080423-Ex16&17-R | cccctccttcatttgttcat | | | | |
| 17 | NM_001080423-Ex16&17-F | agtgggaccaggatgtgagt | 584 | TD54 | no | 1 5 |
| | NM_001080423-Ex16&17-R | cccctccttcatttgttcat | | | | |
| 18 | NM_001080423-Ex18-F | gaaatcagagcgtgagtgactg | 399 | TD54 | no | 1 5 |
| | NM_001080423-Ex18-R | ccaatgtcacccacagtgtc | | | | |
| 19 | NM_001080423-Ex19-F | gtttttggcagctgggttta | 380 | TD54 | no | 1 5 |
| | NM_001080423-Ex19-R | ggctgcagtgaggactctgt | | | | |
| 20 | NM_001080423-Ex20-F | gagtttccagaccaggctga | 390 | TD54 | yes (0 75) | 1 5 |
| | NM_001080423-Ex20-R | taaaggatgccaggagagga | | | | |
| 21 | NM_001080423-Ex21-F | gtagatgcaggcccagagag | 450 | TD54 | yes (0 75) | 1 5 |
| | NM_001080423-Ex21-R | ggctctgctcagtgcttctc | | | | |
| 22* | NM_001080423-Ex22-F | cagaactgccaccgaggac | 391 | TD54 | yes (0 75) | 0* |
| | NM_001080423-Ex22-R | ccaggtgagtgcatacgtaca | | | | |
| 23 | NM_001080423-Ex23-F | ggttcatgacttcccactcc | 394 | TD54 | yes (0 75) | 1 5 |
| | NM_001080423-Ex23-R | ctctggagtcccacgacagt | | | | |
| 24 | NM_001080423-Ex24-F | taagggctccgtcgtacttg | 469 | TD54 | no | 1 5 |
| | NM_001080423-Ex24-R | ttcacaggcgatagatgcag | | | | |
| 25 | NM_001080423-Ex25-F | acatcacatgcctcttgctc | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_001080423-Ex25-R | gtgtctgggagccaggatct | | | | |

*GRIP2 Ex22 –

(1) Used platinum Taq DNA polymerase high fidelity from Invitrogen instead of Taq DNA Polymerase Recombinant.

(2) Used $MgSO_4$ (1.5mM) instead of $MgCl_2$

| C3orf19 (NM_016474) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | NM_016474-Ex1-F | tacgcaccacactctgacct | 299 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex1-R | gaaacaagctaggtcgacattc | | | | |
| 2 | NM_016474-Ex2-F | ttgtccatttagcggaacttg | 360 | TD56 | yes (0 75) | 2 5 |
| | NM_016474-Ex2-R | catttcatgcagatgttaaatacaa | | | | |
| 3 | NM_016474-Ex3+BF698254-Ex3-F | aagaactcatgatctcctttgttt | 483 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex3+BF698254-Ex3-R | cacaccatgggaaaacctct | | | | |
| 4 | NM_016474-Ex4-F | ctgtcaggcagggagtgtg | 248 | TD54 | yes (0 75) | 2 |
| | NM_016474-Ex4-R | tgacgtctaaattttctctctactcct | | | | |
| 5 | NM_016474-Ex5+CN278581-Ex5-F | caccaggcaatgaatgtgac | 552 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex5+CN278581-Ex5-R | caggcactgggtttccttc | | | | |
| 6 | NM_016474-Ex6+CD58017981-Ex6-F | ccccccttcactgttcaccta | 386 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex6+CD58017981-Ex6-R | cagtcaggttgttccccaag | | | | |
| 7 | NM_016474-Ex7-F | tgggtcagaagcctctgtagt | 396 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex7-R | caaagcaggatctttaatttcca | | | | |
| 8 | NM_016474-Ex8-F | aggaaatgtggcttgagtgt | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex8-R | ggcaactacctttatgcaattc | | | | |
| 9 | NM_016474-Ex9-F | agaggtggcctgaagatcaa | 389 | TD54 | no | 1 5 |
| | NM_016474-Ex9-R | tcctgccagtacctcaaagg | | | | |
| 10 | NM_016474-Ex10-F | ctgcctgggacatgttaggt | 387 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex10-R | cacaaggcagcacgaagtta | | | | |
| 11 | NM_016474-Ex11a-F | agttttgctgttgccattctt | 572 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex11a-R | aatccaacccgaagtctgtg | | | | |
| | NM_016474-Ex11b-F | cacgctgacttgggtttgta | 494 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex11b-R | tgtgattacttccgggtcct | | | | |
| | NM_016474-Ex11c-F | ggaccataggtcacgaggaa | 481 | TD54 | no | 1 5 |
| | NM_016474-Ex11c-R | ggcacactgctgactttctg | | | | |
| | NM_016474-Ex11d-F | attcaggagcctgacttgga | 474 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex11d-R | aagtctgagatctgccccctta | | | | |
| | NM_016474-Ex11e-F | agtgccccatcctcattaca | 466 | TD54 | yes (0 75) | 1 5 |
| | NM_016474-Ex11e-R | tcgtagctttgtggtcaatctt | | | | |

| C3orf20 (NM_032137) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | NM_032137-Ex1-F | tgtgacatgcagatgaggtag | 250 | TD54 | no | 1 5 |
| | NM_032137-Ex1-R | caggatgactgcagtagtttca | | | | |
| 2 | NM_032137-Ex2-F | gatggtgcatcttttgatcact | 379 | TD54 | no | 1 5 |
| | NM_032137-Ex2-R | atggttccaggaaagcttca | | | | |
| 3 | NM_032137-Ex3a-F | ctgatggtgtctctgggtagg | 487 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex3a-R | acaaaggtgggttccatgag | | | | |
| | NM_032137-Ex3b-F | catcagtgacccttcagtgc | 398 | TD54 | no | 1 5 |
| | NM_032137-Ex3b-R | gttcccagcatgtgcctct | | | | |
| 4 | NM_032137-Ex4-F | ttcctgggccatttacagag | 294 | TD54 | no | 1 5 |
| | NM_032137-Ex4-R | ctttccattgcattcctgct | | | | |
| 5 | NM_032137-Ex5-F | caatgcggcactttacagtatg | 380 | TD54 | no | 1 5 |
| | NM_032137-Ex5-R | gcagagccgctatgtatcct | | | | |
| 6 | NM_032137-Ex6-F | ccaggtgacactgatgcaaa | 297 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex6-R | tccactcacatatacacagggact | | | | |
| 7 | NM_032137-Ex7-F | ggcaagggctattcttagacg | 697 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex7-R | tgaactgaacgtgaggcttg | | | | |
| 8 | NM_032137-Ex8-F | gctaggtcatcgttgggttc | 497 | TD54 | no | 1 5 |
| | NM_032137-Ex8-R | ctgcagaaacagcctcagc | | | | |
| 9 | NM_032137-Ex9-F | agcatcagcaggtgttaagg | 400 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex9-R | tgggcttttaatgctaggc | | | | |
| 10 | NM_032137-Ex10-F | tcttctccatgaagcggaat | 378 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex10-R | taagagcaaaggccctgaga | | | | |
| 11 | NM_032137-Ex11-F | ctctgaaccggagcaagtct | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex11-R | tctgtcccttgagatgacctg | | | | |
| 12 | NM_032137-Ex12-F | gaggtgccggagaatcaat | 464 | TD54 | no | 1 5 |
| | NM_032137-Ex12-R | ttgggttcactatcctgtgc | | | | |
| 13 | NM_032137-Ex13-F | gcctttgaccagcaggag | 494 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex13-R | gcaggtgctcagccagtc | | | | |
| 14 | NM_032137-Ex14-F | tgtacagggctccaggactc | 400 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex14-R | gtgctgggtgactcttggtc | | | | |
| 15 | NM_032137-Ex15-F | ctgtggcgctgttctagtga | 395 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex15-R | gtggaggcttcccacctc | | | | |
| 16 | NM_032137-Ex16&17a-F | aggctcccagaacactgatg | 585 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex16&17a-R | atgccgaatgtgtgattcat | | | | |
| 17 | NM_032137-Ex16&17b-F | tgcacctgttaaaggccact | 592 | TD54 | yes (0 75) | 1 5 |
| | NM_032137-Ex16&17b-R | acgtgacctgggacaagtct | | | | |

| FGD5 (NM_152536) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | NM_152536-Ex1a#2F | ggagaccgacgaggattaca | 646 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex1a#2R | AACTTCAGGGCAGTCTGTGG | | | | |
| | NM_152536-Ex1bF | GATGCTGAGGACACCAGTGA | 673 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex1bR | CACCCTCTGCCTCACAGTTC | | | | |
| | NM_152536-Ex1cF | AGGAAGAACACCAGCACGAG | 670 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex1cR | TCCGGTCAACTTCCAGAATC | | | | |
| | NM_152536-Ex1dF | GCATGTGGATGTGAACGTGT | 700 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex1dR | gtcccatttcacggatgaag | | | | |
| 2 | NM_152536-Ex2F | atgaatgggacccctggtag | 298 | TD54 | no | 1 5 |
| | NM_152536-Ex2R | caccaagggaaggagcagat | | | | |
| 3 | NM_152536-Ex3&4F | gcttggaacttgcctccag | 400 | TD54 | no | 1 5 |
| | NM_152536-Ex3&4R | ccaggcctctgttaatggaa | | | | |
| 4 | NM_152536-Ex3&4F | gcttggaacttgcctccag | 400 | TD54 | no | 1 5 |
| | NM_152536-Ex3&4R | ccaggcctctgttaatggaa | | | | |
| 5 | NM_152536-Ex5&6F | ggtgcatgtggtcatctctg | 839 | TD54 | no | 1 5 |
| | NM_152536-Ex5&6R | gatttgcattaagcccttcg | | | | |
| 6 | NM_152536-Ex5&6F | ggtgcatgtggtcatctctg | 839 | TD54 | no | 1 5 |
| | NM_152536-Ex5&6R | gatttgcattaagcccttcg | | | | |
| 7 | NM_152536-Ex7F | gctgctgtgagcatggtg | 249 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex7R | agaggcaaggtcaccacagt | | | | |
| 8 | NM_152536-Ex8F | gatgcttcattcttccaactct | 250 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex8R | tcctgggccttggaattg | | | | |
| 9 | NM_152536-Ex9F | gaggcctgtgacctgacact | 233 | TD54 | no | 1 5 |
| | NM_152536-Ex9R | ctaggagtggggcatcagaa | | | | |
| 10 | NM_152536-Ex10F | gatcccagtcctgccttgta | 346 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex10R | tccatgggtagacagcacct | | | | |
| 11 | NM_152536-Ex11F | atgatccaagccctgctcta | 500 | TD54 | no | 1 5 |
| | NM_152536-Ex11R | tcattcaaataacaaatgagtggtg | | | | |
| 12 | NM_152536-Ex12F | gggatcctttgaggacagaa | 299 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex12R | cttgctgtagcccaacactg | | | | |
| 13 | NM_152536-Ex13F | gtggacatggtccctctgtt | 241 | TD54 | no | 1 5 |
| | NM_152536-Ex13R | ttctgaaccttcccatgctc | | | | |
| 14 | NM_152536-Ex14F | aaggggcatctgagtgagaa | 277 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex14R | gtgtgagtggcaacagcaat | | | | |
| 15 | NM_152536-Ex15F | aggagaggcctttcacatca | 396 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex15R | cctcactctggggcttacag | | | | |
| 16 | NM_152536-Ex16F | aaagtaggggacacgcacag | 365 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex16R | tgaaccatggaccacagaga | | | | |
| 17 | NM_152536-Ex17F | acagcccagtccctgctac | 461 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex17R | ctgactccgtctacccagga | | | | |
| 18 | NM_152536-Ex18F | gaccttatccagtgtcaaagca | 391 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex18R | aagacttgactgccaccagaa | | | | |
| 19 | NM_152536-Ex19F | ccacacagaggcactcttca | 396 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex19R | agttctgcagttctgctatgg | | | | |
| 20 | NM_152536-Ex20aF | tccaggccttttgctgaata | 817 | TD54 | yes (0 75) | 1 5 |
| | NM_152536-Ex20aR | ACAGCCAGGTCTGCACAAAT | | | | |
| | NM_152536-Ex20bF | GCTGCAACCTCCAATTTTGT | 839 | TD54 | yes (0 75) | 2 5 |
| | NM_152536-Ex20bR | tgcaagtaagtccagagatgg | | | | |

| NR2C2 (NM_003298) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | NM_003298-Ex1#2-F | gggtcacgaactctgaccttt | 586 | TD54 | yes (0 75) | 1 5 |
| | NM_003298-Ex1#2-F | ctccgtcgagaacaaaatgg | | | | |
| 2 | NM-003298-Ex2-F | gccacagtgaaagcaagagg | 550 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex2-R | cccaaaaggcacaactcttt | | | | |
| 3 | NM-003298-Ex3-F | caaagggatattgcttataaagtgc | 296 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex3-R | tggctttttgattcagcttg | | | | |
| 4 | NM-003298-Ex4-F | ggtttttggcagtcaccact | 375 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex4-F | gcctcgaaggcagatgtaag | | | | |
| 5 | NM-003298-Ex5-F | ctcaagtgatccacccacct | 298 | TD58 | yes (0 75) | 1 5 |
| | NM-003298-Ex5-F | ctaccaccccacctcacaac | | | | |
| 6 | NM-003298-Ex6-F | tgagacttgcacaggaactga | 400 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex6-F | caaaatgaagcctgagtgcat | | | | |
| 7 | NM-003298-Ex7-F | ccctgcagtaatggatttgg | 358 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex7-F | aacacaacgctggaggctat | | | | |
| 8 | NM-003298-Ex8-F | tgtggaaggacaggtaaatgc | 275 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex8-F | tctctgaaggccaggaagaa | | | | |
| 9 | NM-003298-Ex9-F | gctgtttttgaaccagcatt | 300 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex9-F | gctgttttccaccctcacaa | | | | |
| 10 | NM-003298-Ex10-F | gtgaaaccccatctctacgg | 386 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex10-F | aaaactgcagatcggctttg | | | | |
| 11 | NM-003298-Ex11-F | agcatgacagggatacatgttg | 299 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex11-F | caaactagacaggccctcaga | | | | |
| 12 | NM-003298-Ex12-F | tgggaaatgctgggacttag | 386 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex12-F | gctgtgggaaactgaggcta | | | | |
| 13 | NM-003298-Ex13&14a-F | aatggtcccatttctttcca | 684 | TD58 | yes (0 75) | 1 5 |
| | NM-003298-Ex13&14a-R | cctcttcatctcctgcgaac | | | | |
| 14 | NM-003298-Ex13&14a-F | aatggtcccatttctttcca | 684 | TD58 | yes (0 75) | 1 5 |
| | NM-003298-Ex13&14a-R | cctcttcatctcctgcgaac | | | | |
| | NM-003298-Ex13&14b-F | agactacaccatacaaaagtgtgc | 600 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex13&14b-R | tcaaagactgatctagtttgggta | | | | |
| | NM-003298-Ex13&14c-F | gccagatgtccagaaccaat | 497 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex13&14c-R | tgtgatccaatggaggctct | | | | |
| 15 | NM-003298-Ex15a-F | aggagcctttgctgagctgt | 596 | TD58 | yes (0 75) | 1 5 |
| | NM-003298-Ex15a-R | CACATACTATTCTCTCAGGTTCTTTCA | | | | |
| | NM-003298-Ex15b-F | GCCTTTTGATGAAGCAGCAG | 593 | TD58 | yes (0 75) | 1 5 |
| | NM-003298-Ex15b-R | GAACCCCAGGTTCTAGAGGAG | | | | |
| | NM-003298-Ex15c-F | TGTGTGCACATGTTGTGGAG | 685 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex15c-R | CTTGCCAACAGAGATGCTGA | | | | |
| | NM-003298-Ex15d-F | TGCAGCTCCATCATAACTGTG | 599 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex15d-R | ACCTGTGAGTCAGGGGAGTG | | | | |
| | NM-003298-Ex15e-F | GAACAGTGGCATGTGGAAGA | 673 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex15e-R | AAGGATCAAAACGACGGAAG | | | | |
| | NM-003298-Ex15f-F | CCCTGATGTGTTTCAGTTTCA | 697 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex15f-R | AAGCCCTGTGAAAGTTGGAG | | | | |
| | NM-003298-Ex15g-F | TGGTTTCTGAGATCCTCTTGG | 691 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex15g-R | AAGCCCAGAACAGCAAAAGA | | | | |
| | NM-003298-Ex15h-F | TGAAAGACCAAATTAGTTGAGCA | 700 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex15h-R | GATGAGACTTCCTATGACCTGGA | | | | |
| | NM-003298-Ex15i-F | AATCAGCAGTTCATGCCACA | 700 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex15i-R | TCACCCAAAGGAAAATGGAG | | | | |
| | NM-003298-Ex15j-F | GCATTGTGCGTGTACAGGTT | 693 | TD58 | yes (0 75) | 1 5 |
| | NM-003298-Ex15j-R | TTTAACTGGCGATAGAAAGATGG | | | | |
| | NM-003298-Ex15k-F | CGTCACGGTCAGAGATTCAG | 661 | TD54 | yes (0 75) | 1 5 |
| | NM-003298-Ex15k-R | CAGGGAAGCTGAAGGCTAAA | | | | |
| | NM-003298-Ex15l-F | GCTCATAGACCTGGGAAGCA | 657 | TD58 | yes (0 75) | 1 5 |
| | NM-003298-Ex15l-R | AGCAGCATTTGAGTCCATGA | | | | |
| | NM-003298-Ex15m-F | TGTTCAACATTTATTGATTGATAGACT | 828 | TD58 | yes (0 75) | 1 5 |
| | NM-003298-Ex15m-R | aacacaagagagcgctgcat | | | | |

| MRPS25 (NM_022497) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | Exon1-NM_022497-F | gcctacaagtcccagagtgc | 400 | TD54 | no | 1 5 |
| | Exon1-NM_022497-R | gactcgggacctcacgttac | | | | |
| 2 | Exon2-NM_022497-F | tgcagtacaggttgggataca | 467 | TD54 | no | 1 5 |
| | Exon2-NM_022497-R | aaaggtgccaggagtcacag | | | | |
| 3 | Exon3-NM_022497-F | tgcgctgtgagttctgtttc | 457 | TD54 | no | 1 5 |
| | Exon3-NM_022497-R | caagctgagattcccaggag | | | | |
| 4 | Exon4a-NM_022497-F | ttgtctcggctgagaaactg | 583 | TD54 | no | 1 5 |
| | Exon4a-NM_022497-R | tgtggccttaacctctcagg | | | | |
| | Exon4b-NM_022497-F | atatccctgctctgctggtc | 688 | TD54 | no | 1 5 |
| | Exon4b-NM_022497-R | aagtgacagtaacatcagtgtcctatg | | | | |
| | Exon4c-NM_022497-F | GGTGTCTATAGCTGGCTCTGC | 700 | TD54 | no | 1 5 |
| | Exon4c-NM_022497-R | CCTGGAAAGCTTTTAGGACTTG | | | | |
| | Exon4d-NM_022497-F | cacggacccctagaaactca | 663 | TD54 | no | 1 5 |
| | Exon4d-NM_022497-R | cactgcagatgaaagccaaa | | | | |
| | Exon4e-NM_022497-F | TCTTGACCTTGATACCACCTG | 597 | TD54 | no | 1 5 |
| | Exon4e-NM_022497-R | TGCTTGTGCAAGTAAAATTAAGAA | | | | |
| | Exon4f-NM_022497-F | ccctgtaggctacttgatcctg | 681 | TD54 | no | 1 5 |
| | Exon4f-NM_022497-R | ccctggcaggacagtcttta | | | | |
| | Exon4g-NM_022497-F | ggcctgtctgtaatggcatc | 599 | TD54 | no | 1 5 |
| | Exon4g-NM_022497-R | cagctttcttctccctgtgaa | | | | |
| | Exon4h-NM_022497-F | TTCAAGACTGAAGACATTGATTAGA | 693 | TD54 | no | 1 5 |
| | Exon4h-NM_022497-R | aaagtatctttgcggggagaa | | | | |

| ZFYVE20 (NM_022340) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | NM_022340-Ex1-F | ctcgacaggcgtctaccag | 236 | TD54 | no | 1 5 |
| | NM_022340-Ex1-R | cgtctccgttctggaggagt | | | | |
| 2 | NM_022340-Ex2-F | gtccaattcggtggttgtgt | 397 | TD54 | no | 1 5 |
| | NM_022340-Ex2-R | atgcactcaccctaacgac | | | | |
| 3 | NM_022340-Ex3-F | cgcccattagagataagtgctc | 437 | TD54 | no | 1 5 |
| | NM_022340-Ex3-R | tagaccaagcaagctgcaaa | | | | |
| 4 | NM_022340-Ex4-F | gcttgcttggtctaaggatttg | 561 | TD54 | no | 1 5 |
| | NM_022340-Ex4-R | gtggcaacaccaaagctgat | | | | |
| 5 | NM_022340-Ex5-F | tgcttgctcctgcattttac | 384 | TD54 | no | 1 5 |
| | NM_022340-Ex5-R | gcctcctcagccaactctta | | | | |
| 6 | NM_022340-Ex6-F | tgttgaccaacctgcttttg | 377 | TD54 | no | 1 5 |
| | NM_022340-Ex6-R | cccaggctagaactgtccac | | | | |
| 7 | NM_022340-Ex7&8-F | tttgtgattttgtggcgttg | 576 | TD54 | no | 1 5 |
| | NM_022340-Ex7&8-R | tgaatgaagattcccatgagc | | | | |
| 8 | NM_022340-Ex7&8-F | tttgtgattttgtggcgttg | 576 | TD54 | no | |
| | NM_022340-Ex7&8-R | tgaatgaagattcccatgagc | | | | |
| 9 | NM_022340-Ex9-F | cacagctggagcctgatctt | 469 | TD54 | no | 1 5 |
| | NM_022340-Ex9-R | cgaacaacttcccaaagcat | | | | |
| 10 | NM_022340-Ex10-F | acaattaatgtgttagaaaagtgaaca | 395 | TD54 | no | 1 5 |
| | NM_022340-Ex10-R | gatggtagcctgtaccccact | | | | |
| 11 | NM_022340-Ex11-F | gtctgaaatgccacctttagaga | 298 | TD54 | no | 1 5 |
| | NM_022340-Ex11-R | gagtgcaggtctgctttcg | | | | |
| 12 | NM_022340-Ex12-F | catgaaagcctggcaactc | 287 | TD56 | yes (0 75) | 2 5 |
| | NM_022340-Ex12-R | aaacaagtcattggcatgga | | | | |
| 13 | NM_022340-Ex13-F | gagaacagagtggctcagctt | 300 | TD54 | no | 1 5 |
| | NM_022340-Ex13-R | actacaaccgcacacaacct | | | | |
| 14 | NM_022340-Ex14a-F | gcacagggatagagggacaa | 492 | TD56 | yes (0 75) | 2 5 |
| | NM_022340-Ex14a-R | TCCCTTTCTCGTTCCAACTC | | | | |
| | NM_022340-Ex14b-F | GACGAGTATGACCAGCAGCA | 690 | TD54 | no | 1 5 |
| | NM_022340-Ex14b-R | GTCCATCTCAAAGGGGTTGA | | | | |
| | NM_022340-Ex14c-F | CAGACAGCCCAGCTCCTAAC | 687 | TD54 | no | 1 5 |
| | NM_022340-Ex14c-R | TAGCAAATGGACCCAAGAGG | | | | |
| | NM_022340-Ex14d-F | TCAGCATTCAGTTGCTGCTT | 566 | TD54 | no | 1 5 |
| | NM_022340-Ex14d-R | CGGCCACCACTCTTATTTTC | | | | |
| | NM_022340-Ex14e-F | AGGTGGGGTGATGGGAATA | 697 | TD54 | no | 1 5 |
| | NM_022340-Ex14e-R | GGCTGATCAAATAGCCTTGG | | | | |
| | NM_022340-Ex14f-F | GCCGGGTCTCTAATTTAGGAAT | 600 | TD54 | no | 1 5 |
| | NM_022340-Ex14f-R | AAATTAACAATTTCCCTTTCTGG | | | | |
| | NM_022340-Ex14g-F | CCGCCCTGATGAGATATTTT | 699 | TD54 | no | 1 5 |
| | NM_022340-Ex14g-R | AGAAATGTCATGGGCCTCAG | | | | |
| | NM_022340-Ex14h-F | AACCCTGCTGGTTGTTATGG | 599 | TD58 | yes (0 75) | 1 5 |
| | NM_022340-Ex14h-R | AGCGATTCTCGTGCCTCAG | | | | |
| | NM_022340-Ex14i-F | TCCTGTAATCCCAGCACTTTG | 697 | TD58 | yes (0 75) | 1 5 |
| | NM_022340-Ex14i-R | CATCTGATGCCTGAATCCCTA | | | | |
| | NM_022340-Ex14j-F | GGTTGACAGCAGGCAAACTT | 581 | TD54 | no | 1 5 |
| | NM_022340-Ex14j-R | ttgtaacacatgtgcacacagac | | | | |

| CAPN7 (NM_014296) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | NM_014296-Ex1-F | gccctcagctgctcaatc | 678 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex1-R | gacgagggaaaagcaaagc | | | | |
| 2 | NM_014296-Ex2-F | ttttgtcattcatggaaacca | 339 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex2-R | gtggaatgagcagcaacaga | | | | |
| 3 | NM_014296-Ex3-F | aaatgaagaatacagggttttgg | 367 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex3-R | tgttaaaattcttttattcaaggtg | | | | |
| 4 | NM_014296-Ex4-F | tgtttgttctcaatttcagcaag | 286 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex4-R | gaggtttgttactgcagcctttt | | | | |
| 5 | NM_014296-Ex5-F | tgcttatctttggaagtagctga | 463 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex5-R | tgaatacaattcagcaaacctaatg | | | | |
| 6 | NM_014296-Ex6-F | tggttttgtgtcttggtgtca | 498 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex6-R | ccacccctctattccaacaa | | | | |
| 7 | NM_014296-Ex7-F | agccagatttgaccagaagg | 396 | TD56 | yes (0 75) | 2 |
| | NM_014296-Ex7-R | tttccaattacacttctgtacctga | | | | |
| 8 | NM_014296-Ex8-F | tgttgaaattacgtgaatgaatatga | 242 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex8-R | tgaatgatatgccctctcca | | | | |
| 9 | NM_014296-Ex9-F | atatacctgaatgaaatgtgctgtat | 250 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex9-R | tgatttcactcaagattcaaagg | | | | |
| 10 | NM_014296-Ex10-F | cgtcagcatattgtgcgtta | 396 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex10-R | tcatcctgtttcaggggttg | | | | |
| 11 | NM_014296-Ex11-F | gctgctctggggatttaaga | 283 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex11-R | cttggcctcccatagtgct | | | | |
| 12 | NM_014296-Ex12-F | tcttcaggggtaggaacagc | 397 | TD56 | yes (0 75) | 2 |
| | NM_014296-Ex12-R | ttttcaacggtagagatcaagtt | | | | |
| 13 | NM_014296-Ex13&14-F | ggcagccatttacagttttg | 699 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex13&14-R | gtcaactgatccacccacct | | | | |
| 14 | NM_014296-Ex13&14-F | ggcagccatttacagttttg | 699 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex13&14-R | gtcaactgatccacccacct | | | | |
| 15 | NM_014296-Ex15-F | ggatagcttaattcacttttttgaaat | 300 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex15-R | cgtatgtctttcacaatgttctaact | | | | |
| 16 | NM_014296-Ex16-F | tttttgtggagagatttagagca | 262 | TD54 | no | 2 |
| | NM_014296-Ex16-R | aaatgtgtcatttttcaacttaatca | | | | |
| 17 | NM_014296-Ex17-F | cagttaaggagatgtccatcca | 332 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex17-R | tccagataataccacttacaaagca | | | | |
| 18 | NM_014296-Ex18a-AK093742-Ex1-F | tgctctgctttgtaagtggtatt | 647 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex18a-AK093742-Ex1-R | ttcccagctagatgagagcag | | | | |
| 18 | NM_014296-Ex18#2-F | tgacttgtgtatgccaactgatt | 280 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex18#2-R | attcctgtgaagcctttgct | | | | |
| 19 | NM_014296-Ex19-F | agaggggttggagacacaga | 400 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex19-R | gattgtcctttcctcactcca | | | | |
| 20 | NM_014296-Ex20&21a-F | aaacttagtatgacatctgcacttca | 681 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex20&21a-R | acaaaatggcccctaaaatg | | | | |
| 21 | NM_014296-Ex20&21a-F | aaacttagtatgacatctgcacttca | 681 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex20&21a-R | acaaaatggcccctaaaatg | | | | |
| | NM_014296-Ex21b-F | aaggacgcaaatcttcagga | 594 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex21b-R | tgaatgggagataacctcaaca | | | | |
| | NM_014296-Ex21c-F | caaatgaattgtgccaccataa | 581 | TD54 | yes (0 75) | 1 5 |
| | NM_014296-Ex21c-R | tctaacttaaactatgcaatatccgta | | | | |
| | NM_014296-Ex21d-F | cataaaaacccatcaggacag | 491 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex21d-R | tttgatggcagggattaggt | | | | |
| | NM_014296-Ex21e-F | ggtggacaaatttaacaagttttct | 487 | TD56 | yes (0 75) | 1 5 |
| | NM_014296-Ex21e-R | cttgtgaaccctgggcaaa | | | | |
| mRNA1 | BI464220-Ex2-F | cctcctgagtagctgggact | 299 | TD56 | yes (0 75) | 2 |
| | BI464220-Ex2-R | tggtgtttccattctcctca | | | | |

| SH3BP5 (NM_004844 and NM_001018009) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer Name | Primer Sequence | Amplicon Size | Anneal Temp | Betaine (final conc (M)) | $MgCl_2$ (final conc (mM)) |
| 1 | NM_001018009-Ex1-F | gactgacagttctgcattgct | 571 | TD54 | no | 1 5 |
| | NM_001018009-Ex1-R | GCACTTTTATAAACCTAAACTCTTCC | | | | |
| 1* | NM_004844-Ex1-F | gacctggatagctgggactg | 496 | TD54 | yes (0 75) | 1 5 |
| | NM_004844-Ex1-R | gtacgcgtagacaccgacct | | | | |
| 2 | NM_001018009-Ex2-F | gtcaaaagggacaccctcaa | 380 | TD54 | no | 1 5 |
| | NM_001018009-Ex2-R | gggtccagcaataacactcc | | | | |
| 3 | NM_001018009-Ex3-F | cggtgctgctctagacactg | 385 | TD54 | yes (0 75) | 1 5 |
| | NM_001018009-Ex3-R | tcgagcttctaaccttcagca | | | | |
| 4 | NM_001018009-Ex4-F | aggcagagatgagcttgtcc | 390 | TD54 | no | 1 5 |
| | NM_001018009-Ex4-R | agggtgttttcaggcaacat | | | | |
| 5 | NM_001018009-Ex5-F | tattccaagcctgggctgtt | 280 | TD54 | yes (0 75) | 1 5 |
| | NM_001018009-Ex5-R | ggtggtccttgagtgtgtga | | | | |
| 6 | NM_001018009-Ex6-F | agtggtagctgatgcctcttg | 223 | TD54 | no | 1 5 |
| | NM_001018009-Ex6-R | ggcccatgtgatactctgct | | | | |
| 7 | NM_001018009-Ex7-F | tgatgggcagaaaatgtacttg | 421 | TD54 | no | 1 5 |
| | NM_001018009-Ex7-R | tgcccaactctgagacctttt | | | | |
| 8 | NM_001018009-Ex8-F | caggcttgggaagtcacatt | 498 | TD54 | no | 1 5 |
| | NM_001018009-Ex8-R | ctggcagaggtatggaatgg | | | | |
| 9 | NM_001018009-Ex9a-F | gggctgagttagctgactggt | 596 | TD54 | no | 1 5 |
| | NM_001018009-Ex9a-R | TGTTCCACAAAGGCTGTGAA | | | | |
| | NM_001018009-Ex9b-F | AATGAAGTTTCAGTGACCTTGAG | 982 | TD54 | no | 1 5 |
| | NM_001018009-Ex9b-R | CAGTGGCAACCATTCTTCAC | | | | |
| | NM_001018009-Ex9c-F | GGGGAGAAGGTTCTACAGCA | 700 | TD54 | yes (0 75) | 1 5 |
| | NM_001018009-Ex9c-R | aagtagaattcaaaccttggtaaaca | | | | |

*SH3BP5 has two isoforms, each with a different exon 1.

# Appendix 4

## Mutation Surveyor

Mutation Surveyor is advanced software for analyzing sequencing data, which compares your sequences to a reference sequence and aids in mutation detection. The Mutation Surveyor manual can be used to understand the program. In addition, stated below are several tips that may aid in Mutation Surveyor use.

**How reference sequences were downloaded for this study:**

Go to the National Center for Biotechnology Information (NCBI) homepage (http://www.ncbi.nlm.nih.gov/). Under the search options go to 'CoreNucleotide', then type in the accession number of the gene of interest, and press 'Go'. Once the search results appear, click once on the 'Links' tab that is directly to the right of the desired accession number. A list of several options will appear, select 'Map Viewer'. Once in Map Viewer look on the right hand side of the page just below 'Maps and Options', 'Download/View Sequence/Evidence' should be an option, click on the 'Download/View Sequence/Evidence' link. Here change the "Strand" to plus or negative depending on the gene of interest, then click 'Change Region/Strand'. Also, change the 'Sequence Format' to GenBank. On the same page under the 'This chromosome region corresponds to the contig region(s)' section click 'Display'. That will display a report for the genomic area of interest (which contains your gene of interest). In the 'Features' section, three options should be checked (STS, tRNA and microRNA). The only option not checked should be 'SNP', check that and then press 'Refresh'. Finally go to the

internet Toolbar, click 'File' and then 'Save as...'. Save this report in a desired computer folder as a text file under encoding 'Unicode (UTF-8)'. For simplicity, the name of the file can be the gene accession number.

**Mutation Surveyor Basics:**

Analyzing sequences of a particular gene in Mutation Surveyor for the first time requires to first save the downloaded reference sequence text file (described above) as a GenBank '*.gbk' file. The most efficient way to do so is to open Mutation Surveyor, click on 'Tools', then 'GBK file editor...'. Once the GBK file editor is open, open the desired text file by either clicking on 'Open' at the bottom of the editor page or clicking 'File' in the toolbar, then 'Open'. Once the folder in which your text file is saved in is open, in order to see your text file you may have to change the File type to Text (*.txt) because the default setting is GenBank file (*.gbk). Once the file is open there should be four tabs across the top of the page, General & Reference, Features 1, Features 2, and Sequence. Look in the 'Features 2' tab and note the 'Gene' information. If there are multiple genes in the downloaded text file a list of all the genes in that contig is available. You have to select your gene of interest because the default setting selects the first gene in contig to be analyzed. Once your gene of interest is selected make sure the proper protein sequence is recorded, then click 'Save As' at the bottom of the page. This will convert your text file to a Genbank (gbk) file. Save the file into the same folder as the text file. Remember the saved location because this is the reference (gbk) file that will always be used to analyze sequences of that particular gene. Close the 'GBK file editor'.

In order to analyze sequences click on 'File' then 'Open Files'. By clicking 'Add' in the 'GenBank Sequence File (Optional)' section you can upload your reference (gbk) file. Then add your raw sequencing data to the 'Sample File' section by clicking 'Add' and locating the raw sequencing data. Both forward and reverse sequences can be added to this section. Press 'OK' once everything is uploaded. Go to 'Process' and click 'Run'. If forward and reverse sequences are added together then they will be divided into different contigs. A table of the variants detected in each sample is shown. For more information see the Mutation Surveyor Manuel.

# Appendix 5

**Sequencing variants identified by screening the 20 *ARVD5* positional candidate genes.** The ARVD5 ancestral haplotype is highlighted yellow. Microsatellite markers are blue.

| Locus | Accession number | Primer Exon name | Variation Name | | Amino acid change | Genomic number of variant | Samples | BB04--66 | | WP05-218 | | CH04-81 | | GT04-141 | | GP05-219 | | MH05-244 | | MS06-114 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | IVS and mRNA naming | HGV name | | | Family | 64 | | 840 | | 64 | | 453 | | 840 | | 64 | | 69 | |
| | | | | | | | Clinical Status | affected | | affected | | affected | | affected | | unaffected | | unaffected | | unaffected | |
| | | | | | | | Chromosome | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| IQSEC1 | NM_014869 | 14c | 3'UTR A>G | 2892+1060A>G | | 12914829 | | A | A | nd | nd | A | A | A | A | A | A | nd | nd | A | G |
| IQSEC1 | NM_014869 | 14c | 3'UTR C>T | 2892+1343 C>T | | 12914546 | | T | T | T | T | T | T | T | T | T | T | T | T | T | T |
| IQSEC1 | NM_014869 | 9 | IVS9+60 T>C | 2358+60T>C | | 12929868 | | C | C | nd | nd | C | C | C | C | C | C | C | C | C | C |
| IQSEC1 | NM_014869 | 8 | IVS8+73C>T | 2232+73 C>T | | 12931531 | | T | T | T | C | T | T | T | T | T | C | T | C | T | T |
| Marker | *D3S3610* | | | Start 12980656 | End | 12980996 | | 246 | 246 | 256 | 242 | nd | nd | 246 | 242 | 256 | 242 | 246 | 242 | nd | nd |
| IQSEC1 | NM_014869 | 1 | IVS1+73C>G | 65+73C>G | | 12983814 | | nd | nd | nd | nd | C | C | C | G | C | C | C | C | C | C |
| Marker | *D3S2403* | | | Start 13147397 | End | 13147709 | | 251 | 279 | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd |
| NUP210 | NM_024923 | 40d | 3'UTR T>C | 5664+1195 T>C | | 13332986 | | C | C | C | T | C | C | C | T | C | T | C | T | C | T |
| NUP210 | NM_024923 | 40c | 3'UTR G>A | 5664+1051 G>A | | 13333130 | | A | A | A | G | A | A | A | G | A | G | A | G | A | G |
| NUP210 | NM_024923 | 40c | 3'UTR A>C | 5664+1010 A>C | | 13333171 | | C | C | C | A | C | C | C | A | C | A | C | A | C | A |
| NUP210 | NM_024923 | 40c | 3'UTR C>T | 5664+977 C>T | | 13333204 | | T | T | T | C | T | T | T | C | T | C | T | C | T | C |
| NUP210 | NM_024923 | 40b | 3'UTR A>T | 5664+365A>T | | 13333816 | | A | A | A | A | A | A | A | T | A | A | A | A | A | A |
| NUP210 | NM_024923 | 38&39 | IVS39+59 G>A | 5563+59G>A | | 13335513 | | A | A | A | G | A | A | A | G | A | G | A | G | A | G |
| NUP210 | NM_024923 | 38&39 | IVS39+9 C>G | 5563+9C>G | | 13335563 | | C | G | C | G | C | G | C | G | C | C | C | G | C | C |
| NUP210 | NM_024923 | 37 | IVS37+25 G>A | 5383+25G>A | | 13336238 | | A | A | A | G | A | A | A | G | A | G | A | G | A | G |
| NUP210 | NM_024923 | 37 | 5339 G>A | 5339 G>A | V>M | 13336287 | | A | A | A | G | A | A | A | G | A | G | A | G | A | G |
| NUP210 | NM_024923 | 37 | 5225 T>C | 5225 T>C | L>S | 13336391 | | C | C | C | C | C | C | C | C | C | C | C | T | C | T |
| NUP210 | NM_024923 | 34 | 4701 C>A | 4701 C>A | T>T | 13339876 | | A | A | A | G | A | A | A | G | A | G | A | G | A | G |
| NUP210 | NM_024923 | 33 | 4582 C>T | 4582 C>T | R>W | 13342357 | | C | C | C | C | C | C | C | T | C | C | C | C | C | C |
| NUP210 | NM_024923 | 33 | 4533 G>A | 4533 G>A | S>S | 13342406 | | A | A | A | G | A | A | A | G | A | G | A | G | A | G |
| NUP210 | NM_024923 | 32 | 4332C>T | 4332C>T | C>C | 13343892 | | C | C | C | T | C | C | C | T | C | T | C | T | C | T |
| NUP210 | NM_024923 | 26 | 3489 G>A | 3489 G>A | E>E | 13354400 | | G | G | G | A | G | G | G | A | G | A | G | A | G | A |
| NUP210 | NM_024923 | 26 | IVS25 -35 C>T | 3472 -35 C>T | | 13354452 | | C | C | C | C | C | C | C | T | C | T | C | C | C | C |
| NUP210 | NM_024923 | 22&23 | 3048 T>C | 3048 T>C | F>F | 13358540 | | T | T | T | C | T | T | T | C | T | C | T | C | T | C |
| NUP210 | NM_024923 | 22&23 | IVS21 -26 C>T | 2965 -26 C>T | | 13358649 | | C | C | C | C | C | C | C | C | C | T | C | C | C | C |
| NUP210 | NM_024923 | 21 | IVS21 +143 C>T | 2964 +143 C>T | | 13359532 | | C | C | C | T | C | C | C | C | C | C | C | C | C | C |
| NUP210 | NM_024923 | 17 | 2461 C>G | 2461 C>G | P>A | 13370475 | | C | C | C | C | C | C | C | G | C | G | C | C | C | C |
| NUP210 | NM_024923 | 17 | 2357 G>T | 2357 G>T | R>L | 13370579 | | G | G | G | T | G | G | G | T | G | T | G | T | G | T |
| NUP210 | NM_024923 | 16 | 2264 C>T | 2264 C>T | A>V | 13374786 | | C | C | C | T | C | C | C | C | C | C | C | T | C | T |
| NUP210 | NM_024923 | 16 | IVS15 -53 A>G | 2155 -53 A>G | | 13374948 | | A | A | A | G | A | A | A | A | A | A | A | G | A | G |
| NUP210 | NM_024923 | 14 | 1822 A>G | 1822 A>G | I>V | 13382556 | | A | A | A | G | A | A | A | G | A | G | A | G | A | G |
| NUP210 | NM_024923 | 9 | IVS8 -64 ins | 1046 -64 ins | | 13394126 | | wt | wt | wt | ins | wt | wt | wt | wt | wt | wt | wt | ins | wt | ins |
| NUP210 | NM_024923 | 8 | IVS8 +29 G>T | 1045 +29 G>T | | 13395383 | | G | T | G | G | G | G | G | G | G | G | G | G | G | T |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NUP210 | NM_024923 | 7 | IVS7 +39 G>A | 976 +39 G>A | | 13396024 | | G | G | G | G | G | G | G | G | G | A | G | G | G |
| NUP210 | NM_024923 | 7 | 889 G>A | 889 G>A | A>T | 13396150 | | G | G | G | A | G | G | G | A | G | A | G | A | G |
| NUP210 | NM_024923 | 7 | IVS6 -98 A>G | 818 -98 A>G | | 13396320 | | A | A | A | G | A | A | A | G | A | G | A | G | A |
| NUP210 | NM_024923 | 6 | IVS5 -91 C>A | 685 -91 C>A | | 13402998 | | C | C | C | A | C | C | C | C | C | C | C | A | C |
| NUP210 | NM_024923 | 3 | IVS3 +27 T>C | 436 +27 T>C | | 13413830 | | T | T | T | C | T | T | T | C | T | C | T | C | T |
| NUP210 | NM_024923 | 3 | IVS2 -17 T>C | 305 -17 T>C | | 13414005 | | T | T | T | C | T | T | T | T | T | T | T | T | T |
| NUP210 | NM_024923 | 1 | IVS1+130G>A | 166 +130G>A | | 13436430 | | G | G | G | A | G | G | G | G | G | G | G | G | G |
| HDAC11 | NM_024827 | 2 | IVS2 +79 C>T | 151 +79 C>T | | 13497973 | | C | C | C | C | C | T | C | C | C | C | C | T | C |
| HDAC11 | NM_024827 | 3 | IVS3 +24 G>T | 252 +24 G>T | | 13500088 | | G | G | G | G | G | G | G | G | G | G | G | G | G |
| HDAC11 | NM_024827 | 4 | IVS4 +19 insG | 369 +19 insG | | 13513371 | | ins | wt | ins | ins | ins | wt | ins | wt | wt | wt | wt | wt | wt |
| HDAC11 | NM_024827 | 7 | IVS6 -97 C>T | 490 -97 C>T | | 13518274 | | C | C | C | C | C | C | C | T | C | C | C | C | C |
| HDAC11 | NM_024827 | 8 | IVS7 -74 G>A | 553 -74 G>A | | 13519310 | | G | G | G | G | G | G | G | A | G | G | G | G | G |
| FBLN2 | DA856838 | EST 4 | del GT | | | 13586337 | | del | del | del | wt | del | wt | del | wt | wt | wt | wt | wt | del |
| FBLN2 | DA856838 | EST 4 | G>A | | | 13586424 | | G | G | G | G | G | G | G | G | G | A | G | G | G |
| FBLN2 | NM_001004019 | Ex2C#2 | del** | | | 13587794 | | del | del | del | del | del | del | del | del | del | del | del | del | del |
| FBLN2 | NM_001004019 | Ex2C#2 | 1081 A>G | | | 13587937 | | A | A | A | A | A | A | A | A | A | G | A | A | A |
| FBLN2 | BC111426 | EST1 | G>A | | | 13627967 | | G | G | G | A | G | A | G | G | G | G | G | G | G |
| FBLN2 | BC111426 | EST1 | A>G | | | 13628401 | | A | A | A | G | A | G | A | A | A | G | A | A | A |
| Marker | *D3S1516* | | | Start 13628628 | End | 13629103 | | 347 | 359 | 351 | 359 | 347 | 351 | 347 | 335 | 351 | 343 | 351 | 363 | nd |
| FBLN2 | NM_001004019 | 5 | IVS5 +18 C>T | | | 13630683 | | C | C | C | C | C | C | C | C | C | T | C | C | C |
| FBLN2 | BF368851 | EST 6 | C>G | | | 13637585 | | G | G | G | G | G | G | G | G | G | G | G | C | G |
| FBLN2 | NM_001004019 | 11 | IVS11 +68 T>G | | | 13644538 | | G | G | G | G | G | G | G | G | G | G | G | T | G |
| FBLN2 | NM_001004019 | 12 | 2646 A>G | | | 13645482 | | G | G | G | G | G | G | G | G | G | A | G | A | G |
| FBLN2 | NM_001004019 | 12 | 2701 A>G | | | 13645537 | | G | G | G | G | G | G | G | G | G | A | G | A | G |
| FBLN2 | NM_001004019 | 13 | 2826 T>C | | | 13645777 | | C | C | C | C | C | C | C | C | C | C | C | T | C |
| FBLN2 | NM_001004019 | 13 | IVS13 +45 T>G | 2842 +45 T>G | | 13645838 | | G | G | G | G | G | G | G | G | T | G | T | G | T |
| FBLN2 | NM_001004019 | 15 | IVS15 +77G>C | 3085 +77 G>C | | 13647393 | | G | C | G | C | G | C | G | C | G | G | G | C | G |
| FBLN2 | W70042 | EST 2 | A>T | | | 13649190 | | T | A | T | A | T | A | T | A | A | A | T | A | A |
| FBLN2 | NM_001004019 | Ex 17 | IVS17 +105 G>T | 3338 +105 G>T | | 13653174 | | G | G | G | T | G | T | G | G | T | T | G | T* | G |
| FBLN2 | NM_001004019 | 18a | IVS17 -18 G>A | 3339 -18 G>A | | 13654045 | | G | G | G | A | G | G | G | G | A | A | G | G | G |
| FBLN2 | NM_001004019 | 18a | 3480 G>A | | | 13654204 | | G | G | G | A | G | G | G | G | A | A | G | G | G |
| FBLN2 | NM_001004019 | 18a | 3612 C>T | | | 13654336 | | C | C | C | T | C | C | C | C | T | T | C | C | C |
| FBLN2 | NM_001004019 | 18b-3'UTR | G>A | | | 13654616 | | G | A | G | A | G | G | nd | nd | G | G | G | A | G |
| FBLN2 | NM_001004019 | 18b-3'UTR | T>A | | | 13654689 | | T | A | T | T | T | T | nd | nd | T | T | T | T | T |
| FBLN2 | NM_001004019 | 18b-3'UTR | C>A | | | 13654696 | | C | A | C | C | C | C | nd | nd | C | C | C | C | C |
| FBLN2 | NM_001004019 | 18b-3'UTR | C>A | | | 13654709 | | C | A | C | C | C | A | nd | nd | C | C | C | C | C |
| FBLN2 | NM_001004019 | 18b-3'UTR | C>T | | | 13654830 | | C | T | C | T | C | T | nd | nd | T | T | C | C | C |
| FBLN2 | NM_001004019 | 18b-3'UTR | del TA | | | 13654852 | | wt | del | wt | wt | wt | del | nd | nd | wt | wt | wt | wt | wt |
| FBLN2 | NM_001004019 | 18b-3'UTR | T>A | | | 13654856 | | T | T | T | A | T | T | nd | nd | A | A | T | T | T |
| Marker | *D3S3608* | | | Start 13670236 | End | 13679541 | | 165 | 165 | 165 | 165 | 165 | 175 | 165 | 165 | 165 | 165 | nd | nd | nd |
| WNT7A | NM_004625 | 4a#1 | IVS3 -37 T>C | 571 -37T>C | End | 13835958 | | T | T | T | C | T | T | T | T | T | T | T | C | T |
| Marker | *D3S2385* | | | Start 13853945 | | 13854287 | | 146 | 142 | nd | nd | nd | nd | 146 | 146 | nd | nd | nd | nd | nd |
| WNT7A | NM_004625 | 3 | 459 T>C | 459 T>C | S>S | 13871141 | | C | T | C | C | C | T | C | T | C | T | T | T | T |
| WNT7A | NM_004625 | 3 | 315 G>A | 315 G>A | A>A | 13871285 | | A | G | A | A | A | G | A | G | A | G | G | G | G |
| WNT7A | NM_004625 | 2 | IVS2 +37 C>A | 298 +37 C>A | | 13891408 | | C | A | C | C | C | C | nd | nd | nd | nd | C | A | C |

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WNT7A | AA405144 | EST1-Ex2 | A>G | | | 13892334 | | A | G | A | G | A | A | A | G | G | G | A | G | A | G |
| WNT7A | AA405144 | EST1-Ex1 | delG | | | 13893611 | | wt | wt | wt | delG | wt | wt | wt | wt | wt | wt | wt | wt | wt | wt |
| WNT7A | NM_004625 | 1b | 5'UTR C>A | 1 -164 C>A | | 13896478 | | C | C | C | A | C | C | C | C | C | C | C | C | C | C |
| Marker | *D3S3602* | | Start 13900968 | | End | 13901215 | | 113 | 119 | 113 | 111 | 113 | 113 | 113 | 123 | 119 | 119 | 119 | 107 | nd | nd |
| Marker | *D3S1585* | | Start 13916682 | | End | 13916861 | | 118 | 126 | 118 | 118 | 118 | 118 | 118 | 126 | 126 | 126 | 118 | 138 | nd | nd |
| TPRXL | AK092426 | Ex2 | IVS1 -9 ins/delTTT | ATTT microsatelitte | | 14079871 | | 8 | 6 | 8 | 9 | 8 | 8 | 8 | 6 | 6 | 6 | 8 | 6 | 8 | 7 |
| TPRXL | AK092426 | Ex2 | IVS2 +81 A>G | 5' UTR | | 14080090 | | A | G | A | A | A | A | A | G | G | G | G | G | nd | nd |
| CHCHD4 | NM_144636 | BC033775 Ex 3 | 3'UTR A>C | | | 14129232 | | A | A | A | A | A | A | A | C | C | C | A | C | C | C |
| CHCHD4 | NM_144636 | BC033775 Ex 2 | IVS3 +36 G>C | 160 +36 G>C | | 14132891 | | G | G | G | G | G | G | G | G | G | G | C | G | G | G |
| CHCHD4 | NM_144636 | BC033775 Ex 2 | IVS3 +13 T>C | 160 +13 T>C | | 14132914 | | T | T | T | T | T | T | T | T | T | T | C | T | T | T |
| CHCHD4 | NM_144636 | BC033775 Ex 1 | 5'UTR A>G | | | 14141281 | | G | G | G | G | G | G | G | G | G | G | G | G | A | G |
| TMEM43 | NM_024334 | 4 | IVS3 -106 C>T | 298 -106 C>T | | 14147975 | | C | C | C | C | C | C | C | T | C | T | C | C | C | C |
| TMEM43 | NM_024334 | 5&6 | IVS5 +51 T>C | 442 +51 T>C | | 14149147 | | T | T | T | T | T | T | T | T | nd | nd | T | T | C | C |
| TMEM43 | NM_024334 | 7 | IVS6 -90 G>A | 513 -90 G>A | | 14150150 | | G | G | G | G | G | G | G | G | G | G | G | A | G | G |
| TMEM43 | NM_024334 | 7 | 536 T>C | | M>T | 14150263 | | T | T | T | T | T | T | T | C | C | C | T | T | T | C |
| TMEM43 | NM_024334 | 8&9 | IVS8 +55 G>A | 705 +55 G>A | | 14151447 | | G | G | G | G | G | G | G | A | A | A | G | G | G | A |
| TMEM43 | NM_024334 | 10 | IVS9 -56 G>A | 781 -56 G>A | | 14152252 | | G | G | G | G | G | G | G | A | A | A | G | G | G | A |
| TMEM43 | NM_024334 | 11 | IVS10 -47 C>T | 883 -47 C>T | | 14155634 | | C | C | C | C | C | C | C | C | C | C | C | T | C | T |
| TMEM43 | NM_024334 | 12a | 1073 C>T | | S>L | 14158166 | | T | C | T | C | T | C | T | C | C | C | C | C | C | C |
| TMEM43 | NM_024334 | 12a | 3' UTR T>C | c.1203 +115 T>C | | 14158411 | | T | T | T | T | T | T | T | C | C | C | C | C | C | C |
| TMEM43 | NM_024334 | 12b | 3' UTR C>T | | | 14158573 | | C | C | C | C | C | T | C | C | C | C | C | C | C | C |
| TMEM43 | NM_024334 | 12b | 3' UTR A>G | | | 14158759 | | A | A | A | A | A | A | A | A | A | A | A | G | A | G |
| TMEM43 | NM_024334 | 12c | 3' UTR A>C | | | 14159075 | | A | A | A | A | A | A | A | A | A | A | A | C | A | C |
| TMEM43 | NM_024334 | 12d | 3' UTR T>C | | | 14159670 | | T | T | T | T | T | T | T | T | T | T | T | C | T | C |
| TMEM43 | NM_024334 | 12d | 3' UTR T>C | | | 14159720 | | T | T | T | T | T | T | T | T | T | T | T | C | T | C |
| TMEM43 | NM_024334 | 12e | IVS12 +58 ins/del | 1203 +1942 | | 14160264 | | 12 | 11 | 12 | 11 | 12 | 11 | 12 | 11 | 13 | 14 | 11 | 13 | 12 | 14 |
| XPC | NM_004628 | 16b | IVS16 +30 ins/del T | 2823 +821 | | 14161621 | | 12 | 13 | 12 | 13 | 12 | 13 | 12 | 14 | nd | nd | nd | nd | 12 | 14 |
| XPC | NM_004628 | 16b | 3'UTR G>C | c.2823 +684 G>C | | 14161759 | | G | C | G | C | G | C | G | C | C | C | nd | nd | C | C |
| XPC | NM_004628 | 16b | 3'UTR A>G | | | 14161824 | | A | A | A | A | A | A | A | A | A | A | A | A | A | G |
| XPC | NM_004628 | 16b | 3'UTR T>A | | | 14161831 | | T | T | T | T | T | T | T | T | T | T | T | T | T | A |
| XPC | NM_004628 | 16a | 3'UTR C>G | | | 14162346 | | C | C | C | C | C | C | C | C | C | C | C | G | C | G |
| XPC | NM_004628 | 16a | 2815 C>A | | Q>K | 14162450 | | C | C | C | C | C | C | C | A | A | A | C | A | A | A |
| XPC | NM_004628 | 16a | IVS15 -40 A>G | 2605 -40 A>G | | 14162700 | | A | A | A | A | A | A | A | G | G | G | A | G | G | G |
| XPC | NM_004628 | 16a | IVS15 -51 A>G | 2605 -51 A>G | | 14162711 | | A | A | A | A | A | A | A | G | G | G | A | A | A | G |
| XPC | NM_004628 | 15 | IVS15 +90 G>T | 2604 +90 G>T | | 14163701 | | G | G | G | G | G | G | G | T | T | T | nd | nd | T | T |
| XPC | NM_004628 | 15 | IVS15 +28 C>G | 2604 +28 C>G | | 14163763 | | C | C | C | C | C | C | C | G | G | G | nd | nd | G | G |
| XPC | NM_004628 | 14 | IVS14 +68 T>C | 2514 +68 T>C | | 14164341 | | T | T | T | T | T | T | T | C | C | C | T | C | C | C |
| XPC | NM_004628 | 12&13 | IVS 12 -6 A>C | 2251 -6 A>C | | 14165238 | | A | A | A | A | A | A | A | C | C | C | A | C | C | C |
| XPC | NM_004628 | 12&13 | IVS 12 +46 C>G | 2250 +46 C>G | | 14165269 | | C | C | C | C | C | C | C | G | G | G | C | G | G | G |
| XPC | NM_004628 | 11 | 2061 G>A | | R>R | 14168890 | | G | G | G | G | G | G | G | A | A | A | G | G | G | A |
| XPC | NM_004628 | 10 | del C | | | 14172877 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| XPC | NM_004628 | 10 | 1881 T>A | | A>A | 14172989 | | A | A | A | A | A | A | A | A | A | A | A | A | A | A |
| XPC | NM_004628 | 9b | 1496 C>T | | A>V | 14174889 | | C | C | C | C | C | C | C | C | C | C | C | T | C | T |
| XPC | NM_004628 | 9a | 1415 A>C | | D>A | 14174970 | | C | C | C | C | C | C | C | C | C | C | C | C | C | C |
| XPC | NM_004628 | 9a | del G | | | 14174971 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| XPC | NM_004628 | 9a | del G | | | 14175043 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| XPC | NM_004628 | 8 | IVS7 -70 A>C | 901 -70 A>C | | 14176404 | | A | A | A | A | A | A | A | C | C | C | A | A | A | C |
| XPC | NM_004628 | 5 | IVS5 +44 C>G | 621 +44 C>G | | 14183629 | | C | C | C | C | C | C | C | G | G | G | C | C | C | G |
| XPC | NM_004628 | 3 | IVS3 +73 G>T | 412 +73 G>T | | 14186869 | | G | G | G | G | nd | nd | G | T | T | T | G | G | G | T |
| XPC | NM_004628 | 1 | G>C 5' UTR | | | 14195099 | | nd | nd | G | G | G | G | G | C | C | C | G | C | C | C |
| LSM3 | NM_014463 | | No SNP's | No SNP's | 14195341 | -14214840 | | | | | | | | | | | | | | | |
| Marker | *D3S1554* | | | Start 14342778 | End | 14343106 | | 129 | 129 | 129 | 129 | 129 | 129 | 129 | 129 | 129 | 131 | 129 | 129 | nd | nd |
| SLC6A6 | NM_003043 | 1 | G>T 5'UTR | 1 - 41125 G>T | | 14419021 | | T | G | T | G | T | G | T | G | T | G | T | G | T | G |
| SLC6A6 | NM_003043 | 1 | del 5'UTR | 1 - 40888 del | | 14419248 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| SLC6A6 | NM_003043 | 1 | del 5'UTR | 1 - 40882 del | | 14419254 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| SLC6A6 | NM_003043 | 1 | del 5'UTR | 1 - 40854 del | | 14419282 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| SLC6A6 | NM_003043 | 1 | del 5'UTR | 1 - 40846 del | | 14419290 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| SLC6A6 | NM_003043 | 1 | del 5'UTR | 1 - 40823 del | | 14419313 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| SLC6A6 | NM_003043 | 1 | del 5'UTR | 1 - 40813 del | | 14419323 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| SLC6A6 | NM_003043 | 1 | del 5'UTR | 1 - 40805 del | | 14419331 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| SLC6A6 | NM_003043 | 1 | del 5'UTR | 1 - 40802 del | | 14419334 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| SLC6A6 | NM_003043 | 2 | G>A 5'UTR | 1 - 27420 | | 14432716 | | A | G | A | G | A | G | A | G | G | G | G | G | G | G |
| SLC6A6 | BC038790 | mRNA1a | C>G 5'UTR | | | 14435728 | | G | G | G | C | G | G | G | G | G | C | G | G | C | C |
| SLC6A6 | BC038790 | mRNA1b | C>T 5'UTR | | | 14435919 | | T | T | nd | nd | T | T | T | C | C | C | T | T | C | C |
| SLC6A6 | BC111489 | 5 | IVS5+46 G>T | 599+46 G>T | | 14464374 | | G | G | G | G | G | T | G | T | G | G | G | T | G | T |
| SLC6A6 | BC111489 | 5 | IVS5+370 A>G | c.599+370 A>G | | 14464698 | | G | G | G | A | G | A | G | A | A | A | A | A | A | A |
| SLC6A6 | AK023516-Ex1a | mRNA2 | IV6-1745 G>A | 733-1745 G>A | | 14481283 | | A | A | A | G | A | A | A | A | G | G | G | A | G | A |
| SLC6A6 | AK023516-Ex1a | mRNA2 | IV6-1541 T>C | 733-1541 T>C | | 14481487 | | C | C | C | T | C | C | C | C | T | T | C | T | C | T |
| SLC6A6 | AK023516-Ex1a | mRNA2 | IV6-1521 C>A | 733-1521 C>A | | 14481507 | | A | A | A | C | A | A | A | A | C | C | A | C | A | C |
| SLC6A6 | AK023516-Ex1a | mRNA2 | IV6-1392 G>A | 733-1392 G>A | | 14481638 | | A | A | A | G | A | A | A | A | G | G | A | G | A | G |
| SLC6A6 | AK023516-Ex1a | mRNA2 | IV6-1363 G>A | 733-1363 G>A | | 14481665 | | A | A | A | G | A | A | A | A | G | G | A | G | A | G |
| SLC6A6 | AK023516-Ex1b | mRNA2 | IV6-1226 A>G | 733-1226 A>G | | 14481802 | | G | G | G | A | G | A | G | A | A | A | A | A | nd | nd |
| SLC6A6 | AK023516-Ex1c | mRNA2 | IV6-280 del A | 733-280 del A | | 14482800 | | del | wt | del | wt | del | wt | del | wt | del | wt | del | wt | nd | nd |
| SLC6A6 | NM_003043 | 15a | 3'UTR del | | | 14501629 | | del | wt | del | wt | del | wt | del | wt | del | wt | del | wt | del | wt |
| SLC6A6 | NM_003043 | 15b | 3'UTR G>A | | | 14502367 | | A | A | A | A | A | A | A | A | A | A | A | A | A | A |
| SLC6A6 | NM_003043 | 15c | 3'UTR C>G | | | 14502940 | | C | C | C | G | C | C | C | C | C | C | C | C | C | C |
| SLC6A6 | NM_003043 | 15c | 3'UTR C>G | | | 14502943 | | G | G | G | G | G | G | G | G | G | G | G | G | G | G |
| SLC6A6 | NM_003043 | 15c | 3'UTR T>C | | | 14503008 | | C | C | C | C | C | T | C | C | C | C | C | C | C | C |
| SLC6A6 | NM_003043 | 15c | 3'UTR T>C | | | 14503032 | | C | T | C | T | C | T | C | T | C | T | C | T | T | T |
| SLC6A6 | NM_003043 | 15c | 3'UTR del | | | 14503092 | | wt | del | wt | del | wt | del | wt | del | wt | del | wt | del | wt | wt |
| SLC6A6 | NM_003043 | 15d | 3'UTR C>T | | | 14503260 | | T | T | T | T | T | C | T | T | T | T | T | T | T | T |
| SLC6A6 | NM_003043 | 15e | 3'UTR G>T | | | 14503788 | | G | G | G | G | G | T | G | G | G | G | G | G | G | G |
| SLC6A6 | NM_003043 | 15g | 3'UTR del | | | 14505435 | | wt | del | wt | del | wt | del | wt | del | wt | del | wt | del | wt | del |
| SLC6A6 | NM_003043 | 15h | 3' UTR A>G | | | 14505725 | | A | A | A | A | A | A | A | A | A | A | A | G | A | A |
| GRIP2 | NM_001080423 | 23 | IVS22 -5 del | 2974 -5 del | | 14513077 | | wt | wt | wt | del | wt | wt | nd | nd | wt | wt | wt | wt | wt | wt |
| GRIP2 | NM_001080423 | 22 | IVS22+20A>G | 2973+20A>G | | 14520058 | | A | A | A | A | A | A | A | A | A | G | A | A | A | A |
| GRIP2 | NM_001080423 | 21 | 2808 C>T | | T>T | 14522186 | | C | C | C | C | C | C | C | T | C | C | C | C | C | C |
| GRIP2 | NM_001080423 | 20 | IVS20 +101 C>T | 2692 +101 C>T | | 14523212 | | C | C | C | T | C | C | C | C | C | C | nd | nd | nd | nd |
| GRIP2 | NM_001080423 | 20 | IVS19 -6 T>C | 2513 -6 T>C | | 14523498 | | T | T | T | C | T | T | T | T | T | T | T | C | nd | nd |
| GRIP2 | NM_001080423 | 18 | 2256 T>C | | S>S | 14526448 | | T | T | T | C | T | T | T | C | T | T | nd | nd | T | C |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GRIP2 | NM_001080423 | 16 | IVS15 -7 T>C | 2015 -7 T>C | | 14527998 | | T | C | T | C | T | T | T | C | T | C | C | C | C | C |
| GRIP2 | NM_001080423 | 14 | 1890 A>G | | A>A | 14530220 | | A | A | A | G | A | A | A | G | A | A | nd | nd | A | G |
| GRIP2 | NM_001080423 | 14 | IVS13 -49 A>G | 1787 -49 A>G | | 14530371 | | A | A | A | G | A | A | nd | nd | nd | nd | nd | nd | A | G |
| GRIP2 | NM_001080423 | 13 | IVS13 +42 ins AACT | 1787 +42 ins AACT | | 14530762 | | wt | wt | wt | wt | wt | wt | wt | ins | wt | wt | nd | nd | wt | wt |
| GRIP2 | NM_001080423 | 13 | 1776 T>C | | S>S | 14530815 | | T | T | T | T | T | T | T | C | T | T | nd | nd | T | T |
| GRIP2 | NM_001080423 | 13 | IVS12 -35 G>A | 1601 -35 G>A | | 14531025 | | G | G | G | G | G | G | G | A | G | G | nd | nd | G | G |
| GRIP2 | NM_001080423 | 10 | 1309 ins | | | 14536633 | | ins | ins | ins | ins | ins | ins | ins | ins | ins | ins | nd | nd | ins | ins |
| GRIP2 | NM_001080423 | 8 | 894 T>C | | S>S | 14538262 | | T | T | T | C | T | T | T | T | T | T | nd | nd | T | C |
| Marker | *D3S3595* | | | Start 14617332 | End | 14617642 | | 265 | 265 | 265 | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd |
| C3orf19 | NM_016474 | 2 | IVS2+122 T>A | 147+122 T>A | | 14671163 | | nd | nd | nd | nd | T | A | nd | nd | nd | nd | T | A | T | A |
| C3orf19 | NM_016474 | 3 | IVS3+45C>A | 248+45 C>A | | 14672185 | | C | C | C | C | C | C | C | C | C | C | C | C | C | A |
| C3orf19 | NM_016474 | 3 | IVS3+182 A>G | 248+182 A>G | | 14672322 | | G | G | G | G | G | G | G | G | G | G | G | A | G | A |
| C3orf19 | NM_016474 | 5 | IVS4-55C>A | 308 -55C>A | | 14677986 | | A | A | A | A | nd | nd | A | A | A | A | nd | nd | A | A |
| C3orf19 | NM_016474 | 5 | IVS5+83C>G & T | 485+83C>G & T | | 14678301 | | C | C | C | C | C | C | C | C | C | C | C | T | C | G |
| C3orf19 | NM_016474 | 5 | /S5+190 delTTACAAT | 485+190 delTTACAATA | | 14678409 | | del | del | del | del | del | del | del | del | del | del | del | del | del | del |
| C3orf19 | NM_016474 | 6 | IVS6+139 insTTGA | 581+139 insGATT | | 14681773 | | wt | wt | wt | wt | wt | wt | wt | wt | wt | wt | ins | ins | ins | ins |
| C3orf19 | NM_016474 | 6 | IVS6+140 A>G | 581 +140 A>G | | 14681774 | | G | G | G | G | G | G | G | G | G | G | G | A | G | A |
| C3orf19 | NM_016474 | 6 | IVS6+141 insTTGG | 581+141 insTTGG | | 14681775 | | wt | ins | wt | ins | wt | ins | wt | ins | ins | ins | wt | wt | wt | wt |
| C3orf19 | NM_016474 | 8 | IVS7 -13 A>G | 724 -13 A>G | | 14683922 | | G | G | G | G | G | G | G | G | G | G | G | G | G | G |
| C3orf19 | NM_016474 | 8 | IVS8 +33 G>C | 819 +33 G>C | | 14684062 | | G | G | G | G | G | G | G | G | G | C | G | G | G | C |
| C3orf19 | NM_016474 | 11C | 3'UTR G>A | 1404+825 G>A | | 14688530 | | A | A | A | A | A | A | A | A | A | A | A | A | A | A |
| C3orf20 | NM_032137 | 2 | 5'UTR C>T | 405-505 C>T | | 14698720 | | C | C | C | C | C | C | C | C | C | T | C | C | C | T |
| C3orf20 | NM_032137 | 2 | 5'UTR A>G | 405-498 A>G | | 14698728 | | A | A | A | A | A | A | A | G | A | A | A | A | A | A |
| C3orf20 | NM_032137 | 2 | 5'UTR C>A | 405-471 C>A | | 14698754 | | A | A | A | A | A | A | A | A | A | A | A | A | A | A |
| C3orf20 | NM_032137 | 3 | 125 G>A | 125G>A | G>D | 14699349 | | G | G | G | G | G | G | G | G | G | G | G | G | G | A |
| C3orf20 | NM_032137 | 3 | 193 G>A | 193 G>A | D>N | 14699417 | | G | G | G | G | G | G | G | G | G | G | G | A | G | A |
| C3orf20 | NM_032137 | 7 | 892 G>A | 892 G>A | A>T | 14720861 | | A | A | A | G | A | G | A | A | A | G | A | G | A | A |
| C3orf20 | NM_032137 | 8 | 1219 G>A | 1219 A>G | I>V | 14730576 | | A | G | A | G | A | G | A | A | A | G | A | G | A | A |
| C3orf20 | NM_032137 | 8 | 1264 G>C | 1264 C>G | L>V | 14730621 | | G | G | G | C | G | C | G | G | G | C | G | C | G | G |
| C3orf20 | NM_032137 | Ex9 | IVS8 -66 A>G | 1314-66 A>G | | 14731734 | | G | G | G | G | G | G | G | G | G | G | G | G | G | G |
| C3orf20 | NM_032137 | 11 | IVS11 +3 A>G | 1690+3A>G | | 14743538 | | A | A | A | G | A | G | A | A | A | G | A | G | A | A |
| C3orf20 | NM_032137 | 11 | IVS11 +23 G>A | 1690+23G>A | | 14743558 | | A | A | A | G | A | G | A | A | A | G | A | G | A | A |
| C3orf20 | NM_032137 | 13 | 2039 G>A | 2039 G>A | R>H | 14773980 | | G | A | G | G | G | A | G | G | G | G | G | G | G | G |
| C3orf20 | NM_032137 | 14 | 2334 G>A | 2334 G>A | V>V | 14776491 | | G | A | G | G | G | A | G | G | G | G | G | G | G | G |
| C3orf20 | NM_032137 | 14 | IVS14 +45 G>A | 2352+45 G>A | | 14776554 | | G | A | G | A | G | A | G | A | G | A | A | A | G | A |
| C3orf20 | NM_032137 | 16 | IVS16 +117G>C | 2630+117G>C | | 14788839 | | C | C | C | G | C | C | C | C | C | G | C | C | C | G |
| FGD5 | NM_152536 | 1c | 934 G>A | 934 G>A | V>M | 14837239 | | A | G | A | G | A | G | A | G | G | G | G | G | G | G |
| FGD5 | NM_152536 | 5&6 | IVS5 +22 G>A | 2186+22 G>A | | 14914202 | | G | A | G | A | G | A | G | A | A | A | A | A | A | A |
| FGD5 | NM_152536 | 5&6 | IVS5 -82 G>A | 2187 -82 G>A | | 14914368 | | G | A | G | A | G | A | G | A | A | A | A | A | A | A |
| FGD5 | NM_152536 | 5&6 | 2220 G>T | 2220 G>T | L>L | 14914483 | | G | T | G | T | G | T | G | T | T | T | T | T | T | T |
| FGD5 | NM_152536 | 8 | IVS8 +64 T>C | 2482+64 T>C | | 14917028 | | T | T | T | C | T | T | T | T | C | C | T | C | C | C |
| FGD5 | NM_152536 | 10 | IVS10 +50 C>T | 2613+50 C>T | | 14924272 | | C | T | C | T | C | T | C | T | T | T | T | T | T | T |
| FGD5 | NM_152536 | 15 | 3072 C>T | 3072 C>T | H>H | 14939047 | | C | T | C | C | C | T | C | C | C | C | C | C | C | C |
| FGD5 | NM_152536 | 16 | IVS15 -74 G>A | 3085-74 G>A | | 14939483 | | A | G | A | G | A | G | A | G | G | G | G | G | G | G |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FGD5 | NM_152536 | 17 | IVS16-49 C>T | 3215-49 C>T | | 14940470 | | T | T | T | T | T | T | T | T | T | C | T | C | T | T |
| FGD5 | NM_152536 | 17 | IVS16-8 A>G | 3215-8 A>G | | 14940511 | | G | G | G | G | G | G | G | G | G | G | G | G | G | G |
| NR2C2 | NM_003298 | 8 | IVS8+70 G>A | 855+70 G>A | | 15040789 | | A | G | A | G | A | A | A | G | G | G | G | G | G | G |
| NR2C2 | NM_003298 | 12 | IVS11-34 C>T | 1290-34 C>T | | 15051147 | | C | C | C | C | C | C | C | T | C | C | C | C | C | T |
| NR2C2 | NM_003298 | 14c | 1635 G>A | 1635 G>A | Q/Q | 15055700 | | G | G | G | G | G | G | G | A | G | G | G | G | G | A |
| NR2C2 | NM_003298 | 15b | 3'UTR A>G | | | 15059873 | | A | A | A | A | A | A | A | A | A | G | A | A | A | A |
| NR2C2 | NM_003298 | 15b | 3'UTR T>A | | | 15059884 | | A | T | A | T | A | A | A | T | T | T | T | T | T | T |
| NR2C2 | NM_003298 | 15e | 3'UTR C>T | | | 15061256 | | C | C | C | C | C | C | C | T | C | C | C | C | C | T |
| NR2C2 | NM_003298 | 15g | 3'UTR insGATA | | | 15062484 | | ins | wt | ins | wt | ins | isn | ins | wt | wt | wt | ins | wt | wt | wt |
| NR2C2 | NM_003298 | 15k | 3'UTR C>G | | | 15064207 | | C | C | C | C | C | G | C | C | C | C | C | C | C | C |
| NR2C2 | NM_003298 | 15k | 3'UTR G>A | | | 15064452 | | G | G | G | G | G | G | G | G | G | G | G | G | G | A |
| MRPS25 | NM_022497 | 4f | 3' UTR A>G | 522 +2704A>G | | 15066248 | | A | A | A | A | A | A | A | A | A | A | A | A | A | G |
| MRPS25 | NM_022497 | 4f | 3' UTR T>C | 522 +2700T>C | | 15066252 | | C | T | C | T | C | C | C | T | T | T | C | T | T | T |
| MRPS25 | NM_022497 | 4d | 3' UTR T>C | 522 +1875T>C | | 15067077 | | T | T | T | T | T | T | T | C | T | T | T | T | T | C |
| MRPS25 | NM_022497 | 4c | 3' UTR G>A | 522+1059G>A | | 15067893 | | A | G | A | G | A | G | A | G | G | G | G | G | G | G |
| MRPS25 | NM_022497 | 4a | 3' UTR A>T | 522+192A>T | | 15068761 | | A | A | A | A | A | A | A | T | A | A | A | A | A | A |
| MRPS25 | NM_022497 | 2 | IVS2 +75 C>T | 241+75 C>T | | 15075805 | | nd | nd | C | T | C | T | C | T | C | C | C | T | C | C |
| ZFYVE20 | NM_022340 | 14g | 3'UTR T>C | | | 15088127 | | C | T | C | T | C | C | C | C | T | T | C | T | C | T |
| ZFYVE20 | NM_022340 | 14g | 3'UTR T>A | | | 15088307 | | T | T | T | T | T | T | T | T | T | T | T | T | T | A |
| ZFYVE20 | NM_022340 | 14g | 3'UTR G>A | | | 15088523 | | A | G | A | G | A | A | A | A | G | G | A | G | A | G |
| ZFYVE20 | NM_022340 | 14e | 3'UTR insA / AA ?? | | | 15089241 | | wt | wt | wt | wt | wt | ins | nd | nd | wt | wt | wt | wt | wt | ins |
| ZFYVE20 | NM_022340 | 14c | 3'UTR C>A | | | 15090274 | | A | C | A | C | A | A | A | A | C | C | A | C | A | C |
| ZFYVE20 | NM_022340 | 14b | 1734 C>G | | S>S | 15090914 | | C | C | C | C | C | C | C | C | C | C | C | C | C | G |
| ZFYVE20 | NM_022340 | 6 | IVS5 -55 delTC | 290 -55 delTC | | 15102527 | | del | wt | del | wt | del | del | del | del | wt | wt | del | wt | del | wt |
| CAPN7 | NM_014296 | 1 | 5'UTR A>G | | | 15222800 | | G | G | nd | nd | nd | nd | nd | nd | G | A | G | G | G | G |
| CAPN7 | NM_014296 | 6 | IVS5-160A>G | 639-160A>G | | 15239826 | | A | A | A | A | A | G | A | A | A | A | A | A | A | A |
| CAPN7 | NM_014296 | 11 | IVS11+68 C>T | 1286 +68 C>T | | 15250552 | | T | C | T | C | T | C | T | C | C | C | C | C | C | C |
| CAPN7 | NM_014296 | 12 | IVS12+150 G>A | 1407 +150G>A | | 15251806 | | G | A | G | A | G | A | G | A | G | A | G | A | G | A |
| CAPN7 | NM_014296 | 13&14 | IVS12-28 T>C | 1408 -28 T>C | | 15256956 | | C | T | C | T | C | T | C | T | T | T | T | T | nd | nd |
| CAPN7 | NM_014296 | 19 | IVS18-78 del | 2074 -78del | | 15263760 | | wt | del | wt | del | wt | del | wt | del | wt | del | wt | del | wt | del |
| SH3BP5 | NM_004844 | 9b | 3'UTR del | 1368+498 del | | 15272099 | | nd | nd | wt | del | nd | nd | wt | del | wt | del | wt | del | wt | del |
| SH3BP5 | NM_004844 | 9a | 3'UTR G>A | 1368+23 G>A | | 15272574 | | G | G | G | G | G | G | G | G | G | A | G | G | G | G |
| SH3BP5 | NM_004844 | 8 | 961 G>A | | V>M | 15273553 | | G | G | G | G | G | G | G | G | G | G | G | A | G | G |
| SH3BP5 | NM_004844 | 6 | IVS5-53 C>T | 627-53 C>T | | 15276367 | | T | C | nd | nd | T | T | T | C | T | C | T | T | T | T |
| SH3BP5 | NM_004844 | 6 | IVS5-79 G>C | 627-79G>C | | 15276393 | | G | C | nd | nd | G | C | G | C | G | C | G | C | G | C |
| SH3BP5 | NM_004844 | 4 | 390 T>C | | R>R | 15286329 | | C | T | C | C | C | C | C | C | T | T | C | T | C | T |
| SH3BP5 | NM_004844 | 4 | IVS3-80 C>A | 331-80 C>A | | 15286468 | | A | A | A | A | A | A | A | A | A | A | A | A | A | A |
| Marker | *D3S3613* | | | Start 15336926 | End | 15337204 | | 193 | 197 | 193 | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd |
| SH3BP5 | NM_001018009 | 1 | 5'UTR A>G | 1-71626 A>G | | 15357874 | | A | A | A | A | A | A | A | A | A | A | A | G | A | A |

# Appendix 6

## Panel 28 of the ABI PRISM Linkage Mapping Set (v2.5-MD10 kit)

| Marker | Label - Dye | Genomic Location (Mb) | Band | Tandem Repeat | Heterozygosity | Allele Size Range |
|---|---|---|---|---|---|---|
| DXS1227 | FAM | 140.6 | Xq27.2 | di | 0.73 | 79-99 |
| DXS990 | FAM | 92.9 | Xq21.32 | di | 0.74 | 122-132 |
| DXS986 | FAM | 79.3 | Xq21.1 | di | 0.77 | 151-181 |
| DXS987 | FAM | 14.5 | Xp22.2 | di | 0.83 | 205-229 |
| DXS993 | FAM | 41 | Xp11.4 | di | 0.79 | 267-293 |
| DXS1073 | FAM | 153.4 | Xq28 | di | 0.8 | 306-334 |
| DXS8091 | VIC | 147.4 | Xq28 | di | 0.78 | 80-102 |
| DXS1106 | VIC | 102.6 | Xq22.2 | di | 0.67 | 126-140 |
| DXS1047 | VIC | 128.9 | Xq25 | di | 0.81 | 156-172 |
| DXS1001 | VIC | 119.7 | Xq24 | di | 0.82 | 191-211 |
| DXS1068 | VIC | 38.8 | Xp11.4 | di | 0.79 | 244-264 |
| DXS1214 | VIC | 31.2 | Xp21.2 | di | 0.79 | 284-298 |
| DXS8055 | VIC | 114.6 | Xq23 | di | 0.65 | 312-324 |
| DXS8051 | NED | 9.3 | Xp22 31 | di | 0.88 | 104-134 |
| DXS8043 | NED | 143.8 | Xq27.3 | di | 0.8 | 146-180 |
| DXS1060 | NED | 5.4 | Xp22.32 | di | 0.84 | 244-268 |
| DXS1226 | NED | 22 9 | Xp22.11 | di | 0 84 | 280-302 |
| DXS991 | NED | 55 5 | Xp11.21 | di | 0.8 | 313 341 |

# Appendix 7

## X chromosome fine mapping markers

| Marker | Label - Dye | Forward/Reverse Primer | Primer Sequen ce | Genomic Location (Mb) | Band | Tandem Repeat | Heterozygosity | Allele Size Range |
|---|---|---|---|---|---|---|---|---|
| DXS1053 | FAM | F | TTAAGGAAGTATGAGGCTCCA | 15 5 | Xp22 2 | di | 0 68 | 194-208 |
| | | R | TTGGTGCAAAAGTAATCACG | | | | | |
| DXS8036 | FAM | F | ATACAAACTGCCCCACTTCC | 17 | Xp22 2 | di | 0 74 | 107-135 |
| | | R | CTCTGGNTCCTGTCCTGG | | | | | |
| DXS1195 | FAM | F | TGCAAATGGACCCAAG | 17 4 | Xp22 13 | di | 0 61 | 233-241 |
| | | R | ATATCACACAGGCACAAATG | | | | | |
| DXS8019 | FAM | F | TTCATTAAAAGCCGTCTTTG | 17 7 | Xp22 13 | di | 0 8 | 156-174 |
| | | R | TGTTGAGTTTCCTCACAGC | | | | | |
| DXS999 | FAM | F | GCTAACAACCTAGACTTCAACC | 18 7 | Xp22 13 | di | 0 75 | 258-278 |
| | | R | CAGTTTCACAATCTCTGCC | | | | | |
| DXS7592 | FAM | F | AGGCAATTCCTAGTGGCTG | 19 4 | Xp22 12 | di | 0 55 | 225-233 |
| | | R | ATTGAGCCTAGATCGCACC | | | | | |
| DXS1229 | FAM | F | TAGAATCAACATAAGGCCCA | 20 5 | Xp22 12 | di | 0 33 | 202 230 |
| | | R | GGATACAATTAGCAGAATCAGACT | | | | | |
| DXS7101 | FAM | F | CGGTAACACAGATAACTGGTTT | 21 6 | Xp22 12 | di | 0 49 | 140-164 |
| | | R | GGAAGCAAATATGCCAAGAT | | | | | |
| DXS7105 | FAM | F | GATACCCAACTCTACACATGATAAA | 22 3 | Xp22 11 | di | 0 65 | 135-159 |
| | | R | CAGGATATTGACATTGATATAATCC | | | | | |
| DXS274 | FAM | F | CCTTGATTTATTGACTCCTAG | 22 55 | Xp22 11 | di | 0 55 | 134-150 |
| | | R | ATATGCTTGCACTATAATTA | | | | | |
| DXS451 | FAM | F | CTTGATCTTCTGAGGAGTGG | 23 2 | Xp22 11 | di | 0 8 | 182-194 |
| | | R | TTATTCCTAGGCTTAGGATTC | | | | | |
| DXS8099 | FAM | F | AGCTGGTTTTGTGATTCTGC | 24 3 | Xp22 11 | di | 0 74 | 105-119 |
| | | R | GGCCTTGAGATGTAGCCA | | | | | |
| DXS8056 | FAM | F | CCTGGGAGGTGGAGGT | 25 1 | Xp21 3 | di | 0 57 | 137 161 |
| | | R | GGGCATAAGTGGCTTCG | | | | | |
| DXS1202 | PET | F | CCCAGCACAGGTGTCA | 26 2 | Xp21 3 | di | 0 81 | 265-285 |
| | | R | GGGTTGTTGCTCCAAAG | | | | | |
| DXS7102 | FAM | F | TAGGCTCTACTGGATCTCCAAC | 26 6 | Xp21 3 | di | 0 66 | 258-272 |
| | | R | TGCTATAAATATTCATGTGCAGA | | | | | |
| DXS8065 | FAM | F | GAGCGACACCCTCGTT | 27 3 | Xp21 3 | di | 0 63 | 137 147 |
| | | R | ATCCCTGGCACATAATGA | | | | | |
| DXS8047 | FAM | F | GTGCCCGGCCCATATTA | 27 6 | Xp21 3 | di | 0 75 | 130-146 |
| | | R | CCAAAGTTTTGTTGCATCCC | | | | | |
| DXS8010 | FAM | F | CTGGCCCAAAGGTTAATTTT | 27 8 | Xp21 3 | di | 0 6 | 151-165 |
| | | R | AAGGTGGAAGGTCCACTG | | | | | |
| AFM318ZC1 | FAM | F | AATTGTTGTTAGTGCAAATGAT | 28 | Xp21 3 | di | n/a | 277 |
| | | R | CTAGTCCCTGCCAGGTG | | | | | |
| AFMB072ZE9 | FAM | F | AGCCACTGCTCTGGAA | 28 2 | Xp21 3 | di | n/a | 192 |
| | | R | TGACATTGAAAATATACACAATTTA | | | | | |
| DXS1218 | VIC | F | AAGACTAAGATTGTTTCAGTTTGTT | 29 1 | Xp21 3 | di | 0 67 | 255 275 |
| | | R | GTGCTTTTGTAATTTTAACCCA | | | | | |
| AFMA082XB9 | FAM | F | TCTTCAGTTATACCTTCTGGC | 29 86 | Xp21 2 | di | n/a | 125 |
| | | R | GAACTCTTGACTTGCAGTTT | | | | | |
| DXS1065 | FAM | F | TAATCTGATCCCCAACCTGC | 29 89 | Xp21 2 | di | 0 21 | 160-164 |
| | | R | GGGACATTGTGAANACAGG | | | | | |
| DXS8039 | FAM | F | AAACTAACAAACCTCTAGCCAG | 30 1 | Xp21 2 | di | 0 8 | 147-165 |
| | | R | TTCATGTCCATGTGAACAG | | | | | |
| DXS985 | FAM | F | TTTCCAAACTTCACTAAAAC | 30 5 | Xp21 2 | di | 0 6 | 133-139 |
| | | R | AAAATGTTGTTGTACTGGTC | | | | | |
| DXS992 | NED | F | AAGAATGGGACTCCATTTCA | 30 7 | Xp21 2 | di | 0 87 | 201-205 |
| | | R | GCTTATCCACTGGGACAGAA | | | | | |
| DXS1234 | FAM | F | GAAAGATTGTAAACTAAAGTGTGC | 31 | Xp21 2 | di | 0 46 | 129 |
| | | R | GGATGCAAAACAATGCGCTGCCTC | | | | | |
| DXS1241 | FAM | F | TGTCTGTCTTCAGTTATATG | 31 41 | Xp21 2 | di | 0 51 | 245 |
| | | R | ATAACTTACCCAAGTCATGT | | | | | |
| DXS1036 | FAM | F | TGCAGTTTATTATGTTTCCACG | 31 68 | Xp21 1 | di | 0 62 | 141-153 |
| | | R | GCCATTGATAAGTGCCAGAT | | | | | |
| DXS1067 | FAM | F | TATGTCCTCAGACTATTCAGATGCC | 31 71 | Xp21 1 | di | 0 64 | 214-230 |
| | | R | CCTCCAGTAACAGATTTGGGTG | | | | | |
| DXS997 | FAM | F | TGGCTTTATTTTAAGAGGAC | 31 79 | Xp21 1 | di | 0 65 | 108-117 |
| | | R | GTTTTCAGTTTCCTGGGT | | | | | |

| Marker | Label - Dye | Forward/Reverse Primer | Primer Sequen ce | Genomic Location (Mb) | Band | Tandem Repeat | Heterozygosity | Allele Size Range |
|---|---|---|---|---|---|---|---|---|
| DXS1219 | FAM | F | TTAATGTTTCANCCAGGTAAAT | 32 | Xp21 1 | dı | 0 55 | 230-246 |
| | | R | GATCACTCCAAAGGATAGATTGT | | | | | |
| DXS1238 | FAM | F | TCCAACATTGGAAATCACATTTCAA | 32 1 | Xp21 1 | dı | 0 85 | 174 |
| | | R | TCATCACAAATAGATGTTTCACAG | | | | | |
| DXS1243 | NED | F | ATGGTTGATACTTTAGAGATTATTG | 33 1 | Xp21 1 | dı | n/a | 67 |
| | | R | GTGTCAAGTAAAATTTATTTACAGC | | | | | |
| DXS1049 | PET | F | CGAGATGGTGCTACTACACG | 35 1 | Xp21 1 | dı | n/a | 192 198 |
| | | R | CATCAAGCTAAAACTCGGTG | | | | | |
| DXS8090 | VIC | F | GGGTGAAATTCCATCACAAA | 36 9 | Xp21 1 | dı | 0 81 | 154-172 |
| | | R | ACAAATGCAGATGTACAAAAAATA | | | | | |
| DXS1069 | FAM | F | AGCCTAACCCACATAACAGC | 38 5 | Xp11 4 | dı | n/a | 254-268 |
| | | R | AGCTACTATATTNACCTTGGTCTTG | | | | | |
| DXS1209 | PET | F | TCTATCACTATATATATGCNATCCC | 84 4 | Xq21 1 | dı | 0 45 | 106-116 |
| | | R | AGATGGAGAACAGATTATTGGT | | | | | |
| DXS6803 | FAM | F | GAAATGTGCTTTGACAGGAA | 86 3 | Xq21 31 | tetra | n/a | 110-126 |
| | | R | CAAAAAGGGACATATGCTACTT | | | | | |
| DXS1217 | VIC | F | TCCAGTAAGTTTGGTCTATATGACG | 88 3 | Xq21 31 | dı | 0 62 | 231-249 |
| | | R | ATGAAGTATCGTATCTGAATCCCG | | | | | |
| DXS6786 | FAM | F | GCATGCACCTATATTCCCAG | 90 3 | Xq21 31 | tri | n/a | 228 |
| | | R | TTTGTGAATTTTAGGATATTTTTCC | | | | | |
| AFM178XC11 | FAM | F | GCTGATTTTCTACTTTAAAACTTGC | 91 | Xq21 31 | dı | n/a | 164 |
| | | R | GCTCAGGTTCCTCTATGTGC | | | | | |
| DXS1203 | FAM | F | CCTGAATTTCCCCAGC | 92 7 | Xq21 32 | dı | 0 65 | 206-220 |
| | | R | TCCCACTTGCCAACTC | | | | | |
| DXS1170 | VIC | F | TGCCATCTATACTAGTCAGGG | 93 3 | Xq21 32 | dı | 0 67 | 137 |
| | | R | CTTTACAATAATAATTGGGTGCT | | | | | |
| DXS8077 | NED | F | ACATTTTTATGACATTAAACACACA | 95 2 | Xq21 33 | dı | 0 6 | 101-121 |
| | | R | GACAAAATTNTCCAGTGAAGTCATT | | | | | |
| DXS454 | PET | F | AGAAGACATAAGGATACTGC | 97 9 | Xq21 33 | dı | 0 66 | 144-145 |
| | | R | GATCCCAACTATTTCTTTCT | | | | | |
| DXS1231 | PET | F | AAAAGCACCCTCATAGGAACC | 98 5 | Xq22 1 | dı | 0 49 | 202-210 |
| | | R | GGCTCTCTGCACCAGCAT | | | | | |
| DXS8089 | VIC | F | TGGTGCNAGGATTTGG | 100 6 | Xq22 1 | dı | 0 81 | 91 111 |
| | | R | TCTTGTATGGCTTCTGGGTC | | | | | |
| DXS8064 | FAM | F | AGAATCGCTTGACCCTTG | 117 2 | Xq24 | dı | 0 56 | 209-225 |
| | | R | CTGATGGCTGCCAACTC | | | | | |

# Appendix 8

## X-linked positional candidate gene primers and amplification conditions

| PHKA2 (NM_000292) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I.D. | Sequence | Amplicon Size | Annealing temp. | Betaine (final conc. (M)) | MgCl (final conc (mM)) |
| 1 | NM_000292-Ex1a-F | agaagggcctgtttctctgg | 398 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex1a-R | GCTGGAGCGACGAAGGTC | | | | |
| | NM_000292-Ex1b-F | GGACGTAAAAGGGGCTTGG | 500 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex1b-R | cctgggaaggagaatgagtt | | | | |
| 2 | NM_000292-Ex2-F | ttaggacaagcatctgttaccc | 299 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex2-R | tgtgggtttttcagttcaagg | | | | |
| 3 | NM_000292-Ex3-F | ctttccagggcagttgctg | 240 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex3-R | cgtcttaagcatgttaactgtcac | | | | |
| 4 | NM_000292-Ex4-F | ctggctcccagaagtagcag | 398 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex4-R | ttaaagcagtgaagcagcaca | | | | |
| 5 | NM_000292-Ex5-F | ctctctgcaacatggcttca | 247 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex5-R | gcgctagtctcgctaacctt | | | | |
| 6 | NM_000292-Ex6-F | tgtggctgcaggaacataaa | 291 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex6-R | ccttgatgctgacttttgaca | | | | |
| 7 | NM_000292-Ex7-F | ttgcttaatgaaaaaggaacacc | 246 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex7-R | actggatcccaggccttc | | | | |
| 8 | NM_000292-Ex8-F | ggcaaccgtattcgaatgtt | 334 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex8-R | acctcatggggaactgagg | | | | |
| 9 | NM_000292-Ex9-F | ctgacagcaaagaacacttgg | 296 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex9-R | tccctttccaatattccaacata | | | | |
| 10 | NM_000292-Ex10-F | cccatgaggcacaatggta | 299 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex10-R | ccgaaacagtgacctaaattctt | | | | |
| 11 | NM_000292-Ex11-F | ttctatgttgcaccgctgtt | 239 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex11-R | aacaggccaagcaatgagtc | | | | |
| 12 | NM_000292-Ex12-F | ggatgagtaccatttcataaatcc | 298 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex12-R | gcctggcagagagtcccta | | | | |
| 13 | NM_000292-Ex13-F | tgaatatgttgagccccaaa | 300 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex13-R | tgctaaagcataggcacataca | | | | |
| 14 | NM_000292-Ex14-F | tgaagagacaagagtcatgagctaac | 249 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex14-R | ccccaaatctagaaccgaga | | | | |
| 15 | NM_000292-Ex15-F | tagcctgagttgcctggatt | 250 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex15-R | ggagaaaagcttaccagacagc | | | | |
| 16&17 | NM_000292-Ex16&17-F | ttgtcagctggtgtttccag | 600 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex16&17-R | gggcagcaaaggactttaaga | | | | |
| 18 | NM_000292-Ex18-F | tctggcagaagcaccacat | 387 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex18-R | ccttgtttttaattatcaatcattgtc | | | | |
| 19 | NM_000292-Ex19-F | tttctacagcttggtttcattg | 299 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex19-R | gcttttaaattatgtccctgcatt | | | | |
| 20 | NM_000292-Ex20-F | gcgaaattccgtccctacta | 573 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex20-R | tggaagagtttggggactca | | | | |
| 21 | NM_000292-Ex21-F | ttccgaaagttccatgccta | 250 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex21-R | cctaagccaatctgttcatgg | | | | |
| 22 | NM_000292-Ex22-F | tgcaaaatgcacagtgatcc | 314 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex22-R | ctgggagattggaagtggaa | | | | |
| 23&24 | NM_000292-Ex23-24-F | agggcaacatgggtcactaa | 586 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex23-24-R | gggtttcagagacagggtga | | | | |
| 25 | NM_000292-Ex25-F | cacagccttcctcagagtgg | 299 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex25-R | aagcagttcttcccctgaga | | | | |
| 26 | NM_000292-Ex26-F | cacctgaccctacctttaccc | 249 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex26-R | ccccacagtgctggttctaa | | | | |
| 27 | NM_000292-Ex27-F | cagagaaggccctcattgtc | 376 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex27-R | ggacaggggtgtgttcagat | | | | |
| 28 | NM_000292-Ex28-F | acggccctggattaagagag | 230 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex28-R | atagagccgccctctacaca | | | | |
| 29 | NM_000292-Ex29-F | ctctgctgctgctttctgtg | 248 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex29-R | tcagaagttgggcatgaggt | | | | |
| 30 | NM_000292-Ex30-F | cgagatggcgcactttcc | 300 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex30-R | tgccttgccaggacagtaa | | | | |
| 31 | NM_000292-Ex31-F | atatgggtttccagggcaga | 236 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex31-R | cagagtccatgcaatgaagc | | | | |
| 32 | NM_000292-Ex32-F | gctacggtcacccttggtta | 349 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex32-R | cagaatctccagcctcaacag | | | | |
| 33 | NM_000292-Ex33a-F | ctcagaaggccaaggctcta | 451 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex33a-R | CAGGGTGAGTGCTACCATTG | | | | |
| | NM_000292-Ex33b-F | TGGCTTCTTATTTGCAGGAA | 399 | TD 54 | yes (0 75) | 1 5 |
| | NM_000292-Ex33b-R | CCAATTTAAGCTGGCGTCTC | | | | |

| GPR64 (NM_001079858) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I D | Sequence | Amplicon Size | Annealing temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_001079858-Ex1F | ctgaaacgggcatatgggta | 561 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858 Ex1R | ctgtgcgagtctgggtctc | | | | |
| 2 | NM_001079858-Ex2F | gctgatgagctcactgcagtatt | 246 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex2R | gggaatttgcaatgaaggag | | | | |
| 3 | NM_001079858 Ex3F | atggcagctgtgatgatctg | 214 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex3R | cacacatgcagataaacagaaaaa | | | | |
| 4 | NM_001079858-Ex4F | agtgaggctgataaacattcaca | 298 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex4R | aagcttcttggaggaaaacaa | | | | |
| 5&6 | NM_001079858 Ex5&6F | cgcttatgggacatgcagat | 399 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858 Ex5&6R | ggaaaccttcaacaccattca | | | | |
| 7 | NM_001079858 Ex7F | atttggggaaaagggaaatg | 208 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex7R | cccagtagactgcacaagca | | | | |
| 8 | NM_001079858 Ex8F | cccagtagactgcacaagca | 208 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858 Ex8R | atttggggaaaagggaaatg | | | | |
| 9 | NM_001079858-Ex9F | tgtcttaaaggtgtcaacagtgtg | 245 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex9R | gaaattataaacaccagccatttg | | | | |
| 10 | NM_001079858 Ex10F | tgacaaagtaaatgtgaataaaaacaa | 269 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex10R | aaataatctgaacctcaaacaaatg | | | | |
| 11 | NM_001079858 Ex11F | ttaggattcagtctttggggta | 294 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex11R | acctttgggctggtctcttt | | | | |
| 12 | NM_001079858-Ex12F | ttcaaagtgaacaccataagcaa | 246 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex12R | atctcccaggaatgcagaac | | | | |
| 13 | NM_001079858 Ex13F | ccctaagggctctcgtgatt | 247 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex13R | ccctttgcatttgcaattta | | | | |
| 14 | NM_001079858-Ex14F | tcagtgtccccttcctcac | 324 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex14R | cccaaggacaaaggagtcaa | | | | |
| 15 | NM_001079858-Ex15F | tgtaggaaccgtgatacttaatgg | 247 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex15R | cccaaaattctaaaactccttagc | | | | |
| 16 | NM_001079858-Ex16F | attgccagaggggaaagact | 582 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex16R | gaatcccgaaaggaaagctc | | | | |
| 17 | NM_001079858-Ex17F | ccaggagtgagtttgccata | 394 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858 Ex17R | agcaacctgagaaaccatga | | | | |
| 18 | NM_001079858-Ex18F | gaattgcttgaacccaggag | 372 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858 Ex18R | ttgcctcaatcatgaaatacaa | | | | |
| 19 | NM_001079858-Ex19F | gcgttttaattattaggctggtg | 399 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex19R | cagccttaagaatattcgagca | | | | |
| 20 | NM_001079858-Ex20F | ttgtccctcgatttctgacc | 342 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex20R | ttctgcctgatctgcctttt | | | | |
| 21 | NM_001079858-Ex21F | gtctatttagggcttacttgcttc | 400 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex21R | CATTCAATCTCCTGTCTTTGAC | | | | |
| 22 | NM_001079858-Ex22F | ATGgtaagttccagtcataacc | 356 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858 Ex22R | tggtgtataaatcagccaca | | | | |
| 23 | NM_001079858 Ex23F | tgaatggatccctagaatacttga | 296 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex23R | ccgaggaaagggcttagtta | | | | |
| 24 | NM_001079858 Ex24F | aggaggcagaggttgcagt | 499 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex24R | cttcatcgtttccacctgct | | | | |
| 25 | NM_001079858-Ex25F | aatcacttgaacccaggag | 350 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex25R | gttttattaggatgatagcaggag | | | | |
| 26 | NM_001079858 Ex26F | ttagagattcattgtatcattttgaac | 400 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858 Ex26R | cattcagtttgtatttgatgctg | | | | |
| 27 | NM_001079858-Ex27F | ttttgtgcaaagtgagcttatga | 244 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex27R | tggatcttaatttatacaatcttttcc | | | | |
| 28 | NM_001079858 Ex28F | cctgaagtgagaaatcagttgc | 397 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858 Ex28R | ggaaccacattttgatccact | | | | |
| 29 | NM_001079858 Ex29aF | tttgagtggccttagggaag | 480 | TD 54 | yes (0 75) | 1 5 |
| | NM_001079858-Ex29aR | CTCCAGATGTTGCCGAGAAT | | | | |

| PDHA1 (NM_000284) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I D. | Sequence | Amplicon Size | Annealing temp. | Betaine (final conc. (M)) | MgCl (final conc (mM)) |
| 1 | NM_000284-Ex1-F | ccctctctggcgctgatac | 436 | TD54 | yes (0.75) | 1 5 |
| | NM_000284-Ex1-R | gcggaggcgaagtaaagg | | | | |
| 2 | NM_000284-Ex2-F | tttccataacttcattctgataactg | 248 | TD54 | yes (0.75) | 1 5 |
| | NM_000284-Ex2-R | gcatgaacatcaaaatgttctattcta | | | | |
| 3 | NM_000284-Ex3-F | tgctcctttgcagtttttca | 369 | TD54 | yes (0 75) | 1.5 |
| | NM_000284-Ex3-R | ggcagggtaaagcccaata | | | | |
| 4 | NM_000284-Ex4-F | tgacaggttctactagcccaca | 367 | TD54 | yes (0 75) | 1 5 |
| | NM_000284-Ex4-R | ccagagcaccaaaggagaga | | | | |
| 5 | NM_000284-Ex5-F | acatatttggggtgcggttt | 248 | TD54 | yes (0 75) | 1.5 |
| | NM_000284-Ex5-R | ttgtctggggctgtgaaagt | | | | |
| 6 | NM_000284-Ex6-F | aaaatgcagggcattaattagtatct | 250 | TD54 | yes (0 75) | 1 5 |
| | NM_000284-Ex6-R | tgctacttctttagaaattcaactgt | | | | |
| 7 | NM_000284-Ex7-F | gcagagcagcagctgttaga | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_000284-Ex7-R | aagggtctgcacaaggacat | | | | |
| 8 | NM_000284-Ex8-F | ttctcatacaggagcaggtca | 241 | TD54 | yes (0 75) | 1 5 |
| | NM_000284-Ex8-R | taacagcttcagcaggcaca | | | | |
| 9 | NM_000284-Ex9-F | taataaagggcctgcgtttg | 250 | TD54 | yes (0 75) | 1 5 |
| | NM_000284-Ex9-R | ccaaaagcaaatactgaaatgtg | | | | |
| 10 | NM_000284-Ex10-F | ctcattgggacaatccaacc | 296 | TD54 | yes (0 75) | 1 5 |
| | NM_000284-Ex10-R | tggaggaacccacacaaact | | | | |
| 11 | NM_000284-Ex11a-F | tcttctgaattagcaactgttcg | 447 | TD54 | yes (0.75) | 1 5 |
| | NM_000284-Ex11a-R | GCATGCAATTACTACCTGCTCA | | | | |

| SH3KBP1 (NM_001024666 and NM_031892) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I.D. | Sequence | Amplicon Size | Annealing temp. | Betaine (final conc. (M)) | MgCl (final conc (mM)) |
| 1a | NM_001024666-Ex1-F | ggtacgggacaggatgctta | 248 | TD54 | yes (0 75) | 1 5 |
| | NM_001024666-Ex1-R | gggcaacaggctacaaaaga | | | | |
| 1b | NM_031892-Ex1-F | gctgaggaggagggGAGA | 456 | - | | - |
| | NM_031892-Ex1-R | CCCTCAAGGCTGAAGTGGAC | | | | |
| 2 | NM_031892-Ex2-F | ctggccttcaagggatactg | 360 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex2-R | tatcacagaggagggcaagc | | | | |
| 3 | NM_031892-Ex3-F | catttccagaaagatggagctta | 283 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex3-R | cacagttcaccagaagtttcca | | | | |
| 4 | NM_031892-Ex4-F | gcatgcctaggaagcctctc | 250 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex4-R | atgggtgggtaagcagtgac | | | | |
| 5 | NM_031892-Ex5-F | gggctctgcatgctacacta | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex5-R | agaagggaggctggagcac | | | | |
| 6 | NM_031892-Ex6-F | ttcctggctggttgtgaagt | 374 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex6-R | aattggaagcctccagcac | | | | |
| 7 | NM_031892-Ex7-F | ccaatttcgagcattttgtc | 295 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex7-R | catttttctgatgcgtttgc | | | | |
| 8 | NM_031892-Ex8-F | tgtgcctttccctgaactta | 250 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex8-R | gtcaacaccccaaagcagtt | | | | |
| 9 | NM_031892-Ex9-F | gaaagggcctccattctctt | 298 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex9-R | caaatgccagtgctcttcaa | | | | |
| 10 | NM_031892-Ex10-F | ctgtggtccttctgtgagagg | 216 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex10-R | aaccaggaatcccgaactct | | | | |
| 11 | NM_031892-Ex11-F | gatttaaatgctggtgggtgt | 229 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex11-R | tctggactagccgtgtaacaga | | | | |
| 12 | NM_031892-Ex12-F | gccttccatgatcaccttca | 356 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex12-R | agttcccctcctgtgttcct | | | | |
| 13 | NM_031892-Ex13-F | ccttaacgtgtgcagcaaga | 250 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex13-R | atgctggttgaaggcaagc | | | | |
| 14 | NM_031892-Ex14-F | tgaggaagagaggtgcttatga | 288 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex14-R | gctgcttgcagaagagtcct | | | | |
| 15 | NM_031892-Ex15-F | atacagcattatattaacgaaagtgg | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex15-R | cctgcactgtttgtgtatctttg | | | | |
| 16 | NM_031892-Ex16-F | ggacagtggtgttaacggaga | 446 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex16-R | gggatccatgctttcagaac | | | | |
| 17 | NM_031892-Ex17-F | ggatcagatggattttgtagtca | 228 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex17-R | aaagccctcacaatgtcctg | | | | |
| 18 | NM_031892-Ex18a-F | cagtcgaccgggttgataat | 431 | TD54 | yes (0 75) | 1 5 |
| | NM_031892-Ex18a-R | CCACTATGGACTAAACTGCCTGA | | | | |

| RSK2 (NM_004586) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I D | Sequence | Amplicon Size | Annealing temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_004586-Ex1-F | AGGGAGGAGGCGGTGAAG | 247 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex1 R | ACCCAAACCAAAACCCTGAG | | | | |
| 2 | NM_004586-Ex2 F | ctgcatactcatatggtggttct | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex2 R | cactgtaaaatgattactctgctca | | | | |
| 3 | NM_004586-Ex3-F | tgggttggtaaaaggtggtt | 400 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex3-R | ttgctggtaggaagactggt | | | | |
| 4 | NM_004586-Ex4 F | acattttaggggaaataatctacagt | 298 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex4-R | ccaaggagtcccatgagc | | | | |
| 5 | NM_004586-Ex5-F | gcaacttcctttggttatgtaatg | 500 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex5-R | tggctaccactcagaaactaactt | | | | |
| 6 | NM_004586-Ex6-F | tccacatggaaactaagataaatca | 371 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex6-R | tttgggtctgattcttccaga | | | | |
| 7 | NM_004586-Ex7-F | aattttgctattcctttcacga | 388 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex7-R | gcagagatgtgaagcacagg | | | | |
| 8 | NM_004586-Ex8-F | aaatgtggtgatctgatgctt | 267 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex8-R | ctacgcctggccctaataca | | | | |
| 9 | NM_004586-Ex9-F | ctgggtgacagagcgagact | 370 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex9-R | ccagtttctttaacattcactgc | | | | |
| 10 | NM_004586-Ex10-F | aaacacttgttatgtagcaattttaga | 286 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex10-R | gattacaggcatgagccaca | | | | |
| 11 | NM_004586-Ex11-F | aaccagtgccatgtgaatga | 692 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex11-R | cgcagtggtgtgatctcagt | | | | |
| 12 | NM_004586-Ex12&13-F | tgaaaacaaaggctgagcaa | 397 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex12&13-R | tggaaaagattttctgatgaaaca | | | | |
| 13 | NM_004586-Ex14-F | ttccagaataagtgaaaaaggtca | 294 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex14-R | tagaaaaagatctattcacagaatgaa | | | | |
| 14 | NM_004586-Ex15-F | atgtagggcatcaacaatgc | 397 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex15-R | acctttagctcacagggttct | | | | |
| 15 | NM_004586-Ex16-F | caatctggccaggtagttgaa | 298 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex16-R | tttcaatgtgagttagacaactgattc | | | | |
| 16 | NM_004586-Ex17-F | tgaatatgaagtgagtttaatttttcc | 366 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex17-R | agcagttattttccatttcagg | | | | |
| 17 | NM_004586-Ex18-F | gcattgcttgacagtttgtga | 380 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex18-R | gaaaggctggcattttatcc | | | | |
| 18 | NM_004586-Ex19-F | tgccttatttggtaaccatgttc | 356 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex19-R | ctgggcaacagagtgagacc | | | | |
| 19 | NM_004586-Ex20-F | cctggtgccttatggagctt | 356 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex20-R | caaaggtttctaactggtagtatttct | | | | |
| 20 | NM_004586-Ex21-F | gttgaccaaagggctgctac | 383 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex21-R | caacagaactggatgcagga | | | | |
| 21 | NM_004586-Ex22a-F | tttgaccagtgtggttattttca | 399 | TD54 | yes (0 75) | 1 5 |
| | NM_004586-Ex22a-R | CAGCAGGAACAGCAGCATTA | | | | |

| KLHL34 (NM_153270) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I.D. | Sequence | Amplicon Size | Annealing temp. | Betaine (final conc. (M)) | MgCl (final conc (mM)) |
| 1 | NM_153270-Ex1a-F | ccctaatcccctaggctttg | 492 | TD54 | yes (0 75) | 1 5 |
| | NM_153270-Ex1a-R | CCATTTCTGCCTGCTCAGTT | | | | |
| | NM_153270-Ex1b F | TACTAGGGGCCTCCGAAAAC | 599 | TD54 | yes (0 75) | 1 5 |
| | NM_153270 Ex1b R | AAGTAGCGCCCACAGAGC | | | | |
| | NM_153270 Ex1c-F | GCCTGCTGGACTTCATCTACA | 482 | TD54 | yes (0 75) | 1 5 |
| | NM_153270-Ex1c-R | GAGGCCAGAGCCCGAGTA | | | | |
| | NM_153270 Ex1d-F | CAGAGTTGCTGGAGCGTGT | 700 | TD54 | yes (0 75) | 1 5 |
| | NM_153270-Ex1d-R | CGACGCAGGTCGTACATCTC | | | | |
| | NM_153270 Ex1e F | CACGCTTGGACGGAAGTG | 698 | TD54 | yes (0 75) | 1 5 |
| | NM_153270-Ex1e-R | ACCTCTCCCTCCCTCTCGT | | | | |
| | NM_153270-Ex1f-F | TAGTGGGCTACGACCTCGAC | 598 | TD54 | yes (0 75) | 1 5 |
| | NM_153270-Ex1f-R | gccaacaattttctgagttttg | | | | |

| SMS (NM_004595) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I D | Sequence | Amplicon Size | Annealing temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 1 | NM_004595-Ex1 F | attagcgaggcctcacacc | 435 | TD54 | no | 1 5 |
| | NM_004595-Ex1 R | GTGGCAGTCCGGGAGAGT | | | | |
| 2 | NM_004595-Ex2 F | aagctcagtcctcaaactgtg | 430 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex2 R | caccaagttccattcccatc | | | | |
| 3 | NM_004595-Ex3-F | gtggtcttttaattgggtga | 480 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex3-R | ctggaaagtaagcaatgaaga | | | | |
| 4 | NM_004595-Ex4 F | cgctaaggttaggtcagcaa | 566 | TD54 | yes (1 5) | 2 |
| | NM_004595-Ex4 R | ctcagctcactgcaacctct | | | | |
| 5 | NM_004595-Ex5-F | ttggattcgttttgtgcttg | 433 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex5-R | tgcatctcaaaaaccagcag | | | | |
| 6 | NM_004595-Ex6-F | ggctggattttcttcccact | 491 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex6-R | cacatttgcctctagtctgtgg | | | | |
| 7 | NM_004595-Ex7 F | ttttgttaaatttgtgatacccaag | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex7 R | aaaagctgttttaataaggcacaa | | | | |
| 8 | NM_004595-Ex8-F | tcctgtagcagttgctgtgg | 384 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex8-R | tccaccatggacctcaattt | | | | |
| 9 | NM_004595-Ex9-F | gcagtgctaggtggatgtga | 390 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex9-R | gatgatgccgctctatcctt | | | | |
| 10 | NM_004595-Ex10-F | tgtgatttgtgaaaatacagaagc | 400 | TD54 | yes (1 5) | 2 |
| | NM_004595-Ex10-R | gctggagaaccagggcttta | | | | |
| 11 | NM_004595-Ex11a-F | aggccagcaaacgaattaca | 365 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex11a-R | GGGGTCAGTGCATTTAGCAG | | | | |
| | NM_004595-Ex11b-F | TGAAAGTCAGCTGAAGGATGG | 383 | TD54 | yes (0 75) | 1 5 |
| | NM_004595-Ex11b-R | aacagaccaggagatactgttgc | | | | |

Platinum Taq DNA polymerase was used for exon 4 and 10

| PHEX (NM_000444) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I.D. | Sequence | Amplicon Size | Annealing temp. | Betaine (final conc. (M)) | MgCl (final conc (mM)) |
| 1 | NM_000444-Ex1-F | actttgctgagggagagcac | 499 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex1-R | ggcaaacagccctatacctg | | | | |
| 2 | NM_000444-Ex2-F | ccgtgtcaacactgagaagga | 294 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex2-R | tggtaaacctcaagcaaacaga | | | | |
| 3 | NM_000444-Ex3-F | aaggcttggaaactggttga | 391 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex3-R | tcttggaaatgaatttagcatgg | | | | |
| 4 | NM_000444-Ex4-F | ttgtgattatcaaatgacttccaa | 299 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex4-R | tcattcctttaacacctttaaaca | | | | |
| 5 | NM_000444-Ex5-F | tcttgagaagatgcacttaaattct | 400 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex5-R | cactaaggcagcatgagtctct | | | | |
| 6 | NM_000444-Ex6-F | aatatggctgggatgcagac | 245 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex6-R | taaagaagcagcggccttag | | | | |
| 7 | NM_000444-Ex7-F | gctattttctgctcttccatgtc | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex7-R | tcattcaggatatactgcctttaata | | | | |
| 8 | NM_000444-Ex8-F | tgttttggcacatgtaggaag | 244 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex8-R | caaatttacacagagatgaaatcca | | | | |
| 9 | NM_000444-Ex9-F | tctgcctttcttacttgctacc | 295 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex9-R | gttttcaaaggatgtgagaagg | | | | |
| 10 | NM_000444-Ex10-F | ggagctttgccaactgtttc | 240 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex10-R | ttccagcatgtaaagttatcca | | | | |
| 11 | NM_000444-Ex11-F | ttcagccatgggttttatcc | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex11-R | gacaatacccacaggccact | | | | |
| 12 | NM_000444-Ex12-F | atgcaagccaattttgttcc | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex12-R | ggggtcattcagagtcaacag | | | | |
| 13 | NM_000444-Ex13-F | gaagggcgcatttctacatc | 243 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex13-R | tgctaggacttgggttagttga | | | | |
| 14 | NM_000444-Ex14-F | ttgctccttcctatgctgaag | 296 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex14-R | ggcaagccagctactctgac | | | | |
| 15 | NM_000444-Ex15-F | gtccaacatccccattgttc | 243 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex15-R | tcatggaaaagttctttgcattt | | | | |
| 16 | NM_000444-Ex16-F | gaggagtgcctttcagatgg | 235 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex16-R | tctgtcctctaccaaagcaaca | | | | |
| 17 | NM_000444-Ex17-F | tgggtatggttattcgattgtt | 298 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex17 R | tgctgctgctatgcaaagat | | | | |
| 18 | NM_000444-Ex18-F | ttttgaaggcttgtcgaggt | 379 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex18-R | ttcagcaggtatggggtagg | | | | |
| 19 | NM_000444-Ex19-F | ttgatgcctcttgctgaatg | 226 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex19-R | tggctatggtatgaattgagg | | | | |
| 20 | NM_000444-Ex20-F | tttgctattcctgtgctagctg | 296 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex20-R | tgctgaaaatcaaagtcatgc | | | | |
| 21 | NM_000444-Ex21-F | ttggagcagttaaaacagcaga | 237 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex21-R | tctggtagagcccttggatg | | | | |
| 22 | NM_000444-Ex22a-F | gaatgccaaccttcttctagc | 496 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex22a R | GGACAATAGCAACAATTTAGCAGA | | | | |
| | NM_000444-Ex22b-F | CAAAGTTGTAGGGCTTATAAAGTGG | 300 | TD54 | yes (0 75) | 1 5 |
| | NM_000444-Ex22b-R | attccacagagcagcaagc | | | | |

| SAT1 (NM_002970) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I.D. | Sequence | Amplicon Size | Annealing temp. | Betaine (final conc. (M)) | MgCl (final conc (mM)) |
| 1 | NM_002970-Ex1-F | gcctcattggatgagaggtc | 416 | TD 54 | yes (0 75) | 1 5 |
| | NM_002970-Ex1-R | cgagacctcaagacgtgaatc | | | | |
| 2 | NM_002970-Ex2-F | gagtggggtgattcacgtct | 397 | TD 54 | yes (0 75) | 1 5 |
| | NM_002970-Ex2-R | AGTGCTCTTTCGGCACTTCT | | | | |
| 3 | NM_002970-Ex3-F | ccagGAGCTGGCTAAATATGA | 367 | TD 54 | yes (0 75) | 1 5 |
| | NM_002970-Ex3-R | tttgaacacttgctggctattt | | | | |
| 4&5 | NM_002970-Ex4&5-F | gcctggactgcactcacata | 492 | TD 54 | yes (0 75) | 1 5 |
| | NM_002970-Ex4&5-R | gacatttcagcctgctctca | | | | |
| 6 | NM_002970-Ex6a-F | ggtcttgagagcaggctgaa | 461 | TD 54 | yes (0 75) | 1 5 |
| | NM_002970-Ex6a-R | CACCACCTTGTTGTTTATCGAA | | | | |
| | NM_002970-Ex6b-F | AATGAGCACCCATTCCAAAG | 359 | TD 54 | yes (0.75) | 1 5 |
| | NM_002970-Ex6b-R | tgttcatccaatgatgtttgc | | | | |

| EIF2S3 (NM_001415) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I.D. | Sequence | Amplicon Size | Annealing temp | Betaine (final conc. (M)) | MgCl (final conc (mM)) |
| 1 | NM_001415-Ex1-F | actaggacccctgcagatga | 225 | TD 54 | yes (0.75) | 1 5 |
| | NM_001415-Ex1-R | catgctcccgacactgacta | | | | |
| 2 | NM_001415-Ex2-F | gcagatggtggcaagatgta | 246 | TD 54 | yes (0.75) | 1 5 |
| | NM_001415-Ex2-R | tcctcatactattcaagtaatcacaca | | | | |
| 3&4 | NM_001415-Ex3&4-F | accttgtgatccacccacct | 595 | TD 54 | yes (0.75) | 1 5 |
| | NM_001415-Ex3&4-R | agccattaggttcaccgttt | | | | |
| 5 | NM_001415-Ex5-F | cgacaagccctgtaaccatt | 293 | TD 54 | yes (0 75) | 1 5 |
| | NM_001415-Ex5-R | tccccttaaatatttcatcttgg | | | | |
| 6 | NM_001415-Ex6-F | caacccataatttgctttgtctc | 400 | TD 54 | yes (0 75) | 1 5 |
| | NM_001415-Ex6-R | gcaaaaggtagggcgataag | | | | |
| 7 | NM_001415-Ex7-F | ttccacttgttgctgacattta | 374 | TD 54 | yes (0 75) | 1 5 |
| | NM_001415-Ex7-R | cacctctgcagaaagaggaaat | | | | |
| 8 | NM_001415-Ex8-F | tggccactattttagaaacattaaca | 298 | TD 54 | yes (0 75) | 1 5 |
| | NM_001415-Ex8-R | tttttcaatttggattgtccat | | | | |
| 9 | NM_001415-Ex9-F | ccagtccctgttggaacttt | 364 | TD 54 | yes (0.75) | 1 5 |
| | NM_001415-Ex9-R | catgattgttccccacccta | | | | |
| 10 | NM_001415-Ex10-F | gcccaggctggtctttaact | 392 | TD 54 | yes (0 75) | 1 5 |
| | NM_001415-Ex10-R | tggtttaacactcccatttgtc | | | | |
| 11 | NM_001415-Ex11-F | ttttatgattgacatatttcccaatag | 364 | TD 54 | yes (0 75) | 1 5 |
| | NM_001415-Ex11-R | tcctcaaggaagattgtgtgc | | | | |
| 12 | NM_001415-Ex12a-F | ttatggaggagcaggctttg | 291 | TD 54 | yes (0.75) | 1 5 |
| | NM_001415-Ex12a-R | CAATTCCCCAATATTGCTTTG | | | | |

| IL1RAPL1 (NM_014271) | | | | | | |
|---|---|---|---|---|---|---|
| Exon | Primer I.D. | Sequence | Amplicon Size | Annealing temp. | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 2 | NM_014271-Ex2-F | attgcaccgatcatgtttga | 286 | TD54 | no | 1 5 |
| | NM_014271-Ex2-R | ccactaatggcacatgtgtaga | | | | |
| 3 | NM_014271-Ex3-F | tgacccaatatggatgctca | 493 | TD54 | no | 1 5 |
| | NM_014271-Ex3-R | agacaacggctttaggcaaa | | | | |
| 4 | NM_014271-Ex4-F | atacactgccttggcagcac | 348 | TD54 | no | 1 5 |
| | NM_014271-Ex4-R | tcacacccatctaagttttctatcc | | | | |
| 5 | NM_014271-Ex5-F | gcaaattgattccccagaga | 486 | TD54 | no | 1 5 |
| | NM_014271-Ex5-R | tgattcaaaagcaaggaggaa | | | | |
| 6 | NM_014271-Ex6-F | gcagtatgttcatcgttacttgg | 299 | TD54 | no | 1 5 |
| | NM_014271-Ex6-R | catcgggttacaatctaggc | | | | |
| 7 | NM_014271-Ex7-F | caaaatgttacctgtcagtttgc | 291 | TD54 | no | 1 5 |
| | NM_014271-Ex7-R | tcctttcagcttactttctattcac | | | | |
| 8 | NM_014271-Ex8-F | gggcataacttaattaacacagca | 448 | TD54 | no | 1 5 |
| | NM_014271-Ex8-R | tcggtggctctaatgcaaat | | | | |
| 9 | NM_014271-Ex9-F | acccgttaacccacatctga | 370 | TD54 | no | 1 5 |
| | NM_014271-Ex9-R | tatacgagctgctgccattg | | | | |
| 10 | NM_014271-Ex10-F | aaaagtgcatgttttcatgtcc | 397 | TD54 | no | 1 5 |
| | NM_014271-Ex10-R | tgtgaacacacaaagacgtttc | | | | |
| 11 | NM_014271-Ex11a-F | ggagaagcaagtcccaaactc | 597 | TD54 | no | 1 5 |
| | NM_014271-Ex11a-R | CGTGTTGTGAAAGGTAGAACG | | | | |
| | NM_014271-Ex11b-F | TGGGGAGCTGCAGACTGT | 480 | TD54 | no | 1 5 |
| | NM_014271-Ex11b-R | CCCTGTTTTTCCTCGTGTTC | | | | |

| TAB3 (NM_152787) | | | | |
|---|---|---|---|---|
| Exon | Primer I.D. | Sequence | Betaine (final conc. (M)) | MgCl (final conc (mM)) |
| 6 | TAB3-Ex6-F | CCTTAGAGGAGATGCCACCA | yes (0.75) | 1.5 |
| | TAB3-Ex6-R | ACATCACACATCACCAGCTCT | | |
| 7 | TAB3-Ex7a-F | TGCATAACTCTCTGGCAAAGC | yes (0.75) | 1.5 |
| | TAB3-Ex7a-R | TCCATCACTTGAGCTATGTACCA | | |
| | TAB3-Ex7b-F | GGATGAATAGAAATCGCCTTT | yes (0.75) | 1.5 |
| | TAB3-Ex7b-R | ATGGAAGATGGCTGTTGAGG | | |
| | TAB3-Ex7c-F | TCCAGCTGTTGTTGCTGCTA | yes (0.75) | 1.5 |
| | TAB3-Ex7c-R | CGGCGTACTCTGAGGAGTTT | | |
| | TAB3-Ex7d-F | CCACAAATTCCAAGCAATCTC | yes (0.75) | 1.5 |
| | TAB3-Ex7d-R | CAACTGGGAAGGTTGCAC | | |
| | TAB3-Ex7e-F | CCACCTCCTTCTCAATGTCC | yes (0.75) | 1.5 |
| | TAB3-Ex7e-R | CACTAAATGGAGGTTTTGGTTGA | | |
| | TAB3-Ex7f-F | CACGCCACCTTCAAGTTCTC | yes (0.75) | 1.5 |
| | TAB3-Ex7f-R | GCTGCCTCACCACAGAAATTA | | |
| 8 | TAB3-Ex8-F | GCATATCGAAATTTTCCAGTCG | yes (0.75) | 1.5 |
| | TAB3-Ex8-R | AAAGAAAAATAAATGACCTGACA | | |
| 9 | TAB3-Ex9-F | GACTAAGGCCGTATTCTTGGT | yes (0.75) | 1.5 |
| | TAB3-Ex9-R | AAATCTACAAAACCAAAGTTATT | | |
| 10 | TAB3-Ex10-F | AATTTTGAAATCCTGGAAAACA | yes (0.75) | 1.5 |
| | TAB3-Ex10-R | CCTCTAGTGGAAATTTCTTCAGA | | |
| 11 | TAB3-Ex11-F | CCTGTTCACTCTTGCAGTTGA | yes (0.75) | 1.5 |
| | TAB3-Ex11-R | TGGACAGCACAGCCTTAGAA | | |
| 12 | TAB3-Ex12-F | CCTCCCTTGATTTTTGTGGT | yes (0.75) | 1.5 |
| | TAB3-Ex12-R | CCCGAGGTTTCCTTTTCTTC | | |

| DMD (X14298) Exons | | | | | |
|---|---|---|---|---|---|
| Exons | Primer I D | Sequence | Annealing temp | Betaine (final conc (M)) | MgCl (final conc (mM)) |
| 45 | NM_004013-Ex2-F | TTCTTTGCCAGTACAACTGCAT | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex2-R | TAGTGCCTTTCACCCTGCTT | | | |
| 46 | NM_004013-Ex3-F | TTTGTGTCCCAGTTTGCATT | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex3-R | CCTAATGGGCAGAAAACCAA | | | |
| 47 | NM_004013-Ex4-F | CTCGGTCAAGTCGCTTCATT | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex4-R | GAAGCACCCAGGAAACAAAA | | | |
| 48 | NM_004013-Ex5-F | TGCTGCTAAAATAACACAAATCA | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex5-R | TGATACCAAATGAGAAAATTCAGTG | | | |
| 49 | NM_004013-Ex6-F | AGGCAGAAATTGATCTGCAA | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex6-R | TAGTCCACGTCAATGGCAAA | | | |
| 50 | NM_004013-Ex7-F | TCCAATAATATATTCACCAAATGGA | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex7-R | TCTCACCCAGTCATCACTTCA | | | |
| 51 | NM_004013-Ex8-F | TTGAAATTGGCTCTTTAGCTTG | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex8-R | GGAGAGTAAAGTGATTGGTGGAA | | | |
| 52 | NM_004013-Ex9-F | TGGCTGGTCTCACAATTGTACT | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex9-R | TGTGAGGGGGATATATGAACTT | | | |
| 53 | NM_004013-Ex10-F | AAGCAACATAAATGTGAGATAACG | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex10-R | GCCTGGGTGACAGTGAGACT | | | |
| 54 | NM_004013 Ex11-F | GACCTGAGGATTCAGAAGCTG | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex11-R | CTCCCAAGCTCCAGTTTAGC | | | |
| 55 | NM_004013-Ex12-F | CAACTCACCCCATTGTTGGT | TD54 | yes (0 75) | 1 5 |
| | NM_004013-Ex12-R | TTGTCCCTGGCTTGTCAGTT | | | |
| 56 | NM-004014 Ex2-F | GCCTGATTGAAATGTCTTTTG | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex2-R | TTTACGTAGACATGTGAGATACCAG | | | |
| 57 | NM-004014 Ex3-F | TCAAGACTGCTTTGTAGTTCACA | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex3-R | TGACCCTTGGGTGAGAAGAG | | | |
| 58 | NM-004014 Ex4-F | GAGATAGAAATTGACCTGGGAGTT | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex4-R | TTTTTCAGCATCTATGTTTAATTTGA | | | |
| 59 | NM-004014 Ex5-F | GGTTACCCTCTTGTTCAACTGT | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex5-R | TTGTGAAAGACGGACTGATTTC | | | |
| 60 | NM-004014 Ex6-F | ACTGGCACTGCACCCTAAAG | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex6-R | TCCTCACAAATATTACCATGAACA | | | |
| 61 | NM-004014 Ex7-F | TTGCTTTAGTGTTCTCAGTCTTGG | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex7-R | ATCCAATTGGCCTTCCTCTT | | | |
| 62 | NM-004014 Ex8-F | TTGTAATTCTGGAGATTAATGTTGTC | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex8-R | TGTAGGCCAGGCTAATGTCG | | | |
| 63 | NM-004014 Ex9-F | TTCAAAAGCAAAAATCATGTTG | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex9-R | GCAAGTAACTTTCACACTGCAAA | | | |
| 64 | NM-004014 Ex10-F | GGCTATTTTTGTGGTGGTTG | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex10-R | GCAACAGGAATTGTGACAGC | | | |
| 65 | NM-004014 Ex11-F | AGCTAGGATTCTCAGAGGAAAAA | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex11-R | GGGCCCTGTTGTAATGCTAA | | | |
| 66 | NM-004014 Ex12-F | AGAGTTCACACATCATTGAGCA | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex12-R | TTGACAAGGAATGGCACAAA | | | |
| 67 | NM-004014 Ex13-F | CCCCACTCTGTGGAAATACTG | TD54 | yes (0 75) | 1 5 |
| | NM-004014 Ex13-R | AAACGAAGCTCTGTGGGTTTT | | | |
| 68 | DMD Ex7-F | ACACCTCCTTTGCCATCTTG | TD54 | yes (0 75) | 1 5 |
| | DMD Ex7-R | CTCCCACTTTGGAGACGGTA | | | |
| 69 | NM_004014-Ex15-F | CGTGGTAGAAGGTTTATTAAAGAGTG | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex15-R | GGGAAGTGACAAATCACAGC | | | |
| 70 | NM_004014-Ex16-F | GTGTCATGGGGCAGAAGACT | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex16-R | GAGGGTGTTCAGCTGAGAGG | | | |
| 71 | NM_004014-Ex17-F | GGTTTGGCTATTGCTTTCCA | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex17-R | CAAATAAACAGAACAAAAGAGAACCA | | | |
| 72 | NM_004014-Ex18-F | AAAAGCATTCTAGGCCATGTG | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex18-R | TCAGACAAGTTTGGGGAAAAA | | | |
| 73 | NM_004014-Ex19-F | TTTCAGGAATGTTCGATTAGGTC | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex19-R | TGATGCTAATTCCTATATCCTGTGC | | | |
| 74 | NM_004014-Ex20-F | TGGTAGATCACAACCTCAGCA | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex20-R | GGCACTTTTCTATGTGTGCAA | | | |
| 75 | NM_004014-Ex21-F | TTGGTGATGATATTCAAATTACTTTTT | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex21-R | TGCTCTCTGAGGTTTAGTTTTGTTT | | | |
| 76 | NM_004014-Ex22-F | TCCTGAGTGTGTATCATATTTTGAA | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex22-R | ACGGCCAAATATTCATGTCC | | | |
| 77 | NM_004014-Ex23-F | AAAAATGTAATCATGGCCCTTT | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex23-R | AGTTGGGTAGGGAAGCGAGT | | | |
| 78 | NM_004014-Ex24-F | TCCAGCAGAGAATAAATGGTCA | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex24-R | AAACGCCAAACATAAAAAGCA | | | |
| 79 | NM_004014-Ex25-F | TTCCCAAATGGCAAAGAAAC | TD54 | yes (0 75) | 1 5 |
| | NM_004014-Ex25-R | AATACCACTACCCTTCACAAAAAT | | | |

# Appendix 9

## Manually designed MLPA probes for non-overlapping *DMD* exons

**Dp116 - Ex1 = Synthetic MLPA D-Probe** Chr.Xp21.2
Synthetic MLPA probe for the detection of exon 1 of the DMD gene, isoform Dp116.
Genbank sequence: NM_004014
Total length probe incl. primers: 63 + 63 = 126 nt

Genomic area around Ex 1 of Dp116. Lower case letters represents intervening sequences, upper cases letter represents the exon. Red letters are non-coding and blue are coding.

acagaatgta agcaaagttg gcattttaaa gcagggctct ttcagtttct 31436337
GGTTTTCTC AGGATTGCTA TGCAACAGGA TCAGTGCTGT AGTGCCCGGT 31436287
TCAAGCTGAA AATGTTACAC AGGAAGACAT ACCATGTAAA Ggtcagattc 31436237
ttctactata ataattttct tgatctgtgt gtatacaagt gaagttgaat 31436187

acagaatgta agcaaagttg gcattttaaa gcagggctct ttcagtttct 31436337
GGTTTTCTC AGGATTGCTA TGCAACAGGA TCAGTGCTGT AGTGCCCGGT 31436287
TCAAGCTGAA AATGTTACAC AGGAAGACAT ACCATGTAAA Ggtcagattc 31436237
ttctactata ataattttct tgatctgtgt gtatacaagt gaagttgaat 31436187

**RPO** : Tm: 74.53 °C., GC% = 38
RHS (5' phosphorylated!!) + **reverse primer sequence (bold)**:
CAAGCTGAAAATGTTACACAGGAAGACATACCATGTAAAG**TCTAGATTGGA TCTTGCTGGCAC**
Total length (including 23 nt PCR primer): 40 + 23 = 63 nt.

**LPO**: Tm= 84.83 °C., GC% = 52
**forward primer sequence (bold)** + LHS:
**GGGTTCCCTAAGGGTTGGA**CTCAGGATTGCTATGCAACAGGATCAGTGCTG TAGTGCCCGGTT
Total length (including 19 nt primer): 44 + 19 = 63 nt

**Dp71 - Ex1 = Synthetic MLPA D-Probe** Chr.Xp21.2
Synthetic MLPA probe for the detection of exon 1 of the DMD gene, isoform Dp71.
Genbank sequence: NM_004015
Total length probe incl. primers: 47 + 49 = 96 nt

Genomic area around Ex 1 of Dp71. Lower case letters represents intervening sequences, upper cases letter represents the exon. Red letters are non-coding and blue are coding.

```
tccgcagtgc tttcagctgt gagcttgggc ggcggcggcg gcggcgctcc  31194946
 CTTTCGGGG AGCCCGGCGG CTCTGGGAAG CTCACTCCTC CACTCGTACC  31194896
CACACTCGAC CGCGGAGCCC TTGCAGCCAT GAGGGAACAG CTCAAAGGgt  31194846
```
aagtggatcg cggccccggc cccgtctgat ggccgcaatc cgtgcacttg 31194796

```
tccgcagtgc tttcagctgt gagcttgggc ggcggcggcg gcggcgctcc  31194946
 CTTTCGGGG AGCCCGGCGG CTCTGGGAAG CTCACTCCTC CACTCGTACC  31194896
CACACTCGAC CGCGGAGCCC TTGCAGCCAT GAGGGAACAG CTCAAAGGgt  31194846
```
aagtggatcg cggccccggc cccgtctgat ggccgcaatc cgtgcacttg 31194796

**RPO** : Tm: 77.08 °C., GC% = 58
RHS (5' phosphorylated!!) + **reverse primer sequence (bold)**:
GCAGCCATGAGGGAACAGCTCAAAGG**TCTAGATTGGATCTTGCTGGCAC**
Total length (including 23 nt PCR primer): 26 + 23 = 49 nt.

**LPO**: Tm= 83.83 °C., GC% = 68
**forward primer sequence (bold)** + LHS:
**GGGTTCCCTAAGGGTTGGA**CGTACCCACACTCGACCGCGGAGCCCTT
Total length (including 19 nt primer): 28 + 19 = 47 nt

# Appendix 10

## Phase 1 – *BRCA1* and *BRCA2* full gene screening results

### (i) Variant detection

Ninety-six probands with hereditary breast cancer from the population of Newfoundland had full gene *BRCA1* and *BRCA2* screening for the detection of causative point mutations and small insertions and deletions. A total of 15 deleterious mutations were identified in the 96 probands, 8 mutations in *BRCA1* (8.3%) and 7 mutations in *BRCA2* (7.3%) (Table A10.1, and Figure A10.1 and A10.2). Eleven of those 15 mutations (73.3%) were identified in high-risk families (7 *BRCA1*, 4 *BRCA2*) and 4 (26.7%) were identified in moderate-risk families (1 *BRCA1*, 3 *BRCA2*) (Table A.10.2). *BRCA1* and *BRCA2* mutations explain 15.6% of those 96 Newfoundland hereditary breast cancer families. Only one of the mutations, *BRCA2* c.4642del4, had not been previously reported (332, 388). Another variant, *BRCA2* 10204A>T, was detected but considered an unclassified variant (389), therefore, it was not included in the list of pathogenic mutations. This variant is in the last exon (exon 27) of the *BRCA2* gene which is not translated, thus its effect on protein function is unclear.

### (ii) Probands with *BRCA1* mutations

Eight probands screened positive for a *BRCA1* mutation (Table A10.1 and A10.2). The average age of diagnosis for a *BRCA1* proband was 41.8 years. All *BRCA1* probands were female. The primary cancer site for six of the eight *BRCA1* probands was the breast, four of which had a second primary cancer (two had a second breast cancer and two were

also diagnosed with a primary ovarian cancer). The remaining two of the eight *BRCA1* probands were diagnosed with ovarian cancer only. Seven of the eight *BRCA1* probands were from high-risk families and one was from a moderate-risk family (Table A10.1 and A10.2), with an average of nine breast or ovarian cancer cases per *BRCA1* family. Forty-two of the 52 female breast cancer cases in the *BRCA1* families were diagnosed under the age of 50 (80.8%). There were 15 ovarian cancer cases in the eight families (an average of approximately two ovarian cancers per family) and one male breast cancer case. All families had other cancers associated with the disease as well (Table A10.1).

**(iii) Probands with *BRCA2* mutations**

Seven probands screened positive for a *BRCA2* mutation (Table A10.1 and A10.2). The average age of diagnosis for a *BRCA2* proband was 47.9 years. All probands, but one, were female. All probands had breast cancer as their primary cancer. Three of the seven probands screened had a second primary cancer in the breast, and one proband had a primary ovarian cancer as well (Table A10.1). Four of the seven probands were from high-risk families and 3 from moderate-risk families (Table A10.1 and A10.2), and there was an average of five breast and ovarian cancer cases per *BRCA2* family. There were a total of 26 breast cancer cases in all seven *BRCA2* families, 16 of which (61.5%) were diagnosed under the age of 50. There were three ovarian cancers diagnosed (less than one case per family) and three male breast cancer cases. Four of the seven *BRCA2* families had other cancers associated with the disease (Table A10.1).

**Table A10.1: *BRCA1* and *BRCA2* mutations detected in phase 1 (conventional full gene screening) and family characteristics.**

| | | | | | Proband | | | | Cancer cases in family | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene | Exon | Mutation | Putative Effect | Mutation Ethnicity (BIC report) | Sex | Age of Onset | Site of Primary | Other Primary | Br <50 | Br 50+ | M Br | Ov-FT | Total | Other Cancers |
| BRCA1 | 2 | c.185delAG | 39 stop | Jewish / Western Europe | F | 61 | Breast | Ovarian | 1 | 1 | 0 | 1 | 3 | Bn, Ln |
| | 7 | c.546G>T | 143 stop | West Europe | F | 28 | Breast | Breast | 8 | 0 | 1 | 1 | 10 | Pr |
| | 11 | c.2190delA | 700 stop | West Europe | F | 33 | Breast | None | 5 | 0 | 0 | 3 | 8 | Co, Ln, Sto |
| | 11 | c.3118delA | 1023 stop | West/Cen/East Europe | F | 48 | Ovary | None | 4 | 1 | 0 | 1 | 6 | Sto |
| | 11 | c.3717C>T | 1200 stop | West Europe | F | 40 | Breast | None | 3 | 1 | 0 | 0 | 4 | Co, Pr |
| | 11 | c.3726C>T | 1203 stop | West Europe | F | 48 | Ovary | None | 6 | 1 | 0 | 8 | 15 | Co, Leu, Ln, Pr |
| | 11 | c.3875del4 | 1262 stop | West Europe | F | 36 | Breast | Breast | 10 | 3 | 0 | 0 | 13 | Co, Pr, Leu, Ln, NHL |
| | 13 | c.4446C>T | 1443 stop | West Europe | F | 40 | Breast | Ovarian | 5 | 3 | 0 | 1 | 9 | Ut, Pr |
| BRCA2 | 5 | c.697delAA | 181 stop | Not reported | F | 36 | Breast | Breast | 4 | 0 | 0 | 0 | 4 | None |
| | 10 | c.2024delTTTAT | 599 stop | West Europe | F | 44 | Breast | Breast | 4 | 3 | 1 | 1 | 9 | La, Pr, Sto |
| | 11 | c.4642delAAGA | 1478 stop | Not reported | F | 52 | Ovarian | Breast | 0 | 1 | 0 | 1 | 2 | None |
| | 11 | c.6137C>A | 1970 stop | West Europe | F | 42 | Breast | None | 3 | 0 | 1 | 1 | 5 | Pr, Co |
| | 11 | c.6293C>G | 2022 stop | West Europe | F | 42 | Breast | Breast | 5 | 2 | 0 | 1 | 8 | Co, Ln, Pr, Sto, Sk |
| | 11 | c.6714delACAA | 2166 stop | West Europe | F | 59 | Breast | None | 0 | 3 | 0 | 0 | 3 | Ln, Sk |
| | 11 | c.6985delCT | 2259 stop | West Europe | M | 60 | Breast | None | 0 | 1 | 1 | 0 | 2 | None |

Abbreviations: Br-Breast, Ov-Ovarian, Bn- Brain, Ln- Lung, Pr-Prostate, Co-Colon, Sto-Stomach, Leu-Leukemia, NHL-Non-Hodgkin's Lymphoma, Ut-Uterine, La-Larynx, Sk-Skin, Tes-Testicular, Thr-Throat, Cx-Cervix

**Table A10.2: Categorization of the 15 *BRCA1* and *BRCA2* mutations found in phase 1 into high- and moderate-risk families**

| Gene | Probands Grouped by age | High-risk families (4 or more cases) | Moderate-risk families (2 to 3 cases) | Total |
|---|---|---|---|---|
| BRCA1 | 20-29 | 1 | 0 | 1 |
| | 30-39 | 2 | 0 | 2 |
| | 40-49 | 4 | 0 | 4 |
| | 50-59 | 0 | 0 | 0 |
| | 60+ | 0 | 1 | 1 |
| | **All ages** | **7** | **1** | **8** |
| BRCA2 | 20-29 | 0 | 0 | 0 |
| | 30-39 | 1 | 0 | 1 |
| | 40-49 | 3 | 0 | 3 |
| | 50-59 | 0 | 2 | 2 |
| | 60+ | 0 | 1 | 1 |
| | **All ages** | **4** | **3** | **7** |
| **Total** | **All ages** | **11** | **4** | **15** |

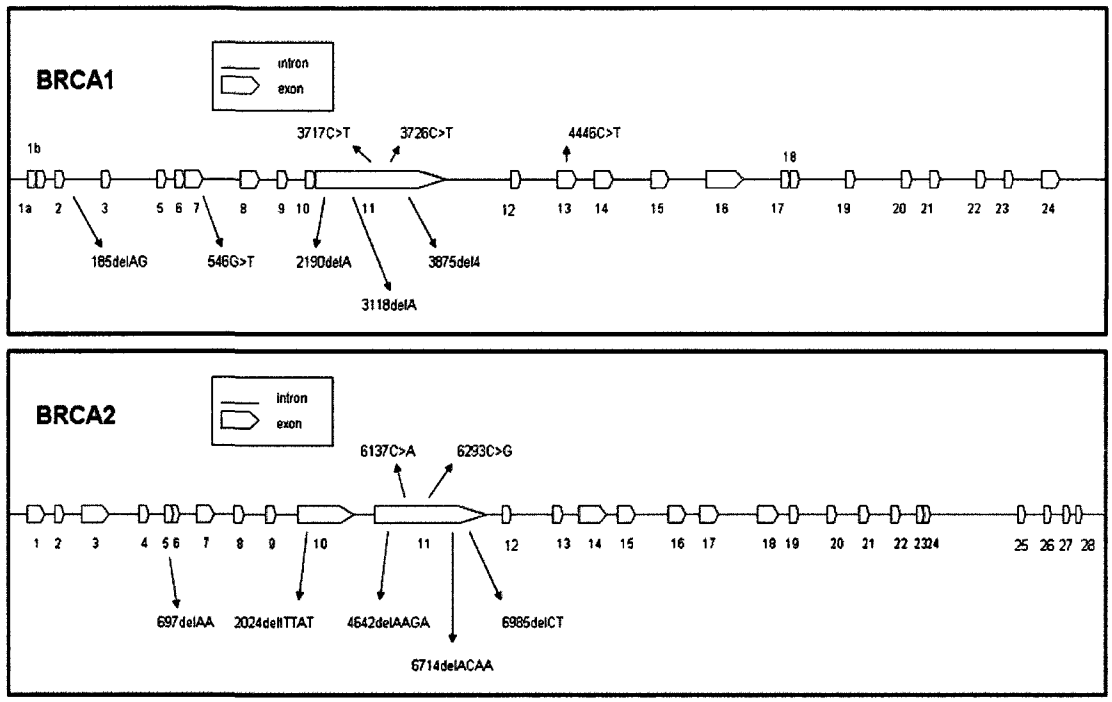



**Figure A10.1: Distribution of the pathogenic mutations found in *BRCA1* and *BRCA2* during phase 1.**

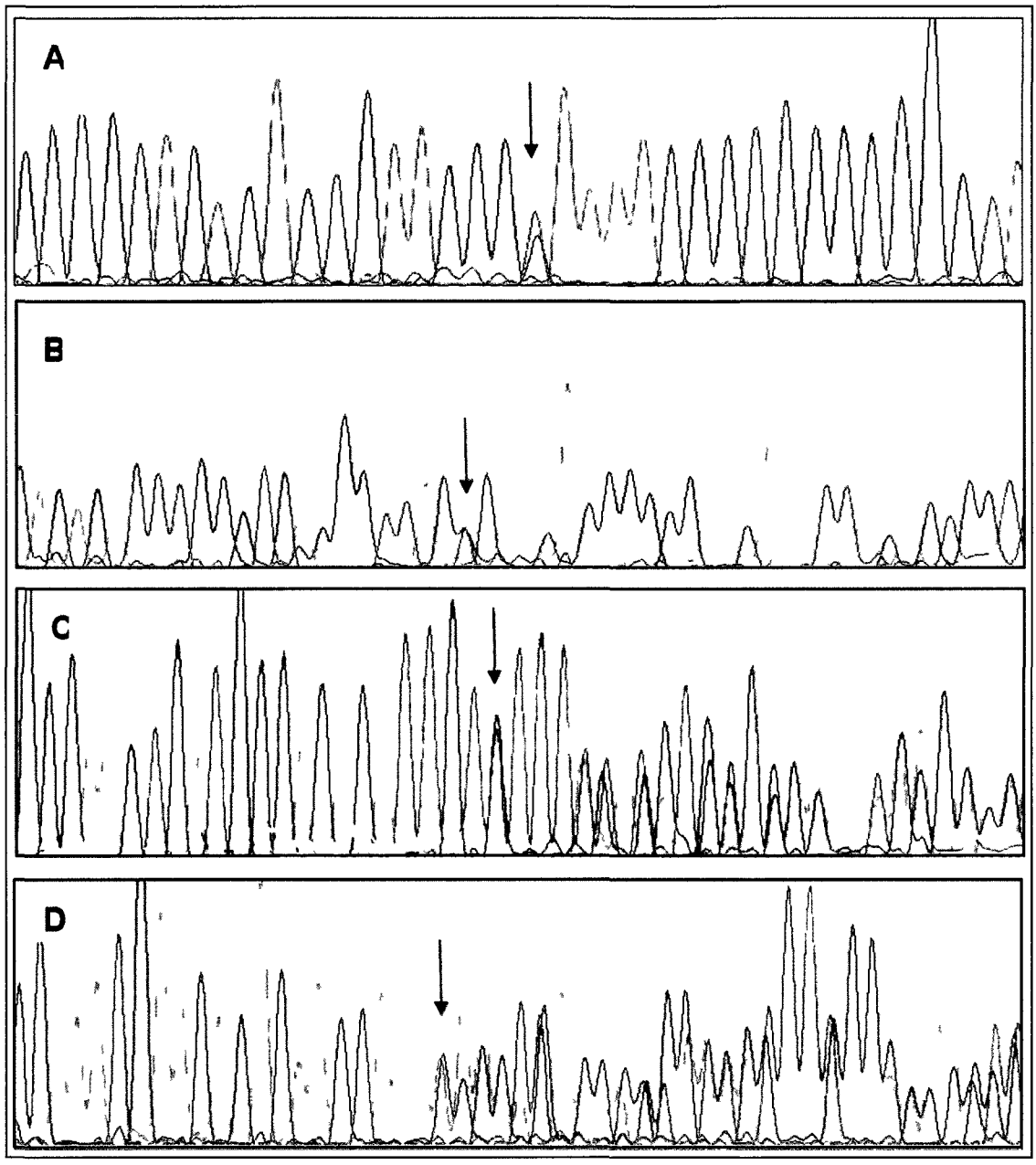



**Figure A10.2: Four electropherograms of *BRCA1* and *BRCA2* Newfoundland mutations.** (A) *BRCA1*, c.546 G>T - transversion (nonsense). (B) *BRCA1* c.3726 C>T - transition (nonsense). (C) *BRCA2* c.6985delCT - frameshift deletion. (D) *BRCA1* c.2190delA – frameshift deletion.

# Appendix 11

## *BRCA1* and *BRCA2* PTT primers and amplification conditions

| Gene | Forward Primer Name | Forward sequence | Reverse Primer Name | Reverse sequence | Size (bp) | Annealing temp. | Extention time | # of cycles |
|---|---|---|---|---|---|---|---|---|
| **BRCA1** | BRCA1ex11ptt792 for | GCTTGTGAATTTTCTGAGACGG | BRCA1ex11ptt4167 rev | CCGTTCCTCTTTCTTCATCATCTG | 3375 | 58 | 4min | 35 |
| **BRCA1 (alt)** | BRCA1ex11ptt792 for | GCTTGTGAATTTTCTGAGACGG | BRCA1ex11ptt2125 rev | TGAGTTGTAGGTTTCTGCTGTGC | 1333 | 54 | 1min30s | 35 |
| | BRCA1ex11ptt1921 for | CACAATTCAAAAGCACCTAAAAA | BRCA1ex11ptt3383 rev | ACCCCTAATCTAAGCATAGATTCA | 1462 | 54 | 1min30s | 35 |
| | BRCA1ex11ptt3061 for | CCACCACTTTTTCCCATCAAGT | BRCA1ex11ptt4183 rev | TATTTTCTTCCAAGCCCGTTCCT | 1122 | 54 | 1min30s | 35 |
| **BRCA2** | BRCA2ex10ptt1061 for | GGCTGCAAAGACCACATTGGAAAG | BRCA2ex10ptt2094 rev | TGCTTCAAACTGGGCTGAACAG | 1033 | 56 | 1min30s | 35 |
| | BRCA2ex11ptt2188 for | GAACCAACTTTGTCCTTAACTAGC | BRCA2ex11ptt4974 rev | GGTCTCACATGCTAATTCAAGGTC | 2786 | 54 | 2min30s | 35 |
| | BRCA2ex11ptt3625 for | CCAAGCTACATATTGCAGAAG | BRCA2ex11ptt7070 rev | CCCACTAAGATAAGGGGCTC | 3445 | 52 | 3min30s | 35 |
| **BRCA2 (alt)** | BRCA2ex11ptt2207 for | GCTAGCTCTTTTGGGACAATTCTGA | BRCA2ex11ptt3511 rev | GCTTGGAAAATAACATCTGAGGGG | 1305 | 56 | 1min30s | 35 |
| | BRCA2ex11ptt3242 for | GTTGGAGGTAGCTTCAGAACAGC | BRCA2ex11ptt4432 rev | TACCATGACATGCTTCTTGAGC | 1191 | 54 | 1min30s | 35 |
| | BRCA2ex11ptt4129 for | ACTACTGGCACTTTTGTTGAAG | BRCA2ex11ptt5237 rev | GCTAAGGCTGAATTTTCAATGACTC | 1109 | | 1min30s | 35 |
| | BRCA2ex11ptt5038 for | CTTGTTTCTATTGAGACTGTGGTG | BRCA2ex11ptt6306 rev | TGTGAGCTGGTCTGAATGTTCG | 1269 | 54 | 1min30s | 35 |
| | BRCA2ex11ptt5911 for | GAGGCATTGGATGATTCAGAGG | BRCA2ex11ptt7067 rev | ACTAAGATAAGGGGCTCTCCTCTT | 1157 | 58 | 1min30s | 35 |

# Appendix 12

## *BRCA1* sequencing and SSCP primer sets

| Exon | Primer Name | Primer | Annealing temp. | Amplicon size |
|---|---|---|---|---|
| 1F | B1.1A-267 M 13F | 5'-TGTAAAACGACGGCCAGTGCAATAAGCCGCAACTGGAA -3' | 60 | 531 |
| 1R | B1.1A+143M13R | 5'-CAGGAAACAGCTATGACCTTCACAACGCCTTACGCCTC -3' | | |
| 2F | B1.2-147M13F | 5'-TGTAAAACGACGGCCAGTAAGGACGTTGTCATTAGTTCTT -3' | TD54 | 334 |
| 2R | B1.2+98M13R | 5'- CAGGAAACAGCTATGACCCATGTCTTTTCTTCCCTAGTATG -3' | | |
| 3F | B1.3-138M13F | 5'-TGTAAAACGACGGCCAGTATATTGAACGAACTTGAGG -3' | 52 | 303 |
| 3R | B1.3+111M13R | 5'-CAGGAAACAGCTATGACCGTTCTCACTTAATTGAAGAAAG -3' | | |
| 5F | B1.5-187M13F | 5'- TGTAAAACGACGGCCAGTAGGTTTTTGCTTATGCAGCA -3' | 58 | 482 |
| 5R | B1.5+217M13R | 5'- CAGGAAACAGCTATGACCGCATTAGAGAAAGGCAGTAAGTTTC -3' | | |
| 6F | B1.6-185M13F | 5'-TGTAAAACGACGGCCAGTTGCCATCACACGGTTTATACA -3' | 52 | 381 |
| 6R | B1.6+107M13R | 5'-CAGGAAACAGCTATGACCCCTAGACTAAAAGGTCTTATCACCA -3' | | |
| 7F | B1.7-101M13F | 5'-TGTAAAACGACGGCCAGTCACAACAAAGAGCATACATAGGG-3' | 58 | 349 |
| 7R | B1.7+108M13R | 5'-CAGGAAACAGCTATGACCGGAGGACTGCTTCTAGCCTG-3' | | |
| 8F | B1.8-94M13F | 5'- TGTAAAACGACGGCCAGTTGTTAGCTGACTGATGATGGT -3' | 58 | 291 |
| 8R | B1.8+45M13R | 5' - CAGGAAACAGCTATGACCTTTGGCAAAACTATAAGATAAGG - 3' | | |
| 9F | B1.9-104M13F | 5'- TGTAAAACGACGGCCAGTTTGTACCTGCCACAGTAGATGCTC - 3' | 56 | 308 |
| 9R | B1.9+84M13R | 5' - CAGGAAACAGCTATGACCCAAACTGCACATACATCCCTGAAC-3' | | |
| 10F | B1.10-89M13F | 5'- TGTAAAACGACGGCCAGTTGGTCAGCTTTCTGTAATCG -3' | 60 | 261 |
| 10R | B1.10+74M13R | 5' - CAGGAAACAGCTATGACCAGGTCCCAAATGGTCTTCAG - 3' | | |
| 11AF | B1.11A-100M13F | 5'-TGTAAAACGACGGCCAGTTAGCCAGTTGGTTGATTTCC -3' | 54 | 401 |
| 11AR | B1.1090M13R | 5'-CAGGAAACAGCTATGACCCTTCCAGCCCATCTGTTATG -3' | | |
| 11BF | B1.844M13F | 5'-TGTAAAACGACGGCCAGTCACAAATACTCATGCCAGCT -3' | 54 | 432 |
| 11BR | B1.1375M13R | 5'-CAGGAAACAGCTATGACCACCTCATTTAGAACGTCCAA -3' | | |
| 11CF | B1.1249M13F | 5'-TGTAAAACGACGGCCAGTGCAGCATTCAGAAAGTTAATGAGTG -3' | 56 | 490 |
| 11CR | B1.1738M13R | 5'-CAGGAAACAGCTATGACCTCCGTTTGGTTAGTTCCCTG -3' | | |
| 11DF | B1.1593M13F | 5'-TGTAAAACGACGGCCAGTATAATACAAGAGCGTCCCCT -3' | 54 | 445 |
| 11DR | B1.2037M13R | 5'-CAGGAAACAGCTATGACCGCAATTCAGTACAATTAGGTGG -3' | | |
| 11EF | B1.1885M13F | 5'-TGTAAAACGACGGCCAGTGCAGTATAAGCAATATGGAACTCG -3' | 54 | 482 |
| 11ER | B1.2368M13R | 5'-CAGGAAACAGCTATGACCATCTTTGGGGTCTTCAGCAT -3' | | |
| 11FF | B1.2232M13F | 5'-TGTAAAACGACGGCCAGTGCACCTGGTTCTTTTACTAAGTGTT -3' | 56 | 515 |
| 11FR | B1.2746M13R | 5'-CAGGAAACAGCTATGACCCCTGGATTTGAAAACGGAGC -3' | | |
| 11GF | B1.2611M13F | 5'-TGTAAAACGACGGCCAGTATCCATTGGGACATGAAGTT -3' | 54 | 467 |
| 11GR | B1.3077M13R | 5'-CAGGAAACAGCTATGACCGATGGGAAAAAGTGGTGGTA -3' | | |
| 11HF | B1.2930M13F | 5'-TGTAAAACGACGGCCAGTGCCAGTTGATAATGCCAAATGTAG -3' | 56 | 507 |
| 11HR | B1.3436M13R | 5'-CAGGAAACAGCTATGACCGGATGCTTACAATTACTTCCAGGA -3' | | |
| 11IF | B1.3318M13F | 5'-TGTAAAACGACGGCCAGTAACATTCAAGCAGAACTAGG -3' | 56 | 472 |
| 11IR | B1.3788M13R | 5'-CAGGAAACAGCTATGACCGAAGCTCTTCATCCTCACTA -3' | | |
| 11JF | B1.3852M13F | 5'-TGTAAAACGACGGCCAGTGCAAAAGCGTCCAGAAAGGA -3' | 56 | 476 |
| 11JR | B1.4127M13R | 5'-CAGGAAACAGCTATGACCACTCAGACCAACTCCCTGGC -3' | | |
| 11KF | B1.3952M13F | 5'-TGTAAAACGACGGCCAGTAGGCATCTCAGGAACATCAC -3' | 54 | 418 |
| 11KR | B1.11K+135M13R | 5'-CAGGAAACAGCTATGACCATGAAAAGCACCTTAGGAGG -3' | | |
| 12F | B1.12-66M13F | 5'- TGTAAAACGACGGCCAGTGTCCTGCCAATGAGAAGAAA -3' | TD60 | 255 |
| 12R | B1.12+68M13R | 5'- CAGGAAACAGCTATGACCTGTCAGCAAACCTAAGAATGT -3' | | |
| 13F | B1.13-92M13F | 5'- TGTAAAACGACGGCCAGTCCATACTAGGTGATTTCAATTCCTG-3' | 58 | 407 |
| 13R | B1.13+78M13R | 5'- CAGGAAACAGCTATGACCTACCCATGTGCTGAGCAAGG-3' | | |

| Exon | Primer Name | Primer Name | Annealing temp. | Amplicon size |
|---|---|---|---|---|
| 14F | B1.14-43M13F | 5'- TGTAAAACGACGGCCAGTCTAACCTGAATTATCACTATCA -3' | 58 | 312 |
| 14R | B1.14+96M13R | 5'- CAGGAAACAGCTATGACCGTGTATAAATGCCTGTATGCA -3' | | |
| 15F | B1.15-54M13F | 5'- TGTAAAACGACGGCCAGTTGGCTGCCCAGCAAGTATG -3' | TD 58 | 338 |
| 15R | B1.15+48M13R | 5'- CAGGAAACAGCTATGACCAACCAGAATATCTTTATGTAGGA -3' | | |
| 16F | B1.16-53M13F | 5'- TGTAAAACGACGGCCAGTAATTCTTAACAGAGACCAGAAC -3' | 52 | 427 |
| 16R | B1.16+39M13R | 5'- CAGGAAACAGCTATGACCAAAACTCTTTCCAGAATGTTGT -3' | | |
| 17F | B1.17-108M13F | 5'-TGTAAAACGACGGCCAGTCGTGCAGGATTGCTACAT -3' | TD54 | 249 |
| 17R | B1.17+53M13R | 5'-CAGGAAACAGCTATGACCCATGTGGTTTTATGCAGCA -3' | | |
| 18F | B1.18-58M13F | 5'- TGTAAAACGACGGCCAGTGGCTCTTTAGCTTCTTAGGAC -3' | 58 | 352 |
| 18R | B1.18+173M13R | 5'- CAGGAAACAGCTATGACCGAGACCCATTTTCCCAGCATC -3' | | |
| 19F | B1.19-87M13F | 5'- TGTAAAACGACGGCCAGTCTGTCATTCTTCCTGTGCTC -3' | TD 56 | 249 |
| 19R | B1.19+61M13R | 5'- CAGGAAACAGCTATGACCCATTGTTAAGGAAAGTGGTGC -3' | | |
| 20F | B1.20-58M13F | 5'- TGTAAAACGACGGCCAGTATATGACGTGTCTGCTCCAC -3' | TD 54 | 380 |
| 20R | B1.20+216M13R | 5'- CAGGAAACAGCTATGACCGGGAATCCAAATTACACAGC -3' | | |
| 21F | B1.21-150M13F | 5'-TGTAAAACGACGGCCAGTGCAGCAGAAATCATCAGGTG -3' | 54 | 416 |
| 21R | B1.21+211M13R | 5'-CAGGAAACAGCTATGACCATGCTCTTGAGAAGGGGGAC -3' | | |
| 22F | B1.22-212M13F | 5'- TGTAAAACGACGGCCAGTgcctatggagactgataacca-3' | 54 to 58 | 378 |
| 22R | B1.22+80M13R | 5'- CAGGAAACAGCTATGACCatagggactgacaggtgcca-3' | | |
| 22F | B1.22-120M13F | 5'- TGTAAAACGACGGCCAGTAGACAGACTCTCCCATTGAG-3' | 60 | 359 |
| 22R | B1.22+185M13R | 5'- CAGGAAACAGCTATGACCAATCCTTACCCATCCCTTAC-3' | | |
| 23F | B1.23-87M13F | 5'- TGTAAAACGACGGCCAGTCAGAGCAAGACCCTGTCTC -3' | 50 | 235 |
| 23R | B1.23+69M13R | 5'- CAGGAAACAGCTATGACCACTGTGCTACTCAAGCACCA -3' | | |
| 24F | B1.24-50M13F | 5'- TGTAAAACGACGGCCAGTATGAATTGACACTAATCTCTGC -3' | 58 | 279 |
| 24R | B1.24+92M13R | 5'- CAGGAAACAGCTATGACCGTAGCCAGGACAGTAGAAGGA -3' | | |

# Appendix 13

## *BRCA2* sequencing and SSCP primer sets

| Exon | Primer Name | Primer | Annealing temp | Amplicon size |
|---|---|---|---|---|
| 2F | B2 2-120M13F | 5 -TGTAAAACGACGGCCAGTTTCCAGGAGATGGGACTGAATTAG -3 | TD58 | 321 |
| 2R | B2 2+94M13R | 5'- CAGGAAACAGCTATGACCAAGCAACACTGTGACGTACTGGG -3' | | |
| 3F | B2 3-125M13F | 5'-TGTAAAACGACGGCCAGTGCCTTAACAAAAGTAATCCATAGTC -3' | TD56 | 489 |
| 3R | B2 3+109M13R | 5 -CAGGAAACAGCTATGACCGAGAGACTGATTTGCCCAG -3' | | |
| 4F | B2 4-111M13F | 5'- TGTAAAACGACGGCCAGTTCTCATTCCCAGTATAGAGGAGAC -3' | TD56 | 433 |
| 4R | B2 4+223M13R | 5 - CAGGAAACAGCTATGACCCAGCCAATTCAACATCACAAG -3' | | |
| 5 6F | B2 5 6-234M13F | 5 -TGTAAAACGACGGCCAGTTACACGGTTTCCAGCAGC -3' | TD56 | 520 |
| 5 6R | B2 5 6+103M13R | 5'-CAGGAAACAGCTATGACCGCAGAATGCTAGGTACAGATTTGTA -3' | | |
| 7F | B2 7-131M13F | 5 -TGTAAAACGACGGCCAGTTCTGTACCTAGCATTCTGCC -3' | TD56 | 449 |
| 7R | B2 7+203M13F | 5'-CAGGAAACAGCTATGACCGACACCACTGGACTACCACT -3' | | |
| 8F | B2 8-267M13F | 5 -TGTAAAACGACGGCCAGTCCAACTCATTGTGGACAGTA -3' | TD58 | 452 |
| 8R | B2.8+135M13R | 5'-CAGGAAACAGCTATGACCCCAGGTTTAGAGACTTTCTCA -3' | | |
| 9F | B2 9-223M13F | 5'- TGTAAAACGACGGCCAGTCCACATCCTAGTGGTGCAAGATTTC -3' | TD58 | 441 |
| 9R | B2 9+106M13R | 5'- CAGGAAACAGCTATGACCCGGTAAACTGAGATCACGGGTG -3' | | |
| 10AF | B2 10-194M13F | 5 -TGTAAAACGACGGCCAGTGCTTTACTAGAAGAACAGGAGAAGG -3' | TD54 | 496 |
| 10AR | B2 1322M13R | 5 -CAGGAAACAGCTATGACCGGATCAGTATCATTTGGTTCCAC -3' | | |
| 10BF | B2 1129M13F | 5'-TGTAAAACGACGGCCAGTGATACCTCTGAAGAAGATAGTT -3' | TD54 | 533 |
| 10BR | B2 1661M13R | 5'-CAGGAAACAGCTATGACCGTATGAGATTCAAGATGCTG -3' | | |
| 10CF | B2 1470M13F | 5'-TGTAAAACGACGGCCAGTGCATATTTCTTCATGTGACCA -3 | TD56 | 449 |
| 10CR | B2 1918M13R | 5'-CAGGAAACAGCTATGACCGCCAGCTTCCATTATCAATTA -3' | | |
| 10DF | B2 1766M13F | 5'-TGTAAAACGACGGCCAGTAAGAGACTTTCAATGCAAGT -3 | TD56 | 548 |
| 10DR | B2 +175M13R | 5'-CAGGAAACAGCTATGACCTTTGAGTGACCTGATTCTAA -3' | | |
| 11AF | B2 11A-95M13F | 5'- TGTAAAACGACGGCCAGTTGTGCCCAAACACTACCTTT -3' | TD54 | 510 |
| 11AR | B2 2552M13R | 5'- CAGGAAACAGCTATGACCGAAGTAGGAGTTAAAATAAGAGTGC -3' | | |
| 11BF | B2 2371M13F | 5'- TGTAAAACGACGGCCAGTGGACAGTGTGAAAATGATCC -3' | TD54 | 510 |
| 11BR | B2 2880M13R | 5'- CAGGAAACAGCTATGACCTGAGAAAAGTTCTTCAGAGTCTG -3' | | |
| 11CF | B2 2720M13F | 5'- TGTAAAACGACGGCCAGTTTGAGCTGTTGCCACCTGA -3' | TD54 | 490 |
| 11CR | B2 3209M13R | 5'- CAGGAAACAGCTATGACCGCCCATTTGTTCATGTAATCAT -3' | | |
| 11DF | B2 3009M13F | 5'-TGTAAAACGACGGCCAGTGGTTTTATATGGAGACACAGG -3' | TD54 | 500 |
| 11DR | B2 3508M13R | 5'-CAGGAAACAGCTATGACCTGGAAAATAACATCTGAGGG -3' | | |
| 11EF | B2 3394M13F | 5'- TGTAAAACGACGGCCAGTCAAAAGAAACTGAGCAAGCC -3' | TD56 | 511 |
| 11ER | B2 3904M13R | 5'- CAGGAAACAGCTATGACCTTGTGCCATGAGCAGAATAA -3' | | |
| 11FF | B2 3790M13F | 5'- TGTAAAACGACGGCCAGTATTAAACGGAAGTTTGCTGG -3' | 54 | 457 |
| 11FR | B2 4247M13R | 5'- CAGGAAACAGCTATGACCCACTGCCATCAAATTCTAAG -3' | | |
| 11GF | B2 3985M13F | 5'-TGTAAAACGACGGCCAGTGCAGAGGTACATCCAATAAG -3' | TD54 | 569 |
| 11GR | B2 4553M13R | 5'-CAGGAAACAGCTATGACCGACTCTTTGGCGACACTAAT -3' | | |
| 11HF | B2 4411M13F | 5'- TGTAAAACGACGGCCAGTGCTCAAGAAGCATGTCATGG -3 | TD54 | 499 |
| 11HR | B2 4908M13R | 5'- CAGGAAACAGCTATGACCGGCTAAAACTGGTGATTTCACT -3' | | |
| 11IF | B2 4791M13F | 5'- TGTAAAACGACGGCCAGTGTTGGGTTTTCATACAGCTA -3' | TD56 | 489 |
| 11IR | B2 5279M13R | 5'- CAGGAAACAGCTATGACCGTCTGACTCACAGAAGTTTTTCTAC -3' | | |
| 11JF | B2 5168M13F | 5'- TGTAAAACGACGGCCAGTCAGCAAAAAGTCCTGCAACT -3' | TD54 | 496 |
| 11JR | B2 5668M13R | 5'- CAGGAAACAGCTATGACCTCCTCAACGCAAATATCTTC -3' | | |
| 11KF | B2 5544M13F | 5'-TGTAAAACGACGGCCAGTTGAGCCAGTATTGAAGAATGTTG -3' | 56 | 505 |
| 11KR | B2 6048M13R | 5'- CAGGAAACAGCTATGACCCTCCAATCCAGACATATTTTGG -3' | | |

| Exon | Primer Name | Primer | Annealing temp | Amplicon size |
|---|---|---|---|---|
| 11LF | B2 5888M13F | 5 - TGTAAAACGACGGCCAGTCGAAAATTATGGCAGGTTGT -3' | TD54 | 522 |
| 11LR | B2 6409M13R | 5 - CAGGAAACAGCTATGACCCTGTACTAAATCCAGAGAAAGCAGA -3' | | |
| 11MF | B2 6213M13F | 5'- TGTAAAACGACGGCCAGTCGCAAGACAAGTGTTTTCTG -3' | TD54 | 383 |
| 11MR | B2 6595M13R | 5 - CAGGAAACAGCTATGACCCCATTTCTGAGTTTACACAGTG -3 | | |
| 11NF | B2 6459M13F | 5'- TGTAAAACGACGGCCAGTGGGAGTGTTAGAGGAATTTG -3' | TD56 | 506 |
| 11NR | B2 6964M13R | 5 - CAGGAAACAGCTATGACCGCAGTTTAGAATCTGTCAGTTC -3 | | |
| 11OF | B2 11 6769M13F | 5 -TGTAAAACGACGGCCAGTGGAAAAGAACAGGCTTCACCTA -3 | TD56 | 449 |
| 11OR | B2 11+1483M13R | 5 -CAGGAAACAGCTATGACCATCAAACCATACTCCCCCAA -3' | | |
| 12F | B2 12-229M13F | 5'-TGTAAAACGACGGCCAGTGACCCTGTTAAATAGTGGTGT -3' | TD58 | 456 |
| 12R | B2 12+132M13R | 5'-CAGGAAACAGCTATGACCTCCTTGATTAGGCACAGTG -3' | | |
| 13F | B2 13-130M13F | 5'- TGTAAAACGACGGCCAGTTATTGAGCATCCGTTACATTCACTG -3' | TD56 | 470 |
| 14R | | 5'- CAGGAAACAGCTATGACCACCACCAAAGGGGGAAAACC -3' | | |
| 14AF | B2 14-192M13F | 5'-TGTAAAACGACGGCCAGTGAGGTAGATTGTAAAGTCAAAGGCT -3' | TD56 | 506 |
| 14AR | B2 7549M13R | 5'-CAGGAAACAGCTATGACCCATGTCCATCAATGTTTTGC -3 | | |
| 14BF | B2 7409M13F | 5'-TGTAAAACGACGGCCAGTGACACTTGATTACTACAGGC -3 | TD54 | 490 |
| 14BR | B2 14+236M13R | 5 -CAGGAAACAGCTATGACCGTAACAGCAGTCCTAGATTG -3' | | |
| 15F | B2 15-173M13F | 5'- TGTAAAACGACGGCCAGTGAGACAGGGTTTCTCCATTTTGG -3' | TD58 | 473 |
| 15R | B2 15+119M13R | 5'- CAGGAAACAGCTATGACCCATCCATTCCTGCACTAATGTGTTC -3' | | |
| 16F | B2 16-206M13F | 5'- TGTAAAACGACGGCCAGTCAGCAGACTGTGGAATGTATGGATC -3 | TD56 | 583 |
| 17R | | 5'- CAGGAAACAGCTATGACCTAGCAATCCCCAAATAGTG -3' | | |
| 17F | B2 17-207M13F | 5'-TGTAAAACGACGGCCAGTCACCATGCTCAGCAATGAAGT -3' | TD58 | 500 |
| 18R | | 5'- CAGGAAACAGCTATGACCAAGCTCAAGAAAGATCTCTGGAC -3 | | |
| 18AF | B2 18-232M13F | 5'-TGTAAAACGACGGCCAGTATCCACTATTTGGGGATTGC -3' | TD56 | 464 |
| 18AR | B2 8436M13R | 5'-CAGGAAACAGCTATGACCGAGGGGAGGATCTAACTGGG -3' | | |
| 18BF | B2 8221M13F | 5'-TGTAAAACGACGGCCAGTGATAGAAGCAGAAGATCGGC -3' | TD54 | 453 |
| 18BR | B2 18+114M13R | 5 -CAGGAAACAGCTATGACCCAGTACATCTAAGAAATTGAGCATC-3' | | |
| 19F | B2 19-144M13F | 5'-TGTAAAACGACGGCCAGTTTTACTGTCTTACTAATCTTCC -3' | TD 56 | 708 |
| 19R | B2 19+407M13R | 5'-CAGGAAACAGCTATGACCTCTCCATCCACTGTAATAAT -3' | | |
| 20F | B2 20-66M13F | 5'- TGTAAAACGACGGCCAGTGCCTGATACAATTAACTTGAATG -3' | TD48 | 365 |
| 21R | | 5 - CAGGAAACAGCTATGACCGCAGCATTTCAACATACTCCTTCC -3 | | |
| 21F | B2 21-82M13F | 5'-TGTAAAACGACGGCCAGTGGTGTTTTATGCTTGGTTC -3' | TD54 | 406 |
| 21R | B2 21+203M13R | 5'-CAGGAAACAGCTATGACCCTCCCATCCTCTACTCGGT -3' | | |
| 22F | B2 22-205M13F | 5'- TGTAAAACGACGGCCAGTGAGGATCATTTTTGCCGTAGGG -3' | TD58 | 539 |
| 22R | | 5'- CAGGAAACAGCTATGACCGCCAACTGGTAGCTCCAACTAATC -3' | | |
| 23F | B2 23-132M13F | 5'-TGTAAAACGACGGCCAGTATCAGAGAAGCAAAATCCAC -3' | TD54 | 507 |
| 23R | B2 23+211M13R | 5'-CAGGAAACAGCTATGACCCGACAAATCCTATTAGGTCC -3' | | |
| 24F | B2 24-139M13F | 5'-TGTAAAACGACGGCCAGTCATACAGTTAGCAGCGACAA -3' | TD54 | 419 |
| 24R | B2 24+141M13R | 5'-CAGGAAACAGCTATGACCGAGGTTCAAAGAGGCTTACTT -3' | | |
| 25F | B2 25-184M13F | 5'- TGTAAAACGACGGCCAGTGTCATTTTGGAAAACCTGAGCTTTC -3' | TD58 | 572 |
| 25R | B2 25+163M13R | 5'- CAGGAAACAGCTATGACCCATTCCCCCATCTCCTGAGG -3' | | |
| 26F | B2 26-206M13F | 5'- TGTAAAACGACGGCCAGTCTCTCCCTATCAGCTAGATTCCCC -3' | TD58 | 491 |
| 26R | B2 26+138M13R | 5'- CAGGAAACAGCTATGACCCTTACAGGAGCCACATAACAACCAC -3' | | |
| 27AF | B2 27A-135M13F | 5'- TGTAAAACGACGGCCAGTacctattaggagttaggggagg -3' | TD58 | 661 |
| 27AR | B2 27A 7812 M13R | 5'- CAGGAAACAGCTATGACCgaaccagacaaaagagcttggg -3' | | |
| 27BF | B2 27B-252M13F | 5'-TGTAAAACGACGGCCAGTCCAAGGAGTTGTGGCACCAA -3' | TD 56 | 530 |
| 27BR | B2 27B+178M13R | 5'-CAGGAAACAGCTATGACCACGGAATCGGGCAAAACGAT -3' | | |

# Appendix 14

## *BRCA2* locus markers

| Marker | Label Dye | Forward/Reverse Primer | Primer Sequence | Genomic Location (Mb) | Band | Tandem Reapeat | Heterozygosity | Allele size range |
|---|---|---|---|---|---|---|---|---|
| D13S1242 | fam | F | GTGCCCAGCCAGATTC | 28 | 13q12.3 | di | na | 198-248 bps |
| | | R | GCCCCCAGTCAGGTTT | | | | | |
| D13S1299 | fam | F | CGATGAGTGACAGGGCT | 29.5 | 13q12.3 | di | na | 199-229 bps |
| | | R | CCTCGTGGGTGGAATAA | | | | | |
| D13S1287 | fam | F | TATGCCAGTATGCCTGCT | 30.6 | 13q12.3 | di | na | 226-238 bps |
| | | R | GTCACATCAGTCCATTTGC | | | | | |
| D13S260 | fam | F | AGATATTGTCTCCGTTCCATGA | 31.3 | 13q13.1 | di | na | 158-173 bps |
| | | R | CCCAGATATAAGGACCTGGCTA | | | | | |
| D13S171 | fam | F | CCTACCATTGACACTCTCAG | 32.2 | 13q13.1 | di | na | 227-241 bps |
| | | R | TAGGGCCATCCATTCT | | | | | |
| D13S220 | fam | F | CCAACATCGGGAACTG | 34.1 | 13q13.2 | di | na | 191-203 bps |
| | | R | TGCATTCTTTAAGTCCATGTC | | | | | |

# Appendix 15

## PCR-based applications for the amplification across the breakpoints of six *BRCA1* genomic rearrangements

| Mutation | Mutation detail | Forward Primer | Reverse Primer | Amplicon size (bp) | | PCR conditions | | |
|---|---|---|---|---|---|---|---|---|
| | | | | WT | Mutant | TD | Betaine | Cycles |
| Del Exon 3 | 1,039 deletion | TTTTTCTCCCCCCCTACCCTG | GCTCAGCATTTGTTACTCAAGCTG | 1,543 | 514 | 60 | Yes | 45 |
| Duplication Exon 13 | 6,018 bp duplication | ATTATTTCCCCCCAGGCTACCCAG | GGTCCATTTCAAAGAAGAGTGTGC | none | 994 | 60 | Yes | 45 |
| Deletion Exon 14-20 | 26,456 bp deletion | GCTCAAAATGAATTACAGT | CTTTTGTCACCAATCAG | 27,180 | 725 | 55 | Yes | 45 |
| Deletion Exon 20 | 3,950 bp deletion | AGATGGAGTCTCGCTCTGTTACCC | TGTATGAATAAACCGCACCCCC | 4,318 | 369 | 60 | Yes | 45 |
| Deletion Exon 20-22 | 11,561 bp deletion | AATGATAAAAGGTTCGGGCACAGTG | GTCTCAAACTCCTGGGCTCAACTTC | 12,513 | 954 | 60 | No | 45 |
| Deletion Exon 22 | 510 bp deletion | TCCCATTGAGAGGTCTTGCT | ACTGTGCTACTCAAGCACCA | 1,750 | 1,241 | 60 | Yes | 45 |

# Appendix 16

## Primers for nested *CHK2* PCR

| Primers | Forward | Reverse | TD | Amplicon |
|---|---|---|---|---|
| External | GTCTAAATGGACGAGTGGGCTG | CAACAGAAACAAGAACTTCAGGCG | 60 | 905bp |
| Internal | TGATCTAGCCTACGTGTCTTCTTGG | AAATCTTGGAGTGCCCAAAACCA | 50 | 124bp |