POPULATION-BASED STUDY OF LYNCH SYNDROME IN NEWFOUNDLAND

PHILLIP A. WILLIAMS

POPULATION-BASED STUDY OF LYNCH SYNDROME IN NEWFOUNDLAND

by

Phillip A. Williams

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Science

Discipline of Genetics/ Faculty of Medicine

Memorial University of Newfoundland

May, 2008

St. John's Newfoundland

Abstract:

Lynch Syndrome is an autosomal dominant condition causing predisposition to various cancers, primarily colorectal cancer (CRC). This syndrome is caused by mutations in DNA mismatch repair (MMR) genes. MMR mutations have been previously identified in the Newfoundland population and the province of Newfoundland and Labrador has one of the highest age standardised rates of CRC in Canada. The Newfoundland Colorectal Cancer Registry (NFCCR) is a population-based registry of CRC cases in Newfoundland from 1999-2003. Patients from the NFCCR were screened for MMR mutations. 13 pathogenic mutations were identified in 740 cases in the NFCCR. This corresponds to an incidence of Lynch Syndrome of 1.8% of CRC cases in the NFCCR. In addition, all published literature concerning Lynch Syndrome was reviewed to construct and maintain a web-based public catalogue of MMR mutations as a resource for determining the pathogenicity of any variants identified in this and future works.

Acknowledgements:

I would like to thank first and foremost Ban Younghusband. I could not have asked for a more knowledgeable and considerate individual for my supervisor and to guide me through this process. He is a wonderful mentor, and I will miss his daily advice. I would also like to thank Mike Woods, without his support this thesis would not be what it is today. I would also like to thank my friends and colleagues in the lab, and the whole Discipline of Genetics. I have many good memories, both work related and not. Finally, I would like to thank my family and friends, for their never ending support for the completion of my degree.

Table of Contents:

Title ..... -i-

Abstract: ..... -ii-

Acknowledgements: ..... -iii-

Table of Contents: ..... -iv-

List of Tables: ..... -vi-

List of Figures: ..... -vi-

List of Abbreviations: ..... -vii-

Chapter 1 ..... -1-
1.1 Introduction: ..... -1-
1.1.1 Colorectal Cancer: ..... -1-
1.1.2 Hereditary Cancer Syndromes: ..... -1-
1.1.2.1 Familial Adenomatous Polyposis ..... -1-
1.1.2.2 MYH-associated polyposis ..... -2-
1.1.2.3 Lynch Syndrome ..... -4-
1.1.2.4 Familial Colorectal Cancer Type X ..... -5-
1.1.2.5 Other Colorectal Cancer Syndromes ..... -5-
1.1.3 History of Lynch Syndrome: ..... -6-
1.1.4 Lynch Syndrome Risk Criteria ..... -8-
1.1.4.1 Amsterdam Criteria ..... -8-
1.1.4.2 Bethesda Criteria: ..... -9-
1.1.4.3 Age and Cancer Modified Amsterdam Criteria: ..... -9-
1.1.5 Mismatch Repair Pathway ..... -10-
1.1.6 Mismatch Repair Genes ..... -11-
1.1.7 Molecular Characteristics of Lynch Syndrome: ..... -12-
1.1.7.1 Microsatellite Instability: ..... -12-
1.1.7.2 Immunohistochemistry ..... -13-
1.2 Objectives: ..... -15-

Chapter 2 ..... -16-
2.0 Population Based Study of Lynch Syndrome in Newfoundland ..... -16-
2.1 Introduction: ..... -16-
2.2 Materials and Methods: ..... -19-
2.2.1 Subjects: ..... -19-

2.2.2 Screening Criteria: ..... -19-
2.2.3 Multiplex Ligation Dependent Probe Amplification: ..... -20-
2.2.4 MLPA Reaction Conditions: ..... -21-
2.2.5 MLPA Fragment Analysis: ..... -23-
2.2.6 Dosage Analysis: ..... -24-
2.2.7 Direct Sequencing for Mutation Detection: ..... -25-
2.2.7.1 PCR DNA Amplification: ..... -25-
2.2.7.2 PCR Product Purification: ..... -26-
2.2.7.3 Sequencing Reaction: ..... -27-
2.2.7.4 Sequencing Reaction Purification: ..... -27-
2.2.7.5 Sequence Analysis: ..... -28-
2.3 Results: ..... -29-
2.3.1 MSH2 ..... -29-
2.3.2 MLH1 ..... -30-
2.4 Discussion: ..... -35-
2.4.1 Mutation or Polymorphism? ..... -40-

Chapter 3 ..... -42-
3.0 Mismatch Repair Variant Database ..... -42-
3.1 Introduction: ..... -42-
3.2 Materials and Methods: ..... -44-
3.3 Results: ..... -46-
3.3.1 Exonic Distribution of Unique Variants by Gene: ..... -48-
3.3.2 Distribution of Unique Variants by Type: ..... -49-
3.4 Discussion ..... -60-

Chapter 4 ..... -63-
4.1 Conclusions ..... -63-

References: ..... -64-

Appendix 1: Bethesda and Revised Bethesda Criteria ..... -69-
Bethesda Criteria (Rodriguez-Bigas *et al.*, 1997): ..... -69-
Revised Bethesda Guidelines (Umar *et al.*, 2004) ..... -70-

Appendix 2: Sequencing Primers ..... -71-
MLH1 Primers: ..... -71-
MSH2 Primers: ..... -73-

List of Tables:

Table 2.1: Mutations and Polymorphisms found in MSH2 32
Table 2.2: Polymorphisms found in MLH1 33
Table 3.1: Alternative MMR gene names 44

List of Figures:

Figure 1.1: GO repair system in prokaryotes 3
Figure 2.1: MLPA Probe Hybridisation and Ligation 22
Figure 2.2: Mutation Analysis Flow Chart 31
Figure 2.3: Molecular Testing Breakdown by IHC and MSI Status 34
Figure 3.1: Screenshot of the Mismatch Repair Genes Variant Database Website 47
Figure 3.2: Frequency of Germline MLH1 Variants Reported in the Literature 51
Figure 3.3: Distribution of Unique MLH1 Variants per 100bp 52
Figure 3.4: Frequency of Germline MSH2 Variants Reported in the Literature 53
Figure 3.5: Distribution of Unique MSH2 Variants per 100bp 54
Figure 3.6: Frequency of Germline MSH6 Variants Reported in the Literature 55
Figure 3.7: Distribution of Unique MSH6 Variants per 100bp 56
Figure 3.8: Unique Germline MLH1 Variants Breakdown by Type 57
Figure 3.9: Unique Germline MSH2 Variants Breakdown by Type 58
Figure 3.10: Unique Germline MSH6 Variants Breakdown by Type 59

List of Abbreviations:

8-OxoG:	8-hydroxyguanine
ACI:	Amsterdam Criteria I
ACII:	Amsterdam Criteria II
ACMAC:	Age and Cancer Modified Amsterdam Criteria
AFAP:	Attenuated Familial Adenomatous Polyposis
APC:	*Adenomatous Polyposis Coli*
BC:	Bethesda Criteria
ExoI:	Exo Nuclease I
DHPLC:	Denaturing High Performance Liquid Chromatograpy
dNTP:	Deoxynucleotide Triphosphate
FAP:	Familial Adenomatous Polyposis
LS:	Lynch Syndrome
HNPCC:	Hereditary Non-Polyposis Colorectal Cancer
HuGE:	Human Genome Epidemiology
ICG-HNPCC:	International Collaborative Group on HNPCC
IHC:	Immunohistochemistry
InSiGHT:	International Society for Hereditary Gastrointestinal Tumours
INTI:	Intermediate I
MAP:	*MYH*-associated polyposis
MLPA:	Multiplex Ligation-dependant Probe Amplification
MMR:	Mismatch Repair
MLH1:	*MutL Homologue 1*
MRC-Holland:	Microbiology Research Centre Holland
MSH2:	*MutS Homologue 2*
MSH6:	*MutS Homologue* 6
MSI:	Microsatellite Instability
MSS:	Microsatellite Stable
MYH:	*MutY Homologue*
NF1:	Neurofibromatosis Type 1
NFCCR:	Newfoundland Colorectal Cancer Registry
PCR:	Polymerase Chain Reaction
PMS2:	*Postmeiotic Segregation increased 2*
RFU:	Relative Fluorescence Units
SAP:	Shrimp Alkaline Phosphatase
SLS:	Sample Loading Solution
UVs:	Unclassified Variants

Chapter 1

1.1 Introduction:

1.1.1 Colorectal Cancer:

Colorectal cancer (CRC) is the third most common cancer in Canada with an estimated incidence of 20,800 cases (13.8% of all cancer cases) for 2007 (Cancer Canada). It is the leading cause of cancer death in non-smokers and is second in lethality only to lung cancer, with an estimated 8,500 deaths (12.2% of all cancer deaths) for 2007 (Cancer Canada). In the Canadian population, the overall lifetime risk of developing CRC is approximately 6% (Cancer Canada). It is estimated that upwards of 20% of all CRCs have a familial component without a clear genetic cause, while around 2-6% are thought to be hereditary (Kemp *et al.*, 2004). The most common syndromes known to cause hereditary CRC are: Familial Adenomatous Polyposis (FAP), Lynch Syndrome (LS) or Hereditary Non-Polyposis Colorectal Cancer (HNPCC) and *MYH*-associated polyposis (MAP); a fourth, Familial Colorectal Cancer Type X, has recently been suggested (Lindor *et al.*, 2005).

1.1.2 Hereditary Cancer Syndromes:

1.1.2.1 Familial Adenomatous Polyposis:

Familial Adenomatous Polyposis (FAP) is inherited in an autosomal dominant pattern. The signature clinical feature of FAP is hundreds or thousands of polyps developing in the late teens or early twenties. Developing this large number of polyps

increases an individual's lifetime risk of CRC to a near certainty. FAP is a very rare disease which accounts for approximately 1% of all CRC, but has a penetrance of nearly 100% (Bisgaard *et al.*, 1994). Germline mutations in the tumour suppressor gene *adenomatous polyposis coli* (*APC*) account for FAP. Depending on the location of the germline mutation in *APC*, another phenotype of Attenuated Familial Adenomatous Polyposis (AFAP) may occur. AFAP is a less severe version of FAP, with fewer polyps (10-100s) and a later age of CRC onset. Mutations causing AFAP are generally localised to the 5', 3' or exon 9 regions of *APC*. They do not down regulate the APC protein to the same degree as FAP mutations hence the attenuated phenotype (Sieber *et al.*, 2006).

1.1.2.2 MYH-associated polyposis:

MYH-associated polyposis (MAP) presents with tens to hundreds of polyps and is inherited in an autosomal recessive pattern (Al-Tassan *et al.*, 2002). MAP accounts for less than 1% of all CRC. A diagnosis of MAP must first exclude FAP, that is, there must be a negative *APC* mutation result, because of the clinical similarity of a large number of polyps. Some MAP CRC patients have been reported with few or no polyps, which can overlap with a Lynch Syndrome phenotype (Wang *et al.*, 2004). MAP is caused by mutations in the *MYH (MutY Homologue)* gene. Oxidative damage to DNA can cause the formation of 8-hydroxyguanine (8-OxoG) from guanine, which pairs with adenine instead of cytosine. Subsequent DNA replication results in the replacement of 8-OxoG with thymine (Figure 1.1). The MYH protein forms a trimer with two additional proteins to form a complex which is responsible for the removal of adenine that is paired with

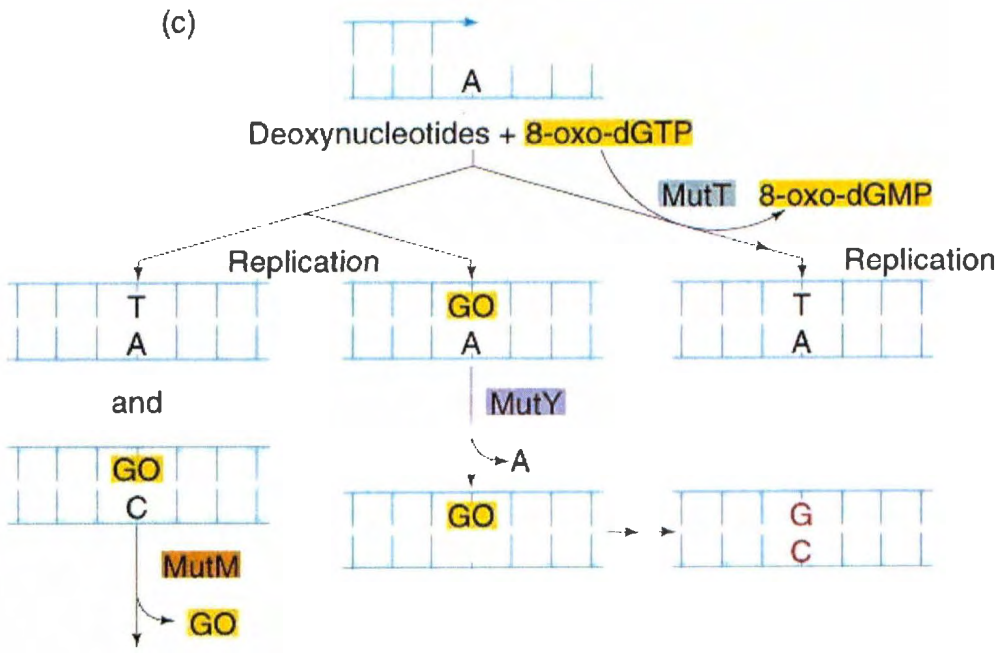

Figure 1.1: GO repair system in prokaryotes: Graphical representation of the 8-hydroxyguanine (GO) repair system in prokaryotes. These three proteins prevent 8-hydroxyguanine from integrating into DNA post-replication. In normal function, MutY removes adenine nucleotides mispaired with 8-hydroxyguanine, MutT removes a phosphate group from 8-Oxo-dGTP, preventing it from being used in DNA synthesis, while MutM excises 8-OxoG from the synthesised DNA strand.

(www.mun.ca/biochem/courses/4103/figures/Griffiths/G19-36c.jpg)

8-OxoG. Mutations in MYH disrupt this removal resulting in the pairing of 8-OxoG with A, and in an accumulation of mutations due to G:T transversions which, if they occur in critical regions of particular genes, may result in carcinoma.

1.1.2.3 Lynch Syndrome:

Lynch Syndrome (LS) or Hereditary Non-Polyposis Colorectal Cancer (HNPCC) is another autosomal dominant condition that increases an individual's lifetime risk of developing CRC, as well as a range of extracolonic cancers. Lynch Syndrome is the most common inherited colon cancer syndrome and is thought to account for anywhere from 2-5% of all CRC. LS patients have been reported to have a lifetime risk of 78% for developing CRC. This is likely to be a very rough estimate as different mutations have varying penetrance and may cause different cancers, depending which gene is mutated. It has been noted previously that mutations is *MLH1* result in CRC, while mutations in *MSH2* are associated with more extra-colonic tumours (Vasen *et al.*, 2001, Bandipalliam *et al.*, 2004). Colon and rectal cancers are the primary cancers in this syndrome but there is also increased risk of developing many extra-colonic tumours including endometrial uterine, gastric, small bowel, pancreas, hepato-biliary, ovarian, kidney, ureter, brain and lymphoma (Lynch *et al.*, 1993). Particularly significant in females is the 43% lifetime risk of developing endometrial cancer. Overall, including colorectal, endometrial and the other extracolonic cancers, there is a cumulative 90% lifetime risk of cancer (Aarnio *et al.*, 1995). Mutations in any one of at least four mismatch repair (MMR) genes are responsible for this syndrome: *MLH1* (*MutL Homologue 1*), *MSH2*

(*MutS Homologue 2*), *MSH6* (*MutS Homologue 6*) and *PMS2* (*Postmeiotic Segregation increased 2*).

1.1.2.4 Familial Colorectal Cancer Type X:

Familial Colorectal Cancer Type X has been used to describe some families who match the most stringent familial criteria for Lynch Syndrome, but lack the signs of a MMR defect such as microsatellite instability (see below) or a MMR gene mutation. These apparent LS families are also noted to have a lower incidence of cancer than LS families, fewer extra-colonic tumours and a later age of onset (Lindor *et al.*, 2005; Woods *et al.*, 2006).

1.1.2.5 Other Colorectal Cancer Syndromes:

Two other rare cancer syndromes, Muir-Torre and Turcot, can also be caused by mutations in the MMR genes. Muir-Torre Syndrome is an autosomal dominant condition which was reported independently by Muir *et al.* in 1967 and Torre in 1968. The syndrome presents with the same cancers as Lynch Syndrome but with additional sebaceous cancers of the skin. Muir-Torre Syndrome is caused by mutations in the MMR genes *MSH2* or *MLH1*. It is usually classified as a sub-type of Lynch Syndrome instead of a separate disease. 'Turcot Syndrome' was first described by Turcot *et al.* in 1959 and is not really a syndrome, but the combination of polyposis or colon cancer and brain tumour. This combination can occur due to mutations in *APC*, *MLH1* or *MSH2*, or homozygous or compound heterozygous mutations in *PMS2* (De Vos *et al.*, 2004).

'Turcot Syndrome' is generally characterised by a milder polyposis than FAP, but with cancer of the central nervous system in addition to CRC. This can occur in one person in an FAP or Lynch Syndrome family.

Another genetic malignancy syndrome that can be caused by MMR mutations is Neurofibromatosis Type 1 (NF1). NF1 is an autosomal dominant condition usually caused by inheriting a mutation in the *NF1* gene and is generally characterised by multiple café au lait spots and fibromatous tumours of the skin, with a wide range of other manifestations (OMIM 162200). It has been shown that homozygous *MLH1* mutations (Ricciardone *et al.*, 1999; Wang *et al.*, 1999) and homozygous *MSH2* mutations (Whiteside *et al.*, 2002) can both cause *de novo* cases of NF1 in children. One study has shown that patients who are MMR deficient due to homozygous *MLH1* mutations have somatic mutations the in *NF1* gene causing an NF1-like phenotype (Wang *et al.*, 2003).

1.1.3 History of Lynch Syndrome:

In 1913, Warthin published a paper on several families he described as "cancer fraternities" which he had followed since 1895. In his paper, he described a Cancer Family "G" which had a high incidence of carcinoma of the colon, endometrium and stomach (Warthin *et al.*, 1913). Warthin revisited "Family G" in 1925 and gave an update on the cancer incidence in an expanded family tree. He stated that the familial susceptibility for gastro-intestinal cancer in males and the "generative organs" in females was inherited in a recessive pattern (Warthin *et al.*, 1925). In 1936 "Family G" was

investigated again by Hauser and Weller and they published an extended family tree. They confirmed the presence of the familial susceptibility in the family, and the primary incidence of gastro-intestinal and endometrial cancer. They also stated that no conclusion concerning the inheritance pattern could be made at that time (Hauser *et al.*, 1936).

In 1966 Lynch *et al.* published a paper entitled "Hereditary Factors in Cancer: Study of Two Large Midwestern Kindreds". They identified two families in the United States that suffered from what Lynch described as a "cancer family syndrome" that had a similar phenotype to Warthin's Family G. Lynch was unsure at the time what the specific cause might be, but postulated that there might be an autosomal dominant inheritance. In 1967, Lynch *et al.* described 6 more "cancer families" including Warthin's Family G. In this paper, the specific phenotype of Lynch Syndrome was becoming clearer. For the first time, Lynch commented on how colon cancer was the primary cancer seen in the syndrome but also described the early age of onset and the occurrence of specific extracolonic tumours. Lynch *et al.* examined "Family G" again in detail in 1970. Cancer of the colon, endometrium, and stomach were reported to predominate, but there was also increased incidence of leukemia and of sarcomas, which were not present in Lynch's "cancer family syndrome". Lynch suggested that modifier genes might influence development of additional cancers in the syndrome (Lynch *et al.*, 1970).

In 2000, a subject from "Family G" was screened for mutations in mismatch repair genes. An *MSH2* mutation, c.646-3T>G, was found that alters the splice acceptor site for exon 4. This change results in a protein that has 8 amino acids inserted between codons 215 and 216 making it functionally inactive (Yan *et al.*, 2000). "Family G" was

reinvestigated in 2005 and predictive mutation testing was performed for all willing members of Warthin's "Cancer Family", a little over 110 years after Warthin first investigated this family (Douglas *et al.*, 2005).

1.1.4 Lynch Syndrome Risk Criteria:

1.1.4.1 Amsterdam Criteria:

In 1990, the International Collaborative Group on HNPCC (ICG-HNPCC) met in Amsterdam, Holland to develop criteria for classifying whether a family was at high risk for Lynch Syndrome. Due to the location of the meeting, these criteria become popularised as the "Amsterdam Criteria" (ACI). The requirements of the ACI are as follows: three family members (one of whom is a first-degree relative of the other two) in two or more successive generations must have had colorectal cancer and at least one case of colorectal cancer should be diagnosed under the age of 50. In addition to this, the diagnosis of Familial Adenomatous Polyposis or Attenuated FAP must be excluded (Vasen *et al.*, 1990). In 1990, the genes responsible for Lynch Syndrome were unknown, so these criteria were primarily used for selecting families for linkage analysis. They were specific, but not sensitive and excluded a number of families with Lynch Syndrome that had a high incidence of extracolonic tumours. In 1999, the ICG-HNPCC met again in Holland to create a modified ACI, which they dubbed the Amsterdam Criteria II or ACII. They modified the ACI keeping the same basic requirements as before, but now included extracolonic tumours (endometrium, stomach, small bowel, renal pelvis, ovary, brain and hepato-biliary) in the Lynch Syndrome spectrum instead of only CRC. The new criteria are used to pre-screen families for genomic sequencing of MMR genes which is

an expensive and time consuming process (Vasen *et al.*, 1999).

1.1.4.2 Bethesda Criteria:

The ACI and ACII criteria are useful for classifying families, but additional guidelines exist for selection of individuals who may be at risk for Lynch Syndrome. These are termed the "Bethesda Criteria" and were developed in 1997 (Rodriguez-Bigas *et al.*, 1997) and modified to the "Revised Bethesda Criteria" in 2004 (Umar *et al.*, 2004). Persons who meet the Bethesda Criteria (BC) are recommended to be screened for classical molecular features of Lynch Syndrome such as microsatellite instability and deficiency of MMR proteins in their tumour cells prior to genomic mutation testing. Using the BC allows for screening to become more efficient in two ways. First, only selected cancer patients that meet BC would be pre-screened for characteristics of Lynch Syndrome, thus saving time and money through not testing non-MMR deficient cases. Secondly, only those persons who both meet the BC and show features of Lynch Syndrome would be tested by the more expensive mutation screening, giving a second level of control. The complete Bethesda and Revised Bethesda guidelines are listed in Appendix 1.

1.1.4.3 Age and Cancer Modified Amsterdam Criteria:

The ACII have been modified following study of the Newfoundland population to identify heritable forms of cancer with a later age of onset and more variable extracolonic tumours. In these new criteria, the age of cancer is modified to under 60. Also, the range

of cancers listed in ACII have been expanded to include the Revised Bethesda Criteria cancers. These new criteria are referred to as Age and Cancer Modified Amsterdam Criteria (ACMAC) (Woods *et al.*, 2005). Since the MMR genes have been identified, these criteria are useful for selection of patients for mutation testing, as well as to identify linkage analysis for new genes that may be responsible for cancer syndromes outside of the mismatch repair pathway.

1.1.5 Mismatch Repair Pathway:

Mutations in the mismatch repair (MMR) genes are responsible for Lynch Syndrome. Mutations in these genes leave an individual with only one functional copy of a particular MMR gene, thus with no second copy if it is somatically mutation, making DNA mismatch repair vulnerable. MMR is responsible for the post-replicative identification and repair of mismatched bases or small insertions/deletions. While a single copy of a MMR gene is sufficient to maintain normal cellular DNA repair, individuals with a sole functional copy leave a cell susceptible to accumulating mutations if this is inactivated. These mutations are not specific, but can affect genes that either inhibit cell proliferation in their normal capacity (tumour suppressing), or genes that can induce cellular proliferation (tumour inducing) thus increasing a person's risk of developing cancer. To date, there are four MMR genes that Lynch Syndrome causing mutations: *MLH1*, *MSH2*, *MSH6*, and *PMS2*. *MLH1* and *MSH2* combined account for approximately 80% of known Lynch Syndrome mutations (Woods *et al.*, 2007).

1.1.6 Mismatch Repair Genes:

The *MSH2* gene is located on chromosome 2p22-21 and was the first gene known to cause Lynch Syndrome when mutated (Fishel, *et al.* 1993, Leach, *et al.* 1993, Peltomaki, *et al.* 1993). The *MSH2* gene spans 16 exons which produces a 3,145 base transcript that encodes a 934 amino acid protein. *MSH6* was first reported as a unknown MMR protein in 1992 by Hughes and Jiracny and later named *GTBP* or G T mismatch Binding Protein (Palombo *et al.*, 1995). It was mapped to 2p16 and also causes Lynch Syndrome when mutated (Papadopoulos *et al.*, 1995). The *MSH6* gene spans 10 exons which produces a 4,255 base transcript that encodes a 1360 amino acid protein. The protein product of *MSH2* forms a heterodimer with the protein product of *MSH6* or *MSH3* to form the MUTSα (Drummond *et al.*, 1994) or MUTSβ (Palombo *et al.*, 1996) complexes respectively. The dimerization domains of MSH2 are in the middle of the protein and at the C-terminus (Bandipalliam, 2007). MUTSα scans DNA for mismatched bases as well as small insertions or deletions while MUTSβ scans for DNA loops caused by insertions or deletions (Palombo, *et al.*, 1996). When these complexes detect a mismatch, insertion or deletion, they recruit additional proteins such as MLH1, PMS2 and EXO1 to first excise and then repair the error. Variations in the *EXO1* gene do not appear to cause Lynch Syndrome (Thompson *et al.*, 2004).

The *MLH1* gene is located on chromosome 3p21-23 and was the second gene discovered to cause Lynch Syndrome when mutated (Bronner *et al.* 1994; Papadopoulos *et al.* 1994). The *MLH1* gene spans 19 exons that produces a 2,524 base transcript that encodes a 756 amino acid protein. The MLH1 protein forms a heterodimer with the

protein product of *PMS2* to form the MUTLα complex. The PMS2 binding domain is located at the C-terminus of the MLH1 protein (Bandipalliam, 2007). The *PMS2* gene is located on chromosome 7p22 and is also responsible for Lynch Syndrome when mutated (Nicolaides *et al.*, 1995). It spans 15 exons which produce a 2,802 base transcript encoding an 862 amino acid protein. The MLH1 protein also forms a heterodimer with the MLH3 protein to form MUTLδ (Cannavo *et al.*, 2005). Some variants in the *MLH3* gene seem to be associated with a weak familial cancer risk but not Lynch Syndrome (Liu *et al.*, 2003). Either the MUTLα or MUTLδ complex is responsible for recognising MUTSα/MUTSβ bound to mismatches. MUTLα/MUTLδ then recruits the EXO1 protein to excise the error and allow DNA polymerase to fill in the gap.

1.1.7 Molecular Characteristics of Lynch Syndrome:

1.1.7.1 Microsatellite Instability:

Microsatellite Instability (MSI) is a hallmark feature of Lynch Syndrome. MSI involves either the shortening or lengthening of short tandem repeats of DNA, called microsatellites. These microsatellites which can be either mononucleotide repeats, di-, tri-, tetra-, penta- etc. When the MMR pathway is compromised, these repetitive elements accumulate insertions or deletions that would normally be repaired. Whether these repeats are microsatellite stable (MSS), or display instability (MSI) can be determined via amplification of the repeat in both normal and tumour DNA and comparing the relative sizes. Specific microsatellites can be amplified and examined to detect expansion or contraction. This detects whether or not the MMR pathway is functional. This is a

simple and efficient system to examine the MMR pathway, but gives no indication of which gene might be impaired or how. While MSI is associated with Lynch Syndrome it can also be seen in sporadic colorectal cancer. A proportion (~15%) of older patients with sporadic CRC will display MSI in tumour DNA. This MSI is commonly due to the epigenetic inactivation of the *MLH1* gene through hypermethylation of the promoter (Cunningham *et al.*, 1998) or may rarely occur because of two somatic mutations of the *MLH1*, *MSH2*, *MSH6*, or *PMS2* genes. Hypermethylation of the *MLH1* promoter is a cause of transcriptional silencing and the incidence of methylation increases with age (Cunningham *et al.*, 1998). A recent study has also shown that *MSH2* can be inactivated by promoter hypermethylation, and this change can be inherited (Chan *et al.*, 2007).

1.1.7.2 Immunohistochemistry:

Detecting the protein products of MMR genes is another useful way to assess MMR pathway function. Immunohistochemistry (IHC) is used to stain cells for the protein products of specific genes. This is done by staining sections of tumour tissue with monoclonal antibodies directed against a particular protein. For Lynch Syndrome these proteins are MLH1, MSH2, MSH6 and PMS2. Negative IHC means that there was no protein product detected in the tumour tissue. This means the gene is inactivated or not producing the correct protein in the tumour, either by germline and/or somatic mutation or epigentic silencing. The absence of a MMR protein in tumour tissues means that this gene has been inactivated, though the exact mechanism is not clear from IHC alone.

IHC and MSI status are key molecular markers for screening CRC families for

LS. These tests can be used as independent risk criteria, but are more useful in parallel. They provide information about the underlying genetic cause of a tumour. MSI status provides information on the status of the MMR pathway, while IHC provides information as to which gene or genes are inactivated, giving direction to mutational studies. Information from both of these key LS markers was used substantially as screening criteria in thesis' work.

1.2 Objectives:

The objectives of this thesis' work were twofold. First, to determine the proportion of CRC in the Newfoundland population resulting from germline mutations of the primary MMR genes: *MSH2* and *MLH1*. Germline mutations in these two genes would indicate the prevalence of Lynch Syndrome in Newfoundland. By using a Newfoundland-based CRC registry, a population based approach can be used that would identify families in the general population outside of those known high risk families. This will allow a better indication of the incidence of LS in the population. The second objective was to catalogue and compile all variants in the MMR genes *MSH2*, *MLH1*, and *MSH6* that have been published in the literature. There were no comprehensive and current databases of variants of LS at the time of this undertaking. This database was developed to provide an invaluable resource for this thesis' work as well as for other researchers of Lynch Syndrome.

Chapter 2

2.0 Population Based Study of Lynch Syndrome in Newfoundland

2.1 Introduction:

Newfoundland and Labrador has the highest age-standardised rate of colorectal cancer of all Canadian provinces for males (82 100,000) and second highest for females (51 100,000). This incidence is reported as an underestimate for Newfoundland and Labrador due to undercounting (Cancer Canada). Several Lynch Syndrome MMR mutations have been previously detected in the Newfoundland population. These include two large deletions in the *MSH2* gene, one being a deletion of Exon 8 (c.1277-?_1386+?del, p.Lys427_Gln462>GlyfsX4) (Woods *et al.*, 2005), and the second a deletion of exons 4 to 16 (c.646-?_2802+?del) (Woods *et al.*, 2005). There is also a mutation which disrupts the intron 5 splice acceptor site of *MSH2*, c.942+3A>T (a.k.a. the "Family C mutation"), leading to incorrect splicing out of exon 5 via post-transcriptional modification (p.Val265_Gln314del) (Green *et al.*, 2002). A promoter mutation in *MLH1*, c.-42C>T, leading to an approximate 33% efficiency in gene transcription has also been reported in the Newfoundland population (Green *et al.*, 2003). In addition there are many families that display the clinical characteristics of Lynch Syndrome in which MMR mutations have not been found, giving rise to the possibility of novel hereditary cancer syndrome genes outside of the MMR pathway.

The island of Newfoundland presents an ideal environment for the study of

genetic diseases. The current population of Newfoundland is approximately 505,000 (www.stats.gov.nl.ca). The island was populated by 20,000-30,000 settlers mainly from South-West England and South-East Ireland between the years of 1760-1820 (Manion JJ, 1977). This makes the genetic structure of the population very homogenous. Settlers came to Newfoundland primarily for the occupation of fishing, and because of this, founded many small "outport" communities all along the island's coast where travel was limited until the 20th century. Due to the small number of settlers, carriers of MMR mutations who settled in Newfoundland have enriched the Newfoundland population with Lynch Syndrome via a founder effect. A founder effect is when a small number of initial settlers with a decreased amount of genetic variation from the original population form a new population, leading to an enrichment of particular alleles in the new population. This limited number of settlers in Newfoundland has also created a limited pool of modifier genes acting on LS as well. This homogenous population, enriched in mutations, may allow study of how mutations behave with specific modifiers, and how these may act to alter phenotype.

It has been proposed that the high colorectal cancer incidence in Newfoundland is caused by a high frequency of Lynch Syndrome due to a population enriched in founder MMR gene mutations. This hypothesis was tested by examining cancer patients who were diagnosed with colorectal cancer in Newfoundland for the five year period of 1999-2003 and determining MMR mutation status for those fulfilling high risk criteria for Lynch Syndrome. All CRC patients in Newfoundland during this period were eligible to be entered into the Newfoundland Colorectal Cancer Registry (NFCCR). The NFCCR

was then used to identify (details reviewed in the Materials and Methods section) those subjects who have characteristics of Lynch Syndrome, and should be tested for MMR mutations. Identification of positive MMR mutation status, indicating Lynch Syndrome, will allow for enhanced clinical care in patients found to be carriers, as well as potentially affected family members.

With a known family mutation, it is easier to screen family members through genetic testing. It is no longer required to perform expensive exon by exon sequencing of multiple genes searching for an unknown mutation, or to offer expensive clinical screening to all family members. When subjects are found to be carriers, specific and intensified preventative clinical screening at younger ages than the general population should allow for earlier detection and treatment of carriers. Family members who were previously "at-risk", but who receive negative genetic test results, will be given piece of mind, as well as ceasing any unnecessary screening other than the general population recommendations.

Another possible benefit of this project is the exclusion of high risk criteria families from being Lynch Syndrome families. If it is known that a family is not a carrier of an MMR gene mutation and still meets AC1 risk criteria, these families would then be ideal candidates in screening for novel genes responsible for Familial Colorectal Cancer Type X.

2.2 Materials and Methods:

2.2.1 Subjects:

Research subjects included in the Newfoundland Colorectal Cancer Registry (NFCCR) were those diagnosed under the age of 75 with primary CRC in the province of Newfoundland and Labrador during the period of 1999-2003. Subjects alive at time of contact gave informed consent, and those previously deceased were consented by proxy. Subjects were asked to provide a blood sample from which genomic DNA was extracted. Also, DNA was extracted from paraffin-embedded tissue samples of patients' colon tumour and normal mucosa for MSI testing. IHC staining of tumour tissue was performed to determine expression of MLH1, MSH2 and MSH6 proteins. Promoter hypermethylation causing epigenetic silencing of *MLH1* was also determined in selected cases. Family histories were obtained and families or subjects were classified by the Amsterdam Criteria or Revised Bethesda Criteria. Subjects who had blood DNA available before June 30th, 2005, and met screening criteria, were all tested for mutations in *MLH1* and *MSH2*.

2.2.2 Screening Criteria:

A hierarchal molecular testing scheme was employed to triage subjects for mutation testing (See Figures 2.2 and 2.3 in the results section). In brief, testing for genomic rearrangements via Multiplex Ligation-dependant Probe Amplification (MLPA), a relatively inexpensive and quick test, was performed first. This was performed on subjects who met family criteria of either AC I, AC II or the highest category of Bethesda:

(Intermediate I (INTI) Appendix 1) and had tumours which were MSI high (MSI in at least 2 of 5 microsatellite markers tested), or IHC deficient for MLH1, MSH2 and or MSH6. Samples which did not show large genomic rearrangements were then tested according to the following criteria: subjects whose tumours were IHC deficient for MSH2, or were ACI or ACII, were all tested for a common splice variant in Newfoundland, c.942+3A>T (referred to as the "Family C" mutation), via sequencing. This is the most common Lynch Syndrome mutation in the published literature and a founder mutation in Newfoundland (Froggatt *et al.*, 1999; Woods *et al.*, 2007). It alters the splice acceptor site for exon 5 of *MSH2* resulting in the in-frame deletion of exon 5 (p.Val265_Gln314del) in the MSH2 protein. Samples that were negative via MLPA and "Family C" testing underwent further testing. Samples that were IHC deficient for either MSH2 or MLH1, or fulfilled the ACI or ACII criteria were sequenced exon by exon in either *MLH1* or *MSH2* from polymerase chain reaction (PCR) products (Mullis *et al.*,1986). These two genes account for greater than 90% of all variants in LS (Woods *et al.*, 2007). Subjects whose tumours were deficient in MLH1 by IHC were only sequenced if there was no hypermethylation of the *MLH1* promoters.

2.2.3 Multiplex Ligation Dependent Probe Amplification:

MLPA is a relatively new procedure developed by Microbiology Research Centre Holland (MRC-Holland) which detects genomic deletions or duplications. In brief, this system works by having two probes for each genomic exon, which are ligated together when annealed to genomic DNA. The ligated probes are then amplified by fluorescent

dye-labelled primers. Comparison of amounts of ligated and amplified probe in test samples, relative to normal controls (intact genomic individuals with no deletions or duplications, therefore two copies of each exon), makes it possible to detect deletions or duplications of entire exons. Upon probe amplification of an exon there will be approximately 50% of amplified product (ligated probe) per deletion, or approximately 150% of the amplified product per exon per duplication as compared to a control with two copies of each gene.

An MLPA kit was available for all exons of *MLH1* and *MSH2*. Exonic rearrangements in *MSH2* and *MLH1* are fairly common (Woods *et al.*, 2007). They are less common in *MSH6* and there are no data concerning large deletions/duplications in *PMS2*, *MLH3*, and *MSH3*. Due to the low cost of the test ($15 per patient), as well as the ease and speed of the procedure, MLPA was used as the initial mutation screening test. Samples were always tested under the same reaction conditions with at least five controls. All samples and controls from each reaction were analysed via the Beckman CEQ-8000 Genetic Analysis System (Beckman Coulter Canada, Inc., 6755 Mississauga Road, Suite 600, Mississauga, Ontario, L5N 7Y2). Any samples that showed a deletion or duplication via MLPA were repeated for confirmation.

2.2.4 MLPA Reaction Conditions:

Samples to be tested by MLPA (Figure 2.1) were amplified according to the manufacturer's protocol, available from www.mlpa.com. In a Eppendorf Mastercycler Gradient thermocycler (Eppendorf, 6670 Campobello Road Mississauga, Ontario, L5N

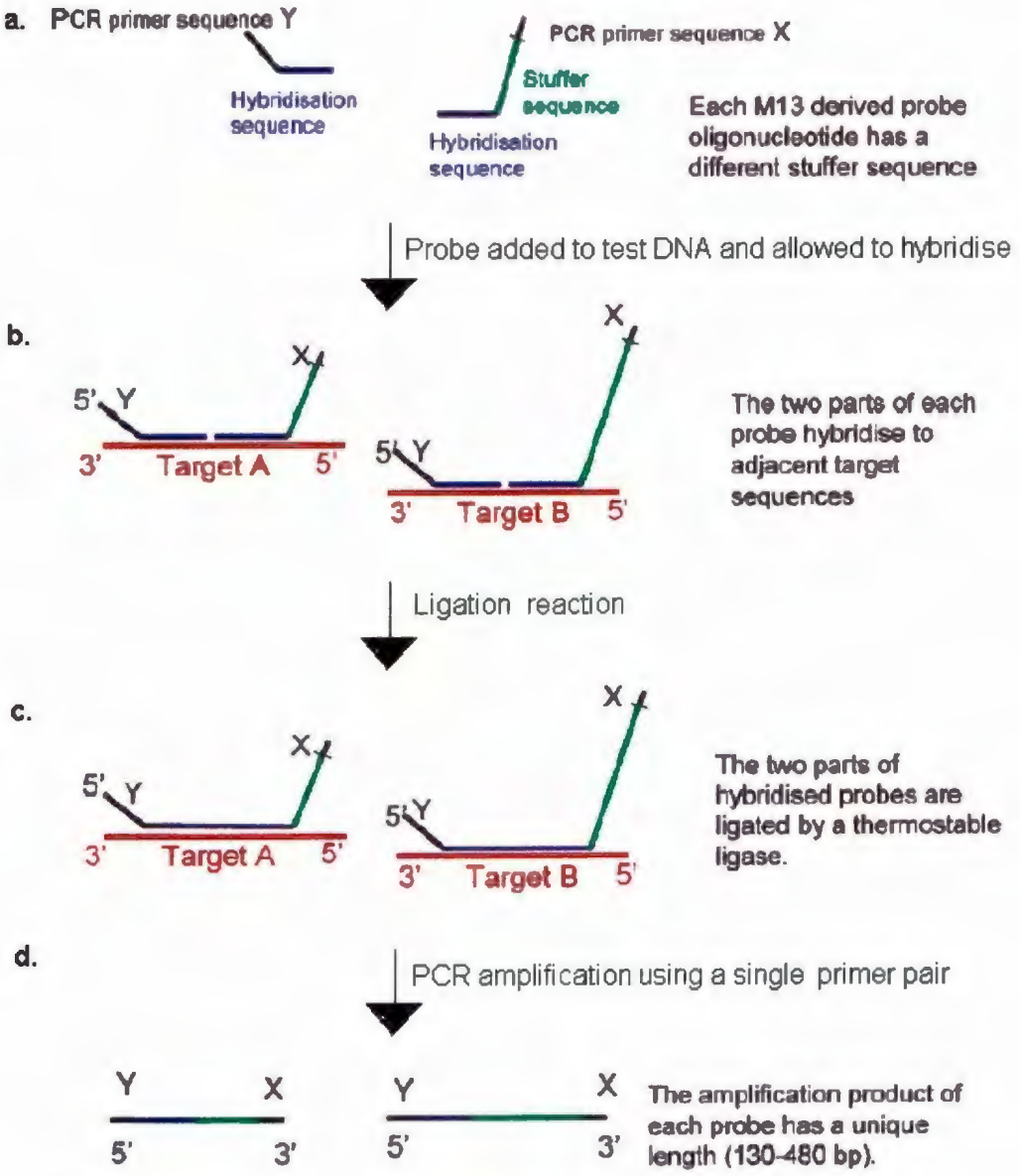

Figure 2.1: MLPA Probe Hybridisation and Ligation: Graphical representation of different size MLPA probes hybridising with DNA and the ligation reaction generating the amplification product where target A is one exon, and target B is a different exon. (http://leedsdna.info/HUGO/2004/Lab_Notes/Image33.gif).

2L8), 100ng of genomic DNA was diluted with 5 µl of TE buffer, denatured at 98°C for 5 minutes, and then cooled to 25°C. At 25°C, 1.5 µl of SALSA Probe mix and MLPA buffer were added, and mixed vigorously by pipetting at least 10 times. Samples were then incubated at 90°C for 1 minute, and the probes were hybridised at 60°C for 16 hours. For the ligation reaction, the temperature was reduced to 54°C and 32 µl of "Ligase-65" mix was added and mixed by pipetting, incubated at 54°C 15 minutes for ligation, and then heated for 5 minutes at 98°C to inactivate the ligase. "Ligase-65" mix was made fresh within an hour of use, stored on ice and consisted of 3 µl of "Ligase-65" buffer A, 3 µl of ligase buffer B, 25 µl of deionised water and 1 µl of Ligase-65 mixed by pipetting. Following ligation, "PCR Protocol Two" (www.mlpa.com) was followed to amplify the ligated probes with fluorescently labelled primers. For "PCR Protocol Two", 4 µl of 10X SALSA PCR buffer, 20 µl deionised water, and 10 µl of MLPA ligation reaction were mixed, and then put into the thermocycler and held at 60°C. While at 60°C, 10 µl Polymerase mix was added, and the PCR conditions were as follows: 30 cycles of 30 seconds at 95°C, 30 seconds at 60°C, 60 seconds at 72°C, followed by a 20 minute incubation at 72°C. Polymerase mix was made fresh within an hour before use, stored on ice and consisted of 2 µl of SALSA PCR-Primers, 2 µl of SALSA Enzyme Dilution buffer, 5.5 µl of deionised water, and 0.5 µl of SALSA Polymerase.

2.2.5 MLPA Fragment Analysis:

After completion of the PCR reaction, 0.7 µl of the PCR reaction, 0.5 µl of the Beckman D1-labelled 60-600bp molecular weight marker and 40 µl of Beckman Sample

Loading Solution (SLS) were mixed and loaded into the CEQ-8000 system for fragment analysis. Settings for the fragment analysis were as follows: capillary temperature of 50°C, denaturation of 90°C for 90 seconds, injection time of 2.0 KV for 60 seconds and runtime of 60 minutes at 4.8 KV.

Each test run was analysed by the CEQ-8000 fragment analysis software according to the following settings: peaks >3% were included; size standard-600 (Beckman nr. 608095); and a slope threshold of 1. These settings included all of the correct peaks for dosage analysis, but also included size standard peaks, as well as background and shoulder peaks. A filter was applied to the analysed data whereby dyes D1, D2 and D3 (which were not probe specific) were excluded, peak areas less than 5000 Relative Fluorescence Units (RFU) were excluded, and peak sizes less than 125nt were excluded. This filter eliminates ambiguous peaks, but each electropherogram was still manually reviewed to include some probe peaks less than 5000RFU, and to eliminate noise peaks greater than 5000RFU due to variations in peak size.

2.2.6 Dosage Analysis:

Data including peak size and height were then copied into an Excel spread sheet developed by recommendations from the www.mlpa.com website. The website provided a template, which was then modified to be applicable for the HNPCC-1(*MLH1* and *MSH2*) kit.

Firstly, for each of the normal controls and test samples the peak area for each

probe was added to give the total area. Each probe area was divided by the total to give the relative fraction for each probe of the total area. For normal controls, this relative fraction was then averaged for each probe over all the normal controls.

Next, each relative probe area for both normals and test samples was divided by the average relative fraction for the normal controls. The average of each relative fraction was then taken for each test and control sample. Finally, each relative fraction was divided by the average relative fraction to give the final result. Output was generated in an excel spreadsheet colour coordinated by dosage.

The final output of analysis was colour-coded for ease of identification: yellow for normal relative dosage, if within 20% of 1.00 ($\geq$0.80-$\leq$1.20); dosages within 20-30% of 1.00 were intermediate results and were coded pink if less than 1.00 ($\geq$0.70-$\leq$0.80), and light blue if 20-30% greater than 1.00 ($\geq$1.20-$\leq$1.30). Output for relative dosage of $\leq$70% or $\geq$130% is indicative of either a deletion or a duplication relative to the control, and displayed as red if $\leq$0.70 and dark blue if $\geq$1.30. MLPA data for each gene was then recorded in a spreadsheet labelled by DNA number.

2.2.7 Direct Sequencing for Mutation Detection:

2.2.7.1 PCR DNA Amplification:

Polymerase chain reaction (PCR) (Mullis *et al.*,1986) was used to amplify all exons and flanking intronic sequences of MMR genes *MLH1* and *MSH2*. PCR products were then purified for genomic sequencing in a Beckman CEQ 8000 Genetic Analysis

System. Primers were ordered from Operon, a division of Qiagen (QIAGEN Inc., 2800 Argentia Road, Unit 7, Mississauga, Ontario, L5N 8L2). PCR reactions were performed in a Biometra T1 thermocycler (Biometra GmbH I. L., Rudolf-Wissell-Straße 30, 37079 Goettingen, Germany).

The reaction conditions for *MLH1* and *MSH2* primers were as follows: 95°C for 2 minutes, 35 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds and elongation at 72°C for 1 minute. Then a final elongation at 72°C for 5 minutes.

Concentrations used for the PCR reactions were optimized for the PlatiniumTaq polymerase kit from Invitrogen (Invitrogen Canada Inc., 2270 Industrial St., Burlington, Ontario, L7P 1A1). Final concentrations per 25µL reactions were as follows: 1X Buffer solution (10X stock provided by the manufacturer, with 20mM $MgCl_2$), 0.6µM of each primer, 0.2mM of each deoxynucleotide triphosphate (dNTPs) and 0.75 units of Taq polymerase. Reaction products were stored at 4°C until they were analysed via electrophoresis on a 2% agarose gel stained with ethidium bromide visualised under UV light for PCR product verification. Primer sequences are summarized Appendix 2.

2.2.7.2 PCR Product Purification:

To facilitate clean up for sequencing reactions, the PCR reactions were purified by Exo Nuclease I (ExoI) and Shrimp Alkaline Phosphatase (SAP) to digest excess primer and dephosphorylate surplus free nucleotides respectively. In the thermocycler,

7µL of PCR product was incubated with 7.5µL deionised water, 0.5µL 1U µL of SAP and 0.5µL of 10U µL ExoI. Reaction conditions were 30 minutes at 37°C followed by a 10 minute inactivation step at 80°C.

2.2.7.3 Sequencing Reaction:

Each amplicon was sequenced in both the reverse and forward directions. Sequencing reactions were set up using 1/4 the recommended sequencing reagents to conserve reagents while still giving reliable results. Reactions were made to the following specifications: Sequencing buffer, 0.325pMol of primer, 1µL of purified PCR product, and 1µL of DTCS Quick Start Mix with final volume of 20µL. Thermocycler conditions were as follows: 30 cycles of 96°C for 20 seconds, 50°C for 20 seconds, and 60°C for 4 minutes. The sequencing reaction was stopped by the addition of 5µL of freshly prepared stop solution. The stop solution had a final concentration of 1.5M Sodium Acetate, 50mM Sodium EDTA at pH 8.0 and 4µg of glycogen and was prepared less than an hour before use.

2.2.7.4 Sequencing Reaction Purification:

After the addition of stop solution, the entire volume of the reaction was mixed with 60µl of 95% ethanol that had been stored at -20°C, inverted 5 times and mixed by vortex before being centrifuged at 4°C and 3000g for 40 minutes. Following this spin, sample plates were inverted, and ethanol decanted. Samples were then washed by the

addition of 200µl of 70% ethanol that had been stored at -20°C, and spun at 4°C and 3000g for 5 minutes. This wash step was repeated twice. Following the second washing, the sample plate was left inverted and centrifuged again at 300rpm to remove any excess ethanol. Samples were left to dry in a desiccator for approximately 15 minutes. Samples were then resuspended in 30µl of SLS, and covered in mineral oil for analysis in the Beckman CEQ-8000 Genetic Analysis System.

2.2.7.5 Sequence Analysis:

The run conditions for analysis of sequence fragments was the LFR-b default program in the CEQ-8000 software. This program consisted of capillary temperature of 57°C, denaturation of 90°C for 120 seconds, injection at 2.0 KV for 15 seconds and runtime of 60 minutes at 6.0 KV. Sequence data were exported and analysed on a remote workstation. The analysis settings were modified from the default as follows: PCR product option was selected, heterozygote detection was enabled, and a reference sequence was included in the analysis.

Forward and reverse sequences for each exon, as well as for the reference sequence, were reviewed and manually compiled using the Beckman CEQ-8000 Genetic Analysis System software. Sequence data were then recorded in a spreadsheet coded by DNA number.

2.3 Results:

There were 779 patients in the NFCCR as of June 30th, 2005 (end date of this thesis' work) and 568 (73%) had genomic DNA available from blood extraction. Of these, 536 (94%) where intact for MLH1, MSH2 and MSH6 proteins by IHC, and 32 (6%) showed a deficiency of at least one MMR protein. This can be broken down to 16 IHC deficient for MLH1 (50% of all IHC), 6 IHC deficient for MSH2 (19%) and 15 IHC deficient for MSH6 (47%). The numbers add up to greater than 100% due to the fact that many of the samples deficient for MSH6 protein were deficient for one of the other MMR proteins as well. Of the 15 IHC deficient for MSH6, only 10 of these were deficient for MSH6 alone. Three samples were deficient additionally for MSH2 and two deficient additionally for MLH1. The complete breakdown of samples can be seen in Figure 2.3.

2.3.1 MSH2:

In total, 115 subjects (Figure 2.2) who had available genomic DNA met criteria (ACI ACII/INT1 or IHC negative for MLH1 or MSH2 or were MSI high) for testing via MLPA. None of these samples showed any duplications or deletions in *MSH2* (Figure 2.2).

There were 22 subjects who met criteria (ACI ACII or IHC negative for MSH2) for testing for the common Newfoundland "Family C" mutation. Of these subjects, 4 were positive for the mutation (Figure 2.2).

There were 14 subjects who met criteria (ACI/ACII or IHC negative for MSH2) for complete sequencing of *MSH2*. Eleven polymorphisms where found, but no pathogenic mutations. A list of polymorphisms and "Family C" mutations found is summarised in Table 2.1.

2.3.2 *MLH1*:

In total, 115 subjects (Figure 2.2) who had available genomic DNA met criteria (ACI/AC1/INT1 or IHC negative for MLH1 or MSH2 or were MSI high) for testing via MLPA None of these samples showed any duplications or deletions in *MLH1* (Figure 2.2).

There were 15 subjects who met criteria (ACI/ACII or IHC negative for MLH1) for complete sequencing of *MLH1*. Seven polymorphisms where found, but no pathogenic mutations. A list of polymorphisms found is summarised in Table 2.2.

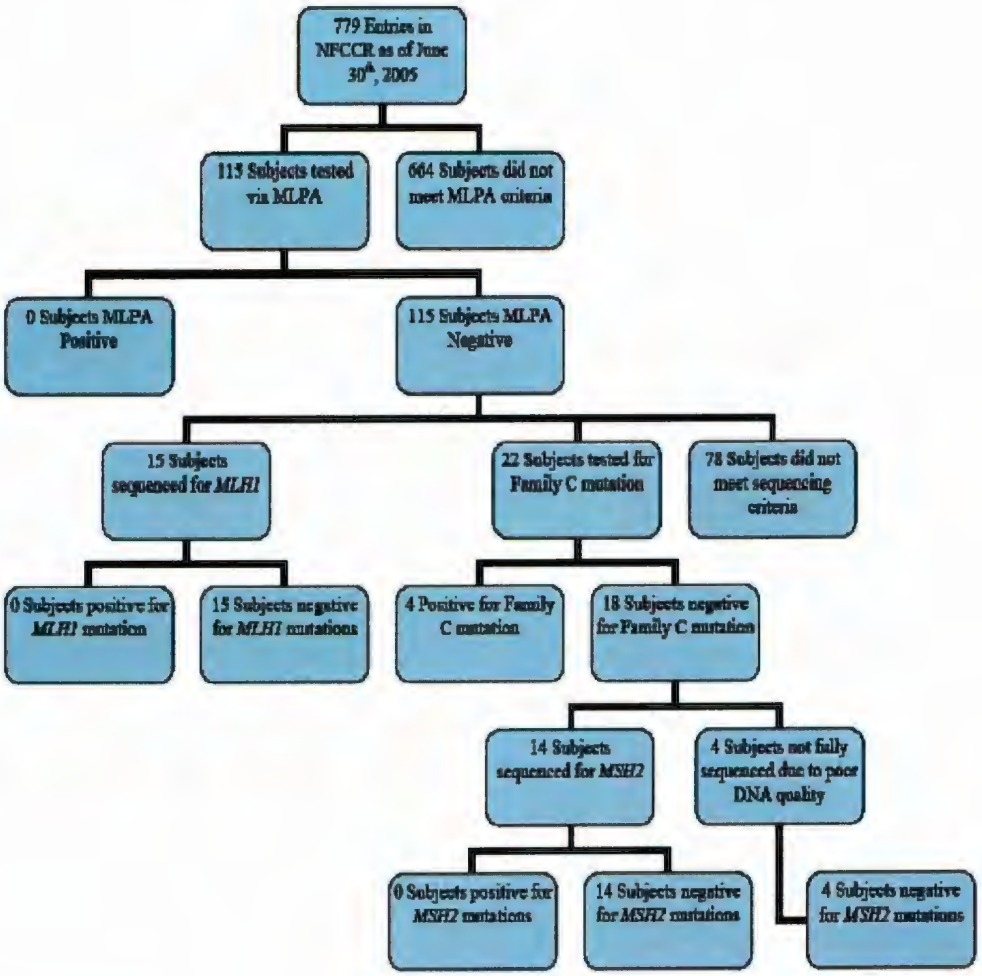

Figure 2.2: Mutation Analysis Flow Chart: Flow chart illustrating samples tested via MLPA and subsequent testing and results.

Table 2.1: Mutations and Polymorphisms found in *MSH2*

Exon/ Intron	Blood DNA ID	Base Change	Predicted Protein Change	Novel
Intron 5	RD-217	c.942+3A>T	p.Val265_Gln314del	No[1]
Intron 5	RD-987	c.942+3A>T	p.Val265_Gln314del	No[1]
Intron 5	RD-1053¤	c.942+3A>T	p.Val265_Gln314del	No[1]
Intron 5	RD-2266	c.942+3A>T	p.Val265_Gln314del	No[1]
Intron 9	RD-238*	c.1511-9A>T	N/A	No[2]
Intron 9	RD-599	c.1511-9A>T	N/A	No[2]
Intron 10	RD-64	c.1661+12G>A	N/A	No[3]
Intron 10	RD-70	c.1661+12G>A	N/A	No[3]
Intron 10	RD-238*	c.1661+12G>A§	N/A	No[3]
Intron 10	RD-347*	c.1661+12G>A	N/A	No[3]
Intron 10	RD-348	c.1661+12G>A§	N/A	No[3]
Intron 10	RD-526*	c.1661+12G>A§	N/A	No[3]
Exon 11	RD-347*	c.1666T>C	p.Leu556Leu	No[4]
Intron 12	RD-526*	c.1759+46A>T	N/A	Yes
Intron 12	RD-1053¤	c.1759+46A>T	N/A	Yes

* -These subjects contain two polymorphisms

§ -This change was homozygous

¤ -This subject contains one polymorphism and the "Family C" mutation

1 - Leach,F.S., *et al.*, 1993, Liu,B., *et al.*, 1994, Froggatt,N.J. *et al.*, 1995, Miyaki,M. *et al.*, 1995, Konishi,M. *et al.*, 1996, Thibodeau,S.N., *et al.*, 1996, Moslein,G. *et al.*, 1996, Lu,S.L. *et al.*, 1996, Lu,S.L. *et al.*, 1996, Pensotti,V. *et al.*, 1997, Viel,A. *et al.*,1997, Wijnen,J. *et al.*, 1997, Viel,A. *et al.*, 1998, Bai,Y.Q. *et al.*, 1999, Bapat,B.V. *et al.*, 1999, Chan,T.L. *et al.*, 1999, Curia,M.C. *et al.*, 1999, de Leon,M.P. *et al.*, 1999, Froggatt,N.J. *et al.*, 1999, Lamberti,C. *et al.*, 1999, Lin,X. *et al.*, 1999, Lynch,H.T. *et al.*, 1999, Syngal,S. *et al.*, 1999, Wahlberg,S. *et al.* 1999, Wang,Q. *et al.*, 1999, Desai,D.C. *et al.*, 2000, Fidalgo,P. *et al.*, 2000, Liu,T. *et al.*, 2000, Martin,R.H.I *et al.*, 2000, Montera,M. *et al.*, 2000, Nomura,S. *et al.*, 2000, Otway,R. *et al.*, 2000, Percesepe,A. *et al.*, 2000, Pistorius,S.R. *et al.*, 2000, Bisgaard,M.L. *et al.*, 2001, Caldes,T. *et al.*, 2001, Holinski-Feder,E. *et al.*, 2001, Jakubowska,A. *et al.*, 2001, Kurzawski,G., *et al.*, 2001, Samowitz,W.S. *et al.*, 2001, Furukawa,T. *et al.*, 2002, Kurzawski,G. *et al* , 2002, Nilbert,M. *et al.*, 2002, Wahlberg,S.S. *et al.*, 2002, Yuen,S.T. *et al.*, 2002, Wagner,A. *et al.*, 2003, Fields,J.Z. *et al.*, 2004, Caldes,T. *et al.*, 2004, Jose,G. *et al* , 2004, Lage,P.A. *et al.*, 2004, Mangold,E., *et al.*, 2004, Ponz,de Leon, *et al.*, 2004, Hampel,H. *et al.*,2005,Stonnorken,A.T. *et al.*, 2005

Table 2.2: Polymorphisms found in *MLH1*

Exon/ Intron	Blood DNA ID	Base Change	Predicted Protein Change	Novel
Exon 8	RD-147	c.655A>G*	p.Ile219Val	No[1]
Exon 8	RD-854	c.655A>G	p.Ile219Val	No[1]
Exon 8	RD-1052	c.655A>G	p.Ile219Val	No[1]
Exon 8	RD-1769	c.655A>G	p.Ile219Val	No[1]
Exon 8	RD-1913	c.655A>G	p.Ile219Val	No[1]
Exon 8	RD-2164	c.655A>G	p.Ile219Val	No[1]
Intron 8	RD-1971	c.677+72G>T	N/A	Yes

* -This change was homozygous

1 - Buerstedde,J.M.*et al.,1995*, Liu et al., 1995, Moslein,G. *et al.*, 1996, Nystrom-Lahti,M. *et al.*, 1996, Nakahara,M. *et al.*, 1997, Tomlinson,I.P. *et al.*, 1997, Herfarth,K.K. *et al.*, 1997, Wu,Y. *et al.*, 1997, Shimodaira,H. *et al.*, 1998, Curia,M.C. *et al.*, 1999, Ghimenti,C. *et al.*, 1999, Hutter,P. *et al.*, 2000, Infante-Rivard,C. *et al.*, 2000, Montera,M. *et al.*, 2000, Ellison,A.R. *et al.*, 2001, Furihata,M. *et al.*, 2001, Baldinu,P. *et al.* 2002, Krajinovic,M. *et al.* 2002, Kurzawski,G. *et al.*, 2002, Murata,H. *et al.*, 2002, Palicio,M. *et al.* 2002, Trojan,J. *et al.*, 2002, Ward,R. *et al.* 2002, Chen-Shtoyerman,R. *et al.*, 2003, Kondo *et al.*, 2003, Mathonnet,G. *et al.*, 2003, Raevaara,T.E. *et al.* 2003, Infante-Rivard,C. *et al.*, 2003, Renkonen,E. *et al.*, 2003, Bagnoli et al.., 2004, Hudler,P. *et al.*, 2004, Kim,J.C. *et al.*, 2004, Liu,S.R. *et al.*, 2004, Liu,S.R. et al., 2004, Rollinson,S. *et al.*, 2004, Tournier,I. *et al.*, 2004, Apessos et al.., 2005, Bacani et al.., 2005, Li,J.H. *et al.*, 2005, Lee et al.., 2005, Oda et al.., 2005, Stanislawska-Sachadyn *et al.*, 2005, Damaraju *et al.*, 2006, Song *et al.*, 2006, Yu *et al.*, 2006

con't from table 2.1

2 - Cunningham,J.M., *et al.*, 2001, Rajkumar,T. *et al.*,2004, Hegde *et al.*, 2005, Velasco *et al.*, 2005, Jung *et al.*, 2006

3 - Wijnen,J. *et al.*, 1994, Wahlberg,S.S. *et al.*, 1997, Farrington,S.M. *et al.*, 1998, Percesepe,A. *et al.*, 1998, Desai,D.C. *et al.*, 2000, Otway,R. *et al.*, 2000, Holinski-Feder,E. *et al.*, 2001, Baldinu,P.*et al.*, 2002, Kurzawski,G. *et al.*, 2002, Scartozzi,M. *et al.*, 2002, Chen-Shtoyerman,R. *et al.*, 2003, Cederquist,K. *et al.*, 2004, Liu,S.R. *et al.*, 2004, Rajkumar,T. *et al.*,2004, Apessos,A. *et al.*., 2005, Hegde et al.., 2005, Lee *et al.*, 2005, Velasco *et al.*, 2005, Jung *et al.*, 2006, Wang *et al.*, 2006

4 - Wehner,M. *et al.*, 1997, Desai,D.C. *et al.*, 2000, Scott,R.J., *et al.*, 2001, Scartozzi,M. *et al.*, 2002, Hendriks,Y. *et al.*, 2003, Auclair *et al.*, 2006, Jung *et al.*, 2006

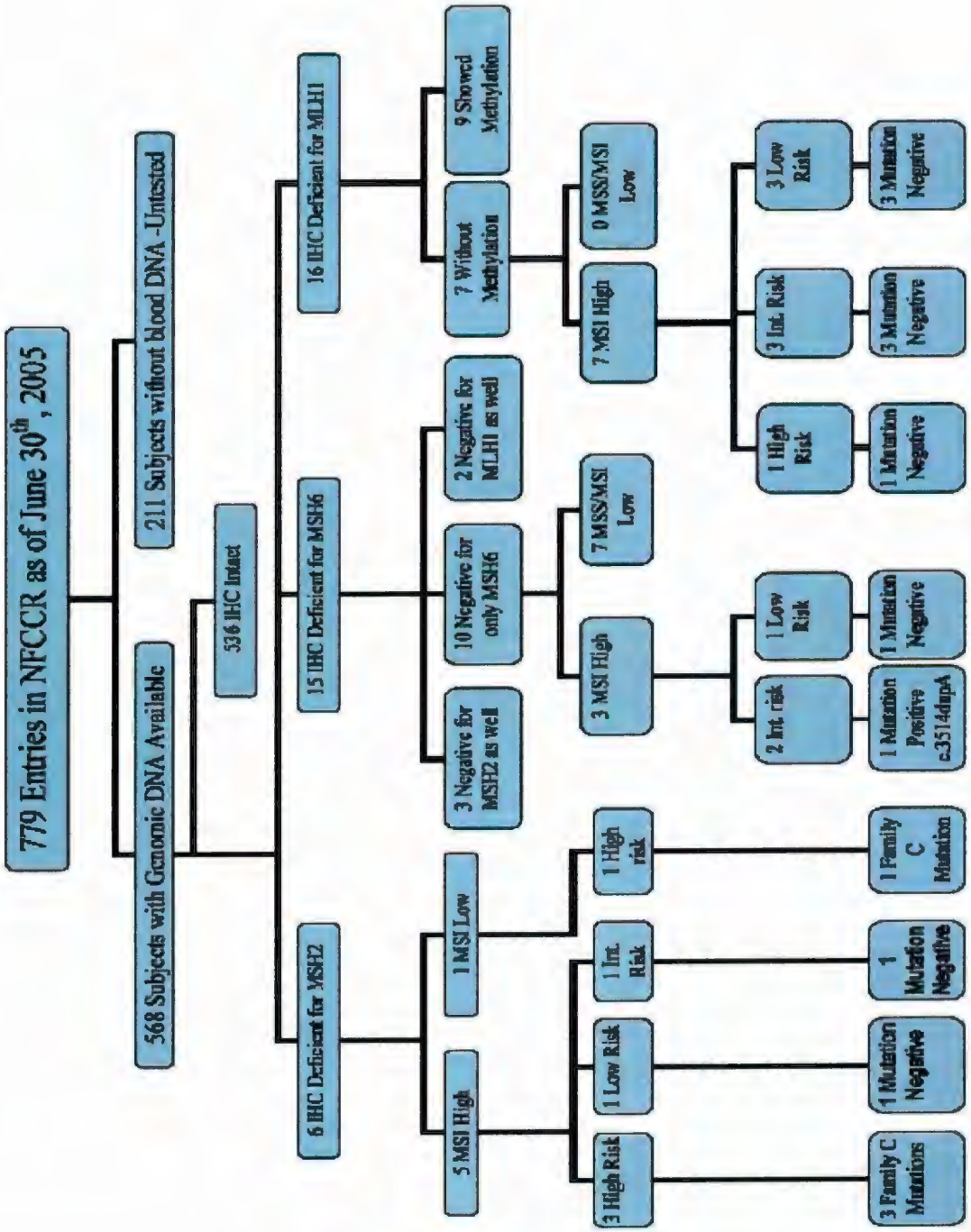

Figure 2.3: Molecular Testing Breakdown by IHC and MSI Status: Flow chart illustrating samples tested based upon IHC and MSI status.

2.4 Discussion:

All four samples from subjects that were at high family risk (i.e. AC1 or AC2), and were IHC deficient for MSH2 were found to carry the *MSH2* "Family C" mutation. One of these subject's tumour was MSI low but still IHC negative for MSH2. The other three samples were all MSI high. It should be noted that all four subjects who were positive for the "Family C" mutation came from families who were known from the genetics clinic to segregate this mutation.

Samples that were deficient for the MLH1 protein (16) can be broken down into those that displayed hypermethylation of the *MLH1* promoter (9) and those that did not (7) (Figure 2.3). Only samples that were not hypermethylated were sequenced. The hypermethylation status was determined by Dr. Roger Green's lab (Roger Green, Personal Communication). Promoter hypermethylation cannot be easily determined to be either mono-allelic (benign) or bi-allelic (inactivating the gene). It was assumed at the time that any samples that were not of high familial risk and displayed methylation were bi-allelic. All of the samples that were deficient for MLH1 and not methylated were MSI high (7). One sample was from an individual with family AC1 and MSI-H risk status but no mutation could be identified in the subject.

Of the 10 MSH6-only deficient samples, three were MSI high and seven were MSI low, or MSS (Figure 2.3). There were no subjects from high risk families among the MSI high samples, but there were two of intermediate risk as defined by the Revised Bethesda Criteria. One of these samples was shown to contain an *MSH6* mutation that was found by another researcher working on the same samples. The *MSH6* mutation was

identified in a subject who was negative for the MSH6 protein, MSI high and Bethesda 1 (See Appendix 1). This mutation was the duplication of an adenine base at position 3514 in exon 6 (c.3514dupA) which caused a frameshift resulting in a premature stop five codons downstream (p.Arg1972LysfsX5) (Amanda Dohey, Personal Communication). This mutation has been previously described in two unrelated families as pathogenic (Wijnen *et al.*, 1999, Plaschke *et al.*, 2004). No other mutations were identified in the other nine MSH6 negative samples.

There were several samples in the NFCCR that were MSI high but IHC intact. Since the completion of this thesis' work IHC has been performed for the PMS2 protein on these samples, and those deficient were then sequenced for *PMS2* mutations by Dr. A. de la Chapelle of the University of Ohio. Subsequently, two *PMS2* mutations were identified in three subjects (Ban Younghusband, Personal Communication). Investigations into the *MYH* gene responsible for *MYH*-associated polyposis (MAP) undertaken after the completion of this thesis' work identified the *MYH* mutations. Two of these subjects were microsatellite stable, while the other had no MSI data available. There were 18 other subjects that were heterozygous for known mutations in *MYH* (Roger Green, Ban Younghusband, Steve Gallinger, Personal Communication). In addition to this *MSH6* mutation, four more "Family C" mutations, and one deletion of exon eight of *MSH2* (c.1277-?_1386+?del, p.Lys427_Gln462delGlyfsX4), a known founder mutation in the Newfoundland population, two mutations have been identified in *PMS2* and three subjects have been identified as homozygotes for know deleterious *MYH* mutations.

Based upon the results of this study, it appears that CRC caused by Lynch Syndrome mutations in the Newfoundland population correlates with a recent USA population based study that found the total incidence of Lynch Syndrome was 2.2% of CRC. This was reported as the highest incidence found in a USA population (Hampel *et al.*, 2005). With the two *PMS2* mutation carriers identified, the *MSH6* mutation, plus the nine *MSH2* mutations found, the frequency of Lynch Syndrome in the NFCCR is 1.8% of the 740 subjects with genomic DNA available. No mutations were found in the seven samples which were IHC deficient for MLH1 and not methylated. The majority (16 of 23) of samples sequenced that displayed IHC deficiency for MLH1, MSH2 or MSH6 contained polymorphisms, regardless of family history or MSI status.

There are several reasons for not finding a greater number of mutations in both the NFCCR and the Newfoundland population in general. First and most important is that this study selected people for MMR mutation testing only from information in the NFCCR database. This registry only includes individuals that had CRC in a specific time interval and agreed to participate in the study. The registry excludes those who declined to participate in the study, and also excludes those patients with Lynch Syndrome who may only have had extra-colonic tumours. Endometrial cancers are more common than CRC in some kindreds and these would not have been included in the NFCCR (Quehenberger *et al.*, 2005; Stuckless *et al.*, 2006). Also, patients who agreed to participate but subsequently died or were too sick to provide a blood sample limited testing again. These are factors which cannot be overcome, and hopefully only exclude a very small percentage of Lynch Syndrome mutation carriers in the Newfoundland

population.

A second explanation for the lower than expected number of mutations identified is that there may be intronic mutations that activate cryptic splice sites that were not sequenced due to distance from the intron exon boundary. A novel intronic variant, c.1759+46A>T, was found in *MSH2* in a subject who was a "Family C" mutation carrier. It is possible that this variant was introduced via recombination or a spontaneous event onto the same chromosome as the "Family C" mutation in this particular family branch. This is likely a polymorphism, as previous evidence has shown that two mutations in *MSH2* can cause a phenotype of NF1 (Whiteside *et al.*, 2002), while this patient showed only a LS phenotype. If both variants were inactivating the same allele the NF1 phenotype would not be present as well. It must be noted that this variant is possibly pathogenic, but doubtful. Another novel intronic variant was found in *MLH1*, c.677+72G>T, whose significance is still unknown. This subject's tumour was IHC deficient for MLH1, MSI high, and not methylated. This subject was of Bethesda 1 risk status indicating colorectal cancer under the age of 50. Both of these intronic variants are far from the splice sites but may activate cryptic splice sites. Future research in the form of expression and functional studies, are needed before these variants can be confidently classified as polymorphisms.

Thirdly, there may be an upstream promoter mutation that would have been missed in routine sequencing. There has previously been an *MLH1* promoter mutation detected in the Newfoundland population by Green *et al.* (2003). Since this particular mutation was known, it was screened for in all samples that were sequenced for *MLH1*

but was not identified in any subjects.

Fourthly, IHC staining that did not correspond with MSI data was repeated and some samples were reclassified. If the correct IHC status was known at the time of this thesis' work, more samples may have been included, and the number of mutations identified may have increased. It must be noted however that ongoing testing of samples from the NFCCR using the updated information has identified no other mutations besides the *MSH6* mutation previously mentioned in this thesis (Amanda Dohey, Personal Communication)

Due to these reasons, I suspect the incidence of Lynch Syndrome in the Newfoundland population is higher than 1.8% making this study an underestimate of Lynch Syndrome in the Newfoundland population. For example, there were 7 samples that were MLH1 deficient, MSI high, did not display hypermethylation and in which no mutation could be identified. There were also two samples which were deficient for MSH6 protein, MSI high and no mutation found. These samples indicate a mismatch repair deficiency of a currently unknown cause, and may include intronic variants affecting splicing which may be missed by exonic sequencing. If mutations could be identified in these 9 samples high risk in addition to the 12 mutations found, the incidence of Lynch Syndrome in NFCCR would increase to 3.9%.

As well as Lynch Syndrome, it is possible that there are other cancer syndromes causing the increased CRC seen in the Newfoundland population. The three subjects homozygous for *MYH* mutations indicated that at least some of this CRC burden is caused by MAP. All samples with genomic DNA in the NFCCR were tested for common

MYH mutations via an allele specific PCR test. Samples that contained a variant were then screened by Denaturing High Performance Liquid Chromatography (DHPLC), a technique which compares migration rates of DNA through a gel and detects variations. Those samples which displayed variants were then sequenced to identify the DNA change. The possibility cannot be excluded that there are novel genes causing CRC in Amsterdam Criteria families via Familial Colorectal Cancer Type X (Lindor *et al.*, 2005). These would have been missed entirely by this study.

In the future, samples from this study which are negative for *MSH2*, *MLH1*, *MSH6* and *PMS2* mutations, at high family risk, and have tumours that are MSS may be useful in genome wide scan studies for detection of novel CRC causing genes. While the high prevalence of CRC in Newfoundland is possibly the result diet and/or a population collection of modifying alleles or low penetrance mutations acting together in addition to Lynch Syndrome, there may be novel genes responsible for Familial Colorectal Cancer Type X phenotype. There have been 15 cases in the NFCCR classified as ACI or ACII and without MMR deficiency (MSS) indicating Familial Colorectal Cancer Type X (Green *et al.*, 2007).

2.4.1 Mutation or Polymorphism?

A recurring issue with the DNA sequencing data generated from this project was that a significant number of variants identified were of unknown effect. Variants that were found could not easily be interpreted as pathogenic or benign. Often, no other family members were available for segregation analysis. Each variant was searched for in

the literature individually and it is very likely that some data were missed. The only functioning MMR variant database at the time, was hosted by the International Society for Hereditary Gastrointestinal Tumours (InSiGHT) on their website. This database includes submitted data only, lacking the majority of variant information concerning LS in the published literature. Therefore, it was decided to construct a database of all published MMR variants as part of this thesis.

Chapter 3

3.0 Mismatch Repair Variant Database

3.1 Introduction:

During mutation screening, missense and other changes were found which could not be easily classified as either pathogenic or polymorphic. Each of these unclassified variants (UVs) required a search in the Lynch Syndrome variants database from www.insight-group.org or an extensive review of the published literature if the genomic variant was not listed in the database. The InSiGHT database is a compilation of submissions of variant data only. While it contains numerous unpublished variants in addition to those published, the submission-only nature of the database means it is largely incomplete. A review by the Human Genome Organisation Society (HUGO) was completed in 2002 on MMR genes *MLH1* and *MSH2* (Mitchell *et al.*, 2002). This study reviewed all papers published up to and including December 31st 2001 and contained all published variants to this date as well as all the unpublished variants in the InSiGHT database. Upon searching for another database or review which was both current and comprehensive, it was found there was none.

To provide a resource for ourselves and for other Lynch Syndrome researchers, the task of constructing a complete, current and comprehensive Lynch Syndrome variant database was undertaken. For the purpose of this thesis, a variant is defined as a genomic alteration reported in the literature which is different from the reference sequence from

ensembl for each gene. The database was to consist of all variants published in the mismatch repair genes *MLH1*, *MSH2* and *MSH6* which were estimated to account for 95-100% of all known Lynch Syndrome variants. Every paper that was published on each of these genes was reviewed, and any paper that was found to contain or possibly contain a variant was catalogued and collected for reference and further review (so as not to miss any potential data). From this catalogue, a variant database was developed with an accompanying website allowing online access.

For the purposes of this thesis, mutation is defined as a genomic change from the reference sequence, know to cause LS. A variation/variant is defined as a genomic change from the reference sequence of unknown or unclear significance. A polymorphism is defined as a genomic change from the reference sequence which is not associated with any effect.

3.2 Materials and Methods:

The three major mismatch repair genes, *MLH1, MSH2*, and *MSH6* are suggested to account for 95-100% of all variants detected in Lynch Syndrome cases, making these genes the obvious choice for the variant database. Using Reference Manager v.10, PubMed was searched for each of these gene names, as well as each alternative name found using NCBI Entrez Gene. For *MSH2*: *MSH2, hMSH2, FCC1, COCA1,* and *HNPCC1* were used. For *MLH1*: *MLH1, hMLH1, FCC2, COCA2,* and *HNPCC2* were used. For *MSH6*: *MSH6, hMSH6, GTBP, HSAP,* and *HNPCC5* were used (Table 3.1). This thesis' work reviewed all the publications up to August 15th, 2005 but the database has been maintained and updated since by Amanda Dohey.

Table 3.1: Alternative MMR gene names

Gene	Keywords
MLH1	MLH1, hMLH1, FCC2*, COCA2* and HNPCC2*
MSH2	MSH2, hMSH2, FCC1*, COCA1* and HNPCC1*
MSH6	MSH6, hMSH6, GTBP, HSAP* and HNPCC5*
PMS2	PMS2, hPMS2, HNPCC4*, PMS2CL* and PMSL2*

* -These names are no longer current and did not result in any publications

All gene searches were then combined into a single list using Reference Manager including complete reference, PubMed hyperlink and abstract. Each abstract was then reviewed carefully for indications of sequence variants. If this was unclear from the abstract alone, the paper was read in PDF form if available, or it was marked to be acquired in print form. All references that either stated they contained variant data or

could possibly contain data of human MMR variants were then combined into a new reference list in which each paper was either downloaded in PDF format or acquired in a hard copy, numbered and assigned a unique reference ID number in Reference Manager. Each paper was then reviewed fully for variants. A small percentage (~6%) of papers were in a language other than English. Some of these listed the variants in proper nomenclature and were able to be transcribed directly into the database. Others required the assistance of foreign language translators who volunteered their time.

For each variant reported, the nucleotide number, codon, and base was confirmed against the published sequence at www.ensembl.org, with any deviations noted in the database. Secondly, the putative protein change, if applicable was checked against the published sequence and any errors noted. After errors were noted, the variant was recorded in the database according to the most recent and updated nomenclature available online at: http://www.genomic.unimelb.edu.au/mdi/mutnomen. This nomenclature is compiled and maintained by the Human Genome Variation Society (HGVS) and is updated regularly. Due to the constant change in the nomenclature, there are frequent differences in the published literature and the HGVS recommendations depending on the publication date. Previously correct nomenclature was updated to the current standard and not noted in the database, as the majority of references before the year 2000 use nomenclature that is outdated and equivocal. Fortunately, most recent publications use current nomenclature which is important for maintaining a standard for comparison of current and future references.

3.3 Results:

PubMed was searched for publications by each of the alternate names for the three major MMR genes. It was found that some of the older names for each of the genes did not yield any results, so they were omitted from future searches. The numbers presented here are current up to June 27th, 2007. For *MSH2*, only the search terms *MSH2* and *hMSH2* were used yielding 2217 abstracts, for *MLH1*, only the search terms *MLH1* and *hMLH1* were used yielding 2394 abstracts. For *MSH6* it was found that in addition to *MSH6* and *hMSH6*, the original name of the gene, *GTBP* also yielded results, requiring the inclusion of this search term. For *MSH6* there were 753 abstracts.

In total, 1306 papers have been identified which contained mutation information and were reviewed thoroughly for MMR mutations. This includes 579 for *MLH1*, 559 for *MSH2*, and 198 for *MSH6*. There were 942 unique variants for *MLH1*, 924 unique for *MSH2* and 281 unique for *MSH6*. The current updated version of the database is available online at: http://www.med.mun.ca/mmrvariants, a screenshot of the home page is seen in Figure 3.1. Genomic variant classification and distribution is summarised in Figures 3.2-3.9. Since 2005, the database has been updated and maintained with the addition of variants from *PMS2*.

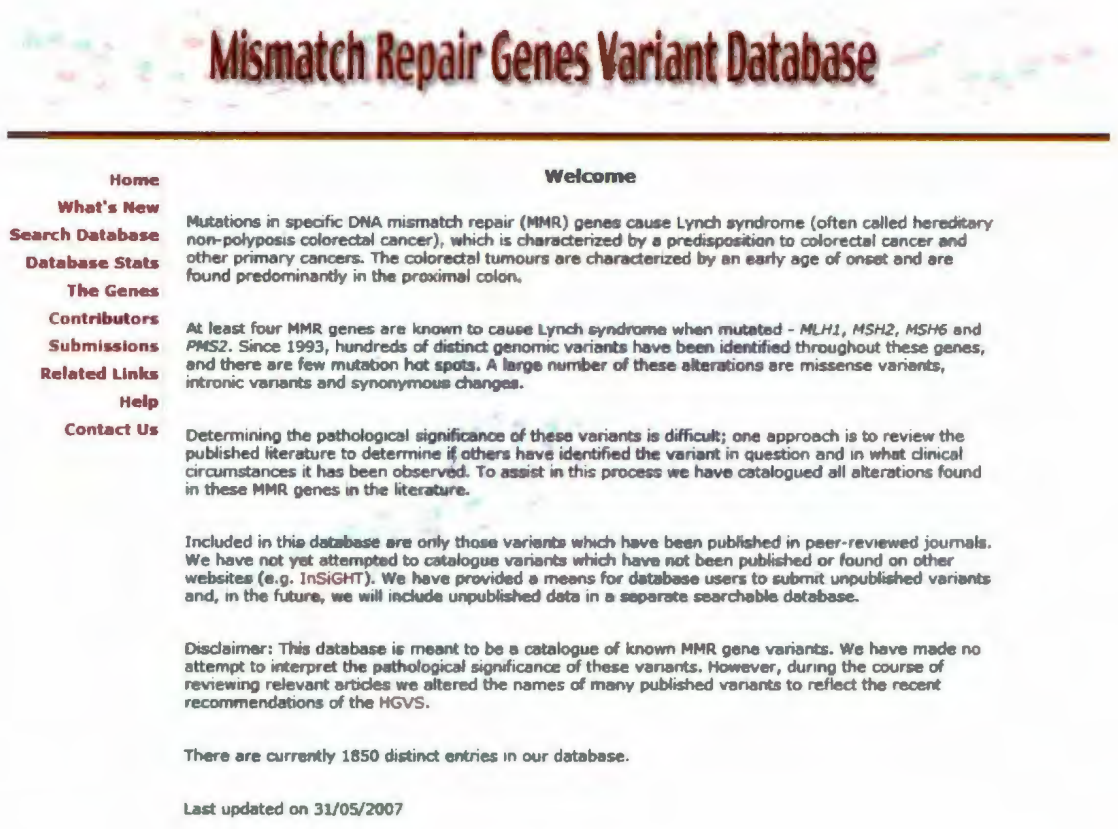

Figure 3.1: Screenshot of the Mismatch Repair Genes Variant Database Website

3.3.1 Exonic Distribution of Unique Variants by Gene:

A "unique variant" was defined as a singular genomic DNA deviation from the reference sequence. Regardless of how many times it had been reported, it is recorded once in these figures. The distribution of unique variants in *MLH1* showed a skewed distribution. Exon 12 had the highest number of variants with 58 unique variants reported while exon 16 had the second largest report of unique variants at 54. Variants in exons 7 and 15 were rare with 25 and 22 unique variants respectively (Figure 3.2). When comparing the distribution of unique variants and the size of the exons (Figure 3.3), it is noted that exon 7 had the most reported variants for its size (58.14 variants/100 bases), while exon 4 had the second highest (55.41 variants/100 bases). Intronic variants were common, with introns 8, 9, 13 and 15 most frequently containing variants.

MSH2 displayed the most variants in exon 3 with 84 unique entries. Exon 12 had the next highest number of variants with 66, followed by exon 13 with 62 (Figure 3.4). When comparing the number of unique variants for the size of the exon, it is seen that exon 11 shows by far the most changes with 53.06 variants per 100 bases, while the final exon in *MSH2*, exon 16, has the least variants at 6.32 variants per 100 bases (Figure 3.5). Overall, the distribution of variants was more regular than those of *MLH1* when comparing unique variants her 100 bases of exon. In general, intronic variants were fewer in *MSH2* as compared with *MLH1*.

Variants in *MSH6* showed an expected skewed distribution of unique variants in exon 4 due to its large relative size. Exon 4 has 112 individual variants, while the next most heterogenous exon, exon 5, only contained 27 unique variants (Figure 3.6). This can

be explained by the relatively large size of exon 4 which has 2645bp when compared to the other exons which range from 82bp to 473bp in size. The next largest exon, exon 1, only has 473bp. When comparing unique variants per 100bp, exon 4 is not found to be more frequently changed (Figure 3.7).

3.3.2 Distribution of Unique Variants by Type:

Differences between the types of variants seen in each of the different genes is apparent as well. Missense changes were slightly more common in *MSH6* and *MLH1* accounting for 27% and 24% of unique variants respectively, and only 17% of *MSH2*. Nonsense changes were evenly distributed among all three genes with 11% in *MSH6*, 7% in *MLH1* and 10% in *MSH2*. Insertions and deletions were slightly more common in *MSH6* (37%) than *MLH1* (22%) and *MSH2* (25%). Silent variations only accounted for 9% of *MSH2* and of *MLH1* changes, yet were responsible for 21% of all *MSH6* changes. Unique variants that affected proper splicing were common in *MLH1*, accounting for 13% of variants, but were less common in *MSH2* at only 6%, and very rare in *MSH6*, accounting for only 2%. Large genomic rearrangements (deletion or duplication of one or more exons) were very common in *MSH2* (33%) and *MLH1* (25%) but were virtually absent from *MSH6* (2%). This is all summarised in Figures 3.8-3.10.

Upon examination of the percentage of transitions versus transversions for the three MMR genes, it was found that transitions (purine to purine or pyrimidine to pyrimidine changes) outnumbered transversions (purine to pyrimidine, or pyrimidine to purine changes) for all three genes, as expected. While there are twice as many possible

transversion changes due to the greater number of bases that can be substituted, transitions are generally more commonly found in genes. This is due to several reasons, including transition changes generally resulting in more conserved amino acid changes due to wobble in the genetic code allowing for silent mutations which are less likely to disrupt function, C–>T transitions occurring frequently in methylated CpG islands concentrated in promoter regions, and differential DNA repair systems which repair transversions more readily than transitions (Strachan and Read, 2004). The results are summarised as follows: for *MLH1* the breakdown was 54% transitions and 46% transversions; for *MSH2,* the breakdown was 55% transitions and 45% transversions; for *MSH6,* the breakdown was 60% transitions and 40% transversions.

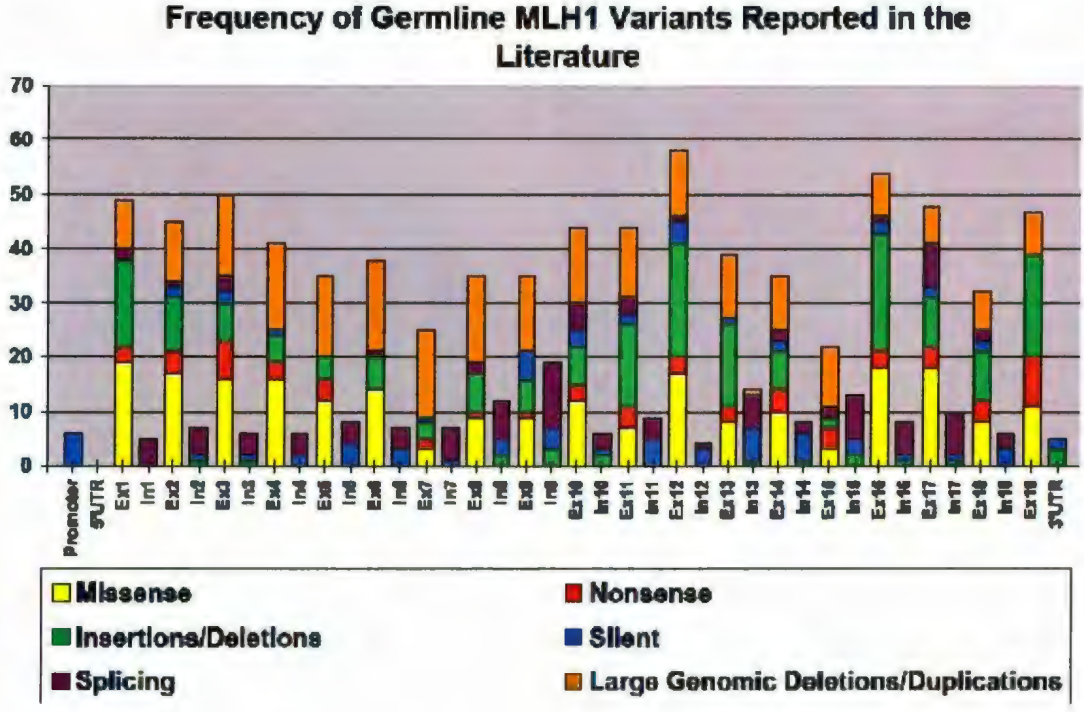

Figure 3.2: Frequency of Germline MLH1 Variants Reported in the Literature: Graphical representation of the breakdown of unique MLH1 variants by type and by location.

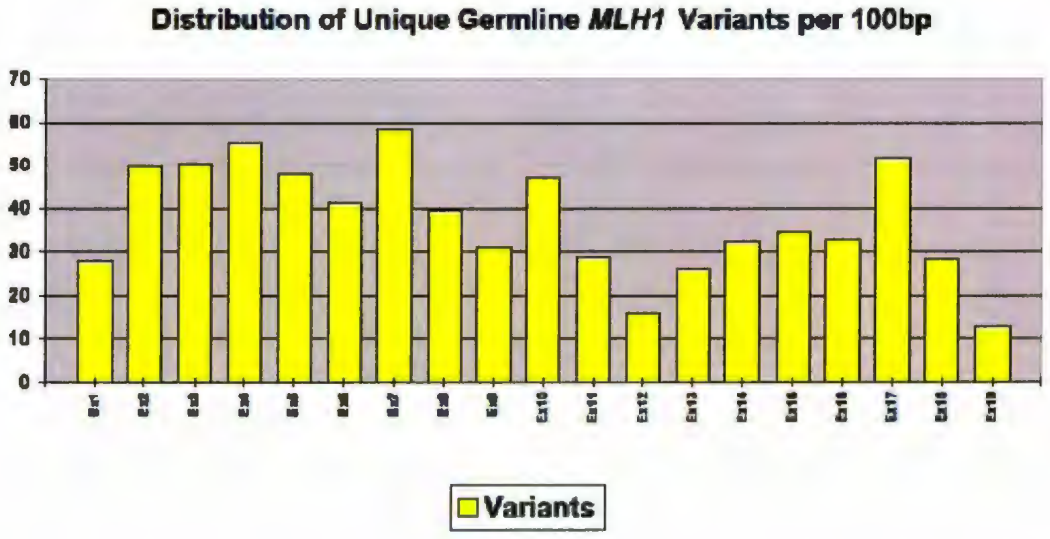

Figure 3.3: Distribution of Unique MLH1 Variants per 100bp: Graphical representation of the breakdown of unique MLH1 variants per 100 base pairs of each exon.

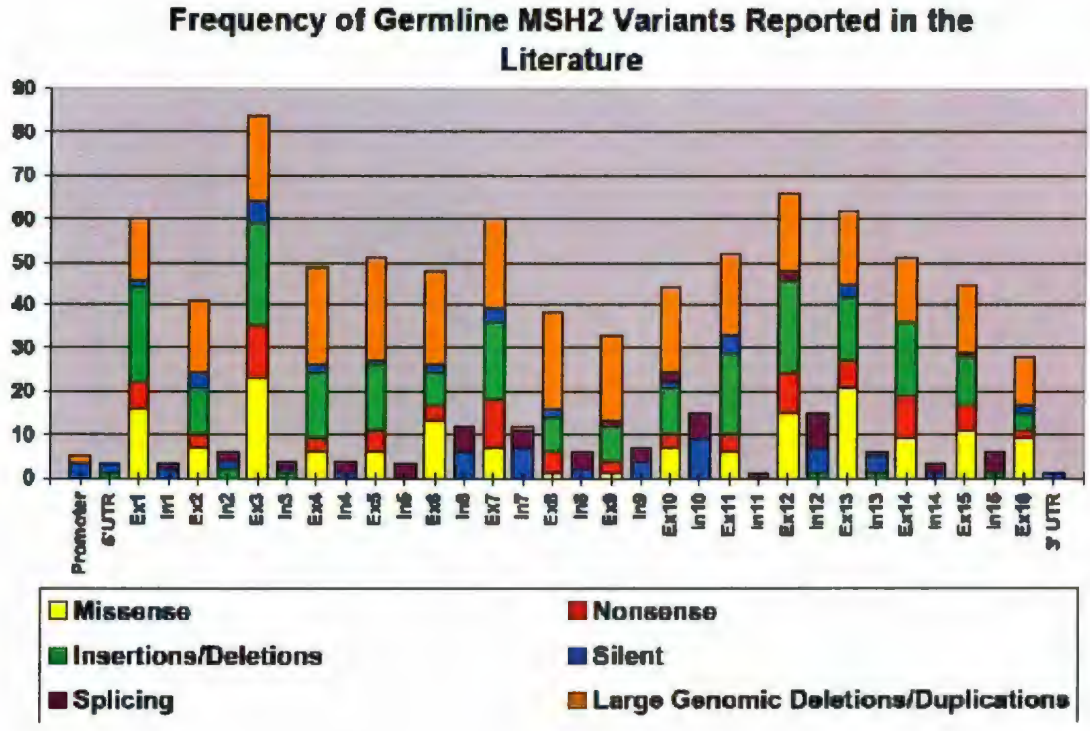

Figure 3.4: Frequency of Germline MSH2 Variants Reported in the Literature:

Graphical representation of the breakdown of Unique MSH2 variants by type and by location.

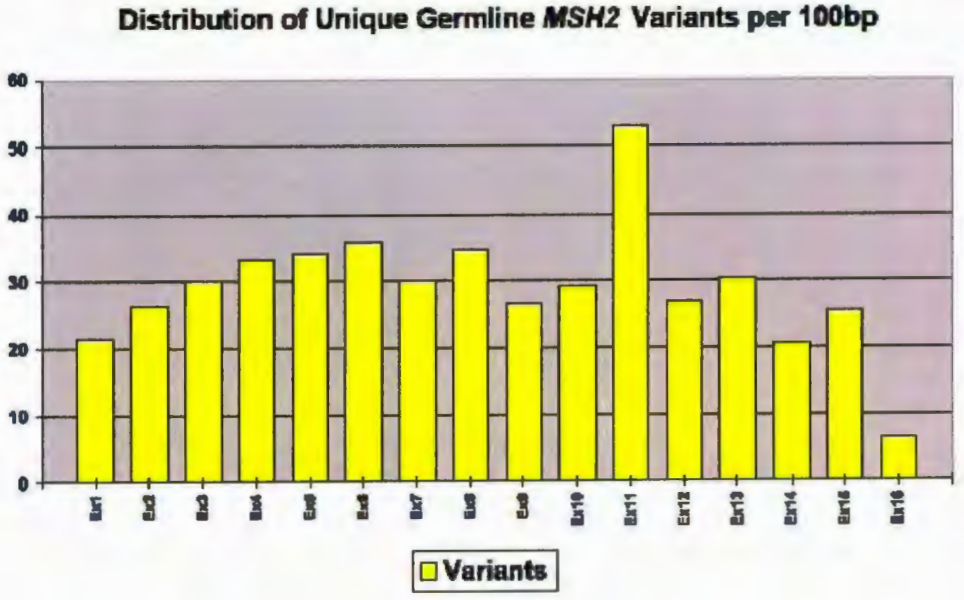

Figure 3.5: Distribution of Unique MSH2 Variants per 100bp: Graphical representation of the breakdown of unique MSH2 variants per 100 base pairs of each exon.

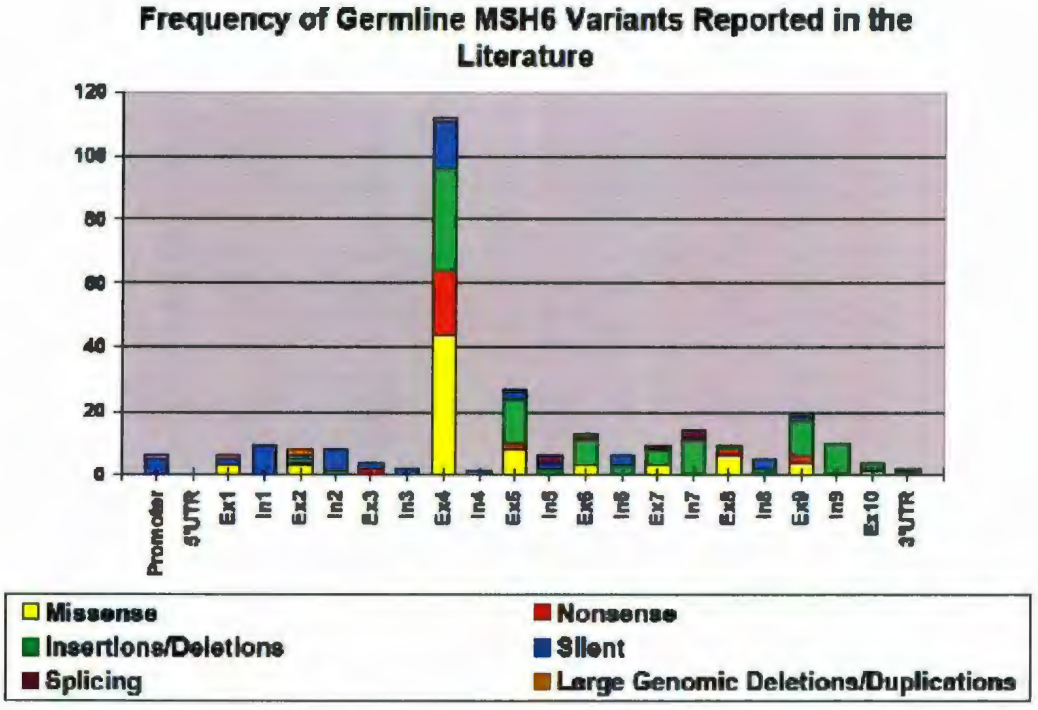

Figure 3.6: Frequency of Germline MSH6 Variants Reported in the Literature:

Graphical representation of the breakdown of unique MSH6 variants by type and by location.

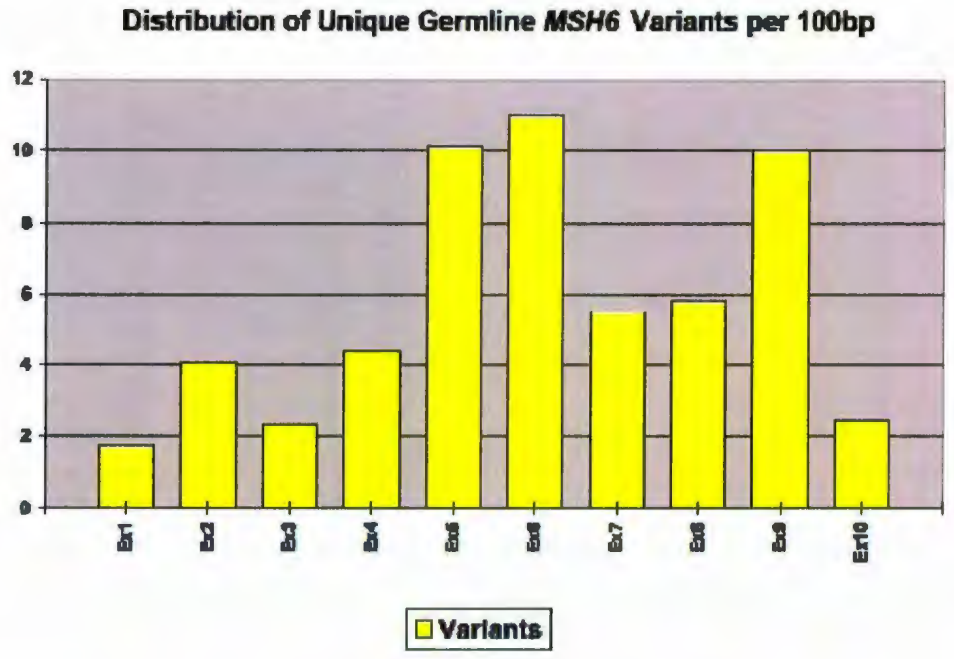

Figure 3.7: Distribution of Unique MSH6 Variants per 100bp: Graphical representation of the breakdown of unique MSH6 variants per 100 base pairs of each exon.

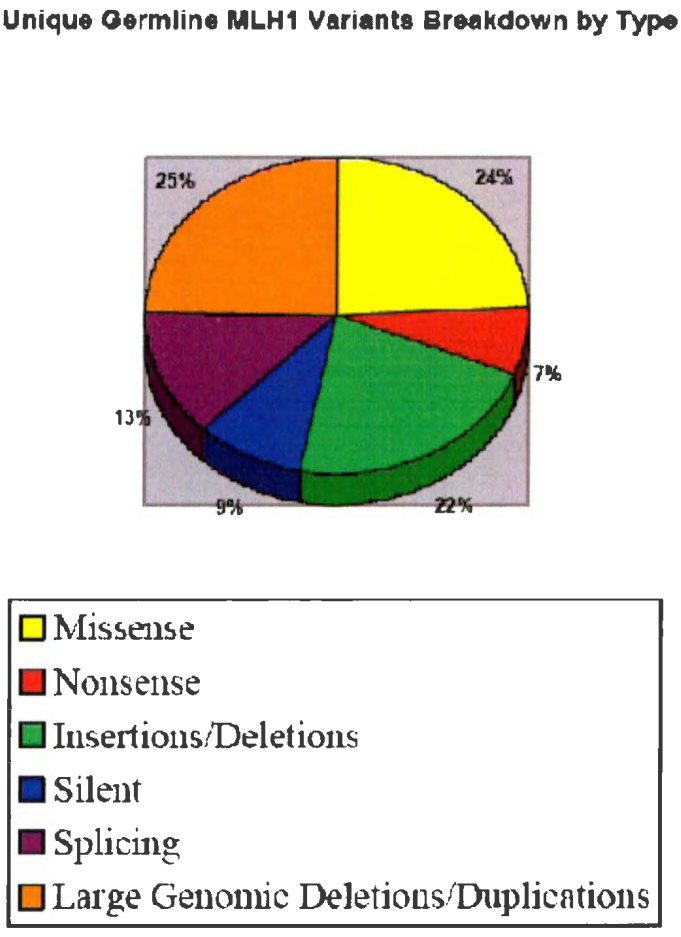

Figure 3.8: Unique Germline MLH1 Variants Breakdown by Type: Graphical representation of the breakdown of unique MLH1 variants as a percent of total, by type of variation.

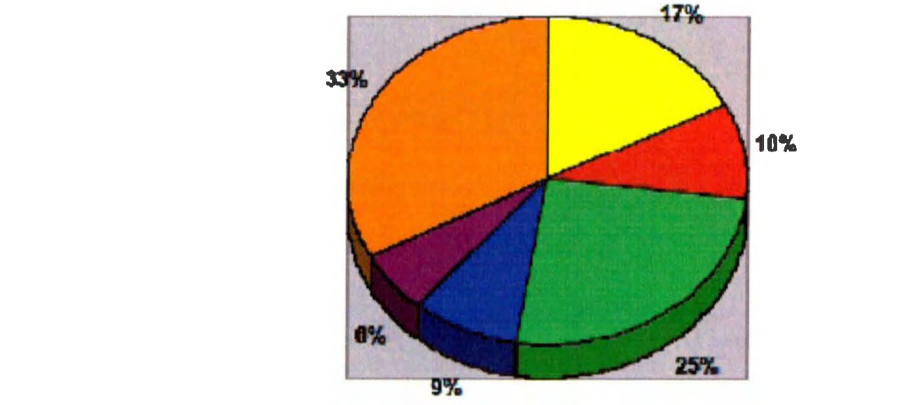

Figure 3.9: Unique Germline MSH2 Variants Breakdown by Type: Graphical representation of the breakdown of unique MSH2 variants as a percent of total, by type of variation.

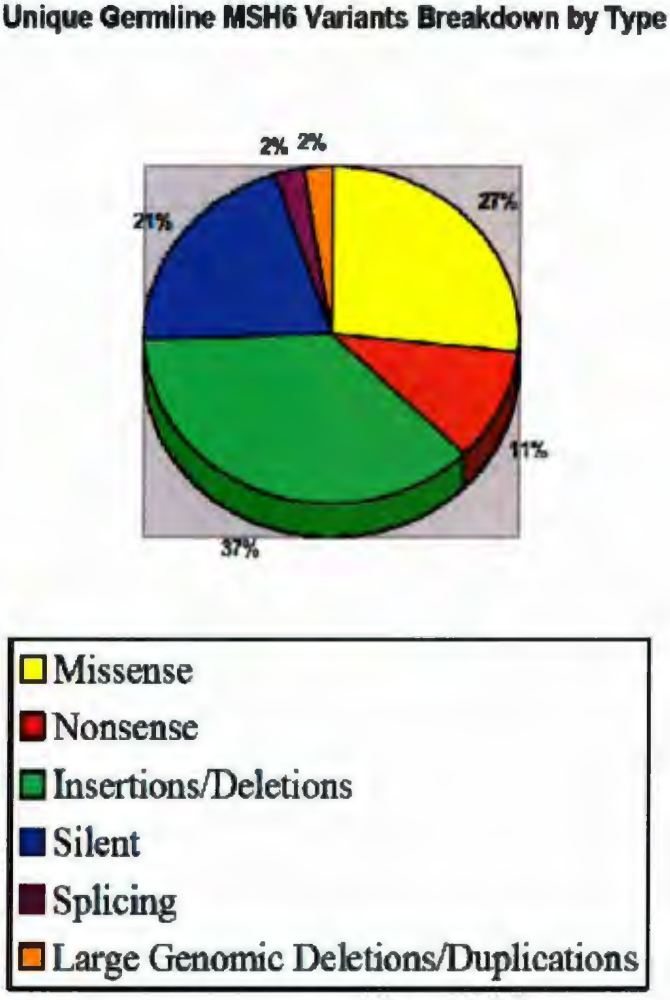

Figure 3.10: Unique Germline MSH6 Variants Breakdown by Type: Graphical representation of the breakdown of unique MSH6 variants as a percent of total, by type of variation.

3.4 Discussion:

There were a number of variants that were frequent in the database. For example the Newfoundland "Family C" mutation in *MSH2* (c.942+3A>T, p.Val265_Gln314del) has been reported in over 70 publications and this mutation has been previously reported to occur frequently *de novo* (Desai *et al.*, 2000). It is almost certain that some of these entries are from either the same family reported multiple times, or multiple seemingly distinct families sharing a common, but unknown, founder. It is impossible to determine from the literature what percentage of reported mutations are unique occurrences or the descendants of a single event.

Large genomic rearrangements are reported to be more common in *MSH2*, but there are similar unique changes between *MLH1* (33% of unique mutations for *MLH1*) and *MSH2* (25% of unique mutations for *MSH2*). The rarity of large genomic rearrangements in *MSH6* (2% of unique mutations for *MSH6*) could be due to the fact that it is less common for *MSH6* to under go genomic rearrangement, or probably due to minimal testing done on *MSH6* regarding rearrangements. Now that there is an MLPA kit available for testing of all exons of *MSH6*, the reported incidence of genomic rearrangements in *MSH6* may increase.

The number of intronic variants reported is most definitely under observed due to the fact that full introns are not normally sequenced. Introns are generally massive compared to the coding regions of genes, and since they are not expressed, variants in these regions are unlikely to be pathogenic. Intronic variants that interfere with splicing are notable exceptions to this, but most splicing mutations occur at the splice site

consensus sequences bordering exons, which are always sequenced due to their proximity to the exon.

During the construction of the MMR variant database, it was noted that a significant number of the reported alterations in the published literature were in error. This was more common in older papers (prior to 2000), but many recent papers carried significant errors as well. Mistakes in nucleotide and codon numbering in the paper compared to the correct reference were the most common. Other common errors included incorrect numbering of new stop codons generated by frameshifts or amino acid changes in missense mutations. All of these mistakes were the result of human error and are recorded in the comments section of the database for each error. Previously correct nomenclature was modified to the current standard, and was not noted in the database. Occasionally, gene names were mixed up and while a particular variant was labelled as being in one gene, examination of the data showed the variant to actually be in another gene. Errors in the published literature were sometimes, but rarely addressed *in errata* published after the initial publication.

Errors in peer reviewed publications imply they were missed by the original authors and reviewers as well as by publication editors. This points to the need for reviewers and writers of manuscripts to get back to basics and to check simple data output. Well-published authors in Lynch Syndrome were no exception to committing nomenclature errors. The literature review for this database highlights the importance of carefully checking and correcting variant and mutation information during the publication process. These errors should have been corrected before the paper was even submitted, let

alone published. Correspondence authors were sometimes contacted if the error was cryptic, but few responded.

The variant database has become a very popular tool since it has been available online. The site has received many positive emails from researchers around the globe and has had over 13,000 visits during its lifetime. There are links to the database on the InSiGHT group's website (in addition to their own database), as well as the HGVS list of human variant databases. The database will be maintained and updated monthly for new variants and current nomenclature by the lab of Dr. Ban Younghusband and Dr. Michael Woods. It appears that this new MMR database will be the principal resource for researchers and clinicians who require information concerning published Lynch Syndrome variants.

Chapter 4

4.1 Conclusions

_____The overall incidence of Lynch Syndrome found in this study of the NFCCR was 1.8% of colorectal cancer cases. A 1.8% incidence rate concurs with previous studies placing estimated incidence of Lynch Syndrome between 2-6% of CRC. There were several subjects whose tumours had molecular characteristics of LS (MSI and/or IHC negative for MMR protein) but in which no deleterious mutation could be identified. In addition, this study only examined in incidence of LS in CRC, excluding extra-colonic tumours that are frequently associated and sometimes more prevalent in LS. Due to these factors the author believes the reported 1.8% incidence LS found in this study is an underestimate of the true incidence of LS in the province. _____

_____The MMR variant database developed and used in this study proved to be an invaluable resource. It allowed the rapid identification and review of all previously published literature concerning LS variants. Searching the database for variants identified in this work, particularly intronic variants of unknown significance, allowed easy access to previous investigations into pathogenicity. This allowed for the efficient labelling of variants as polymorphisms, or pathogenic mutations. The initial creation of the database was labour-intensive and time-consuming but once established and the expertise developed, maintaining and updating was very efficient._____

References:

Aarnio M, Mecklin JP, Aaltonen LA, Nystrom-Lahti M, Jarvinen HJ. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. Int.J.Cancer 1995; 64: 430-433.

Al Tassan N, Eisen T, Maynard J *et al.*. Inherited variants in MYH are unlikely to contribute to the risk of lung carcinoma. Hum.Genet. 2004; 114: 207-210.

Bandipalliam P, Garber J, Kolodner RD, Syngal S. Clinical presentation correlates with the type of mismatch repair gene involved in hereditary nonpolyposis colon cancer. Gastroenterology 2004; 126: 936-937.

Bandipalliam P. Variability in the clinical phenotype among families with HNPCC--the potential importance of the location of the mutation in the gene. Int.J.Cancer 2007; 120: 2275-2277.

Bisgaard ML, Fenger K, Bulow S, Niebuhr E, Mohr J. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. Hum.Mutat. 1994; 3: 121-125.

Bronner CE, Baker SM, Morrison PT *et al.*. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994; 368: 258-261.

Cannavo E, Marra G, Sabates-Bellver J *et al.*. Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. Cancer Res. 2005; 65: 10759-10766.

Chan TL, Yuen ST, Kong CK *et al.*. Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. Nat.Genet. 2006; 38: 1178-1183.

Cunningham JM, Christensen ER, Tester DJ *et al.*. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. Cancer Res. 1998; 58: 3455-3460.

De Vos M, Hayward BE, Picton S, Sheridan E, Bonthron DT. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. Am.J.Hum.Genet. 2004; 74: 954-964.

Desai DC, Lockman JC, Chadwick RB et al. Recurrent germline mutation in MSH2 arises frequently de novo. J.Med.Genet. 2000; 37: 646-652.

Douglas JA, Gruber SB, Meister KA *et al.*. History and molecular genetics of Lynch syndrome in family G: a century later. JAMA 2005; 294: 2195-2202.

Fishel R, Lescoe MK, Rao MR *et al.*. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993; 75: 1027-1038.

Froggatt NJ, 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. J.Med.Genet. 1999; 36: 97-102.

Green J, O'Driscoll M, Barnes A *et al.*. Impact of gender and parent of origin on the phenotypic expression of hereditary nonpolyposis colorectal cancer in a large Newfoundland kindred with a common MSH2 mutation. Dis.Colon Rectum 2002; 45: 1223-1232.

Green RC, Green JS, Buehler SK *et al.*. Very high incidence of familial colorectal cancer in Newfoundland: a comparison with Ontario and 13 other population-based studies. Fam.Cancer 2006.

Hampel H, Frankel WL, Martin E *et al.*. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N.Engl.J.Med. 2005; 352: 1851-1860.

Hauser IJ, & Weller CV. A further report on the cancer family of Warthin. Am J Cancer 1936; 27:434-49.

Hughes MJ, Jiricny J. The purification of a human mismatch-binding protein and identification of its associated ATPase and helicase activities. J.Biol.Chem. 1992; 267: 23876-23882.

Kemp Z, Thirlwell C, Sieber O, Silver A, Tomlinson I. An update on the genetics of colorectal cancer. Hum.Mol.Genet. 2004; 13 Spec No 2: R177-R185.

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

Lindor NM, Rabe K, Petersen GM *et al.*. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. JAMA 2005; 293: 1979-1985.

Liu HX, Zhou XL, Liu T *et al.*. The role of hMLH3 in familial colorectal cancer. Cancer Res. 2003; 63: 1894-1899.

Lynch HT, Shaw MW, Magnuson CW, Larsen AL, Krush AJ. Hereditary factors in cancer. Study of two large midwestern kindreds. Arch.Intern.Med. 1966; 117: 206-212.

Lynch HT, Krush AJ. Heredity and adenocarcinoma of the colon. Gastroenterology 1967; 53: 517-527.

Lynch HT, Krush AJ. Cancer family "G" revisited: 1895-1970. Cancer 1971; 27: 1505-1511.

Lynch HT, Smyrk TC, Watson P *et al.*. Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. Gastroenterology 1993; 104: 1535-1549.

Manion JJ. The peopling of Newfoundland: essays in historical geography. Institute of Social and Economic Research, Memorial University of Newfoundland, St. John's, Newfoundland 1977.

Mitchell RJ, Farrington SM, Dunlop MG, Campbell H. Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review. Am.J.Epidemiol. 2002; 156: 885-902.

Muir EG, Bell AJ, Barlow KA. Multiple primary carcinomata of the colon, duodenum, and larynx associated with kerato-acanthomata of the face. Br.J.Surg. 1967; 54: 191-195.

Nicolaides NC, Carter KC, Shell BK, Papadopoulos N, Vogelstein B, Kinzler KW. Genomic organization of the human PMS2 gene family. Genomics 1995; 30: 195-206.

Palombo F, Gallinari P, Iaccarino I *et al.*. GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. Science 1995; 268: 1912-1914.

Palombo F, Iaccarino I, Nakajima E, Ikejima M, Shimada T, Jiricny J. hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. Curr.Biol. 1996; 6: 1181-1184.

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD. Mutation of a mutL homolog in hereditary colon cancer. Science 1994; 263, 1625-1629.

Papadopoulos N, Nicolaides NC, Liu B *et al.*. Mutations of GTBP in genetically unstable cells. Science 1995; 268: 1915-1917.

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

Plaschke J, Engel C, Kruger S *et al.*. Lower incidence of colorectal cancer and later age of disease onset in 27 families with pathogenic MSH6 germline mutations compared with families with MLH1 or MSH2 mutations: the German Hereditary Nonpolyposis Colorectal Cancer Consortium. J.Clin.Oncol. 2004; 22: 4486-4494.

Quehenberger F, Vasen HF, van Houwelingen HC. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. J.Med.Genet. 2005; 42: 491-496.

Ricciardone MD, Ozcelik T, Cevher B *et al.*. Human MLH1 deficiency predisposes to hematological malignancy and neurofibromatosis type 1. Cancer Res. 1999; 59: 290-293.

Rodriguez-Bigas MA, Boland CR, Hamilton SR *et al.*. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J.Natl.Cancer Inst. 1997; 89: 1758-1762.

Sieber O, Segditsas S, Knudsen A *et al.*. Disease severity and genetic pathways in attenuated familial adenomatous polyposis vary greatly, but depend on the site of the germline mutation. Gut 2006.

Strachan T, Read A. (2004) Human Molecular Genetics. (3rd ed.) New York: Garland Publishing

Stuckless S, Parfrey PS, Woods MO *et al.*. The phenotypic expression of three MSH2 mutations in large Newfoundland families with Lynch syndrome. Fam.Cancer 2006.

Thompson E, Meldrum CJ, Crooks R *et al.*. Hereditary non-polyposis colorectal cancer and the role of hPMS2 and hEXO1 mutations. Clin.Genet. 2004; 65: 215-225.

Torre D. Multiple sebaceous tumors. Arch.Dermatol. 1968; 98: 549-551.

Turcot J, Despres JP, St Pierre F. Malignant tumors of the central nervous system associated with familial polyposis of the colon: report of two cases. Dis.Colon Rectum 1959; 2: 465-468.

Umar A, Boland CR, Terdiman JP *et al.*. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J.Natl.Cancer Inst. 2004; 96: 261-268.

Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 1999; 116: 1453-1456.

Vasen HF, Stormorken A, Menko FH *et al.*. MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: a study of hereditary nonpolyposis colorectal cancer families. J.Clin.Oncol. 2001; 19: 4074-4080.

Wang L, Baudhuin LM, Boardman LA *et al.*. MYH mutations in patients with attenuated and classic polyposis and with youngApril 1, 2008-onset colorectal cancer without polyps. Gastroenterology 2004; 127: 9-16.

Wang Q, Lasset C, Desseigne F *et al.*. Neurofibromatosis and early onset of cancers in hMLH1-deficient children. Cancer Res. 1999; 59: 294-297.

Wang Q, Montmain G, Ruano E *et al.*. Neurofibromatosis type 1 gene as a mutational target in a mismatch repair-deficient cell type. Hum.Genet. 2003; 112: 117-123.

Warthin AS. Heredity with reference to carcinoma. Arch Intern Med 1913: 12:546-55.

Warthin AS. The further study of a cancer family. J Cancer Res 1925: 9:279-86.

Whiteside D, McLeod R, Graham G *et al.*. A homozygous germ-line mutation in the human MSH2 gene predisposes to hematological malignancy and multiple cafe-au-lait spots. Cancer Res. 2002; 62: 359-362.

Wijnen J, de Leeuw W, Vasen H *et al.*. Familial endometrial cancer in female carriers of MSH6 germline mutations. Nat.Genet. 1999; 23: 142-144.

Woods MO, Hyde AJ, Curtis FK *et al.*. High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. Clin.Cancer Res. 2005; 11: 6853-6861.

Woods MO, Williams P, Careen A *et al.*. A new variant database for mismatch repair genes associated with Lynch syndrome. Hum.Mutat. 2007; 28: 669-673.

www.stats.gov.nl.ca

www.cancer.ca

Yan H, Papadopoulos N, Marra G *et al.*. Conversion of diploidy to haploidy. Nature 2000; 403: 723-724.

Appendix 1: Bethesda and Revised Bethesda Criteria

Bethesda Criteria (Rodriguez-Bigas *et al.*, 1997):

Any one of these conditions may be met:

1. Individuals with cancer in families that meet the Amsterdam Criteria

2. Individuals with two HNPCC-related cancers, including synchronous and metachronous colorectal cancers or associated extracolonic cancers*

3. Individuals with colorectal cancer and a first-degree relative with colorectal cancer and/or HNPCC-related extracolonic cancer and/or a colorectal adenoma; one of the cancers diagnosed at age <45 y, and the adenoma diagnosed at age <40 y

4. Individuals with colorectal cancer or endometrial cancer diagnosed at age <45 y

5. Individuals with right-sided colorectal cancer with an undifferentiated pattern (solid/cribriform) on histopathology diagnosed at age <45 y†

6. Individuals with signet-ring-cell-type colorectal cancer diagnosed at age <45 y‡

7. Individuals with adenomas diagnosed at age <40 y

*Endometrial, ovarian, gastric, hepatobiliary, or small-bowel cancer or transitional cell carcinoma of the renal pelvis or ureter.

†Solid cribriform defined as poorly differentiated or undifferentiated carcinoma composed of irregular, solid sheets of large eosinophilic cells and containing small gland-like spaces.

‡Composed of ≥ 50% signet ring cells.

Revised Bethesda Guidelines (Umar *et al.*, 2004)

Tumors from individuals should be tested for MSI in the following situations:

1. Colorectal cancer diagnosed in a patient who is less than 50 years of age.

2. Presence of synchronous, metachronous colorectal, or other HNPCC associated tumors,* regardless of age.

3. Colorectal cancer with the MSI-H† histology‡ diagnosed in a patient who is less than 60 years of age.§

4. Colorectal cancer diagnosed in one or more first-degree relatives with an HNPCC-related tumor, with one of the cancers being diagnosed under age 50 years.

5. Colorectal cancer diagnosed in two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age.

*Hereditary nonpolyposis colorectal cancer (HNPCC)-related tumors include colorectal, endometrial, stomach, ovarian, pancreas, ureter and renal pelvis,biliary tract, and brain (usually glioblastoma as seen in Turcot syndrome) tumors, sebaceous gland adenomas and keratoacanthomas in Muir–Torre syndrome, and carcinoma of the small bowel.

†MSI-H microsatellite instability–high in tumors refers to changes in two or more of the five National Cancer Institute-recommended panels of microsatellite markers.

‡Presence of tumor infiltrating lymphocytes, Crohn's-like lymphocytic reaction, mucinous/signet-ring differentiation, or medullary growth pattern.

§There was no consensus among the Workshop participants on whether to include the age criteria in guideline 3 above; participants voted to keep less than 60 years of age in the guidelines.

Appendix 2: Sequencing Primers

***MLH1* Primers:**

Name:	Locus	Sequence 5'–>3'
E1FA	EXON 1	AGGCACTGAGGTTGATTGGC
E1RB	EXON 1	TCGTAGCCCTTAAGTGAGC
E2FA	EXON 2	AATATGTACATTAGAGTAGTTG
E2RB	EXON 2	CAGAGAAAGGTCCTGACTC
E3FA	EXON 3	AGAGATTTGGAAAATGAGTAAC
E3RB	EXON 3	ACAATGTCATCACAGGAGG
E4FC	EXON 4	GACCCAGCAGTGAGTTTTTC
E4RB	EXON 4	GATTACTCTGAGACCTAGGC
E5FA	EXON 5	GATTTTCTCTTTTCCCCTTGGG
E5RB	EXON 5	CAAACAAAGCTTCAACAATTTAC
E6FA	EXON 6	GGGTTTTATTTTCAAGTACTTCTATG
E6RB	EXON 6	GCTCAGCAACTGTTCAATGTATGAGC
E7FA	EXON 7	CTAGTGTGTGTTTTTGGC
E7RB	EXON 7	CATAACCTTATCTCCACC
E8FC	EXON 8	CAGCCATGAGACAATAAATCC
E8RD	EXON 8	GAAGCATAAAACAAGCCTGTG
E9FC	EXON 9	AGGACCTCAAATGGACCAAGT
E9RD	EXON 9	GGTCCCATAAAATTCCCTGTG
E10FA	EXON 10	CATGACTTTGTGTGAATGTACACC
E10RB	EXON 10	GAGGAGAGCCTGATAGAACATCTG
E11FA	EXON 11	GGGCTTTTTCTCCCCCTCCC
E11RB	EXON 11	AAAATCTGGGCTCTCACG
E12FE	EXON 12	ACAGAAGCTTGATGCATTTC

E12RF	EXON 12	AGAGAAGATGCAAGTGATTCA
E12FG	EXON 12	ATACAGACTTTGCTACCAGGACT
E12R11	EXON 12	GGGGTTGCTGGAAGTAGGTC
E13RC	EXON 13	TGATGCTATTGTGGGTTAGT
E13FA	EXON 13	TGCAACCCACAAAATTTGGC
E14FC	EXON 14	ACTTGGTGTCTCTAGTTCTGGT
E14RD	EXON 14	ACTACCTTCATGCTGCTCTC
E15FC	EXON 15	CCAACTGGTTGTATCTCAAGCAT
E15RB	EXON 15	CGGTCAGTTGAAATGTCAG
E16FA	EXON 16	CATTTGGATGCTCCGTTAAAGC
E16RB	EXON 16	CACCCGGCTGGAAATTTTATTTG
E17FA	EXON 17	GGAAAGGCACTGGAGAAATGGG
E17RB	EXON 17	CCCTCCAGCACACATGCATGTACCG
E18FC	EXON 18	AGTCTGTGATCTCCGTTTAG
E18RD	EXON 18	GATGTATGAGGTCCTGTCCTAGT
E19FA	EXON 19	GACACCAGTGTATGTTGG
E19RB	EXON 19	GAGAAAGAAGAACACATCCC
E19FC	EXON 19	GAGGCTTATGACATCTAATGT
E19RD	EXON 19	AAGAAATTATGTTAAGACACATCT

***MSH2* Primers:**

Name:	Locus	Sequence 5'–>3'
HE1FC	EXON 1	GGCGGGAAACAGCTTAGT
HE1RD	EXON 1	AAGGAGCCGCGCCACAA
H2A	EXON 2	CTTGAACATGTAATATCTCAAATC
HE2RB	EXON 2	CCCATTCTACTATCACAATCT
HE3FC	EXON 3	CATAGAGTTTGGATTTTTCCTTTTTGCT
HE3RD	EXON 3	CTAGGCCTGGAATCTCCTCTATCACTAG
HE4FA	EXON 4	ATTCCTTTTCTCATAGTAGTTT
HE4RB	EXON 4	TTGAGATAAATATGACAGAAATAT
HE5FA	EXON 5	CCAGTGGTATAGAAATCTTCG
HE5RC	EXON 5	TACCTGAAAAAGGTTAAGGGC
HE6FA	EXON 6	GAGCTTGCCATTCTTTCTAT
HE6RB	EXON 6	GGTATAATCATGTGGGTAAC
HE7FA	EXON 7	CTAAAATATTTTACATTAATTCAAG
HE7RB	EXON 7	ATGTGTCCTAAGAGTGAGTC
HE8FC	EXON 8	GATCTTTTTATTTGTTTGTTTTAC
HE8RD	EXON 8	AATATTACATCCACTGTCCAC
HE9FA	EXON 9	GTCTTTACCCATTATTTATAGG
HE9RB	EXON 9	GTATAGACAAAAGAATTATTCC
HE10FA	EXON 10	GTGAGTATGTTGTCATATAATAA
HE10RB	EXON 10	GCATTTAGGGAATTAATAAAGG
HE11FA	EXON 11	CATTATTTGGATGTTTCATAGG
HE11RB	EXON 11	CATGATTTTTCTTCTGTTACCA
HE12FC	EXON 12	TGTAAATTAGGAAATGGGTTTTGA
HE12RB	EXON 12	CAAAACGTTACCCCCACAA
HE13FA	EXON 13	CTTGCTTTCTGATATAATTTGTT

HE13RB	EXON 13	CATGAGAATCTGCAAATATACT
HE14FA	EXON 14	GGCATATCCTTCCCAATGTAT
HE14RB	EXON 14	AGTAAGTTTCCCATTACCAAG
HE15FA	EXON 15	TCTTCTCATGCTGTCCCCT
HE15RB	EXON 15	ATAATAGAGAAGCTAGTTAAAC
HE16FC	EXON 16	CTCATGGGACATTCACATGTGTTTCA
HE16B	EXON 16	TTAAGTTGATAGCCCATGG