ELUCIDATION OF THE OCCURRENCE OF EXTRACOLONIC CANCERS IN LYNCH SYNDROME

Kusha Kalideen

Supervisor:

Professor R. Ramesar

Department of Clinical Laboratory Science

Division of Human Genetics

Presented for M.Sc. (Med)

In the School of Biomedical Sciences

Faculty of Health Sciences

University of Cape Town

2008

DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and to pretend that it is ones own.

2. I have used the American Journal of Human Genetics convention for citation and referencing. Each significant contribution to, and quotation in, this essay/ report/ project/… from the work, or works, of other people has been attributed, and has cited and referenced.

3. This project is my own work **using my own words**.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing if off as his or her own work.

DATE……11 November 2008**……...**

STUDENT NAME……KUSHA KALIDEEN**……...**

STUDENT NUMBER……KLDKUS001**……..**

SIGNATURE……

Table of Contents

Acknowledgments

I would first like to acknowledge my parents for granting me amazing opportunities and for their everlasting encouragement, support and allowing us to pursue our dreams.

I am grateful to my supervisor Professor Ramesar for granting me the opportunity to be a part of his laboratory and for the constructive criticism during the process. I am also grateful to Dr Nicki Tiffin for her priceless help with the bioinformatics analyses.

I am indebted to Alvera Vorster and the members of the Division of Human Genetics for their invaluable advice and encouragement during my Masters.

I acknowledge my family (especially my eldest cousin) and the Ribeiro family for always believing in me and taking care of us. And to my sisters; Avika Kalideen and Marisa Ribeiro for keeping me grounded and trying to make me more assertive

To my friends for their support, encouragement, and always being available as shoulders to cry on or just to vent.

To the National research Foundation (NRF) for financially supporting me through this project.

List of abbreviations

°C – Degrees Celsius

α - Alpha

β- Beta

ACVR2 – Activin receptor type 2 A

AGE – Agarose gel electrophoresis

AR – Androgen receptor

ATP – Adenine triphosphates

BAX – Bcl associated protein X

bp – Base pairs

BRD8 – Bromodomain containing protein 8

CALS - Café au lait spots

Casp5 – Caspase 5

CCS – Childhood Cancer Syndrome

CFS – Cancer Family Syndrome

CI – Confidence Interval

cMs – Coding microsatellite

CRC – Colorectal Cancer

CYP4AII – Cytochrome P450 4AII

DNA – Deoxyribonucleic acid

dNTPs – Deoxyribonucleotide triphosphates

EDTA –

Erα – Estrogen receptor Alpha

Erβ – Estrogen receptor Beta

EtBr – Ethidium bromide

FAP – Familial adenomatous polyposis

GIST – Gastrointestinal Stromal Tumours

hMLH1 – Human Mut L homolog 1

hMSH2 – Human Mut S homolog 2

hMSH3 – Human Mut S homolog 3

hMSH6 – Human Mut S homolog 6

HIC1 – Hypermethylated in Cancer

HGNC – Human Genome Nomenclature Committee

HLA-DOA – HLA class II histocompatability antigen Do Alpha chain precursor

HLA-DMB – Major histocompatibility complex class II DMBeta

HNPCC – Hereditary Non-Polyposis Colorectal Cancer

HSSCC – Hereditary Site Specific Colorectal Cancer

IBD – Identity by descent

ICG-HNPCC – International Collaborative Group for the study of HNPCC

IGF2R – Insulin growth factor receptor type 2

MAML1 – Mastermind like protein 1

MBD4 – Methyl CpG binding domain protein

mg/ ml – Milligrams per millilitre

MLH – Mut L Homolog

mM - milliMolar

MMR – Mismatch repair

MSH – Mut S Homolog

MSI – Microsatellite instability

MSI-H – High microsatellite instability

MSI-L – Low microsatellite instability

Ms-MLPA – Methylation specific Multiplex Ligation dependant Probe Amplification

MSS – Microsatellite stable

MTS – Muire Torre Syndrome

ng – nanograms

ng/ µl – Nanograms per microlitre

NF1 – Neurofibromatosus Type 1

OR – Odds ratio

PCR – Polymerase Chain Reaction

PTEN – Phospatase and Tensin homolog

PTPRU – Receptor type tyrosine protein phosphatase U precursor

PvuII – Proteas vulgaris

RE – Restriction Endonuclease

RER – Replication error phenotype

RPM – Revolutions per minute

RR – Relative Risk

SLE – Systemic Lupus Erythromatosus

SNP – single nucleotide polymorphism

TAF1B – TATA box binding protein associated factor 1B

Taq – Thermus aquaticus

TBE – Tris Borate EDTA

TGFBR2 – Transforming Growth Factor Receptor type 2

TIMP3 – Tissue inhibitor of metalloproteases

Tm – Melting temperature

TP53 – Tumour suppressor protein 53

U – units

UCT – University of Cape Town

µg – micrograms

µl – Micolitre

µM – micromolar

UK – United Kingdom

USA – United Stated of America

V – volts

Y – Years

List of figures

List of tables

Plan of thesis

This thesis consists of five chapters.

Chapter One is a general background into Lynch syndrome, its molecular pathology and extracolonic manifestations. The aims and objectives of this thesis are also provided.

Chapters Two, Three and Four are the experimental chapters. Chapter Two investigates the contribution of a polymorphism in the *hMLH1* gene to the extracolonic cancer spectrum of Lynch syndrome in a large cohort. Chapter Three centres on the search for a uniform genetic signature between Lynch-associated malignancies. A panel of genes was assessed as well as Multiplex Ligation dependant Probe Amplification (MLPA) in a cohort consisting of germline DNA and their tumour DNA counterparts. Chapter Four focuses on the bioinformatics analyses of existing microarray data to supplement the study.

Finally Chapter Five provides an overview of the project and includes potential research prospects.

ABSTRACT

Lynch Syndrome, also known as Hereditary Non-Polyposis Colorectal Cancer (HNPCC) (OMIM #120435), is a familial disorder resulting from mutations within DNA MMR genes. Effective surveillance, diagnosis and treatment of the disorder is complicated due to the phenotypic and genetic heterogeneity of lynch syndrome, which exhibits an autosomal dominant mode of inheritance.

Aim: Determine the molecular pathology of extracolonic cancers in Lynch syndrome and to elucidate whether or not the occurrence of these extracolonic cancers are a direct result of the mismatch repair deficiency.

Materials and Methods: First, a modifier study was performed assessing the effect of a variant within the DNA MMR gene *hMLH1* in a cohort of individuals predisposed to Lynch syndrome in order to examine a potential epistatic effect in the gene. In order to obtain a genetic signature of Lynch associated tumours, germline DNA and corresponding tumour DNA was isolated from Lynch syndrome patients. The genetic material was assessed via a panel of microsatellite rich genes and MS-MLPA. Finally, *in silico* analyses were undertaken assessing microarray data from microsatellite unstable colorectal and endometrial cancers to characterise novel candidate genes.

Results: The modifier study did not prove fruitful as no association was found between the *hMLH1* promoter variant and site of cancer in individuals predisposed to Lynch syndrome. An association was observed heterozygous and homozygous variant genotypes and an increased risk of colorectal cancer, regardless of predisposing mutation ($p = 0.000181$). Two tumour suppressor genes; *HIC1* and *TIMP3*, were found to be methylated in the tumour samples in the germline/ tumour tissue study. This study also showed instability of the *Erβ* gene in the

majority of tumour samples. Bioinformatic analysis utilising existing microarray data resulted in common under-expression of four genes and common over-expression in three genes in microsatellite unstable colorectal and endometrial cancers.

Conclusion: Further investigation into the modifier study and elucidation of a genetic signature in MMR deficient, MSI cancers. The results obtained in this study contribute to the increasing body of knowledge in the field and the various stages of malignancies should be assessed for a more informative result. Genetic and functional studies should be performed on the information garnered from the bioinformatics analysis. Overall evaluation and molecular classification of Lynch syndrome tumours may guide better diagnosis, surveillance and treatment of those at risk.

Chapter 1:

Introduction

1.1. Lynch Syndrome

Lynch Syndrome, also known as Hereditary Non-Polyposis Colorectal Cancer (HNPCC) (OMIM #120435), is a familial disorder. It exhibits an autosomal dominant mode of inheritance and a penetrance of about 80% (Lynch and De La Chapelle 2003). As a result, multiple generations are affected with synchronous and metachronous, colorectal and extracolonic cancers (Douglas *et al.* 2005). Although the incidence of Lynch Syndrome in South Africa is unknown, it is likely to match the international figures which are reported to account for 2 – 5 % of all colorectal cancers (Lynch and De La Chapelle 2003, Cancer Research 2006).

While it is imperative to identify families with Lynch Syndrome (via presymptomatic genetic testing), management and surveillance of the disease is costly, invasive and not always accessible particularly in a resource-poor country. The disease could be detected early and life may subsequently be prolonged with rigorous screening of predisposed individuals. In a developing country, provision of the medical centres required for adequate screening and management procedures are not feasible. A recent study has shown that annual screening for these malignancies allows for a longer cancer-free lifespan and onset of severe disease, which can be more successfully treated and managed by preventing its escalation (Lynch and De La Chapelle 2003).

1.2. History of Lynch Syndrome

In 1913, Dr Albert Scott Warthin first described a family with numerous manifestations of colorectal, endometrial and gastric cancers (Douglas *et al.*

2005). Dr Henry T. Lynch, in 1966, further investigated and expanded on the initial diagnosis made by Dr Warthin (Lynch and Lynch 1997).

To further investigate the disease, an International Collaborative Group for the study of HNPCC (ICG-HNPCC) was formed in 1989 under the leadership of Dr Lynch. In 1990, part of the mandate of the ICG-HNPCC was to identify criteria through which diagnosis of Lynch syndrome could be made (ICG-HNPCC 2004). Known as the Amsterdam criteria, it was originally based on a detailed family history. The criteria subsequently evolved in order to incorporate the inclusion of extracolonic cancers found within the HNPCC tumour spectrum (Table 1) (Lynch *et al.* 2003).

Table 1: Criteria used to identify Lynch families (information obtained from ICG-HNPCC)

Amsterdam Criteria
1. At least three relatives with colorectal cancer, one of whom is a first degree relative of the others
2. Colorectal cancer involved at least two generations
3. One or more of the colorectal cancers presented before the age of 50 years (y)
Modified Amsterdam Criteria
1. At least three relatives diagnosed with Lynch-associated cancers (colorectal, endometrial, small bowel, ureter or renal pelvis cancer); one of whom is a first degree relative of the other two
2. Colorectal cancer involving two generations
3. One or more of the colorectal cancers presented before the age of 50y

These criteria resulted in the identification of large families with Lynch syndrome. It is unfortunately recognised that due to the stringency of the Amsterdam criteria, all families affected with Lynch syndrome may not be acknowledged. Subsequently further studies have found families who do not adhere to the previously defined criteria (Table 1) yet have been shown to carry mutations in genes, known to be implicated in Lynch syndrome. The discrepancy in the identification of families

arises as the families present with later ages of onset and variable penetrance of the disease (Fearnhead *et al.* 2002; Lynch and Lynch 2000; Gonzalez-Aguilera *et al.* 2003).

1.3. Genetic Basis of Lynch Syndrome

Mutations within the DNA Mismatch Repair (MMR) genes, which encode DNA MMR enzymes, are responsible for Lynch Syndrome. Collectively, the DNA MMR enzymes form the DNA MMR system which ensures the fidelity and integrity of the DNA sequence during DNA replication. This regulatory system scans the genome for mismatches, insertions and deletions during DNA replication and repairs any errors (Watson *et al.* 2004, Peltomaki 2005).

1.3.1. DNA MMR system

In 1975, DNA mismatch repair genes were first discovered and characterised in the bacterium, *Escherichia coli*. Between 1988 and 1991, human mismatch repair activity was demonstrated (Lynch and Lynch 1997) and the first DNA MMR gene was described in 1993 (Peltomaki *et al.* 1993). The human DNA mismatch repair genes are named after their bacterial homologs (Snustad and Simmons 2003).

The simplest DNA MMR system is found in prokaryotes and consists of a MutS protein, which scans replicating DNA for errors and a MutL protein, which facilitates the repair of errors (Jun *et al.* 2006; Marti *et al.* 2002). In contrast, higher organisms have several MutS (referred to as MSH) and MutL (referred to as MLH) homologs. For accurate functioning of the system, the DNA MMR homologs work in concert using each other as recognition sites and partners

(Chao and Lipkin 2006). The system is adenine triphosphate (ATP) dependant and the ATP binding sites on each protein are highly conserved in eukaryotes. Comprehensive reviews on this system are available (Kunkel and Erie 2005; Jun *et al.* 2006; Schofield and Hsieh 2003; Marti *et al.* 2002).

The function of the DNA MMR system in humans is schematically represented in Figure 1 below. A MSH protein, *MSH2*, comes into contact with newly synthesized double-stranded DNA and scans it for mismatches. Once an anomaly is encountered, *MSH2* recruits and binds to the appropriate MSH MMR protein in order to rectify the defect. If the mismatch is a single base-pair substitution or a one base-pair insertion or deletion in the DNA sequence, the *MSH2* protein recruits the *MSH6* protein. For insertions or deletions in the sequence from one to four base-pairs in length, *MSH2* recruits an *MSH3* protein. Thereafter, the primary MutL complex, a heterodimer consisting of the proteins *MLH1* and *PMS2*, recognizes the MutS complex and binds to it and catalyses the repair of the DNA (Chao and Lipkin 2006). Additional factors, (*EXO1, PCNA* and *MED1*), essential for effective repair of the DNA are recruited. However, if the DNA damage is too extensive and unable to be effectively repaired, the DNA MMR system induces the cell to undergo apoptosis via the p73 apoptotic pathway (Marti *et al.* 2002; Jacob and Praz 2002). *MLH1* and *PMS1* proteins also form a MutL complex however, the function of this complex is unknown in human cells (Jacob and Praz 2002).

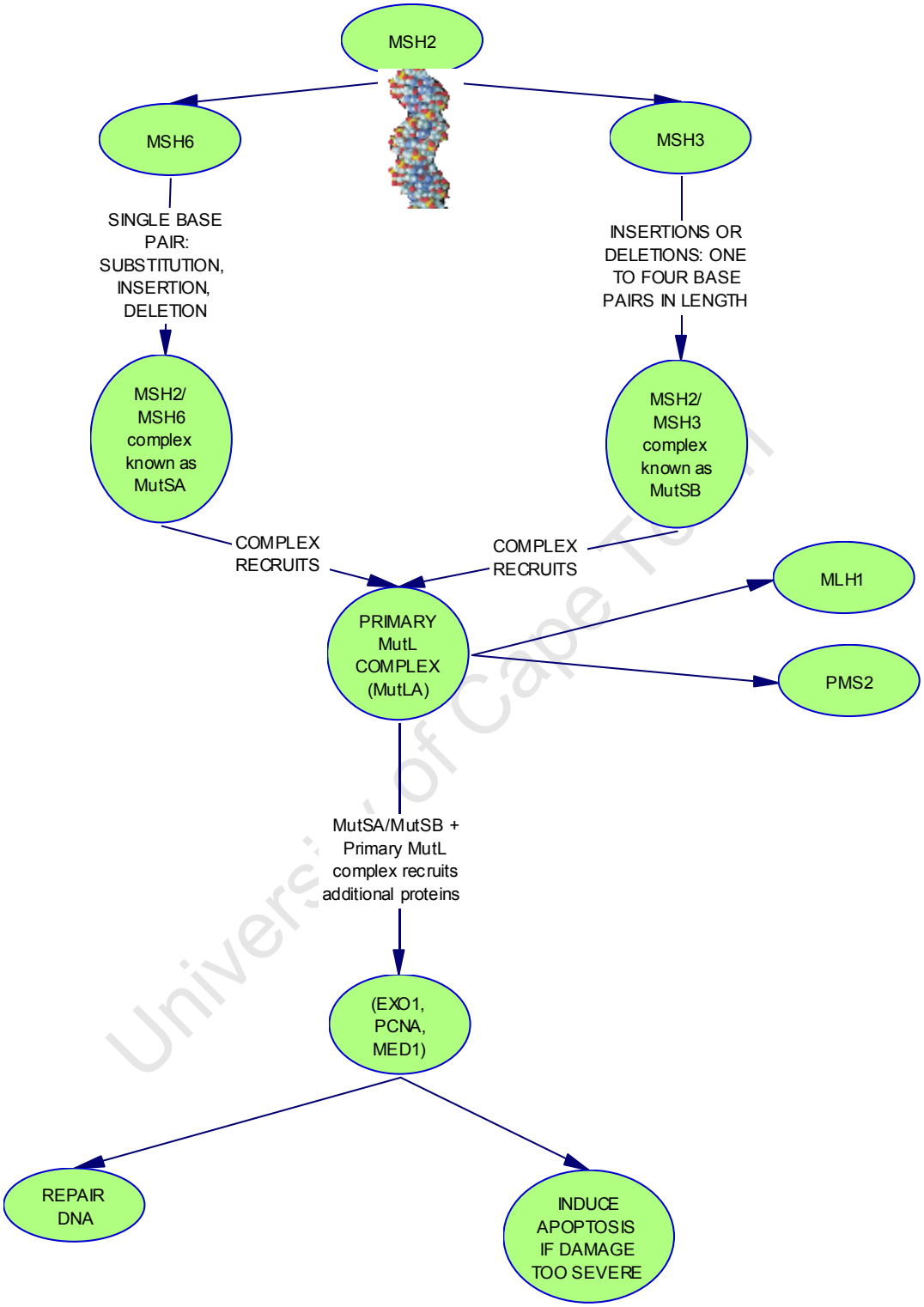


Figure 1: **Schematic representation of the function of the human DNA MMR system. Information obtained from Kunkel and Erie 2005; Jun *et al.* 2006; Schofield and Hsieh 2003; Marti *et al.* 2002; Jacob and Praz 2002.**

.

1.3.2 Defects in the MMR system

Mutations within DNA MMR genes predispose to a cancerous phenotype. The DNA MMR genes act as tumour suppressor genes and require both copies of the gene to be inactive to allow for initiation of malignancy (Chao and Lipkin 2006). Genetic heterogeneity is evident as a mutation in any one of the mismatch repair genes can result in the same phenotypic outcome.

1.3.2.1. Monoallelic Germline Mutations

A monoallelic germline mutation in a DNA MMR gene predisposes to Lynch Syndrome. Loss of the second allele, leads to the lack of expression or a defective protein, consequently resulting in the progression of a malignant phenotype. A major effort has been made internationally to identify disease causing genes and mutations in Lynch syndrome. To date, mutations within the following five DNA MMR genes have been implicated: *hMLH1*, *hMSH2*, *hMSH6*, *hPMS1, hPMS2, hMSH3* (Chao and Lipkin 2006; Lynch and Lynch 2000; Lynch and De la Chapelle 2003).

Mutations in *hMLH1* and *hMSH2* account for majority of disease in Lynch families. The remaining genes, *hMSH6*, *hPMS1* and *hPMS2*, are responsible for only about 10% of the disease (Rowley 2005). All mutations (insertions, deletions, duplications or substitutions) identified internationally with the DNA MMR genes contributing to Lynch Syndrome are deposited in the Insight Mutation database, (*http://www.insight-group.org*) (ICG-HNPCC 2004). Another MMR gene database containing all published mutations in each MMR gene is also available at the following URL*: http://www.med.mun.ca/MMRvariants/*.

1.3.2.2. Biallelic germline mutations

Although homozygous germline mutations within the DNA MMR genes are rare, there have been cases described involving DNA MMR genes *hMLH1*, *hMSH6* and more recently *hPMS2*.

The childhood cancer syndrome (CCS) comprises haematological malignancies, brain tumours and neurofibromatosus type 1 (NF1), evident in individuals with biallelic mutations in a DNA MMR gene (Ricciardone *et al.* 1999; Whiteside *et al.* 2002; De Vos *et al.* 2005; Felton *et al.* 2007; Poley *et al.* 2007; Tan *et al.* 2007; Kruger *et al.* 2008). The appearance of Café-au-lait spots (CALS) at a young age may lead to a misdiagnosis of classic NF1. The current hypothesis with regards to the occurrence of CALS is that during embryonic development, the NF1 gene is more susceptible to mutations as a result of deficient mismatch repair due to the large size of the gene. This is likely to result in the NF1 phenotype.

The composite CCS syndromic phenotype usually has patients presenting with a haematological or brain malignancy in childhood and progressing to colorectal cancer into adolescence. Due to the germline MMR deficiency normal tissues of individuals show microsatellite instability (MSI) (De Vos 2005). Wang *et al.* (1999) describe a family in which children with *hMLH1* deficiency presented with NF1, leukemia and lymphoma before the age of seven years (y).

Two brothers from a consanguineous marriage presented with colorectal cancers at age 11y and 12y respectively (Muller *et al.* 2006). No other disorder was reported and the children showed a biallelic germline mutation in intron twelve of *hMSH2*. The authors state the phenotype is unusual as the characteristic malignancies were not evident. Although no expression or functional studies were performed, they suggest that the mutation could have resulted in low levels of

expression of the functional *MSH2* protein with reduced activity (Muller *et al.* 2006).

It has been suggested that biallelic germline mutations in *hMLH1*, *hMSH2* and *hMSH6* may occur within families who have a clear, autosomal dominantly inherited Lynch history (Menko *et al.* 2004). A homozygous 3 base-pair deletion in exon five of *hMSH6* was described by Menko *et al.* (2004). The child, from a consanguineous marriage, presented with CALS as well as a glioma at age 10y and rectosigmoid cancer at age 12y.

A child, from a consanguineous family, with a homozygous mutation in *hPMS2* and presenting with Turcot syndrome (to be discussed in more detail later) was described by de Vos *et al.* (2005). Another consanguineous family had a different germline homozygous mutation in *hPMS2*. This family had three children, all affected and presenting with CALS. Two of the three children developed rare neuroectodermal tumours while the third developed T-cell lymphoma.

From the data reviewed above, it is evident that a biallelic germline mutation in a DNA MMR gene will inevitably lead to malignancy early in life. The exact malignancy will depend on the nature and severity of the mutation. In some cases, as described by Muller *et al.* (2006), some mutations despite occurring homozygously, may only partially compromise protein function. This residual protein activity may only affect the most susceptible tissues. Therefore, it could be postulated, that only colorectal carcinomas and not the classical haematological malignancies would be prevalent due to the large number of carcinogens in contact with the colorectal tissue and high rate of colorectal cell turnover.

1.3.2.3. Compound heterozygotes

Though rare, there have been reported cases of compound heterozygosity in the DNA MMR genes. De Rosa *et al.* (2002) described a non-consanguineous family in which the offspring had presented with Turcot syndrome, a variant of Lynch syndrome (discussed later). Analysis of the children had shown truncating mutations in each allele of the *hPMS2* gene, inherited from each parent. The mutations in isolation had no clinical phenotype as seen in the parents or their blood relatives.

A compound heterozygote *hMSH6* was recently reported (Plaschke *et al.* 2006) where the patient presented with rectal cancer, at age 19y, and endometrial cancer, at age 24y. Although a non-sense mutation in one allele and a truncating mutation in the other allele were identified, further investigation showed residual expression of *MSH6* in both tumour tissue types. The parents were found to be carriers of the mutations with no obvious related phenotype (Plaschke *et al.* 2006).

Rahner *et al.* (2008) recently described a patient with two compound heterozygous mutations in *hMSH6* (c.1806-1809 del AAAG and c.3226 C>T) who was diagnosed with vitiligo and systemic lupus erythematosus (SLE) at age 16y. At age 17y, the patient was diagnosed with four synchronous colorectal cancers (Rahner *et al.* 2008). The occurrence of auto-immune disorders within the patient is probably fortuitous though it has been documented that tumour-infiltrating lymphocytes are present in all Lynch tumours. The heightened immune response in Lynch tumours have been reported to show a better prognosis. Perhaps the appearance of vitiligo and SLE in the Rahner *et al.* (2008) patient is due to complications as a result of the truncated MSH6 protein.

Non-epithelial malignancies, hyperplastic endometrial tissue and ovarian carcinomas have also been reported in an *hMSH2* compound heterozygous patient. The combination of a frameshift mutation, predicted to result in a truncated protein, and a heterozygous polymorphism were predicted to cause a polarity change in the protein interaction region. Family history uncovered multiple malignancies occurring in other family members including the probands sister (endometrial cancer) and father (colonic and gastric cancer). All malignancies presented within the fourth decade of life (Huang *et al.* 2004). As the cancerous phenotype presented only in adult life, the compound heterozygous mutations within *hMSH2* do not seem as detrimental as those discussed in *hPMS1* and *hMSH6*, which presented during childhood and adolescence.

From the cases reviewed above, it is noteworthy to observe that compound heterozygous mutations within the *hPMS2* gene are more detrimental than those in the *hMSH6* or *hMSH2* genes. However it is not clear from the literature that the mutations reported all translate into equivalent functional compromise.

Compound heterozygous mutations in *hMSH2* resulted in the highest age of onset. Though the mutations occurred within a Lynch family, the patient presented with rare skin tumours. As the skin is a highly proliferative tissue, similar malignancies are expected in MMR deficient individuals. However, skin tumours are extremely rare in Lynch families. The biology underlying the “resistance” of skin to succumb to malignancies in Lynch families is as yet poorly understood.

1.3.3. Epigenetic silencing of a DNA MMR gene

Epigenetic factors are also important in the loss of function of DNA mismatch repair genes. Cancer cells usually have aberrant epigenetic changes during the

progression of the disease and may result in either hypomethylation or hypermethylation. Global hypomethylation (decreased methylation, typically of repetitive sequences) results in chromosomal instability (Wong *et al.* 2007). Alternatively, hypermethylation results in increased methylation of sequences, particularly tumour suppressor genes inducing their silencing. Tumour suppressor genes commonly methylated in tumour tissue include *hMLH1, APC and Rb* – genes directly implicated in the initiation and progression of malignancy (Wong *et al.* 2007).

A vast number of studies of microsatellite instability have shown that a wide range of sporadic cases may exhibit this phenotype (Rowley 2005). In addition loss of expression of the *MLH1* protein has also been noted. These cases, with no family history of malignancy and not meeting the criteria for classification of Lynch syndrome, have been found, in some cases, to have both alleles of the *hMLH1* gene methylated. This methylation-related down-regulation of gene expression results in mimicking a Lynch Syndrome phenotype (Rowley 2005).

More recently, reports of inherited germline epimutations in the *hMLH1* gene described a mother and son who carried same epigenetic marker. This germline epimutation resulted in multiple tumours within the mother and conferred an increased risk of cancer in her son (Hitchins *et al.* 2007). This unusual finding indicates that epigenetic changes in the genome may be passed down through generations, rather than the expectation of epigenetic modifications in parents being “rewired” in the gametes and therefore zygote, as is currently understood.

1.4. Tumourigenesis in MMR deficient cells

Cancers resulting from defects within the DNA MMR system develop from the "mutator pathway" as opposed to the "traditional pathway", which describes carcinogenesis via the loss of heterozygosity in critical genes as proposed by Fearon and Vogelstein (1991). The dubbed "Vogelstein pathway", shown in Figure 2, utilized colorectal cancer as a model for the study of carcinogenesis. The ability to capture various stages of carcinogenesis and to molecularly characterise these changes from normal colonic epithelium to a malignant tumour, allowed the sequence of events to be assessed. It must, however, be noted that it is the accumulation of genetic changes, not the step-wise progression depicted below, that is important in the initiation and accumulation of a malignancy.

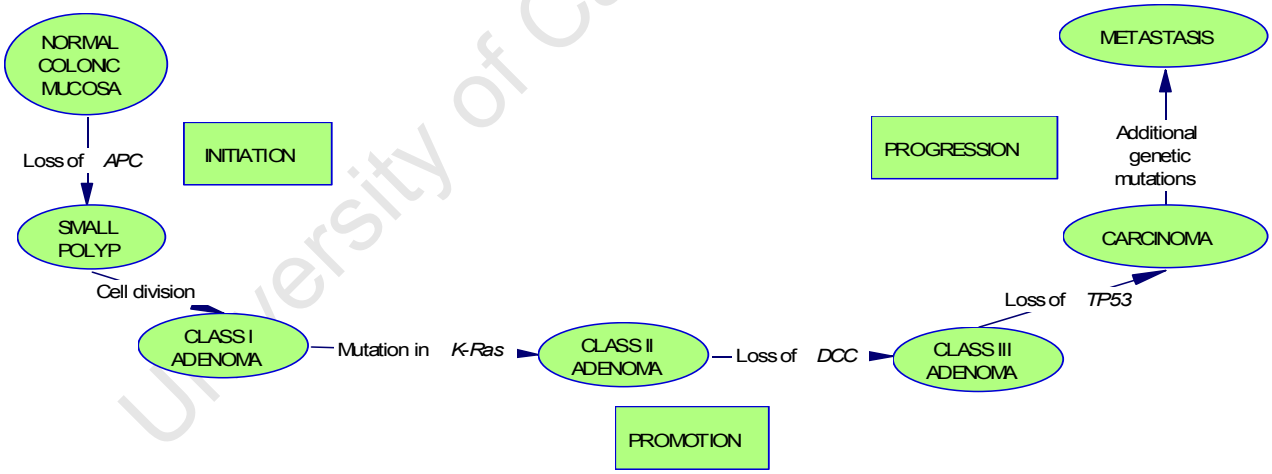


Figure 2: **"Vogelgram" of the adenoma to carcinoma sequence in colon cancer. The conversion of normal epithelium to a hyperproliferative epithelium in the colon results from the loss of the *APC* tumour suppressor gene on chromosome 5. Through the continuation of cell division, a small polyp may develop. An early, or Class I adenoma develops through the loss of DNA methylation in these cells. An intermediate (class II) adenoma develops via mutations within the *K-ras* gene on chromosome 12. The conversion to a late (class III) adenoma results from the loss of the tumour suppressor gene *DCC* (Deleted in Colon Cancer) on chromosome 18. Finally, the loss of *TP53* on chromosome 17 completes the conversion to a carcinoma. Metastasis can occur through additional genetic changes.**

The genetic model described above is based on chromosomal instability and estimated to account for 85% of all colorectal cancer cases (Hung and Chung 2006). While the majority of colorectal cancers originate through the Vogelstein pathway (Figure 2), tumours of Lynch syndrome and some sporadic cancers originate via the "mutator pathway". This pathway is characterized by a high level of microsatellite instability (Iglesias *et al.* 2006).

1.4.1. Mutator Pathway

During cell replication, the entire complement of double-stranded DNA has to be replicated with fidelity and integrity. As is expected, errors of various kinds occur during replication. Microsatellite regions (regions of repeated regions within the genome) are prone to the incorporation of errors during normal DNA replication. Instability of the microsatellites results in the insertion or deletion of repeated sets from either slippage of the DNA polymerase enzymes or through incorrect recombination processes in germ cells. These errors are corrected by a vigilant DNA MMR system. In the absence of an optimal DNA MMR system, cells accumulate mutations in their genome at a faster rate than normal cells (Lynch and Lynch; 1997).

In cells where errors are not corrected, the instability of these microsatellite regions is observed (Huang *et al.* 2004). The exact mechanism of the Replication Error Phenotype (RER) to the initiation of carcinogenesis, is unknown. There is a distinction between the extent of microsatellite instability (MSI) within tumour tissue of patients. The recommended panel of markers used to test for MSI is known as the Betheseda Panel and currently incorporates five highly polymorphic markers. The instability of three or more of the markers constitutes a high microsatellite phenotype (MSI-H). A low microsatellite instability phenotype (MSI-L) represents visible instability of between one and three of the Bethesda markers

being unstable whereas a microsatellite stable phenotype (MSS) represents no marker of the panel being unstable in the tumour tissue (Lynch and De la Chapelle 2003).

It is thought that there are three basic levels of mutation within the mutator pathway (Iglesias *et al.* 2006), shown in Figure 3.

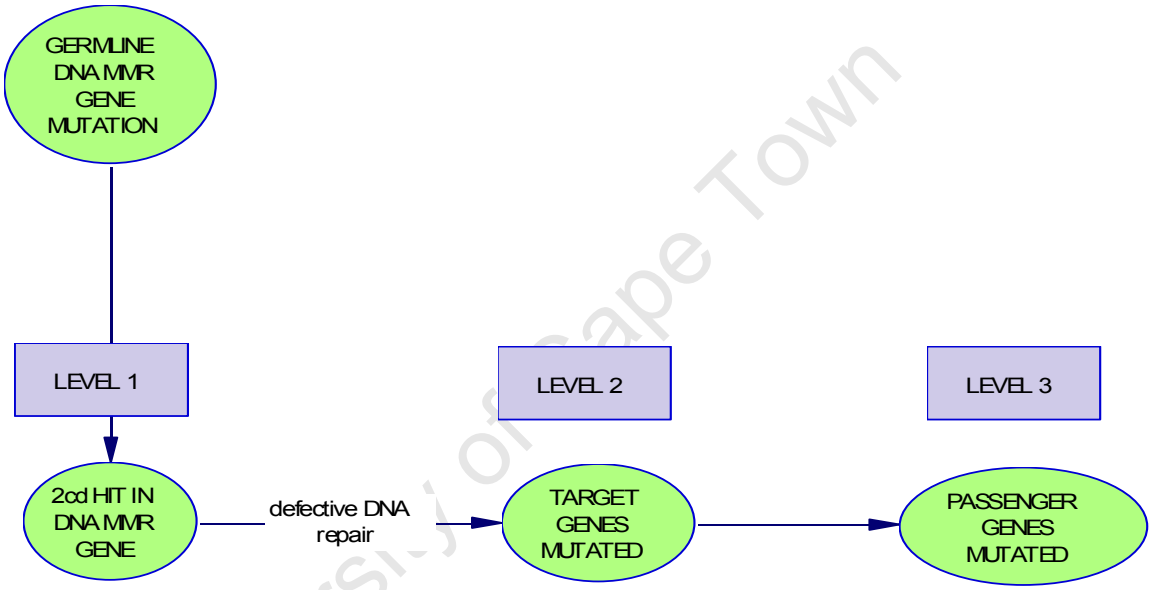


Figure 3.: **The three basic levels of mutation in the mutator pathway.**

A second "hit" occurring in a DNA MMR gene that already has a germline mutation in the first level of the pathway results in defective DNA MMR repair and microsatellite instability. The second level of the pathway occurs when genes that affect tumour development have been influenced by the microsatellite instability. The final level of mutation in the mutator pathway includes all other altered genes which do not affect, or enhance the development of a tumour (Iglesias *et al.* 2006).

Due to the alteration of many microsatellite regions, it has been difficult to predict which of those mutations play a causal role in the development of the malignancy. Duval *et al.* (2001) statistically examined instability of coding sequences in high MSI human colorectal cancers. A sample of 25 candidate genes, each with a minimum of eight mononucleotide tracts, were chosen and analysed for mutations in these cancers. Thereafter a mutation order was established using the maximum likelihood statistical model. The study found that the next probable mutation in the mutator pathway (after loss of the DNA MMR gene) is the MMR gene *hMSH3*. This gene is not directly associated with Lynch syndrome (Duval *et al.* 2003). The *MSH3* protein forms, with *MSH2*, a complex known as MutSb which is involved in the repair of single base-pair deletion and mutations (Figure 1).

Genetic instability in the mutator pathway results in mutations in the genome occurring more frequently due to a decreased rate of mutations being repaired (Cotran *et al.* 1999). There is an accelerated rate of carcinogenesis of tumours developing from the mutator pathway and this is thought to be due to the rapid accumulation of mutations (Fernandez-Peralta *et al.* 2005). It has been estimated that the normal progression from an adenoma to carcinoma for sporadic cases of colorectal cancer occurs within 8 -10 years, whereas the adenoma to carcinoma progression in colorectal cancer from Lynch syndrome patients can occur within 2 – 3 years (Grady 2005; Lynch and De La Chapelle 2003).

1.4.2. Interaction between Vogelstein and Mutator Pathways

Salahshor *et al.* (1999) aimed to elucidate the difference between the MSI-H and MSS pathways in a cohort of colorectal cancer patients. Mutations within the typical Vogelstein pathways genes and *TGFBR2* were examined within the tumour tissues. Traditional Vogelstein pathway genes *APC* and *TP53* were found not be

mutated in any of the MSI-H samples yet there was frameshift mutations within *TGFBR2* (Salahshor *et al.* 1999).

There are certain families who clinically fulfil the Amsterdam criteria (Table 1) yet do not show MMR deficiency in tumour tissue. The tumour tissue from these families show distinctive molecular features such as mutations within *K-ras* genes and the mutation profile of Vogelstein pathway genes mimic that of sporadic malignancies.

An integral part of the Vogelstein pathway is the mutation of *TP53*. This gene has been implicated in almost every malignancy and is an integral part in the optimal functioning of any cell. The *TP53* gene has a mononucleotide repeat within the coding region and due to the nature of the mutator phenotype; it is expected that *TP53* will be affected by the MMR deficiency (Woerner *et al.* 2005. Shitoh *et al.* (1998) demonstrated instability of the *TP53* gene as a hallmark of the replication error phenotype in colorectal carcinomas (both sporadic and in Lynch syndrome).

1.5. Tumour Spectrum in Lynch Syndrome

Colorectal cancer is the archetypal cancer in Lynch syndrome and usually presents within the fourth decade of life yet, the occurrence of extracolonic cancers has also been observed (Lynch and De La Chapelle 2003). Consequently, the disease has been phenotypically characterized into two syndromes; namely Lynch syndrome I and Lynch syndrome II.

Both syndromes, Lynch I and Lynch II, are primarily diagnosed by: an autosomal dominant mode of inheritance, an early age of onset and a germline mutation within a DNA MMR gene (Douglas *et al.* 2005). Lynch Syndrome I is also known

as Hereditary Site Specific Colon Cancer (HSSCC) and describes those predisposed families who present only with colon cancer, predominantly right-sided colon cancer.

Extracolonic cancers are a hallmark of Lynch syndrome II, often referred to as Cancer Family Syndrome (CFS) (Lynch and Lynch 1997). The recognition of extracolonic cancers within the tumour spectrum has led to the revision of the criteria traditionally used to identify affected pedigrees (Table 1).

The screening and development of endometrial cancer within Lynch patients has been recognized and this malignancy is considered to present in the predisposed female member of a Lynch kindred. There has been controversy as to the inclusion of other extracolonic cancers within the Lynch tumour spectrum as there may be cases in which the tissue specific malignancy may be fortuitous or may segregate within the particular family as a result of a mutation in another gene causing susceptibility. To date, the inclusion of tissue sites in the Lynch tumour spectrum are based on elevated risks in individuals to a particular type of cancer. (Watson and Riley 2005). Gastric, small bowel, hepatobiliary tract, upper urological tract, ovarian, endometrial and brain cancers were defined as part of the HNPCC tumour spectrum (Watson and Riley 2005).

1.5.1. Gynaecologic cancers

Endometrial cancer was one of the malignancies first described by Dr Warthin in the analysis of Family G and together with colorectal cancer; it is the second most prominent cancer in Lynch syndrome (Banno *et al.* 2004, Douglas *et al.* 2005). Lu *et al.* (2005) documented that in females, the development of endometrial cancer, and to a lesser extent ovarian cancer may equal or even exceed the risk of colon cancer in Lynch syndrome.

There have been multiple studies attempting to correlate the occurrence of endometrial cancer with the primary germline mutation (Gonzalez-Aguilera *et al.* 2003, Plaschke *et al.* 2004, Goodfellow *et al.* 2003). It has been suggested that the female members of families segregating with *hMSH2* mutations present with endometrial cancer more frequently than any other Lynch-associated cancer (Gonzalez-Aguilera *et al.* 2003) yet other studies have shown the presence of this malignancy regardless of the germline MMR gene or mutation. In particular, a study performed in 2003 demonstrated that female patients with *hMSH6* mutations present with endometrial cancer significantly earlier than women in the general population ($p = 0.04$) (Goodfellow *et al.* 2003). Malander *et al.* (2005) found truncating mutations in *hMSH6* confer a high risk for the development of a gynaecologic cancer. There is no consensus as to the genotype – phenotype relationship in Lynch syndrome and it is recommended that females harbouring a predisposing mutation to the disorder undergo routine gynaecologic screening (ICG-HNPCC 2004).

Kuismanen *et al.* (2002) suggested that the genesis of gastrointestinal and endometrial cancers in patients predisposed to Lynch syndrome, occur through different routes. Interestingly, however, both malignancies are characterised by MSI. The study focused on MSI and mutations within certain microsatellite-rich genes and found that changes in different target loci could distinguish between colorectal and endometrial cancer (Kuismanen *et al.* 2002).

Pathologically, Van den Bos *et al.* (2004) demonstrated that Lynch-related endometrial cancers differ considerably from sporadic endometrial cancers. Lynch-related endometrial cancers show poor differentiation, Crohn-like lymphoid reactions, lymphoangioinvasive growth and tumour-infiltrating lymphocytes; pathologically equivalent to Lynch-related colorectal cancer (Van den Bos *et al.* 2004).

Molecularly, Lynch-related endometrial cancer and sporadic endometrial cancer seemed to be comparable (Rijcken *et al.* 2007). The group examined the expression of various cell cycle and apoptosis proteins and only found the expression of Cyclin B1, a cell cycle regulatory gene, is more highly expressed in Lynch endometrial cancer. The apoptosis related protein *BAX* was significantly absent in three Lynch cancers when compared to the presence of the protein in sporadic cancers ($p = 0.013$). The study also suggested a trend toward a higher proliferation rate in Lynch cancers. Although the Lynch-related endometrial cancer were determined through family history and inclusion by the revised Amsterdam criteria, upon further investigation only ten of eighteen samples were shown to have DNA MMR defects (six samples had mutations in *hMLH1*, two in *hMSH2* and two in *hMSH6*) (Rijcken *et al.* 2007). Unfortunately there is no mention of pathology or IHC investigations of the samples where the DNA MMR genes were genuinely compromised or not functional.

Ovarian Cancer

Female Lynch patients have an eight to fifteen percent risk of development of ovarian, an estimated ten-fold increase in risk compared to that of the general population (Song *et al.* 2006). Interestingly Geisler *et al.* (2003) reported that 16.8% of all ovarian carcinomas have defects in the DNA MMR system. Suzuki *et al.* (2001) showed that defects in the DNA MMR system are involved in the development and progression of ovarian malignancies – notably in the genetic alterations introduced by the defective system. In a large multicenter study of common variants in MMR genes there was an observed association between a specific variant in the DNA MMR gene *hPMS2* ($p = 0.046$) whereas variants in the other DNA MMR genes (*hMLH1*, *hMSH2*, *hMLH3, hMSH3*, *hMSH6* and *hPMS1*) showed no such association (Song *et al.* 2006). In addition to hereditary MMR

mutations in ovarian carcinomas, hypermethylation of the *hMLH1* promoter has also been identified (Geisler *et al.* 2001).

1.5.2. Hepatobiliary Cancers

The hepatobiliary tract is composed of the pancreas, biliary system and liver. Lynch *et al.* (1985b) first described the occurrence of pancreatic cancer in a Lynch family, evident through five generations.

Genetics of Hepatobiliary Cancers

In particular, liver, or hepatocellular, cancers have many environmental risk factors. For example, chronic viral hepatitis confers a high risk of hepatocellular cancer. Alpha1-antitypsin deficiency, Non-alcoholic fatty liver disease, haemochromatosis as well as alcohol and exposure to toxins may also result in a malignancy (Teufel *et al.* 2007).

Moinzadeh *et al.* (2005) found amplifications of certain chromosomal positions (1p, 8q, 6p, 17q) in the majority of sporadic hepatobiliary malignancies. In addition, deletions of chromosomal regions 16q, 4q, 17p and 13q were also found. At the gene level, point mutations in the tumour suppressor gene *TP53* are evident (Teufel *et al.* 2007). Mutations within the components of the Wnt, Ras and TGFB-signalling pathways have also been shown to promote the development of hepatocellular carcinoma. The Wnt pathway functions in cell to cell interaction and the regulation of proliferation. In addition mutations within the tumour suppressor gene *Rb* has been implicated in the disease (Moinzadeh *et al.* 2005).

Though the initiation of sporadic hepatocellular carcinomas is currently unknown and probably due to external factors, the genetic progression of the cancer

involves all the major components in the disease as found in the genetic model proposed by Fearon and Vogelstein (1991). There are also components of the tumourigenesis pathway similar to the mutator pathway as seen by the mutation of the TGFB signalling pathway.

1.5.3. Gastric Cancers

Gastric cancer was originally described in association with Lynch syndrome in Family G in 1913 (Douglas *et al.* 2005). Though gastric cancer is rare in Lynch families described to date, studies have Chinese, Korean and Brazilian descent have a higher risk for both gastric and hepatic cancer as compared to that of endometrial and small intestinal cancer (Cai *et al.* 2006, Wang *et al.* 2006).

A recent study by Gylling *et al.* (2007) found that MSI gastric cancers from both MMR-positive and MMR-deficient individuals showed significant mutations within repeat-tracts of genes ($p < 0.001$) and a higher methylation index than that of sporadic MSS gastric cancers. Additionally, there was a significant association between frame-shift mutations in microsatellite rich regions of growth regulatory genes ($p = 0.0001$) as well in MSI tumours (Gylling *et al.* 2007). The Gylling *et al.* (2007) study provides further insight into the molecular differences between MSI and MSS malignancies. However, the genetic signature of MSI tumours – regardless of site of cancer – still remains unanswered.

Genetics of Gastric Cancer

Environmental or external factors contribute considerably in the development of all diseases and gastric cancer is no exception. The bacterium *Helicobacter pylori* is an important risk factor for the development of gastric cancer.

Microsatellite instability in gastric carcinomas (Garay *et al.* 2004) is rare although there have been cases of MSI-H gastric carcinomas documented, in which there is frequent loss of a DNA MMR gene, *hMLH1* and *hMSH2*. Additionally, mutations within *TGBR2, BAX, IGF2R, E2F4, hMSH3* and *hMSH6* are frequent. These mutations are commonly found in Lynch colorectal cancers (El-Rifai and Powell 2002, Rafau and Powell 2002).

Sporadic gastric cancers are found to have mutations in *TP53, APC*, *B-catenin* and overexpression of growth factors and receptors (such as *c-MET, HER-2/neu* and *EGFR/K-sam*) – the molecular pathology similar to the Vogelstein pathway (El-Rifai and Powell 2002).

1.5.4. Small intestinal cancers

A study by Love first included cancers of the small intestine in the Lynch spectrum of cancers. The inclusion was then verified by Lynch *et al.* (1985a). In 1998, an international collaboration into the epidemiology of intestinal cancers in Lynch families showed that this malignancy had an age of onset almost two decades before that of the general population (Rodriguez-Bigas *et al.* 1998). In addition, small intestinal cancers showed a gender bias towards males.

Genetics of Small intestinal cancers

Small intestinal cancers comprise of cancers within the ileum, duodenum and ileum and are a common occurrence in familial adenomatous polyposis (FAP) – a hereditary form of colorectal cancer. However, these malignancies are rare in the general population (Lynch *et al.* 1985a). In Lynch syndrome, the majority of cancers are found within the jejunum.

Malignancies of the gastrointestinal tract may be attributed to hereditary conditions such as Lynch syndrome or FAP. Gastrointestinal Stromal tumours (GIST) may be another type of tumour occurring throughout the intestinal tract. The genesis of small intestinal tumour is similar to that of colonic carcinomas, using the Vogelstein pathway as a model.

1.5.5. Upper Uroepithelial Tract

Loss of *hMSH2* has been found in 60% of microsatellite instable upper uroeptihelial cancers (Ericson *et al.* 2005) and thus there are recommendations to screen families with this germline mutation for urinary tract carcinomas. Various studies have shown that Lynch families have an increased risk for the development of transitional cell carcinoma of the ureter, renal pelvis and calyx. These studies also concur that these patients do not have an increased risk for the development of bladder cancer (Sijmons *et a.* 1998, Lynch *et al.* 2003, Hartmann *et al.* 2003). The life-time for the development of an upper uroepithelial tract cancer in Lynch patients is estimated to be between four and ten percent.

Genetics of Upper Uroeptihelial Tract Cancer

The occurrence of an upper uroepithelial tract malignancy is rare. As with most cancers, involvement of *TP53* and *Rb* pathways is evident (McGrath *et al.* 2006).

Ruopret *et al.* (2005) reviewed upper uroeptithelial tract cancers and found that MSI and expression of MSH2 could be effectively used prognostically. Loss of expression of E-cadherin was also found to contribute to the overall survival rate (). When abnormal karyotypes were further studied, loss of chromosome nine was found to be prevalent in 205 upper urinary tract and bladder carcinomas reviewed

by Fadl-Elmula (2005). Tumour suppressor genes found on chromosome nine include apoptosis related genes such as p16.

Wu *et al.* (2007) reported that individuals with single nucleotide polymorphisms (SNPs) with in DNA repair genes and cell cycle control controls were predisposed to the development of bladder cancer. Yet, there was a bias toward certain environmental effects (for example smoking); together the environmental effect coupled with a high-risk allele was found to have a multiplicative risk factor for the individual.

Mutations within detoxification genes have also been implicated in the predisposition to and the development of bladder cancer (McGrath *et al.* 2006).

1.5.6. Brain cancer

Brain cancer is extremely rare in Lynch families. A study by Vasen *et al.* (1996) showed a statistically significant increase ($p = 0.04$) in the risk of brain tumours in Lynch families compared to the background population.

Mutations within the MMR gene *hPMS2* have been implicated in Turcot syndrome, a rare variant of Lynch syndrome. Turcot syndrome is characterized by manifestation of cancer in both the colorectum and brain (Lynch *et al.* 2003).

Genetics of Brain Cancer

Loss of chromosome 17 and in particular the *TP53* which resides there is an important factor in the development of brain malignancies. Gliomas are the most common brain cancer and in addition to *TP53*, overexpression of growth factors

has been described (Louis 2006). Several other tumour suppressor genes have been implicated in the development and initiation of brain cancers and these are thoroughly reviewed in Louis (2006).

1.5.6.1.Turcot syndrome

The appearance of colorectal carcinomas coupled with brain cancers is characterized as Turcot syndrome. Mutations within the *hPMS2* gene have been associated with this syndrome (Hamilton *et al.* 1995).

1.5.7. Muire Torre Syndrome

Muire Torre Syndrome (MTS) is a rare variant of Lynch syndrome. This syndrome is characterised by sebaceous adenomas, sebaceous carcinomas and multiple keratocanthomas coupled with colorectal cancer (Lynch *et al.* 2003). Mutations within *hMSH2* have been associated with MTS.

1.5.8. Genotype – Phenotype correlation

Due to the diverse number of genes and cancerous manifestations that are implicated in this disorder, there have been several investigations into determining the genotype – phenotype relationship of the mismatch repair mutation and occurrence of specific cancers in families (Grady 2005).

A number of these studies (Plaschke *et al.* 2004; Rodriguez – Moranta *et al.* 2006; Wagner *et al.* 2003) have shown that HNPCC kindreds with *hMSH2*

mutations seem to have a later age of onset of colorectal cancer. Female members have a lower risk for CRC but have a higher risk for endometrial cancer. *hMSH2* mutation carriers also seem to have a higher risk of extracolonic cancers; that is cancers of the upper urothelial tract, ovary, and gastric cancer (Vasen *et al.* 1996). Vasen *et al.* (2001) showed that *hMSH2* mutation carriers had a higher lifetime risk of developing any cancer in comparison to *hMLH1* mutation carriers ($P < 0.01$). In addition, Gonzalez – Aguilera *et al.* (2003) investigated the effects of mutations within different DNA mismatch repair genes on the development of extracolonic cancers in HNPCC. Goecke *et al.* (2006) studied the German HNPCC cohort and reported that in males with an *hMSH2* predisposing mutation, there was a correlation to prostate cancer, a cancer not found within the Lynch tumour spectrum.

Germline mutations resulting in the expression of defective proteins were shown to exhibit an increase in the frequency of extracolonic manifestations when compared to germline mutations that resulted in only monoallelic expression of the wild type *hMLH1* allele in a Danish cohort (Jager *et al.* 1997). In addition, the group reported an intronic *hMLH1* germline mutation, which resulted in silencing of the allele. This particular mutation presented with a reduced frequency of extracolonic cancers when compared to other mutations in the same cohort (Jager *et al.* 1997). The authors hypothesise two potential disease modifying mechanisms: the first mechanism involves the presence of a modifier gene close to *hMLH1* that can potentially compensate for MMR. The second mechanism involves a dosage effect in which the mutant allele is silenced due to the mutation. The intronic germline mutation reported in the study seems to adhere to the proposed second mutation as individuals with a mutation that results in expression of a defective protein exhibited a high number of extracolonic malignancies (Jager *et al.* 1997). However, the mutation that results in a defective protein is the *hMSH2* gene – of which various studies have implicated in an increase in extracolonic malignancies.

The cumulative incidence of extracolonic cancers manifesting within HNPCC families was further investigated by Lin *et al.* (1998). This study showed significant evidence in associating extracolonic cancers to individuals with *hMSH2* germline mutations (48% *hMSH2* vs 11% *hMLH1* patients with extracolonic cancers, p = 0.016).

More recently, a genotype-phenotype association was attempted using a family with an *hMLH1* mutation and a family with an *hMSH2* mutation (Havarlsson *et al.* 2006). Tumours from the family members were characterized and no correlation was obtained. In fact, both inter-tumour and intra-tumour heterogeneity was observed histopathologically and immunohistochemically (Halvarsson *et al.* 2006).

In a recent study into the gene-related cancer spectrum in families with Lynch syndrome, Geary *et al. (*2007) analysed all malignancies reported in a cohort of Lynch syndrome families. The study concluded that colorectal and endometrial cancers were prominent in these families, regardless of primary mutation. *hMSH2* mutation carriers seemed to have a broader tumour spectrum of cancers than *hMLH1* mutation carriers. It was also evident that certain cancer types clustered within families.

The inclusion of cancers in the tumour spectrum of Lynch syndrome is based on incidences as opposed to the molecular pathogenesis of malignancy at each anatomic site. Monglat-Artus *et al.* (2006) describe a patient with a *hMSH2* germline mutation who presented with multiple primary cancers of the colon, upper urinary tract and endocervix. These cancers were shown to be MSI and loss of the *MSH2* protein was demonstrated. In addition, the patient also presented with a renal papillary cancer. On further investigation, this cancer was unrelated to the DNA MMR deficiency as the *MSH2* protein was present and the tumour was MSS (Monglat-Artus *et al.* 2006).

Miyaki *et al.* (1997) found that mutations in *hMSH6* can lead to a more benign form of Lynch Syndrome and thus may be under-diagnosed or present much later than usual It has been further documented that mutations in *hMSH6* in Lynch syndrome families result in a greater frequency of gynecologic cancers among the female members of those pedigrees (Cederquist *et al.* 2005).

There is a vast phenotypic heterogeneity seen within Lynch tumours, both intra- and inter-individual. Halvarsson *et al.* (2006) reported variable histopathology and immunohistochemical expression of *B-catenin* in Lynch related colorectal, endometrial and ovarian cancers. Due to the identical germline mutation in the MMR gene, it was hypothesized that independent modifier genes, which were affected by frame-shift mutations due to the replication error (RER) phenotype (microsatellite instability), influence the histopathology and morphology of these Lynch tumours (Halvarsson *et al.* 2006). As indicated in other studies, the modifier genes affected are most likely to be tissue specific.

When correlated with its function, the promiscuity of the spectrum of tumours seen in *hMSH2* mutation positive individuals seems obvious. A *MSH2* complex (either the *MSH2/MSH6* complex or *MSH2/MLH3* complex) binds to the replicating DNA and scans for any defects (Figure1) (Chao and Lipkin 2006). Thus, anomalies within the final product should result in a wider variety of tumours; the *MSH2* protein is one of the first MMR proteins in the pathway and is critical in the efficient functioning of the system in every cell type.

Germline mutations resulting in the expression of defective proteins were shown to exhibit an increase in the frequency of extracolonic manifestations when compared to germline mutations that resulted in only monoallelic expression of the wild type *hMLH1* allele in a Danish cohort (Jager *et al.* 1997). A study by Watson *et al.*

(2008) aimed to statistically characterise the risk of extracolonic malignancies in a large, pooled cohort of mutation carriers (either *hMLH1* or *hMSH2* mutation carriers in four study centres (Finland, USA, Holland and Denmark). Overall, Watson *et al.* (2008) found individuals with MSH2 mutations to be at a higher risk for urologic cancer (males) ($p<0.0001$) and ovarian cancer (females) ($p<0.006$).

In light of this, analysis of published founder mutations within the *hMLH1*, *hMSH2* and *hMSH6* genes were attempted with the aim of possibly elucidating mutations occurring within a specific area leading to certain malignancies. The genes *hMLH1* and *hMSH2*, accounting for the majority of HNPCC, show a variety of founder mutations globally. The results of this can be seen in Tables 2 and 3 (*hMLH1* founder mutations and *hMSH2* founder mutations respectively). Mutations were regarded as founder mutations if haplotype studies were performed and the mutation could successfully be dated back to at least 5 generations.

Overall, the founder mutations found in both genes (*hMLH1* and *hMSH2*) result in an array of tumours and some of the tumours encountered do not fall into the HNPCC tumour spectrum (Table 2, Table 3). When compared to the results of the *hMLH1* founder mutations, mutations within the *MSH2* gene are more promiscuous. Every mutation the h*MSH2* gene results in at least three different tumour sites whereas each mutation within the *hMLH1* gene results in at least one tumour site.

Examination of the effect of founder mutations within populations is essential in order to extract vital information regarding the biology of the disease. A variety of cancers were found with each founding mutation and therefore a specific region of the protein mutated resulting in a specific range of malignancies is not possible. In addition, the founding mutations do not comply with the assumption that *hMSH2* mutations result in extracolonic cancers, more so than *hMLH1* however, it was

observed that extracolonic cancers occur more frequently (not exclusively) in patients with *hMSH2* mutations. A possible setback was the lack of clinical data in respect to each founding mutation as the information was mined from publications.

Table 2.: ***hMLH1*** **founder mutations and tumour spectrum**

EXONS	6 - 11				13 – 19					
DOMAIN	MutS homologs interaction domain				*PMS1*/ *PMS2*/ *MLH3* interaction domain					
MUTATION	g454-1a	g.IVS6+3SA >G	Del 546 – 2 AG	1.8kb deletion	C1528T	c.1757-1758insC	3.5kb deletion	c.1831delAT	c.G2141A	c.2269-2270
RESULT	Deletion exon 6	Truncated protein	Deletion exon 7, skipping exon 6	Deletion exons 10 and 11	Predicted Truncated protein	Truncated protein	Truncated protein	Truncated protein	Inframe deletion AA at 578 – 632	Predicted chain elongation
TUMOUR SPECTRUM										
CRC	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
GYN	Y	Y				Y	Y	Y		
GAS	Y	Y				Y	Y	Y		
UUT	Y							Y		
SMB	Y				Y	Y				
HEP	Y	Y				Y	Y			
BRA						Y				
PRO										
BRE		Y			Y					
SKI					Y					
LUN					Y					
GI		Y (tracheal)								
SPIN										
HAE										
COUNTRY	Finland	Italian Quebec	USA	China	RSA	Korea	Finland	Italian Quebec	Switzerland	Italy
REF	Nystom-Lahti *et al.* 1995	Thiffault *et al.* 2004	Tanko *et al.* 2002	Chan *et al.* 2001		Shin *et al.* 2004	Nystom – lahti *et al.* 1995	Thiffault *et al.* 2004	Hutter *et al.* 1998	Caluseriu *et al.* 2004

Y = yes, tumour present. CRC = Colorectal, GYN = gynaecological (ovarian, endometrial, uterine), GAS = gastric, UUT = upper uroepithelial tract (bladder, kidney), SMB = small bowel, HEP = hepatobiliary tract (pancreas, liver), BRA = brain, PRO = prostate, BRE = breast, SKI = skin, LUN = lung, GI = gastrointestinal tract (tracheal, oesphogeal), SPIN = Spinal cord, HAE = haematological (leukemia, lymphoma). Yellow represents tumours within the HNPCC tumour spectrum.

Table 3.: ***hMSH2*** **founder mutations and tumour spectrum**

EXONS	1 – 6					7 - 12			
DOMAIN	DNA BINDING SITE					*MSH3*/ *MSH6* INTERACTION			
MUTATION	Deletion exons 1 – 6	Deletion exons 1 – 6	c. A942+3T	c.A943+3T	Deletion exon 4 – 16	Deletion exon 8	c.1452 – 1455 deletion AATG	G1906C	T2063G
RESULT	Truncated protein	Truncated protein	Deletion exon 5	Deletion exon 5	Truncated protein	Truncated protein	Truncated protein	Pro336Ala	M688R
TUMOUR									
CRC	Y	Y	Y	Y	Y	Y	Y	Y	Y
GYN	Y	Y	Y	Y	Y	Y	Y	Y	Y
GAS		Y	Y	Y		Y	Y	Y	Y
UUT	Y	Y	Y	Y	Y	Y		Y	Y
SMB	Y	Y	Y	Y					Y
HEP			Y						
BRA			Y			Y	Y		Y
PRO			Y			Y			Y
BRE			Y	Y		Y	Y		Y
SKI		Y	Y	Y	Y	Y			Y
LUN									
GI			Y		Y				Y
SPIN			Y						
HAE								Y	Y
COUNTRY	USA	Italy	Newfoundland	North American English	Newfoundland	Newfoundland	China	Ashkenazi Jewish	Tenerife
REFERENCE	Wagner *et al.* 2003	Stella *et al.* 2007	Stuckless *et al.* 2006	Frogatt *et al.* 1999	Stuckless *et al.* 2006	Stuckless *et al.* 2006	Chan *et al.* 2004	Foulkes *et al.* 2004	Medina – Arana *et al.* 2006

Y = yes, tumour present. CRC = Colorectal, GYN = gynaecological (ovarian, endometrial, uterine), GAS = gastric, UUT = upper uroepithelial tract (bladder, kidney), SMB = small bowel, HEP = hepatobiliary tract (pancreas, liver), BRA = brain, PRO = prostate, BRE = breast, SKI = skin, LUN = lung, GI = gastrointestinal tract (tracheal, oesphogeal), SPIN = Spinal cord, HAE = haematological (leukemia, lymphoma). Yellow represents tumours within the HNPCC tumour spectrum.

1.6. Pathology of Lynch Syndrome Cancers

Colonic and extracolonic tumours formed in Lynch syndrome are poorly differentiated and exhibit a large amount of mucoid and signet-cell features as well as infiltrating lymphocytes within the tumor. A feature of these tumors is that MSI is evident and the DNA MMR gene mutated is absent through immunohistochemical staining (Lynch and Lynch 2000, Jass 2007). Although MSI can occur in other non-familial forms of cancers (for example sporadic colon cancers with *hMLH1* epigeneticaly silenced), it is not a consistent feature (Hitchins *et al.* 2007).

It is evident that there are differences in early tumourigenic events between Lynch syndrome colonic adenomas in comparison to their sporadic counterparts (Rijcken *et al.* 2007). Apart from the clear pathological observations, more molecular differences are being elucidated. Rijcken *et al.* (2007) showed that certain cell cycle and apoptosis proteins have differential expression in Lynch syndrome colonic adenomas compared to sporadic colonic adenomas. Studies investigating occurrence of endometrial cancer in Lynch patients have also shown molecular differences when compared to sporadic endometrial carcinomas (Lax 2004, Van den Bos *et al.* 2004, Ricjken *et al.* 2005).

In contrast, a study profiling the target gene mutation profile in colorectal, gastric and endometrial high MSI tumours, demonstrated that there are tissue specific differences (Duval *et al.* 2002). However, the study did find that there are certain target genes mutated in all the tumours studied. Once these genes have been mutated, it was hypothesized that the other mutations were dependant on the location of the tumour (Duval *et al.* 2002). The tumour samples used in the study were not necessarily from Lynch syndrome patients but were chosen on the basis of their MSI, thus the samples may have been MMR deficient due to epigenetic silencing. *TGFBR2, BAX, IGF2R, MSH3, MSH6* were all proposed to be target

genes for mutation in MSI-H tumours, regardless of site of cancer (Duval *et al.* 2001, Duval *et al.* 2002).

Apart from the differences in the site of cancer in individuals predisposed to Lynch Syndrome, the age of onset of the cancer may differ significantly. Although most individuals present in the fourth decade of life, there are cases in which the disease presents in the second or third decade of life (Westaphalen *et al.* 2005).

1.7. Research in Lynch Syndrome

Since the setting of criteria for identifying the Lynch phenotype (Amsterdam and Bethesda criteria) large Mendelian families were followed up. These families facilitated the localisation and ultimate identification of the genes involved in the phenotype. The genes involved were identified between 1993 and 1995 (Rowley 2005).

In biological systems, usually the optimal functioning of a gene or protein depends not only on its integrity but also on the molecules (or the responsible genes) upstream of it in a regulatory or biochemical pathway, or on factors that enhance or reduce its function. These factors are now being thoroughly studied and these include low penetrance genes (or modifier) genes (Ilyas *et al.* 1999, Dribble and McCabe 2000, De Jong *et al.* 2002).

The majority of studies in Lynch syndrome have focused on the variability seen in the age of onset of symptoms in predisposed families. The occurrence of specific extracolonic cancers and the molecular pathology of these cancers in terms of a DNA MMR gene deficiency are yet to be elucidated.

1.8. Aims and Objectives

The aim of the study is ultimately to determine the molecular pathology of extracolonic cancers in Lynch syndrome and to elucidate whether or not the occurrence of these extracolonic cancers is a direct result of the mismatch repair deficiency.

The working hypothesis of this study relies on the mechanism of DNA MMR activity. As discussed previously, mutations within DNA MMR genes that consequently result in a breakdown of the repair system exhibit unstable regions in microsatellite rich areas of the genome. Manipulating this fact, a panel of microsatellite rich genes will be assessed in germline DNA and the corresponding tumour DNA of the available Lynch-associated malignancies. This cohort will also be subjected to Methylation specific – Multiplex Ligation-dependant Probe Amplification (Ms-MLPA) in order to attempt to characterize a pattern of methylation in these patients. The objective of this study is to evaluate and potentially molecularly characterize Lynch-associated tumours. If successful, this method may be employed in pathology laboratories and other diagnostic centres to appropriately diagnose and initiate treatment. Furthermore, the results of this study may direct future drug development.

Prior to the microsatellite-rich genes and Ms-MLPA study, a potential modifier for site of cancer will be assessed in a large cohort. Both patients predisposed to Lynch syndrome and controls will be assessed for the modifier genotypes. Potential modifiers for site of cancer may be possible diagnostic and management essentials by directly treatment and screening regimes.

Various techniques will be learned and implemented in the study.

Akin to many cancers prevalent today, the earlier the disease is diagnosed, the better the prognosis and in individuals with germline mutations thus resulting in better the management of the symptoms. It is therefore imperative to diagnose the disease and identify predisposed families as early as possible to allow for the potentially effective preventative measures.

Chapter 2:

hMLH1 -93G>A Promoter polymorphism as a phenotypic modifier of Lynch Syndrome

2.1 Introduction

The current study is focused on the occurrence of extracolonic cancers in Lynch syndrome. Phenotypic and genotypic heterogeneity are evident in Lynch syndrome, further complicating both diagnosis and treatment. The phenotypic heterogeneity observed in the syndrome has implications on the screening regimens and management of the disease.

Low penetrance or modifying genes are genes which secondarily influence the ultimate phenotype of a disease. These genes may potentially be markers for the development and severity of diseases (Kotnis *et al.* 2004). SNPs in these genes may affect functional activity and may thus pose an increased risk to cancerous growth (Kotnis *et al.* 2004). Genetic variation in key pathways may hold the key to explaining the initiation of malignancy, rate of carcinogenesis and the resistance or susceptibility of the individual (Kotnis *et al.* 2005).

Although many studies in Lynch syndrome, have investigated associations between the variation in the age of onset of malignancy and modifier genes (Bala and Peltomaki 2000; Maillet *et al.* 2000; Felix *et al.* 2006), no known modifier studies have been found with respect to variations in the site of cancer. The variations in the site of cancer may, in part, be explained by the primary predisposing mutation in the DNA MMR genes. Mutations in certain DNA MMR genes have been shown to predispose extracolonic cancers within families (Vasen

et al. 1996). However, this is not always the case. Although originally though to be mostly due to mutations in *hMSH2* and *hMSH6*, extracolonic cancers have been reported in carriers of germline mutations in h*MLH1* (Nystrom-Lahti *et al.* 1995; Thiffault *et al.* 2004), thus blurring the lines between the conservative phenotype (solely colorectal cancer) and the promiscuous phenotype (widest range of extracolonic cancers).

Recently, a promoter polymorphism in the DNA MMR gene *hMLH1* gene has been implicated in an increased risk of colorectal and endometrial cancers (Harley *et al.* 2008). This c.-93G>A polymorphism, occurs 93 bases away from the transcription start site of the gene.

The *hMLH1* c.-93G>A polymorphism has also been associated with an increased risk of microsatellite unstable colorectal cancer in two different populations (Raptis *et al.* 2007). The case-control study was undertaken in Ontario and New Foundland and the OR for the each populations' heterozygosity were 1.84 (95% CI: 1.20 – 2.83) (p = 0.001) and 2.56 (95%CI: 1.14 – 5.75) (p = 0.003) respectively.

More recently, Harley *et al.* (2008) reported an association between the *hMLH1* c.-93G>A polymorphism and risk for epithelial ovarian cancer (OR: 1.5, p = 0.0005). The study was performed on a mixed white North American population (Caucasian and Hispanic) in which the variant is present in at least 31% of the female population (Harley *et al.* 2008).

Beiner *et al.* (2006) examined the effect of *hMLH1* and *hMSH2* variants on the risk of developing endometrial cancer in a cohort of patients already diagnosed with endometrial cancer. The study found that women who carry at least one of the *hMLH1* -93 A alleles, are at a significantly higher risk than those without the

variant. Heterozygous (GA) women have an odds ratio (OR) of 1.4 (P = 0.008) and homozygous variant (AA) women have an OR of 1.9 (P = 0.009); whereas homozygous wildtype (GG) women have an OR of 1.

The aim of this particular study is to assess the effect of the *hMLH1* c.-93G>A polymorphism in a cohort of Lynch syndrome patients in order to investigate an association between the variant and site of cancer as well as age of onset of the affected patients.

2.2 Materials and Methods

2.2.1 Cohort selection

The cohort for this study was selected from the Colorectal Cancer Database in the Division of Human Genetics, at the University of Cape Town (UCT). The database contains information on individuals mostly from the Western Cape Province of South Africa, and was initiated in 1987 when the first family with Lynch syndrome was identified (Ramesar *et al.* 2000).

The database currently contains information on individuals recruited for investigation of DNA MMR genes (under 50 years (y) of age), or if they had a family history of colorectal cancer (CRC). In some instances, unaffected family members are also in the database. The largest families are usually those with a mutation already identified in the proband (usually affected). Information archived in the database includes that relating to demographic, clinical and pathological aspects.

For inclusion into the cohort, patients had to have a predisposing mutation in the *hMLH1* gene, specifically the C1528T founder mutation previously described in this cohort (Ramesar *et al.* 2000). Ethical approval of the cancer study was obtained. Patient consent was granted for the use of biological samples to be utilized in research studies by every patient. An example of the consent form is provided in Appendix 1.

Ninety-three individuals with the C1528T germline mutation in the *hMLH1* gene were included in the study. Of the individuals selected, 38 were affected with cancer and 55 were, at the time of study, unaffected (cancer-free).

As a control group, 97 individuals were chosen. Most of these patients did present with a Lynch associated malignancy at a young age and were subsequently recruited into the Lynch study cohort but on further investigation, no germline mutation within one of the three DNA MMR genes (*hMLH1*, *hMSH2* and *hMSH6*) was identified. Also included in the control cohort are those individuals who are related to a proband with a known predisposing germline DNA MMR gene mutation. In this control group, 77 were affected with cancer and 20 individuals were unaffected (cancer -free) at time of study.

2.2.2. DNA Isolation and Integrity

Genomic DNA was isolated from peripheral blood lymphocytes using the Puregene™ DNA Isolation Kit (Gentra Systems, USA). Two samples of blood were obtained by venepuncture from the patient in 10ml EDTA tubes. The DNA was isolated from both the samples, by the manufacturers' details.

The quality and quantity of the DNA was measured using a NanoDrop 1000 spectrophotometer. The concentration of the stock samples was calculated using NanoDrop1000 software (NanoDrop Technologies). The amount of protein and organic solvent contamination within the sample was reflected by the ratio of the wavelength for maximum absorbance of DNA (A_{260}) and the wavelength for the maximum absorbance of protein (A_{280}) or organic solvents (A_{230}). Viable stock solutions were diluted to a final working concentration of 100 nanograms per microlitre (ng/µl) for use in further investigation. The diluted samples were assessed for integrity by gel electrophoresis. Two hundred nanograms of the diluted sample was mixed with three microlitres of loading dye (Appendix 2.1) and then loaded onto a one percent agarose gel (Appendix 2.2). The gel was electrophoresed at 120 volts (V) for 30 minutes and visualised on a UviPro UviGold Transilluminator (Whitehead Scientific, USA).

2.2.3. Identification of Gene Sequence and Primer design

Information regarding the gene sequence, SNPs identified in the sequence and information about the gene were accessed from the NCBI Gene website (URL: http://www.ncbi.nih.gov) (Version GI:89161205) and the Ensembl website (http://www.ensembl.org) (Version Ensembl 48). Information from these websites were merged to create an annotated gene sequence. The program used for the annotation was Annotv9, created by Dr Rebello (Division of Human Genetics, UCT, Personal communication).

For analysis of the genomic regions of interest target fragments of these genes required amplification by the Polymerase Chain Reaction (PCR) (Mullis *et al.* 1986). The genomic region of interest in the present study was identified from the annotated gene sequence and the portion of the gene to be amplified was defined

(a minimum of 60 bases on either side of the region of interest was included). Ideally, the amplified product should be between 100 and 500 base-pairs long.

The selected region was interrogated with the primer design program, Primer 3 at the following URL: http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi. Primer 3 (Version 2004) is a free web-based program developed and supplied by the Whitehead Institute for Biomedical Research (Rosen and Skaletsky 2000). The following parameters were placed on the design of the primer using Primer 3: length of the primer was specified to be between 18 and 22 bases, the melting temperature of the primer between 55 degrees Celsius (°C) and 60°C, the maximum number of repeats to be included as three, and mispriming was not required.

The Primer 3 output was then analysed using the free web-based program IDT Oligo Analyzer. (URL: http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/). For the analysis of the primer pair, each half was assessed for the characteristics of the primer (e.g. melting temperature of the primer, length and GC content). The ability of the primer to form dimers with itself and the second primer in the reaction as well as the degree of self-complementarity of the primer was also assessed. Finally, a NCBI BLAST search was also performed to ensure the primer would not bind to other regions on the genome The Nucleotide-Nucleotide BLAST search was performed online at the NCBI webpage. The program searches sequenced genomes (in this case, the human genome was specifically chosen) for homologous areas to which the primer can alternatively bind (Altschul *et al.* 1990). The program can be found at the following URL: http://www.ncbi/nih.com/BLAST. The finalised primer set (seen in Table 4) was obtained from the Department of Molecular Cell Biology at UCT.

Table 4: Primer sequences

PRIMER	SEQUENCE	LENGTH	Tm	%GC	OD/ml
*hMLH1*_EXON1_F	GCTGTCCAATCAATAGCTGCCGC	23	59	57	4.265
*hMLH1*_EXON1_R	GCTGTCCAATCAATAGCTGCCGC	21	58	62	4.672

2.2.4. Polymerase Chain Reaction

2.2.4.1. PCR Optimization

Prior to amplification of the cohort and mutation detection the PCR was optimized to determine the optimum temperature, concentration of reagents and cycling conditions. The first optimization procedure performed was to run the reaction across a temperature gradient. For each PCR reaction a control sample was used in a standard PCR reaction as described. A volume of 25 microlitres (µl) of identical PCR reaction mix was placed in twelve PCR tubes. The Multigene PCR machine (Whitehead Scientific, USA) was set at a 1°C interval starts from a minimum temperature of 50°C.

2.2.4.2. Amplification

All reactions were made up to a final volume of 25µl. A standard reaction contained 100ng of template DNA, 0.1 units (U) of GoTaq DNA polymerase (Promega, USA), 1 x GoTaq Polymerase Buffer (pH 8.5, 1.5mM $MgCl_2$ per reaction), a final concentration of 200 micromolar (µM) deoxyribonucleotide (dNTPs) (Bioline, USA), a final concentration of 0.4µM of each primer (forward and reverse) and distilled water (Sabex) made up to the final volume.

The standard cycling conditions used for the PCR included: an initial denaturing step at 95° C for 5 minutes, 30 cycles of amplification, which involved a denaturing step at 95° C for 30 seconds; an annealing temperature of 57° C for 30 seconds and an elongation step at 72°C for 40 seconds. An additional elongation step at 72°C for 7 minutes ensured completion of the reaction.

To ensure the correct-sized fragment was obtained after amplification, a molecular weight marker was electrophoresed alongside the amplified products. The GeneRuler 100 base-pair (bp) molecular weight marker (Fermentas, USA) was employed for all resolution of fragments on an agarose gel. The marker was used at a concentration of 0.05 micrograms (µg)/ µl (Appendix 3).

2.2.5. Resolution of DNA fragments

PCR products and products of restriction enzyme digestion were detected by Agarose Gel Electrophoresis (AGE). AGE functions by exploiting the charge of DNA. DNA is a negatively charged molecule and when placed in an electric field, migrates towards the positive electrode, with smaller DNA molecules moving faster through the gel. DNA fragments migrate through a porous agarose gel on the basis of size (Sambrook *et al.* 1989). Low electroendosmosis agarose was supplied from Roche Diagnostics (Roche, Switzerland). Tris Borate EDTA (TBE) electrophoresis buffer (Appendix 2.3) was used.

A molecular weight marker, GeneRuler 100 base-pair (bp) molecular weight marker (Fermentas, USA), was loaded in the first lane to ensure the PCR products are of the correct size.

DNA bands were detected in the gel by the addition of Sybr Gold (Molecular Probes Invitrogen, USA) to the loading dye (Appendix 2.1) in order to obtain a working concentration of 0.04µg/ µl. This mixture was thoroughly vortexed and kept in the dark.

Sybr Gold (Molecular Probes Invitrogen, USA) intercalates between bases in the DNA molecule and as it is fluorescent, it can be easily detected with a transilluminator. However, due to the intercalation of the molecule into the DNA, migration of the amplicon will be distorted. As a result the fragments may migrate slower through the porous gels and seem larger than expected. Appendix 2.4. shows the Sybr Gold staining procedure.

Fragments obtained from preliminary experiments (integrity of cohort and temperature gradient) were detected by the use of Ethidium Bromide (EtBr) (Sigma, USA). The fluorescent dye EtBr was directly added during the preparation of the agarose gel (Appendix 2.5) to a final concentration of 0.25 milligrams (mg)/ milliliter (ml). The dye intercalates between base-pairs in the DNA molecule (Sambrook *et al.* 1989).

Agarose gels were electrophoresed at 160V for at least 30 minutes before visualization. The fragments were detected with the UviPro UviGold Transilluminator (Whitehead Scientific, USA).

2.2.6. Restriction Enzyme digestion

The *hMLH1* c.-93G>A variant in the promoter region of the *hMLH1* gene was investigated by restriction enzyme digestion. The web-based program, Webcutter,

was used to find the possible restriction endonuclease sites. Webcutter is an online service (URL: http://ma.lungberg.gu.se/cgi-bin/cutter2). An example of the output of the program can be found in the Appendix 4.

The restriction enzyme *PvuII,* isolated from the bacterium *Proteus vulgaris,* recognizes the site 5' CAGCTG 3' on a DNA sequence. Cleavage by the enzyme occurs in the middle of the site, between the G and C residues. The recognition site, present in the sequence of interest, is destroyed by the polymorphism investigated in this study. As a result, there is no cleavage of the sequence by the enzyme. The resulting products can be seen in Table 5.

PCR products were checked for amplification prior to digestion using a 1% agarose gel. A fifth of the PCR product together with 3µl of loading buffer (Appendix 2.1) was loaded and the gel was electrophoresed at 160V for 30 minutes. DNA was detected with Sybr Gold (Molecular Probes Invitrogen, USA) and AGE was performed; unless otherwise stated.

Preceding digestion of amplified DNA in the entire cohort, the function of the enzyme was tested on control samples of known cutting or non-cutting status. A standard digestion reaction consisted of 10µl PCR product, 1 x Buffer (corresponding to the enzyme used), 1 U of enzyme *PvuII* (Fermentas) and made up to a final volume of 20µl with distilled water (Sabex); unless otherwise stated. Digestion reactions were incubated overnight at 37°C. Post incubation; the products, together with loading buffer, were loaded onto a 3% agarose gel.

Table 5: Fragments expected in the restriction endonuclease (RE) digestion

GENE	R.E.	SITE	INITIAL PCR PRODUCT	FRAGMENTS EXPECTED	
hMLH1 *c.* -93 G>A	*PvuII* (37°C)	5' CAGCTG 3'	340bp	Wild-type (GG)	285bp, 55bp
				Heterozygous (GA)	340bp, 285bp, 55bp
				Mutant (AA)	340bp

2.2.7. Direct sequencing

Control samples utilized in the optimization procedures for the PCR and detection of the *hMLH1* promoter variant were of known status. The samples were amplified and then verified by direct sequencing. An automated version of the dideoxy-terminator method of sequencing (originally described by Sanger *et al.* 1977) was used.

Following PCR and AGE, the fragment was analysed and cycle sequencing was performed on the Perkin Elmer (Applied Biosystems, USA). The brightness of the DNA fragment in the agarose gel determined the amount of DNA to use in the cycle sequencing reaction: bright DNA bands correlated with 3µl for the sequencing reaction whereas faint DNA bands produced readable sequences with 5µl. Contents of the cycle sequencing reaction were as follows: 1 x terminator mix (Applied Biosystems, USA), less than 3.2µM of one of the primers, 3µl DNA, 1 x dilution buffer (Applied Biosystems, USA) and distilled water (Sabex) to a final

volume of 20µl. The standard cycling reaction involved a single denaturing cycle at 98°C for 5 minutes followed by 25 cycles of 95°C for 15 seconds, 50°C for 10 seconds and a final step of 60°C for four minutes.

Following cycle sequencing, the reaction was cleaned by Ethanol precipitation. The products of the cycle sequencing reaction were transferred to a 150µl Eppendorf microfuge tube. Two microlitres of Sodium acetate (3mM pH 5.5) and 50ul of 100% ethanol were added to the eppendorf. The solution was mixed well and incubated overnight at 4°C. Post incubation, the sample was centrifuged for 10 minutes at 10000 revolutions per minute (rpm). The supernatant was removed and 50ul of 70% ethanol was added to resuspend the pellet. The sample was centrifuged again for 10 minutes at 10000rpm and the supernatant was removed. The pellet was air-dried for 3 hours and then resuspended in Hi-Di™ Formamide (Applied Biosystems, UK). Samples were placed on the ABI Prism 3100 Gene Analyzer (Applied Biosystems, USA) and automated electrophoresis commenced. The 3100 Data Collection Software Version 1.1(Applied Biosystems, USA) was used to analyze the results of the sequencing reaction.

2.2.8. Statistical analysis

The mutation positive group (cases) and the mutation negative group (controls) were subdivided into affection status (“affected with cancer” and “unaffected (cancer-free)”) and gender. On examining the cohort by ethnicity, it was found that other than mixed ancestry, alternate ethnic groups made up a small proportion. Subsequently ethnic division was not applied.

All statistical tests were performed in Microsoft Excel (XP) and MedCalc. The odds ratio (OR), and relative risk (RR) ratios were performed on all data, where

possible. The Fisher's Exact test was also performed. Statistical results lower than or equal to p = 0.005 was considered as significant.

2.3 Results

2.3.1. Cohort Analysis

The study consisted of 93 cases (mutation positive individuals) and 97 controls (mutation negative individuals). Table 6 below represents a breakdown of the cohort into various subgroups. The cases consisted of 38 individuals affected with a Lynch associated malignancy (40.9% of the cohort), of which 47.4% (18 individuals) were female. Of the group of 38 affected mutation-positive individuals, 37 were of mixed ancestry.

In the control cohort, a total of 77 individuals were affected with a Lynch-associated malignancy (79.4%). With regard to affection status in the control cohort, 38 females (49.4%) were present. With regard to ethnicity, 62.3% (48 individuals) of were of mixed ancestry. Eleven individuals were of Black ethnicity, nine were Caucasian and the ethnicity was unknown for nine individuals.

Table 6: Cohort stratification

STATUS		NUMBER OF INDIVIDUALS
CASES: MUTATION POSITIVE (C1528T) INDIVIDUALS		93
AFFECTED WITH CANCER		38
MALE	20	
FEMALE	18	
BLACK	1	
CAUCASIAN	0	
MIXED ANCESTRY	37	
UNCLASSIFIED		
NOT AFFECTED WITH CANCER		55
MALE	21	
FEMALE	34	
BLACK	0	
CAUCASIAN	1	
MIXED ANCESTRY	54	
UNCLASSIFIED		
CONTROLS: MUTATION NEGATIVE INDIVIDUALS		97
AFFECTED WITH CANCER		77
MALE	39	
FEMALE	38	
BLACK	11	
CAUCASIAN	9	
MIXED ANCESTRY	48	
UNCLASSIFIED	9	
NOT AFFECTED WITH CANCER		20
MALE	7	
FEMALE	13	
BLACK	1	
CAUCASIAN	6	
MIXED ANCESTRY	10	
UNCLASSIFIED	3	

2.3.2. Integrity gel

Prior to amplification and digestion of the DNA obtained from the cohort, the concentration of the DNA and its integrity was assessed. DNA with a concentration of 100 ng/ µl was utilised.

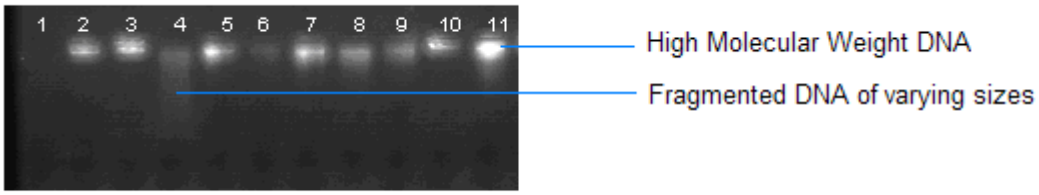


Figure 4: **AGE of the integrity of DNA samples DNA samples resolved on a 1% (w/v) agarose gel. The gel was electrophoresed at 160V for 30 minutes. The detection agent was ethidium bromide. Lanes 1 - 11: DNA Samples**

Figure 4 shows the integrity of a selected series of samples in the cohort. Lanes 1 and 6 shows no evidence of a DNA band. The samples electrophoresed in Lanes 2, 3, 5, 7, 8, 9 and 10, though the DNA is faint, show good integrity as a single band of is visible at the top of the well. Lane 4 shows fragmented DNA and this sample may not be viable for the study. Lane 11 shows DNA with very good integrity.

2.3.3. Temperature gradient

As a primary optimization step, a temperature gradient was performed on a control sample of DNA. A range of annealing temperatures, taking into account the

calculated primer melting temperatures, were chosen (50°C – 61°C) and the presence or absence of the correct sized product was observed.

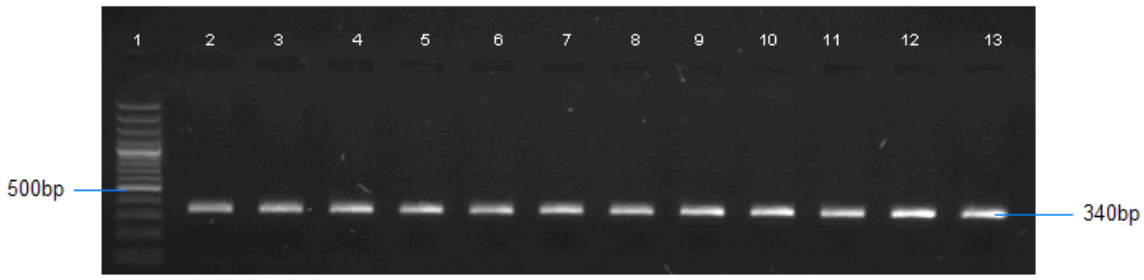


Figure 5: **AGE of the *hMLH1* temperature gradient resolved on a 1% (w/v) agarose gel. The gel was electrophoresed at 160 volts for 45 minutes. The detection agent was ethidium bromide. Lane 1: 100bp GeneRuler Molecular weight marker (Fermentas, USA). Lanes 2 – 13: PCR products from annealing temperatures 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 and 61°C, respectively.**

As observed in Figure 5 above, the fragment is adequately amplified at every annealing temperature. No non-specificity of the PCR was noted and no primer dimers were visible. The annealing temperature eventually used was 57°C, corresponding to Lane 9 in Figure 5.

2.3.4. Restriction Endonuclease digestion

In order to detect the SNP in the promoter region of the *hMLH1* gene, restriction endonuclease (RE) digestion was used. The G>A polymorphism at -93bp from the transcription start site destroys a *PvuII* recognition site.

Following amplification of the cohort and digestion of the amplicons with *PvuII* overnight at 37°C, resulting products were resolved by AGE. An example of the agarose gel can be seen in Figure 6.

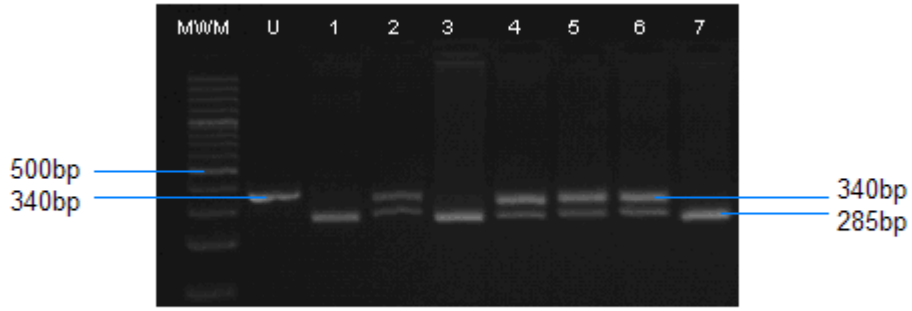


Figure 6: **AGE of restriction endonuclease digestion of *hMLH1* on a 3% (w/v) agarose gel. Products were detected with Sybr Gold. Electrophoresis was performed at 160V for 90 minutes. MWM: 100bp GeneRuler Molecular weight marker (Fermentas, USA). U: Uncut sample. Lanes 1 – 7: Digested samples.**

A number of heterozygous individuals were identified in Figure 6. Lanes 2, 4 – 6 showed the digested bands of the PCR product therefore reflecting heterozygous individuals. The enzyme can only cleave the wild type sequence and thus results in half the fragment being cleaved while the other half not. The remaining individuals are wild type due to the presence of a cut sample. Lane 1 represents an uncut sample. This sample serves as a reference of the amplified sample at 340bp. Lanes 1, 3 and 7 represent homozygous variant samples seeing that both alleles have been cleaved.

2.3.5. Distribution of Genotypes and Alleles within the study cohorts

The results of the restriction enzyme digest have been tabulated below (Table 7).

Table 7: ***hMLH1*** **promoter genotypes in the cohort under investigation**

STATUS	GENOTYPE			TOTAL
	HOMOZYGOUS WILD TYPE	HETEROZYGOUS	HOMOZYGOUS VARIANT	
	GG	GA	AA	
MUTATION POSITIVE				**93**
AFFECTED	*33*	*5*	*0*	*38*
MALE	17	3	0	20
FEMALE	16	2	0	18
UNAFFECTED	*43*	*10*	*2*	*55*
MALE	16	4	1	21
FEMALE	27	6	1	34
TOTAL	*76*	*15*	*2*	*93*
MUTATION NEGATIVE				**97**
AFFECTED	*43*	*31*	*3*	*77*
MALE	20	18	1	39
FEMALE	23	13	2	38
UNAFFECTED	*13*	*5*	*2*	*20*
MALE	5	0	2	7
FEMALE	8	5	0	13
TOTAL	*56*	*36*	*5*	*97*

2.3.6. Statistical Analysis

To test for any association between site of cancer and the *hMLH1* promoter genotype, the OR and RR of the heterozygous and homozygous variant genotypes were calculated. Significance was calculated by the Fisher exact test (MedCalc). Table 8 below shows the results of the site of cancer statistical analyses.

Table 8: Statistical analysis of h*MLH1* promoter genotypes in the cohort under investigation

SITE OF CANCER	GENOTYPE	MUTATION POSITIVE	MUTATION NEGATIVE	OR (95% CI)	P VALUE	RELATIVE RISK (95% CI)
Colorectal	GG	28	43			
	GA/ AA	2/ 0	31/ 3	1.9295 (0.6297-5.912)	P = 0.0001	1.2028 (0.8471 – 1.7079)
Endometrial	GG	3	0		Not applicable	
	GA/ AA	1	0			
Other*	GG	3	0		Not applicable	
	GA/ AA	1	0			
No cancer	GG	43	13			
	GA/ AA	10/ 2	5/ 2	1.9295 (0.6297-5.912)	P = 0.3676	1.2028 (0.8471 – 1.7079)

*Other consists of breast (2), renal (1), skin (1) carcinomas

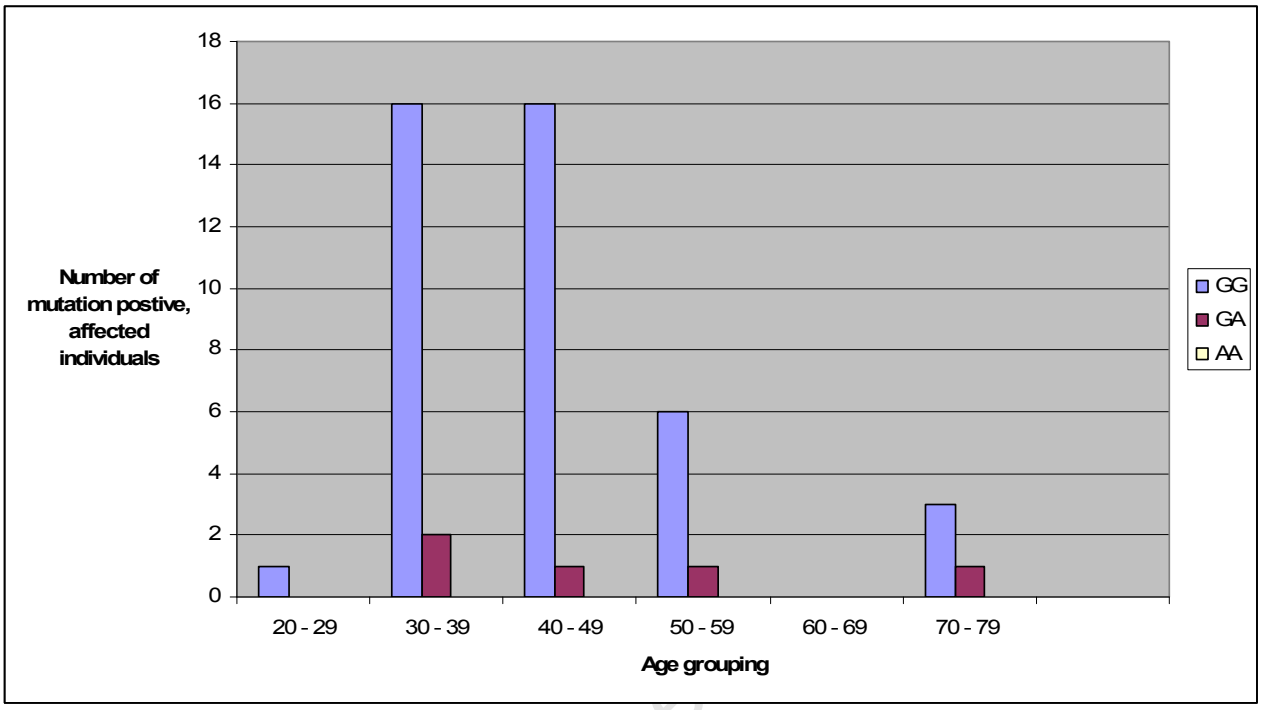


Figure 7: **Graphical representation of the distribution of the genotypes within the affected cases (mutation positive) cohort.**

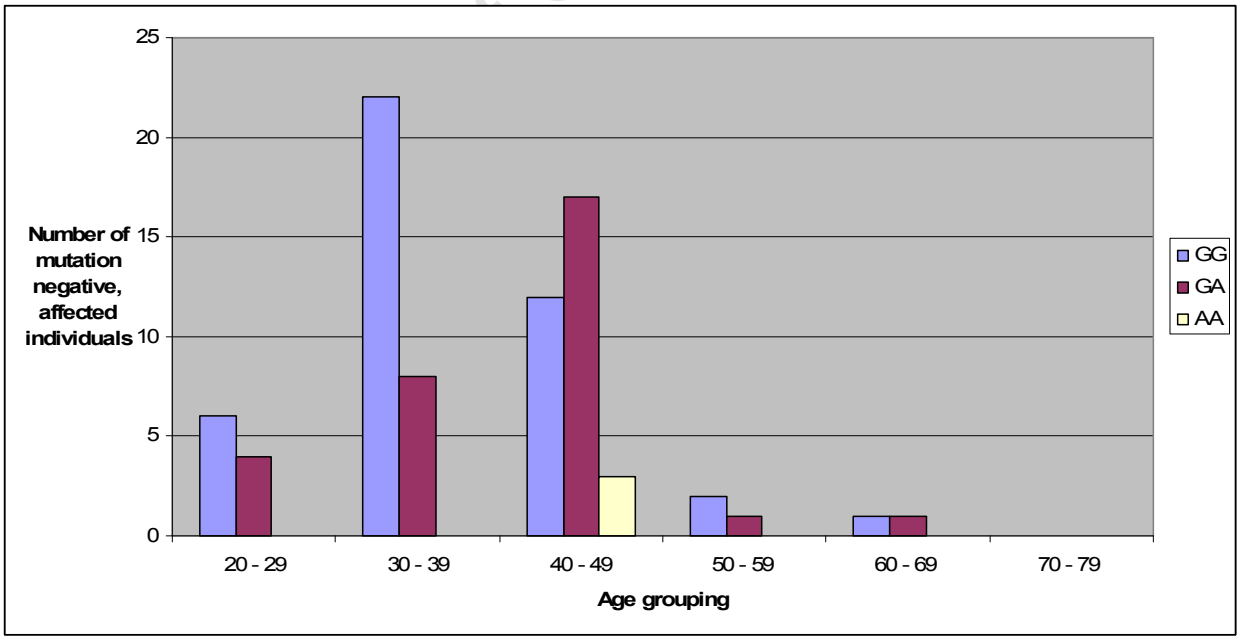


Figure 8: **Graphical representation of the distribution of the genotypes within the affected control (mutation negative affected) cohort.**

The age of onset was also assessed in the cohorts. The results of the age stratification can be observed in Figures 7 and 8. As observed in Figure 7, within the mutation positive affected cohort, no homozygous variant genotype was evident. A large proportion of the cases fell between the age groups of 30y and 50y. Typical age of onset for Lynch syndrome patients is between 35y and 45y (Lynch and De La Chapelle, 2003) and the cohort data corresponds to this. There are only a few heterozygous genotypes evident in the case cohort. In comparison, Figure 8 shows the number of patients with heterozygous genotypes outnumbers the homozygous wild type genotypes in the 40 – 49 age group in the mutation negative cohort. There are also two homozygous variant genotypes present in this age group.

2.4 Discussion

The aim of this study was to ascertain whether or not a SNP in the promoter of the *hMLH1* gene was associated with the site of malignancy in a control cohort and a cohort of individuals predisposed to Lynch syndrome. In addition, an analysis of the age of onset of the cohort was also assessed with respect to their *hMLH1* promoter genotype. The polymorphism was detected with the cohorts by restriction enzyme digestion, which followed successful amplification of the *hMLH1* product.

Lynch syndrome is phenotypically heterogenous – however, there is no molecular basis for the tumour spectrum. In light of this, the molecular pathology of each of the tumour sites and corresponding germline features are being investigated and follow in the next chapter. However, the effects of modifying genes are not to be ignored. As a result, the *hMLH1* promoter – implicated in recent studies (Harley *et al.* 2008, Beiner *et al.* 2006) as being associated with an increased risk of

colorectal, endometrial or ovarian cancer – was investigated in our cohort as a potential modifier of site of cancer.

Due to the small sampling size of our cohort and the rigid inclusion methods (only individuals with a specific *hMLH1* mutation was chosen for the case cohort), only colorectal and gynaecologic cancers (namely endometrial and ovarian) were prevalent in both the cases and controls. The mutation positive cases did contain two breast carcinomas, a renal carcinoma and a skin malignancy. Statistical analyses could not be performed on the endometrial and "other" sites of cancer as none of these malignancies were present in the mutation negative cohort (Table 8.). The heterozygous genotype and homozygous genotypes were pooled for all statistical analysis due to small numbers.

As the site of malignancy, colorectal cancer has the highest number of affected individuals, in both the cases and controls. This is due to the fact that most of these individuals are probands and have been initially recruited into the program for DNA MMR gene screening.

Due to the skewed data, the OR for the wild type genotype (GG) and the occurrence of colorectal cancer is 11.0698 (95% CI: 2.46 – 49.78; p = 0.000181). The statistics imply that mutation positive individuals homozygous genotype (GG) are at an 11 times increased risk of colorectal cancer. However, it is known the mutation positive individuals are already predisposed to this type of cancer and this may therefore create a bias toward the mutation positive individuals. Though the result reached significance, the sample numbers are too small to justify the result and larger study cohorts are required. In addition, there is a large discrepancy in the limits of the confidence interval.

In contrast, the RR (the relative risk of carrying the GG and having colorectal cancer in the cases (mutation positive group) compared to the controls (mutation negative group) is 1.2028 (95%CI: 0.8471 – 1.7079). This indicates that the mutation positive group is 1.2 times more likely not to develop a malignancy when compared to the case cohort.

The protective effect of the SNP has not been described. Functional studies, previously performed on the *hMLH1* gene, have placed the SNP within the promoter region (Ito *et al.* 1999). The promoter region is crucial for effective transcription of the gene and is it hypothesized that the c.-93G>A SNP may dysregulate transcriptional activity of the gene. In addition, the SNP is located within a CpG island (Deng *et al.* 2001), which is commonly methylated in sporadic, microsatellite unstable cancers. It may therefore be possible that the SNP may "protect" the control group, as shown in the no-cancer statistic, by preventing methylation of the site resulting in legitimate transcription of the gene.

Previous studies (Beiner *et al.* 2006; Raptis *et al.* 2007; Harley *et al.* 2008) have reported an association between the *hMLH1* promoter polymorphism G>A, at position -93, and an increased risk in either the development of ovarian or colorectal cancer in the general population. As a result, these studies disprove the hypothesis of the protective effect of the SNP; though a larger cohort should be assessed for clarification of the results observed in this study.

As mentioned previously, the observed results may be due to a number of limiting factors such as bias in the cohort selection, small sample sizes and lack of follow-up in the mutation negative cohorts (specifically due to the age of onset). These factors are further discussed below.

2.4.1. Limitations of the study

All individuals screened in the study were randomly chosen. The only basis for selection was positive confirmation of the C1528T mutation in the *hMLH1* gene for the mutation positive individuals and no known germline mutation in the DNA MMR genes, *hMLH1*, *hMSH2* and *hMSH6,* despite intensive screening.

A bias is evident in the mutation negative cohort as there are a larger proportion of affected individuals when compared to the mutation positive cohort. The mutation positive cohort consists of not only the proband, who is usually affected, but also family members, as yet unaffected but who have been shown to have the predisposing mutation.

In contrast, the mutation negative cohort consists of patients with a Lynch associated cancer and in most cases an early onset of disease – in some cases, family members have been recruited for the <50y study. However, on molecular investigation, no Lynch syndrome predisposing mutation has been found.

The age of onset of affected individuals were taken at age of diagnosis. For those unaffected, the age was calculated to date. For mutation positive individuals, this method of calculating the current age was adequate as these individuals routinely undergo surveillance. However, the surveillance program may introduce a bias in the study as it may potentially delay age of onset of a malignancy. In mutation negative individuals, this method may introduce a bias in the study as these individuals do not undergo the surveillance programs and therefore current affection status is unknown.

Only primary malignancies were regarded in the analyses of the study. Metastases and secondary malignancies were ignored.

The statistical tests were performed as though the individuals in each group were unrelated. However, some of the individuals are blood relatives. Therefore, more stringent statistical tests should have been applied taking the Identity by Descent (IBD) of some of the individuals into account. However, due to the lack of significance of the study, IBD statistics did not need to be applied.

2.4.2. Future prospects

A large, multicentre, national study would ideally be performed in order to obtain clearer results as to the effect of the SNP in the general population, as well as those already predisposed to Lynch syndrome.

Harley *et al.* (2008), Raptis *et al.* (2007) and Beiner *et al.* (2006) have performed such studies and successfully shown an increased risk of various malignancies in those carrying the polymorphism in the general population. This knowledge is of great importance as it provides a guideline to those individuals carrying the variant in terms of screening and management programmes therefore limiting strain on health-care providers.

Chapter 3:

Elucidation of a genetic signature

3.1. Introduction

Pathologically colonic and extracolonic tumours arising in Lynch syndrome patients are similar. These tumours are poorly differentiated, have a large number of tumour infiltrating lymphocytes and signet-cell features (Jass 2007). Molecularly, in individuals with a predisposing mutation in one of the DNA MMR genes, acquisition of a second hit to that gene is hypothesized to result in the initiation and subsequently the progression of malignancy at the site at which the second hit was acquired (Knudson 1971; Chao and Lipkin 2006).

The tumourigenic process occurs via the mutator pathway, discussed previously in Section 1.4.1. In essence, the study assesses the molecular hallmarks of Lynch – associated malignancies. To investigate whether the spectrum of extracolonic cancers in Lynch syndrome are a direct result of the DNA MMR deficiency, the mutation profile of these cancers was assessed. The working hypothesis of the study assumed that should all malignancies be initially generated from the mutator pathway, resulting tumours ought to demonstrate the same genetic signature, regardless of the site of cancer.

Microsatellite Rich Panel of Genes

The instability of repeated sequences in Lynch syndrome, due to the mutator pathway, is well documented. Numerous studies have focused on the volatility of such sequences in microsatellite unstable cancers as mutational targets (Duval *et*

al. 2001, Miyaki *et al.* 2001, Duval *et al.* 2002, Vilkki *et al.* 2002, Woerner *et al.* 2005, Jung *et al.* 2006, Bertholon *et al.* 2006).

It is imperative to distinguish between target mutations (occurring at the second level of the mutator pathway) and passenger mutations (those mutations which do not directly contribute to the carcinogenic phenotype). Five criteria used in order to define target genes were proposed at a Bethesda Consensus Meeting in 1997. The criteria for the genes state that: the genes should have a high rate of frequency, biallelic inactivation, a role in the suppressor pathway, the occurrence of alterations within the same pathway in MSI-negative tumours and functional data (Boland *et al.* 1998).

To date, there are no official genes adhering to these criteria, though there are certain genes which appear to be mutated at a higher frequency in Lynch associated malignancies than in their sporadic counterparts. For example, the Transforming growth factor receptor type 2 gene (*TGFBR2*), the Bcl-apoptosis related protein X (*BAX*) gene, the phospotensin homolog (*PTEN*), as well as the DNA MMR genes *hMSH3* and *hMSH6* have all been implicated in the mutator pathway (Zhou *et al.* 2002, Fernandez-Peralta *et al.* 2005), though the exact sequence of the mutation pathway is yet to be elucidated.

Miyaki *et al.* (2001) examined 29 MSI-H colorectal cancers from Lynch syndrome patients for alterations of microsatellites in a number of genes. The study, analysing different pathological stages of the tumours, concluded that the number of altered genes increased as the tumour progressed through malignancy and that alterations upstream of the gene was also evident (Miyaki *et al.* 2001). Both Duval *et al.* (2002) and Woerner *et al.* (2005) utilised statistical algorithms to select for target genes somatically mutated in Lynch syndrome. Somatically acquired mutations are hypothesised to be tissue specific in MSI-H cancers by

Duval *et al.* (2002), though the authors have stated there are a few genes commonly altered. Woerner *et al.* (2005) utilised the Selective Target Gene algorithm, developed by the authors, to obtain 26 microsatellites occurring within coding regions of genes (this is known as coding microsatellites (cMS)). Through examination of MSI-H colorectal cancers, the authors found an increase of target gene mutations through the evolution of the malignancy from an adenoma to a carcinoma. Mutations within certain genes, such as *BAX* and *TGFBR2*, showed instability at early stages of tumourigenesis while others (for example *PTHL3*) showed instability more frequently in carcinomas (Woerner *et al.* 2005). These studies have attempted to classify the sequence of mutations within the mutator pathway. It is evident from these studies that MSI is prevalent in Lynch-associated cancers however exact gene mutations are still elusive. The aim of this study is to find a consensus in genes commonly mutated in order to obtain a genetic signature of Lynch – associated tumours. Elucidation of the genetic signature will provide information on the nature of the tumour which will in turn be useful in diagnosis, treatment and management.

The presence and degree of MSI in tumour tissue is currently assessed by means of testing via the Bethesda panel of markers. The Bethesda panel is a range of five highly polymorphic markers. These markers encompass two mononucleotide and three dinucleotide repeats, have been internationally utilised since its recommendation by the National Cancer Institute in 1977 (Murphy *et al.* 2006).

Methylation Specific Multiplex Ligation – Dependent Probe Amplification (MS - MLPA)

Ms-MLPA was first described in 2005 by Nygren *et al.* (2005). An overview of the method is discussed in section 3.2.5. The technique has been successfully

described utilising DNA isolated from paraffin-embedded/ formalin-fixed tumour tissue (Henken *et al.* 2007). In addition, the original MLPA technique is accepted as a simple and reliable method of detection of changes within DNA sequences (copy number changes, deletions and insertions) for DNA derived from blood, amniotic fluid (Nygren *et al.* 2005). An advantage of the Ms-MLPA is that in addition to the methylation status of the sample, copy number changes can also be recorded. Together, the methylation status and the copy number changes may prove informative for prognostic results and further management of the patient.

Promoter methylation of 26 tumour suppressor genes is examined in the Ms-MLPA kit. Worsham *et al.* (2006) and Henkel *et al.* (2007) successfully utilised the Ms-MLPA technique to identify genes targeted in cervical and head and neck squamous cell carcinomas respectively. Epigenetic changes, such as methylation of promoters, of tumour suppressor genes results in silencing of that gene. As a result, the cell checkpoint mechanisms are compromised hence providing a suitable environment for tumourigenesis.

The possible elucidation of the genetic signature is the goal of this chapter. Both germline DNA and corresponding tumour material will be assessed to provide both inter- and intra- sample analysis. Two methods are utilised to obtain a genetic signature; namely, the stability of a panel of microsatellite rich genes Methylation Specific Multiplex Ligation-Dependant Probe Amplification (Ms-MLPA).

3.2. Materials and Methods

3.2.1. Cohort selection

The cohort for the study was selected from the Colorectal Cancer Database housed in the Division of Human Genetics, at UCT. As a pre-requisite for inclusion into this study, patients had to be predisposed to Lynch syndrome with the founder C1528T mutation in the *hMLH1* gene. Additional criteria for inclusion into the study were: that the patient be affected with a Lynch syndrome-associated cancer, of which the pathological grading be Dukes stage B. Patient consent was granted for the use of biological samples by every patient.

3.2.2. Selection of microsatellite rich genes

For a thorough elucidation of microsatellite genes, initially literature was mined. Microsatellite sequences that have been implicated as mutational targets in a number of published reports were noted (Duval *et al.* 2001; Miyaki *et al.* 2001; Duval *et al.* 2002; Vikki *et al.* 2002; Woerner *et al.* 2005; Jung *et al.* 2006; Bertholon *et al.* 2006).

The genes obtained from the published studies were then interrogated. The type of repeat (mononucleotide, dinucleotide and trinucleotide), function of the gene product (and biological pathways) and observed mutations were identified.

Thereafter, a bioinformatic tool, SUSPECTS (http://www.genetics.med.ed.ac.uk/suspects/), was employed to determine lists of genes common in Lynch syndrome and in sporadic forms of its extracolonic manifestations. The lists were then manually curated and the list of genes

identified from the literature was compared. The SUSPECTS program selects genes by scoring each gene according to expression libraries and Gene Ontology by relating them to the disease of interest. Microsatellite rich genes were included in the study if the genes were found to be implicated in Lynch syndrome and its extracolonic manifestations as well as its functional process.

3.2.3. DNA Isolation

3.2.3.1. Germline DNA

DNA was isolated from peripheral blood lymphocytes, as described in Chapter 2, Section 2.2.2.

3.2.3.2. Formalin-fixed paraffin embedded tumour tissue

Sections of archived tumour material were obtained from the Department of Anatomical Pathology (UCT). Requested tumour material was received as 5 µm sections. A haematoxylin and eosin stain accompanied each slide. The Roche High Pure PCR Template Preparation Kit (Roche, Switzerland) was employed in the isolation of DNA from formalin fixed, paraffin embedded tumour tissue. A few amendments were made to the original protocol, supplied by the manufacturer. These will be further described below.

Initially, each slide was soaked in 100% xylol (Roche, Switzerland) for 45 minutes, to deparaffinise the slide. The tissue section was then submerged in a series of varied ethanol concentrations for 10 seconds each, in order to dehydrate the tissue. Dehydration was carried out in a stepwise fashion with 100% ethanol (Merck, Germany), followed by; 80% ethanol, 60% ethanol and 40% ethanol.

Finally, the tissue section was submerged in double distilled water to ensure rehydration of the material.

The tumour material was then scraped off the slide with a razor blade. The material was placed in a 1.5ml microcentrifuge tube. Tissue lysis buffer (100µl) (Roche, Switzerland) and Proteinase K (20µl) (Roche, Switxerland) were added to the tissue followed by vortexing and incubation at 37°C for at least eight hours.

Subsequently, 10µl of Proteinase K was added to the microcentrifuge tube and the sample was incubated at 55°C for at least one hour. A volume of 100µl binding buffer (Roche, Switzerland) was then added to the sample and vortexed followed by incubation at 72°C for 10 minutes. Thereafter, 50µl of isopropanol (Merck, Germany) was added to the sample and thoroughly mixed. The sample was then pipetted into the upper reservoir of a High Pure filter collection tube (Roche, Switzerland) and subsequently centrifuged for one minute at 8000 rpm. The collection tube, with the flow through was then discarded and the filter tube was combined with a new collection tube.

A volume of 250µl Inhibitor removal buffer (Roche, Switzerland) was added to the upper reservoir of the filter tube and the sample was centrifuged at 8000 rpm for one minute. The filter tube was placed in a new collection tube to which 250µl of the Wash buffer was added to the filter tube. The sample was centrifuged at 8000 rpm for one minute. The collection tube and flow through was discarded and the Wash buffer step was repeated and the collection tube was retained. The filter tube and collection tube were centrifuged for ten seconds at maximum speed to remove the residual Wash buffer from the filter tube. The collection tube was then discarded and the filter tube was inserted in to a clean 1.5ml tube. Elution buffer (100µl) was pre-warmed to 70°C and added to the filter tube, which was

subsequently centrifuged at 8000 rpm for 1 minute. The eluent was stored at 4°C until further use.

3.2.4. Primer design and Polymerase Chain Reaction (PCR)

3.2.4.1. Multiplex PCR

Multiplex PCR involves the simultaneous resolution of multiple fragments, of differing length DNA. Primers amplifying different regions of the genome are added to the reaction mix and thus different regions are amplified. These primers were designed and analysed via the programs mentioned above (Section 2.2.3).

The freeware, web-based program Autodimer (http://www.cstl.nist.gov/div831/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm) was utilised in order to determine the compatibility of the primer pairs in a multiplex reaction. Secondary structures of the primers (hairpins, homodimers and heterodimer formations) were analysed with Autodimer. The core was decreased from seven to three, in order to increase the stringency of analysis. All other default parameters were applied. Further information regarding these primers can be seen in Table 9.

Primers were ordered from the Department of Molecular and Cell Biology (UCT). The forward primer of each gene was tagged fluorescently, either with a Hex or Fam molecule. The fluorescent markers on the primers allowed for automated electrophoresis of the PCR products. All reactions were performed on the PX2 Thermocycler (Thermo Electron Corporation, USA).

Table 9: Information on primers utilised in the project

GENE	CHR	SEQ	Tm	LENGTH	FRAGMENT (bp)
*ACVR2*_F	2	Hex - CACTTGTTGTAGGGTCAGT	49	19	130
*ACVR2*_R		GCAACAAAAATATACTTCAGGGC	52	23	
AR_F	X	Hex - GTTCCAGAGCGTGCGCGAGTGATCCAG	66	27	273
AR_R		TGTGGGGCCTCTACGATGGGCTT	61	23	
*BAX*_F	19	Fam- CACTTGTTGTAGGGTCAGT	53.1	19	131
*BAX*_R		CACTCGCTCAGCTTCTT	50	17	
CASP5_F	11	Hex - ATGTGTGTTATTCGCTGGAG	50	20	340
CASP5_R		CTTTCGCAAAGAGTCTACCAAG	53	22	
ERA_F	6	Fam -GTATCAGAGTCTATTAGGCTG	53	21	402
ERA_R		CCA AAC ATA ACC TCA GGT CA	52	20	
ERB _F	14	Hex - GGTAAACCATGGTCTGTACC	56	20	177
ERB _R		CAAAATGTTGAATGAGTGGGCCT	53	23	
IGF2R_F	6	Hex - ATGCCTCATGTCATCATACACG	50	22	370
IGF2R_R		ACAGTAGAAGAAGATGGCTGTG	50	22	
MBD4_F	3	Fam -ATGCTGAAAGTGAACCTGTTGC	53	22	244

Table 10 contd.: Information on primers utilised in the project

GENE	CHR	SEQ	Tm	LENGTH	FRAGMENT (bp)
MBD4_R		CTCGTTGTGTTCTGAGTCTTTG	53	22	
MSH3_F	5	Fam -CTGGATGATGCTGTAAATGTTG	51	22	339
MSH3_R		CTCATGGCTACTTGGTACTTGT	52	22	
MSH6_F	2	Hex - GGTGATGGTCCTATGTGTCG	54	20	220
MSH6_R		TTTGGTCCAGTAACAAGCACAC	53	22	
PTEN_EXON7_F	10	Fam -CCACACGACGGGAAGACAAG	56	20	363
PTEN_EXON7_R		TTCTCAGTTAACCATCCTTGTTAAG	53	25	
TAF1B_F	2	Fam -CAAATAAAAGCCCTCAACCGG	52	22	469
TAF1B_R		GGTTAATTATATGTACTCACTG	50	21	
*TGFBR2*_F	3	Fam -CTAGAGACAGTTTGCCATGACC	55	22	187
*TGFBR2*_R		TGTTGTCATTGCACTCATCAGA	51	22	
TP53_F	17	Hex - CTAAAAGGAAATCTCACCCCATC	53	23	461
TP53_R		AGAATGTAAAAGATGTTGACCCTTC	53	25	

Table 11: Components of the PCR reaction

REAGENT	CONCENTRATION
Gene 1: Forward Primer	20uM
Gene 1: Reverse Primer	20uM
dNTPs	200uM
Buffer	1 x
Taq	1 U
DNA	100ng/ul
dH2O	Up to 25ul

Three multiplex PCRs were performed. Multiplex 1 contained the following genes: *AVCR2, hMSH6, AR, CASP5* and *IGF2R*. Multiplex 2 contained *TGFBR2, MBD4*, *hMSH3, PTEN, TAF1b* and *Erβ* . The final multiplex reaction contained the genes *TP53, Erα* and *BAX*. The components of a standard PCR can be seen in Table 10. Additional primer pairs for the genes were added to the reaction mix and the final amount of distilled water (SA*BAX*) was adjusted accordingly.

The cycling conditions for the multiplex reactions were as follows: an initial denaturation step at 95°C for 5 minutes; 35 cycles of: 95°C for 30 seconds, annealing temperature for 30 seconds, 72°C for 40 seconds. Finally an extension period of 72°C for 7 minutes completed the reaction. The annealing temperature of each multiplex reaction was unique. Multiplex 1 had an annealing temperature of 53°C, Multiplex 2 had an annealing temperature of 49°C and Multiplex 3 had an annealing temperature of 55°C.

Tumour DNA PCR

Successful amplification of the tumour samples required additional optimization. Briefly, 500ng/µl of tumour DNA per 25µl reaction was utilised in the study. The cycling conditions for amplification of tumour DNA were slightly altered. The reaction commences with an initial denaturation step at 95°C for 5 minutes; 40 cycles of: 95°C for 30 seconds, annealing temperature for 30 seconds, 72°C for 40 seconds; and 72°C for 7 minutes.

3.2.4. Genotyping

In preparation for analysis through the ABI Prism 3100 Genetic Analyser (Applied Biosystems, UK), the samples were checked on an agarose gel and diluted one in ten if necessary. One microlitre of the sample (either neat or diluted) was placed together with 8.6µl Hi-Di™ Formamide (Applied Biosystems, UK) and 0.4µl GeneScan™ 500 ROX™ Size Standard (Applied Biosystems, UK) in a microtitre plate. Hi-Di™ Formamide (Applied Biosystems, UK) stabilises single strands of denatured DNA. GeneScan™ 500 ROX™ Size Standard (Applied Biosystems, UK) is fluorescently labeled and allows for correct sizing of the DNA fragment (the sizing can be found in Appendix 5) .

The samples were then denatured at 95°C for 5 minutes on the Hybaid Touchdown thermocycler (Hybaid Limited, UK), and snap-frozen on ice. Thereafter, the samples were placed in the ABI Prism 3100 Genetic Analyser, and the fragments were resolved by automated fluorescent-based, capillary electrophoresis. The Applied Biosystems GeneScan software program was used for the electrophoresis and the parameters were as follows: 3100 project, dye set D, run module Genescan36POP4.DEF and analysis module GS2500Analysis.gsp.

The data (FSA output files) were subsequently analysed using the Gene Mapper software (Applied Biosystems). Bins were created in the regions of interest, with a 10 base-pair region on either side of the actual fragment size.

3.2.5. Methylation Specific Multiplex Ligation – Dependent Probe Amplification (MS - MLPA)

The Ms-MLPA ME100B kit was purchased from MRC Holland and was performed as per the quantification protocol provided (MRC-Holland, Version 09; 2008). An overview of the procedure is provided in Figure 9.

Briefly, there are four parts to the reaction: firstly DNA is denaturated and there is hybridisation of the provided SALSA probes, a second stage of ligation and digestion, the third stage is a PCR and finally separation of the amplification products by automated electrophoresis.

The DNA sample, at a concentration of at least 100ng (germline and tumour DNA) (resuspended in distilled water), was heated at 98°C for 10 minutes. Thereafter, the samples were allowed to cool at 25°C for two minutes before 3µl of a mixture of the SALSA-Probe Mix (1.5µl) and MLPA Buffer (1.5µl) was added to each DNA sample. The samples were mixed before incubation at 95°C for one minute and 60°C for 16 hours in the GeneAmp 500 thermocycler (Applied Biosystems, UK).

Following incubation, DNA samples were brought to room temperature (25°C). 13µl of the buffer mixture was added to each tube (3µl Ligase buffer and 10µl distilled water). This mixture was thoroughly mixed before 10µl of the sample was transferred to a second tube, labelled M. Both tubes were incubated at 49°C for a minute. Thereafter, while at 49°C, 10µl of the Ligase-65 mixture (1.5µl Ligase-65

buffer B, 8.25µl distilled water, 0.25µl Ligase-65 enzyme) was added to the first tube – this served as the copy number test. At the same temperature, 10 µl of the ligase-digestion mix (1.5µl Ligase-65 buffer B, 7.75µl distilled water, 0.25µl Ligase-65 enzyme and 0.5µl *HhaI* enzyme (Promega R6441, 10/ µl)) was added to the tube labelled M. Both tubes were incubated at 49°C for 30 minutes followed by 98°C for 5 minutes.

The next stage of the Ms-MLPA procedure was PCR. Each DNA sample has two reactions within the Ms-MLPA; the copy number test and the methylation test (which involves digestion of the sample with *HhaI*). The PCR buffer mixture (2µl SALSA PCR buffer and 13µl distilled water) was added to new tubes (each sample has two new tubes). Five microlitres of the MLPA digestion-ligation or the MLPA ligation reaction was placed in the tubes with the PCR buffer. The tubes were then placed on ice while 5µl of the Polymerase mix (1µl SALSA PCR-Primers, 1µl SALSA enzyme dilution buffer, 2.75µl distilled water and 0.25µl SALSA Polymerase) was added to each tube. With the addition of the Polymerase mix into each tube, the tube was immediately placed in a preheated thermocycler (72°C). Once all tubes were placed in the thermocycler, the PCR was started with the following cycling conditions: 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 60 seconds; 1 cycle of 72°C for 20 minutes.

Following the PCR, samples were placed in the ABI 3100 for automated electrophoresis. One microlitre of the sample was placed together with 8.6µl Hi-Di™ Formamide (Applied Biosystems, UK) and 0.4µl GeneScan™ 500 ROX™ Size Standard (Applied Biosystems, UK) (Appendix 5) in a 96-well microtitre plate. The samples were then denatured at 95°C for 5 minutes on the Hybaid Touchdown thermocycler (Hybaid Limited, UK), and snap-frozen on ice. Thereafter, the samples were placed in the ABI Prism 3100 Genetic Analyser.

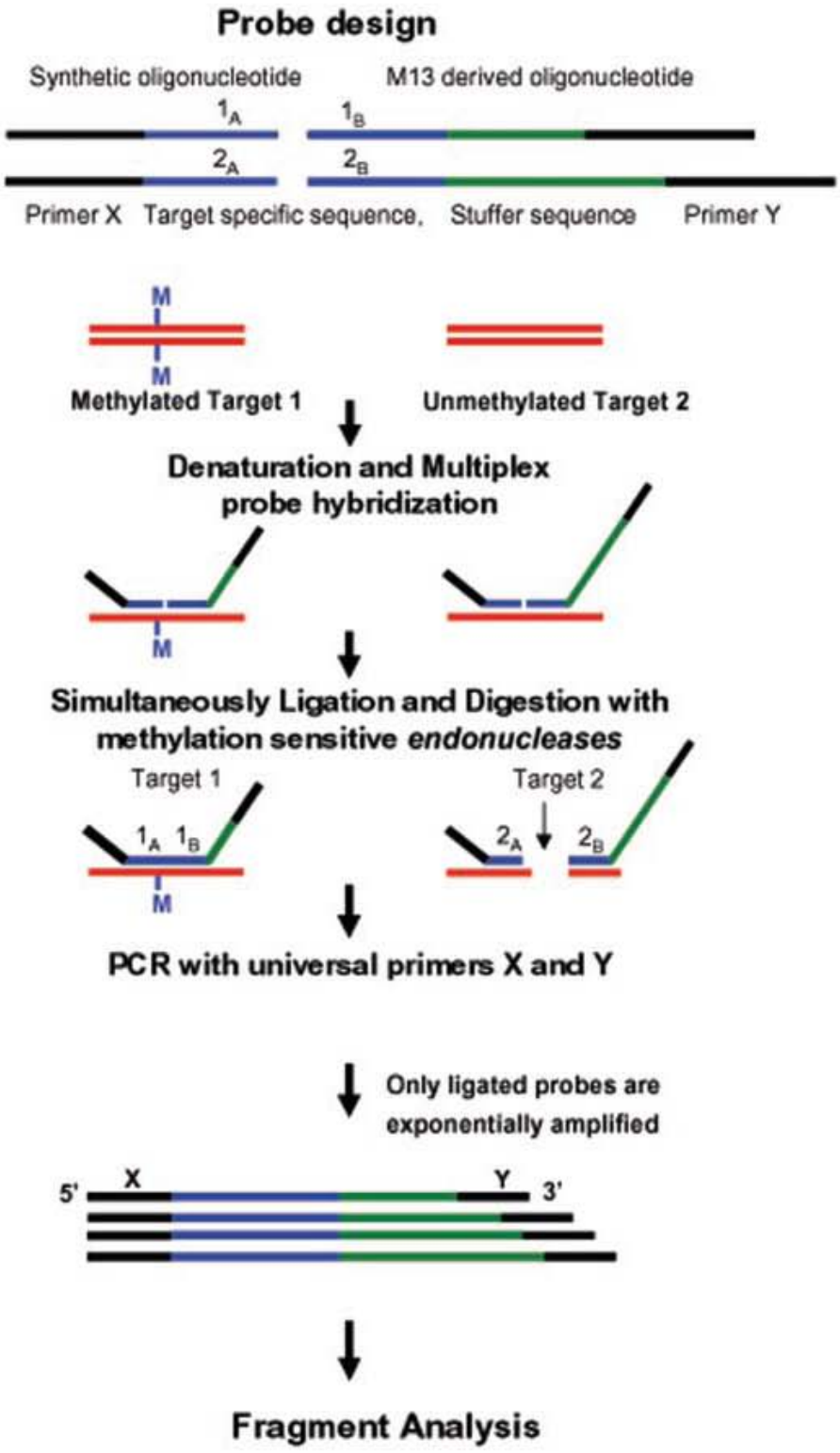


Figure 9: **Diagrammatic overview of the Ms-MLPA procedure, obtained from Nygren *et al.* (2005).**

Analysis of results

Ms-MLPA results were initially analysed and manually curated with GeneMapper software (Applied Biosystems, version 3). The sizing of each sample was exported from the GeneMapper software (Applied Biosystems, version 3) and input into the Coffalyser version 8 (obtained off the MRC-Holland website) for further analysis. The Methyaltion Status analysis tab was selected for final results.

3.3. Results

3.3.1. Cohort selection

Seven samples were finally chosen for the tumour analysis, which is represented in the table below (Table 11). For both, colorectal and endometrial carcinomas, a minimum of three samples each were selected. A single patient presented with a renal carcinoma. All samples had the same predisposing mutation in the *hMLH1* gene and were pathologically graded as Dukes B.

Table 12: Cohort utilized in the study

DNA NUMBER	SITE
NPC 77.11ELI	Endometrial
NPC 1.232ELK	Endometrial
NPC 82.2LOU	Endometrial
NPC 1.93JOH	Colorectal (Proximal colon)
NPC 1.52MOR	Colorectal (Hepatic flexure)
NPC 101.1EDI	Colorectal (Hepatic flexure)
NPC 1.248FAN	Renal

3.3.2. DNA Isolation

All DNA obtained for the study were analysed prior to use. For DNA isolated from both blood and tumour tissue, the quality of the DNA was examined by assessment of the migration of the DNA in an agarose gel. The quantity of the DNA was assessed by spectrophotometry.

The quantity and quality of the DNA isolated from the tumour tissue is represented in the table below (Table 12). The integrity gel of the DNA isolated from tumour tissue can be seen in Figure 10.

Table 13: Quality of the DNA isolated from the tumour tissue

SAMPLE	CONCENTRATION (ng/µl)	$^{260}/_{280}$	$^{260}/_{230}$
NPC 1.232	12.5	1.57	-0.55
NPC 1.248	188.5	1.86	2.96
NPC 1.52	37.4	1.66	2.83
NPC 1.93	162.1	1.81	3.32
NPC 77.11	29.6	1.76	2.14
NPC 82.2	69.7	1.76	0.94
NPC 101.1	38.7	1.78	9.77

Ideally, the concentration of DNA isolated from tumour tissue would be at least 100ng/ µl. However, due to the preservation of the tumour tissue and loss of very small sheared fragments of DNA during isolation, the concentration of these samples are low. The isolated samples are quite pure, evident by the $^{260/}_{280}$ ratio

which for all samples, barring NPC 1.232, are around 1.8. This indicates that there is hardly any protein contamination. Unfortunately, the $^{260}/_{230}$ ratio which analysis organic solvent contamination, is variable ranging from -0.55 (NPC 1.232) to 9.77 (NPC 101.1) (Table 12). This indicates that there may be some contamination of reagents used in the isolation procedure.

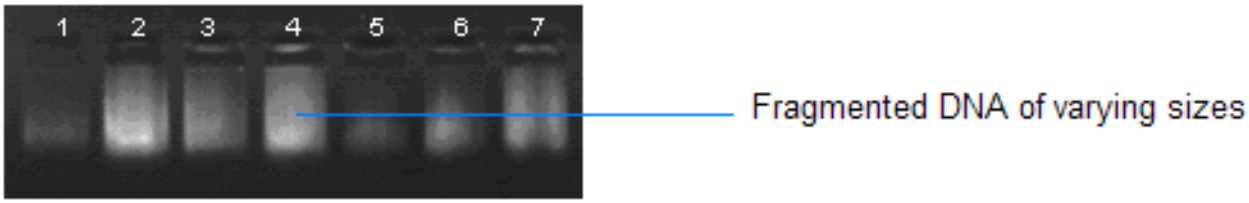


Figure 10: **AGE of DNA isolated from tumour tissue. DNA was electrophoresed on a 2% (w/v) agarose gel and detected with Sybr Gold. Electrophoresis occurred at 120V for 30 minutes. Lane 1: NPC 1.232, Lane 2: NPC 1.248, Lane 3: NPC 1.52, Lane 4: NPC 1.93, Lane 5: NPC 77.11, Lane 6: NPC 82.2, Lane 7: NPC 101.1**

Figure 10 shows the integrity gel of DNA isolated from tumour tissue. All lanes show fragmented DNA. This is expected due to the preservation, storage and isolation of DNA from the slide. Lanes 1, 5 and 6 show very faint fragments and these correspond to very low DNA concentrations (12ng/µl, 29.6ng/µl and 69.7ng/µl respectively) (Table 12). However, Lanes 3 and 7 corresponds to samples NPC 1.52 and NPC 101.1 which have lower concentrations that NPC 82.2 (Lane 6) yet have brighter smears of DNA. This may possibly due to pipetting error whereby not enough sample was placed into the wells of the gel or the sample was not properly mixed with the loading dye causing a small amount of the sample to escape the well prior to the commencement of electrophoresis.

3.3.3. Selection of microsatellite rich genes

The genes selected for the study can be seen in Table 13. Information about the gene, gene function, the microsatellite under investigation and previous studies on the microsatellite are also provided. As previously mentioned, the genes were primarily chosen through literature and then subjected to bioinformatic analysis to narrow down the list of genes. In total, 14 genes were chosen for analysis. All but two of the genes had cMS. One of the genes (Estrogen receptor Alpha) had a promoter microsatellite while the other had a microsatellite found in an intron (Estrogen receptor Beta).

Table 14: List of genes selected for the study.

GENE NAME	GENE	REPEAT	POSITION	OBSERVED MUTATIONS	GENE FUNCTION	MUTANT GENE FUNCTION	REF
Activin receptor type 2 A	*ACVR2*	(A)8	*Exon 10*	Expansion/ Contraction of repeat	Transcriptional regulator, TGFB signaling pathway	Unknown	1
Androgen receptor	*AR*	(CAG)n	*Exon 1*	Expansion/ Contraction of repeat	Transcriptional Regulator	Increased/ Decreased transcription of target genes	2
Bcl-associated protein X	*BAX*	(G)8	*Exon 3*		Proapototic gene	Unknown	3
Caspase-5	*CASP5*	(A)10	*Exon 2*	Deletions 1bp (common)/ 2bp, insertions 1bp	Cleave pro-caspase-3 thus activating caspase-3 (involved in apoptotic cell death)	Growth advantage of cells through inhibition of apoptosis	3
Estrogen receptor Alpha	*ERα*	(TA)	*Promoter*	Expansion/ Contraction of repeat	Transcriptional regulator	Unknown	4
Estrogen Receptor Beta	*Erβ*	(CA)	*Intron 5*	Expansion/ Contraction of repeat	Transcriptional regulator	Unknown	4
Insulin growth factor receptor 2	*IGF2R*	(G)8	*Exon 28*	Expansion/ Contraction of repeat	Activate TGFB	Possible premature stop codon resulting in the prevention of cytotoxic T cell mediated apoptosis	3
Methyl CpG binding domain protein	*MBD4*	(A)10	*Exon 3*	Expansion/ Contraction of repeat	DNA repair	Genomic instability increased?	3
	hMSH3	(A)8	*Exon 8*	Deletion 1bp/ 2bp, insertion 1bp	DNA repair	More severe replication error	3
	hMSH6	(C)8	*Exon 5*	Frameshift as above	DNA repair	More severe replication error	3
Phospatase and tensin homolog	*PTEN*	(A)6	*Exon 7*	Expansion/ Contraction of repeat	Signaling	Unknown	6

Table 15 contd: **List of genes selected for the study.**

GENE NAME	GENE	REPEAT	POSITION	OBSERVED MUTATIONS	GENE FUNCTION	MUTANT GENE FUNCTION	REF
TATA box binding protein associated factor 1B	*TAF1B*	(A) 10	*Exon 8*	Expansion/ Contraction of repeat	Transcription regulation	Unknown	1
Transforming growth factor receptor type 2	*TGFBR2*	(A)10	*Exon 3*	Deletions 1bp/ 2bp, insertions 1bp	Mediates growth inhibition of epithelial cells by TGFB	Growth advantage to cells	3
Tumour suppressor P53	*TP53*	(A)11	*Exon 11*	Unknown	Apoptosis	Unknown	7

1. Jung et al, (2006)
2. Gottlieb B *et al.* (2004)
3. Miyaki *et al.* (2001).
4. Slattery *et al.* (2005)
5. Woerner *et al.* (2005)
6. Zhou *et al.* (2002)
7. Shitoh *et al.* (1998)

3.3.4. Genotyping

To assess the alleles in the 14 microsatellite rich genes of interest, multiplex PCR was performed on the samples. Thereafter, the samples were electrophoresed on a three percent agarose gel in order to ensure the reaction was successful.

Samples that were correctly amplified were diluted $^{1}/_{10}$ if the PCR products were too bright. Thereafter, the samples were pooled before 1µl of the sample was taken through genotyping.

Results obtained from the genotyping analysis are provided in Table 14. The various alleles observed for each gene studied were allocated a numerical number. The allele sizes and details are provided in Appendix 6. Figure 11 represents sample NPC 1.248 genotyped for the *MBD4* gene. As a single peak is observed, NPC 1.248 is homozygous. The allele size, 224bp, represents Allele 2.

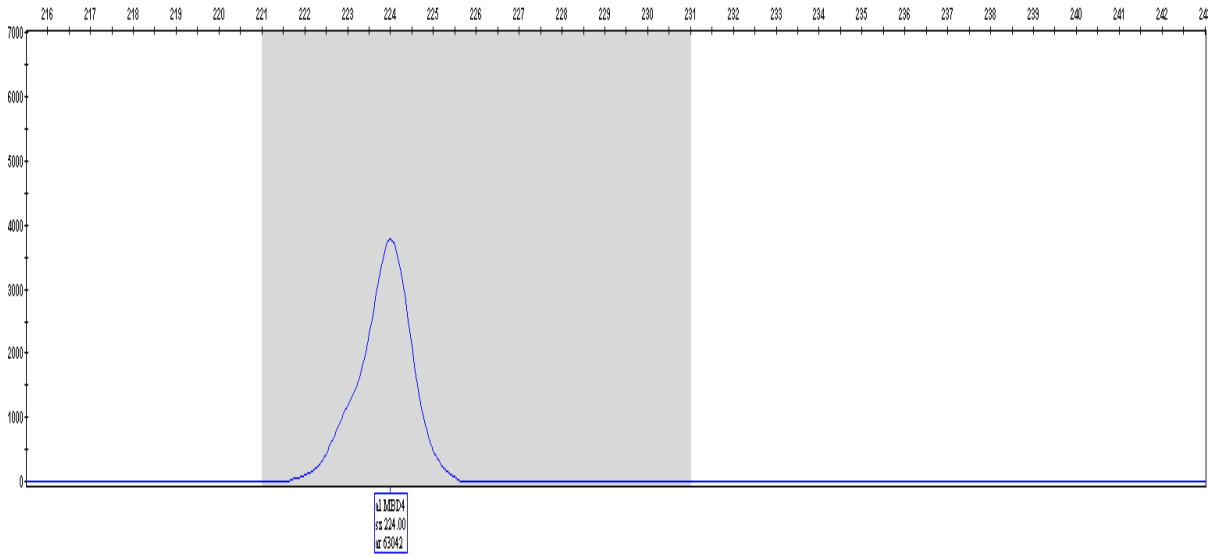


Figure 11: Example of the GeneMapper peak for the MBD4 gene for NPC 1.248.

Table 16: **Results of the genotyping of 14 microsatellite rich genes and their alleles. G represents germline DNA. T represents the tumour DNA of the sample. U represents instability in the gene. Numbers 1 – 5 represent the allele sizes. These can be viewed in Appendix 6.**

SITE	Endometrial												Colorectal												Renal			
SAMPLE	NPC 1.232				NPC 77.11				NPC 82.2				NPC 1.52				NPC 1.93				NPC 101.1				NPC 1.248			
	G		T		G		T		G		T		G		T		G		T		G		T		G		T	
ACVR2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1
Erβ	4	5	**U**		3	1	3	1	4	4	**U**		2	2	**U**		3	3	**U**		3	3	**U**		4	1	**U**	
MSH6	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
AR	1	1	1	1	5	5	5	5	4	4	4	4	2	2	2	2	2	2	2	2	3	3	3	3	2	2	2	2
CASP5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IGF2R	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TP53	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
BAX	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3
TGFBR2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MBD4	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2
MSH3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PTEN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ERa	2	2	2	2	1	1	1	1	4	4	4	4	2	2	2	2	1	1	1	1	1	1	1	1	3	3	3	3
TAF1B	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Six of the genes (*CASP5, IGF2R, TGFBR2, MSH3, PTEN* and *TAF1B*) under investigation, displayed the same genotype in all seven samples. *ACVR2*, highlighted in red, demonstrated a different genotype in the individual who presented with renal cancer when compared to the other six samples.

The majority of the genes in the tumour samples remained static. The *Erβ* gene did however show instability in six of the seven samples. An example of the instability can be seen in Figure 12.

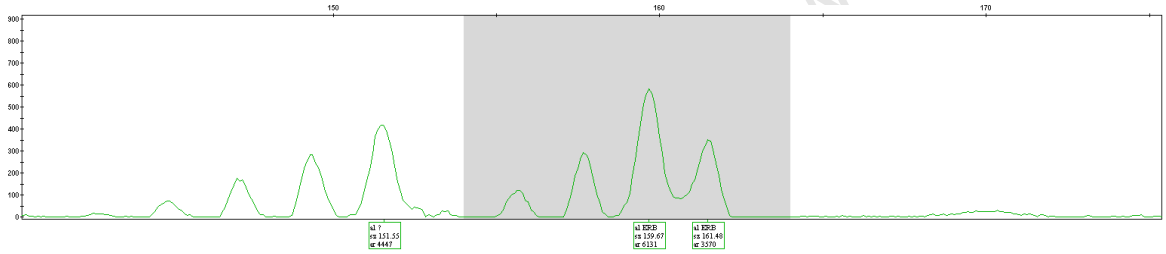


Figure 12: Example of the instability seen in the *Erβ* gene for sample NPC 1.232.

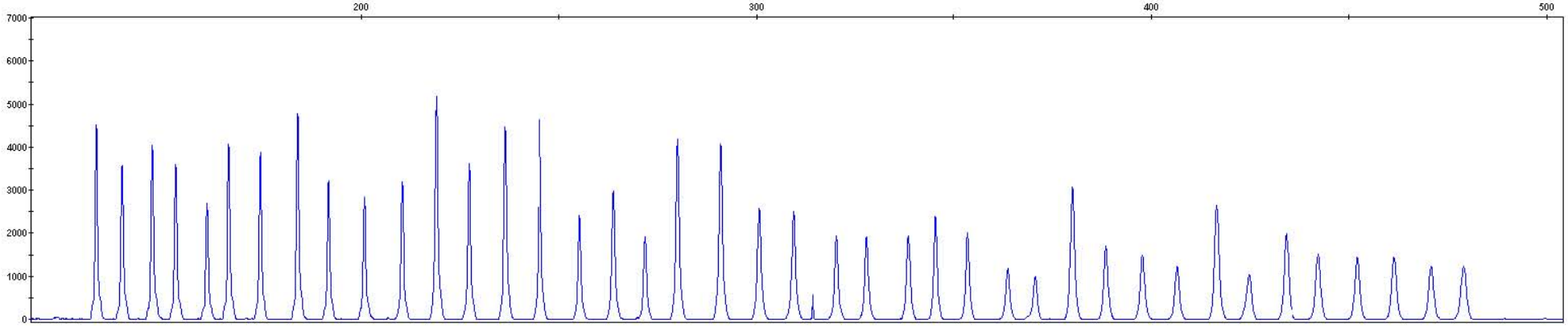
3.3.5. Methylation Specific Multiplex Ligation – Dependent Probe Amplification (MS - MLPA)

The results of the methylation status of each sample and the corresponding tumour tissue are tabulated (Table 15). Figure 13 shows the GeneMapper output of the germline sample of NPC 82.2.

Methylation status of the tumour was established via the ratio of the cut sample to the reference (uncut) sample of the Ms-MLPA. The calculations of the ratios were automatically performed by the Coffalyser software. The classification of methylation was categorised into unmethylated (ratio of less than 0.25), hemi-methylated (ratio of between 0.25 and 0.75) and methylated (ratio greater than 0.75). Table 15 contains the ratios of the germline and tumour DNA analysis for each sample. The table is also colour-coded: red represents no methylation, green represents hemi-methylation and blue represents homozygous methylation.

The Ms-MLPA kit contains 14 control probes. These probes should be amplified in all samples. There are 26 tumour suppressor gene probes in the kit, corresponding to 24 tumour suppressor genes (two genes, *MLH1* and *RASSF1* contain two probes). In normal, healthy samples these genes should not be methylated (either heterozygous methylation or homozygous methylation) and have a ratio of less than 0.25. This can be seen from the results obtained from the germline samples of NPC 1.232, NPC 1.248 NPC 1.52, NPC 82.2 and NPC 101.1 (Table 15). When an allele of the tumour suppressor gene is methylated, that allele cannot be digested by the restriction enzyme *HhaI*. As a result, that allele is amplified in the PCR and when the result of the cut sample is compared to the reference sample, that gene would have a ratio greater than 0.25.

A.



B.

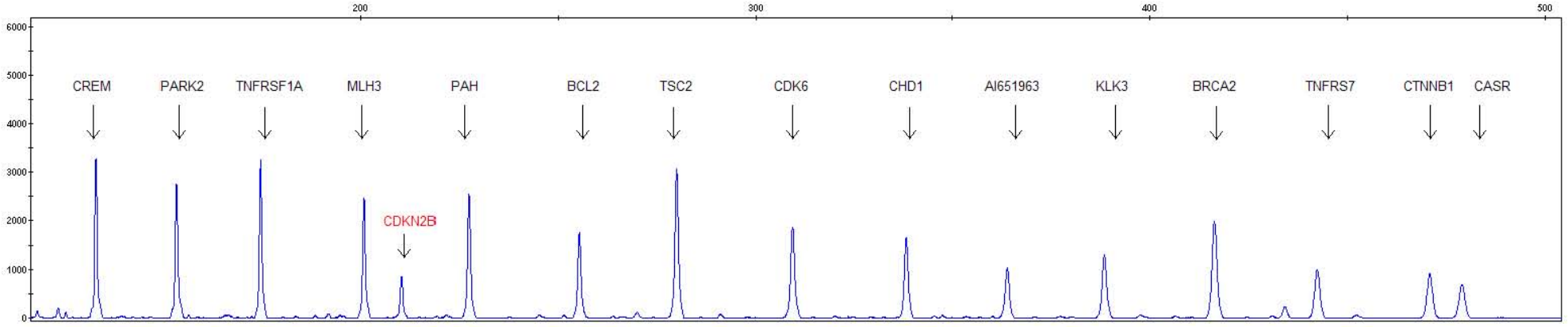


Figure 13: **Ms-MLPA of germline DNA sample NPC 82.2. Panel A represents the uncut, control sample showing 41 peaks corresponding to the 41 probes in the kit. Panel B represents the cut sample showing the 15 control probes, labeled in black. The tumour suppressor gene CDKN2B is visible on the electropherogram, labeled in red however after normalization and ratio analysis, the peak was fortuitous.**

Table 17: **Ratio Chart of the Ms-MLPA results for the germline and corresponding tumour samples. Red blocks represent no product. Green blocks represent amplification of a single allele and blue blocks represent amplification of both alleles.**

SAMPLE	NPC 1.232		NPC 1.248		NPC 1.52		NPC 1.93		NPC 77.11		NPC 82.2		NPC 101.1	
MALIGNANCY	Endometrial		Renal		Colorectal		Colorectal		Endometrial		Endometrial		Colorectal	
	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR
TP73	0	0	0.02	0.62	0	0.02	0	0.39	0.02	0.02	0	0.03	0.04	0.02
CASP8	0	0	0	0.89	0	0.03	0.01	0.55	0.02	0.06	0	0.37	0.03	0.03
VHL	0	0	0	32.43	0.09	0.07	0	0.46	0.01	0.07	0.03	0.06	0.01	0.05
RARB	0	0	0	0.73	0.02	0.02	0	0.4	0.02	0.02	0.02	0.04	0	0.01
MLH1 166bp	0	0	0	0.67	0	0.03	1.01	0.53	0.03	0.02	0	0.04	0	0.91
MLH1 463bp	0.01	0	0.02	0.14	0	0.12	0.05	0.32	0	0.02	0	0.34	0	0.27
RASSF1 328bp	0.01	0	0.01	0.34	0	0.03	1	0.31	0.11	0.04	0.02	0.03	0.01	0.84
RASSF1 382bp	0.01	0.09	0.03	0.48	0.02	0.12	0.09	0.22	0	0.06	0.02	0.06	0.03	0.05
FHIT	0.01	0.09	0	0.36	0	0	0.03	0.24	0	0.01	0	0	0	0.01
APC	0.01	0	0.01	0.35	0	0	0.12	0.25	0	0.04	0.02	0.04	0.02	0.02
ESR1	0.01	0.23	0.01	0.62	0	0.04	0.06	0.72	0.01	0.07	0.02	0.07	0	0.01
CDKN2A	0.01	0.15	0.01	0.51	0	0	0.04	0.39	0.43	0.07	0	0.04	0.02	0.06
CDKN2B	0.01	0.1	0	0.58	0.09	0.06	0.01	0.42	0.03	0.1	0.03	0.09	0.05	0.06
DAPK1	0.02	0.12	0.04	0.28	0.09	0	0.05	0.35	0	0.1	0	0.09	0.04	0.11
PTEN	0.02	0	0.04	0.54	0.01	0.07	0	0.17	0	0.05	0.01	0.06	0.01	0.06
CD44	0.02	0.29	0.06	0.57	0	0.09	0	0.24	0.02	0.1	0	0.33	0.04	0.06
GSTP1	0.02	0.13	0	0.53	0.15	0.1	0.09	0.51	0.87	0.2	0.06	0.21	0.15	0.09
ATM	0.02	0.13	0	0.74	0.14	0.1	0	0.77	0.05	0.13	0	0.2	0.05	0.03
IGSF4	0.03	0	0	16.34	0.1	0.08	0.01	1.16	2.01	0.11	0	0.1	0.07	0.45
CDKN1B	0.03	0	0.03	26.61	0	0.02	0.01	0.91	2.83	0.05	0.04	0.04	0	0.04
CHFR	0.04	0	0.04	0.64	0.03	0.05	0	0.65	0.01	0.04	0.04	0.1	0.01	0.35
BRCA2	0.04	0.19	0.06	0.82	0.03	0.09	0.04	0.23	0	0.1	0.03	0.14	0.07	0.08
CDH13	0.05	0.07	0.06	0.54	0.14	0	0.12	0.39	0	0.04	0.02	0.05	0.01	0.07
HIC1	0.06	0.3	0.09	0.56	0.31	0.27	0.32	0.77	3.38	0.27	0.15	0.33	0.17	0.23
BRCA1	0.08	0	0.11	0.86	0.09	0.15	0.05	0.39	0.09	0.27	0.05	0.38	0.11	0.11
TIMP3	0.12	0.29	0.19	1.01	0.95	0.48	0.69	0.4	0.2	0.58	0.36	0.57	0.19	0.85

CONTROL PROBES														
SAMPLE	NPC 1.232		NPC 1.248		NPC 1.52		NPC 1.93		NPC 77.11		NPC 82.2		NPC 101.1	
	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR	GERMLINE	TUMOUR
CTNNB1	0.73	1.18	0.64	0.9	1.44	1.13	0.68	0.65	0.6	0.69	1.01	0.63	1	0.84
CASR	0.84	0.04	0.94	0.54	1.52	1.22	1.13	0.35	1	0.45	1.16	0.42	0.77	1
PARK2	0.86	0.82	0.79	1.01	1.84	1	0.97	1.08	78.36	0.92	1.06	0.47	1.12	1.16
CDK6	0.95	0.67	0.98	0.91	0.98	0.79	1.05	1.17	0.93	0.74	0.96	0.94	1	1.41
CREM	0.96	1.63	0.89	1.03	1.25	1.15	1.11	0.51	1.65	0.79	1	0.78	1.08	0.85
AI651963	0.98	1.19	1.14	0.7	1.44	0.61	2.52	1.93	117.67	1.12	1.02	1.09	0.87	0.84
TNFRSF1A	1	0.92	0.85	0.9	0.8	1.01	1.48	1.55	0.81	0.93	1	0.71	0.85	1.28
TNFRSF7	1	1.07	1	1	0.98	0.9	0.28	2.58	14.77	1.77	0.71	1.56	0.75	0.94
PAH	1.06	0.58	1.21	1.17	1	0.57	0.45	4.65	1.88	1.33	1.19	1.44	1.09	1.14
BRCA2	1.08	1	1.16	1.09	0.84	0.02	1.31	0.85	0.01	1	1.16	1.17	0.75	1.28
MLH3	1.1	1	1.07	0.92	1	1.01	0.62	1.17	0.24	1.65	0.95	1.51	0.92	0.75
TSC2	1.14	0.72	1.28	1.11	0.97	0.95	0.56	0.81	1.03	0.73	1	1.05	1.07	1.37
CDH1	1.16	1.84	1.31	1	0.93	1.4	1	1	6.92	1.39	0.98	0.91	1.34	0.96
BCL2	1.17	1.27	0.93	1.14	0.73	1.11	0.91	0.78	0.04	1.47	0.85	1.61	0.82	0.98
KLK3	1.34	1.05	1.18	1.86	2.57	0.02	2.59	0.9	0.85	1.05	1.19	1	1.49	1.33

3.4. Discussion

The aim of the study was to determine a genetic signature of Lynch-associated malignancies by distinguishing molecular similarities between pathologically similar colorectal, endometrial and renal cancers. A panel of microsatellite rich genes and Ms-MLPA were the techniques used.

3.4.1. Microsatellite rich genes

Three genes involved in the apoptosis pathway were analysed in the study. These genes, BCl-associated protein X (*BAX*), Caspase 5 (*CASP5*) and Tumour protein 53 (*TP53*) have coding microsatellites, which may deem these genes susceptible to microsatellite instability in the presence of defective mismatch repair. However, these genes remained static between germline and tumour samples.

Activin receptor type 2A (*ACVR2*) is a transcriptional regulation gene and also involved in the Transforming Growth Factor signalling pathway. This gene contains two microsatellite repeats. The specific repeat examined in this project is found in exon 10 and is reported to be frequently mutated in colon cancer (Jung *et al.* 2006). Deacu *et al.* (2004) demonstrated that restoration the *ACVR2* gene within deficient cell lines allowed the *TGFBR2* pathway to be restored. To date, there are no reports on the association of the *ACVR2* gene and renal cancer. The results of the *ACVR2* gene showed that the individual who presented with renal cancer had an alternate genotype of the gene when compared to the other sex members of the cohort. The individual was homozygous for the genotype, of which the product was significantly shorter (100 base-pairs) than the product size (112 base-pairs).

The transforming growth factor receptor type 2 (*TGFBR2*) gene is involved in the regulation of growth and proliferation of cells and contains a tract of ten adenine bases in exon three of the gene. Expansion and contraction of this tract is noted in most microsatellite unstable colorectal cancers (Woerner *et al.* 2005, Jung *et al.* 2006) and is thought to confer a growth advantage to the cancerous cells. There are numerous reports which implicate microsatellite instability of the *TGFBR2* gene in Lynch syndrome cancers. Unfortunately, in this study, the *TGFBR2* gene was found to be stable within the tumour tissue of the genes assessed.

Two DNA MMR genes were assessed and no somatic mutations were observed in the tumour tissue. The genes *hMSH3* and *hMSH6* have coding microsatellites in exon eight and five respectively. Both genes have a tract of eight repeats; *hMSH3* has a tract of adenine whereas *hMSH6* has a tract of cytosine residues. Expansion or contraction of these repeats may lead to impaired function of the gene which in turn will lead to impaired DNA repair (Miyaki *et al.* 2001). Mutations within the *hMSH6* gene, particularly point mutations, are implicated in Lynch syndrome; however, mutations within the gene may lead to a variation in the age of onset of the disorder. In addition, *hMSH6* mutations have been associated to have an increase in the risk of the gynaecologic cancers (Cederquist *et al.* 2005).

Three steroid hormone genes, the androgen receptor (AR) and the estrogen receptors alpha and beta (Era and *Erβ)*, were assessed. The results of the study hoped to provide further insight into the tissue specific nature of the extracolonic cancer spectrum. Recent publications (Frogatt *et al.* 1999, Campbell-Thompson *et al.* 2001, Slattery *et al.* 2006) have found associations between these three genes, specifically the length of microsatellite regions within the genes and the occurrence of colorectal cancer. In addition, the androgen receptor has also been associated to Lynch –associated ovarian cancer (Santarosa *et al.* 2002).

In 2001, Campbell-Thompson *et al.* investigated the expression of the estrogen receptor subtypes and isoforms in colon tumours, normal colonic mucosa and colon cancer cell lines. This study was performed in order to elucidate the protective effect of female hormones. In Lynch syndrome, the lifetime risk of developing colon cancer is much lower in females than in their male counterparts (Froggatt *et al.* 1999). The Campell-Thompson *et al.* (2001) study and a more recent study (Slattery *et al.* 2005) have shown that the expression of *Erβ* is prominent in the colon and decreased levels of the gene in females are associated with colonic malignancies. Six of the seven tumour tissue samples analysed in the current study showed instability at this receptor. *Erβ* is involved in the regulation of gene expression (Deroo and Korach. 2006). It is hypothesised that microsatellite instability observed in the *Erβ* gene may lead to abnormal transcriptional activation and erroneously result in the silencing of tumour suppressor genes or conversely an increase the transcription of proliferative genes (Slattery *et al.* 2005). MSI evident at this gene in the tumour samples examined may be informative as to a potential marker for Lynch-associated malignancies. However, this result needs to be validated and an investigation of *Erβ* in a control tumour set (of sporadic malignancies) is required.

The study involving the microsatellite rich genes unfortunately did not prove fruitful apart from one gene. The lack of microsatellite instability seen in the panel of genes assessed may be due to the stage of the cancer. All the tumours assessed were Dukes Stage 2 cancers and perhaps the rate of instability increases with the increased number of mutations randomly introduced into the cancerous tissue. Also, tumour tissue is highly heterogeneous and several sections from varying tumour position would have been ideally examined. However, previous studies (Jung *et al.* 2006, Vilkki *et al.* 2002) have successfully uncovered mutations within microsatellite rich genes while utilising Dukes stage 2 cancers. A limitation to the study was that the tumours were analysed by different pathologists and during DNA isolation from tumour tissue, the tumour sample was not micro-dissected. These factors may have introduced a bias into the study.

3.4.2. Ms-MLPA

The MS-MLPA was performed in order to potentially identify a defining signature of DNA MMR negative malignancies. The Ms-MLPA kit analyses the gene copy number and the methylation status of the promoter regions of tumour suppressor genes.

Two of the tumour suppressor genes, analysed by the Ms-MLPA kit for promoter methylation, were found to be methylated (either hemi-methylation of one allele or complete methylation) in all the tumour samples. These genes are *HIC1* and *TIMP3*.

The *HIC1* gene is also known as the Hyper-methylated in Cancer Protein and encodes a transcription factor. Functional interaction of *HIC1* with the tumour suppressor gene *TP53* has been described (Fleuriel *et al.* 2008). Feng *et al.* (2008) has shown that methylation of four tumour genes; namely *HIC1*, *RASSF1*, *CABIN1* and *MINT31*; is associated with ovarian carcinoma. Ovarian carcinoma is a member of the tumour spectrum of Lynch syndrome unfortunately, there was no ovarian tissue malignancy available to analyse in the current study. It was observed by Feng *et al.* (2008) that methylation of the *HIC1* and *RASSF1* genes occurred early in the evolution of the malignancy. The tumour tissue analysed in this study corresponded to the Dukes B pathological grading and therefore correlate with the Feng *et al.* (2008) study in terms of the early methylation of *HIC1*.

Joensuu *et al.* (2008) investigated the methylation of 24 tumour suppressor genes in a cohort of Lynch syndrome patients who presented with either an endometrial or colorectal carcinoma. Four tumour suppressor genes, one of which was *TIMP3*, was found to be methylated in both types of carcinomas. The other genes included

CHD13, APC and *GSTP1*. Mutations within the tissue inhibitor of metalloproteases 3 (*TIMP3*) are known to result in Sorsby fundus dystrophy. Methylation of the gene has been implicated in breast, uterine, oesphogeal, pancreatic, renal and colorectal carcinomas (Lui *et al.* 2005, Smith *et al.* 2008). It is evident methylation of *TIMP3* is critical in the tumourigenic process however, in studies performed on the methylation of *TIMP3* in the various malignancies described above; the MMR status of each carcinoma was not mentioned. Therefore it is not clear whether *TIMP3* methylation is common to all malignancies (familial and sporadic) and all tumourigenic pathways (Vogelstein and mutator).

NPC 1.248 is the only sample which presented with a renal carcinoma. The tumour DNA shows vast methylation of all bar two tumour suppressor gene promoters. A control gene, *CTNNB1*, showed hemi-methylation in the germline DNA. The excessive methylation of the tumour DNA may be due vast heterogeneity of the tumour tissue and may be a false positive result as the tumour section was not micro-dissected prior to isolation. The heterogeneity of the tumour section could also have resulted in certain areas of the tumour being a pathologically higher grading than Dukes B.

In a number of samples, both tumour and germline, there is hemi-methylation or loss of amplification of the control samples. In the tumour DNA, it is possible that due to the degradation of DNA during preservation and isolation, these sequences of these genes may have been sheared and therefore amplification may have been restricted. The germline DNA sample for NPC 77.11 shows no amplification of three control probes; *BRCA2, MLH3* and *BCL2*. As the control probes are expected to successfully amplify in all samples, especially in germline DNA, the lack of amplification of these control probes is unknown. In addition, three germline samples (NPC 1.248, NPC 1.93, NPC 77.11) have hemi-methylation of *CTNNB1*. Perhaps the amplification of these control probes were not as successful as the reference samples and as a result, upon calculation of the ratios, these control

probe ratios seemed to fall within the lower ratio categories causing the automated program to regard them as either hemi-methylated or not amplified.

Methylation of the tumour samples of the tumour suppressor genes *HIC1* and *TIMP3* gene may provide impetus to further investigate a genetic signature in Lynch syndrome tumours.

3.4.3. Future work

Verification of the promoter methylation of *HIC1* and *TIMP3* in the samples by bisulphite sequence is required. Malignancies at different pathological stages taken through both the microsatellite panel of genes and the Ms-MLPA may be more informative about the tumourigenic process. A setback in the study was the ability in obtaining archived tumour material from mutation positive individuals. Perhaps, had there been a larger tumour tissue cohort and tumour tissue at various stages of tumour development, the results may have been more informative. In future, the assessment of these genes may be beneficial when tracking the path of mutations within the mutator pathway.

Chapter 4: Bioinformatic analysis

4.1. Introduction

This project aims to elucidate the occurrence of extracolonic malignancies in Lynch syndrome by focusing on the characterization of tumours from the susceptible anatomic sites.

In order to supplement the study, bioinformatic analyses of existing, available microarray data was assessed. The aim was to provide further insight into potential mechanisms for the initiation of malignancy in the susceptible anatomic sites. To date, the identification of candidate genes allowing the proliferation of cancerous tissue in only certain anatomic sites have been elusive.

4.2. Materials and Methods

4.2.1. Overview

Figure 14 represents a diagrammatic overview of the bioinformatics approach. First, the ONCOMINE database was accessed and microarray data on microsatellite unstable colorectal and endometrial tissue was obtained. The common under and over expressed genes in these cancer profiles were acquired. Thereafter, the common candidate genes were analysed in isolation and through their relationships in terms of biological pathways and interactions.

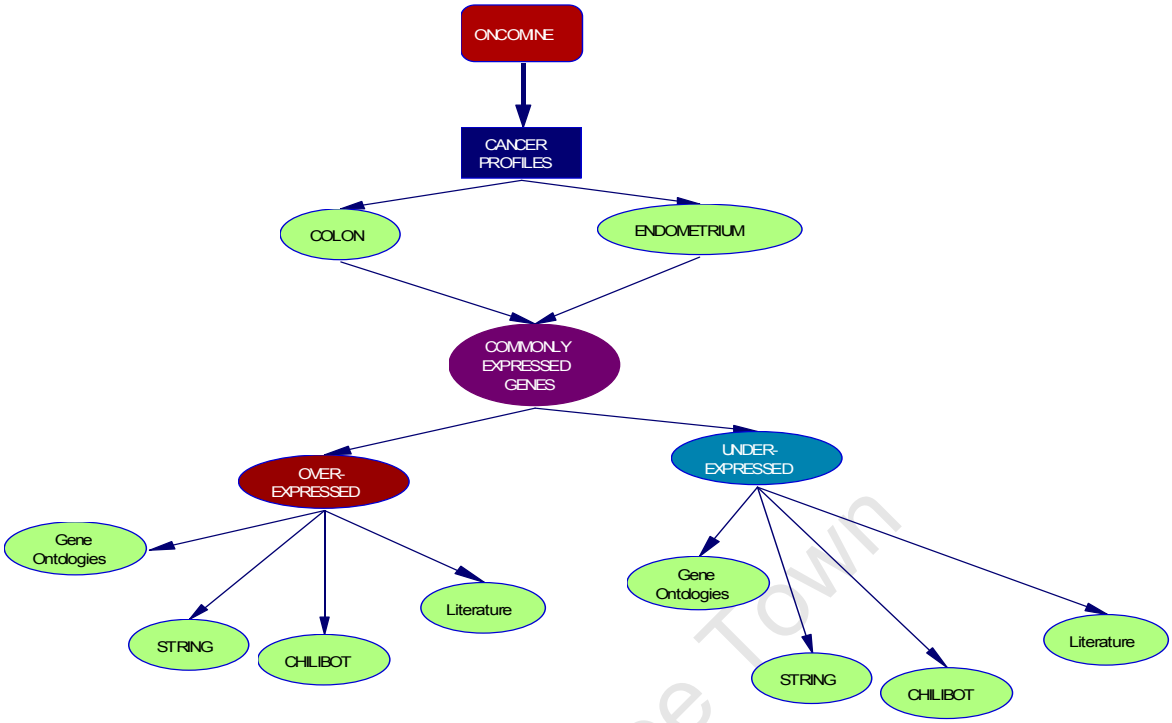


Figure 14: Overview of the bioinformatics workflow

4.2.2. ONCOMINE

Microarray data was retrieved from the ONCOMINE database (http://www.oncomine.org) (Rhodes *et al.* 2007). The paper-centric database provides data from previously performed and published cancer studies. Initially data representing microsatellite status was identified. The studies of interest compared the level of gene expression between microsatellite stable tumours and those exhibiting MSI. Due to the stringent criteria only three investigations were available for study. The studies chosen were; Watanabe *et al.* (2006), Koinuma *et al.* (2004) and Mutter *et al.* (2001). Watanabe *et al.* (2006) and Koinuma *et al.* (2004) investigated colorectal tissue samples whereas the Mutter *et al.* (2001) study analysed endometrial samples. Under-expressed and over-expressed genes was obtained for each study and exported to a text file.

4.2.3. Conversion of gene names to Ensembl identifiers

The gene lists exported from the ONCOMINE database were Human Genome Nomenclature Committee (HGNC) gene symbols. Prior to analysis, the gene lists were converted into Ensembl identifiers by mining the Ensembl site with Biomart. The database used for the mining was the Ensembl 50 database with the Homosapiens NCBI 36 dataset. Under the filters the gene identifier list was selected and the HGNC symbol was chosen. The gene lists obtained from ONCOMINE were added to the available section. Thereafter, the results tab was selected and the lists of Ensembl identifiers were chosen to be exported via a text document.

Upon manual curation of the resulting lists, it was found that there were duplicates. The duplicates were therefore removed utilising the Excel Unique and Duplicate Remover.

4.2.4. Collation of gene lists and scoring matrix

The gene lists were analysed in Microsoft Excel. A scoring matrix was obtained from Dr Tiffin (personal communication). The scoring matrix was written using Visual Basic, by A.J. Swift (2007).

4.2.4.1. ONCOMINE

The gene lists from the microarray studies were placed in an Excel document containing the scoring matrix. Under-expressed and over-expressed gene lists were examined in isolation. The exported gene list for each study was placed in

individual columns. The scoring matrix the analysed the list of genes input and scored each gene on the number of times the gene appeared in each column list. Each gene, for each of the two anatomic sites in question, was weighted equally. The results of the scoring matrix were manually curated and genes with a final score of three were then selected as commonly expressed in all the three anatomic sites.

4.2.5. Automated analysis of common under-expressed and over-expressed genes

Genes found to be commonly under-expressed or over-expressed in the microsatellite unstable colorectal and endometrial tissue were investigated. Initally, information regarding the genes was obtained by accessing the Ensembl website. Functional annotation of the genes was obtained using the Gene ontologies (http://www.geneontologies.org).

Known relationships between the genes were examined by utilising the CHILIBOT website (http://www.chilibot.net) (Chen and Sharp 2004). This data-mining program identifies the relationships between the genes by extracting information via literature. Protein interactions were then interrogated via the web-based Search Tool for Retrieval of Interacting Proteins (STRING) (http://string.embl.de) (von Mering *et al.* 2007). Where possible, biological pathways where examined by interrogating the Reactome (http://www.reactome.org) (Vastrik *et al.* 2007) and Kegg databases.

4.3. Results

4.3.1. Up and Down regulated genes in ONCOMINE

Three studies from the ONCOMINE database were selected for further study in this investigation. Microarray data showing both down-regulated and up-regulated genes in tumour tissue comparing microsatellite stable and microsatellite unstable status was obtained. Only those genes with a cut-off p – value of 0.05 were selected for further analysis. Table 16 shows the results of the selection.

Table 18: Summary of the three studies utilized in this investigation

STUDY	TISSUE TYPE	# MSS SAMPLES	# MSI SAMPLES	# UP REGULATED GENES	# DOWN REGULATED GENES
Watanabe	Colorectal	51	33	13495	8862
Koinuma	Colorectal	10	10	1639	1437
Mutter	Endometrium	5	5	195	130

4.3.2. Comparison of the ONCOMINE commonly expressed genes.

In total four genes were commonly under-expressed in the colorectal and endometrial studies. Three genes were commonly over-expressed. The results can be seen in Table 17.

Table 19: Common genes

ENSEMBL ID	GENE SYMBOL	GENE NAME	GO TERMS
UNDEREXPRESSED			
ENSG00000187048	*CYP4AII*	Cytochrome P450 4AII (Fatty acid omega hydroxylase)	Iron ion binding (Heme binding) Mono-oxygenase activity Fatty acid metabolism Electron carrier activity
ENSG00000161021	*MAML1*	Mastermind like protein 1	Transcription co-activator activity Notch signalling pathway
ENSG00000060656	*PTPRU*	Receptor type tyrosine protein phosphatase U precursor (pancreatic carcinoma phosphatase 2)	Protein tyrosine phosphatase activity
ENSG00000112624	*KIAA0240*	Uncharacterised protein	
OVEREXPRESSED			
ENSG00000112983	*BRD8*	Bromodomain containing protein 8	Hydroxylase activity Protein dimerisation activity Regulation of cell growth Transcription factor activity
ENSG00000204252	*HLA-DOA*	HLA class II histocompatibility antigen Do Alpha chain precursor	Antigen processing and presentation Immune response
ENSG00000204258	*HLA-DMB*	Major histocompatibility complex class II DMBeta	Immune response Antigen processing

CHILIBOT

There was no relevant literature on the relationships between the four common under-expressed genes. In terms of the over-expressed genes, there was only a relationship between HLA-DOA and HLA-DMB.

STRING

Similarly there were no interactions between the gene products of the under-expressed genes. There was only a protein interaction between HLA-DOA and HLA-DMB in the over-expressed gene cohort, shown in Figure 15. The lines linking HLA-DOA and HLA-DMB indicate there is a putative interaction of these proteins though an experimental assay has not been performed. There is no evidence of an interaction of the BRD8 protein with HLA-DOA or HLA-DMB.

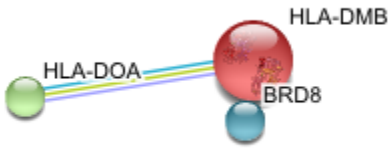


Figure 15: **STRING output of the relationship of the over-expressed genes common to the colorectal and endometrial MSI tumour tissue.**

SEQUENCE ANALYSIS

Interrogation of the seven gene sequences (NCBI, Ensembl) showed no microsatellite rich regions within the genes.

4.4. Discussion

Analysis of existing microarray data from cancer profiles were assessed in order to potentially expose novel candidate genes which may explain the occurrence of extracolonic malignancies in Lynch syndrome. Gene expressions in Lynch syndrome susceptible tissue sites were analysed.

Microarray data was obtained for microsatellite unstable and microsatellite stable tumour tissue from both colorectum and endometrium. Comparison of under-expressed and over-expressed genes in these sites was performed. Genes common to both sites were obtained for both under-expression and over-expression. The cohort of commonly expressed genes were analysed in terms of functional ability, literature mining and protein interactions.

Four genes were found to be commonly under-expressed in the microsatellite unstable colorectal and endometrial tissue compared to microsatellite stable tumours of those sites. One of the genes, *KIAA0240*, is yet to be characterised. *CYP4AII* is involved in fatty acid metabolism and iron binding (Table 17). A literature search could not find an association between the *CYP4AII* gene and cancer.

MAML1 is mainly involved in the Notch signalling pathway though a *MAML1* – *TP53* interaction has been recently described (Zhou *et al.* 2007). The authors showed the *MAML1* protein is a constituent of an activator complex which binds to response elements within target genes that are regulated by *TP53* (Zhou *et al.* 2007). In addition, Alves-Guerra *et al.* (2007) found *MAML1* to be part of the Wnt signalling pathway and a co-activator of *β-catenin*. *In vivo* experiments showed the *MAML1* protein being recruited by *β-catenin* for interaction with the cell cycle proteins *CCND1* and *c-myc*. Further experiments by this group demonstrated that

knockdown of *MAML1* in colorectal cancer cell lines resulted in cell death. The group hypothesised cell death was due to the disruption of *β-catenin* and the failure to recruit cell cycle proliferation proteins *CCND1* and *c-myc* (Alves-Guerra *et al.* 2007). Under-expression of *MAML1* observed in this study can be postulated to result in better prognosis seen MSI malignancies. The decreased concentration of *MAML1* will decrease its interaction with *β-catenin* and in turn decrease *CCND1* and *c-myc* thus slowing down the proliferation process.

Mori *et al.* (2004) examined MSI-H colorectal tumours and found that the *PTPRU* gene was frequently methylated in predominantly right-sided malignancies. Methylation of the gene has been implicated in pancreatic carcinomas and lymphocytic leukaemia (Motiwala *et al.* 2007). *PTPRU* was commonly under-expressed in the current study and this may be attributed to partial methylation of the gene in MSI colorectal and endometrial malignancies.

Of the over-expressed genes, the highly polymorphic immune response genes *HLA-DOA* and *HLA-DMB* are of interest. Lynch syndrome patients are reported to have a better prognosis than their sporadic counterparts (Jass 2007). It is hypothesised that the better prognosis is a result of the large numbers of tumour infiltrating lymphocytes which are a hallmark of the Lynch malignancies. BRD8 is involved in apoptosis and DNA repair (GeneCards 2006). However, in the midst of a DNA repair deficient mechanism, the *BRD8* protein may be non-functional. Assuming the MSI malignancies are due to a DNA MMR deficiency, perhaps the over-expression of *BRD8* is the cell's mechanism to try to regulate the DNA repair mechanism.

ONCOMINE

The ONCOMINE site provided a comprehensive set of cancer microarray data. ONCOMINE is paper centric and provides a surplus of information on the study

under investigation. Unfortunately, a criticism of the site is that spelling errors were found to be present and when searching for specific terms, this may lead to a lack of results due to the grammatical errors.

Due to the specific nature of the study, only data from microsatellite stable and microsatellite unstable colorectal and endometrial tissue was able to be analysed. Ideally, all the Lynch susceptible sites would have been analysed. The extracolonic tumour spectrum in Lynch syndrome consists of a plethora of anatomic sites and more microarray data involving these sites would be useful in a study such as this. This lack of information therefore is a major limitation of the study.

Two of the studies investigated (Koinuma *et al.* 2004, Watanabe *et al.* 2006) examined colorectal tissue. The Watanabe *et al.* (2006) study had more significantly differentially expressed genes than the Koinuma *et al.* (2004) study. This could be attributed to the larger number of samples utilised in the Watanabe *et al.* (2006) study. The endometrial tissue study performed by Mutter *et al.* (2001) used an older microarray chip than the colorectal studies. Therefore, perhaps a larger cohort of commonly differentially expressed genes may be uncovered when the same microarray chip is utilised.

In conclusion, the results of the bioinformatics analyses of the microarray data is encouraging though the limitations discussed above should be considered. Alternate biological functions and interactions must be taken into account when deciphering the susceptibility of specific tissue sites in Lynch syndrome.

Chapter 5: Concluding remarks

The aim of the project was to elucidate the occurrence of extracolonic malignancies in Lynch syndrome. Numerous techniques were utilised in order to attempt to obtain an answer to the study.

5.1. Modifier study

The essence of the project analyses the site of cancer within Lynch syndrome. As a result, the effect of modifier genes cannot be discounted. To date, there are no studies reporting on the variation in the site of cancer and therefore no potential candidate genes.

A promoter variant within the h*MLH1* gene has been recently associated with an increased risk of endometrial and ovarian cancer within the general population. The promoter variant was analysed in this study and was not associated with either differing sites of cancer or age of onset within the cohort under investigation. It was found that significantly more individuals with an existing mutation in the hMLH1 (predisposing Lynch syndrome) ($p = 0.000181$) were found to be at a significantly increase risk for the development of colorectal cancer though there is a bias introduced in the study as a result of the DNA MMR existing mutation.

5.2. Genetic signature

It was hypothesized that malignancies originating from the same point, in other words from a defective DNA MMR gene and following the same tumourigenic pathway, should display an identical or at the very least, a similar genetic

signature. In individuals predisposed to Lynch syndrome, genetic material from tumour tissue was assessed and compared to its germline DNA to document the extent of the alterations within the tumour tissue. Malignancies from various anatomic sites were examined in order to elucidate molecular similarities between the tumours.

A panel of microsatellite rich genes and Ms-MLPA of tumour suppressor genes were utilised. Results of the 14 microsatellite rich genes showed instability of the *Erβ* gene was present in all but one of the tumour samples. The Ms-MLPA profile of all tumour samples exhibited methylation of two tumour suppressor genes, *HIC1* and *TIMP3*. Though the sample size was small and limited to three sites of cancer, these putative findings provide impetus for further study into the genetic signature of Lynch-associated malignancies. A study examining the evolution of tumourigenesis in Lynch syndrome patients will be beneficial to understanding the intricacies of the mutator phenotype. Use of the panel of microsatellite rich genes and the Ms-MLPA was valuable in the illumination of a genetic signature, though microarray experiments will provide more information.

5.3. In silico analyses

In an attempt to potentially uncover candidate genes for site of cancer, bioinformatics analyses were undertaken. The web based cancer profile database, ONCOMINE was utilised to gain information on microarray data focusing on microsatellite status of colorectal and endometrial tumour tissue. Lists of under-expressed and over-expressed genes from the microarray study that were common to both tissue sites were collated. Four genes (*KIAA0240*, *CYP4AII*, *MAML1*, and *PTPRU*) were commonly found to be under-expressed in microsatellite unstable colorectal and endometrial tissue. Three genes (*BRD8*, *HLA-DOA*, *and HLA-DMB*) were found to be commonly over-expressed. Though these results may be fortuitous due to the limited sample size (only colorectal and

endometrial tissue utilised), genomic and proteomic interrogation of these genes are required to determine their validity in vivo.

5.4. Final remarks

This study provided insight into the complexity of Lynch syndrome associated tumours. Perhaps the promiscuity of the disease may be due to the microenvironment and external environmental factors and not the genetic material in isolation – as was hypothesized in this project.

The importance of the biological pathways in their inherent environment should provide more answers into the susceptibility of certain tissue sites. In addition, alternate biological mechanisms, for example the role of the immune and vascular systems as well as the effect of carcinogens and lifestyle choices, should be thoroughly investigated.

The work undertaken in this project encompassed a plethora of technologies such as familiarization with comprehensive and molecular databases, cohort selection, genotyping, molecular mutation detection techniques, Ms-MLPA, data analysis and *in silico* analyses. .

WEBSITES UTILISED:

http://www.insight-group.org

http://www.med.mun.ca/MMRvariants/.

http://www.ncbi.nih.gov

http://www.ensembl.org

http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi.

http://www.idtdna.com/analyzer/ Applications/OligoAnalyzer/)

http://www.ncbi/nih.com/BLAST

http://ma.lungberg.gu.se/cgi-bin/cutter2)

http://www.genetics.med.ed.ac.uk/suspects/

http://www.cstl.nist.gov/div831/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm

http://www.mrc-holland.com

http://www.oncomine.org

http://www.geneontologies.org

http://www.chilibot.net

http://string.embl.de

http://www.reactome.org

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410

AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques. 2004 Aug, 37(2):226-31

Bala, S., and Peltomaki, P. (2001) Cyclin D1 as a genetic modifier in hereditary non-polyposis colorectal cancer. Cancer Res, *61*: 6042 – 6045.

Banno, K., Susumu, N., Yanokura, M. *et al.* (2004) Association of HNPCC and endometrial cancers. Int J Clin Oncol, *9*: 262 – 269.

Beiner, M.E., Rosen, B., Fyles, A. *et al.* (2006) Endometrial cancer risk is associated with variants of the mismatch repair genes *hMLH1* and *hMSH2.* Cancer Epidemiol Biomarkers Prev, *15(9)*: 1636 – 1640.

Bertholon, J., Wang, Q., Galmrinin, C.M., Puisieux, A. (2006) Mutational targets in colorectal cancer with microsatellite instability. Fam Cancer *5*: 29 - 34

Boland, C.R., Thibodeau, S.N., Hamilton, S.R. *et al.* (1998) A national cancer institute workshop on microsatellite instability for cancer detection and familial predisposition: Development of International criteria in colorectal cancer. Cancer Research, *58*: 5248 – 5257.

Cai, S.J., Xu, Y., Cai, G.X. *et al.* (2003) Clinical characteristics and diagnosis of patients with hereditary nonpolyposis colorectal cancer. World J Gasteroenterol; *9 (2)*: 284 – 287.

Caluseriu, O., Di Gregorio, C., Lucci-Cordisco, E. *et al.* (2004) A founder *HMLH1* mutation in families from the districts of Modena and Reggio-Emilia in northern Italy with hereditary non-polyposis colorectal cancer associated with protein elongation and instability. J Med Genet *41(3):* e34.

Campbell – Thompson, M., Lynch, J., Bhardwaj, B. (2001) Expression of estrogen receptor subtypes and ErB isoforms in colon cancer. Cancer Res *61:* 632 - 640

Cederquist, K., Emanuelsson, M., Wiklund, F. *et al.* (2005) Two Swedish founder *hMSH6* mutations, one nonsense and one missense conferring high cumulative risk of Lynch syndrome. Clin genet *68*: 533 – 540.

Chan, T.L., Chan, Y.W., Ho, J.W. *et al.* (2004) *hMSH2* c.1452-1455delAATg is a founder mutation and important cause of hereditary non-polyposis colorectal cancer in the southern Chinese population. Am J Hum Gen *74(5)*: 1035 – 1042

Chan, T.L., Yuen, S.T., Ho, J.W. *et al.* (2001) A novel 1.8-kb deletion of *hMLH1* mimicking alternative splicing: a founder mutation in the Chinese population. Oncogene *20(23):* 2976 - 2981

Chao, E., and Lipkin, S.M. (2006) Molecular models for the tissue specificity of DNA mismatch repair deficient carcinogenesis. Nucleic Acids Research *34(3):* 840 – 852

Chen, H., and Sharp, B.M. (2004) Content-rich biological network constructed by mining PubMed abstracts. BMC Bioinfomatics *5*: 127

Cotran, R.S., Kumar, V., Collins, T. (1999) Pathologic basis of disease. 6th edition. W.B. Saunders Company, United States of America

Deau, E., Mari, X., Sato, F., *et al.* (2004) Activin type II receptor restoration in ACVR2 deficient colon cacner cells induces the transforming growth factor beta response pathway genes. Cancer Res *64(21)*: 7690 - 7696

De Jong, M.M., Nolte, I.M., te Meerman, G.J., *et al.* (2002) Low-penetrance Genes and their Involvement in Colorectal Cancer Susceptibility. Cancer Epidemiol Biomarkers Prev *11*: 1332 – 1352

De La Rosa, M., Fasano, C., Panariello, L., *et al.* (2000) Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene. Oncogene *19(13):* 1719 - 1723

De Vos, M., Hayward, B., Bonthron, D.T. *et al.* (2005) Phenotype associated with recessively inherited mutations in DNA mismatch repair (MMR) genes. Biochemical Society Transactions, *33(4):* 718 – 720

Deng, G., Chen, A., Pong, E. *et al.* (2001) Methylation in *hMLH1* promoter interferes with its binding to transcription factor CBF and inhibits gene expression. Oncogene, *20*: 7120 – 7127

Deroo, B.J., and Korach, K.S. (2006) Estrogen receptors and human disease. J Clin Invest *116(3):* 561 – 570

Douglas, J.A., Gruber, S.B., Meister, K.A. *et al.* (2005) History and Molecular Genetics of Lynch Syndrome in Family G: A century later. JAMA; *294*: 2195 - 2202

Dribble, K.M., and McCabe, E.R.B. (2000) Modifier genes convert “simple” Mendelian disorders to complex traits. Molecular genetics and Metabolism *71:* 43 – 50.

Duval, A., Rolland, S., Compoint, A. (2001) Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. Human Molecular Genetics *10(5):* 513 – 518

Duval, A., Reperant, M., Compoint, A., *et al.* (2002) Target gene mutation profile differs between gastrointestinal and endometrial tumours with mismatch repair deficiency. Cancer Res *62:* 1609 - 1612

Ericson, K.M., Isinger, A.P., Isfoss, B.L., Nilbert, M.C. (2005) Low frequency of defective mismatch repair in a population-based series of upper urothelial carcinoma. BMC Cancer *1*: 5

El-Rifai, W., Powell, S.M. (2002) Molecular and biologic basis of upper gastrointestinal malignancy – Gastric carcinoma. Surg Oncol Clin N Am; *11 (2):* 273 – 291

Fadl-Elmula I. (2005) Chromosomal changes in uroepithelial carcinomas. Cell chr *7*: 4

Fernandez-Peralta, A.M., Nejda, N., Oliart, S. *et al.* (2005) Significance of mutations in *TGFBR2* and *BAX* in neoplastic progression and patient outcome in sporadic colorectal tumours with high microsatellite instability. Cancer Genet Cytogenet; *157*: 18 - 24

Fearnhead, N.S., Wilding, J.L., Bodmer, W.F. (2002) Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumourigenesis. Br Med Bull; *64*: 27 - 43

Fearon, E.R., and Vogelstein, B. (1990) A genetic model for colorectal carcinogenesis. Cell *61 (5):* 759 - 767

Felix, R., Bodmer, W., Fearnhead, N.S. *et al.* (2006) GSTM1 and GSTT1 polymorphisms as modifiers of age at diagnosis of hereditary nonpoylposis colorectal cancer (HNPCC) in a homogenous cohort of individuals carrying a single predisposing mutation. Mut research 602(1-2): 175 - 181

Felton, K.E.A., Gilchrist, D.M., Andrew, S.E. *et al.* (2007) Constitutive deficiency in DNA mismatch repair. Clin Genet; *71*: 483 – 498

Feng, Q., Defterreos, G., Hawes, S.E. *et al.* (2008) DNA hypermethylation, Her-2/neu overexpression and p53 mutations in ovarian carcinoma. Gynaecol Oncol, Epub ahead of print

Fleuriel, C., Touka, M., Boulay, G *et al.* (2008) HIC1 (Hypermethylated in Cancer 1) epigenetic silencing in tumours. Int J Biochem Cell Biol, Epub ahead of print

Foulkes, W.D., Thiffault, I., Gruber, S.B. *et al.* (2002) The founder mutation *HMSH2**1906G→C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. Am J Hum Gen *71(6):* 1395 – 1412

Froggatt, N.J., Green, J., Brassett, C. *et al.* (1999) A common *HMSH2* mutation in English and North-American HNPCC families: origin, phenotypic expression and sex specific differences in colorectal cancer. Journal of Medical Genetics *36(2):* 97 – 102

Garay, J., Bravo, J.C., Correa, P., Schneider, B.G. (2004) Infrequency of microsatellite instability in complete gastric intestinal metaplasia. Hum Pathol; *35 (1):* 102 – 106

Geary, J., Sasieni, P., Houlston, R. *et al.* (2007) Gene-related cancer spectrum in families with hereditary non-polyposis colorectal cancer. Fam Cancer *7(3):* 163 - 172

Geisler, J.P., Goodfellow, M.J., Sood, A.K. *et al.* (2003) Mismatch repair gene expression defects contribute to microsatellite instability in ovarian carcinoma. Cancer; *98 (10):* 2199 - 2206

GENE CARD: Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D.: GeneCards: encyclopedia for genes, proteins and diseases. Weizmann Institute of Science, Bioinformatics Unit and Genome Center (Rehovot, Israel), 1997.
Date accessed: 20 October 2008

Goecke, T., Schulmann, K., Engel, C. *et al.* (2006) Genotype-Phenotype comparison of German *hMLH1* and *hMSH2* mutation carriers clinically affected with Lynch syndrome: a report by the German HNPCC consortium. J Clin Oncol *24(26):* 4285 - 4292

Gonzalez-Aguilera, J.J., Nejda, N., Fernandez, F.J., *et al.* (2003) Genetic alterations and MSI status in primary, synchronous, and metachronous tumors in a family with hereditary nonpolyposis colorectal cancer (HNPCC). Am J Clin Oncol. *26(4)*: 386-391

Gottlieb, B., Beitel, L.K., Wu, J.H., Trofiro, M. (2004) The androgen receptor gene mutations database (ARDB): 2004 update. Hum Mutat *23(6)* 527 – 533

Grady, W.M. (2005) Molecular basis for subdividing hereditary colon cancer? Gut. *54:* 1676 – 1678

Gylling, A.H., Nieminen, T.T., Abdel-Rahman, W.M., *et al.* (2008) Differential caner predisposition in Lynch syndrome: insights from molecular analysis of brain and urinary tract tumours. Carcinogenesis *29(7):* 1351 – 1359

Halvarsson, B., Muller, W., Planck, M. *et al.* (2006) Phenotypic heterogeneity in hereditary nonpolyposis colorectal cancer: identical germline mutations associated with variable tumour morphology and immunohistochemical expression. J Clin Pathol *60(7*); 781 - 786

Hamilton, S.R., Liu, B., Parsons, R.E. *et al.* (1995) The molecular basis of Turcot syndrome. NEJM *332(13):* 839 – 847

Hartmann, A., Dietmaier, W., Hofstader, F. *et al.* (2003) Urothelial carcinoma of the upper urinary tract: Inverted growth pattern is predictive of microsatellite instability. Hum pathol; *34*: 222 – 227

Harley, I., Rosen, B., Risch, H.A. *et al.* (2008) Ovarian cancer risk is associated with a common variant in the promoter sequence of the mismatch repair gene *hMLH1*. Gynaecologic oncology; *109*: 384 – 387

Henken, F.E., Wilting S.M., Overmeer, R.M., *et al.* (2007) Sequential gene promoter methylation during HPV-induced cervical carcinogenesis. Br J Cancer; *10:* 1 – 8

Hitchins, M.P., Wong, J.J., Suthers, G. *et al.* (2007) Inheritance of a cancer associated *MLH1* germline epimutation. NEJM; *356 (7)*: 697 - 705

Huang, R.L., Chao, C.F., Ching, D.C. *et al.* (2004) Multiple epithelial and nonepithelial tumours in hereditary nonpolyposis colorectal cancer: characterization of germline and somatic mutations of the *MSH2* gene and heterogeneity of replication error phenotypes. Cancer Genet Cytogenet; *153:* 108 - 114

Hutter, P., Couturier, A., Membrez, V. *et al.* (1998) Excess of *hMLH1* germline mutations in Swiss families with hereditary non-polyposis colorectal cancer. Int J Cancer *78(6):* 680 – 684

Iglesias, D., Fernandez-Peralta, A.M., Nejda, N. *et al.* (2000) RISI, a gene with trinucleotide repeats, is a target in the mutator pathway of colorectal tumourigenesis. Cancer Genet Cytogenet; *167:* 138 - 144

Ilyas, M., Straub, J., Tomlinson, J.P., Bodmer, W.F. (1999) Genetic pathways in colorectal and other cancers. Eur J Cancer *35(14):* 1986 - 2002

International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. Guidelines for management recommended by ICG-HNPCC. 2004. Accessed online: www.insight–group.org Date accessed: 24 March 2007

Ito, E., Yanagiswa, Y., Iwahashi, Y. *et al.* (1999) A core promoter and a frequent single-nucleotide polymorphism of the mismatch repair gene *hMLH1*. Biochem Biophys Res Commun, *256*: 488 - 494

Jacob, S., and Praz, F. (2002) DNA mismatch repair defects: role in colorectal carcinogenesis. Biochimie, *84*: 27 – 47

Jager, A.C., Bisgaard, M.L., Myrhoj, T. *et al.* (1997) Reduced frequency of extracolonic cancer in hereditary nonpolyposis colorectal cancer families with monoallelic *hMLH1* expression. Am J Hum Gen, *61*: 129 – 138

Jass, J. (2007) Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathol; *50*: 113 – 130

Joensuu, E.T., Abdel-Rahman, W.M., Ollikainen, M., *et al.* (2005) Epigenetic signatures of familialcancer are characteristic of tumour type and family category. Cancer Res *68(12):* 4597 - 4605

Jun, S.H, Kim T.G., Ban, C. (2006) DNA Mismatch Repair system: Classical and fresh roles. FEBS Journal *273*: 1609 – 1619

Jung, B., Smith, E.J., Doctolero, R.T., *et al.* (2006) Influence of target gene mutations on survival, stage and histology in sporadic microsatellite unstable colon cancers. Int J cancer *118:* 2509 – 2513

Koinuma, K., Yamashita, Y., Liu, W., *et al.* (2006) Epigenetic silencing of AXIN2 in colorectal carcinomas with microsatellite instability. Oncogene 25(1): 139 - 146

Kotnis, A., Sarin, R., Mulherkar, R. (2005) Genotype, phenotype and cancer: Role of low penetrance genes and environment in tumor susceptibility. J Biosci *309(1)*, 101 – 110

Knudson (1971) Mutation and cancer: statistical study of retinoblastoma. PNAS 68 (4): 820 - 823

Kruger, S., Kinzel, M., Walldorf, C. *et al.* (2008) Homozygous PMS2 germline mutations in two families with early-onset haematological malignancy, brain tumours, HNPCC-associated tumours, and signs of neurofibromatosus type 1. Eur J Hum Genet.;*16(1)*:62-72.

Kunkel, T.A., Erie, D.A. (2005) DNA Mismatch Repair. Annu. Rev. Biochem. *74:* 681 - 710

Kuismanen, S.A., Moisio, A.I., Schweizer, P. *et al.* (2002) Endometrial and colorectal tumours from patients with hereditary nonpolyposis colon cancer display different patterns of microsatellite instability. Am J Pathol; *160 (6):* 1953 – 1980

Lax, S.F. (2004) Molecular genetics pathways in various types of endometrial carcinoma: from a phenotypical to a molecular based classification. Virchows Arch; *444:* 213 – 223

Lin, K.M., Shahidharan, M., Thorson, A.G. *et al.* (1998) Cumulative incidence of colorectal and extracolonic cancers in *hMLH1* and *hMSH2* mutation carriers of Hereditary Nonpolyposis Colorectal Cancer. Journal of Gastrointestinal Surgery, *2*: 67 – 71

Louis, D.N. (2006) Molecular pathology of malignant gliomas. Annu Rev Path; *1*: 97 - 117

Lu, K.H., and Broaddus, R.R. (2005) Gynaecologic Tumours in Hereditary NonPolyposis Colorectal Cancer, We know they are common – Now what? Gynaecologic Oncol; *82*: 221 - 222

Lui, E.L., Loo, W.T., Zhu, L., *et al.* (2005) DNA hypermethyaltion of the TIMP3 gene in invasive breast ductal carcinoma. Biomed pharmacother *59(2):* 5363 - 5365

Lynch, H.T, Schulke, G.S., Kinberling, W.J. *et al.* (1985a) Hereditary Nonpolyposis Colorectal Cancer (Lynch I and II). I: Clinical description of resource; II: Biomarker studies. Cancer; *56:* 934 - 938

Lynch, H.T., Smryk, T., Lynch, J.F. (1985b) Molecular genetics and clinical pathology features of Hereditary NonPolyposis Colorectal Cancer. Oncol; *55*: 103 - 108

Lynch, H.T., Taylor, R.J., Lynch, J.F. *et al.* (2003) Multiple primary cancer, including transitional cell carcinoma of the upper uroepithelial tract in a multigeneration HNPCC family: Molecular, Genetic, Diagnostic and Management Implications. Am J Gasteroenterol; *98 (3):* 664 - 670

Lynch, H.T., Lynch, J.F (2000) Hereditary Non-polyposis Colorectal Cancer. Sem Surg Oncol; *18*: 305 - 313

Lynch, H.T., and De la Chapelle, A. (2003) Hereditary Colorectal Cancer. NEJM *348*: 919 - 932

Lynch, H.T. and Lynch, J.F. (1997) An Update of Hereditary Nonpolyposis Colorectal Cancer. Cancer Genet Cytogenet. *93*: 84 - 99

Malander, S., Rambech, E., Krisstoffeson, U. *et al.* (2005) The contribution of hereditary nonpolyposis colorectal cancer to the development of ovarian cancer. Gynaecologic Oncol; *101 (7):* 238 - 243

Maillet, P., Chappuis, P.O., Vaudan, G., *et al.* (2000) A polymorphism in the *ATM* gene modulates the penetrance of hereditary non-polyposis colorectal cancer. Int. J. cancer. *88,* 928-931

Marti, T.M., Kunz, C., Fleck, O. (2002) DNA Mismatch repair and mutation avoidance pathways. Journal of Cellular Physiology *191*: 28 – 41

McGrath, M., Michoud, D., De Vivo, I. (2006) Polymorphisms in GSTT1, GSTM1, NAT1 and NAT2 genes and bladder cancer risk in men and women. BMC Cancer; *6*: 239

Mecklin, J.P., and Jarvinen, H.J. (1991) Tumour spectrum in Cancer Family Syndrome (Hereditary Non-Polyposis Colorectal Cancer). Cancer; *68*: 1109 - 1112

Medina-Arana, V., Barrios, Y., Fernandez-Peralta, A. (2006) New founding mutation in *hMSH2* associated with hereditary non-polyposis colorectal cancer on the Island of Tenerife. Cancer Letters *244(2*): 268 - 273

Menko, F.H., Kaspers, G.L., Meijer, G.A. (2004) A homozygous *hMSH6* mutation in a child with café-au-lait spots, oligodendroglioma and rectal cancer. Fam Cancer; 3 (2): 123 – 127

Miyaki, M., Iijima, T., Shiba, K., *et al.* (2001) Alterations of repeated sequences in 5' upstream and coding regions in colorectal tumours from patietns with hereditary nonpolyposis colorectal cancer and Turcot syndrome. *Oncogene 20:* 5215 - 5218

Miyaki, M., Konihi, M., Tanaka, K. (1997) Germline mutation of *hMSH6* as the cause of hereditary non-polyposis colorectal cancer. Nat Genet. *17:* 271 – 272

Mongiat-Artus, P., Miquel, C., Flejou, L. *et al.* (2006) Spectrum of molecular alterations in colorectal, upper urinary tract, endocervical, and renal carcinomas arising in a patient with hereditary non-polyposis colorectal cancer. Virchows Arch; *449*: 238 – 243

Moinzadeh, P., Breuhahn, K., Stutzer, H., Skirmacher, P. (2005) Chromosome alterations in human hepatocellular carcinomas correlate with aetiology and

histological grade--results of an explorative CGH meta-analysis. Br J Cancer; *92 (5)*: 935 - 841

Muller, A., Schackert, A.K., Lange, B. *et al.* (2005) A novel *hMSH2* germline mutation in homozygous state in two brothers with colorectal cancers diagnosed at the age of 11 and 12 years. Am J Med Genet A; *140 (3):* 195 – 199

Mullis, K., Faloona, F., Scharf, S. *et al.* (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol. *1*: 263 – 273

Murphy, K.,M., Zhang, S., Geiger, T. *et al.* (2006) Compensation of the microsatellite instability analysis systen and Bethesda panel for determination of MSI in colorectal cancers. J Molec Diag *8*: 30

Mutter, G.L., Baak, J.P.A., Fitzgerald, J.T., *et al.* (2001) Global gene expression changes of constitutive and hormonally regulated genes during endometrial neoplastic transformation. Gynaecol Oncol *83:* 177 - 185

Nystrom-Lahti, M., Kristo, P., Nicolaides, N.C. *et al.* (1995) Founding mutations and Alu-mediated recombination in hereditary colon cancer. Nature medicine *1(11):* 1203 – 1206

Nygren, A.O.H., Ameziane, N., Duarte, H.M.B., *et al.* (2005) Methyaltion-Specific MLPA (Ms-MLPA): simulataneous detection of CpG mehtyaltion and copy number changes of up to 40 sequences. Nuc Acids Res *33(14)*: e128

Peltomaki, P. (1993) Genetic mapping of a locus predisposing to colorectal cancer. Science 260 (5109): 810 - 812

Peltomaki, P. (2005) Lynch syndrome genes. Fam Cancer; *4*: 227 - 232

Plaschke, J., Engel, C., Kruger, C. *et al.* (2004) Lower Incidence of Colorectal Cancer and Later Age of Disease Onset in 27 Families With Pathogenic *hMSH6* Germline Mutations Compared With Families With *hMLH1* or *hMSH2* Mutations: The German Hereditary Nonpolyposis Colorectal Cancer Consortium. Journal of clinical oncology; *22*:4486-4494.

Poley, J.W., Wagner, A., Hoogmans, M.M. *et al.* (2007) Biallelic germline mutations of mismatch-repair genes: a possible cause for multiple pediatric malignancies. Cancer; *109 (11):* 2349 – 2356

Rafau, W.E., Powell, S.M. (2002) Molecular biology of gastric cancer. Seminars in Radiation Oncology; *12 (2):* 128 - 140

Rahner, N., Hoeffler, G., Hogenauer, C. *et al.* (2005) Compound heterozygosity for two MSH6 mutations in a patient with early onset colorectal cancer, vitiligo and systemic lupus erythematosus. Am J Med Genet A; *146 (10):* 1314 - 1319

Ramesar, R.S., Madden, M.V., Felix, R., *et al.* (2000) Molecular genetics improves the management of hereditary non-polyposis colorectal cancer. S Afr Med J. *90(7):* 709 – 714

Raptis, S., Mrkonjis, M., Green, R.C. (2007) MLH1 -93G>A promoter polymorphism and the risk of microsatellite unstable colorectal cancer. J natl Cancer Inst 99(6): 463 - 474

Ricciardone, M.D., Ozcelik, T., Cevher, B. *et al.* (1999) Human *MLH1* deficiency predisposes to haematological malignancy and Neurofibromatosus Type I. Cancer Res; *59*: 290 - 293

Rijcken, F.E.M., Koornstra, J.J., Van der Sluis, T. *et al.* (2005) Early carcinogenic events in HNPCC adenomas: Differences with sporadic adenomas. Dig Dis Sci; *53(6):* 1660 – 1668

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisnov, V., *et al.* (2007) Oncomine 3.0: genes, pathways, and networks in a collection of 18 000 cancer gene expression profiles. Neoplasia *9(2):* 166 - 180

Rodrigues-Bigas, M.A., Vasen, H.F.A., Lynch, H.T. *et al.* (1998) Characteristics of small bowel carcinoma in HNPCC. Am Cancer Soc; *83:* 240 - 244

Rodrıguez-Moranta, F., Castells, A., Andreu, M. *et al.* (2006) Clinical Performance of Original and Revised Bethesda Guidelines for the Identification of *HMSH2/HMLH1* Gene Carriers in Patients with Newly Diagnosed Colorectal Cancer: Proposal of a New and Simpler Set of Recommendations. American Journal of Gastroenterology; *101*:1104–1111

Rosen, S. and Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

Roupret, M., Azzouzi, A.R., Cussenot, O. (2005) Microsatellite instability and transitional cell carcinoma of the upper urinary tract. BJU Int; *96 (4):* 489 - 492

Rowley, P.T. (2005) Inherited susceptibility to colorectal cancer. Annu Rev Med; *56*: 539 - 554

Salahshor, S., Kressner, U., Pahlman, L. *et al.* (1999) Colorectal cancer with and without microsatellite instability involves different genes. Genes Chromosomes Cancer; *26(3):* 247 - 252

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) Molecular Cloning – A Laboratory Manual, 2cd edition. Cold Spring Harbour Laboratory Press, United States of America

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

Santarosa, M., Bidoll, E., Gallo, A (2002) Polymorphis CAG repeat length within the androgen receptor gene: Identification of a subgroup of patients with increased risk of ovarian cancer. Oncol rep *9*: 639 - 644

Scholfield, M.J., and Hsieh, P. (2003) DNA Mismatch Repair: Molecualr mechanisms and biological function. Annu. Rev. Microbiol *57*: 579 – 605

Sijmons, R.H., Kiemeney, L.A.L.M., Witjies, J.A., Vasen, H.F.A. (1998) Urinary Tract cancer and Hereditary Nonpolyposis colorectal cancer: Risks and screening options. J Urolog; *160*: 466 - 470

Shin, Y.K., Heo, S.C., Shin, J.H. *et al.* (2004) Germline mutations in *HMLH1*, *HMSH2* and *HMSH6* in Korean hereditary non-polyposis colorectal cancer. Human Mutation *24(4*): 351 – 355

Shitoh, K., Konishi, F., Masubuchi, S., Senba, S., Tsukamato, T., Kanazawa, K. (1998) Important microsatellite markers in the investigation of replication errors in carcinomas. Jpn J Clin Oncol, 28 (9): 538 - 541

Slattery, M.L., Sweeney, C., Murtaugh, M., *et al.* (2005) Associations between Er Alpha, Er Beta and AR in colon and rectal cancer. Cancer epidemiol biomarkers prev *14(12):* 2936 - 2942

Smith, E., De Young, D.J., Tian, Z.Q. (2008) Methylation of TIMP3 in oesophageal squamous cell carcinoma. World J Gasteroenterol *14(2):* 203 – 210

Snustad, D.P., and Simmons, M.J. (2003) Principles of genetics. 3rd edition. John Wiley and Sons. United States of America.

Song, H., Ramus, S.J., Quaye, L. *et al.* (2006) Common variants in mismatch repair genes and risk of invasive ovarian cancer. Carcinogenesis; *27(11):* 2235 - 2242

Stella, A., Surdo, N.C., Lastella, P. *et al.* (2007) Germline novel *HMSH2* deletions and founder *HMSH2* deletion with anticipation effects in HNPCC. Clinical Genetics *71(2):* 130 – 139

Stuckless, S., Parfrey, P.S., Woods, M.O. *et al.* (2007) The phenotypic expression of three *HMSH2* mutations in large Newfoundland families with Lynch syndrome. Familial Cancer *6(1):* 1 – 12

Suzuki, M., Usui, N., Furugen, Y., Mitsuhasu, N. (2001) Pyrimidine nucleoside phosphorylase acitivity in normal tissues of the uterus and ovary and in benign and malignant lesions of these organs. Int J Clin Oncol*; 6 (1):* 19 - 24

Tan, T.Y., Orme, L.M., Lynch, E. *et al.* (2007) Biallelic PMS2 mutations and a distictive childhood cancer syndrome. Ped Heamat Oncol;

Tanko, Q., Franklin, B., Lynch, H.T. *et al.* (2002) A h*HMLH1* genomic mutation and associated novel mRNA defects in a hereditary non-polyposis colorectal cancer family. Mutation research *503 (1-2*): 37 - 42

Teufel, A., Staib, F., Kanzler, S. *et al.* (2007) Genetics of hepatocellular carcinoma. Worls J Gastroenterol; *13 (16):* 2271 - 2282

Thiffault, I., Foulkes, W.D., Marcus, V.A. *et al.* (2004) Putative origin of two *HMLH1* mutations in Italian Quebec hereditary non-polyposis colorectal cancer families. Clinical Genetics *66(2):* 137 - 143

Van den Bos, M., Van den Hoven, M., Jongegan, E. *et al.* (2004) More differences between HNPCC-related and sporadic carcinomas from teh endometrium as compared to the colon. Am J Surg pathol; *28 (6)*: 706 - 711

Vasen, H.F.A., Sanders, E.A.C.M., Taal, B.G. *et al.* (1996) The risk of brain tumours in HNPCC. Int J Cancer; *65*: 422 - 425

Vasen, H.F.A., Stormorken, A., Menko, FH. *et al.* (2001) MSH2 mutation carriers are at a higher risk of cancer than *MLH1* mutation carriers: A study of HNPCC families. J Clin Oncol; *19*: 4074 – 4080

Vastrik, I., D'Eustachio, P., Schmidt, E. *et al.* (2007) Reactome: a knowledgebase of biological pathways and processes. Genome Biol *8:* R39

Vilkki, S., Launonen, V., Karhu, A., *et al.* (2002) Screening for microsatellite instability target genes in colorectal cancers. J Med Genet 39; 785 - 789

Von Mering, C., Jensen, L.J., Kuhn, M. *et al.* (2007) STRING 7 – recent developments in the intergration and prediction of protein interactions. Nucleic Acids Res *35*: 358 - 362

Wagner, A., Barrows, A., Wijnen, J.T. *et al.* (2003) Molecular analysis of hereditary non-polyposis colorectal cancer in the United States: high mutation detection rate among clinically selected families and characterization of an American founder genomic deletion of the *HMSH2* gene. Am J Hum Genet *72(5):* 1088 - 1100

Wang, Q., Lasset, C., Desseigne, F. *et al.* (1999) Neurofibromatosus and early onset cancers in h*HMLH1*- deficient children. Cancer Research *59(2*): 294 - 297

Wang, X.L., Yuan, Y., Zhang, S.Z. *et al.* (2006) Clinical and genetic characteristics of Chinese hereditary nonpolyposis families. World J Gasteroent; *12 (25):* 4074 - 4077

Watanabe, T., Kobunai, T., Toda, E*., et al.* (2006) Distal colorectal cancers with microsatellite instability (MSI) display distinct gene expression profiles that are different from proximal MSI cancers. Cancer Res 66(20): 9804 - 9808

Watson, J.D., Hopkins, N.H., Roberts, J.W. *et al.* (2004) Molecular biology of the gene. Fifth edition. Benjamin Cummings, USA.

Watson, P., Riley, B. (2005) The tumour spectrum in lynch Syndrome. Familial Cancer, *4:* 245 – 248

Watson, P.; Vasen, HFA.; Mecklin, JP. *et al.* (2008) The risk of extracolonic and extraendometrial cancer in the Lynch syndrome. Int J cancer: 123 (2): 444 - 449

Westaphalen, A.A., Russel, A.M., Buser, M. *et al.* (2005) Evidence for anticipation in hereditary nonpolyposis colorectal cancer. Hum Genet; *116 (6):* 461 - 465

Whiteside, D., McLeod, R., Graham, G. *et al.* (2002) A homozygous germline mutation in the human *HMSH2* gene predisposes to haematological malignancy and café-au-lait spots. Cancer research *62:* 359 – 362

Woerner, S.M., Kloor, M., Mueller, A., *et al.* (2005) Microsatellite instability of selective target genes in HNPCC – associated colon adenomas. Oncogene *24:* 2525 - 2535

Wong, J.J., Hawkins, N.J., Ward, R.L. (2007) Colorectal cancer: a model for epigenetic tumourigenesis. Gut; *56 (1):* 140 – 148

Worsham, M.J., Chen, K.M., Meduri, V., *et al.* (2006) Epigenetic events of disease progression in head and neck squamous cell carcinoma. Arch Otolaryngol Head Neck Surg *132:* 668 - 677

Wu, X., Lin, X., Dinney ,C.P. *et al.* (2007) Genetic polymorphism in bladder cancer. Front Biosci; *12:* 192 - 213

Zhou, X.P., Loukola, A., Salovaara, R. e*t al.* (2002) PTEN mutational spectra , expression levels and subcellular localisation in microsatellite stable and instable cancers. Am J Pathol; *161*: 439 – 447

APPENDICES

APPENDIX 1

Example of a patient consent form

APPENDIX 2

2.1.

Loading dye

40% Sucrose solution

Bromophenol blue

2.2.

1% agarose gel (100ml)

1g agarose

100ml 1 x TBE

Heat until the agarose has dissolved.

When cool, pour into gel tray and allow to set.

2.3.

10 x TBE

216g Tris

110g Boric acid

14.8g EDTA

Made up to 2 litres with distilled water

2.4.

Staining with Sybr Gold

Add 2µl of Sybr Gold to 1ml of loading dye

Vortex

Store in dark when not in use

2.5.

Staining with Ethidium Bromide

Add 3µl of EtBr per 50ml of agarose gel when cool.

Swirl around to incorporate the EtBr into the gel.

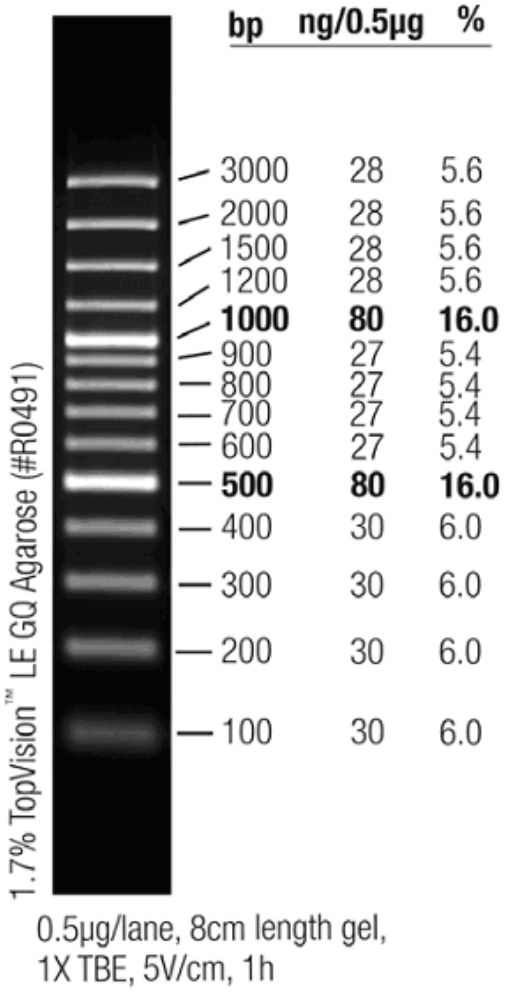
Pour into gel tray and allow to set

APPENDIX 3

GeneRuler 100bp MOLECULAR WEIGHT MARKER (FERMENTAS, USA)

Dilute to a working stock of 0.05µg/ µl

Load 10ul DNA Ladder mixed with 3ul loading dye



APPENDIX 4

WEBCUTTER OUTPUT FOR *hMLH1* PROMOTER VARIANT:

hMLH1 wild type

340 base-pairs

Eco57I NspBII
NspBII SfcI *PvuII* Eco57I
gctgtccaatcaatagctgccgctgaagggtggggctggatggcgtaagctacagctgaaggaagaacgtgagca base-pairs
cgacaggttagttatcgacggcgacttcccaccccgacctaccgcattcgatgtcgacttccttcttgcactcgt 1 to 75
MspA1I BstSFI
MspA1I

BsiI EcoT14I Eco64I BbiII
Alw21I StyI Ksp632I Hin1I
AspHI MslI Eco57I XbaI Eco130I Eam1104I Msp17I
cgaggcactgaggtgattggctgaaggcacttccgttgagcatctagacgtttccttggctcttctggcgccaaa base-pairs
gctccgtgactccactaaccgacttccgtgaaggcaactcgtagatctgcaaaggaaccgagaagaccgcggttt 76 to 150
Bbv12I ErhI SapI KasI
BsiHKAI BssT1I EarI BshNI
BssSI BanI AccB1I

EheI
NarI BbeI
Bsp143II
atgtcgttcgtggcaggggttattcggcggctggacgagacagtggtgaaccgcatcgcggcgggggaagttatc base-pairs
tacagcaagcaccgtccccaataagccgccgacctgctctgtcaccacttggcgtagcgccgccccttcaatag 151 to 225
Hsp92I HaeII
BsaHI BstH2I
AcyI
AflII
EaeI BanII MspCI
NspBII Eco24I BfrI
cagcggccagctaatgctatcaaagagatgattgagaactggtacggagggagtcgagccgggctcacttaaggg base-pairs
gtcgccggtcgattacgatagtttctctactaactcttgaccatgcctccctcagctcggcccgagtgaattccc 226 to 300
MspA1I FriOI Vha464I
CfrI BspTI

Bst98I

ctacgacttaacgggccgcgtcactcaatggcgcggacac base-pairs
gatgctgaattgcccggcgcagtgagttaccgcgcctgtg 301 to 340

PvuII	1	55	cag/ctg	More info

hMLH1 variant

340 base-pairs

Eco57I
NspBII SfcI
gctgtccaatcaatagctgccgctgaagggtggggctggatggcgtaagctacagctaaaggaagaacgtgagca base-pairs
cgacaggttagttatcgacggcgacttcccaccccgacctaccgcattcgatgtcgatttccttcttgcactcgt 1 to 75
MspA1I BstSFI

BsiI EcoT14I Eco64I BbiII
Alw21I StyI Ksp632I Hin1I
AspHI MslI Eco57I XbaI Eco130I Eam1104I Msp17I
cgaggcactgaggtgattggctgaaggcacttccgttgagcatctagacgtttccttggctcttctggcgccaaa base-pairs
gctccgtgactccactaaccgacttccgtgaaggcaactcgtagatctgcaaaggaaccgagaagaccgcggttt 76 to 150
Bbv12I ErhI SapI KasI
BsiHKAI BssT1I EarI BshNI
BssSI BanI AccB1I

EheI
NarI BbeI
Bsp143II
atgtcgttcgtggcaggggttattcggcggctggacgagacagtggtgaaccgcatcgcggcgggggaagttatc base-pairs
tacagcaagcaccgtccccaataagccgccgacctgctctgtcaccacttggcgtagcgccgccccccttcaatag 151 to 225
Hsp92I HaeII
BsaHI BstH2I
AcyI
AflII
EaeI BanII MspCI
NspBII Eco24I BfrI
cagcggccagctaatgctatcaaagagatgattgagaactggtacggagggagtcgagccgggctcacttaaggg base-pairs
gtcgccggtcgattacgatagtttctctactaactcttgaccatgcctccctcagctcggcccgagtgaattccc 226 to 300
MspA1I FriOI Vha464I
CfrI BspTI
Bst98I

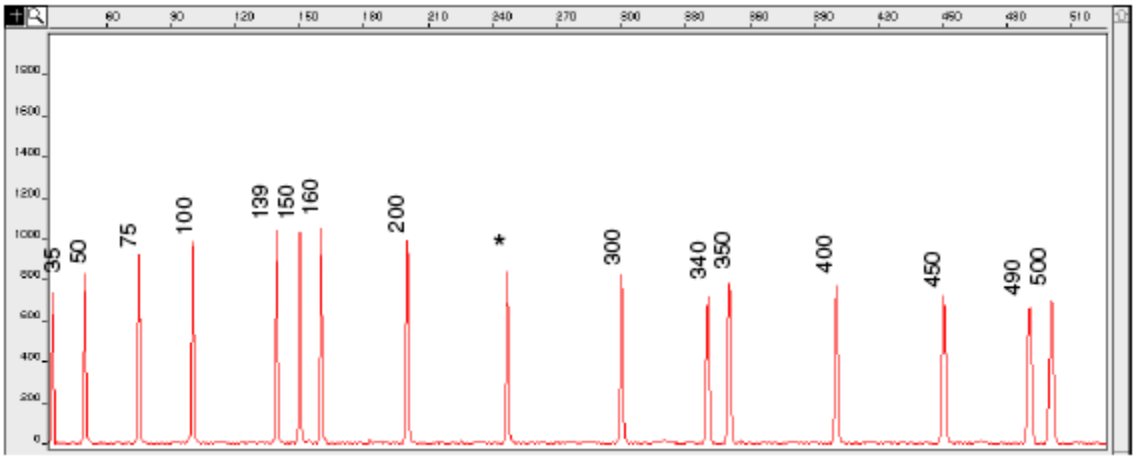
ctacgacttaacgggccgcgtcactcaatggcgcggacac base-pairs
gatgctgaattgcccggcgcagtgagttaccgcgcctgtg 301 to 340

The following endonucleases were selected but don't cut this sequence:

PspOMI, PstI, PstNHI, PvuI, *PvuII*, RcaI, RsrII, SacI, SacII, SalI,

APPENDIX 5

Rox 500 Sizing



APPENDIX 6

Allele sizes from Genotyping

Gene	Repeat	Normal	Size (bp)	Size (bp)				
				Allele 1	Allele 2	Allele 3	Allele 4	Allele 5
ACVR2	A	8	112	100	112			
Erβ	CA		159	153	157	159	161	165
MSH6	C	8	202	200	201			
AR	CAG		255	225	231	237	245	256
CASP5	A	10	322	321				
IGF2R	G	8	352	350				
TP53	A	11	443	441	442			
BAX	G	8	113	106	108	114	119	
TGFBR2	A	10	174	165				
MBD4	A	10	226	223	224			
MSH3	A	8	321	320				
PTEN	A	6	345	347				
ERa	TA		384	376	392	394	396	
TAF1B			451	441	442			