

https://dspace.org/

Pathogenic mutations and novel variants in MLH1 and MSH2 in a South African colon cancer cohort

Davison, Kenneth Mark

2013-03-19
http://hdl.handle.net/10539/12571

Downloaded from a DSpace repository. https://dspace.org/

Pathogenic mutations and novel variants in *MLH1* and *MSH2* in a South African colon cancer cohort

Kenneth Mark Davison



Dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in complete fulfilment of the requirements for the degree of Master of Science in Medicine (MSc. Med.) in Human Genetics

2012

Declaration

I, Kenneth Mark Davison, declare that this dissertation is my own work. This dissertation is being submitted for the Degree Master of Science in Medicine in Human Genetics at the University of the Witwatersrand, Johannesburg. This dissertation has not been submitted before for any degree or for examination at this or any other University.

Kenneth Mark Davison	Date

Dedication

I would like to dedicate this dissertation to all those friends and family that have helped me throughout the years:

- My mother, father, brother and sister for the support, encouragement and ideas that you provided were invaluable
- My wife Saskia Johnson, your love and support have been instrumental in helping me through the challenges of the last few years
- My friends, there are too many to mention but I am blessed to have you all around for support, every bit was great

Abstract

Identification of mutations in the mismatch repair genes of hereditary non-polyposis colorectal cancer (HNPCC) families can lead to improved management and screening of affected family members. This study aimed to characterise the mutation profile of *MLH1* and *MSH2* in a South African colorectal cancer cohort. Twenty patient samples were screened for mutations in *MLH1* and *MSH2* using Sanger sequencing and Multiplex Ligation-dependent Probe Amplification (MLPA). Three previously reported pathogenic mutations were found using Sanger sequencing, two in *MLH1* (c.454-13A>G and c.731_734delGTTA) and one in *MSH2* (c.2006-6T>C). Six novel variants were detected using Sanger sequencing, two in *MLH1* and four in *MSH2*. Further investigation of the novel variants strongly suggests that one variant in *MLH1* (c.885-1G>A) is pathogenic and two have an unknown contribution towards disease. Molecular diagnostic screening of mutations in *MLH1* and *MSH2* has the potential to improve surveillance and management of HNPCC in South Africa.

Acknowledgements

I would like to acknowledge the following for their contributions towards this study:

The funding sources that made this research possible:

- NHLS Research Trust
- Wits Faculty Research Committee

The laboratories that provided valuable input:

- Regional Molecular Genetics Service at Manchester Academic Health Science Centre, St Mary's Hospital, Manchester
- Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury
- National Health Laboratory Service Molecular Genetics Laboratory, Department of Human Genetics, Braamfontein
- National Health Laboratory Service Molecular Diagnostic Division, Department of Human Genetics, Braamfontein

My supervisors:

- Prof. Amanda Krause
- Dr. Robyn Kerr

Shelley Macaulay and the genetic counsellors at the NHLS, Braamfontein

Contents

Declaration ii
Dedication iii
Abstract iv
Acknowledgements v
List of Tables xi
List of Figures xii
Abbreviations xiv
1. Introduction 1
1.1. Cancer 1
1.1.1. Multi-step Tumourigenesis 3
1.2. Genetics of Cancer 3
1.2.1. Oncogenes 3
1.2.1.1. Oncogene Activation 3
1.2.2. Tumour Suppressor Genes 4
1.2.2.1. Silencing of TSGs 4
1.2.2.2. Knudson's Two Hit Hypothesis 4
1.2.3. Stability Genes 5
1.3. Hereditary Cancer 6
1.3.1. Familial Adenomatous Polyposis 7
1.3.2. HNPCC 7
1.3.2.1. Features of HNPCC 8
1.3.2.2. Variations o 10
1.3.2.3. Diagnostic Guidelines for HNPCC 10
1.3.2.3.1. Amsterdam Criteria 10
1.3.2.3.2. Amsterdam II Criteria 11
1.3.2.3.3. Bethesda Guidelines 12
1.3.2.3.4. Revised Bethesda Criteria 13
1.3.3. Nomenclature: HNPCC or Lynch Syndrome? 13
1.4. Mismatch Repair Genes 14
1.4.1. *MLH1* 14
1.4.2. *MSH2* 15
1.4.3. *MSH6* 15
1.4.4. *PMS2* 16

1.4.5. Recurrent and Founder Mutations 16
1.4.6. Functions of the MMR Genes 17
1.4.7. Mechanism of Tumourigenesis in HNPCC 18
1.4.7.1. Coding Microsatellites 18
1.4.7.2. Apoptosis 19
1.4.8. Tissue Specificity of HNPCC 19
1.5. Ethnic Group Differences in CRC Compared to Caucasians 20
1.5.1. Asians 21
1.5.2. African Americans 21
1.5.3. Africans 21
1.5.3.1. South African Black Population 22
1.6. Genetic Testing for HNPCC 22
1.6.1. MSI Testing 22
1.6.2. Immunohistochemistry 23
1.6.3. Direct Sequencing and MLPA 24
1.6.4. Improved Clinical Management 25
1.6.5. Diagnostic Services 25
1.7. Aims and Objectives 26
2. Subjects, Materials and Methods 27
2.1. Subjects 29
2.1.1. Samples Present in the Laboratory 29
2.1.2. Sample Collection through the Clinical Diagnostic Service 29
2.1.3. Summary of Samples Collected for Study 30
2.1.4. Control Samples 31
2.1.5. Ethics Approval 31
2.2. Materials and Methods 31
2.2.1. Multiplex Ligation-dependent Probe Amplification Analysis 31
2.2.2. Polymerase Chain Reaction 33
2.2.2.1. PCR Primers for PCR Reactions 33
2.2.2.2. PCR Conditions for DNA Amplification 34
2.2.2.3. Agarose Gel Electrophoresis of PCR Products 34
2.2.2.4. PCR Product Clean Up 34
2.2.3. Sanger Sequencing 35
2.2.3.1. Universal Primers for Sanger Sequencing 35

2.2.3.2. Sanger Sequencing Product Clean Up 36
2.2.3.3. The 3130xl Genetic Analyser 37
2.2.4. Mutation Analysis of Sequencing and MLPA Data 37
2.2.4.1. Sanger Sequencing Data Analysis 37
2.2.4.2. MLPA Data Analysis 38
2.2.4.3. Determining Reporting History of Variants 38
2.2.5. Determining Pathogenicity of Novel Variants 39
2.2.5.1. Restriction Fragment Length Polymorphism Analysis 40
2.2.5.1.1. RFLP Conditions 41
2.2.5.1.2. RFLP for *MLH1* Variants 41
2.2.5.1.3. RFLP Analysis for *MSH2* Variants 41
2.2.5.2. High Resolution Melt Analysis 42
2.2.5.3. FASTSNP 44
3. Results 45
3.1. MLPA Analysis for *MLH1* and *MSH2* 45
3.2. Variants detected using Sanger Sequencing 45
3.2.1. Variants Detected in *MLH1* 47
3.2.1.1. Previously Reported Non-Pathogenic Variants in *MLH1* 47
3.2.1.2. Previously Reported Pathogenic Variants in *MLH1* 47
3.2.1.2.1. The *MLH1* c.454-13A>G Splice Mutation 49
3.2.1.2.2. The *MLH1* c.731_734delGTTA Frameshift Mutation 49
3.2.1.3. Novel Variants Detected in *MLH1* 50
3.2.2. Variants Detected in *MSH2* 51
3.2.2.1. Positive Control 51
3.2.2.2. Previously Reported Non-Pathogenic Variants in *MSH2* 52
3.2.2.3. Previously Reported Pathogenic Variants in *MSH2* 53
3.2.2.4. Novel Variants Detected in *MSH2* 53
3.2.3. Summary of Variants Detected by Sanger Sequencing 55
3.3. Determining Pathogenicity of the Novel Variants 55
3.3.1. Restriction Fragment Length Polymorphisms 55
3.3.1.1. *MLH1* c.885-1G>A variant 55
3.3.1.2. *MSH2* c.-185C>A variant 56
3.3.1.3. *MSH2* c.1662-43C>T variant 56
3.3.1.4. *MSH2* c.1662-10C>T variant 56

3.3.2. High Resolution Melt Analysis 57
3.3.2.1. *MLH1* c.1732-19T>A variant 57
3.3.2.2. *MSH2* c.367-16T>A variant 57
3.3.3. FASTSNP Analysis 58
3.3.4. Summary of Investigation of Pathogenicity for Novel Variants 59
3.4. Characterising the Mutation Spectrum 59
3.4.1.1. Variants Separated by Ethnicity 59
4. Discussion 62
4.1. Patients Fulfilling Inclusion Criteria 62
4.1.1. Clinical Management of HNPCC Patients 63
4.2. MLPA Analysis 64
4.3. Variants Detected using Sanger Sequencing 64
4.3.1. Number of Variants in Study Cohort 65
4.3.2. Previously Reported Variants in *MLH1* 66
4.3.2.1. The *MLH1* c.655A>G (pIle219Val) variant 66
4.3.2.2. The *MLH1* c.731_734delGTTA variant 66
4.3.3. Previously Reported Variants in *MSH2* 67
4.3.3.1. The *MSH2* c.1811C>A (p.Ala604Asp) variant 67
4.3.3.2. The *MSH2* c.2006-6T>C mutation 67
4.3.3.3. The *MSH2* c.573C>T (p.Leu191Leu) variant 68
4.4. Determining the Pathogenicity of the Novel Variants 68
4.4.1. The *MLH1* c.885-1G>A mutation 68
4.4.2. The *MLH1* c.1732-19T>A variant 69
4.4.3. The *MSH2* c.-185C>A variant 69
4.4.4. The *MSH2* c.367-16T>A variant 69
4.4.5. The *MSH2* c.1662-10C>T variant 70
4.4.6. The *MSH2* c.1662-43C>T variant 70
4.4.7. CRC in Patients with No Pathogenic Mutations 70
4.4.8. Molecular Diagnostic Screening of HNPCC Patients 71
4.5. Characterising the Mutation Spectrum 72
4.5.1. Mutation Spectrum of *MLH1* 72
4.5.2. Mutation Spectrum of *MSH2* 72
4.5.3. Mutation Spectrum with regard to Ethnicity 72
4.5.4. Mutation Spectrum of Pathogenic Mutations 72

4.6. Limitations and Future Directions 73
5. Conclusion 76
Appendix A: Cancer Statistics 78
Appendix B: Solutions 83
Appendix C: Consent Forms 85
Appendix D: Ethics Clearance Certificate 90
Appendix E: PCR Primers for *MLH1* and *MSH2* 91
Appendix F: Protocols Used 97
F.1 MLPA Protocol 97
F.2 Protocol for MLPA on the 3130xl Genetic Analyser 98
F.3 PCR Reaction and cycling conditions 100
F.4 Minelute 96 UF PCR Purification Kit Protocol 101
F.5 Nucleospin Extract II Clean-up Protocol: 101
F.6 Sanger Sequencing 101
F.7 DyeEx 96 Dye-Terminator Removal Protocol 102
F.8 DyeEx 2.0 Dye-Terminator Removal Modified Protocol 102
F.9 Protocol for Sanger Sequencing using the 3130xl Genetic Analyser 103
F.10 Retrieval of Data 104
F.11 Using the LOVD InSiGHT Database 105
F.12 High Resolution Melt Procedure 106
Appendix G: Patient Pedigrees 107
Appendix H: Suppliers of Equipment and Reagents 109
References: 110
Electronic References 116

List of Tables

Table 1.1: Lifetime Risk of Individual Cancers in Families with an Identified MMR Mutation 8
Table 1.2: Percentage of HNPCC cases caused by mutations in each MMR gene 14
Table 1.3: Founder mutations of *MLH1* and *MSH2* found in different populations 17
Table 2.1: Summary of the patients and their inclusion criteria 30
Table 2.2: Restriction enzymes used to investigate the frequencies of the novel variants 40
Table 2.3: Summary of the RFLP reactions for all enzymes 41
Table 3.1: Summary of all variants detected in each patient in *MLH1* 46
Table 3.2: Summary all variants detected in each patient in *MSH2* 46
Table 3.3: Summary of the previously reported non-pathogenic variants in *MLH1* that were detected in the South African cohort 48
Table 3.4: Summary of the previously reported pathogenic variants that were detected in *MLH1* in the South African cohort 48
Table 3.5: Summary of the novel variants detected in *MLH1* in the South African cohort 50
Table 3.6: Summary of the previously reported non-pathogenic variants in *MSH2* in the South African cohort 52
Table 3.7: Summary of the reporting history for the previously reported pathogenic mutation in *MSH2* detected in this study 53
Table 3.8: Summary of the novel variants detected in *MSH2* in this study 53
Table 3.9: Summary of the allele frequencies that were determined by RFLP for each variant 55
Table 3.10: Summary of allele frequencies for the two variants that were investigated using HRM 58
Table 3.11: Adaptation of the FASTSNP results for the *MSH2* c.-185C>A variant. 59
Table 3.12: Summary of the pathogenic status for the seven novel variants detected in the South African cohort 59
Table 3.13: Summary of the variants and the ethnic groups they were detected in. 60
Table 3.14: Summary of patients and the pathogenic mutations found in this study 61
Table 4.1: Summary of mutation detection rates from previous studies which were investigating *MLH1* and *MSH2* 65
Table A.1: Estimated global incidence of cancer for both genders of all ages 78
Table A.2: Estimated global cancer mortality for both sexes of all ages 79
Table A.3: Summary rates for top 5 cancers in South Africa by sex 81
Table A.4: Summary statistics for colorectal cancer, 1998 and 1999 82
Table E.1: PCR Primers for *MLH1* and PCR Product Sizes 91
Table E.2: PCR primers for MSH2 and the PCR product sizes. 92
Table E.3: DNA Sequence of *MLH1* 93
Table E.4: DNA Sequence of *MSH2* 95
Table F.1: Hybridisation Mix for MLPA Technique 97
Table F.2: Ligation mix for MLPA Technique 97
Table F.3: Pre-PCR reaction mix for MLPA Technique 97
Table F.4: Polymerase Mix for MLPA Technique 98
Table F.5: PCR Conditions for MLPA Technique 98
Table F.6: PCR reaction mix used for amplification of all exons for 1X reaction 100
Table F.7: PCR conditions for amplification of all reactions 100
Table F.8: Sanger Sequencing PCR Reaction mix 102
Table F.9: Cycle Sequencing PCR Protocol 102

Table F.10: HRM master mix that was used for all HRM reactions 106
Table H.1: Table of Suppliers of Equipment and Reagents 109

List of Figures

Figure 1.1: Phenotypic Features of HNPCC 9
Figure 1.2: Amsterdam Criteria for the Clinical Diagnosis of HNPCC 11
Figure 1.3: Amsterdam II Criteria used for the Clinical Diagnosis of HNPCC 11
Figure 1.4: Bethesda Guidelines for Genetic Testing of MMR Genes 12
Figure 1.5: Revised Bethesda Guidelines for Genetic Testing of MMR Genes 13
Figure 1.6: Diagram showing how the MMR proteins interact to repair mismatches in the DNA 19
Figure 2.1: Flow Diagram outlining the approach and actions taken in this study 28
Figure 2.2: The MLPA reaction. 32
Figure 2.3: Schematic representation of PCR and sequencing reactions in this study 36
Figure 2.4: Diagram of the expected RFLP analysis using *HpyCH*4III to screen for the *MLH1* c.885-1G>A variant 41
Figure 2.5: Diagram showing the expected RFLP fragments when *Taq*I is used to investigate the *MSH2* c.1662-10C>T variant 42
Figure 2.6: Diagram showing the expected RFLP fragments when *Rsa*I is used to investigate the *MSH2* c.1662-43C>T variant 42
Figure 2.7: The expected RFLP fragments when *MluI* is used to investigate the *MSH2* c.-185C>A variant located within the regulatory region 43
Figure 2.8: The HRM normalised melting curve obtained from control samples 43
Figure 3.1: The MLPA result generated using the spread sheet from NGRL. 45
Figure 3.2: The *MLH1* c.655A>G variant (p.I219V), which leads to a missense mutation, in exon 8 47
Figure 3.3: The *MLH1* c.454-13A>G variant of intron 5 found in Patient 104 49
Figure 3.4: The variant data obtained from LOVD for the *MLH1* c.731_734delGTTA variant 49
Figure 3.5: The *MLH1* c.731_734delGTTA mutation in exon 9 50
Figure 3.6: The *MLH1* c.885-1G>A novel variant located within splice acceptor site of intron 10. 50
Figure 3.7: The *MLH1* c.1732-19T>A novel variant that was detected in intron 15 51
Figure 3.8: The *MSH2* c.1811C>A novel variant in exon 12. 52
Figure 3.9: The *MSH2* c.2006-6T>C pathogenic variant of intron 12 53
Figure 3.10: The *MSH2* -185C>A novel variant in the regulatory region 54
Figure 3.11: The *MSH2* c.367-16T>A novel variant of intron 2 54
Figure 3.12: The *MSH2* c.1662-43C>T novel variant located within intron 10 54
Figure 3.13: The *MSH2* c.1662-10C>T novel variant located within intron 10. 54
Figure 3.14: RFLP analysis with *HpyCH*4III performed on patient 81 56
Figure 3.15: RFLP gel picture from using *Taq*I to investigate the *MSH2* c.1662-10C>T variants in patients 235 and 238. 57
Figure 3.16: The HRM normalised melting curve for the assay of *MLH1* c.1732-19T>A variant 58
Figure A.1: Percentage distribution of 10 most common cancers in South Africa by gender, 1998 and 1999, all ages (NCR, 2004) 80

Figure G.1: The pedigree for black patient 205 107
Figure G.2: Pedigree for black patient 235. 107
Figure G.3: Pedigree for black African patient 238 107
Figure G.4: Pedigree for patient 136 108
Figure G.5: The pedigree for patient 150 where the *MSH2* c.1811C>A variant was found 108

Abbreviations

5’	Five Prime
3’	Three Prime
%	Percentage
α	Alpha
µ	Micro
µg	Microgram
µl	Microlitre
A	Adenosine
APC	Adenomatous Polyposis Coli
ASR	Age Standardized Rate
bp	Base-Pair
C	Cytosine
cm	Centimetre
CRC	Colorectal Cancer
dATP	Deoxyadenosine-5’-Triphosphate
dbSNP	Single Nucleotide Polymorphism Database
dCTP	Deoxycytidine-5’-Triphosphate
ddH_2O	Double Distilled Water
dGTP	Deoxyguanosine-5’-Triphosphate
dNTP	Deoxyribonucleotide Triphosphate
dTTP	Deoxythymidine-5’-Triphosphate
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetra-Acetic Acid
EtBr	Ethidium Bromide
FAP	Familial Adenomatous Polyposis

FASTSNP	Functional Analysis and Selection Tool for Single Nucleotide Polymorphisms
G	Guanine
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HRM	High Resolution Melt
IARC	International Agency for Research on Cancer
IHC	Immunohistochemistry
kb	Kilobases
LOVD	Leiden Open Variation Database
M	Molar
$MgCl_2$	Magnesium Chloride
MLH1	mutL Homolog 1
MLPA	Multiplex Ligation-dependent Probe Amplification
MMR	Mismatch Repair
mRNA	Messenger RNA
MSH2	mutS Homolog 2
MSI	Microsatellite Instability
MSI-H	High microsatellite instability
MSI-L	Low microsatellite instability
n	Nano
NCR	National Cancer Registry
ng	Nanograms
NGRL	National Genetics Reference Laboratory
NHLS	National Heath Laboratory Service
PCR	Polymerase Chain Reaction
PMS1	Postmeiotic Segregation 1
PMS2	Postmeiotic Segregation 2
RFLP	Restriction Fragment Length Polyorphism

RNA	Ribonucleic Acid
Rpm	Revolutions per Minute
SOP	Standard Operating Procedure
T	Thymine
Taq	*Thermus* aquaticus
TBE	Tris, Boric Acid, EDTA
TE	Tris-EDTA
tRNA	Transfer RNA
U	Units
V	Volts
WHO	World Health Organization

1. Introduction

Cancer generally affects older people, as internal and external risk factors accumulate over a period of time that lead to a cell growth advantage (Vogelstein and Kinzler, 2004). However, there are cancer syndromes that co-segregate within families that affect younger individuals. A clear pattern of inheritance has been described for colorectal cancer (CRC) and many cancers (Garber and Offit, 2005). One such cancer syndrome is hereditary non-polyposis colorectal cancer (HNPCC), which is the focus of this study.

Section 1.1 will provide a description of cancer and the processes that lead to tumourigenesis before briefly describing the incidence and mortality rates of cancer. The genetic basis of cancer and the role of specific classes of genes in tumourigenesis will be covered in Section 1.2. Section 1.3 will then focus on hereditary cancer before leading into hereditary CRC. HNPCC and the features of the disease will be described in Section 1.3 as well as the variations of HNPCC, the clinical criteria for the diagnosis and nomenclature for the disease.

HNPCC is caused by defects in the mismatch repair (MMR) genes. The MMR genes, their functions and role in cancer predisposition will be described in Section 1.4. Because there are different ethnic groups that are included in this study, the differences among the ethnic groups relative to CRC in Caucasians will be described in Section 1.5. The methods for molecular detection of HNPCC by will be described in Section 1.6. Finally, this study was established to characterise the mutation spectrum of HNPCC in South Africa since it has not been well characterised in South Africa. Characterisation of the mutation spectrum will allow for the establishment of molecular diagnostic testing in South Africa, which will be necessary to provide a better healthcare service to South Africans. Since *MLH1* and *MSH2* account for approximately 90% of HNPCC mutations, these genes were chosen to be screened first for mutations. The aims and objectives of this study are stated in Section 1.7.

1.1. Cancer

Cancer is a disease characterized by the uncontrolled growth of cells within the body, usually leading to the formation of a tumour. These 'rogue' cells are also able to spread to other parts of the body where they invade other tissues in a process known as metastasis. It is the two characteristics of uncontrolled growth and the ability to spread throughout the body that define a cell as cancerous (Alberts et al., 2002).

According to Ferlay et al. (2010), in 2008, there were approximately 12.7 million new cases of cancer and there were approximately 7.6 million deaths worldwide, 64% of which occurred in developing nations. The most common cancers worldwide are lung cancer, breast cancer, CRC and prostate

cancer. This information can be found at GLOBOCAN (http://globocan.iarc.fr). This study will focus on CRC which accounted for 1.2 million cases of all the cancer in 2008. The global incidence and mortality statistics for cancer can be seen in Appendix A, under Tables A.1 and A.2.

CRC is the third most diagnosed cancer in males and the fourth most diagnosed cancer in females. In 2008, there were approximately 609 000 deaths from CRC among both sexes at all ages. It is also indicated that more males develop CRC than females, perhaps due to differences in risk factors (Ferlay et al., 2010; Cunningham et al., 2010). There are several changing risk factors, such as diet and obesity that are leading to increasing incidence in many developing nations (Cunningham et al., 2010; Jemal et al., 2011).

As CRC is one of the most common cancers in the world, its early identification and management is important in order to lower mortality. Early detection strategies aim to identify a tumour before it grows too large and metastasises, which is vital for preventing mortality. Evidence of the success of this strategy can be seen by the lower mortality rate in developed countries where early detection and treatment is available (Etzioni et al., 2003; Cunningham et al., 2010; Jemal et al., 2011).

Detecting a tumour before it grows too large is important, but how does a tumour grow? Tumour growth is not caused by one single factor but rather by a series of factors that interact with the cell that leads toward uncontrolled growth. These contributing factors are internal (DNA mutations, hormones and immune responses) and external (UV light, tobacco, and infectious organisms) to the cell (Strachan and Read, 2004). When these factors interact they lead to mutations in the DNA of a cell. As mutations accumulate in the cell, normal functioning becomes disrupted in the proliferation, apoptosis or repair pathways (Vogelstein and Kinzler, 2004).

A cell colony that grows uncontrollably will lead to the development of a neoplasm, which comes from the Greek for ‘new growth’. A neoplasm that does not invade surrounding cells but is still dividing rapidly is considered to be a benign tumour. Benign tumours can be removed easily with surgery depending on the tumour position and the patient has a vastly improved prognosis if a benign tumour is recognised and removed before it becomes malignant (Alberts et al., 2002; Etzioni et al., 2003).

However, the prognosis for a patient decreases as the cell colony progresses from a benign tumour to a malignant tumour. A malignant tumour is one that is localised but has invaded the surrounding tissues (Alberts et al., 2002; Etzioni et al., 2003). It is often the metastasic effect of a tumour on the body that leads to death (Nguyen et al., 2009). The progression from a normal cell, to a benign neoplasm and finally to a malignant tumour is a feature of cancer known as multi-step tumourigenesis.

1.1.1. Multi-step Tumourigenesis

Multi-step tumourigenesis can be described using observations from the colorectal tract in a process known as the adenoma-carcinoma sequence. An adenoma, which is a benign tumour, can develop after several mutations have accumulated in different genes. However, the adenoma may never acquire the further mutations necessary to develop into a carcinoma (malignant tumour). Conversely, the adenoma could develop into a carcinoma very quickly if the necessary mutations are acquired in a short space of time (Cunningham et al., 2010).

In this way, CRC exemplifies the theory that cancer requires many different processes and mutations to occur before a tumour becomes malignant. Hence, multiple steps are required for a neoplasm to develop into a malignant tumour (de la Chapelle, 2004). However, once a tumour becomes malignant it is able to metastasise.

1.2.Genetics of Cancer

There are three main classes of genes involved in cancer development: oncogenes, tumour suppressor genes and stability genes.

1.2.1. Oncogenes

The oncogenes are genes that increase the number of cells by stimulating cell growth. The genes involved in normal promotion of cell proliferation are referred to as proto-oncogenes in the cell. However, the term oncogene is ascribed to these genes when they acquire mutations that lead to activation (Strachan and Read, 2004; Vogelstein and Kinzler, 2004).

Oncogenes' role in cancer development is through control of proliferation pathways, including the secretion and reception of growth factors, signal transduction systems, transcription factor and cell cycle regulators (Strachan and Read, 2004; Vogelstein and Kinzler, 2004; Croce, 2008; Boland and Goel, 2010). There are over 100 known oncogenes vital for normal growth and development. A lack of function of these genes would prevent a group of cells from growing (Alberts et al., 2002; Vogelstein and Kinzler, 2004). The activation of these genes can lead to excessive gene products or production of modified/novel products. Since oncogenes exhibit dominant effects, the activation of one allele is necessary for tumourigenesis (Strachan and Read, 2004; Vogelstein and Kinzler, 2004).

1.2.1.1. Oncogene Activation

There are three means by which oncogenes can be activated within the cell. Amplification may result in excessive amounts of gene products, which leads to an increase in cell growth (Vogelstein and Kinzler, 2004; Albertson, 2006; Croce, 2008). Proto-oncogenes can also translocate to transcriptionally active regions where they are overexpressed, or fuse with other genes to create

novel chimeric genes whose products leads to excess cell proliferation (Mitelman, Johansson and Mertens, 2007; Croce, 2008).

Point mutations can occur when mistakes in DNA replication are not corrected. Mutations can occur in regulatory regions which can lead to overexpression of a gene, or they can affect protein and enzyme activity in a way that enhances their activity. Furthermore, genes coding for cell surface receptors may be rendered constitutively active via point mutations (Albertson, 2006; Croce, 2008).

1.2.2. Tumour Suppressor Genes

Tumour Suppressor Genes (TSGs) are genes that function in inhibiting cell growth or provide check points to ensure fidelity of replication. TSGs act in an autosomal recessive fashion, which is different from the autosomal dominant pattern seen in proto-oncogenes. The homozygous loss-of-function of a TSG will lead to increased growth of cells (Alberts et al., 2002; Vogelstein and Kinzler, 2004).

1.2.2.1. Silencing of TSGs

Point mutations in crucial regions can lead to the silencing of a TSG through loss of expression, mRNA or protein degradation or truncated proteins (Strachan and Read, 2004; Vogelstein and Kinzler, 2004). TSGs can also be silenced by insertions or deletions in a chromosome, which can eliminate an entire gene or alter the gene sequence dramatically (Strachan and Read, 2004; Vogelstein and Kinzler, 2004). Epigenetic silencing occurs through methylation of a regulatory region and has been implicated in many cases of cancer. Methylation occurs frequently throughout the genome but aberrant methylation has been implicated in disease (Strachan and Read, 2004; Vogelstein and Kinzler, 2004; Hitchins et al., 2009).

Generally, TSGs are able to function adequately even if one copy is non-functional and the other allele is functional. This is when there is heterozygosity of a gene. The loss of heterozygosity is one manner by which TSG lead to tumourigenesis. When the one functional copy is eliminated, the functions of the TSG are abolished in the cell. It has been found that individuals can be predisposed to cancer syndromes if they inherit one non-functional copy of a TSG; this phenomenon has been described as Knudson's Two Hit Hypothesis.

1.2.2.2. Knudson's Two Hit Hypothesis

Alfred Knudson proposed his 'two hit' hypothesis based on evidence from various cancer studies. Knudson was investigating how many 'hits' were required for a detectable cancer to develop. To do this, he studied cancers that were specific to children (Knudson, 1971).

Knudson (1971) reasoned that cancers require somatic mutations, thus the longer you live, the more mutations you develop and hence older people are more likely to be affected. However, cancers in children have had relatively little time to develop multiple somatic mutations. Knudson (1971) determined that two 'hits' were required for cancer development. One allele was inherited as non-functional and the other allele acquired a loss-of-function somatic mutation.

The two-hit hypothesis can be applied to many hereditary cancers where abolition of the TSG, or the stability gene, leads to the development of the cancer. Both TSG and stability gene classes act in an autosomal recessive manner. Since a cancer-causing mutation is already present in every cell of the body, the person carrying the mutation is at a higher risk of developing cancer (Knudson, 2001; Vogelstein and Kinzler, 2004). Predisposition to cancer syndromes will be described in more detail in Section 1.3.

1.2.3. Stability Genes

The third class of genes implicated in carcinogenesis are the stability genes, also known as 'caretaker genes' and these genes are responsible for genomic stability (Vogelstein and Kinzler, 2004). The MMR genes are a subclass of the stability genes and are the focus of this study. The basic function of the stability genes, and the MMR genes, will be described before a brief description of their involvement in cancer development will be given.

Stability genes code for proteins that are responsible for repairing mistakes that are made by DNA polymerase during DNA synthesis. A DNA polymerase has a very low error rate when synthesising DNA. However, highly repetitive regions, such as microsatellites, are prone to errors during replication through polymerase slippage (Jiricny and Nyström-Lahti, 2000; Strachan and Read, 2004; Boland and Goel, 2010).

Polymerase slippage occurs when there is incorrect pairing between the complementary strands of DNA and this leads to base/base mismatches, or the formation of insertions or deletions which create an insertion/deletion loop (Strachan and Read, 2004). Normally these errors are detected by the stability genes and corrected. If not corrected, a mutation will occur and remain in the genome, which in turn may lead to carcinogenesis after an accumulation of mutations (Jiricny and Nyström-Lahti, 2000; Strachan and Read, 2004; Klug et al., 2006; da Silva et al., 2009; Cunningham et al., 2010).

The stability genes include the nucleotide-excision repair genes, base-excision repair genes and the MMR genes. Defects in the stability genes can lead to genomic instability, which is manifested in two forms. The first commonly seen form is chromosomal instability and the second form is microsatellite instability (MSI) (Strachan and Read, 2004; Cunningham et al., 2010). MSI is an

important aspect of mutations in the MMR genes, whereby MSI is a marker for tumourigenesis (see Section 1.4.7) and can be used as a preliminary screening test for HNPCC (described in Section 1.6.1).

The stability genes are vital for correcting mutation errors and preventing them from being passed on through successive mitotic divisions. Generally, stability genes have been found to behave in an autosomal recessive fashion, indicating that both alleles need to be inactivated in order to achieve an increased mutation rate. When mutations occur in the stability genes that lead to their inactivation, much like the TSGs, a higher mutation rate is observed throughout the genome. This indicates that all genes in the genome will be affected by inactivation of the stability genes, including the oncogenes and TSGs (Vogelstein and Kinzler, 2004; Boland et al., 2008). The stability genes play an important, indirect role in cancer development.

Vogelstein and Kinzler (2004) state that mutations, in any of the above three classes, can occur in the germ line. Germ line mutations can lead to cancer predispositions in individuals and their families. These predispositions increase the lifetime risk of developing a particular cancer (Frank, 2004; Vogelstein and Kinzler, 2004).

1.3.Hereditary Cancer

It is estimated that about 20% of all CRC worldwide has a hereditary component. This is quite significant considering there are over 1.2 million new cases of CRC every year (de la Chapelle, 2004). Hereditary cancer syndromes are often the result of inheriting germ line mutations in genes that play a role in tumourigenesis. Some of these syndromes, such as HNPCC, are also known to have associated cancers that develop in other organs, such as the endometrium (Garber & Offit, 2005; Lynch et al., 2006; Cunningham et al., 2010).

Genes that predispose individuals to cancer tend to affect cell death pathways, cell-cycle regulation pathways or DNA repair pathways. Germ line mutations in oncogenes are generally non-viable as the resultant phenotype is too severe. However, cancers involving inherited mutations in TSGs and stability genes are more common (Fletcher and Houlston, 2010). Cancer predisposing genes that are highly penetrant follow a single-gene Mendelian pattern of inheritance. Mutations in these genes tend to be physiologically recessive but they are inherited in a dominant fashion, via Knudson's two hit hypothesis. A mutation will be inherited in one allele but the individual eventually acquires the mutation in the second allele. Only once both alleles are non-functional will cancer arise (Knudson, 1971; Knudson, 2001; Frank, 2004, Fletcher and Houlston, 2010).

The effect of the hereditary components in CRC is dependent on the penetrance of the particular genes that are involved (Cunningham et al., 2010, Fletcher and Houlston, 2010). CRC may seem familial but may be chance clustering since it is a common disease. A seemingly sporadic case may

be hereditary but may have been misclassified due to small family size, poor diagnostics or incomplete family history (de la Chapelle, 2004; Jasperson et al., 2010).

CRC syndromes, such as Familial Adenomatous Polyposis (FAP) and HNPCC, have been vital to the understanding of CRC since several highly penetrant genes have been identified in these syndromes. These two CRC syndromes are the most common hereditary variants of CRC, accounting for approximately 1% and 5% of all CRC cases respectively. The genes implicated in these syndromes are also involved in sporadic cases of CRC and their identification has led to better understanding of sporadic and hereditary CRC (de la Chapelle, 2004; Jasperson et al., 2010).

The following sections will describe both these diseases in more detail. Although FAP is caused by mutations in the *APC* tumour suppressor gene and HNPCC is caused by mutations in the MMR genes, both diseases follow the two-hit hypothesis.

1.3.1. Familial Adenomatous Polyposis

FAP is a hereditary CRC syndrome characterised by hundreds to thousands of adenomas that present in the colon of affected individuals. These adenomas usually develop in childhood and there is practically a 100% lifetime risk that at least one of the adenomas will develop into a carcinoma. Of all the CRC cases, FAP is estimated to account for about 0.2%-1% (de la Chapelle, 2004; Gryfe, 2009).

FAP is caused by germ-line mutations in the tumour suppressor gene *Adenomatous Polyposis Coli* (*APC*) gene. FAP is inherited in an autosomal dominant manner via the two-hit hypothesis (de la Chapelle, 2004; Gryfe, 2009). Additionally, somatic mutations of the *APC* gene are known to be involved in sporadic CRC, leading to this gene being considered the 'gatekeeper' of neoplasia in the colorectal tract (Gryfe, 2009). It has also been noticed that *APC* mutations occurs early in carcinogenesis, often being the initiator of the adenomatous process (de la Chapelle, 2004)

1.3.2. HNPCC

HNPCC is a group of hereditary cancer syndromes that frequently manifest in the colorectal tract. HNPCC is the most common form of hereditary CRC, with an incidence estimated to be between 2% to 5% of all new CRC cases diagnosed annually (Lynch et al., 2006). When pathogenic mutations in the MMR genes are detected, HNPCC is reclassified as Lynch Syndrome (Vasen et al., 2007; Cunningham et al., 2010).

Families with HNPCC are often identified by an increased number of cases of CRC, endometrial cancer and a number of other cancers that are associated with HNPCC (ovaries, small bowel, hepato-biliary tract, pancreas, upper uro-epithelial tract, brain, kidney and stomach) as can be seen in Table

1.1. HNPCC is relatively common for a lethal, autosomal dominant disease. However, HNPCC generally does not affect the reproductive fitness of a carrier before they have children, leading to the mutations being transferred to children (Frank, 2004; Garber and Offit, 2005; Lynch et al., 2006).

Table 1.1: Lifetime Risk of Individual Cancers in Families with an Identified MMR Mutation (Vasen et al., 2007)

Type of Cancer	Lifetime Risk
Colorectal Cancer (Men)	28-75%
Colorectal Cancer (Women)	24-52%
Endometrial Cancer	27-71%
Ovarian Cancer	3-13%
Gastric Cancer	2-13%
Urinary Tract Cancer	1-12%
Brain Tumour	1-4%
Small Bowel Cancer	4-7%

1.3.2.1. Features of HNPCC

HNPCC receives its name from the paucity of polyposis observed in patients. This is in contrast to the high number of polyps observed in other hereditary CRC syndromes such as FAP where hundreds to thousands of polyps are observed in the colon. Therefore, polyps (or adenomas) are observed in HNPCC but there are significantly fewer polyps, and the few that develop are more likely to develop into carcinomas (Jasperson et al., 2010).

HNPCC has distinct features which allow it to be differentiated from other CRC syndromes and which can be seen in Figure 1.1. HNPCC follows an autosomal dominant inheritance pattern (Lynch et al., 2006). It can also be noted that HNPCC is a highly penetrant disease, indicating that an MMR mutation carrier has a higher lifetime risk of developing colonic or extra-colonic tumours than the general population (Lynch et al., 2006).

In the general Caucasian population, the average age for sporadic CRC development is about 65. However, the average age of cancer development is about 45 years in HNPCC patients (Lynch et al, 2006). It is this feature which the Bethesda criteria (described in Section 1.3.2.3.3) uses to determine if an individual should be sent for further investigation.

Between 70% and 85%, of colonic tumours develop in the proximal colon, with the estimated lifetime risk of CRC between 25% and 75% in carriers (de la Chapelle, 2004; Jasperson et al., 2010). The lifetime risk of developing colonic and extra-colonic tumours is given in Table 1.1. The risk of

developing the other extra-colonic cancers is not as high as CRC but identification of people with MMR mutations can allow those individuals to detect these cancers early (Vasen et al., 2007). In addition, the risk of developing an extra-colonic cancer increases with age (Watson et al., 2008).

Autosomal dominant inheritance pattern
Younger age of onset (<50) of CRC compared to average age of CRC in general population (≈65)
Between 70% and 85% of tumours develop in the proximal colon
Accelerated carcinogenesis
High risk of synchronous and meta-chronous tumours
Increased risk of malignancy at specific extra colonic locations
Sebaceous adenomas, sebaceous carcinomas and keratoacanthomas in Muir Torre Syndrome
Pathology of tumour is poorly differentiated, mucinous, signet rings and excess of infiltrating lymphocytes within tumour
Improved prognosis versus other CRC

Figure 1.1: Phenotypic Features of HNPCC (Lynch et al., 2006)

HNPCC is marked by a significant excess of synchronous and meta-chronous CRCs. Synchronous tumours are present at the same time whereas meta-chronous tumours are present within 10 years of each other. It is estimated that about 30% of patients will develop a second primary CRC within 10 years of developing the first primary CRC (Lynch et al., 2006). It is also known that progression from adenoma to carcinoma is accelerated in HNPCC patients. An adenoma may develop into a carcinoma within 2 years of development whereas this process usually takes about 8 years in sporadic CRC (Lynch et al., 2006). This is a feature of HNPCC that makes identification and frequent screening of carriers very important.

The tumours in HNPCC patients tend to be poorly differentiated and contain an excess of mucoid and signet-ring cells. The tumours also display an excess of infiltrating lymphocytes among the cells which has been linked to the improved prognosis in HNPCC due to the lymphocytes suspected of having cytotoxic effects (Lynch et al., 2006; Boland et al., 2008; Boland and Goel, 2010).

MSI is an indicator that the MMR genes are non-functional; a feature that can be used in screening (described in Section 1.6.1). A tumour that is deficient of its MMR enzymes exhibit MSI as they are unable to correct the errors in the microsatellites. Since the errors are not corrected, the mutations accumulate within the cell. It is estimated that approximately 12% to 17% of all CRCs are deficient in the MMR proteins (Boland and Goel, 2010; Cunningham et al., 2010).

Interestingly, MSI in CRC has been found to be a good prognostic factor and molecular marker. Tumours that have a high frequency of MSI maintain their diploid state but also have certain other features. These features include a more proximal colon localization, poor differentiation, contain tumour infiltrating lymphocytes, are less likely to invade and are less likely to contain mutations in *KRAS* or *p53* (Popat et al., 2005; Boland and Goel, 2010).

1.3.2.2. Variations of HNPCC

There have been variations of HNPCC reported in literature. These variations are related to HNPCC as the MMR genes are non-functional in patients, but they display characteristics that are not typically observed in HNPCC families. Muir-Torre syndrome is characterised by sebaceous gland tumours and individuals that present with these skin lesions should receive screening for MSI or Immunohistochemistry (Ponti and de Leon, 2005, Lynch et al., 2006, South et al., 2008). Turcot syndrome, a rare variant of HNPCC, is characterised by the occurrence of a primary malignancy in the central nervous system as well as a primary CRC (Hamilton et al., 1995, Lebrun et al., 2007).

1.3.2.3. Diagnostic Guidelines for HNPCC

Since HNPCC has specific features, it is possible to identify the disease and differentiate it from other cancers such as sporadic cancers and other hereditary cancer syndromes. The Amsterdam criteria and the Bethesda criteria, as well as the revisions for both criteria, were developed to aid both research and the diagnosis of families with HNPCC and their subsequent management.

1.3.2.3.1. Amsterdam Criteria

Prior to 1990, there were no formal guidelines used to identify HNPCC. This was a hindrance as it prevented many groups from performing uniform research on HNPCC. The International Collaborative Group on HNPCC (ICG-HNPCC) held a meeting in Amsterdam in 1990 and identified the need to perform collaborative research studies on uniformly selected cohorts and address various issues related to HNPCC. Minimum guidelines which could be used to identify families in a clinical setting were established, known as the Amsterdam criteria (seen in Figure 1.2) (Vasen et al., 1991; Vasen et al., 1999). This was particularly important as the genetic basis of HNPCC had not been established yet (Umar et al., 2004a).

The establishment of the Amsterdam criteria had a significant effect as many researchers accepted and used the criteria to identify families with HNPCC. The Amsterdam criteria were instrumental in establishing the association between HNPCC and defects in the MMR genes (Vasen et al., 1999; Umar et al., 2004a). However, it was soon realized that the Amsterdam criteria were designed for specificity, and not sensitivity, as many families did not fulfil the Amsterdam criteria since it did not account for extra colonic tumours (Vasen et al., 1999; Umar et al., 2004a).

At least three relatives should have histologically verified CRC
One of the three relatives should be a first degree relative of the two
Familial Adenomatous Polyposis should have been excluded
At least two successive generations should be affected
At least one of the relatives should be diagnosed prior to the age of 50

Figure 1.2: Amsterdam Criteria for the Clinical Diagnosis of HNPCC (Vasen et al., 1991)

1.3.2.3.2. Amsterdam II Criteria

ICG-HNPCC agreed to broaden the criteria to include extra colonic tumours but they also realised the need to not move too far away from the original criteria as many clinicians had adopted them. It was felt that the criteria needed to remain clinical as many families did not have access to DNA testing or mutation analysis (Vasen et al., 1999).

The ICG-HNPCC decided that it would include the more 'common' extra colonic cancers from all populations. The group wanted to ensure specificity of the criteria and felt that including tumours that were common in only one population could lower the specificity of the criteria (Vasen et al., 1999). In 1999, the Amsterdam II criteria were finalised and accepted by the group. The Amsterdam II criteria, as well as the HNPCC associated cancers, are presented in Figure 1.3.

There should be at least three relatives with an HNPCC associated cancer*
One should be a first degree relative of the other two
At least two successive generations should be affected
At least one family member should be diagnosed before the age of 50
Familial Adenomatous Polyposis should be excluded in the CRC cases
Tumours should be verified by pathological examination

***HNPCC associated cancers: Endometrial, ovarian, gastric, renal pelvis/ureter, brain, hepatobiliary tract, skin and small bowel cancers**

Figure 1.3: Amsterdam II Criteria used for the Clinical Diagnosis of HNPCC (Vasen et al., 1999)

The Amsterdam II criteria allowed for an expanded, yet simple, nomenclature that could be used to identify families to be included in studies across the globe (Vasen et al., 1999). With the implementation of the Amsterdam II criteria, it was possible to identify and characterise many more families with HNPCC. Additionally, all Amsterdam I positive families will still be positive under the

Amsterdam II criteria. The Amsterdam II criteria are particularly important for this study as any patient that fulfilled the Amsterdam II criteria was included in this study.

1.3.2.3.3. Bethesda Guidelines

A genetic diagnosis for HNPCC became necessary after the association between HNPCC and the MMR genes had been established (Umar et al., 2004a; Umar et al., 2004b). The National Cancer Institute (NCI) held a workshop in Bethesda in 1996 to discuss the role of genetics in the pathology of HNPCC. At the conclusion of the workshop, the NCI had developed criteria, known as the Bethesda guidelines, which could be used to identify which individuals should be sent for genetic testing for HNPCC. The Bethesda guidelines, included aspects of the Amsterdam criteria as well as pathological evaluations (Rodriguez-Bigas et al., 1997; Umar et al., 2004a).

The Bethesda guidelines, which can be seen in Figure 1.4, are of particular significance as they recommend that individuals who meet any one of the seven criteria should be considered for genetic testing. The Amsterdam criteria require that all the criteria are met to make a diagnosis of HNPCC (Rodriguez-Bigas et al., 1997; Umar et al., 2004a).

Individuals with cancer in families that meet the Amsterdam criteria
Individuals with two HNPCC-related cancers, including synchronous and meta-chronous CRCs
Individuals with CRC and a first degree relative with CRC or an HNPCC related tumour
Individuals with CRC or endometrial cancer younger than 45
Individuals with right-sided CRC and undifferentiated pattern diagnosed <45 years
Individuals with signet-ring-cell type CRC diagnosed under the age of 45
Individuals with adenomas diagnosed younger than 40 years of age

Figure 1.4: Bethesda Guidelines for Genetic Testing of MMR Genes (Rodriguez-Bigas et al., 1997)

Research on HNPCC had shown that about 90% of patients that fulfilled the Amsterdam criteria also showed MSI. Thus, the group felt that microsatellite testing would be an appropriate genetic screening test and then mutation detection in the MMR genes could be performed to find the pathogenic mutations. Identifying pathogenic mutations will allow other family members to be screened for the mutation, improving management and detection of CRC (Rodriguez-Bigas et al., 1997; Umar et al., 2004a).

1.3.2.3.4. Revised Bethesda Criteria

In 2003, the NCI met again with findings of additional research that was performed so that they could update the Bethesda Guidelines. It was found that the Bethesda criteria were very sensitive but not very specific when it came to determining patients with MSI in their tumours (Terdiman et al., 2001). It was also found that family history was important in the diagnosis of HNPCC, even when the diagnosis is made at a young age. Studies indicated that a poor family history may be due to non-paternity, adoption, new or recurrent mutations, denial of a family history and even small family size (Terdiman et al., 2001; Umar et al., 2004a; Umar et al., 2004b). The NCI updated the Bethesda guidelines to account for some of the research findings, giving rise to the revised Bethesda guidelines, which can be seen in Table 1.6 (Umar et al., 2004b).

The revised Bethesda guidelines place less emphasis on histology but they do keep the family history criteria. Genetic testing allows high risk individuals to be identified and screened regularly for disease (Umar et al., 2004a; Umar et al., 2004b; Vasen et al., 2007). For this study, the revised Bethesda criteria were used to identify patients who developed CRC under the age of 50 years. The selection criteria are discussed in more detail in Section 2.1.

CRC diagnosed in a person younger than 50 years of age
Presence of synchronous or meta-chronous CRC and HNPCC-related tumours
CRC with MSI-H histology younger than 60 years of age
CRC in one or more first degree relatives with an HNPCC related tumour, with one of the cancers being diagnosed under 50 years of age
CRC diagnosed in two or more first or second degree relative with HNPCC related tumours, regardless of age

Figure 1.5: Revised Bethesda Guidelines for Genetic Testing of MMR Genes (Umar et al., 2004b)

Both the Amsterdam II criteria and the Bethesda guidelines are useful for detecting HNPCC but both do yield false positives and false negatives since they are both imperfect criteria (Lynch et al., 2006; Vasen et al., 2007). Screening of the MMR genes is used to confirm the diagnosis of HNPCC in individuals that fulfil either the Amsterdam II criteria or the Bethesda guidelines. Once a mutation in a family is found, targeted screening of other family members can be offered.

1.3.3. Nomenclature: HNPCC or Lynch Syndrome?

The name HNPCC was made a standard in 1989 at a meeting in Amsterdam in order to allow doctors to identify the disease more readily. By 2004, the NCI proposed that “Lynch Syndrome” should be used, as it was before the standardization to HNPCC. It was felt that HNPCC was inappropriate as it excluded the extra-colonic tumours observed in patients (Vasen et al., 1999; Vasen et al., 2007). It was then proposed that Lynch Syndrome should be used for patients who had an identified MMR mutation or had high microsatellite instability (MSI-H). A further proposal was that families fulfilling

the Amsterdam criteria but who lacked evidence of defective MMR should be referred to as having familial CRC (Vasen et al., 2007).

For the current study, patients who fulfil the revised Bethesda criteria will be referred to as having suspected HNPCC. Patients who fulfil the Amsterdam II criteria are considered to have a clinical diagnosis of HNPCC and shall be referred to as such. Since Lynch Syndrome is the term given to patients with an identified mutation or MSI-H, it is not appropriate to use this term except in patients with known MMR mutations.

1.4.Mismatch Repair Genes

HNPCC is known to be caused by a loss of function in one of the MMR genes, also known as the 'mutator phenotype'. There are usually one of four genes that are commonly mutated in HNPCC and they are *MutL homolog 1* (*MLH1), MutS homolog 2 (MSH2), MutS homolog 6 (MSH6)* or *post-meiotic segregation increased 2* (*PMS2)*. Of these four genes, *MLH1* and *MSH2* are the most commonly mutated genes in HNPCC, accounting for about 40% and 50% of all HNPCC cases, respectively (de la Chapelle, 2004; Peltomäki, 2005; Lynch et al.,2006; Jasperson et al., 2010). These two genes are the focus of this study because they are the most commonly mutated genes. A summary of the contribution towards HNPCC from each MMR gene can be seen in Table 1.2.

Table 1.2: Percentage of HNPCC cases caused by mutations in each MMR gene (Peltomäki, 2005)

Gene	Percentage of HNPCC cases
MLH1	**40%**
MSH2	**50%**
MSH6	**~10%**
PMS2	**<1%**

1.4.1. *MLH1*

The *MLH1* locus had been known to be involved in HNPCC for many years through linkage studies and it is a highly penetrant gene (Nyström-Lahti et al., 1994, Kolodner et al., 1995). *MLH1* is located at chromosome 3p21 and contains 19 exons that span over 58 kilobases, with the MLH1 protein containing 756 amino acids. Interestingly, each of the major genes in HNPCC has been found to be associated with specific tumours. Families with *MLH1* mutations tend to have a higher incidence of CRC than families with mutations in the other genes (Kolodner et al., 1995; Peltomäki, 2005; da Silva et al., 2009)

Characterization of *MLH1* was an important step in understanding HNPCC as it gave further insights into the aetiology of HNPCC and allowed for enhanced mutation detection. After characterizing

MLH1, Kolodner et al. (1995) performed mutation screening in a large family who were shown to have possible *MLH1* mutations by linkage analysis. Using direct DNA sequencing of the gene, it was found that an insertion led to a frame shift mutation which produced a truncated MLH1 protein and it was this mutation that was responsible for HNPCC in that family.

MLH1 has also been found to be epigenetically silenced in cases of HNPCC. Gazzoli et al. (2002) reported a case of HNPCC where methylation of one allele was present in the blood, but both alleles were methylated in the tumour. The pathogenesis suggested a constitutional *MLH1* epimutation. Hesson et al. (2010) indicate that many cases with a similar pathogenesis have been reported but that the epimutation may not be present in the germ line. However, those individuals are predisposed to HNPCC by way of methylation of the *MLH1* locus in somatic tissue, proposed to be caused by *cis*-acting elements in the same region as *MLH1*.

1.4.2. *MSH2*

The *MSH2* gene is located at chromosome 2p16 and contains 16 exons spanning over 73kb. The MSH2 protein has been found to contain 934 amino acids (Kolodner et al., 1994). Similarly to *MLH1*, *MSH2* is highly penetrant, but mutations in *MSH2* have been found to be more associated with extra-colonic tumours and Muir-Torre features (Kolodner et al., 1994; Peltomäki, 2005; Lynch et al., 2006; da Silva et al., 2009).

Germ line mutations in *MSH2* tend to lead to abolition of the protein. These mutations include premature stop codons, altered splice sites or rearrangements (Boland et al., 2008). However, epimutations have also been discovered. Terminal mutations in the *EPCAM* gene, located upstream of *MSH2*, lead to an abnormal EPCAM-MSH2 fusion transcript and *MSH2* promoter methylation (Hesson et al., 2010). This EPCAM mutation leads to reduced expression of the *MSH2* gene. This *EPCAM* mutation confers the same risk of tumour development as *MSH2* mutations in familial cancer (Hesson et al., 2010).

1.4.3. *MSH6*

The *MSH6* gene is located at chromosome 2p15 and contains 10 exons and is approximately 20kb in size. The MSH6 protein, also known as GTBP, is a 153kDa protein consisting of 1360 amino acids (Acharya et al., 1996). The mean age of onset of CRC in families with *MSH6* mutations is 56 years, as opposed to 45 years in classical HNPCC. This attenuated phenotype is the result of compensation provided by MutS homolog 3 (MSH3) in the cell (Acharya et al., 1996; da Silva et al., 2009; Talseth-Palmer et al., 2010). In addition to later age of onset, women in families with *MSH6* mutations tend to develop endometrial cancers more frequently, with a 71% lifetime risk by age 70 (Peltomäki, 2005; da Silva et al., 2009; Talseth-Palmer et al., 2010).

Although *MSH6* is responsible for roughly 10% of cases, it is estimated that many cases involving *MSH6* are overlooked. One reason for under-representation in these families is due to lack of routine testing of *MSH6* in many laboratories. Another reason for under-representation is that families with *MSH6* mutations have an attenuated phenotype that does not fulfil the Amsterdam criteria, or the Bethesda guidelines. There are reports that there is a high incidence of *MSH6* mutation carriers in Amsterdam negative families (Boland et al., 2008; da Silva et al., 2009; Sjursen et al., 2010; Talseth-Palmer et al., 2010).

1.4.4. *PMS2*

The *PMS2* locus at chromosome 7p22 contains 15 exons and is 16kb in size (Nicolaides et al., 1994). *PMS2* has been found to be mutated in families with HNPCC but is reported in less than 1% of all HNPCC cases. *PMS2* is however, also thought to be under-represented due to the attenuated phenotype of older age of onset and weak family history (Peltomäki, 2005; Lynch et al., 2006; Boland et al., 2008; da Silva et al., 2009).

Paradoxically, *PMS2* mutations are also thought to lead to childhood onset of tumours when two non-functional alleles are inherited. One challenge facing *PMS2* management is the limited number of observed cases, which is one reason why there is no clear phenotype described for this gene. It has also been suggested that PMS2 is associated with variable (Peltomäki, 2005; Lynch et al., 2006; Boland et al., 2008; da Silva et al., 2009).

1.4.5. Recurrent and Founder Mutations

These above four MMR genes, *MLH1, MSH2, MSH6* and *PMS2*, are the main genes involved in the aetiology of HNPCC. There is a wide range of mutations found in HNPCC families. However, it has been found that occasionally the same mutations occur in families that are unrelated. There are two explanations for these occurrences. The first explanation is that these mutations are recurrent mutations that occur within gene 'hot spots'. These hot-spots are prone to mutations, which often appear *de novo* in individuals (de la Chapelle, 2004; Lynch et al., 2006).

The second explanation is founder mutations, which are usually introduced into a population by one individual and retained through genetic drift. There are several factors that determine if a mutation will persist in a population, including rapid population growth, population bottlenecks, isolation and chance. Founder mutations can be observed in isolated populations that have grown from a small group of individuals (de la Chapelle and Wright, 1998; de la Chapelle, 2004; Strachan and Read, 2004).

Founder mutations have significant impact for individuals and molecular genetic diagnosis (Lynch et al., 2006). As a first step in molecular diagnosis, those individuals from that population can be tested

for the founder mutation as it provides a first line target for mutation detection, rather than whole gene sequencing. Targeted mutation analysis will save costs and facilitate better management (Lynch et al., 2006).

Although there are many mutations reported for each of the MMR genes, there have been found to be several founder mutations in various populations. A summary of some of the known founder mutations can be seen in Table 1.3 below. A founder mutation has been detected in South African mixed ancestry individuals (Goldberg et al., 1998; Stupart et al., 2009).

Table 1.3: Founder mutations of *MLH1* and *MSH2* found in different populations

Population	Gene	Mutation	References
South African Mixed Ancestry	*MLH1*	c.1528C>T in Exon 13	Goldberg et al., 1998
Ashkenazi Jews	*MSH2*	c.1906G>C	Foulkes et al., 2002
Chinese	*MSH2*	c.1452-1455delAATG	Chan et al., 2004
Finnish	*MLH1*	Deletion of Exon 16	Nyström-Lahti et al., 1995
Italians	*MSH2*	Deletion of Exons 1-6	Stella et al., 2007
USA Caucasians (American Founder Mutation*)	*MSH2*	Deletion of Exons 1-6	Clendenning et al., 2008

*This mutation was traced back to a German that settled in the USA in the 18th century (Clendenning et al., 2008)

1.4.6. Functions of the MMR Genes

The MMR gene products do not function independently of one another. Instead, they interact together in a complex pathway which is responsible for correcting errors and even play a role in apoptosis if damage is not repaired. The following section will describe how these gene products interact within the cell.

The MMR system is responsible for detecting and correcting DNA replication errors that occur on the newly synthesized strand of DNA. As mentioned above, DNA polymerase occasionally makes a mistake when synthesising DNA and inserts an incorrect nucleotide. This can lead to a base mismatch or an insertion-deletion loop (Boland and Goel, 2010).

Within the cell, the predominant heterodimer for detecting mismatches on the DNA strand is known as MutSα. The heterodimer MutSα is made up of MSH2 and MSH6 proteins. MutSα is able to bind to base/base mismatches as well as 1bp insertion-deletion loops and repair them (Peltomäki, 2005).

A second mismatch recognition protein known as MutSβ consists of MSH2 and MSH3. The MutSβ complex functions to effectively eliminate 2-8bp unpaired nucleotides, which form an insertion-deletion loop that the complex recognizes. The MutSα and MutSβ vary in their affinities due to the

MSH6 or MSH3 protein bound to MSH2, with the MSH3 complex being able to bind to larger insertion-deletions loops than the MSH6 complex (Peltomäki, 2005; Boland and Goel, 2010).

The functions of the MutL complexes are not as well-known or understood as the MutS complexes. The MutLα complex, which consists of MLH1 and PMS2, is also vital to the functioning of the MMR system. This complex is considered important in bridging mismatch recognition with downstream repair processes, such as long-patch excision (Peltomäki, 2005; Boland and Goel, 2010).

The MLH1 protein is also able to dimerize with PMS1, forming the MutLβ complex. This complex has been shown to suppress mutagenesis in Saccharomyces cerevisiae (yeast), but the function of this complex has not been fully elucidated in humans (Boland and Goel, 2010). Similarly, the MutLγ complex, composed of MLH1 and MLH3, suppresses insertion-deletion loops in yeast but has an unknown function in humans (Boland and Goel, 2010). A diagram of these protein interactions is given in Figure 1.1.

1.4.7. Mechanism of Tumourigenesis in HNPCC

If the MMR proteins do not identify and correct mismatches, the DNA synthesized subsequently will contain mutations. When MMR activity is eliminated, mutations occur at a higher rate as the cells proliferate. Point mutations that arise by single base mismatches could lead to changes in protein structure or function, premature stop codons or complete silencing of genes. The mutation of genes involved in cell proliferation can lead to the elimination of entire regulatory systems (Chao and Lipkin, 2006; Boland et al., 2008; Boland and Goel, 2010).

1.4.7.1. Coding Microsatellites

Coding microsatellites are microsatellite repeats that are located within exons. *MSH2*, *MSH6* and *PMS2* each contain mononucleotide repeats with at least seven adenines, which may be prone to insertions and deletions (Boland and Goel, 2010). It has been found that many genes involved in cell growth (including *TGFBR2*, *BAX* and *CASP-5*), contain coding microsatellites that are frequently mutated in cells with MSI (Chao and Lipkin, 2006, Boland et al., 2008).

The above indicates that the MMR system is responsible for maintaining the functions of other genes by preventing frame-shift mutations in coding microsatellites. Additionally, the MMR system may become more defective as more repair components are abolished. The MMR components are also thought to be important in signalling apoptosis when DNA damage cannot be repaired (Chao and Lipkin, 2006; O'Brien and Brown, 2006; Boland and Goel, 2010).

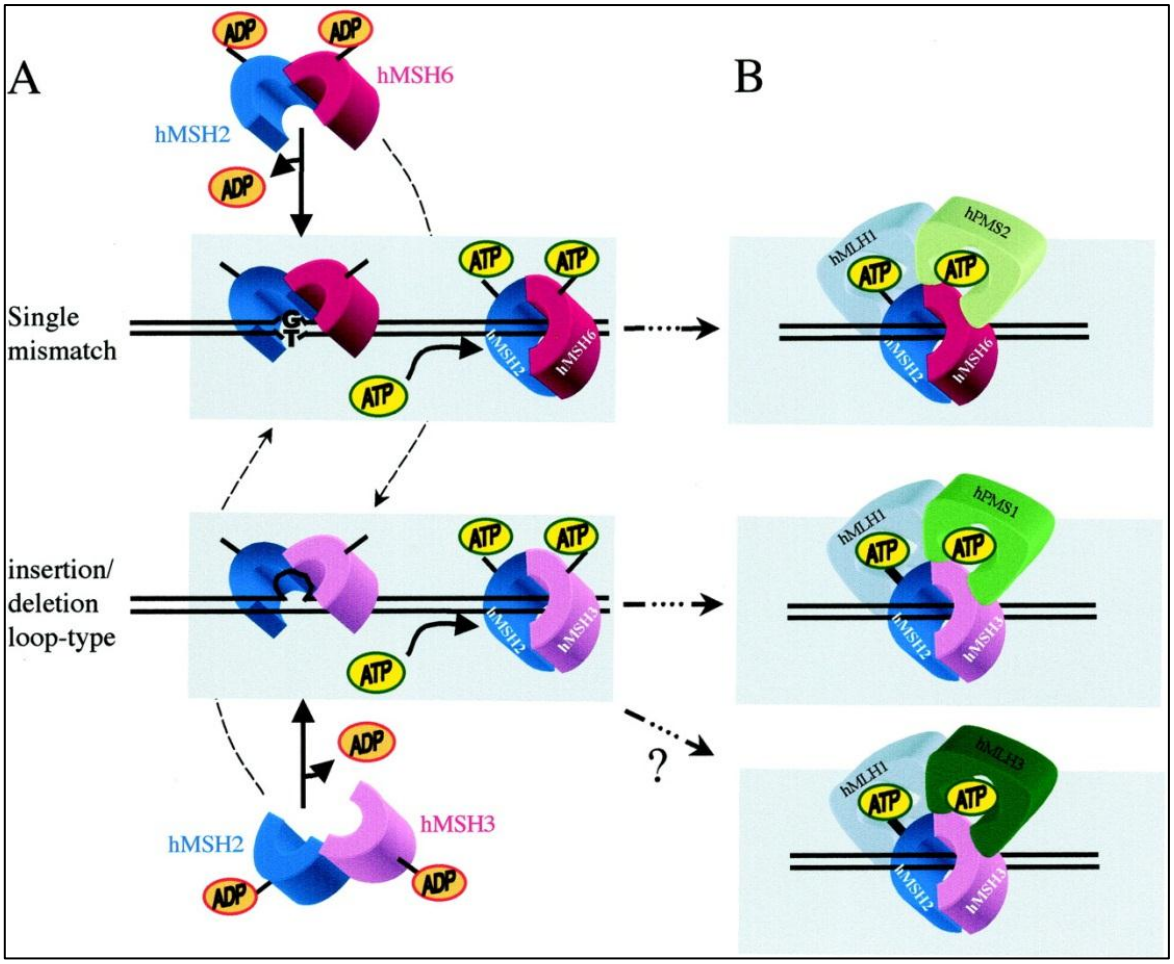


Figure 1.6: Diagram showing how the MMR proteins interact to repair mismatches in the DNA (Fishel, 2001). The MutSα heterodimer (hMSH2 and hMSH6) binds to the DNA and detects single base mismatches. The MutLα complex (hMLH1 and hPMS2) binds to the MutSα complex. The MutSβ heterodimer (hMSH2 and hMSH3) binds to insertion/ deletion loops in the DNA. The MutLβ complex binds to the MutSβ complex. In this figure, the 'h' in the protein name indicates a human protein.

1.4.7.2. Apoptosis

This mechanism for apoptosis is not clear but it is thought that the MMR proteins may use direct signalling by recruiting additional proteins to the cell. Alternatively, indirect signalling through excessive double strand breaks could lead to toxicity, and subsequently apoptosis. An important step for tumourigenesis is the abolition of apoptosis pathways. The *BAX* gene has an important role in signalling apoptosis and is commonly abolished in MSI tumours, which leads to a growth advantage for a cell (Chao and Lipkin, 2006; O'Brien and Brown, 2006).

1.4.8. Tissue Specificity of HNPCC

HNPCC is characterised by tumours that develop predominantly in the colon, but that there are a number of extra-colonic tumours that are linked to MMR deficiency. This raises an important question, why does defective MMR predominantly lead to tumours in the colon? Models have been proposed to understand the tissue specificity involved in HNPCC.

One mechanism by which the tumours develop in the colorectal tract regards cell proliferation rates. The epithelial cells in the gastrointestinal tract have the highest proliferation rate of all cells in the body, being replaced every three to five days. Cells that proliferate at a higher rate are more susceptible to mutations when the MMR genes are mutated. This one factor can provide a very plausible reason for CRC development in HNPCC but does not account for cancers in slower proliferating cells like those of the urinary tract or no tumours in tissues such as bone marrow or respiratory tract (Chao and Lipkin, 2006).

Environmental factors such as tobacco smoking, diet and pharmaceuticals are considered to play important roles in carcinogenesis. Diets high in fat and low in folate and calcium ('Western diets') are associated with an increased incidence of CRC. These diets are also associated with increased Body Mass Indexes (BMIs), which may be an important factor in colorectal tumour development (Watson et al., 2004; Chao and Lipkin, 2006; Botma et al., 2010).

Tumour development in the endometrium is not well understood. The endometrium undergoes monthly cycling in response to hormones, which starts with rapid proliferation of cells and ends with menstruation. After the phase of rapid proliferation, apoptosis occurs in the newly synthesised cells. The loss of apoptosis could lead to the cells not undergoing the regular cycling and persisting. It is believed that rapid changes in proliferation rates and loss of apoptosis could be an important factor in endometrial and ovarian cancer (Chao and Lipkin, 2006).

Overall, there are many important factors that play a role in the tissue-specific tumourigenesis observed in HNPCC. These factors include mutations in genes, diet, hormones and exposure to mutagens.

1.5.Ethnic Group Differences in CRC Compared to Caucasians

This section will describe some of the differences in CRC, with regard to incidence and histology, seen among the different ethnic groups compared to Caucasians. This section is given to highlight that different populations show slightly different features for CRC and HNPCC, which may be seen in the South African population. The Asian population will be compared to Caucasians first, then the African Americans and finally the Africans. The tumour spectrum of families with CRC does vary between geographic locations. This is thought to be attributed to differences in risk factors in the different region's each ethnic group inhabits. However, it has been found that an ethnic group that has migrated to a new region will adopt the same risk of CRC as that regions population because they are exposed to the similar risk factors (Anderson et al., 2003).

1.5.1. Asians

For this literature review, Asians with Chinese and Japanese heritages have been compared to Caucasians as there have been a number of studies on these populations. It has been found that Japanese Americans have an increased risk (1.40) of developing CRC relative to Caucasians, even after accounting for various risk factors. The Japanese Americans had the highest rate of positive family history for CRC, which suggests that genetic factors may be important in the observed cases (Ollberding et al., 2011). Further studies indicate that the number of sporadic CRC patients under 50 is increasing year on year in Asian populations (Chew et al., 2009).

When directly observing HNPCC patients of Asian descent, their tumours are generally located in the distal colon. This localization of HNPCC tumours differs from Caucasians, as well as Africans, who present primarily with tumours in the proximal colon (Wei et al., 2011). With regard to the extra-colonic tumours, stomach cancer is more predominant in Asian HNPCC families compared to Caucasian HNPCC families (Wei et al., 2011). Asian families with HNPCC also show an increased risk of developing endometrial cancer relative to Caucasian patients (Ollberding et al., 2011).

1.5.2. African Americans

It has been reported that there is a higher lifetime risk of CRC and that survival among African Americans with CRC is lower than in Caucasians. The higher incidence rate of CRC may be attributed to various environmental risk factors, including diet and screening rates (Agrawal et al., 2005). The 5 year survival rate in African Americans (53%) is lower than the 5 year survival rate of Caucasians (63%) for all stages of cancer, probably due to a lack of early screening (Anderson et al., 2003; Agrawal et al., 2005). These numbers give more support to early screening and detection of CRC.

The proportion of African Americans who present with CRC under the age of fifty (10.6%) is higher than the proportion of Caucasians (5.5%) that present under the age of 50 (Agrawal et al., 2005). African Americans also show a higher degree of proximal tumours but the rate of tumourigenesis is not as accelerated compared to Caucasians (Anderson et al., 2003; Agrawal et al., 2005).

African American kindred's with HNPCC have been reported. The average age of diagnosis is lower than the population average in HNPCC families (Agrawal et al., 2005). Agrawal et al. (2005) also highlight another study on African Americans which found that 43% of patients had MSI-H tumours and that 77% of these tumours were located in the proximal colon.

1.5.3. Africans

It is indicated that there is a higher prevalence of MSI in black African patients (10-30%) than in Caucasian patients (5-15%) with regard to sporadic CRC (Anderson et al., 2003). There is speculation that the increased prevalence of MSI CRC is the result of germ line mutations in the MMR genes.

HNPCC has been detected in African populations, with unique germ line mutations accounting for several cases (Anderson et al., 2003). Unfortunately, there is very little literature covering CRC in Africa and there is very little information about HNPCC in African populations (Anderson et al., 2003; Cronjé et al., 2009; Hitchins et al., 2011).

1.5.3.1. South African Black Population

In South Africa it is estimated that one in four men and one in five women, will develop cancer during their lifetime (National Cancer Registry (NCR), 2004). The latest cancer statistics in South Africa show that CRC is the sixth most common cancer (NCR, 2011). Interestingly, the Age Standardised Rate (ASR) figures indicate that CRC is more common in Caucasians under the age of 50 than in the black and mixed ancestry population (NCR, 2011). More information on Cancer rates in South Africa can be seen in Appendix A under Figure A.1 as well as Tables A.3 and A.4. The incidence rate for CRC is predicted to increase every year in South Africa, partly due to the spread of the Western diet (Anderson et al., 2003; NCR, 2004). Interestingly, Wentink et al. (2010) found that CRC was rare in rural parts of South Africa, where the Western diet may not have penetrated.

In South Africans, the black patients exhibit similar phenotypes to the African American patients, namely lower age of diagnosis and more proximal localisation of CRC (Anderson et al., 2003; Cronjé et al., 2009; Wentink et al., 2010). It is indicated that about 26-41% of black patients in South Africa who presented with CRC were under the age of 50, which is a high proportion of patients that fulfil the Bethesda criteria (Cronjé et al., 2009; Wentink et al., 2010). Studies have found that younger patients show a tendency to develop proximal CRC with deficient MLH1, MSH2 or MSH6 proteins. They concluded that HNPCC could be a strong underlying factor for CRC in young black patients seen at local hospitals and that genetic testing should be offered (Cronjé et al., 2009).

1.6. Genetic Testing for HNPCC

Testing for MSI and loss of MMR proteins expression can be used to screen colorectal patients for HNPCC. However, the best strategy for identification of HNPCC is to screen for germ line mutations using complete sequencing and Multiplex Ligation-dependent Probe Amplification (MLPA) of the MMR genes (Boland and Goel, 2010). Once Lynch Syndrome is confirmed, the family members can be tested using targeted sequencing to determine if individuals are at an increased risk of developing cancer. This section will discuss the molecular genetic testing strategies for screening and diagnostic identification of HNPCC.

1.6.1. MSI Testing

MSI occurs in approximately 12-17% of all CRC's. Cases of sporadic CRC with MSI tend to be in older patients who, additionally, do not show familial clustering of CRC. MSI in sporadic CRC is often the result of biallelic methylation of MMR genes, typically *MLH1* (Boland and Goel, 2010). However, MSI is present in more than 90% of cases of HNPCC due to the deficiency of the MMR proteins, as described in Section 1.4.7.1 (Boland et al., 2008; Boland and Goel, 2010).

MSI testing allows one to determine whether a microsatellite has expanded and can be performed simply and cheaply using PCR techniques (Boland and Goel, 2010). The National Cancer Institute (NCI) has endorsed a panel of five microsatellite markers which include two mononucleotide microsatellites (BAT25 and BAT26) and three dinucleotide microsatellites (D2S123, D5S346 and D17S250) (Vasen et al., 2007; Pineda et al., 2009; Boland and Goel, 2010). The tumour DNA is compared to the blood DNA to determine if the microsatellites show expansions or deletions. If there are no differences among any microsatellites, then the tumour is microsatellite stable. If one marker shows a change in size, then the tumour has low MSI (MSI-L). If two or more markers are unstable, then the tumour is said to have high MSI (MSI-H) (Pineda et al., 2009; Boland and Goel, 2010).

MSI testing has been shown to be very sensitive for predicting the presence of MMR mutations. This testing can be very useful in families that have a high probability of having HNPCC, even when protein expression seems normal or in individuals who meet the Bethesda criteria to determine whether they should proceed to genetic analysis. The protein could be detected using IHC, but it may be non-functional due to missense mutations. MSI testing can be cost effective as it determines whether a tumour is caused by MMR deficiency before sequencing the MMR genes (Vasen et al., 2007).

MSI testing also has disadvantages for testing. MSI testing requires a high proportion of tumour tissue in order to get good results. MSI testing has also been regarded as not being useful for *MSH6* mutation carriers due to the lower level of MSI seen in their tumours. Furthermore, MSI may be acquired through somatic mutations in sporadic tumours, which would yield a false positive for inherited MMR gene testing (Pineda et al., 2010).

1.6.2. Immunohistochemistry

IHC is the process of using antibodies, prepared against specific proteins, to determine if the protein is present in tissue. This process is achieved by labelling an antibody, then allowing it to bind to proteins in tissue. If the protein is not present, the antibodies cannot bind and are washed away (Strachan and Read, 2004). IHC has been an important histological tool for the molecular diagnosis of HNPCC.

When using IHC for the detection of loss of MMR, antibodies against MLH1 and MSH2 are usually used. However, to get a more complete result, antibodies against MSH6 and PMS2 should also be used (Vasen et al., 2007). Since MSH2 and MLH1 are obligatory components in the MMR system, absence of these proteins leads to instability of their protein partners. If the MLH1 protein is absent, the IHC shows the loss of staining for both MHL1 and PMS2 and similarly, if MSH2 is absent, for MSH2 and MSH6 (Peltomäki, 2005; Boland et al., 2008).

IHC has been shown to be very sensitive for detecting MMR expression but there is a higher rate of false negatives than with MSI testing. These false negatives are the result of detecting non-functional proteins or protein fragments. However, IHC allows one to detect which proteins are abolished, which can be used to direct mutation testing. This has been thought to be particularly useful in *MSH6* mutation carriers (Lynch et al., 2006; Vasen et al., 2007). IHC has a lower cost and is also widely available in many pathology laboratories (Pineda et al., 2010; Prof. A. Krause, University of the Witwatersrand, Johannesburg, Personal communication, 2011). Although there are laboratories that provide IHC testing in South Africa, the testing is not cheaply or easily available. IHC would have been an ideal inclusion in this study. The IHC status was only known in the positive control in this study (described in Section 2.1.1)

IHC also has several limitations which can affect its usefulness. These limitations include a lower sensitivity for detecting MLH1 expression. IHC does not distinguish between germ line mutations or hyper-methylation as a mechanism for loss of expression. It can also be challenging to interpret the results from IHC staining, especially if there are no internal controls (Lynch et al., 2006; Pineda et al., 2010).

Both MSI testing and IHC have strengths and limitations but both have been shown to be very useful for detecting loss of MMR function in a tumour. Thus either test can be offered to patients as a first step to determine MMR status. However, there seems to be consensus that IHC is the more beneficial as it can direct mutation testing. Ultimately, the next step should be mutation detection using sequencing and MLPA if MSI or IHC is positive (Lynch et al., 2006; Vasen et al., 2007; Boland and Goel, 2010).

1.6.3. Direct Sequencing and MLPA

Individuals that fulfil the Bethesda criteria can be screened using MSI testing or IHC but the ideal course of action is to sequence their DNA and determine if they harbour any germ line mutations in the MMR genes. Sequencing the MMR genes is the gold standard method for mutation detection and can be complemented by MLPA analysis (Boland and Goel, 2010).

Sequencing of the exons of these genes is most useful when using primers that include a portion of the intron/exon boundary. This allows the entire exon to be sequenced as well as the splice sites, where mutations are known to occur (Boland and Goel, 2010). However, some pathogenic mutations are small and may be missed by sequencing.

MLPA allows for the detections of large mutations that include the deletion or duplication of whole exons. If an exon is deleted and the breakpoint occurs outside of the sequencing primer binding site, then the sequencing analysis will not show any mutation as only the wildtype allele is sequenced and

not the mutant allele. MLPA allows for the quantification of each exon, thus determining if an exon is missing or duplicated. Large deletions are not rare in HNPCC, with more than a third of mutations in *MSH2* believed to be large exon deletions (Schouten et al., 2002; Boland and Goel, 2010). Furthermore, the mutations detected by MLPA are almost certainly pathogenic as they affect large portions of coding DNA (Lynch et al., 2006).

1.6.4. Improved Clinical Management

The most important reason for finding the pathogenic germ line mutation is that it can aid family members with clinical management. Morbidity and mortality of individuals can be lowered by lowering their risk factors and providing adequate surveillance for affected members. If one family member is found to harbour a pathogenic mutation, family members can have the same region tested for the presence of that mutation. Furthermore, members who are negative for mutations can avoid regular colon tumour screening (Lynch et al., 2006; Boland and Goel, 2010).

The best process for surveillance of the colon is full colonoscopy, due to the tumours developing primarily in the proximal colon. Colonoscopy has been shown to have lifesaving benefits and can reduce the incidence of CRC by up to 90% (Anderson et al., 2003; Lynch et al., 2006). Unfortunately, this procedure can be uncomfortable as it is invasive. If an individual is found to be a mutation carrier, and thus have an increased risk of cancer, the recommendation is to undergo regular colonoscopy every 1-2 years. Alternatively, prophylactic colectomy can be performed to minimise the risk of developing CRC (Lynch et al., 2006).

In addition to colonoscopy, women who carry a mutation are at an increased risk of endometrial cancer. Screening strategies for endometrial cancer are less well established than those for colon cancer. However, endometrial sampling and transvaginal ultrasound are found to detect endometrial cancers at an early stage, which is vital for effective treatment (Järvinen et al., 2009). Furthermore, prophylactic oophorectomy and hysterectomy can be performed on a carrier once they have had children to prevent the endometrial and ovarian cancers from developing (Järvinen et al., 2009).

1.6.5. Diagnostic Services

Early detection is shown to prevent mortality in cancer and awareness of risk can save lives. Unfortunately, early detection hinges very heavily on good diagnostic facilities to identify at risk individuals. This can be particularly important in South Africa where there are limited resources. Mutation screening can benefit the health system as it could allow resources to be directed to individuals that need them, thus providing a more efficient use of the resources (Goldberg et al., 1998).

In South Africa, the ability to diagnose a mutation carrier within molecular genetic laboratories is very limited. Patients who present with clinical features suggestive of HNPCC do not have a molecular diagnostic facility in South Africa where they can be tested. Patients have to have their samples sent to overseas laboratories for testing, which can cost approximately £500 per gene (£1=R13.18, 2 June 2012). Unfortunately, this cost is one that many South Africans cannot afford (Prof. A. Krause, University of the Witwatersrand, Johannesburg, Personal communication, 2011).

In order to look for disease-causing mutations in HNPCC, HNPCC testing needs to be set up locally. Testing for mutations, particularly recurrent and founder mutations, may considerably improve the molecular diagnosis of HNPCC in the South African population. Once individuals have been diagnosed as having an increased risk, they can be screened regularly, which will allow for early detection and surgery can be offered.

1.7.Aims and Objectives

The aim of this project is to establish the mutation profile for HNPCC in *MLH1* and *MSH2* in the South African population. The South African populations have not been extensively studied and may contain novel variants that have not been reported previously or founder mutations. This study will investigate the exons and intron/exon boundaries of *MLH1* and *MSH2* to determine the molecular aetiology of HNPCC. These two genes have been chosen for this study as they account for approximately 90% of HNPCC mutations in other populations. It is important to determine if mutations are common in a population to provide a cheap, effective diagnostic service. There may be possible founder mutations since there are studies to suggest that there is a higher than expected proportion of young black people with CRC, which may be attributable to HNPCC (Cronjé et al., 2009).

The objectives of the following study were:

1. To collect samples from patients that fulfilled the inclusion criteria for this study. A DNA bank at the NHLS, Braamfontein, was used to identify previously banked samples. Patients were also collected prospectively through local hospitals. Any patient that fulfilled the Amsterdam II criteria was included in this study. Since no black patients were Amsterdam II positive, the age criterion from the revised Bethesda guidelines was also used. These criteria would allow patients to be selected from various ethnic groups within South Africa
2. To perform MLPA of the *MLH1* and *MSH*2 genes in the patients that were selected to determine if there are any large deletions or duplications present.
3. To perform Sanger sequencing on all the exons and intron/exon boundaries of the *MLH1* and *MSH2* genes to determine the presence of any small insertions, deletions or single base changes.
4. To determine the pathogenicity of variants detected in *MLH1* and *MSH2* by bio-informatic and lab-based methodology.
5. To characterise the mutation spectrum in patients and determine if any common pathogenic mutations, or common variants, exist.

2. Subjects, Materials and Methods

This chapter will describe all the subjects, materials and methods that were used during this study. Section 2.1 will describe patient identification and inclusion criteria for the study while Section 2.2 will describe the MLPA method, which was used to determine whether large deletions or duplications were present in *MLH1* or *MSH2*. Section 2.3 will describe how DNA was amplified using the polymerase chain reaction (PCR), which was necessary to perform Sanger sequencing, described in Section 2.4. This section will describe how sequencing of the exons and intron/exon boundaries of the genes *MLH1* and *MSH2* was performed using the PCR products as template DNA.

The analysis of MLPA and Sanger sequencing data is given in Section 2.5. Additionally, this section will describe the methods used to correctly name the variants, and determine reporting history, using the LOVD database at the International Society for Gastrointestinal Hereditary Tumours (InSiGHT; www.InSiGHT-Group.org) and the UCSC Genome Browser and the BLAST-Like Alignment Tool (BLAT). Section 2.6 will describe the methods used to elucidate the frequencies of the novel variants that were found, using restriction fragment length polymorphisms and high resolution melt. A flow chart showing the materials and methods can be seen in Figure 2.1. All the solutions used in this study can be found in Appendix B.

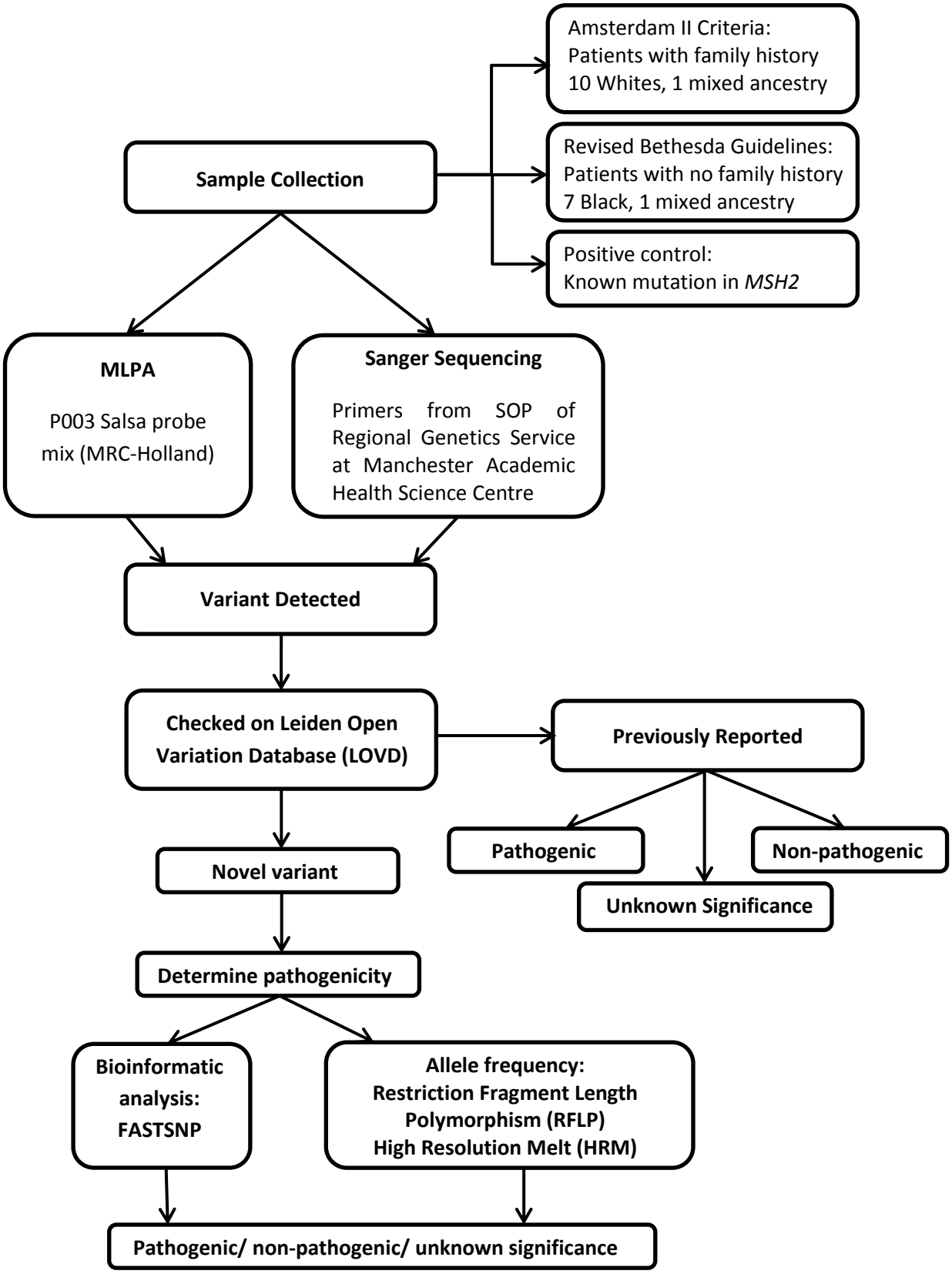


Figure 2.1: Flow Diagram outlining the approach and actions taken in this study

2.1. Subjects

Twenty patient samples were included in this study. The Amsterdam II criteria were used to select patients. However, the black patients available did not fulfil the Amsterdam II criteria, so it was decided to use the revised Bethesda criteria to increase the number of black and mixed ancestry patients in this study. A summary of the patients collected in this study can be seen in Table 2.1. Patient samples were ascertained in two ways:

1) Samples already present in the laboratory from patients who fulfilled the clinical criteria for HNPCC
2) Sample collection through the clinical diagnostic service

2.1.1. Samples Present in the Laboratory

Over the past 15 years, 25 patients who presented with CRC and had a positive family history of CRC were seen through the Genetic Counselling Service, NHLS, Braamfontein. At the time, genetic diagnostic testing was not available in South Africa and the DNA from these patients was stored in the DNA bank at the NHLS, Braamfontein. Both the collection of patient blood, and DNA extraction, were performed by the Clinical Diagnostic Service in the Department of Human Genetics, NHLS and the University of the Witwatersrand. These patients had previously given informed consent for inclusion in further CRC research. Thirteen of these 25 samples were patients that fulfilled inclusion criteria for this study.

Of these 13 samples from the DNA bank, 11 patients fulfilled the Amsterdam II criteria. Ten were white patients and one was a mixed ancestry patient. Due to the higher number of white patients with family history in the DNA bank, it was possible to exclude white patients from the DNA bank who did not fulfil the Amsterdam II criteria. One black patient and one mixed ancestry patient from the DNA bank were included using the revised Bethesda guidelines as they developed CRC under the age of 50.

One of the white patients (150) that fulfilled the Amsterdam II criteria had a known *MSH2* mutation, which was identified by the Wessex Regional Genetics Laboratory in Salisbury. The mutation was identified using Sanger sequencing of the exons and intron/exon boundaries of *MSH2*. Previous IHC results on patient 150 showed a lack of staining for MSH2. Unfortunately, the IHC status of the other patients in this study was not known. Patient 150 was included as a positive control to validate the sequencing method.

2.1.2. Sample Collection through the Clinical Diagnostic Service

Samples were collected at Garden City Clinic, Chris Hani Baragwanath Hospital and Donald Gordon Medical Centre. Patients were identified through the Clinical Diagnostic Service and blood was collected once the patient had been counselled and informed consent for research was attained (Appendix C shows the information and consent forms that the patients signed).

Six black patients were identified through the clinical diagnostic service. These patients were diagnosed with CRC under the age of 50 and were included in the study. One white patient, who had a family history of CRC, was identified through the Clinical Diagnostic Service.

Table 2.1: Summary of the patients and their inclusion criteria

Ethnic Group	Patient Number	Age at Diagnosis	Gender	Location of Tumour	Inclusion Criteria Met
Black	104	49	Male	Unknown (3rd recurrence)	Revised Bethesda Guidelines
Black	205	28	Male	Sigmoid colon	Revised Bethesda Guidelines
Black	235	21	Male	Rectum	Revised Bethesda Guidelines
Black	238	49	Male	Rectum	Revised Bethesda Guidelines
Black	247	47	Male	Rectum (Recurrence)	Revised Bethesda Guidelines
Black	267	47	Female	Rectum	Revised Bethesda Guidelines
Black	290	48	Female	Rectum	Revised Bethesda Guidelines
White	32	50	Male	Sigmoid colon	Amsterdam II Criteria
White	42	64	Female	Unspecified (colectomy)	Amsterdam II Criteria
White	52	37	Male	Upper gastro-intestinal tract	Amsterdam II Criteria
White	61	67	Male	Distal colon	Amsterdam II Criteria
White	81	29	Male	Proximal colon	Amsterdam II Criteria
White	86	40	Female	Transverse colon (recurrence)	Amsterdam II Criteria
White	89	49	Female	Proximal colon	Amsterdam II Criteria
White	136	56	Female	Transverse colon	Amsterdam II Criteria
White	150	44	Male	Sigmoid colon	Amsterdam II Criteria
White	190	50	Male	Duodenum	Amsterdam II Criteria
White	253	28	Female	Rectum	Amsterdam II Criteria
Mixed Ancestry	13	45	Male	Unspecified (Colon)	Amsterdam II Criteria
Mixed Ancestry	175	41	Female	Rectum	Revised Bethesda Guidelines

Average Ages: Black=41.28; White=46.72; Mixed Ancestry=43.00

2.1.3. Summary of Samples Collected for Study

In total, 20 patients were identified for inclusion in this study. Twelve of the 20 patients fulfilled the Amsterdam II criteria (11 white, one mixed ancestry) and eight patients fulfilled the revised Bethesda guidelines (Seven black, one mixed ancestry). One of the Amsterdam II positive patients had a known mutation in *MSH2* and was included as a positive control.

Patient DNA was extracted using the salting out technique. However, this technique was performed by the Molecular Diagnostic Division in the Department of Human Genetics at the NHLS, Braamfontein as they needed to maintain quality control measures on the patient DNA. An aliquot of DNA was then given for this study and the DNA concentration was standardised for PCR and MLPA applications. The DNA concentration is particularly important in MLPA, which is a quantitative

technique. The DNA concentration was measured using a ND-1000 Nanodrop spectrophotometer (Thermo Scientific, USA) and then diluted to 50ng/ul of DNA in ddH_2O.

2.1.4. Control Samples

Sanger sequencing detected seven novel variants in the study cohort (Section 3.2). These variants were investigated further using either RFLP or HRM which were performed on ethnically matched, random population control samples, collected under the "Genetic Diseases studies in Southern African populations: Control samples" (ethics approval: M050706) in the Department of Human Genetics, NHLS, Braamfontein. These samples were collected at the NHLS in Braamfontein. The people that gave these samples consented to their samples being used as population controls and their ethnicity was self-reported. Forty five random white, unlabelled control samples and 45 random black, unlabelled control samples were used. In addition, these control samples were also used in the MLPA reactions.

2.1.5. Ethics Approval

Ethics approval for this study was obtained from the Wits Human Research Ethics Committee (HREC), approval number M10612. The ethics clearance certificate can be seen in Appendix D. As stated previously, informed consent was required from all patients whose samples were included in this study. The samples were tested in the laboratory as anonymised aliquots using only the patient numbers. Due to the nature of the research and the utility of the information to a particular family, the supervisors of this study were able to identify the patients through their corresponding numbers, for the purposes of reporting results back. The results from this study will be given back to the patients or their families by clinicians or genetic counsellors of the Clinical Division.

2.2. Materials and Methods

2.2.1. Multiplex Ligation-dependent Probe Amplification Analysis

MLPA is a relatively simple, high throughput and robust technique developed by the company MRC-Holland. MLPA allows detection of DNA copy number changes of up to 50 probes in a single reaction. The technique is based on the semi-quantitative PCR principle and can be applied to detecting copy number changes (Schouten et al., 2002). In this way it is ideal for looking for large deletion or duplication mutations. For this study, the MLPA technique was used to determine if any exons were deleted or duplicated in *MLH1* or *MSH2*.

The MLPA technique uses a specific probe design to detect DNA copy number changes. Each probe consists of two oligonucleotides, which are ligated together during the MLPA reaction. The first, shorter, oligonucleotide contains a target specific, hybridisation sequence (20-40nt) and a universal sequence which is identical for all probes at the 5' end. The longer oligonucleotide contains a target specific sequence (20-40nt), a stuffer sequence of variable length and the universal sequence at the

3’ end. The universal sequences at the 3’ and 5’ ends allow PCR amplification using only one PCR primer pair (Schouten et al., 2002). The MLPA reaction is shown in Figure 2.2.

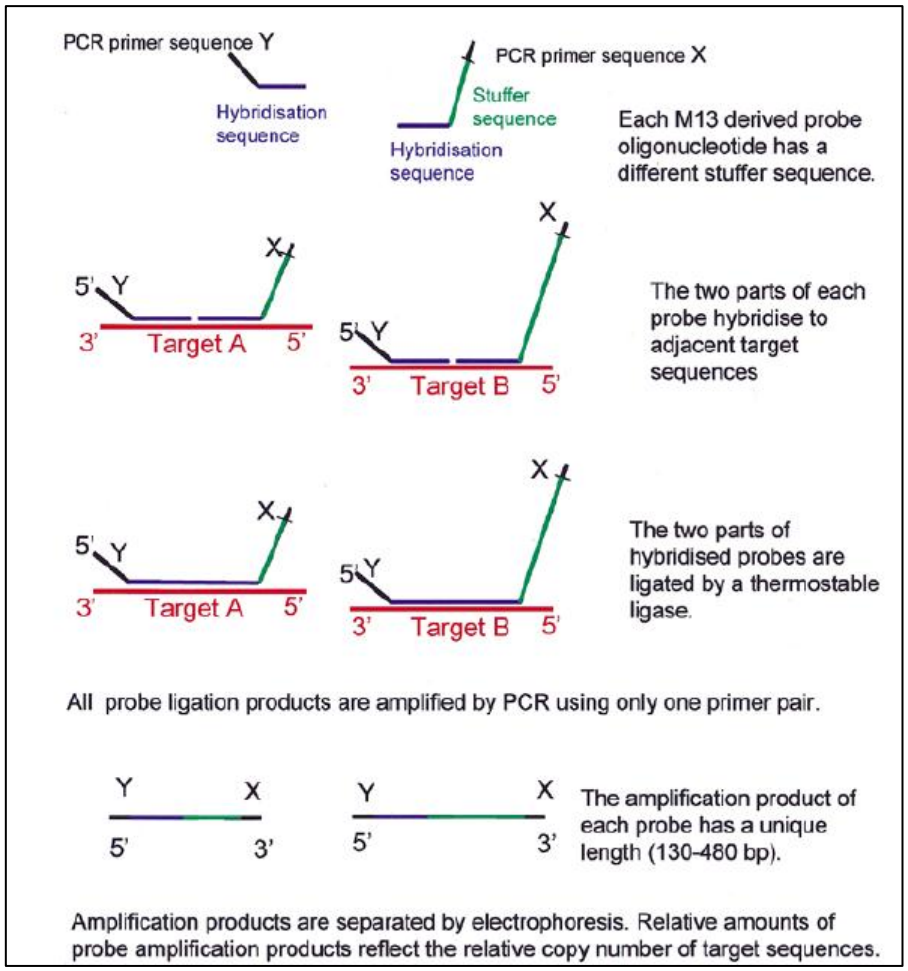


Figure 2.2: The MLPA reaction. This figure outlines the steps in an MLPA reaction. The hybridisation sequences bind to the target DNA and are then ligated together. The hybridised products are then amplified in a PCR reaction, with each amplicons having a unique length (Schouten et al., 2002)

MLPA is particularly useful when trying to detect large deletions or duplications. Large deletions where the breakpoints occur outside of the PCR primer will not be detected by Sanger sequencing. If the PCR primers are unable to bind to an allele because the breakpoints fall outside of the primers, the electropherogram will not show any deletion since only a single allele is amplified. MLPA is a quantitative technique, therefore if an allele is not present, the probe will not bind and this lack-of-binding will be detected through the quantitative PCR reaction. Similarly, an extra allele will be detected during quantification as an additional probe binds to that allele (Schouten et al., 2002).

MLPA and Sanger sequencing complement each other during mutation detection. If a patient has a homozygous deletion of an exon, or more than one exon, Sanger sequencing would appear to fail as the alleles are not present. However, MLPA will detect these deletions as control probes will allow the operator to determine whether an exon is deleted, or if the MLPA reaction failed. The Sanger sequencing will detect the single base changes and small deletions or duplications that MLPA is

unable to detect (Schouten et al., 2002; Boland and Goel, 2010). Somatic mutations in cancer can be problematic due to the diversity of mutations in a cancer cell, but this study is aiming to detect germline mutations by investigating blood, which should not contain many somatic mutations.

The P003-B2 Probe mix (SALSA MLPA Kit P003-B2 MLH1/MSH2) contains probes for the 19 exons of *MLH1* and the 16 exons of *MSH2*. This kit contains two probes for the most 3' exon of the *EPCAM* gene, which has been shown to be implicated in HNPCC (Ligtenberg, 2009). Furthermore, the P003-B2 kit contains nine probes that bind to other chromosomal positions and are used as reference probes. The details of the probe mix can be found at www.mrc-holland.com. The MLPA technique was performed following the manufacturer's instructions (given in Appendix F.1). Each reaction mix for the MLPA technique is given from Table F.1 to F.5.

The MLPA technique requires control samples (5-10) to be run at the same time as the patient samples. This not only allows the results to be quantified, but also lowers the error rate. The genotype of these control samples was unknown, but they were subsequently confirmed to be wild types during MLPA analysis. The MLPA samples were collected into a 96-well plate and placed inside the 3130xl Genetic Analyser (Applied Biosystems, USA). The Foundation Data Collection v3.0 software (Applied Biosystems, USA) was used to initiate the MLPA runs and collected the raw data (Appendix F.2).

Nineteen of the 20 patient samples were included in the MLPA analysis. Patient 136 was not included in the MLPA analysis as this sample was included late in the study. MLPA for a single sample would be too expensive as it is necessary to run at least five control samples with the samples being analysed.

2.2.2. Polymerase Chain Reaction

PCR is a technique that allows for the amplification of a specific segment of DNA (Mullis and Faloona, 1987). The following sections will describe the methods used for PCR amplification in this study.

2.2.2.1. PCR Primers for PCR Reactions

The primer sequences used for *MLH1* and *MSH2* were sent to us courtesy of the Regional Molecular Genetics Service at Manchester Academic Health Science Centre, St Mary's Hospital, Manchester. The primer sequences and primer binding sites are given in Appendix E. These primers were used to amplify each exon as well as the intron/exon boundaries of both genes. The primers were checked for specificity of binding using the Primer-BLAST tool at the National Centre for Biotechnology Information (Rozen and Skaletsky, 2000; NCBI, 2011). The primers were synthesised by Inqaba Biotec (South Africa) originally and Whitehead Scientific (South Africa) synthesised the primers for exons 7 and 8 of *MLH1* after the original primers were found to be faulty.

The primers were designed to include universal tails at the 5' and 3' ends of the PCR products which allow for easier set up of the Sanger sequencing technique. This was facilitated by a single forward or reverse primer being used for the Sanger sequencing reaction. The utility of the universal primers is explained in more detail in Section 2.4.1.

2.2.2.2. PCR Conditions for DNA Amplification

The same PCR conditions were used for each primer pair in every reaction mix. The PCR reaction mix is shown in Appendix F (Table F.6). The annealing temperature for the primers (55°C) was given in the Standard Operating Procedure (SOP) from the Regional Molecular Genetics Service, Manchester. The PCR reaction mixes were placed into an Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, USA) and the cycling conditions from the SOP were used (Appendix F, Table F.7). Negative controls were run with each batch of samples to check for contamination. The conditions for the thermo-cycling can be seen in Table F.7.

Initially, a temperature gradient was used to determine the optimal annealing temperature. It was found that the 55°C from the SOP was the optimal temperature for annealing. During the gradient PCR, the PCR reaction for exon 7 and 8 of *MLH1* did not produce a single band. Further troubleshooting involved $MgCl_2$, temperature and primer gradient PCR. These did not resolve the problem and new primers were ordered from a different manufacturer (as mentioned in Section 2.2.2.1). These new primers produced a single, correct sized band at 55°C.

2.2.2.3. Agarose Gel Electrophoresis of PCR Products

Amplification and band sizes of all the PCR reactions, as specified in Appendix E, were confirmed by visualising products on a 3% (w/v) agarose gel (Bioline, UK) containing 3% (v/v) Ethidium Bromide (Sigma-Aldrich, UK). For each PCR product, 2µl was mixed with 2µl of Ficoll dye before the mixture was loaded into the wells of the agarose gel with a multichannel pipette. The PCR product was electrophoresed for approximately one hour at 5V/cm. Finally, the product was visualized using a G:Box Vacutec Gel Documentation system with GeneSnap v.6.08 software (Syngene, UK). A 1KB+ DNA ladder (Invitrogen, USA) was used to size the bands in the gel.

2.2.2.4. PCR Product Clean Up

Following confirmation of the correctly sized PCR products, the PCR products were cleaned up to remove the unused primers as well as excess reagents. It is important to remove these primers and reagents before Sanger sequencing as they can interfere in the sequencing process. PCR clean-up was performed using either the Qiagen Minelute 96 UF PCR purification kit (Qiagen, Germany) or the Machery-Nagel Nucleospin Extract II column clean-up kit (Machery-Nagel, Germany), both of which filtered the PCR products through a membrane and an elution buffer was used to collect the purified PCR products.

The Minelute 96 UF PCR purification kit was primarily used due to the high-throughput efficiency of the 96-well plate system. This kit was used following the manufacturer's instructions (Appendix F.4). The Nucleospin Extract II kit was used for individual samples or small batches. This kit was used following a protocol (Appendix F.5) that was modified to use a lower volume of reagents. The purified PCR products were placed into clean 96-well plates or strip tubes before being used in Sanger sequencing.

2.2.3. Sanger Sequencing

Sanger Sequencing was used to determine the nucleotide sequence of DNA. This sequencing method is based on the chain-termination method which utilises dideoxyribonucleotide triphosphates (ddNTPs). The ddNTPs are analogues of deoxyribonucleotide triphosphates (dNTPs) which are incorporated into the growing DNA chain by DNA polymerase. However, the ddNTPs lack a 3' hydroxyl group, which prevents the next dNTP from joining the growing chain. Separating the fragments by size allows one to elucidate the sequence of the DNA (Sanger et al., 1977).

The Sanger sequencing technique is useful for detecting base substitutions, small insertion or deletion mutations. Sanger sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) for all the *MLH1* and *MSH2* amplicons. This process was performed using a modified reaction mix, which is given in Appendix F (Table F.8). This reaction has been modified to use only 4µl of BigDye v3.1 instead of the 8µl recommended by the manufacturer. This was found to be the lowest volume of BigDye v3.1 which would yield good quality results. This was performed to consume fewer reagents, and thus save costs. The reaction mix was placed in a GeneAmp 2720 Thermal Cycler (Applied Biosystems, USA) and a cycle sequencing protocol (given in Table F.9) was followed.

2.2.3.1. Universal Primers for Sanger Sequencing

A universal sequencing primer was used in the reaction mix to facilitate easier set up of sequencing. The universal primer was complementary to either the 5' or 3' tail that had been included in the design of every PCR primer. Inclusion of a universal primer sequence optimises the sequencing process as it allows the operator to use a single sequencing primer for all reactions. This is particularly useful when sequencing the 18 primer pairs of *MLH1* and 16 primer pairs of *MSH2*. Without the universal primer, each exon to be sequenced would require an individual primer. Figure 2.3 shows how the universal tail and primer is used during both PCR reactions.

Generally, only the universal primer complementary to the 5' tail (the forward sequencing primer) was used and was sufficient to generate good sequencing data. Due to a microsatellite within the exon 12 in *MLH1*, a second, internal forward sequencing primer was needed to obtain good sequencing data. Bidirectional sequencing using the reverse primer was undertaken if the forward

sequencing data was unclear, or to confirm the presence of a variant. This is the routine approach used by the Molecular Diagnostic Division in the Department of Human Genetics, NHLS, Braamfontein. If the variant was still present after sequencing from both directions, it was considered to be a confirmed variant.

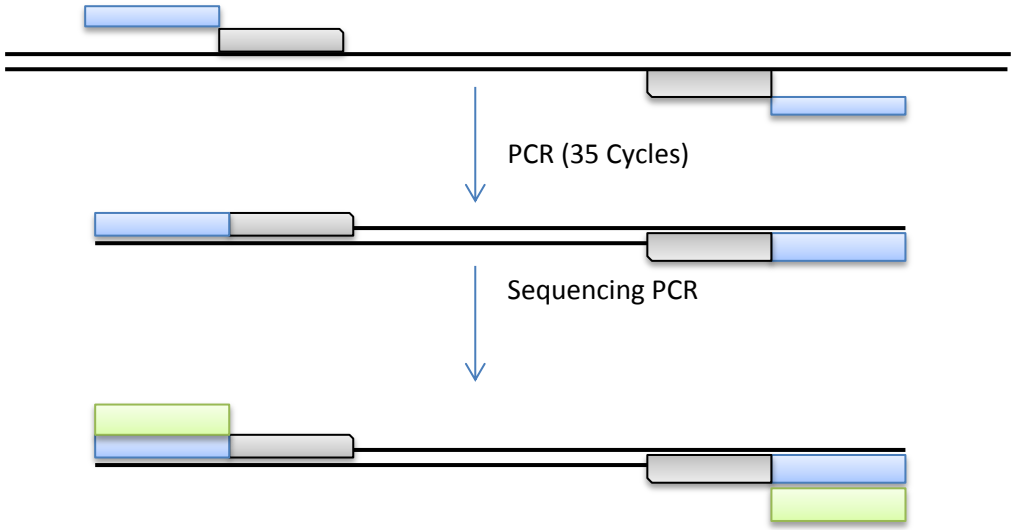


Figure 2.3: Schematic representation of PCR and sequencing reactions in this study. Universal primers (shown in blue) were included at the 5' and 3' ends of the primers (grey). The universal primers do not bind to the template DNA but are included in the amplicons. A universal sequencing primer (shown in green) can be used to sequence from either the 5' or 3' tail.

2.2.3.2. Sanger Sequencing Product Clean Up

Once the Sanger sequencing reaction was completed, it was important to remove excess BigDye v3.1 reagents from the sequencing products as they could interfere with the interpretation of the sequence in the Genetic Analyser. Sanger sequencing clean-up was performed using either the Qiagen DyeEx 96 or the Qiagen DyeEx 2.0 spin kits (Both from Qiagen, Germany). These kits use a gel bed to remove excess reagents in conjunction with a centrifugal force.

The DyeEx 96 and DyeEx 2.0 kit were used following the manufacturer's instructions. The DyeEx 96-well plates allowed for high throughput clean-up while the DyeEx 2.0 could be used for smaller batches. The protocols for each kit are given in Appendix F.7 and Appendix F.8, respectively. The cleaned Sanger sequencing product was collected and vacuum dried in a Concentrator 5301 (Eppendorf, Germany) before undiluted Hi-Di Formamide (Applied Biosystems, USA) was used to re-suspend the DNA products in preparation for the 3130xl Genetic Analyser.

2.2.3.3. The 3130xl Genetic Analyser

The Applied Biosystems 3130xl Genetic Analyser uses a gel matrix, located within the capillaries, to separate Sanger sequencing product fragments by size. The fragments then pass a laser which detects which fluorescent terminator is present on the particular fragment. Since they separate by size, the smallest fragments (those where the terminator is closest to the sequencing primer) pass the laser before larger fragments. The sequence of the DNA is then determined in the order that the fluorescent terminators pass the laser. This information is captured by the genetic analyser and processed by the Foundation Data Collection v3.0 software package. The 3130xl genetic analyser is able to process 16 samples in one run, with six runs needed to process a 96-well plate.

Before initializing a sequencing run, the Sanger sequencing product in the 96-well plate was denatured at 95°C, for two minutes. The 1x buffer and ddH_2O troughs were checked to ensure that they were at the correct level and the polymer was checked to make sure it was not empty or bubbles were present. The 96-well plate was loaded into the 3130xl Genetic Analyser and the Foundation Data Collection v3.0 software was used to create and start each run (given in Appendix F.9). Section 2.5 will describe the analysis of the Sanger sequencing and MLPA data.

2.2.4. Mutation Analysis of Sequencing and MLPA Data

Once Sanger sequencing and MLPA were completed, data analysis was undertaken to determine if there were any variants present. The MLPA and sequencing data were collected and analysed differently.

2.2.4.1. Sanger Sequencing Data Analysis

The Sequencing Analysis 5.2 software (Applied Biosystems, USA) was used to interpret the sequence of each sample. This program interprets each sequence automatically and uses a base caller to give the DNA its sequence based on the electropherogram present. The sequences can then be saved and analysed. The procedure for this can be seen in Appendix F.10.

The sequences were transferred and analysed using the DNAStar Seqman II software package (Lasergene, USA). Each electropherogram could be viewed in Seqman II and corrections could be made if there were errors in base calling. Errors were checked by inspection of the electropherograms. It was important to check peak intensities, and 'pull-ups' were inspected by observing the peak patterns to ensure that the software had called the correct bases.

The sequences needed to be aligned with a reference sequence in order to determine if there were any variants present. A reference sequence for each region was obtained from the NCBI Genome Browser (NG_007109 and NG_007110 for *MLH1* and *MSH2*, respectively, obtained on 21/07/2011) using the RefSeq that was independent of genome builds (NCBI, 2011). The Seqman II programme

aligned the sequences with the reference sequences and then highlighted any mismatches. By reading through each sequence, it was possible to find variations from the RefSeq sequence.

2.2.4.2. MLPA Data Analysis

MLPA analysis required a panel to be designed in the GeneMapper software v4.0 (Applied Biosystems, USA). This panel needed to be designed so that the software could calculate the relative quantities of each amplicon. Since the size of each amplicon in the P003-B2 MLPA kit is known, it was possible to input the expected sizes of each amplicon into the panel. The amplicons separate by size in the capillaries of the 3130xl Genetic Analyser, allowing the software to determine which amplicon was being detected by the 3130xl Genetic Analyser. As the amplicons passed through the laser, the intensity of each was detected and captured by the software and recorded in the panel. It is the relative intensities of each amplicon from the samples and the control samples that is used to determine the allele copy numbers (Schouten et al., 2002). The procedure to retrieve the MLPA data can be seen in Appendix F.10. The GeneMapper v4.0 software was used to analyse the raw data and determine if the correct size calling occurred for each allele. The data from the GeneMapper v4.0 software was then exported to await analysis using spread sheets from the National Genetics Reference Laboratory.

The National Genetics Reference Laboratory (NGRL) at Manchester provides Excel spread sheets that have been preformatted for easier analysis and interpretation of MLPA results (NGRL, 2009). These spread sheets are available from their website (www.NGRL.org.uk). The P003-B1 HNPCC spread sheet was downloaded to analyse the MLPA data collected. The P003-B2 kit used in this study is an updated P003-B1 kit, with two control probes being replaced in the newer kit. The spread sheet was formatted to allow easy input of data from the GeneMapper software, thus making it ideal for use in this study. The final data is presented as a graph showing the dosage of each exon in the MLPA probe mix.

Once all the data from MLPA and Sanger sequencing was collected, it was possible to check the reporting history of the variants found using the LOVD database at InSiGHT.

2.2.4.3. Determining Reporting History of Variants

All variants that were found were checked against one large curated database, The LOVD database (v2.0, Build 32b; www.LOVD.nl) at InSiGHT (InSiGHT, 2011). The LOVD database aims to provide a freely available and flexible tool for the collection and display of genetic variants (Fokkema et al., 2011). The LOVD database at InSiGHT is a curated database that has been modified and optimised to display the data from *MLH1*, *MSH2* and eight other genes commonly involved in hereditary CRC (InSiGHT, 2011).

InSiGHT provide a coding reference sequence which allowed for an international standardised numbering of nucleotides, allowing for the correct positioning of a variant when identified. The numbering in the variant name refers to the nucleotide relative to the ATG start site, where the 'A' is the first base of the coding sequence (InSiGHT, 2011). Once the variant was correctly named, it could be checked against the database to determine if it is novel or has been reported previously.

The LOVD database at InSiGHT contains variants that have been published as well as variants that have been directly submitted to the database but not previously published. The database also displays information from other, merged databases such as the MMR Gene Unclassified Variants Database (Ou et al., 2008) and the Mismatch Repair Genes Variant Database at the Memorial University of Newfoundland (Woods et al., 2007). If a variant has already been reported in the database, it may be known to be pathogenic or non-pathogenic. This allows variants to be classified as novel, pathogenic, non-pathogenic or of unknown significance. The procedure for using and evaluating variants on the InSiGHT database can be seen in Appendix F.11.

If a variant was previously reported on the database, it is possible that it would have been submitted to the Single Nucleotide Polymorphism database (dbSNP, Build 135) (Sherry et al., 2001). dbSNP is the world's largest database of variations and contains information on variants that have been reported. Variants submitted to dbSNP receive a reference SNP number which allows the variant to be tracked.

The LOVD database contained links to the UCSC Genome Browser for all previously reported variants (Kent et al., 2002a). The UCSC Genome Browser contained information on the variant position and the reference SNP number from dbSNP (if available). UCSC also contained the BLAST-like Alignment Tool (BLAT), which allowed the variant sequences to be checked to ensure that they had not been reported previously on other databases (Kent, 2002b). The BLAT searches were performed by inserting the sequence into the BLAT tool and determining whether the sequences had been reported previously by giving a percentage match of the queried region (BLAST-Like Alignment Tool (BLAT), 2012).

2.2.5. Determining Pathogenicity of Novel Variants

Six novel variants with an unknown functional significance were detected using the Sanger sequencing method. Since these variants had an unknown contribution toward disease, elucidation of their potential for pathogenicity was sought. The allele frequencies for each allele were determined in a sample of normal controls to elucidate their potential for pathogenicity, using RFLP or HRM. If a variant occurs only in one patient, it is suggestive that the variant is possibly pathogenic as more common variants are unlikely to be pathogenic (Strachan and Read, 2004). The allele frequencies from the control samples were calculated by dividing the number of times a variant allele was observed by the total number of alleles in the assay.

Bioinformatic analysis was used to determine whether one variant (*MSH2* c.-185C>A) may have a possible pathogenic effect. This variant was investigated using bioinformatic tools as it is located upstream of the transcription start site. Transcription factor binding site prediction, such as that found in FASTSNP (fastsnp.ibms.sinica.edu.tw), is a simple and robust method for predicting whether a variant will have a functional effect (Yuan et al., 2006). The other variants were located near intron/exon boundaries and may create or abolish splice sites. However, *in silico* tools were unable to reliably reproduce results for splice site abolition or creation.

2.2.5.1. Restriction Fragment Length Polymorphism Analysis

RFLP analysis uses restriction endonucleases to determine whether a restriction site is present, or not. This is particularly useful for genotyping as one can design an assay with an appropriate restriction endonuclease and determine the genotype of the DNA, depending on whether or not the restriction site is present or absent (Strachan and Read, 2004). For this study, the same PCR primers for previous PCR reactions were used to amplify the region of interest. Restriction endonucleases were then chosen firstly based on whether they had a restriction site at the variant position, and secondly whether they were present in the lab or could be ordered. After the restriction enzymes were chosen, the PCR products were checked for additional cut sites in the product by searching the previously mentioned RefSeq (2.2.4.1) from NCBI (NCBI, 2011). No additional cut sites were expected to be present in the PCR products.

RFLP analysis was performed for four of the novel variants detected in this study to determine the frequency of the variant in the normal population. Other novel variants were not investigated using RFLP as RFLP assays could not be designed due to restriction endonucleases not being available to cut at the site of the variant. The novel variants that could not be investigated by RFLP were investigated using HRM. The four variants that were investigated using RFLP are listed in Table 2.2.

Table 2.2: Restriction enzymes used to investigate the frequencies of the novel variants

Gene	Variant	Restriction Enzyme	Incubation Temperature	Recognition Site	Effect of Variant on Recognition Site
MLH1	c.885-1G>A	*HpyCH*4III	37°C	ACNGT	Abolish
MSH2	c.1662-10C>T	*Taq*I	65°C*	TCGA	Abolish
MSH2	c.1662-43C>T	*Rsa*I	37°C	GTAC	Abolish
MSH2	c.-185C>A	*Mlu*I	37°C	ACGCGT	Create

***10% efficiency at 37°C**
***Taq*I (Thermo Fisher Scientific, USA)**
***HpyCH*4III, *Rsa*I & *Mlu*I (New England Biolabs, USA)**

2.2.5.1.1. RFLP Conditions

The RFLP analysis was performed using similar conditions for all reactions, which can be seen in Table 2.3. The samples were incubated at 37°C. Although *Taq*I has an incubation temperature of 65°C, it is still able to digest at 37°C. The RFLP reactions were performed to confirm presence of variants in the patient DNA. Each variant were screened in the normal, ethnically matched control samples, a no-enzyme control (referred to as a 'no-cut' control) to ensure no contamination of enzymes, and the patient allele to confirm the presence of the variant. The samples were checked using a 3% (w/v) agarose gel at 5V/cm with ethidium bromide and a 1kb+ DNA ladder was used to determine the size of the digested fragments.

Table 2.3: Summary of the RFLP reactions for all enzymes

Reagent	Final Concentration	Volume (1X)
Restriction Enzyme	1U/ul	1ul
Appropriate Buffer	1X	2.5ul
ddH_2O		11.5ul
PCR Product		10ul
Total		25ul

2.2.5.1.2. RFLP for *MLH1* Variants

The *MLH1* c.885-1G>A variant was detected in a white patient. The PCR primers for exon 9 of *MLH1* were used to amplify the region around this variant. The amplicon was then digested with *HpyCh*4III. The enzyme will cut the wildtype allele of the PCR product, which is 247bp in size. Therefore, following digestion, if the 'wildtype' sequence is present, a 77bp fragment and 169bp fragment will be present. Figure 2.4 shows the expected fragment lengths for the wildtype and variant sites.

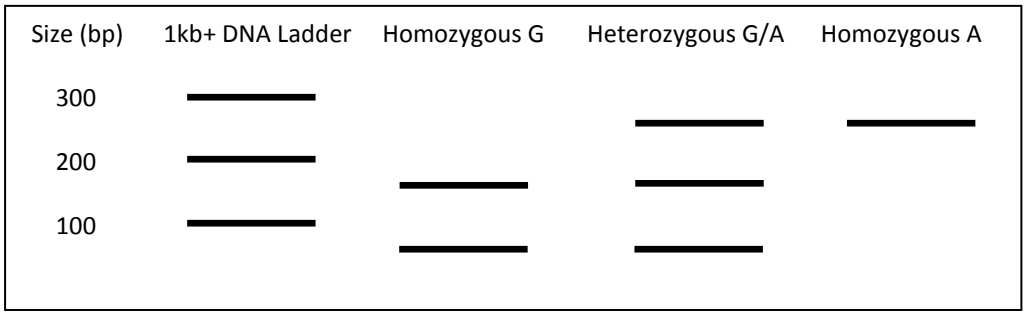


Figure 2.4: Diagram of the expected RFLP analysis using *HpyCH*4III to screen for the *MLH1* c.885-1G>A variant. *HpyCH*4III will recognise and cut the wildtype allele.

2.2.5.1.3. RFLP Analysis for *MSH2* Variants

The *MSH2* c.1662-10C>T novel variant was investigated further as it was found in two black patients. The primers for exon 11 of *MSH2* were used to amplify the PCR product, which give a product of 246bp. *Taq*I was used to investigate this variant as it recognises the wildtype TCGA site. *Taq*I generates two fragments of 80bp and 166bp in wildtype samples. The expected RFLP fragments are given in Figure 2.5.

The *MSH2* c.1662-43C>T variant was found in one black patient. The PCR primers for exon 11 of *MSH2* were used to amplify the region surrounding this variant, giving a PCR product size of 246bp. This variant was investigated using *Rsa*I, which cuts the wildtype C allele, producing bands of 47bp and 199bp. The expected RFLP pattern for each allele can be seen below, in Figure 2.6.

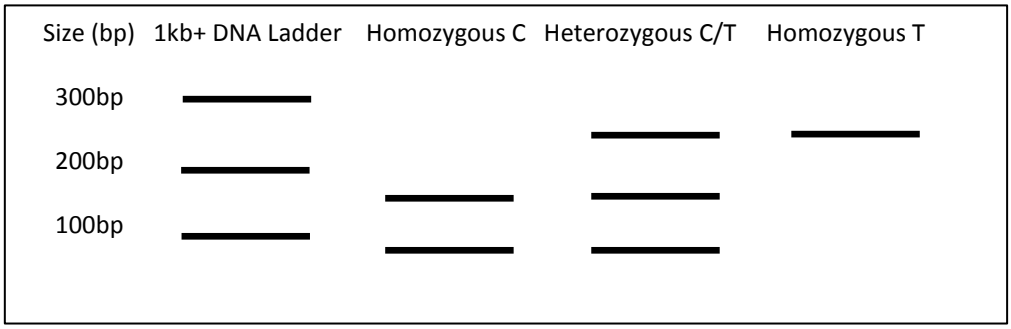


Figure 2.5: Diagram showing the expected RFLP fragments when *Taq*I is used to investigate the *MSH2* c.1662-10C>T variant. *Taq*I will recognise the wildtype allele

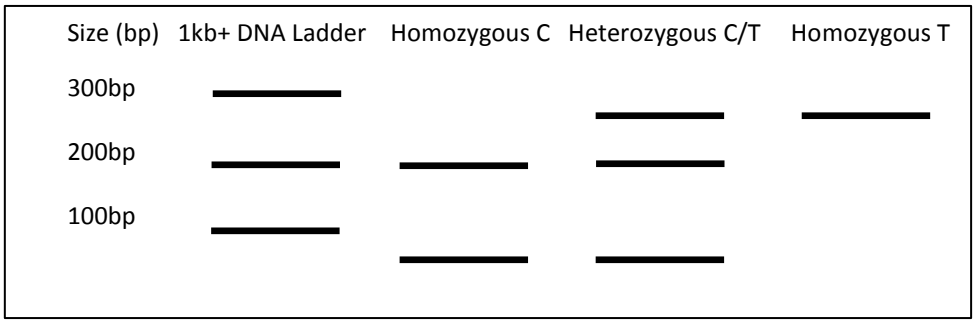


Figure 2.6: Diagram showing the expected RFLP fragments when *Rsa*I is used to investigate the *MSH2* c.1662-43C>T variant. RsaI recognises the wildtype GTAC site.

The *MSH2* c.-185C>A variant was investigated further after it was detected in one black patient. The PCR primers for exon one of *MSH2* were used to amplify the DNA, giving a PCR product of 535bp in size. *MluI* would cut the DNA if the variant C allele was present, leading to two bands of 35bp and 500bp. The expected RFLP bands are displayed in Figure 2.7.

2.2.5.2. High Resolution Melt Analysis

HRM is a technique used to genotype DNA using fluorescent dyes, melt-curve analysis and other statistical techniques. HRM uses PCR techniques with a fluorescent, intercalating dye to perform real-time PCR. Immediately after the real-time PCR stage, the PCR products are melted and the fluorescence is measured. The technique is so sensitive, that a single base pair change in the PCR product could alter the melting curve, even in PCR products as large as 300bp (Gundry et al., 2003; Wittner et al., 2003). The sensitivity of the technique allows one to determine the genotype of the

sample by comparing the normalised melting curve of a sample with that of a control, as shown in Figure 2.8.

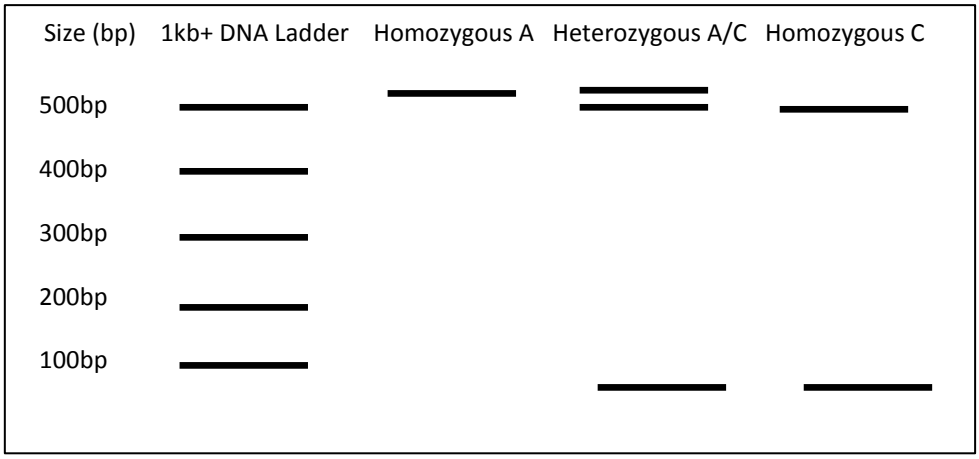


Figure 2.7: The expected RFLP fragments when *MluI* is used to investigate the *MSH2* c.-185C>A variant located within the regulatory region. *MluI* will cut the DNA if the variant is present.

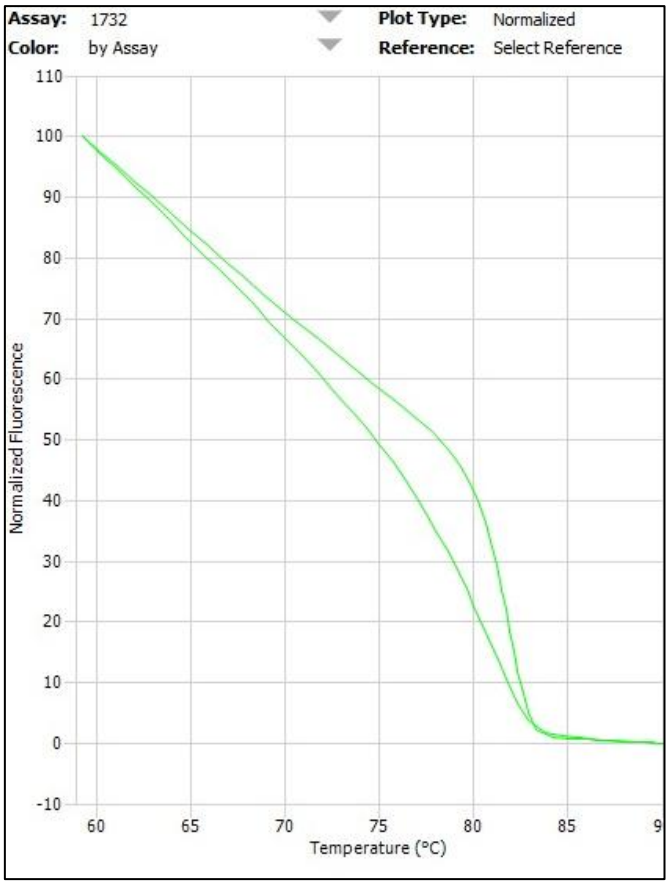


Figure 2.8: The HRM normalised melting curve obtained from control samples. The wildtype allele is the top curve and the variant (*MLH1* c.1732-19T>A is the bottom curve. Random samples would be compared to these curves to determine allele frequency

HRM analysis was used to determine the allele frequencies of two novel variants (*MSH2* c.367-16T>A and *MLH1* c.1732-19T>A) in 25 random samples from the ethnically matched populations (described

in Section 2.1.4). HRM was chosen to analyse these variants as the restriction enzymes required to investigate these variants were not available in the lab and could not be ordered. Only 25 samples were used as the HRM reagents are costly and were limited. The PCR primers used to amplify the variants were the same PCR primers used for previous PCR reactions. These gave amplicons sizes of 284bp (*MLH1* c.1732-19T>A) and 391bp (*MSH2* c. 367-16T>A).

HRM was performed using the Bioline SensiMix HRM kit (Bioline, USA) in the Roche LightCycler 480 (Roche, Switzerland) following the manufacturer's instructions (given in Appendix F.12). This kit contains EvaGreen fluorescent dye. SYBR Green is not considered a good dye for HRM as it affects the melting of the double stranded DNA; it inhibits PCR and affects the signal intensities as liberated dyes redistribute to the double stranded DNA (Gundry et al., 2003). EvaGreen is able to provide similar spectra to SYBR Green, but does not have the abovementioned drawbacks.

The normalised melting curve was obtained by performing HRM on two samples with known genotypes. The positive control for a heterozygous genotype was the patient sample that was found to harbour the variant. The normal homozygote was also a patient sample that was known to be a homozygous normal after sequencing. When the HRM analysis was completed, a normalised melting curve, which accounts for differences in temperature between wells, for the samples and controls was generated by the software, using the information from the fluorescence as the amplicons melted.

2.2.5.3. FASTSNP

The Functional Analysis and Selection Tool for Single Nucleotide Polymorphisms (FASTSNP) is a bioinformatics tool that is used to predict whether a particular variant will have a functional effect on the protein. The tool analyses non-coding variants, non-synonymous variants and synonymous variants (Yuan et al., 2006).

FASTSNP was used to evaluate whether the *MSH2* c.-185C>A variant could have a possible functional effect with regard to changes in mRNA transcription by altering transcription factor binding sites. The 300bp region upstream of the transcription start site was submitted to FASTSNP using the novel SNP analysis tool. The position of the variant was given as well as the base change.

Since this is a non-coding region, FASTSNP will analyse the region to determine if the variant alters the binding site of any transcription factors, which may affect mRNA transcription. FASTSNP analysed the sequence and the variant against the TFSEARCH database to predict transcription factor binding affinities. Altering mRNA transcription may lead to a lower or higher expression of mRNA, which could affect the functioning of the gene (Strachan and Read, 2004; Yuan et al., 2006).

3. Results

This chapter will describe all the results that were found during this study. Section 3.1 will describe the results of the MLPA analysis. Section 3.2 will describe all the variants that were found during Sanger sequencing of the exons of *MLH1* and *MSH2*. This will include all the variants that were found and their positions in the gene as determined using the InSiGHT coding reference sequence. Section 3.3 will describe the results of elucidating the pathogenicity of the novel variants that were detected using Sanger sequencing. Section 3.4 will characterise the mutation spectrum of the *MLH1* and *MSH2* variants in this cohort.

3.1. MLPA Analysis for *MLH1* and *MSH2*

MLPA analysis is an important technique to perform as it has been found that more than 6% of *MSH2* mutations are large deletions and that large deletions also occur in *MLH1* (Wijnen et al., 1998; Boland and Goel, 2010). Gille et al. (2002) found that 29% of their cohort had mutations that were detected using MLPA, 22% of which were found to be pathogenic. MLPA was performed on 19 patients and five random white samples as controls. The MLPA technique found no large deletions or duplications in any of the exons tested in the P003-B2 kit. The MLPA graphs that were generated were all similar and an example of the graphs can be seen in Figure 3.1.

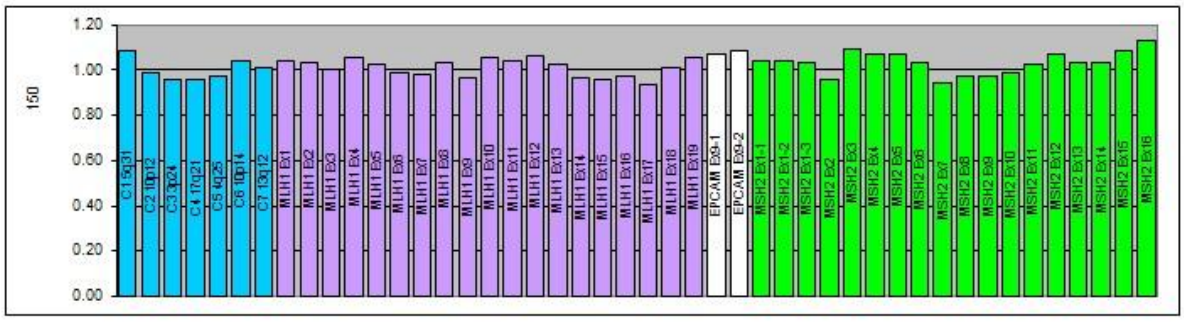


Figure 3.1: The MLPA result generated using the spread sheet from NGRL. Since all patients tested for MLPA were found to have no large insertions, deletions or duplications, all the patients MLPA results were similar to this one, with each allele being ≈1.0. If one allele was deleted, the bar would have been ≈0.5. If both alleles were deleted, the bar would be at 0. If one allele was duplicated, the bar would be ≈1.5, and if both alleles were duplicated, it would have been ≈2.0.

3.2. Variants detected using Sanger Sequencing

All 19 exons of *MLH1* and all 16 exons of *MSH2* were successfully sequenced and produced good data. Variants were detected from sequencing the forward primer and these variants were then confirmed by sequencing using the reverse primer. The variants present on both electropherograms were considered confirmed. The expected mutation from the positive control was detected. An additional 21 different variants were detected in the patient samples. All the variants that were detected are given in Table 3.1 for *MLH1* and Table 3.2 for *MSH2*.

Table 3.1: Summary of all variants detected in each patient in *MLH1*. The dbSNP numbers are available from NCBI (Database for Single Nucleotide Polymorphisms (dbSNP), 2012)

Gene	Variant	dbSNP Number	Location	Patient Number																		
				13 (MA)	32	42	52	61	81	86	89	136	150	190	253	104	175 (MA)	205	235	238	247	267	290
MLH1	c.381-41A>G	rs4647245	Intron 4																		X		X
MLH1	c.454-51T>C	rs4647255	Intron 5													X							
MLH1	c.454-13A>G		Intron 5													X							
MLH1	c.655A>G	rs1799977	Exon 8			X		X			X			X								X	
MLH1	c.731_734delGTTA		Exon 9									X											
MLH1	c.885-1G>A		Intron 10						X														
MLH1	c.1039-8T>A	rs193922367	Intron 11				X				X									X			
MLH1	c.1039-7delT		Intron 11							X				X					X				X
MLH1	c.1558+14G>A	rs41562513	Intron 13														X						
MLH1	c.1668-19A>G	rs9876116	Intron 14			X		X			X			X				X	X		X	X	X
MLH1	c.1732-19T>A		Intron 15															X					

Patient Number: Green= Amsterdam II Positive, Blue= Bethesda positive
MA: Mixed Ancestry Patients. Other green blocks are white patients and other blue blocks are black patients
Novel variants: Orange blocks
Previously reported pathogenic variants: Red blocks

Table 3.2: Summary all variants detected in each patient in *MSH2*. The dbSNP numbers are available from NCBI (dbSNP, 2012)

Gene	Variant	dbSNP Number	Location	Patient																		
				13 (MA)	32	42	52	61	81	86	89	136	150	190	253	104	175 (MA)	205	235	238	247	267	290
MSH2	c.-185C>A		Upstream															X					
MSH2	c.211+9C>G	rs2303426	Intron 1			X			X						X			X	X	X	X		
MSH2	c.367-16T>A		Intron 2					X															
MSH2	c.573C>T	rs1800151	Exon 3															X					
MSH2	c.942+16_942+29delA_{13}		Intron 5															X					
MSH2	c.1511-9T>A	rs12998837	Intron 9						X														
MSH2	c.1661+12G>A	rs3732183	Intron 10						X					X				X	X	X	X		X
MSH2	c.1662-43C>T		Intron 10													X							
MSH2	c.1662-10C>T		Intron 10																X	X			
MSH2	c.1811C>A		Exon 12										X										
MSH2	c.2006-6T>C	rs2303428	Intron 12											X									

Patient Number: Green= Amsterdam II Positive, Blue= Bethesda positive
MA: Mixed Ancestry Patients. Other green blocks are white patients and other blue blocks are black patients
Novel variants: Orange blocks
Previously reported pathogenic variants: Red blocks

A total of 22 variants were detected in this study, including the positive control. All the variants were checked to determine reporting history (InSiGHT, 2011; BLAT, 2012). Of the 22 variants, 12 had been previously reported as non-pathogenic and four were previously reported as pathogenic mutations. Six novel variants were detected. All the variants, except the *MSH2* pathogenic mutation, c.1811C>A, were detected in the heterozygous state. The *MSH2* c.1811C>A mutation was detected as a homozygote, although previous investigation showed that this is a heterozygous patient. This finding is discussed in more detail in Section 4.3.3.1.

3.2.1. Variants Detected in *MLH1*

There were 11 variants found in *MLH1* (Table 3.1). Of these variants, two were located within exons (c.655A>G and c.731_734delGTTA) and the other nine variants were located within introns. The 11 variants detected in *MLH1* were screened through the LOVD at InSiGHT (InSiGHT, 2011). Nine of the variants in *MLH1* were previously reported and two were novel variants. Seven of the previously reported variants were found to be non-pathogenic while two were previously reported as pathogenic mutations.

3.2.1.1. Previously Reported Non-Pathogenic Variants in *MLH1*

The seven previously reported non-pathogenic variants are summarised in Table 3.3. All of these variants have been reported as benign or have a minor effect. The c.655A>G variant (p.Ile219Val) was found in five patients in this study (four whites and one black), suggesting that it is a common variant in the population of South Africa. The electropherogram for the c.655A>G variant can be seen in Figure 3.2. This variant has been observed in other studies (Table 3.3) and is largely considered benign. However, Campbell et al. (2009) indicate that this variant may lead to an increased risk of CRC in conjunction with the Western diet. The other non-pathogenic variants in *MLH1* are benign and will not be investigated further.

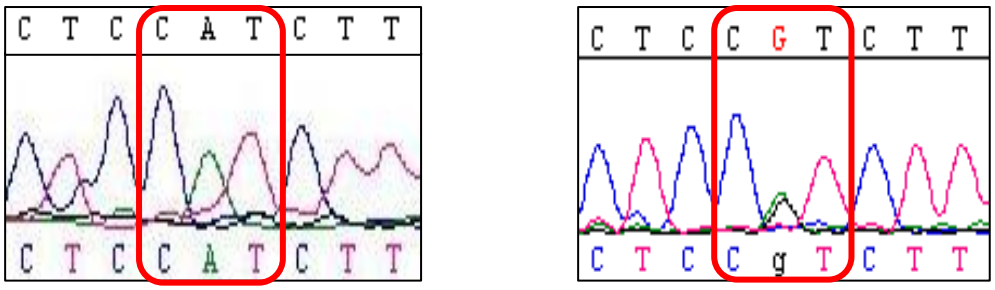


Figure 3.2: The *MLH1* c.655A>G variant (p.I219V), which leads to a missense mutation, in exon 8. The wildtype DNA sequence is seen on the left while the variant can be seen on the right.

3.2.1.2. Previously Reported Pathogenic Variants in *MLH1*

Two of the variants detected in *MLH1* (c.454-13A>G and c.731_734delGTTA) were previously reported as pathogenic. These variants are summarised in Table 3.4. These two previously reported

variants are confirmed as pathogenic mutations, which indicates that patient 104 and patient 136 have confirmed Lynch Syndrome. The implications of these mutations in discussed in Section 4.1.1.

Table 3.3: Summary of the previously reported non-pathogenic variants in *MLH1* that were detected in the South African cohort

Variant	Location	Reference	Comment
c.381-41A>G	Intron 4	Zavodna et al., 2006	Non-pathogenic
c.454-51T>C	Intron 5	Nystrom-Lahti et al., 1996 Wu et al., 1997 Fredriksson et al., 2006	Non-pathogenic
c.655A>G (p.Ile219Val)	Exon 8	Liu et al., 1995 Nystrom-Lahti et al., 1996 Ward et al., 2002 Fredriksson et al., 2006 Takahashi et al., 2007 Capellá et al., 2008 Christensen et al., 2008 Campbell et al., 2009	Associated with increased risk of CRC in patients with high Western diet intake (Campbell et al., 2009)
c.1039-8T>A	Intron 11	Wang et al., 1998	Non-pathogenic
c.1039-7delT	Intron 11	Hudler et al., 2004	Non-pathogenic
c.1558+14G>A	Intron 13	Kowalski et al., 1997 Auclair et al., 2006 Fredriksson et al., 2006 Christensen et al., 2008	Non-pathogenic
c.1668-19A>G	Intron 14	Nystrom-Lahti et al., 1996 Hudler et al., 2004 Auclair et al., 2006 Fredriksson et al., 2006 Capellá et al., 2008 Christensen et al., 2008	Non-pathogenic

Table 3.4: Summary of the previously reported pathogenic variants that were detected in *MLH1* in the South African cohort

Mutation	Location	Patient and Ethnicity	References	Comment
c.454-13A>G	Intron 5	104, Black	Lagerstedt Robinson et al., 2007	Aberrant splicing of exon 6 (Lagerstedt Robinson et al., 2007)
c.731_734delGTTA	Exon 9	136, White	Ramesar et al. (InSiGHT, 2011)	4bp deletion leading to a truncated protein (InSiGHT, 2011)

3.2.1.2.1. The *MLH1* c.454-13A>G Splice Mutation

The *MLH1* c.454-13A>G mutation was detected in patient 104 who is a black patient who fulfilled the age criterion of the Bethesda guidelines. This mutation was heterozygous in patient 104 and has been previously reported to show aberrant splicing of exon 6 in *MLH1* (Lagerstedt Robinson et al., 2007). The electropherogram for this mutation can be seen in Figure 3.3.

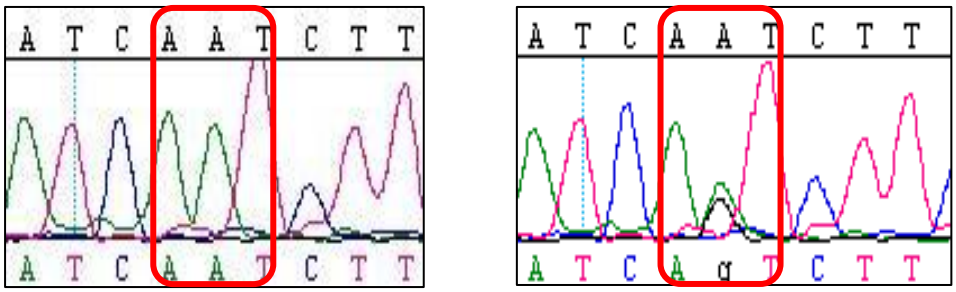


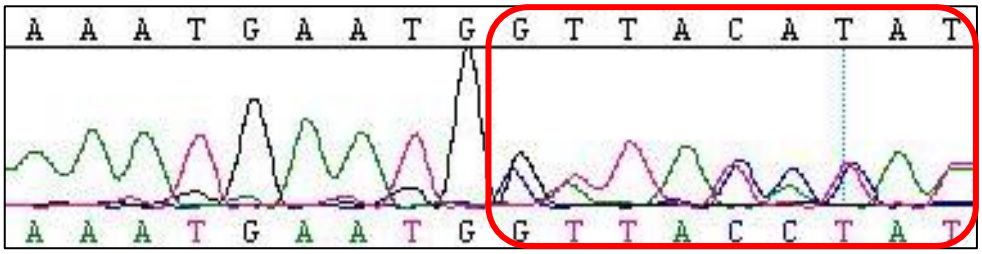
Figure 3.3: The *MLH1* c.454-13A>G variant of intron 5 found in Patient 104. The wildtype sequence is displayed on the left while the image on the right shows the heterozygous variant seen in Patient 104

3.2.1.2.2. The *MLH1* c.731_734delGTTA Frameshift Mutation

The *MLH1* c.731_734delGTTA mutation was found and confirmed as a heterozygous 4bp deletion in exon 9 of *MLH1* in patient 136, a white patient who fulfilled the Amsterdam II criteria. This mutation has been reported on the InSiGHT database by Ramesar et al. (unpublished data). However, the InSiGHT entry has this variant as c.731_733delGTTA. This nomenclature is not correct as it is a four base pair deletion and not a three base pair deletion. This variant will be called as c.731_734delGTTA in this study. The variant data from the LOVD is given in Figure 3.4. The mutation has been reported as a deletion that leads to a frame shift, which results in a premature stop codon downstream, as given in the abbreviated nomenclature, p.fsX. This gives a truncated protein product (InSiGHT, 2011). Figure 3.5 shows the electropherograms for this mutation.

Allele	Parent #1
Reported pathogenicity	Pathogenic
Concluded pathogenicity	Unknown
Exon	09
DNA change	c.731_733delGTTA (View in UCSC Genome Browser)
DNA_reported	731delGTTA
RNA change	r.(?)
Protein	p.fsX
DB-ID	MLH1_00870
Variant remarks	-
Frequency	-
Detection/Template	DNA
Detection/Technique	SEQ
Variant reference	Ramesar et al., (unpublished) 2004

Figure 3.4: The variant data that was obtained from LOVD for the *MLH1* c.731_734delGTTA variant (InSiGHT, 2011). This image shows that the variant is named as c.731_733delGTTA, which is incorrect as it is a four base pair deletion



Reverse primer for patient 136

Figure 3.5: The *MLH1* c.731_734delGTTA mutation in exon 9. This sequence was inspected to ensure that the correct base calling was made. The reverse primer was used to confirm the presence of the mutation and can be seen in this image.

3.2.1.3. Novel Variants Detected in *MLH1*

Two novel variants were detected in *MLH1*, which are shown in Table 3.5. These variants had an unknown contribution toward disease. Section 3.3 will describe the results of determining the allele frequency of each variant and determining the pathogenic status for novel variants in *MLH1* and *MSH2*. The electropherograms for c.885-1G>A and c.1732-19T>A can be seen in Figures 3.6 and 3.7, respectively.

Table 3.5: Summary of the novel variants detected in *MLH1* in the South African cohort

Variant	Location	Patient	Ethnicity	Inclusion Criteria
c.885-1G>A	Intron 10	81	White	Amsterdam II
c.1732-19T>A	Intron 15	205	Black	Revised Bethesda

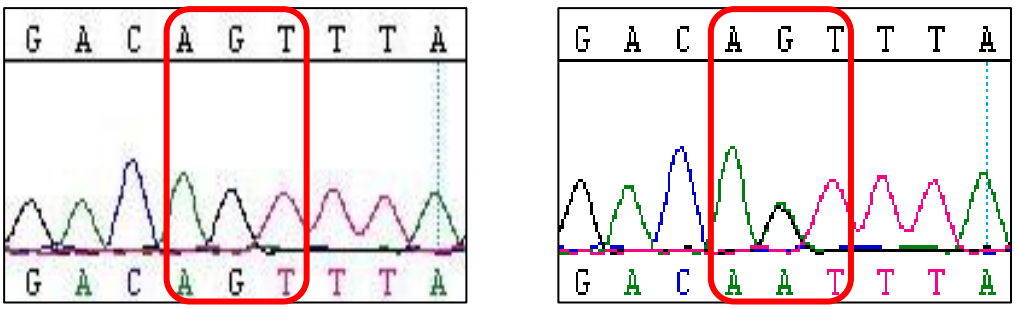


Figure 3.6: The *MLH1* c.885-1G>A novel variant located within the splice acceptor site of intron 10. The wildtype DNA sequence is shown on the left. The image on the right shows the heterozygous mutation that was found in patient 81.

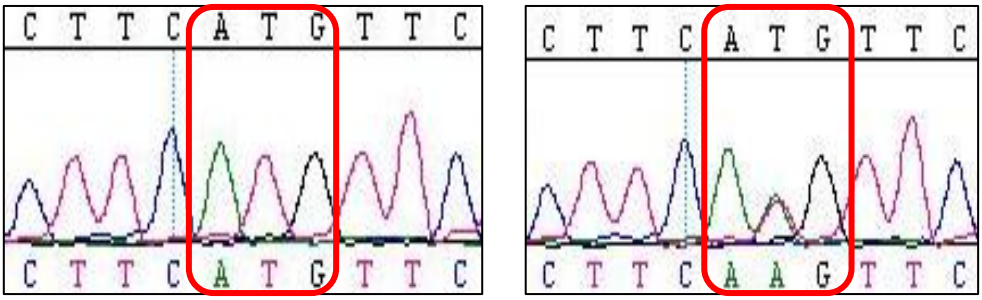


Figure 3.7: The *MLH1* c.1732-19T>A novel variant that was detected in intron 15. The wildtype DNA sequence is seen on the left and the variant is seen on the right

3.2.2. Variants Detected in *MSH2*

Eleven variants were detected in *MSH2* using the Sanger sequencing technique. Two of these variants were found within exons (c.573C>T and c.1811C>A) while one was found upstream in the regulatory region (c.-185C>A). The other eight variants were located within the introns of the *MSH2* gene. One of the variants in *MSH2* was the known positive control, which was a pathogenic mutation (Described in Section 3.2.2.1). A second variant was found to be previously reported as pathogenic. Five of the variants detected in *MSH2* were previously reported as non-pathogenic while four variants were found to be novel.

3.2.2.1. Positive Control

The c.1811C>A missense mutation (p.Ala604Asp) had previously been detected in exon 12 of *MSH2* in white patient 150, who fulfilled the Amsterdam II criteria. This variant was detected by the Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury and has only been reported by this group. This variant was characterised as a pathogenic mutation following IHC evidence, bioinformatics tools and disease co-segregation analysis using the patient's sibling, who was also affected by colon cancer.

The patient was included to validate the technique of Sanger sequencing. Figure 3.8 shows the electropherogram for the c.1811C>A variant that was detected in patient 150. The report by the Wessex Regional Genetics Laboratory indicated that the patient carried a heterozygous mutation. However, Figure 3.8 shows a homozygous mutation and this is discussed in Section 4.3.3.1.

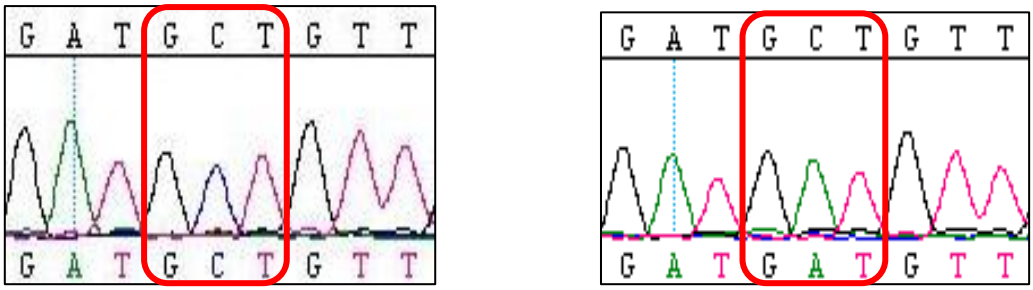


Figure 3.8: The *MSH2* c.1811C>A novel variant in exon 12. This missense mutation was detected by the Wessex Regional Genetics Laboratory and was included in this study as a positive control. The image on the left shows the wildtype DNA sequence and the image on the right shows the variant.

3.2.2.2. Previously Reported Non-Pathogenic Variants in *MSH2*

In *MSH2*, there were five variants that were previously reported as non-pathogenic. The c.573C>T (p.Leu191Leu) variant is located within exon 3 of *MSH2* but has been reported as a non-pathogenic synonymous mutation (Auclair et al., 2006). The other non-pathogenic variants were reported as benign and will not be discussed further in this section but are summarized in Table 3.6.

Table 3.6: Summary of the previously reported non-pathogenic variants in *MSH2* in the South African cohort

Variant	Location	Reference	Comment
c.211+9C>G	Intron 1	Weber et al., 1997 Wang et al., 2006 Capellá et al., 2008	Non-pathogenic
c.573C>T (p.Leu191Leu)	Exon 3	Farrington et al., 1998 Auclair et al., 2006	Non-pathogenic synonymous substitution
c.942+16_942+18delA_{13}	Intron 5	Weber et al., 1997	Non-pathogenic
c.1511-9T>A	Intron 9	Borresen et al., 1995	Non-pathogenic
c.1661+12G>A	Intron 10	Wijnen, Fodde and Khan, 1994 Farrington et al., 1998 Wang et al., 2006 Tournier et al., 2008 Park et al., 2008	Non-pathogenic

3.2.2.3. Previously Reported Pathogenic Variants in *MSH2*

Only one previously reported pathogenic variant was detected in *MSH2* and is summarised in Table 3.7. The mutation electropherogram can be seen in Figure 3.9. The implications of these pathogenic mutations on the families are discussed in Section 4.1.1.

Table 3.7: Summary of the reporting history for the previously reported pathogenic mutation in *MSH2* detected in this study

Variant	Location	Patient and Ethnicity	Reference	Comment
c.2006-6T>C	Intron 12	190, White	Brentnall et al., 1995 Farrington et al., 1998 Beiner et al., 2006 Wang et al., 2006 Capellá et al., 2008 Christensen et al., 2008 Tournier et al., 2008	Aberrant splicing of exon 13 (Brentnall et al., 1995; Tournier et al., 2008) Increased risk of endometrial cancer (Beiner et al., 2006)

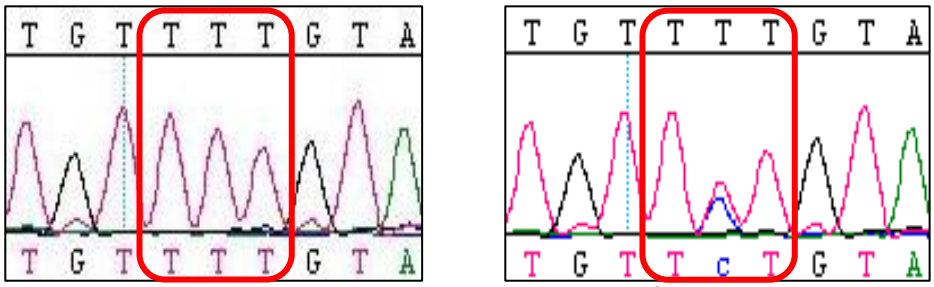


Figure 3.9: The *MSH2* c.2006-6T>C pathogenic variant of intron 12. The wildtype DNA sequence for the region is seen on the left. The image on the right shows the heterozygous variant

3.2.2.4. Novel Variants Detected in *MSH2*

Four novel variants were detected in *MSH2* in this study. As in 3.2.1.3, these novel variants have an unknown contribution toward disease but Section 3.3 will describe the elucidation of their pathogenicity. Table 3.8 shows a summary of the novel variants detected in *MSH2*. The electropherograms for each variant can be seen in Figures 3.10-3.13.

Table 3.8: Summary of the novel variants detected in *MSH2* in this study

Variant	Location	Patient	Ethnicity	Inclusion Criteria
c.-185C>A	Upstream	205	Black	Revised Bethesda
c.367-16T>A	Intron 2	61	White	Amsterdam II
c.1662-43C>T	Intron 10	104	Black	Revised Bethesda
c.1662-10C>T	Intron 10	235, 238	Black, Black	Revised Bethesda (Both)

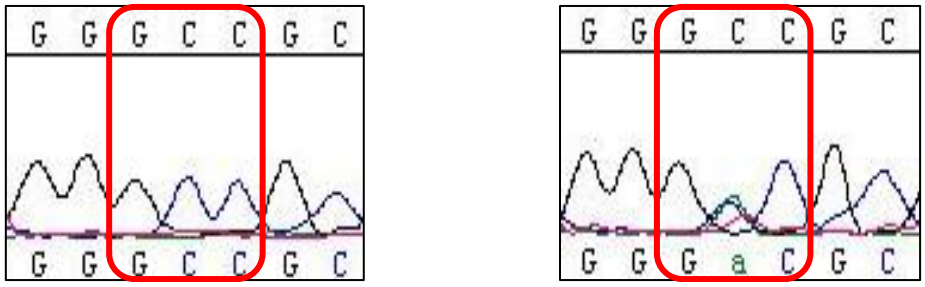


Figure 3.10: The *MSH2* -185C>A novel variant in the regulatory region. The image on the left shows the wildtype DNA sequence and the image on the right shows the variant sequence found in patient 205

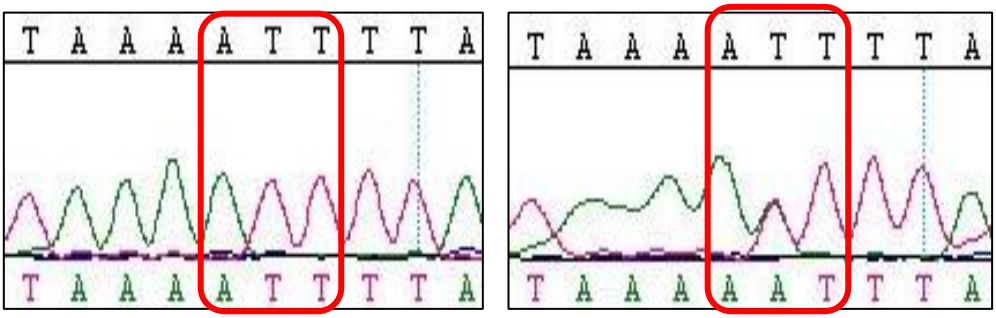


Figure 3.11: The *MSH2* c.367-16T>A novel variant of intron 2. The image on the left shows the wildtype DNA sequence and the image on the right shows the heterozygous variant in patient 61.

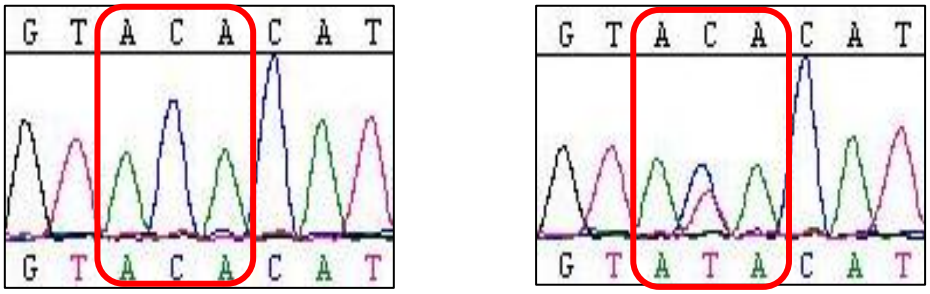


Figure 3.12: The *MSH2* c.1662-43C>T novel variant located within intron 10. The image on the right shows the variant as it was detected in patient 104 and the image on the left shows the wildtype DNA sequence.

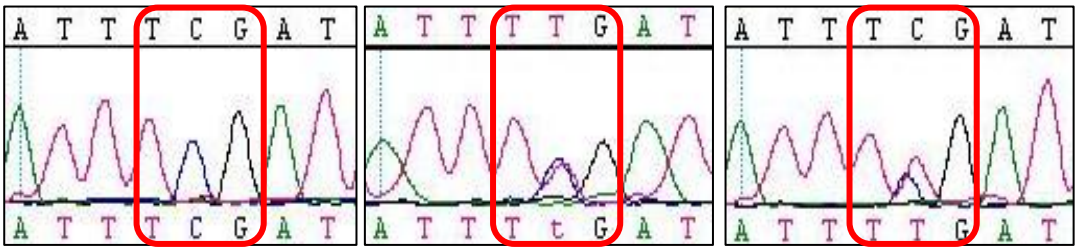


Figure 3.13: The *MSH2* c.1662-10C>T novel variant located within intron 10. The image on the left shows the wildtype DNA sequence. The images in the middle and on the right show the variants that were detected in two different black African patients, 235 and 238, respectively.

3.2.3. Summary of Variants Detected by Sanger Sequencing

In summary, 22 variants were detected in *MLH1* and *MSH2* in this study. Four variants were previously reported as pathogenic, one of which was a positive control, and 12 were previously reported as non-pathogenic. The other six variants detected are novel variants that have an unknown contribution towards disease. The novel variants are given in Table 3.5 (*MLH1*) and Table 3.8 (*MSH2*). Since these variants had an unknown contribution toward disease, it is important to investigate these variants further.

3.3.Determining Pathogenicity of the Novel Variants

The previously reported variants had known functions since they were previously found and their contribution toward disease had been determined. This study found six novel variants that had an unknown effect and further investigation was undertaken. The allele frequency of the novel variants was investigated using ethnically match, random control samples. Four variants were investigated using RFLP and two with HRM. The bioinformatic tool FASTSNP was used to investigate the upstream *MSH2* c.-185C>A variant.

3.3.1. Restriction Fragment Length Polymorphisms

The allele frequencies for four variants were determined using restriction enzymes. A summary of the allele frequencies of each variant from the RFLP assays is given in Table 3.9. Each individual variant is described below with regard to their pathogenicity as determined in this study.

Table 3.9: Summary of the allele frequencies that were determined by RFLP for each variant. If variants were not detected in controls, the allele frequency would remain as 1.00. The *MSH2* c.1662-10C>T was in both the heterozygous and homozygous state for the T allele, in control samples

Gene	Variant	Ethnic Group	Number (Controls)	Allele Frequency			
				A	C	G	T
MLH1	c.885-1G>A	White	45	0.00	-	1.00	-
MSH2	c.-185C>A	Black	45	0.00	1.00	-	-
MSH2	c.1662-43C>T	Black	45	-	1.00	-	0.00
MSH2	c.1662-10C>T	Black	45	-	0.66	-	0.34

3.3.1.1. *MLH1* c.885-1G>A variant

The *MLH1* c.885-1G>A variant was screened in patient 81 as well as 45 random white patients using the *HpyCH*4III restriction enzyme. The presence of this variant in patient 81 was confirmed using RFLP analysis and the variant was not detected in any of the 45 random white samples that were screened. Figure 3.14 shows the gel picture of the DNA ladder, patient 81 and the no cut control.

The *MLH1* c.885-1G>A variant is located on the splice acceptor site of exon 9 in *MLH1* and was not detected in control samples. Based on this evidence, the *MLH1* c.885-1G>A variant is likely to be pathogenic. The implication of this finding on the family is discussed further in Section 4.1.1.

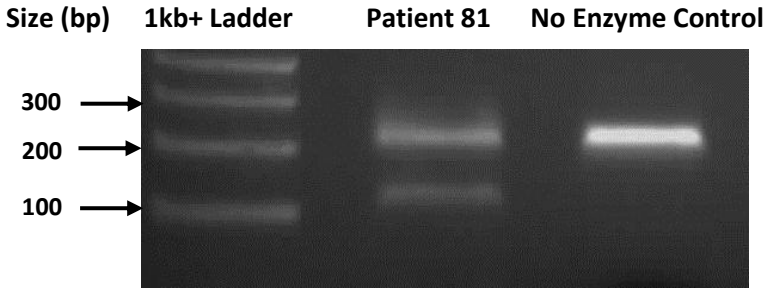


Figure 3.14: RFLP analysis with *HpyCH*4III performed on patient 81. This gel picture shows that patient 81 was heterozygous for c.885-1G>A as the presence of two alleles is confirmed. The 'no enzyme control' was included to ensure no contamination was present

3.3.1.2. *MSH2* c.-185C>A variant

The *MSH2* c.-185C>A variant was investigated using the *Mlu*I restriction enzyme after it was detected in patient 205. This variant was not detected in any of the random black samples tested (Table 3.9). The low allele frequency (C=1.00; A=0.00) shows that the variant is unlikely to be polymorphic, and may be pathogenic. The bioinformatics tool FASTSNP was used to determine whether this variant may have a putative functional effect (Section 3.3.3.

3.3.1.3. *MSH2* c.1662-43C>T variant

The *MSH2* c.1662-43C>T was detected in the black patient 104, who was previously shown to have the pathogenic *MLH1* c.454-13A>G mutation. This variant in *MSH2* was screened using *Rsa*I in 45 random black samples. This variant was not found in any of the random black samples (Table 3.9). This is suggestive that this variant is non-pathogenic.

3.3.1.4. *MSH2* c.1662-10C>T variant

The *MSH2* c.1662-10C>T variant was detected in two black individuals (235 and 238) and was investigated further using *Taq*I to determine the allele frequency of this variant. The c.1662-10C>T variant was found to be common (C=0.66; T=0.34) in the black population. It must be stated that the allele frequency seen here may not reflect homozygotes as the efficiency of TaqI was reduced due to the low incubation temperature. However, it is shown that this variant is polymorphic as the enzyme has cut a number of random samples. This strongly suggests that this variant is non-pathogenic. Figure 3.15 shows the RFLP gel picture that was obtained for patients 235 and 238, which confirmed the presence of the variant.

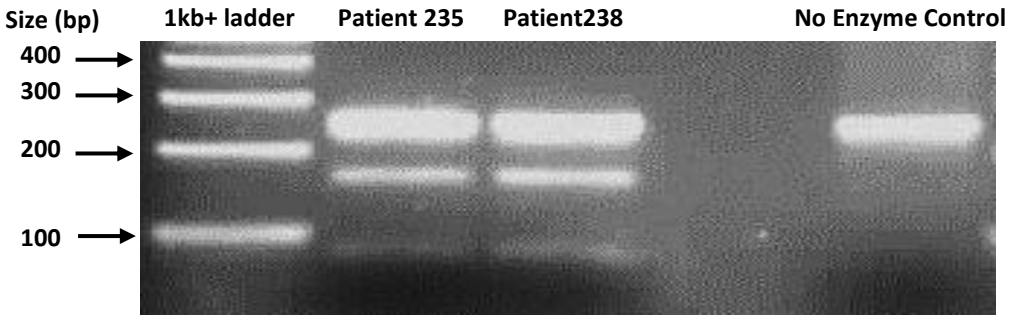


Figure 3.15: RFLP gel picture from using *Taq*I to investigate the *MSH2* c.1662-10C>T variants in patients 235 and 238.

3.3.2. High Resolution Melt Analysis

Two novel variants (*MLH1* c.1732-19T>A and *MSH2* c.367-16T>A) were investigated using the HRM technique. RFLP assays could not be designed for these two variants as no restriction enzymes corresponded to their positions. HRM was used to determine the allele frequencies of these two variants, and the results are summarised in Table 3.10.

3.3.2.1. *MLH1* c.1732-19T>A variant

The *MLH1* c.1732-19T>A variant was detected in the black patient 205 and was investigated using HRM. Comparing the HRM normalised melting curves of the 25 random black samples with the control samples, it was possible to determine the allele frequencies (T=0.94, A=0.06) as three heterozygous control samples were found to have this allele. This indicates that this variant may be common in the African population. The HRM melting curve with the controls and one heterozygous and one homozygous sample can be seen in Figure 3.16.

3.3.2.2. *MSH2* c.367-16T>A variant

The *MSH2* c.367A>C variant was detected in the white patient 61 and was investigated using the HRM technique. Unfortunately, the control samples did not produce distinct normalised melting curves, making analysis of variants impossible. This result is discussed in Section 4.4.2.2 below. There are no allele frequency data for this variant, which means that this variant has unknown significance in the white population.

Table 3.10: Summary of allele frequencies for the two variants that were investigated using HRM. Three heterozygous control samples were detected among the control samples (3/50=0.06)

Gene	Variant	Ethnic Group	Number (Controls)	Allele Frequency			
				A	C	G	T
MLH1	c.1732-19T>A	Black	25	0.06	-	0.94	-
MSH2	c.367-16T>A	White	25	-	-	-	-

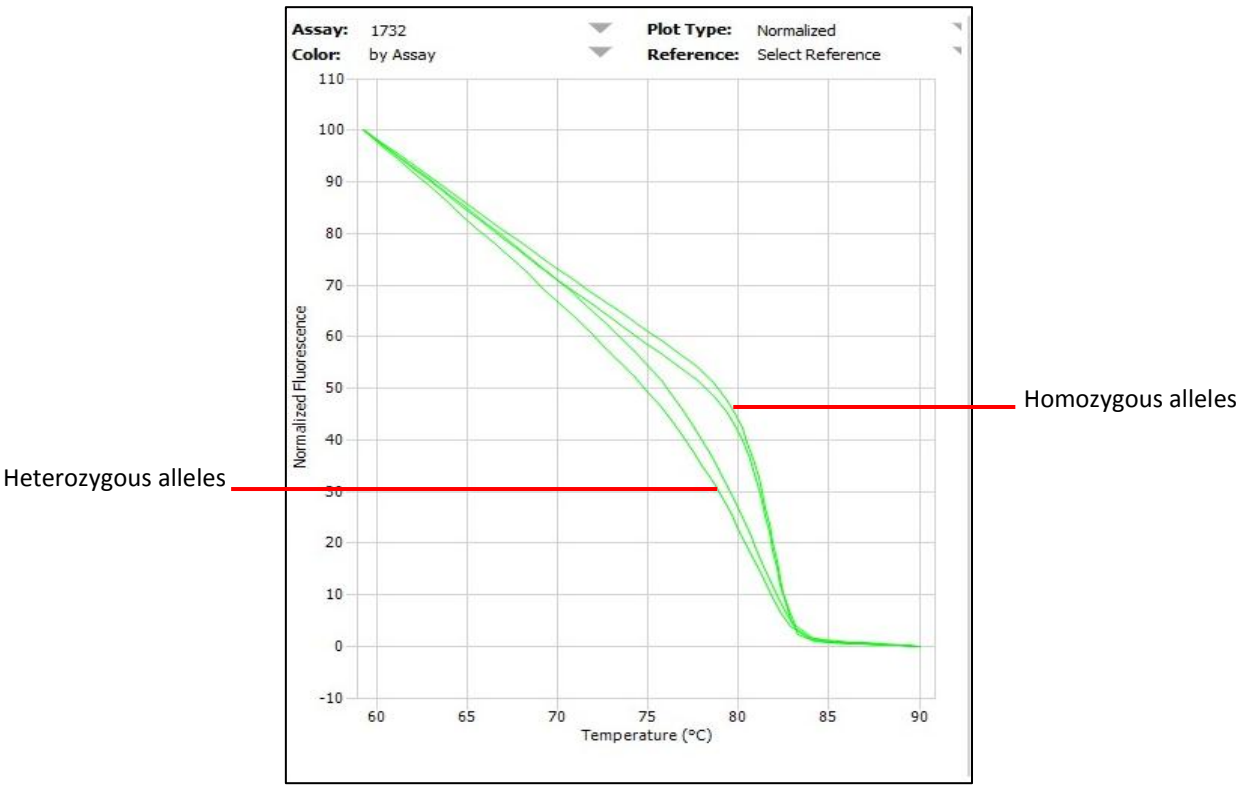


Figure 3.16: The HRM normalised melting curve for the assay of *MLH1* c.1732-19T>A variant. The top curve shows the normalised curve for the homozygous control and random samples. The bottom curve shows the normalised curve for the heterozygous control and random sample

3.3.3. FASTSNP Analysis

The *MSH2* c.-185C>A was investigated using the FASTSNP bioinformatics tool which would allow for the prediction of changes to transcription factoring binding sites. FASTSNP predicts that there would be no change to transcription factor binding at the position of the variant. Table 3.11 shows the FASTSNP result for the *MSH2* c.-185C>A variant, which indicates that this variant may not have a functional effect and may not be pathogenic.

The *MSH2* c.-185C>A variant was found to have a low frequency in the control samples and FASTSNP indicates that this variant has no functional effect. This variant is considered to have an unknown contribution toward disease.

Table 3.11: Adaptation of the FASTSNP results for the *MSH2* c.-185C>A variant. This tool predicts that neither allele will bind transcription factors and as such predicts that there will be no change to transcription factor binding. The entry column will name the transcription factors that bind to that particular sequence over the allele. Since the entry column does not change, it indicates that the variant has no effect (Yuan et al., 2006)

Transcription Regulatory (Predicted result of transcription factor binding sites change: No)		
TFSEARCH		
Allele	**Transcription Factor Binding Sites**	**Entry**
C	AGCGTGCGCGGGAAGCTGGG**C**CGCGTCTGCTTATGATTGGT	
A	AGCGTGCGCGGGAAGCTGGG**A**CGCGTCTGCTTATGATTGGT	

3.3.4. Summary of Investigation of Pathogenicity for Novel Variants

Elucidation of the pathogenic status for the six novel variants was undertaken in this study. A summary of the determined status is given in Table 3.12. Only the *MLH1* c.885-1G>A variant is considered to be pathogenic. The *MSH2* c.-185C>A variant and the *MSH2* c.367-16T>A variant have an unknown contribution toward disease while the other three variants are likely to be non-pathogenic.

Table 3.12: Summary of the pathogenic status for the seven novel variants detected in the South African cohort

Gene	Variant	Location	Status
MLH1	c.885-1G>A	Intron 10	Pathogenic
MLH1	c.1732-19T>A	Intron 15	Non-pathogenic
MSH2	c.-185C>A	Upstream	Unknown
MSH2	c.367-16T>A	Intron 2	Unknown
MSH2	c.1662-10C>T	Intron 10	Non-pathogenic
MSH2	c.1662-43C>T	Intron 10	Non-pathogenic

3.4.Characterising the Mutation Spectrum

In total, 22 variants were detected in the MMR genes *MLH1* and *MSH2* in a cohort of 20 patients. 11 of these variants were in *MLH1* and 11 were in *MSH2*. There were six novel variants that have not been previously published or reported. In *MLH1*, two novel variants were detected (c.885-1G>A and c.1732-19T>A). In *MSH2*, there are four novel variants (c.-185C>A, c.367-16T>A, c.1662-43C>T and c.1662-10C>T). A positive control with a known mutation in *MSH2* c.1811C>A was included in the study. Nine of the variants found in this study were present in more than one patient, indicating that 13 variants were detected in only one individual each.

3.4.1.1. Variants Separated by Ethnicity

The variants were separated by ethnicity to determine if there were any variants that were more common in a specific ethnic group. Any individual that fulfilled the Amsterdam II criteria was included in this study, which led to 12 white patients and one mixed ancestry patient (N=13). Due to

the paucity of family history, black and mixed ancestry patients were included using the revised Bethesda guidelines, which led to six black patients and one mixed ancestry patient being included in this study (N=7). This could show which variants are common in the South African white population and which are common in the black population of South Africa. The *MLH1* c.381-41A>G and *MSH2* c.1662-10C>T variants were in more than one black patient and not in the white or mixed ancestry patients. Table 3.13 gives a summary of all the variants detected in this study as well as the ethnic group of the patients each variant was detected in.

Table 3.13: Summary of the variants and the ethnic groups they were detected in. All these variants were detected as heterozygotes, or were known to be heterozygous in the case of *MSH2* c.1811C>A.

Gene	Variant	Observed	Ethnicity		
			Black (N=7)	White (N=11)	Mixed Ancestry (N=2)
MLH1	**c.381-41A>G**	2	2		
MLH1	**c.454-51T>C**	1	1		
MLH1	**c.454-13A>G**	1	1		
MLH1	**c.655A>G**	5	1	4	
MLH1	**c.731-734delGTTA**	1		1	
MLH1	**c.885-1G>A**	1		1	
MLH1	**c.1039-8T>A**	3	1	2	
MLH1	**c.1039-7delT**	4	2	2	
MLH1	**c.1558+14G>A**	1			1
MLH1	**c.1668-19A>G**	9	5	4	
MLH1	**c.1732-19T>A**	1	1		
MSH2	**c.-185C>A**	1	1		
MSH2	**c.211+9C>G**	7	4	3	
MSH2	**c.367-16T>A**	1		1	
MSH2	**c.573C>T**	1	1		
MSH2	**c.942+16_942+23del A_{13}**	1	1		
MSH2	**c.1511-9T>A**	1		1	
MSH2	**c.1661+12G>A**	7	5	2	
MSH2	**c.1662-43C>T**	1	1		
MSH2	**c.1662-10C>T**	2	2		
MSH2	**c.1811C>A**	1		1	
MSH2	**c.2006-6T>C**	1		1	
Total		53	29	23	1

Pathogenic mutations: Red Blocks
Non-Pathogenic variants: Green Blocks
Unknown contribution variants: Blue Blocks

A total of 53 variants were detected, making up 22 different variants. The black study cohort (N=7) harboured 29 variants (54%) while 24 variants (45%) were observed in the white individuals (N=11) and the mixed ancestry individuals (N=2) were found to have one variant (1%). The *MLH1* c.1668-19A>G variant was the most commonly detected variant in the entire cohort (45% of cohort) and

was always detected in patients with the *MLH1* c.655A>G variant. The next most commonly detected variants were *MSH2* c.211+9C>G (35%) and *MSH2* c.1661+12G>A (35%).

Five pathogenic mutations (*MSH2* c.1811C>A, *MLH1* c.885-1G>A, *MLH1* c.454-13A>G, *MLH1* c.731_734delGTTA and *MSH2* c.2006-6T>C) have been identified in this cohort, but all are unique to one patient (Shown in Table 3.14). Therefore, there does not appear to be any common, disease causing mutations in the South African population. However, there are common, non-pathogenic variants among both black and white patients.

Table 3.14: Summary of patients and the pathogenic mutations found in this study

Ethnic Group	Pt. #	Age	Location of Tumour	Inclusion Criteria Met	Pathogenic Mutations	
					MLH1	*MSH2*
Black	104	49	Unknown (3rd recurrence)	Revised Bethesda Guidelines	c.454-13A>G	
Black	205	28	Sigmoid colon	Revised Bethesda Guidelines		
Black	235	21	Rectum	Revised Bethesda Guidelines		
Black	238	49	Rectum	Revised Bethesda Guidelines		
Black	247	47	Rectum (Recurrence)	Revised Bethesda Guidelines		
Black	267	47	Rectum	Revised Bethesda Guidelines		
Black	290	48	Rectum	Revised Bethesda Guidelines		
White	32	50	Sigmoid colon	Amsterdam II Criteria		
White	42	64	Unspecified (colectomy)	Amsterdam II Criteria		
White	52	37	Upper gastro-intestinal tract	Amsterdam II Criteria		
White	61	67	Distal colon	Amsterdam II Criteria		
White	81	29	Proximal colon	Amsterdam II Criteria	c.885-1G>A	
White	86	40	Transverse colon (recurrence)	Amsterdam II Criteria		
White	89	49	Proximal colon	Amsterdam II Criteria		
White	136	56	Transverse colon	Amsterdam II Criteria	c.731_734del GTTA	
White	150	44	Sigmoid colon	Amsterdam II Criteria		c.1811C>A
White	190	50	Duodenum	Amsterdam II Criteria		c.2006-6T>C
White	253	28	Rectum	Amsterdam II Criteria		
Mixed Ancestry	13	45	Unspecified (Colon)	Amsterdam II Criteria		
Mixed Ancestry	175	41	Rectum	Revised Bethesda Guidelines		

Pt. #= Patient Number
Age given is Age of Diagnosis

4. Discussion

This chapter will discuss the above results in relation to the aim of determining the mutation profile of HNPCC in the South African population. Section 4.1 will provide a discussion on patient collection using the different inclusion criteria and the variant profile that was found in each group. Section 4.1 will also discuss possibilities as to why some patients that fulfilled the inclusion criteria were not found to have mutations. Section 4.2 will discuss the MLPA results and whether MLPA testing is necessary in the South African population. Section 4.3 will provide a discussion on the results that were found using Sanger sequencing. This will include a discussion on the known and novel variants as well as a discussion on the allele frequency analyses. Finally, section 4.4 will discuss the limitations of this study and the possible studies that can be undertaken in the future.

4.1. Patients Fulfilling Inclusion Criteria

All patients that fulfilled the Amsterdam II criteria were included in this study. Since more white patients in the DNA bank had extensive family histories, those who were not Amsterdam II positive could be excluded. This could have created a sampling bias, as the Amsterdam II criteria are more powerful criteria for detecting HNPCC (Lynch et al., 2006). It was found that 27% of Amsterdam II criteria positive patients in this study (not including the positive control) harboured pathogenic mutations. This is in contrast to the 13% of revised Bethesda patients that harboured pathogenic mutations.

The revised Bethesda guidelines were used to select black patients as none of the black patients available fulfilled the Amsterdam II criteria. Figures G.1 to G.3 in Appendix G show the pedigrees for three of the black patients (205, 235 and 238). When compared to the Amsterdam II positive pedigrees seen in Figure G.4 and G.5, it becomes noticeable that family history is not well known in these families. Since the family history among black patients was generally less significant, the revised Bethesda guidelines may be the most ideal criteria for selecting black patients for molecular diagnostic screening.

The poor family history among the black patients may be attributed to several factors. A major factor would be that many black families may have had a diet rich in grains, fruits and vegetables, which would have lowered their risk of CRC (Carrera-Bastos et al., 2005). Exposure to the Western diet will have increased their risk for CRC (Cunningham et al., 2010). Other factors for the poor family history include language barriers, fears of isolation or prejudice and reluctance to follow-up after the medical diagnosis (Cronjé et al., 2009).

The average age of CRC diagnosis in this study for black patients was 41, whereas the average age for white patients was 46. This indicates that black patients who develop CRC are younger than white patients. However, the average ages may have been affected by the different inclusion criteria

as no black patients were older than 50 years. CRC development under the age of 50 is more suggestive of a genetic predisposition (Anderson et al., 2003; Lynch et al., 2006). This is an important consideration for future research and molecular diagnostic screening services as it could allow improved research and clinical management.

Patient 104 was included in this study as their medical history showed that they had a third recurrence of colon cancer prior to the age of 50 and no family history was available. The previously reported pathogenic variant *MLH1* c.454-13A>G was detected in this individual. The value of using the revised Bethesda guidelines has been shown as this family can benefit from targeted screening in the future.

This study has demonstrated that both the Amsterdam II criteria and the revised Bethesda guidelines may be effective at identifying patients with MMR gene mutations, but a larger cohort would be needed to establish truly accurate detection rates.

4.1.1. Clinical Management of HNPCC Patients

As mentioned in Section 1.6.4, identification of patients with HNPCC and Lynch syndrome is important as it allows for improved clinical management of patients and their families through predictive testing (Lynch et al., 2006; Boland and Goel, 2010). Patients in this study were found to harbour pathogenic mutations (Patients 81, 104, 136, 150 and 190) and will need to be managed in order to limit mortality. Since these patients consented to research, their samples will need to be screened by the Molecular Diagnostic Division in the Department of Human Genetics to verify the mutations and ensure they do not report a false positive back to the patient.

These families will need to be counselled as their increased risk of CRC, and other associated cancers, will need to be explained. Since a mutation is known, family members can have targeted screening of the mutation to determine their risk status. High risk individuals can opt for preventative colectomy or regular screening to detect the cancer early (Etzioni et al., 2003; Lynch et al., 2006; Boland and Goel, 2010).

Previous studies on HNPCC in South Africa have largely been conducted in the Western and Northern Cape provinces, where the *MLH1* c.1528C>T mutation has been characterised in the Mixed Ancestry population (Ramesar et al., 2000; Stupart et al., 2009; Wentink et al., 2010; Hitchins et al., 2011). This group have been able to show the importance of targeted screening and improved clinical management in families with a known pathogenic mutation (Goldberg et al., 1998). Improved clinical management of HNPCC should be present throughout South Africa and not limited to one or two provinces or families.

4.2.MLPA Analysis

The MLPA analysis on this cohort showed that there were no large deletions or duplications detected in this study. At first glance, it may seem that MLPA may not be a worthwhile technique to perform in the South African population as none of these patients had any large deletions or duplications. However, the number of patients used in this study is small and a bigger sample size may have resulted in a better indication of the rate of detection of large deletions or duplications.

Gille et al. (2002) performed the MLPA technique on 126 CRC families and they found 17 families that contained large exon deletions, of which, two families had complete deletions of *MLH1*. This shows that less than 15% of the families used in this larger study had large deletions that MLPA could detect. Boland and Goel (2010) estimate that 33% of all the mutations detected in *MSH2* are large exon deletions. Considering this information, MLPA should continue to be performed in larger cohorts.

The complementary detection strategy of MLPA and sequencing has been effective at detecting mutations (Lynch et al., 2006). The MLPA technique should continue to be used in future studies in the South African population unless future evidence suggests that large deletions are very rare or not present in this population.

4.3.Variants Detected using Sanger Sequencing

Sanger sequencing was used to detect small insertions, deletions and substitutions within the DNA of the patients that were screened in this study. In total, 53 variants were detected among 20 patients. This number was made up of 22 different variants. One of the variants was a known pathogenic mutation in a positive control. 14 variants had been published previously (Tables 3.3, 3.4, 3.6 and 3.7) while one has been reported in the curated LOVD open database but has not been published (InSiGHT, 2011). Six of the variants (two in *MLH1* and four in *MSH2*) detected in this study are novel variants, reported for the first time in this study and whose possible functional effects were elucidated in this study (Table 3.12).

The novel variants were detected using the forward primer, and then confirmed using the reverse primer. This is the methodology that was adopted for this study as it is the method used by the Molecular Genetics Laboratory at the NHLS, Braamfontein to detect pathogenic variants. This method does not account for PCR synthesis errors which could be carried over into Sanger sequencing and may give a false positive result. Mutations thus need to be confirmed on a new PCR amplicon. Future studies should be mindful to perform both bidirectional sequencing and repeat the entire procedure on a patient found to have a variant.

4.3.1. Number of Variants in Study Cohort

Excluding the positive control, 17 of 19 patients were found to have variants (89%). Of these patients, three were found to have previously reported pathogenic mutations and one novel pathogenic mutation was found in this study. This shows that four of 19 patients were found to have pathogenic mutations (21%). The patients with pathogenic mutations are considered to have Lynch syndrome.

This detection rate is higher than some previous similar studies that were performed in Asians and Caucasians (Shin et al., 2004; Woods et al., 2010, Wei et al., 2011) but, excluding the known mutation, is similar to a number of other studies (Weber et al., 1997; Farrington et al., 1998; Weber et al., 1999; Gille et al., 2002; Wang et al., 2006). These detection rates can be seen in Table 4.1. These studies screened *MLH1* and *MSH2* for mutations using sequencing. In some studies, MLPA was used to complement sequencing, or was the sole technique for detecting mutations (Gille et al., 2002; Woods et al., 2010). The similarity in detection rates among similar studies shows that the research performed in this study is consistent with other research groups.

Table 4.1: Summary of mutation detection rates from previous studies which were investigating *MLH1* and *MSH2*. The studies given below used similar inclusion criteria for their patients, thus making comparisons easier

Study	Population Group	Inclusion Criteria	Cohort Size	Detection Rate
Weber et al., 1997	Caucasian, African American	Amsterdam II, Bethesda guidelines	40	25%
Farrington et al., 1998	Caucasian	Amsterdam II, revised Bethesda	76	30%
Weber et al., 1999	African American	Amsterdam I, Revised Bethesda	11	27%
Gille et al., 2002	Caucasian	Amsterdam II, Bethesda guidelines	126	22%
Shin et al., 2004	Korean	Amsterdam II, Revised Bethesda	164	13%
Wang et al., 2006	Chinese	Amsterdam II	31	22%
Woods et al., 2010	Caucasian	Amsterdam II, Revised Bethesda, FAP criteria	750	2.7%
Wei et al., 2011	Chinese	Amsterdam II, Bethesda guidelines	98	11%
This study	White, Black and Mixed Ancestry	Amsterdam II, Revised Bethesda	20	21%

There were 29 variants detected in the seven black patients, with each of these patients having at least two variants detected. Patient 205 was found to harbour the most variants, with seven variants, while patient 267 was found to have two variants, both in *MSH2*. Weber et al. (1999) also

found a high percentage (90%) of variants in African American patients and this study is consistent with that finding. The high variant rate is not surprising since the African population has been shown to have greater genetic diversity than whites and Asians (Campbell and Tishkoff, 2008).

4.3.2. Previously Reported Variants in *MLH1*

11 variants were found in *MLH1*. Seven variants were previously reported as non-pathogenic, two were previously reported as pathogenic and two were novel variants. Generally, there is consensus as to whether a previously reported variant is pathogenic or benign. However, the *MLH1* c.655A>G variant will be discussed as some authors suggest a possible role in carcinogenesis, despite being described as non-pathogenic by many other studies. The *MLH1* c.731_734delGTTA mutation will also be discussed as it was previously reported by the South African researcher, Prof. Raj Ramesar.

4.3.2.1. The *MLH1* c.655A>G (pIle219Val) variant

The *MLH1* c.655A>G (p.Ile219Val) mutation is a missense mutation that was found in five patients (four Amsterdam II positive whites and one Bethesda positive black) in this study. The *MLH1* c.655A>G variant been reported in many other individuals in different populations and has been considered benign (Farrington et al., 1998; Ward et al., 2002). However, other studies indicate that this variant is not completely benign (Fredriksson et al., 2006; Takahashi et al., 2007 Campbell et al., 2009).

The *MLH1* p.Ile219Val mutation has been linked to prostate cancer and an increased risk of ulcerative colitis (Bagnoli et al., 2004; Fredriksson et al., 2006). Furthermore, it has been indicated that the p.Ile219Val mutation leads to a lower level of MLH1 protein expression, approximately 60% of the wild-type level (Takahashi et al., 2007). Campbell et al. (2009) found that this variant was associated with CRC in individuals that reported a high intake of Western diets. Patients who are identified with this variant should be informed of the elevated risk of ulcerative colitis and cancer in conjunction with a Western diet. These findings also show that a variant should be investigated fully to properly determine the role of the variant in carcinogenesis.

4.3.2.2. The *MLH1* c.731_734delGTTA variant

The *MLH1* c.731_734delGTTA variant was identified in an Amsterdam II positive family in exon 9 of *MLH1*. The pedigree for this patient is given in Figure G.4 in Appendix G. This mutation leads to a non-functional protein, which indicates that it is pathogenic and is likely to be a significant factor in the development of CRC in this patient (InSiGHT, 2011).This mutation has been reported on the InSiGHT database by Prof. Raj Ramesar, from the University of Cape Town (UCT) in South Africa (InSiGHT, 2011).

It is not known if this variant was found in the same family or in a different family from that of Prof. Ramesar's. This variant has been detected by two different South African groups, which could imply that the same family was tested, or that a possible common/founder mutation exists in South Africa. Collaboration with the UCT group will be needed to identify patient pedigrees and to determine whether this mutation is localised to a family or is common in the population.

4.3.3. Previously Reported Variants in *MSH2*

11 variants were detected in *MSH2*, one of which was a positive control that was included to validate the Sanger sequencing method. The findings from the control sample will be discussed in more detail in Section 4.3.3.1. Five previously reported non-pathogenic variants were detected in *MSH2* in this study while a previously reported pathogenic mutation (*MSH2* c.2006-6T>C) was detected. This mutation will be discussed as there have been studies which indicate that it is benign and others which indicate that it is pathogenic. The synonymous mutation, *MSH2* c.573C>T, will be discussed briefly to demonstrate the potential role of synonymous mutations in disease.

4.3.3.1. The *MSH2* c.1811C>A (p.Ala604Asp) variant

The *MSH2* c.1811C>A missense mutation is located in Exon 12 of the *MSH2* gene and was used as a positive control. The pedigree for this patient can be seen in Figure G.5. This patient and their sibling both showed a loss of staining for MSH2 in IHC and were both affected by CRC. This variant was also found in both patients, which implies strongly that this is the pathogenic mutation associated with the disease. Both this patient and their sibling were tested by the Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury but the sibling was not included in this study.

This mutation was detected as a heterozygote by the Wessex Regional Genetics Laboratory but was found to be homozygous in this study (Figure 3.8). As a homozygote for MMR mutations would be expected to present with cancer in their childhood, the most likely explanation is that there has been allele dropout (Lynch et al., 2006). The single peak seen may be caused by only one allele amplifying, which indicates that one primer may not be binding to the template DNA. This may be caused by a polymorphism in the primer binding region. This suggests that further research needs to be conducted to determine which polymorphisms are present in the South African population, as these could alter the results of sequencing data and give false negatives in patients. This could lead to high costs in finding pathogenic mutations in other genes, or hamper family management if it is assumed that the CRC was sporadic because a pathogenic mutation could not be detected. The positive control in this study was unable to validate the sequencing method, but it was able highlight the importance of additional research into polymorphisms in the primer binding domains of these genes.

4.3.3.2. The *MSH2* c.2006-6T>C mutation

The c.2006-6T>C mutation of *MSH2*, which is located in intron 12, has also been reported previously (Farrington et al., 1998; Wang et al., 2006; Christensen et al., 2008; Tournier et al., 2008). This

mutation has also been reported as being non-pathogenic in HNPCC (Wang et al., 2006). However, Tournier et al. (2008) performed a functional ex vivo splicing assay using a pCAS vector that was transfected into HeLa cell lines. They found that this mutation leads to partial exon 13 skipping in the mRNA of *MSH2*. Beiner et al. (2006) found that this variant was also associated with an increased risk of endometrial cancer.

4.3.3.3. The *MSH2* c.573C>T (p.Leu191Leu) variant

The c.573C>T (p.Leu191Leu) variant was found in exon 3 of *MSH2* in black patient 205. This variant has been reported previously but has been shown to have no effect on the amino acid sequence in the MSH2 protein (Auclair et al., 2006). Although this variant is considered to be benign, it has been reported that synonymous mutations are not always benign and may play a role in pathogenicity (Parmley and Hurst, 2007).

Synonymous mutations can affect gene expression in a number of ways. Codon bias in the translation of mRNA has been observed, where the most widely expressed codons correlate with the most widely available tRNA species. The codon could be replaced by a less common tRNA species, which could affect the rate of mRNA translation. Furthermore, synonymous mutations may also lead to changes in mRNA stability, translation and degradation and alternative splicing of exons (Parmley and Hurst, 2007). Although this variant has been found to be non-pathogenic, it must be remembered that not all synonymous mutations are silent and this variant warrants further investigation.

4.4. Determining the Pathogenicity of the Novel Variants

The six novel variants were investigated further to determine whether they may have possible pathogenic effect. The allele frequency for six variants was determined using RFLP or HRM in random control samples (Section 3.3). In addition, FASTSNP was used to predict whether the variant upstream of the transcription start site in *MSH2* had a functional effect. The likely pathogenic status of each variant is summarised in Table 3.12. Since these variants have been detected for the first time in this study, they will be submitted to dbSNP at the NCBI and they will also be reported to the LOVD database. The allele frequencies obtained will also be submitted for these novel variants

4.4.1. The *MLH1* c.885-1G>A mutation

The *MLH1* c.885-1G>A variant was suspected of being pathogenic as it is located on the splice acceptor site of intron 10. Mutations in the splice acceptor site may abolish splice site recognition by the spliceosome, which would lead to aberrant splicing (Venables, 2004). The AG motif in splice acceptor sites is highly conserved and is important for correct splicing of the mRNA. Mutations in the splice acceptor site frequently leads to pathogenicity by abolishing the splice recognition site, which leads to aberrant splicing (Strachan and Read, 2004; Venables, 2004). Furthermore, a pathogenic mutation is known to occur at the *MLH1* c.885-2 position (Terdiman et al., 2001).

RFLP analysis was undertaken to determine the frequency of the variant in the random white population. The *MLH1* c.885-1G>A variant was confirmed in patient 81 but was not detected in any of the random samples, suggesting that the variant is rare in the white population. The evidence for this variant strongly suggests that it is pathogenic. The family of patient 81 can be screened for the presence of this mutation to determine whether they are at risk of tumourigenesis.

4.4.2. The *MLH1* c.1732-19T>A variant

The *MLH1* c.1732-19T>A variant was investigated using the HRM technique and was found in control samples. This technique uses the melting temperature to determine which allele is present at a particular position since the allele will alter the melting temperature (Gundry et al., 2003). HRM is not very specific and other variants within the PCR amplicons may affect the melting temperature of the amplicons, particularly in a large PCR product such as that used in this study (Gundry et al., 2003). Further investigations may be needed to validate the allele frequency findings for this variant and determine whether it is truly polymorphic.

4.4.3. The *MSH2* c.-185C>A variant

The *MSH2* c.-185C>A variant was investigated using both RFLP and bioinformatic prediction with FASTSNP. This variant could affect transcription by increasing or decreasing the affinity for a transcription factor in that region, which would ultimately have a functional effect (Strachan and Read, 2004). FASTSNP predicted that this variant does not impact the binding affinity of any transcription factors for either allele.

The variant allele was found to be absent (C=1.00, A=0.00) in the control population. This shows that this variant is probably not polymorphic in the South African population. These results do not provide any evidence on the functional effect of this variant. Therefore, this variant is considered to have unknown significance and further investigation will be needed to determine whether it is pathogenic or non-pathogenic. This will be important research to conduct as it will guide the family on future screening programs.

4.4.4. The *MSH2* c.367-16T>A variant

The *MSH2* c.367-16T>A variant was investigated using the HRM technique. The HRM results for this variant did not produce detectably different normalised melting curves for the known controls, which made analysis impossible. HRM is dependent on detecting the melting temperature of PCR amplicons, which is measured in real-time.

Unfortunately, PCR product size does affect the HRM analysis. A longer PCR product becomes more stable and the effect of a variant on the overall structure will become weaker as the size of the

product increases (Gundry et al., 2003). The PCR product used for analysing the *MSH2* c.367-16T>A variant was approximately 300bp in size, which was too large to yield any meaningful results with the HRM technique. However, this does confirm that the PCR product size is an important consideration in HRM and this mistake should be avoided in the future.

4.4.5. The *MSH2* c.1662-10C>T variant

The *MSH2* c.1662-10C>T variant was detected in two unrelated black individuals. This variant appears to be linked to the *MSH2* c.211+9C>G variant and the *MSH2* c.1661+12G>A variant, which are both non-pathogenic and previously reported (Wijnen et al., 1994; Weber et al., 1997). Although this evidence is suggestive that the variant is non-pathogenic, it was investigated further using RFLP analysis to determine the frequency of the variant.

This variant was found to be polymorphic in the black population of South Africa (C=0.66, T=0.34) (Table 3.9). This finding is highly suggestive that this variant is non-pathogenic. Interestingly, this variant appears to be very common in the black population of South Africa but has not been reported or published previously in other populations, to the best of our knowledge.

4.4.6. The *MSH2* c.1662-43C>T variant

The *MSH2* c.1662-43C>T variant was investigated using RFLP analysis in the random black population. This variant was not found (C=1.00, T=0.00), which shows that this variant is not likely to be polymorphic. This variant was detected in a patient that was found to harbour the previously reported pathogenic *MLH1* c.454-13A>G mutation.

The low allele frequency may indicate that the variant is pathogenic but it is not a reliable indicator of pathogenicity. If an allele is not common, it does not confirm that it is pathogenic. This patient was found to have a pathogenic mutation present, and it is improbable that a person carries a pathogenic mutation in both genes (Strachan and Read, 2004).

4.4.7. CRC in Patients with No Pathogenic Mutations

It is important to consider that not all of the variants that were found are pathogenic. Only four of the variants found in this study are pathogenic. There are two variants with an unknown contribution to disease. These unknown variants may be pathogenic but they may also have no functional impact. There are several reasons for disease progression in patients who were found not to have pathogenic mutations, or whose variants are benign.

The first reason is that they may not have had a pathogenic mutation in *MLH1* or *MSH2*, thus the pathogenic mutation would not have been detected in this study. Approximately 90% of all

mutations detected in HNPCC genes are found in *MLH1* and *MSH2*. Mutations in *MSH6* and *PMS2* account for the majority of the remaining 10% (da Silva et al., 2009). The mutation spectrum of HNPCC mutations has not been determined in South Africa, therefore it is possible that there are a higher proportion of mutations in other MMR genes.

A second reason that could explain why no pathogenic variants were detected in these patients is that there may not be mutations in the MMR genes. Approximately 30-50% of cases of suspected HNPCC, or families that fulfil the Amsterdam II criteria, do not have confirmed mutations in the MMR genes (de la Chapelle, 2004; Umar et al., 2004a; Hitchins et al., 2011). These patients may have mutations in other genes involved in CRC, such as *BRAF* or *KRAS*. However, it is important to remember that these patients may also have sporadic CRC. These patients may have developed sporadic CRC through lifestyle factors, including a more Western diet that is high in red meat and fats (Cunningham et al., 2010). Further studies and a larger cohort need to be used to determine the accurate hereditary CRC rates.

Germ line mutations in the 3' region of the *EPCAM* gene have been found to lead to HNPCC (Jasperson et al., 2010). Patients with these mutations have been found to have epigenetically silenced *MSH2* caused by hyper-methylation of the *MSH2* promoter region, which leads to loss of expression of *MSH2*, and subsequently CRC (Jasperson et al., 2010). Hyper-methylation of the *MLH1* promoter region has also been shown in cases of HNPCC and recent evidence suggests that constitutional epimutations may account for these cases. However, this mechanism of disease is rare but investigation in patients who are not found to harbour any MMR gene mutations may detect this specific type of mutation (Hitchins and Ward, 2009; Hitchins et al., 2011).

4.4.8. Molecular Diagnostic Screening of HNPCC Patients

Although pathogenic mutations were found in this study, a cheaper first line screening method would be more preferable than full gene sequencing. As was seen in this study, 21% of patients were found to have mutations in *MLH1* or *MSH2*. A screening procedure such as MSI testing or IHC should be offered to screen patients prior to gene sequencing. Patients who show MSI or loss of staining in IHC should be offered full gene sequencing and MLPA (Lynch et al., 2006).

Patients who are found to not have pathogenic mutations in the *MLH1* or *MSH2* genes may still need to be assessed to determine if they have constitutional epimutations, or mutations in other genes such as *MSH6, PMS2* or *EPCAM* (Lynch et al., 2006; Hitchens et al., 2011). IHC testing may be preferable as it can be used to direct screening in the MMR genes (Lynch et al., 2006).

4.5.Characterising the Mutation Spectrum

Fifty three variants were detected among 20 patients. This number was made up of 22 different variants that were detected among 20 patients. No common pathogenic mutations or founder mutations in the white or black population of South Africa were identified in this study. However, this study only tested 20 individuals, which is not enough to determine the complete mutation spectrum. There were a number of other common variants that were detected.

4.5.1. Mutation Spectrum of *MLH1*

The variant *MLH1* c.1668-19A>G (rs9876116) was detected in nine (five blacks and four whites) patients and is a common polymorphism in the South African population (A=0.66, G=0.34), and has been reported by other groups (dbSNP, 2012). All the patients that were found to have the *MLH1* c.655A>G (four white patients and one black patient) variant were also found to have the *MLH1* c.1668-19A>G variant, suggesting that these variants co-occur and may be in linkage disequilibrium, particularly in the white population. It is possible that an ancestral chromosome carried the *MLH1* c.1668-19A>G variant, and the *MLH1* c.655A>G variant arose later on this chromosome.

4.5.2. Mutation Spectrum of *MSH2*

The *MSH2* c.211+9C>G, *MSH2* c.1661+12G>A and *MSH2* c.1662-10C>T variants were all present in patients 235 and 238. It is possible that these variants may occur on one haplotype in the black population. However, the black patients 205 and 247, as well as the white patient 81, had the *MSH2* c.211+9C>G and the *MSH2* c.1661+12G>A variants. This suggests that the *MSH2* c.1662-10C>T could have arisen on this haplotype but it does seem that *MSH2* c.211+9C>G and *MSH2* c.1661+12G>A are in the same haplotype. Unfortunately, the families were not studied so it is challenging to assign a haplotype to each patient or be certain that these variants occurred on the same chromosome.

4.5.3. Mutation Spectrum with regard to Ethnicity

The *MLH1* c.381-41A>G variant and the *MSH2* c.1662-10C>T variant were only detected in black patients. The *MLH1* c.381-41A>G variant has been previously reported in the Slovakian population (Zavodna et al., 2006). This shows that this variant is not unique to black South Africans. The *MSH2* c.1662-10C>T variant has not been previously reported, yet is thought to be polymorphic among black South Africans after RFLP analysis revealed a high allele frequency (Table 3.9). All other variants that were detected in more than one patient were present in both black and white patients, showing that these variants are not localised to one ethnic group.

4.5.4. Mutation Spectrum of Pathogenic Mutations

In the four pathogenic mutations found in this study, none are common mutations and there does not seem to be any linkage with any other variants. Three of the pathogenic mutations were detected in white patients, and the fourth in a black patient. This is not surprising as the white patients fulfilled the Amsterdam II criteria, which is more stringent for the detection of HNPCC.

These findings show that MMR gene mutations occur in black and white race but more investigation is needed to determine the full spectrum of mutations in South Africa.

Determining the mutation spectrum could have significant cost saving benefits for patients. If a mutation is known to be common in a specific ethnic group, a diagnostic laboratory could attempt to find that mutation first with a targeted screening technique. This is akin to the targeted screening offered to family members of a known mutation carrier. If the common mutation is not present, the patient can then proceed with general, more expensive screening techniques such as sequencing. Unfortunately, a much larger research cohort will be needed to determine the full spectrum of mutations.

4.6.Limitations and Future Directions

This study collected 20 patients for analysis. Two of these patients were of mixed ancestry, seven were black and the remaining 11 were whites. As mentioned in Section 4.1, a selection bias was created during the selection of patients. White patients that did not fulfil the Amsterdam II criteria were excluded. This would increase the probability of mutations being found in white patients as the Amsterdam II criteria are more stringent criteria (Lynch et al., 2006). This also indicates that more patients included under the revised Bethesda guidelines would be needed to properly ascertain the mutation spectrum of the MMR genes in South Africa.

This study hoped to find common pathogenic mutations but none were found in this cohort, suggesting that there may not be any common or founder mutations. A mutation in this study was previously reported by Prof. Ramesar at UCT, which may be a possible common mutation or a mutation in a large family. A larger cohort size would have possibly allowed more common variants to be detected. This is particularly needed to determine the mutation spectrum within each ethnic group. This study did not focus on a specific ethnic group, which may have a different mutation spectrum to another ethnic group. Large numbers of each ethnic group should be included in a future study which could determine adequately the mutation spectrum for each ethnic group, which may reveal common or founder mutations which could be used in targeted screening.

One of the greatest challenges this study faced was ascertaining patients that fulfilled the inclusion criteria. Future efforts could focus on more intense communication or centralising the cancer patient database, which would allow for easier research into cancer related topics. A further limitation was the high cost of Sanger sequencing and MLPA, although MLPA costs per patient are reduced if more patients are present. A high throughput sequencing platform could be utilised to screen more patients at a lower cost per patient. This would ensure results are achieved quicker and more cost-effectively (Shendure and Ji, 2008).

The mixed ancestry mutation profile has been studied more extensively at UCT and a common *MLH1* c.1528C>T mutation has been found in the mixed ancestry population in the Western and Northern Cape of South Africa (Ramesar et al., 2000; Stupart et al., 2009). However, a wider profile can be determined by studying mixed ancestry and black individuals around Gauteng, and eventually all nine provinces in South Africa. A larger study, or even a combined study, will also allow possible founder mutations or common mutations to be determined, which could allow more effective screening of the disease.

This study focused on the screening of *MLH1* and *MSH2*. These genes account for 90% of all MMR gene mutations in HNPCC in other populations. However, these genes may not be involved to the same extent in HNPCC in South Africa. Screening of *MSH6* and *PMS2* will need to be performed to determine if these genes are mutated in CRC patients. Furthermore, *EPCAM* has been shown to be mutated in cases of HNPCC and should be investigated in the South African population (Jasperson et al., 2010).

With regard to mutation detection, Sanger sequencing and MLPA are useful techniques for determining mutations in the DNA of a patient. These techniques can determine both small mutations and whole exon deletions/duplications. To complement this strategy, extracting mRNA could provide valuable insights into the functional effects of variants, especially since so many mutations are located within introns (Strachan and Read, 2004).

Collecting the mRNA will allow future researchers to determine if these intronic variants lead to aberrant splicing. However, RNA studies are challenging to perform as mRNA is not as stable as DNA, with collection, extraction and storage affecting the mRNA. In addition, mRNA extraction should follow soon after the blood is drawn to ensure better mRNA quality (Strachan and Read, 2004; Marteau et al., 2005). In South Africa, many medical facilities are not equipped to extract mRNA, while transportation and storage of samples is very challenging.

Full gene sequencing of the MMR genes is important in determining whether a family has HNPCC. However, this technique is expensive and other screening strategies may be important in reducing costs. IHC has the ability to determine which protein is not expressed in a tumour and may provide direction for gene screening (Lynch et al., 2006). Additionally, IHC testing can be used to screen other disorders that affect protein production. IHC testing is currently available at UCT, but this service must be provided nationwide.

MSI is also a feature of tumours that are deficient in MMR function. MSI testing has also been shown to be a valuable tool in identifying HNPCC since loss of the MMR genes leads to MSI. However, MSI also occurs in sporadic CRC, which will yield a false positive for germ line MMR defects (Lynch et al.,

2006). Setting up MSI testing and IHC testing in local laboratories could potentially improve identification and management of HNPCC families.

Future studies that aim to assess the functions of the novel variants, and even the reported variants, will be important in determining the precise mutation profile of HNPCC patients in South Africa. Other studies have been performed to determine functional effects (Takahashi et al., 2007; Tournier et al., 2008) and bioinformatic prediction (Arnold et al., 2009). Either these techniques, or similar techniques, can be performed in South Africa. It may be particularly important to establish functional assays as the black population could potentially yield a large number of novel variants with unknown function due to the genetic diversity of this population (Campbell and Tishkoff, 2008).

Disease co-segregation studies can also be used to determine if a variant is pathogenic. This was performed in patient 150 and their sibling, who was also affected with colon cancer. Both siblings showed a loss of MSH2 staining in IHC and both had the *MSH2* c.1811C>A variant in *MSH2*, which strongly suggests that this variant is pathogenic. This approach can be used when more than one family member is affected by a disease (Strachan and Read, 2004).

Determining the population frequencies of these variants and their allele frequencies is another study that can be performed. This will provide further information about the diverse African genome and could be submitted to dbSNP or the HapMap project. As seen with the *MSH2* c.1662-10C>T and *MLH1* c.1732-19T>A variants, novel variants were found that are polymorphic in South Africa. These results will also be submitted to dbSNP. More studies must be conducted in this population to gain further insights into African population genetics.

5. Conclusion

HNPCC accounts for approximately 5% of all CRC worldwide (Jasperson et al., 2010). In South Africa, the proportion of CRC cases attributed to HNPCC is unknown and the mutation profile is not well established. However, it is suspected that a higher proportion of CRC in South Africa is attributed to HNPCC (Cronjé et al., 2009). This study aimed to establish the mutation profile of HNPCC in *MLH1* and *MSH2* in the South African population as these genes have been implicated in 90% of MMR gene mutations. Twenty patients were collected in total, 12 of which fulfilled the Amsterdam II criteria (11 white, 1 mixed ancestry), and the remaining eight fulfilling the revised Bethesda guidelines (seven black, 1 mixed ancestry). The patients were analysed using MLPA and Sanger sequencing to identify which variants were present in *MLH1* and *MSH2*.

MLPA did not detect any large deletions or duplications, but this technique should continue to be used to screen *MLH1* and *MSH2*. This was a small cohort and there may be large mutations present in other patients. Using Sanger sequencing, a total of 22 different variants were detected. Eleven variants were detected in *MLH1* and the other 11 in *MSH2*. A known pathogenic mutation in *MSH2* from a positive control was detected. However, this control highlighted that allele drop-out is present and this could be the result of polymorphisms in the primer binding regions. The polymorphisms need to be investigated as they could lead to false negatives, which could affect the mutation detection and clinical management of HNPCC patients. Fifteen previously reported variants were detected, of which three were pathogenic. Six novel variants, of unknown pathogenicity, were detected and investigated further using RFLP, HRM or FASTSNP.

From the six novel variants, the *MLH1* c.885-1G>A variant was thought likely to be pathogenic as it was located on a conserved splice acceptor site and was not present in 45 normal samples. Two novel variants, *MSH2* c.-185C>A and *MSH2* c.367-16T>A, had an unknown contribution toward disease and would require further experimentation to determine if they are pathogenic. Three novel variants were thought likely to be non-pathogenic.

In total, four pathogenic mutations were detected in 19 patients, giving a detection rate of 21%. This detection rate is comparable to other similar studies performed in Caucasians, Asians and African Americans. The Amsterdam II criteria had a greater predictive value for identifying patients with mutations in *MLH1* and *MSH2* as three of the pathogenic mutations were detected among Amsterdam II positive patients. One pathogenic mutation was detected in a Bethesda positive black patient. Both the Amsterdam II criteria and the Bethesda guidelines should continue to be used as guidelines for choosing CRC patients for genetic testing.

Identification of inherited pathogenic mutations in HNPCC is very important as it can allow for improved management of a patient and their family. Family members of a proband can be screened for the pathogenic mutation. If a family member is found to be a carrier of a mutation, that person

can undergo regular colon screening via colonoscopy, or endometrial screening (Lynch et al, 2006; Järvinen et al., 2009). Early detection of cancer is vital to effective treatment (Etzioni, 2003). Alternatively, an at-risk family member can opt to have prophylactic colectomy (Lynch et al., 2006).

There were no common pathogenic mutations detected in this study but further research in a larger cohort may reveal common pathogenic mutations. Mutations in other genes, such as *MSH6* and *PMS2* may contribute to HNPCC cases in South Africa. Although these genes account for 10% of HNPCC cases worldwide, they may account for a higher proportion of HNPCC cases in South Africa (Peltomäki, 2005; da Silva et al., 2009). Screening for mutations in *MSH6*, *PMS2* and *EPCAM* in the future will allow for a more complete mutation spectrum to be established.

Establishing a molecular diagnostic service in South Africa for HNPCC testing is vital to improving the treatment and management of this disease. Currently, testing can be performed in research laboratories or samples must be sent overseas. The general public in South Africa cannot afford to send samples overseas, so a local service is necessary. Determining the mutation profile of MMR defects in South Africa will be an important step in establishing the molecular diagnostic service. This would allow for targeted screening of common or founder mutations as a first line, if they are present in the population.

In summary, this study attempted to characterise the mutation spectrum of *MLH1* and *MSH2* in the South African population. A novel pathogenic mutation (*MLH1* c.885-1G>A) and four previously reported pathogenic mutations (*MLH1* c.454-13A>G; *MLH1* c.731_734delGTTA; *MSH2* c.1811C>A and *MSH2* c.2006-6T>C) were detected as well as five other novel variants (*MLH1* c.1732-19T>A; *MSH2* c,-185C>A; *MSH2* c.367-16T>A; *MSH2* c.1662-43C>T and *MSH2* c.1662-10C>T). No common mutations were detected but future studies with a larger cohort, investigating *MLH1, MSH2, MSH6, PMS2* and *EPCAM*, would better determine the mutation profile of these genes in South Africa.

Appendix A: Cancer Statistics

Table A.1: Estimated global incidence of cancer for both genders of all ages

Cancer	Numbers	Crude Rate	ASR	Cumulative Risk
All cancers excl. non-melanoma skin cancer	12662554	187.6	181.6	18.70
Lung	1608055	23.8	23.0	2.78
Colorectum	1235108	18.3	17.3	1.97
Stomach	988602	14.6	14.1	1.67
Liver	749744	11.1	10.8	1.24
Oesophagus	481645	7.1	7.0	0.86
Bladder	382660	5.7	5.3	0.60
Non-Hodgkin lymphoma	356431	5.3	5.1	0.54
Leukaemia	350434	5.2	5.0	0.47
Kidney	273518	4.1	4.0	0.46
Pancreas	278684	4.1	3.9	0.44
Lip, oral cavity	263020	3.9	3.8	0.45
Brain, nervous system	237913	3.5	3.5	0.35
Thyroid	213179	3.2	3.1	0.31
Melanoma of skin	199627	3.0	2.8	0.30
Larynx	150677	2.2	2.2	0.28
Gallbladder	145203	2.2	2.0	0.22
Other pharynx	136622	2.0	2.0	0.25
Multiple myeloma	102826	1.5	1.4	0.17
Nasopharynx	84441	1.3	1.2	0.13
Hodgkin lymphoma	67919	1.0	1.0	0.08
Kaposi sarcoma	34477	0.5	0.5	0.04

Crude and age-standardized rate (ASR) per 100,000.
Cumulative risk [0-74], percent
Statistics ordered by Cancer name
Adapted from GLOBOCAN, 2008 and Ferlay et al., 2010

Table A.2: Estimated global cancer mortality for both sexes of all ages

Cancer	Numbers	Crude Rate	ASR	Cumulative Risk
All cancers excl. non-melanoma skin cancer	7564802	112.1	106.1	11.16
Lung	1376579	20.4	19.4	2.30
Stomach	737419	10.9	10.3	1.19
Liver	695726	10.3	10.0	1.14
Colorectum	609051	9.0	8.2	0.88
Oesophagus	406533	6.0	5.8	0.70
Pancreas	266669	4.0	3.7	0.41
Leukaemia	257161	3.8	3.6	0.33
Non-Hodgkin lymphoma	191599	2.8	2.7	0.27
Brain, nervous system	174880	2.6	2.5	0.27
Bladder	150282	2.2	2.0	0.20
Lip, oral cavity	127654	1.9	1.9	0.22
Kidney	116368	1.7	1.6	0.18
Gallbladder	109587	1.6	1.5	0.16
Other pharynx	95550	1.4	1.4	0.17
Larynx	81892	1.2	1.2	0.15
Multiple myeloma	72453	1.1	1.0	0.11
Nasopharynx	51609	0.8	0.8	0.08
Melanoma of skin	46372	0.7	0.6	0.07
Thyroid	35383	0.5	0.5	0.06
Hodgkin lymphoma	29902	0.4	0.4	0.04
Kaposi sarcoma	29377	0.4	0.4	0.03

Crude and age-standardised rate (ASR) per 100,000.
Cumulative risk [0-74], percent
Statistics ordered by Cancer name
Adapted from GLOBOCAN, 2008 and Ferlay et al., 2010

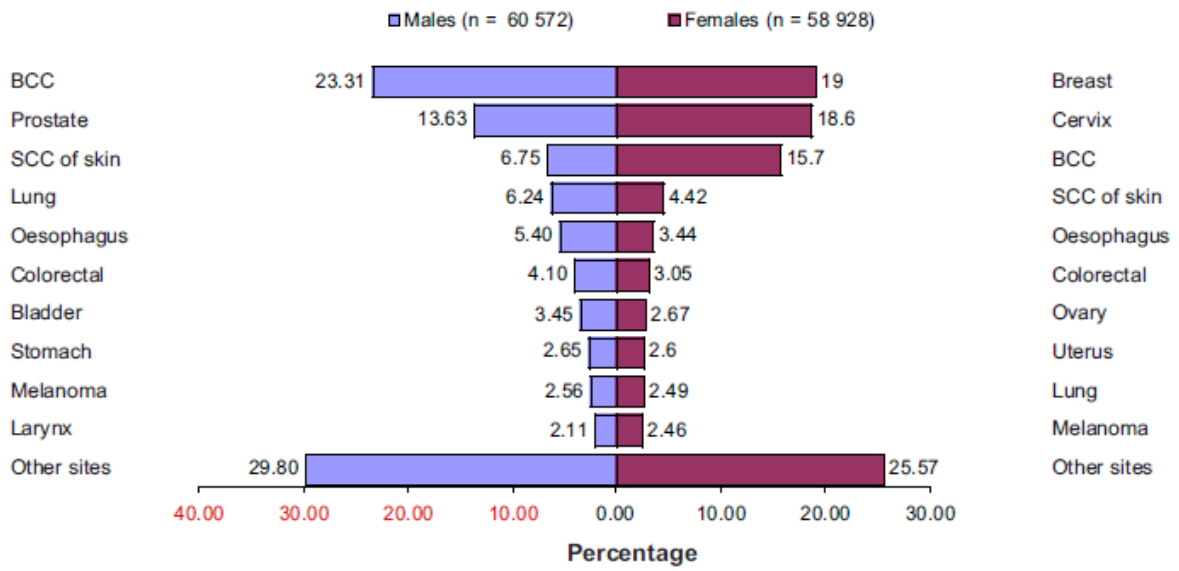


Figure A.1: Percentage distribution of 10 most common cancers in South Africa by gender, 1998 and 1999, all ages (NCR, 2004)

Table A.3: Summary rates for top 5 cancers in South Africa by sex

Pop / Sex	1998			1999			2004		
	Cancer	ASR	LR	Cancer	ASR	LR	Cancer	ASR	LR
Females	Cervix	34.43	26	Breast	33.41	27	Breast	30.07	29
	Breast	32.70	27	Cervix	28.69	31	Cervix	24.40	36
	CRC	5.83	154	CRC	6.61	131	CRC	5.52	150
	Oes	5.95	136	Oes	5.49	143	Uterus	5.77	128
	Ovary	4.91	166	Uterus	5.09	146	Melanoma	3.89	229
	All	**136.74**	**6**	**All**	**134.86**	**6**	**All**	**116.10**	**8**
Males	Prostate	37.59	22	Prostate	34.12	24	Prostate	34.37	23
	Lung	15.18	52	Lung	13.56	59	Lung	9.94	79
	Oes	12.56	65	Oes	11.33	73	Oes	7.56	107
	CRC	9.44	92	CRC	9.74	83	CRC	8.61	98
	Bladder	8.63	98	Bladder	8.24	104	Bladder	6.77	121
	All	**156.18**	**4**	**All**	**148.87**	**5**	**All**	**128.36**	**6**

ASR *Age standardized rate per 100 000*
ASR excludes Basal Cell Carcinoma of the skin and Squamous Cell Carcinoma of skin
LR *Lifetime risk (0-74 years) of developing a cancer expressed as 1 in X number of people*
Oes= Oesophagus
Table adapted from NCR, 2004 & NCR, 2011

Table A.4: Summary statistics for colorectal cancer, 1998 and 1999

Pop/Sex	N(Obs)	N(Adj)	Percent	Crude	ASR	95%LCL	95%UCL	Cumrisk	LR
Females, 1998									
Asian	46	47	5.09	8.56	10.45	7.44	13.46	1.29	78
Black	226	240	25.00	1.4	2.04	1.77	2.30	0.22	455
Mixed Ancestry	102	109	11.28	5.66	7.99	6.46	9.51	0.92	109
White	530	564	58.63	22.21	14.99	13.70	16.28	1.73	58
Males, 1998									
Asian	44	44	3.97	8.22	11.01	7.83	14.19	1.07	94
Black	271	287	24.48	1.75	3.02	2.65	3.39	0.31	323
Mixed Ancestry	130	138	11.74	7.52	13.76	11.40	16.12	1.75	58
White	662	699	59.80	28.34	23.74	21.97	25.51	2.89	35
Females, 1999									
Asian	35	35	3.31	6.29	7.31	4.87	9.75	1.02	99
Black	268	284	25.31	1.62	2.31	2.03	2.58	0.26	385
Mixed Ancestry	125	132	11.8	6.75	9.66	8	11.33	1.13	89
White	631	668	59.58	26.23	17.52	16.16	18.88	2.09	48
Males, 1999									
Asian	53	56	4.57	10.33	14.28	10.4	18.16	1.98	51
Black	278	301	23.97	1.79	2.96	2.6	3.31	0.35	286
Mixed Ancestry	128	138	11.03	7.42	14.09	11.7	16.48	1.8	56
White	701	752	60.43	30.46	25.44	23.64	27.23	3.27	31
Females, 2004									
Asian	46	50	6.73	8.71	9.14	6.58	11.70	1.34	74
Black	253	272	2.22	1.45	2.10	1.84	2.36	0.25	392
Mixed Ancestry	133	142	4.08	6.80	9.34	7.76	10.92	1.10	90
White	516	550	4.66	24.36	14.78	13.48	16.07	1.75	57
Males, 2004									
Asian	59	68	10.95	12.19	15.61	11.76	19.46	2.09	48
Black	288	315	3.70	1.73	3.08	2.72	3.44	0.34	292
Mixed Ancestry	165	179	5.19	8.94	15.69	13.26	18.13	1.72	58
White	582	633	4.24	29.10	21.49	19.77	23.20	2.60	38

N(Obs): Number of cases observed
N(Adj): Observed cases adjusted for unknown population group
Percent: Percentage of all site cancers
Crude: Number of cases / population per 100 000
ASR: Age standardised rate per 100 000 (World standard population)
95%LCL: 95% Lower confidence limit for ASR
95%UCL: 95% Upper confidence limit for ASR
Cumrisk: Cumulative lifetime incidence risk (0-74 years)
LR: Lifetime risk (0-74 years) of developing a cancer expressed as 1 in X number of people
Adapted from NCR, 2004 and NCR, 2011

Appendix B: Solutions

1 Kb+ DNA marker

11µl ladder

10µl Ficoll loading dye

73µl 1x TE

3% Agarose Gel

12g Agarose powder

400ml 1 x TBE

12µl EtBr (10mg/ml)

TE Buffer (10mM Tris-HCl, pH 7.4; 1mM EDTA)

0.5ml 2M Tris-HCl (pH7.4)

20µl EDTA (pH 8.0)

99.48ml dH_20

Autoclaved and stored at room temperature

2M Tris-HCl pH 7.4

242.2g Tris base

700ml dH_20

pH adjusted by adding concentrated HCl

Final volume made to 1l using dH_20

0.5 Ethylenediamine Tetre-acetic Acid (EDTA)

93.05g Na2.EDTA.$2H_2O$

300ml dH_2O

pH adjusted to 8.0 with 10M NaOH

Final volume made to 500ml with dH_2O

Deoxyribonucleotide triphosphate (dNTP) 10x stock solution (1.25mM)

125µl of 10mM dNTP (dATP, dGTP, dTTP and dCTP)

500 µl of ddH_2O to make up to 1ml

Ten aliquots of 100µl made and stored at -20°C

Concentration in PCR reaction mix is 0.125mM

Ethidium Bromide (EtBr) (10mg/ml)

Available from Sigma-Aldrich as an aqueous solution (10mg/ml)

Caution: EtBr is a powerful mutagen. Always wear gloves when handling gels or solutions that contain EtBr

Ficoll-bromophenol blue loading dye (100ml)

50g of 50% sucrose crystals

0.1ml 0.5M EDTA (pH 7.0)

0.1g 0.1% bromophenol blue dye

10g 10% Ficoll powder

Make up to 100ml with ddH2O before aliquot are made into Eppendorf tubes

Appendix C: Consent Forms

The National Health Laboratory Service

University of the Witwatersrand, School Of Pathology
Division Human Genetics

Hospital Street, Johannesburg 2001
Telephone: +27-11-489-9224/9223/9211

PO Box 1038, Johannesburg 2000
Telefax: +27-11-489-9226 or +27-11-489-9209

Prof A Christianson 489-9211/9239 Prof A Krause 489-9219 Prof M Ramsay 489-9214

Prof H Soodyall 489-9208 Dr T Lane 489-9221

Information Sheet

Study Title

Investigating the Molecular Etiology of Hereditary Non-Polyposis Colorectal Cancer in South Africa

Investigator: Kenneth Davison: MSc. (Med) Research Student

Good day, I am Kenneth Davison and I am part of the Division of Human Genetics, National Health Laboratory Service and University of the Witwatersrand and we are conducting research on a type of inherited colon cancer called hereditary nonpolyposis colorectal cancer (HNPCC). The study aims to look for changes, known as mutations, in the genes of individuals who are affected by HNPCC. This will contribute towards the development of a test to diagnose HNPCC in South Africans, which could lead to improvements in the treatment and management of people who have, or are at risk of having, HNPCC.

Colorectal cancer (CRC) is one of the most common types of cancer worldwide with HNPCC accounting for about 5% of all colorectal cancers. Most cases of CRC are sporadic. This means that CRC occurs when many mutations, which happen by chance, build up in the cells of the colon. However, in about 5% of CRC, a mutation is passed from a parent to their child and is present at birth. When one inherits a mutation from their parents, they have an increased risk of developing CRC and will probably develop CRC at a younger age. Generally, sporadic CRC is found in individuals over the age of 60. When CRC is diagnosed under the age of 50 it is often the result of an inherited mutation.

The National Health Laboratory Service

University of the Witwatersrand, School Of Pathology
Division Human Genetics

Hospital Street, Johannesburg 2001
Telephone: +27-11-489-9224/9223/9211

PO Box 1038, Johannesburg 2000
Telefax: +27-11-489-9226 or +27-11-489-9209

Prof A Christianson 489-9211/9239 Prof A Krause 489-9219 Prof M Ramsay 489-9214

Prof H Soodyall 489-9208 Dr T Lane 489-9221

It has been shown that there are four major genes involved in HNPCC and when any of these genes have a mutation; one has an increased risk of getting CRC. The four major genes are known as *MLH1*, *MSH2*, *MSH6* and *PMS2* and these genes are involved in repairing DNA when mutations occur. Their job is to repair mutations but when a mutation happens in them, then they are unable to repair the DNA properly and this could lead to CRC. HNPCC is a cancer that is found in families, meaning that someone in your family either already has the same cancer, or is at a high risk of getting the same cancer as yours. It is also possible that you or your family members could be at risk of getting other related cancers.

We would like to invite you to participate in this study by allowing us to take about two teaspoons (5-10ml) of blood from you. Taking blood is not harmful to you but it does involve slight discomfort when the needle goes through the skin. We will extract DNA from the white blood cells and screen the four genes for possible disease causing mutations.

If an inherited mutation is identified, it may have important consequences for your other family members. If a mutation is found in these genes, it means that other members of your family may be at risk of developing CRC. They can take the test which will be developed to see if they also have the mutation and if they are at risk. If they are also affected, they will be advised to have regular screening of the colon so that the cancer can be caught early. If the cancer is found early, it is easier to treat and manage and affected people are much more likely to survive.

The National Health Laboratory Service

University of the Witwatersrand, School Of Pathology
Division Human Genetics

Hospital Street, Johannesburg 2001
Telephone: +27-11-489-9224/9223/9211

PO Box 1038, Johannesburg 2000
Telefax: +27-11-489-9226 or +27-11-489-9209

Prof A Christianson 489-9211/9239 Prof A Krause 489-9219 Prof M Ramsay 489-9214

Prof H Soodyall 489-9208 Dr T Lane 489-9221

There will be no cost to you for the collection of your blood and all your personal information will be kept confidential. We would also like to compare your family history to your genetic condition. We would also like to keep your DNA for future research related to HNPCC but you have the right to choose whether your DNA can be used for future studies or not.

This study might not help you directly. It will not make you better. Your treatment will continue just as before. You also have the right to withdraw from the study at any time and your withdrawal will not affect present or future treatments. This project will help us to understand more about HNPCC which may help your family and other families in the future

If you have any questions about the study, you can contact me or my colleagues at:

Prof. Amanda Krause: 011 489 9219

Dr. Robyn Kerr: 011 489 9344 Cell: 083 601 6303

Mr. Kenneth Davison: 011 489 9225 Cell: 082 579 4056

Ms. Anisa Keshav (Ethics): 011 717 1234

The National Health Laboratory Service

University of the Witwatersrand, School Of Pathology
Division Human Genetics

Hospital Street, Johannesburg 2001
Telephone: +27-11-489-9224/9223/9211

PO Box 1038, Johannesburg 2000
Telefax: +27-11-489-9226 or +27-11-489-9209

Prof A Christianson 489-9211/9239 Prof A Krause 489-9219 Prof M Ramsay 489-9214

Prof H Soodyall 489-9208 Dr T Lane 489-9221

Consent Form

Investigating the Molecular Etiology of Hereditary Non-Polyposis Colon Cancer in South Africa

Surname: …….

Personal Name: ……..

Date of Birth: ……..

1. The research has been explained to me and I understand that the test involves taking a sample of blood from myself
2. The information collected about me will be kept confidential
3. I understand why the study is being done and that it may have direct benefits for me / my child and my extended family. However, there may also be no direct benefits. The study will help researchers understand HNPCC so that they can develop treatments and ways to prevent HNPCC
4. I have had sufficient opportunity to ask questions about the research and I have decided to participate in the study without coercion
5. I understand that I do not have to take part in this study. If I choose not to participate it will not affect the way I am treated at the hospital / clinic. Similarly, if I choose to withdraw from the study at any time, it will not affect any future treatment I will require

My decision for the sample after the study is completed: (Please mark block with an x)

☐ I give permission for the researcher to view my clinical notes

☐ I give permission for the researcher to receive copies of my histology reports and scans

☐ I would like to be notified of positive results

☐ If possible, my samples(s) should be stored for future analysis in my interest, only on my request

☐ My samples(s) can also be used for medical research

- ☐ With my name
- ☐ Without my name (anonymous). This means that I can't be informed about eventual results

The National Health Laboratory Service

University of the Witwatersrand, School Of Pathology
Division Human Genetics

Hospital Street, Johannesburg 2001
Telephone: +27-11-489-9224/9223/9211

PO Box 1038, Johannesburg 2000
Telefax: +27-11-489-9226 or +27-11-489-9209

Prof A Christianson 489-9211/9239 **Prof A Krause 489-9219** **Prof M Ramsay 489-9214**

Prof H Soodyall 489-9208 **Dr T Lane 489-9221**

☐ My samples must be discarded after this study is completed

Signed on _____day of_____20___ at _____

Patient Name: _____ Signature: _____

Witness Name: _____ Signature: _____

Researcher Name: Mr. K. Davison Signature: _____

Appendix D: Ethics Clearance Certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Mr Kenneth Davidson

CLEARANCE CERTIFICATE	**M10612**
PROJECT	Investigating the Molecular Etiology of Hereditary-Nonpolyposis Colorectal Cancer in South Africa
INVESTIGATORS	Mr Kenneth Davidson.
DEPARTMENT	Division of Human Genetics
DATE CONSIDERED	26/06/2010
DECISION OF THE COMMITTEE*	Approved unconditionally

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 05/08/2010 **CHAIRPERSON** .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable
cc: Supervisor : Prof A Krause

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**
PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES...

Appendix E: PCR Primers for *MLH1* and *MSH2*

Table E.1: PCR Primers for *MLH1* and PCR Product Sizes. The universal tails on the primers are underlined. The sequencing primers are complementary to the universal tails. The PCR product sizes are given.

***MLH1* Single Exon Tailed Primers**			
Exon	**Forward (5'-3')**	**Reverse (5'-3')**	**PCR Product Size**
1	GTAGCGCGACGGCCAGTACTGAGGTGATTGGCTGA	CAGGGCGCAGCGATGACGGCCCGTTAAGTCGTAGCC	270
2	GTAGCGCGACGGCCAGTGTACATTAGAGTAGTTGCAGAC	CAGGGCGCAGCGATGACCCAGAACAGAGAAAGGTCCTGA	248
3	GTAGCGCGACGGCCAGTAGAGATTTGGAAAAATGAGTAAC	CAGGGCGCAGCGATGACACAATGTCATCACAGGAGG	242
4	GTAGCGCGACGGCCAGTGTGACCCAGCAGTGAGTT	CAGGGCGCAGCGATGACCCTAGGCAAAAAATACATTT	228
5	GTAGCGCGACGGCCAGTGATTTTCTCTTTTCCCCTT	CAGGGCGCAGCGATGACCCAATATTTATACAAACAAAGCT	236
6	GTAGCGCGACGGCCAGTCTTTTGCCAGGACATCTT	CAGGGCGCAGCGATGACCTAGAACACATTACTTTGATGAC	261
7&8	GTAGCGCGACGGCCAGTAAAAGGGGGCTCTGACATCT	CAGGGCGCAGCGATGACATGTGATGGAATGATAAACC	451
9	GTAGCGCGACGGCCAGTTTCTTTTGTAATGTTTGAGTTTTGAG	CAGGGCGCAGCGATGACCCTGTGAGTGGATTTCCCATG	247
10	GTAGCGCGACGGCCAGTACCCCTCAGGACAGTTTT	CAGGGCGCAGCGATGACTCTGTTCCTTGTGAGTCTTG	220
11	GTAGCGCGACGGCCAGTCCACATACACCATATGTGGGC	CAGGGCGCAGCGATGACAAGGCCCCAGAGAAGTAG	303
12	CTCCATTTGGGGACCTGTAT Needs an internal fwd sequencing primer: *TTTAATACAGACTTTGCTACCAGGAC*	CAGGGCGCAGCGATGACTAAAGAGTAGCTGTACTTTTCCCAA	545
13	GTAGCGCGACGGCCAGTCTGCACTTCCTTTTCTTCAT	CAGGGCGCAGCGATGACCCCTATGCATCCCAGGCA	228
14	GTAGCGCGACGGCCAGTCTATTACTTACCTGTTTTTTGG	CAGGGCGCAGCGATGACGTAGCTCTGCTTGTTCACAC	220
15	GTAGCGCGACGGCCAGTAACTGGTTGTATCTCAAGCATG	CAGGGCGCAGCGATGACGTGGAGAGCTACTATTTTCAG	227
16	GTAGCGCGACGGCCAGTCAGGCTTCATTTGGATGCTCCG	CAGGGCGCAGCGATGACCAGAAGTATAAGAATGGCTGTC	284
17	GTAGCGCGACGGCCAGTTGACAGCATTATTTCTTGTTC	CAGGGCGCAGCGATGACCATGTACCGAAATGCTTAGTA	210
18	GTAGCGCGACGGCCAGTTCTGTGATCTCCGTTTAGAATGA	CAGGGCGCAGCGATGACGTGCCAGTGTGCATCACC	246
19	GTAGCGCGACGGCCAGTACAGGGAGGCTTATGACATC	CAGGGCGCAGCGATGACCACATCCCACAGTGCATAAA	266

Table E.2: PCR primers for MSH2 and the PCR product sizes. The universal tails on the primers are underlined. The sequencing primers are complementary to the universal tails. The PCR product sizes are given.

***MSH2* Single Exon Tailed Primers**			
Exon	**Forward**	**Reverse**	**PCR Size**
1	GTAGCGCGACGGCCAGTTAGCTAAAGTCACCAGCG	CAGGGCGCAGCGATGACTCTGAGGCGGGAAAGGAG	535
2	TTTTTCTTCCCAGTCTCG	CAGGGCGCAGCGATGACTTTTCTACTCTTAAAAAAATAAGTA	329
3	GTAGCGCGACGGCCAGTGTTCAAGAGTTTGTTAAATT	CAGGGCGCAGCGATGACGGAATCTCCTCTATCACTAG	391
4	GTAGCGCGACGGCCAGTTTCTTATTCCTTTTCTCATAGTAG	CAGGGCGCAGCGATGACGAGATAAATATGACAGAAATATCC	344
5	GTAGCGCGACGGCCAGTTGGATCCAGTGGTATAGAAAT	ACCATTCAACATTTTTAACC	265
6	GTAGCGCGACGGCCAGTAGCTTGCCATTCTTTCTATT	CAGGGCGCAGCGATGACATAATCATGTGGGTAACTGC	267
7	GTAGCGCGACGGCCAGTGCTTAGTTGATAAATTTTAATTT	CAGGGCGCAGCGATGACATTGTATGAGTTGAAGGAAAA	345
8	GTAGCGCGACGGCCAGTTTTGTATTCTGTAAAATGAGATC	CAGGGCGCAGCGATGACAGTGGCCTTTGCTTTTT	247
9	GTAGCGCGACGGCCAGTTTTCTGTCTTTACCCATTATT	CAGGGCGCAGCGATGACCTGGCATCAATATTTTGTCT	283
10	GTAGCGCGACGGCCAGTAATGGTAGTAGGTATTTATGGAA	CAGGGCGCAGCGATGACCACATCATGTTAGAGCATTTAGGG	301
11	GTAGCGCGACGGCCAGTTGTTTCACGTAGTACACATTGCTT	CAGGGCGCAGCGATGACCCAGGTGACATTCAGAACAT	246
12	GTAGCGCGACGGCCAGTTTATTATTCAGTATTCCTGTGTAC	CAGGGCGCAGCGATGACAAACGTTACCCCCACAAAG	379
13	GTAGCGCGACGGCCAGTGCTTTCTGATATAATTTGTT	CAGGGCGCAGCGATGACCAGAGACATACATTTCTATCTTCA	311
14	GTAGCGCGACGGCCAGTTATGTGATGGGAAATTTCATG	CAGGGCGCAGCGATGACTCCCATTACCAAGTTCTGA	364
15	GTAGCGCGACGGCCAGTGCTGTCTCTTCTCATGCTG	CAGGGCGCAGCGATGACAACCTTCATCTTAGTGTCCTG	332
16	GTAGCGCGACGGCCAGTGTGTGATATGTTTAGATGGAAA	CAGGGCGCAGCGATGACTTACAGACAATAGCTTATCAATAT	356

Underlines: Universal tails for primers

Table E.3: DNA Sequence of *MLH1*. The PCR primer binding sites are shown

MLH1 Exon	DNA Sequence
1	1 tggctggatg ctaagctaca gctgaaggaa gaacgtgagc acgaggcact gaggtgattg 61 gctgaaggca cttccgttga gcatctagac gtttccttgg ctcttctggc gccaaaATGT 121 CGTTCGTGGC AGGGGTTATT CGGCGGCTGG ACGAGACAGT GGTGAACCGC ATCGCGGCGG 181 GGGAAGTTAT CCAGCGGCCA GCTAATGCTA TCAAAGAGAT GATTGAGAAC TGgtacggag 241 ggagtcgagc cgggctcact taaggctac gacttaacgg gccgcgtcac tcaatggcgc
2	1 aaacacgtta atgaggcact attgtttgta tttggagttt gttatcattg cttggctcat 61 attaaaatat gtacattaga gtagttgcag actgataaat tattttctgt ttgatttgcc 121 agTTTAGATG CAAAATCCAC AAGTATTCAA GTGATTGTTA AAGAGGGAGG CCTGAAGTTG 181 ATTCAGATCC AAGACAATGG CACCGGGATC AGGgtaagta aaacctcaaa gtagcaggat 241 gtttgtgcgc ttcatggaag agtcaggacc tttctctgtt ctggaaacta ggcttttgca
3	1 tttcctggat taatcaagaa atggaattca aagagatttg gaaaaatgagt aacatgatt 61 atttactcat ctttttggta tctaacagAA AGAAGATCTG GATATTGTAT GTGAAAGGTT 121 CACTACTAGT AAACTGCAGT CCTTTGAGGA TTTAGCCAGT ATTTCTACCT ATGGCTTTCG 181 AGGTGAGgta agctaaagat tcaagaaatg tgtaaaatat cctcctgtga tgacattgtc
4	1 tggaagcagc agncagataa cctttccctt tggtgaggtg acagtgggtg acccagcagt 61 gagttttct ttcagtctat tttcttttct tccttagGCT TTGGCCAGCA TAAGCCATGT 121 GGCTCATGTT ACTATTACAA CGAAAACAGC TGATGGAAAG TGTGCATACA Ggtatagtgc 181 tgacttcttt tactcatata tattcattct gaaatgtatt ttttgcctag gtctcagagt
5	1 aaacaggaaa atgataaata attaagataa tatataatat ggctatattt taatctattg 61 ttgatatgat tttctctttt ccccttggga ttagtatcta tctctctact ggatattaat 121 ttgttatatt ttctcattag AGCAAGTTAC TCAGATGGAA AACTGAAAGC CCCTCCTAAA 181 CCATGTGCTG GCAATCAAGG GACCCAGATC ACGgtaagaa tggtacatgg gagagtaaat 241 tgttgaagct ttgtttgtat aaatattgga ataaaaaata aaattgcttc taagttttca
6	1 agtcagtgct tagaactgtg ctgttggtaa atataattgg attcactatc ttaagacctc 61 gcttttgcca ggacatcttg ggttttattt tcaagtactt ctatgaattt acaagaaaaa 121 tcaatcttct gttcagGTGG AGGACCTTTT TTACAACATA GCCACGAGGA GAAAAGCTTT 181 AAAAAATCCA AGTGAAGAAT ATGGGAAAAT TTTGGAAGTT GTTGGCAGgt acagtccaaa 241 atctgggagt gggtctctga gatttgtcat caaagtaatg tgttctagtg ctcatacat
7	1 gagatatttt tctttatatg tacctttgtt tagattactc aactccacta atttatttaa 61 ctaaaagggg gctctgacat ctagtgtgtg tttttggcaa ctcttttctt actcttttgt 121 ttttcttttc cagGTATTCA GTACACAATG CAGGCATTAG TTTCTCAGTT AAAAAAgtaa 181 gttcttggtt tatgggggat ggttttgttt tatgaaaaga aaaaagggga tttttaatag 241 tttgctggtg gagataaggt tatg
8	1 atgtttcagt ctcagccatg agacaataaa tccttgtgtc ttctgctgtt tgtttatcag 61 CAAGGAGAGA CAGTAGCTGA TGTTAGGACA CTACCCAATG CCTCAACCGT GGACAATATT 121 CGCTCCATCT TTGGAAATGC TGTTAGTCGg tatgtcgata acctatataa aaaaatcttt 181 tacatttatt atcttggttt atcattccat cacattattt tggaaccttt caagatatta
9	1 ccttgtgttt ttaaattctg attcttttgt aatgtttgag ttttgagtat tttcaaaagc 61 ttcagaatct cttttctaat agAGAACTGA TAGAAATTGG ATGTGAGGAT AAAACCCTAG 121 CCTTCAAAAT GAATGGTTAC ATATCCAATG CAAACTACTC AGTGAAGAAG TGCATCTTCT 181 TACTCTTCAT CAACCgtaag ttaaaaagaa ccacatggga aatccactca caggaaacac
10	1 gcatatcact acagaaatgt ctttcctgag gtgatttcat gactttgtgt gaatgtacac 61 ctgtgacctc acccctcagg acagttttga actggttgct ttctttttat tgtttagATC 121 GTCTGGTAGA ATCAACTTCC TTGAGAAAAG CCATAGAAAC AGTGTATGCA GCCTATTTGC 181 CCAAAAACAC ACACCCATTC CTGTACCTCA Ggtaatgtag caccaaactc ctcaaccaag 241 actcacaagg aacagatgtt ctatcaggct ctcctctttg aaagagatga gcatgctaat
11	1 gatcctgagg ttttgaccac tgtgtcatct ggcctcaaat cttctggcca ccacatacac 61 catatgtggg cttttttctcc ccctcccact atctaaggta attgttctct cttatttcc 121 tgacagTTTA GAAATCAGTC CCCAGAATGT GGATGTTAAT GTGCACCCCA CAAAGCATGA 181 AGTTCACTTC CTGCACGAGG AGAGCATCCT GGAGCGGGTG CAGCAGCACA TCGAGAGCAA 241 GCTCCTGGGC TCCAATTCCT CCAGGATGTA CTTCACCCAG gtcagggcgc ttctcatcca 301 gctacttctc tggggcctt t gaaatgtgcc cggccagacg tgagagccca gatttttgc
12	1 gataattata cctcatacta gcttctttct tagtactgct ccatttgggg acctgtatat 61 ctatacttct tattctgagt ctctccacta tatatatata tatatatata tatttttttt 121 ttttttttttt tttaatacag ACTTTGCTAC CAGGACTTGC TGGCCCCTCT GGGGAGATGG 181 TTAAATCCAC AACAAGTCTG ACCTCGTCTT CTACTTCTGG AAGTAGTGAT AAGGTCTATG

	241 CCCACCAGAT GGTTCGTACA GATTCCCGGG AACAGAAGCT TGATGCATTT CTGCAGCCTC 301 TGAGCAAACC CCTGTCCAGT CAGCCCCAGG CCATTGTCAC AGAGGATAAG ACAGATATTT 361 CTAGTGGCAG GGCTAGGCAG CAAGATGAGG AGATGCTTGA ACTCCCAGCC CCTGCTGAAG 421 TGGCTGCCAA AAATCAGAGC TTGGAGGGGG ATACAACAAA GGGGACTTCA GAAATGTCAG 481 AGAAGAGAGG ACCTACTTCC AGCAACCCCA Ggtatggcct ttttgggaaaa gtacagccta 541 cctcctttat tctgtaataa aactgccttc taactttggc ttttcatgaa tcacttgca
13	1 aaatgcaacc cacaaaattt ggctaagttt aaaaacaaga ataataatga tctgcacttc 61 cttttcttca ttgcagAAAG AGACATCGGG AAGATTCTGA TGTGGAAATG GTGGAAGATG 121 ATTCCCGAAA GGAAATGACT GCAGCTTGTA CCCCCCGGAG AAGGATCATT AACCTCACTA 181 GTGTTTTGAG TCTCCAGGAA GAAATTAATG AGCAGGGACA TGAGGgtacg taaacgctgt 241 ggcctgcctg ggatgcatag ggcctcaact gccaaggttt tggaaatgga gaaagcagtc
14	1 tgcctggtgc tttggtcaat gaagtggggt tggtaggatt ctattactta cctgtttttt 61 ggttttattt tttgttttgc agTTCTCCGG GAGATGTTGC ATAACCACTC CTTCGTGGGC 121 TGTGTGAATC CTCAGTGGGC CTTGGCACAG CATCAAACCA AGTTATACCT TCTCAACACC 181 ACCAAGCTTA Ggtaaatcag ctgagtgtgt gaacaagcag agctactaca acaatggtcc
15	1 aaccagattc cacagccagg cagaactatt tctgtctcat ccatgtttca gggattactt 61 ctcccatttt gtcccaactg gttgtatctc aagcatgaat tcagcttttc cttaaagtca 121 cttcattttt attttcagTG AAGAACTGTT CTACCAGATA CTCATTTATG ATTTTGCCAA 181 TTTTGGTGTT CTCAGGTTAT CGgtaagttt agatcctttt cacttctgac atttcaactg 241 atcgtttctg aaaatagtag ctctccacta ata
16	1 catttatggt ttctcacctg ccattctgat agtggattct tgggaattca ggcttcattt 61 ggatgctccg ttaaagcttg ctccttcatg ttcttgcttc ttcctagGAG CCAGCACCGC 121 TCTTTGACCT TGCCATGCTT GCCTTAGATA GTCCAGAGAG TGGCTGGACA GAGGAAGATG 181 GTCCCAAAGA AGGACTTGCT GAATACATTG TTGAGTTTCT GAAGAAGAAG GCTGAGATGC 241 TTGCAGACTA TTTCTCTTTG GAAATTGATG AGgtgtgaca gccattctta tacttctgtt
17	1 ggcagatagg agcacaaggc ctgggaaagc actggagaaa tgggatttgt ttaaactatg 61 acagcattat ttcttgttcc cttgtccttt ttcctgcaag cagGAAGGGA ACCTGATTGG 121 ATTACCCCTT CTGATTGACA ACTATGTGCC CCCTTTGGAG GGACTGCCTA TCTTCATTCT 181 TCGACTAGCC ACTGAGgtca gtgatcaagc agatactaag catttcggta catgcatgtg
18	1 attcattccc agcaatattc agcagtccca tttacagttt taacgcctaa agtatcacat 61 ttcgtttttt agctttaagt agtctgtgat ctccgtttag aatgagaatg tttaaattcg 121 tacctatttt gaggtattga atttctttgg accagGTGAA TTGGGACGAA GAAAAGGAAT 181 GTTTTGAAAG CCTCAGTAAA GAATGCGCTA TGTTCTATTC CATCCGGAAG CAGTACATAT 241 CTGAGGAGTC GACCCTCTCA GGCCAGCAGg tacagtggtg atgcacactg gcacccagg
19	1 aagccctaga taacaccaag tctttccaga cccagtgcac atcccatcag ccaggacacc 61 agtgtatgtt gggatgcaaa cagggaggct tatgacatct aatgtgtttt ccagAGTGAA 121 GTGCCTGGCT CCATTCCAAA CTCCTGGAAG TGGACTGTGG AACACATTGT CTATAAAGCC 181 TTGCGCTCAC ACATTCTGCC TCCTAAACAT TTCACAGAAG ATGGAAATAT CCTGCAGCTT 241 GCTAACCTGC CTGATCTATA CAAAGTCTTT GAGAGGTGTT AAatatggtt atttatgcac 301 tgtgggatgt gttcttcttt ctctgtattc cgatacaaag tgttgtatca aagtgtgat

Table E.4: DNA Sequence of *MSH2*. The PCR primer binding sites are shown

MSH2 Exon	DNA Sequence
1	1 cagcatgcgc agtagctaaa gtcaccagcg tgcgcgggaa gctgggccgc gtctgcttat 61 gattggttgc cgcggcagac tcccacccac cgaaacgcag ccctggaagc tgattgggtg 121 tggtcgccgt ggccggacgc cgctcggggg acgtgggagg ggaGGCGGGA AACAGCTTAG 181 TGGGTGTGGG GTCGCGCATT TTCTTCAACC AGGAGGTGAG GAGGTTTCGA CATGGCGGTG 241 CAGCCGAAGG AGACGCTGCA GTTGGAGAGC GCGGCCGAGG TCGGCTTCGT GCGCTTCTTT 301 CAGGGCATGC CGGAGAAGCC GACCACCACA GTGCGCCTTT TCGACCGGGG CGACTTCTAT 361 ACGGCGCACG GCGAGGACGC GCTGCTGGCC GCCCGGGAGG TGTTCAAGAC CCAGGGGGTG 421 ATCAAGTACA TGGGGCCGGC AGgtgagggc cgggacggcg cgtgctgggg agggacccgg 481 ggccttgtgg cgcggctcct ttcccgcctc agagagtggg cggtgagcag cctctccagt
2	1 aagtccagta agctct**tttt tcttcccagt ctcg**ggtatg tctttatcag cagcatgaag 61 tccagctaat acagtgcttg aacatgtaat atctcaaatc tgtaatgtac ttttttttt 121 tttaagGAGC AAAGAATCTG CAGAGTGTTG TGCTTAGTAA AATGAATTTT GAATCTTTTG 181 TAAAAGATCT TCTTCTGGTT CGTCAGTATA GAGTTGAAGT TTATAAGAAT AGAGCTGGAA 241 ATAAGGCATC CAAGGAGAAT GATTGGTATT TGGCATATAA GGtaattatc ttccttttta 301 atttacttat ttttttaaga gtagaaaaat aaaaatgtga agaatttaat tgtgttttta
3	1 attaataagg ttcatagagt ttggattttt ccttttgct tataaaattt taaagtatgt 61 tcaagagttt gttaaatttt taaaatttta tttttactta ggCTTCTCCT GGCAATCTCT 121 CTCAGTTTGA AGATATTCTC TTTGGTAACA ATGATATGTC AGCTTCCATT GGTGTTGTGG 181 GTGTTAAAAT GTCCGCAGTT GATGGCCAGA GACAGGTTGG AGTTGGGTAT GTGGATTCCA 241 TACAGAGGAA ACTAGGACTG TGTGAATTCC CTGATAATGA TCAGTTCTCC AATCTTGAGG 301 CTCTCCTCAT CCAGATTGGA CCAAAGGAAT GTGTTTTACC CGGAGGAGAG ACTGCTGGAG 361 ACATGGGGAA ACTGAGACAG gtaagcaaat tgagtctagt gatagaggag attccaggcc
4	1 aaacatttaa tgtaggtgaa tctgttatca ctatagttat caatgttata attttcattt 61 ttgcttttct tattcctttt ctcatagtag tttaaactat ttctttcaaa atagATAATT 121 CAAAGAGGAG GAATTCTGAT CACAGAAAGA AAAAAAGCTG ACTTTTCCAC AAAAGACATT 181 TATCAGGACC TCAACCGGTT GTTGAAAGGC AAAAAGGGAG AGCAGATGAA TAGTGCTGTA 241 TTGCCAGAAA TGGAGAATCA Ggtacatgga ttataaatgt gaattacaat atatataatg 301 taaatatgta atatataata aataatatgt aaactatagt gacttttag aaggatattt 361 ctgtcatatt tatctcaaaa acctgtgtat caatgatat
5	tggatc 1 ccagtggtat agaaatcttc gattttaaa ttcttaattt tagGTTGCAG TTTCATCACT 61 GTCTGCGGTA ATCAAGTTTT TAGAACTCTT ATCAGATGAT TCCAACTTTG GACAGTTTGA 121 ACTGACTACT TTTGACTTCA GCCAGTATAT GAAATTGGAT ATTGCAGCAG TCAGAGCCCT 181 TAACCTTTTT CAGgtaaaaa aaaaaaaaaa aaaaaaaaaa aggttaaaa atgttgaatg 241 gttaaaaaat gttttcattg acatatactg aagaagctta taaaggagct aaaatatttt
6	1 tcatggcgta gtaangtttt cactaatgag cttgccattc tttctatttt attttttgtt 61 tactagGGTT CTGTTGAAGA TACCACTGGC TCTCAGTCTC TGGCTGCCTT GCTGAATAAG 121 TGTAAAACCC CTCAAGGACA AAGACTTGTT AACCAGTGGA TTAAGCAGCC TCTCATGGAT 181 AAGAACAGAA TAGAGGAGAG gtatgttatt agtttatact ttcgttagtt ttatgtaacc 241 tgcagttacc cacatgatta taccacttat tgtaatatgc agttttggaa gtatatgtta
7	1 ctgatttagt tgagacttac gtgcttagtt gataaatttt aatttttatac taaaatattt 61 tacattaatt caagttaatt tatttcagAT TGAATTTAGT GGAAGCTTTT GTAGAAGATG 121 CAGAATTGAG GCAGACTTTA CAAGAAGATT TACTTCGTCG ATTCCCAGAT CTTAACCGAC 181 TTGCCAAGAA GTTTCAAAGA CAAGCAGCAA ACTTACAAGA TTGTTACCGA CTCTATCAGG 241 GTATAAATCA ACTACCTAAT GTTATACAGG CTCTGGAAAA ACATGAAGgt aacaagtgat 301 tttgtttttt tgttttcctt caactcatac aatatatact tggcaatgtg ctgtcctcat
8	1 aaaattttat gatttgtatt ctgtaaaatg agatcttttt atttgtttgt tttactactt 61 tcttttagGA AAACACCAGA AATTATTGTT GGCAGTTTTT GTGACTCCTC TTACTGATCT 121 TCGTTCTGAC TTCTCCAAGT TTCAGGAAAT GATAGAAACA ACTTTAGATA TGGATCAGgt 181 atgcaatata ctttttaatt taagcagtag ttatttttaa aaagcaaagg ccacttaag
9	1 cagtaaaatt taagtgggag gaaatatttg ctttataatt tctgtcttta cccattattt 61 ataggatttt gtcactttgt tctgtttgca gGTGGAAAAC CATGAATTCC TTGTAAAACC 121 TTCATTTGAT CCTAATCTCA GTGAATTAAG AGAAATAATG AATGACTTGG AAAAGAAGAT 181 GCAGTCAACA TTAATAAGTG CAGCCAGAGA TCTTGgtaag aatgggtcat tggaggttgg 241 aataattctt ttgtctatac actgtataga caaaatattg atgccagaat tattttataa
10	1 tttttcacta tttaaataat tttagaatta cattgaaaaa tggtagtagg tatttatgga 61 atacttttttc ttttcttctt gtttatcaag GCTTGGACCC TGGCAAACAG ATTAAACTGG 121 ATTCCAGTGC ACAGTTTGGA TATTACTTTC GTGTAACCTG TAAGGAAGAA AAAGTCCTTC

	181 GTAACAATAA AAACTTTAGT ACTGTAGATA TCCAGAAGAA TGGTGTTAAA TTTACCAACA 241 Ggtttgcaag tcattattat atttttaacc ctttattaat tccctaaatg ctctaacatg 301 atgtgaatgt tctatgataa gttttactaa tgtagtcatc aggtaagagt caagctttct
11	1 ttccatttta attatcatta tttggatgtt tcataggata ctttggatat gtttcacgta 61 gtacacattg cttctagtac acattttaat atttttaata aaactgttat ttcgatttgc 121 agCAAATTGA CTTCTTTAAA TGAAGAGTAT ACCAAAAATA AAACAGAATA TGAAGAAGCC 181 CAGGATGCCA TTGTTAAAGA AATTGTCAAT ATTTCTTCAG gtaaacttaa tagaactaat 241 aatgttctga atgtcacctg gcttttggta acagaagaaa aatcatgata tttgaagtgt
12	1 tttaaaaaag tgtaaattag gaaatgggtt ttgaattccc aaatgggggg attaaatgta 61 tttttacggc ttatatctgt ttattattca gtattcctgt gtacatttc tgtttttatt 121 TTTATACAGG CTATGTAGAA CCAATGCAGA CACTCAATGA TGTGTTAGCT CAGCTAGATG 181 CTGTTGTCAG CTTTGCTCAC GTGTCAAATG GAGCACCTGT TCCATATGTA CGACCAGCCA 241 TTTTGGAGAA AGGACAAGGA AGAATTATAT TAAAAGCATC CAGGCATGCT TGTGTTGAAG 301 TTCAAGATGA AATTGCATTT ATTCCTAATG ACGTATACTT TGAAAAAGAT AAACAGATGT 361 TCCACATCAT TACTGgtaaa aaacctggtt tttgggcttt gtgggggtaa cgtttgtttt
13	1 tcagtgtaca gtttaggact aacaatccat ttattagtag cagaaagaag tttaaaatct 61 tgctttctga tataatttgt tttgtagGCC CCAATATGGG AGGTAAATCA ACATATATTC 121 GACAAACTGG GGTGATAGTA CTCATGGCCC AAATTGGGTG TTTTGTGCCA TGTGAGTCAG 181 CAGAAGTGTC CATTGTGGAC TGCATCTTAG CCCGAGTAGG GGCTGGTGAC AGTCAATTGA 241 AAGGAGTCTC CACGTTCATG GCTGAAATGT TGGAAACTGC TTCTATCCTC AGgtaagtgc 301 atctcctagt ccctgaaga tagaaatgta tgtctctgtc ctgtgagaag gaaaagtata
14	1 taagatgatc cattatgttt gtggcatatc cttcccaatg tattgtcata attttgtttt 61 tgtatgtgta tgttaccaca tttatgtga tgggaaattt catgtaatta tgtgcttcag 121 GTCTGCAACC AAAGATTCAT TAATAATCAT AGATGAATTG GGAAGAGGAA CTTCTACCTA 181 CGATGGATTT GGGTTAGCAT GGGCTATATC AGAATACATT GCAACAAAGA TTGGTGCTTT 241 TTGCATGTTT GCAACCCATT TTCATGAACT TACTGCCTTG GCCAATCAGA TACCAACTGT 301 TAATAATCTA CATGTCACAG CACTCACCAC TGAAGAGACC TTAACTATGC TTTATCAGGT 361 GAAGAAAGgt atgtactatt ggagtactct aaattcagaa cttggtaatg ggaaacttac
15	1 cgaggtgaga ggataaatcc attacataaa ttgctgtctc ttctcatgct gtcccctcac 61 gcttccccaa atttcttata gGTGTCTGTG ATCAAAGTTT TGGGATTCAT GTTGCAGAGC 121 TTGCTAATTT CCCTAAGCAT GTAATAGAGT GTGCTAAACA GAAAGCCCTG GAACTTGAGG 181 AGTTTCAGTA TATTGGAGAA TCGCAAGGAT ATGATATCAT GGAACCAGCA GCAAAGAAGT 241 GCTATCTGGA AAGAGAGgtt tgtcagtttg ttttcatagt ttaacttagc ttctctatta 301 ttacataaac aggacactaa gatgaaggtt ttttgtcgtc gtttgtttcc ctctgtgttt
16	1 tgtgggagga gtttgagacc accctgggcn ccatagtgag accctcttct ctcaaaatat 61 gaaaaataaa aaaaaaaatt tttaaatgtg tgatatgttt agatggaaat gnnacaattt 121 gtcactctct cacatgactt ttagaaaaga tattttaatt actcatggga cattcacatg 181 tgtttcagca aggtgaaaAA ATTATTCAGG AGTTCCTGTC CAAGGTGAAA CAAATGCCCT 241 TTACTGAAAT GTCAGAAGAA AACATCACAA TAAAGTTAAA ACAGCTAAAA GCTGAAGTAA 301 TAGCAAAGAA TAATAGCTTT GTAAATGAAA TCATTTCACG AATAAAAGTT ACTACGTGAa 361 aaatcccagt aatggaatga aggtaatatt gataagctat tgtctgtaat agttttatat

Appendix F: Protocols Used

F.1 MLPA Protocol

1. Performed using the SALSA P003-B2 MLPA probe mix (MRC-Holland, Holland)
2. Dilute DNA to 40ng/µl with TE Buffer
3. Aliquot 5µl to a 0.5ml PCR tube and allow to re-suspend for 45 minutes
4. Denature DNA at 98°C for 5 minutes
5. Add the 3µl of the hybridisation mix to (Table F.1) each DNA sample

Table F.1: Hybridisation Mix for MLPA Technique

Reagent	1x Volume
SALSA Probe mix	1.5µl
MLPA Buffer	1.5µl
Total	**3µl**

6. Hybridise at 95°C for 1 minute, then 60°C for 18 hours and hold at 54°C
7. Add Ligation mix (Table F.2) to the 8µl of the hybridised products

Table F.2: Ligation mix for MLPA Technique

Reagent	1x Volume
Ligase-65 Buffer A	3µl
Ligase-65 Buffer B	3µl
ddH_2O	25µl
Ligase-65	1µl
Hybridised Product	8
Total	40µl

8. Ligate at 54°C for 15 minutes, then denature the ligase at 98°C for 5 minutes
9. Prepare the hot-start PCR reaction mix (Table F.3)

Table F.3: Pre-PCR reaction mix for MLPA Technique

Reagent	1x Volume
MLPA Ligated Products	10µl
SALSA PCR Buffer	4µl
ddH_2O	26µl
Total	40µl

10. Heat the pre-PCR reaction mix to 60°C in an Applied Biosystems 2720 Thermal cycler (Applied Biosystems, USA)

11. Add the polymerase mix (Table F.4) to the ligated products

Table F.4: Polymerase Mix for MLPA Technique

Reagent	1x Volume
SALSA PCR Primers	2µl
SALSA Enzyme Dilution Buffer	2µl
ddH2O	5.5µl
SALSA Polymerase	0.5µl
Total	10µl

12. Start the PCR using the conditions given in Table F.5:

Table F.5: PCR Conditions for MLPA Technique

Temperature	Time	Cycles
95°C	30 Seconds	35 Cycles
60°C	30 Seconds	
72°C	1 minute	
72°C	20 minutes	

13. Load 1µl of PCR product with 0.5µl Liz-500 marker (Applied Biosystems, USA) and 8.5µl of Hi-Di Formamide into a 96-well optical plate
14. Load plate into 3130xl Genetic Analyser and follow Protocol for MLPA on the 3130xl Genetic Analyser

F.2 Protocol for MLPA on the 3130xl Genetic Analyser

Creating a Panel for the MLPA Run

1. Open GeneMapper v4.0 Software (Applied Biosystems, USA)
2. Create a new Panel for the P003 MLPA kits:
    a. Click 'Panel Manager'→ 'File'→ 'New Kit'
    b. Highlight the new panel in the window on the left
    c. Click 'File'→ New Panel
    d. Highlight new panel
    e. Click 'File'→ New Marker
    f. Give the marker a name, size and the dye colour it will be
    g. Then highlight the new marker
    h. Click 'Bins'→ Add Bin
    i. Highlight bin area→ Give bin a name, location and size range
    j. Repeat until all markers have been named and sized
3. Then need to ensure that the Panel uses correct analysis method:
    a. Tools→ GeneMapper Manager→ Analysis Methods

b. Open the 'Allele' tab
c. Select 'HNPCC' as the Bin Set to use

4. Can Adjust the peak detection if peaks not called:
    a. Tools→ GeneMapper Manager→ Analysis Methods
    b. Open the 'Peak Detector' tab
    c. Select Advanced
    d. Lower the peak intensities value to detect lower peaks

Preparing the Sample Sheet for the MLPA Run

1. Open the Foundations Data Collection v3.0 software (Applied Biosystems, USA)
2. To create a plate record:
    a. GA Instruments→ga3130xl→Plate Manager
    b. Click 'New' icon
    c. A new window will open to input information for the run:
        i. Name the plate so it can save results under that name
        ii. Can provide a description of that plate
        iii. Application is set to '*GeneMapper-ABI-3130XL*'
        iv. Plate type is set to 96 well plate
        v. Owner name and operator are added
        vi. Click 'OK'
3. The sample sheet will open
4. Add the sample names to the corresponding wells
5. Select following parameters into the sample sheet:
    a. Sample Type is set to Sample
    b. Priority remains at 100
    c. Size Standard is set to GS500(-250)LIZ
    d. Panel is set to 'P003' (Panel created for this particular MLPA kit)
    e. Analysis Method is set to 'HNPCC P003 MLPA' (Created for this MLPA kit)
    f. SNP Set, Study and User Defined columns left blank
    g. Results group is *Molecular_Genetics*
    h. Instrument Protocol is set to 'MLPA_G5_Dye'
    i. Save the sample sheet and close

Preparing the MLPA Run

1. Place the 96-well plate into the 3130xl Genetic Analyser
1. Click on 'Run Scheduler' icon
    a. GA Instruments→ ga3130xl→3130xl→ Run Scheduler
2. Update the list to include the newly created sample sheets by clicking 'Find All'
3. Find the name of the sample sheet for the plate
4. Highlight the name of your plate and then link it to the plate by clicking on the yellow onscreen plate that corresponds to your plate (Slot A or B)
5. Can check you have linked the correct plate and runs by clicking 'Run View'
    a. GA Instruments→ ga3130xl→3130xl→ Run Scheduler→ Run View

6. To start the run, click the green 'play' arrow in the top task bar and confirm that you want to start by clicking 'OK'

F.3 PCR Reaction and cycling conditions

1. A master mix for the PCR reaction is prepared with all reagents given in Table F.6, except the DNA
2. Appropriate volumes are aliquoted into an Eppendorf tube
3. 19ul of the reaction mix are aliquoted into a well in strip tubes or a single PCR tube
4. 1ul of DNA are aliquoted into each tube

Table F.6: PCR reaction mix used for amplification of all exons for 1X reaction

Reagent	Volume (µl)	Stock	Final
Buffer	2	10X	1X
Forward Primer	1	10µM	1µM
Reverse Primer	1	10µM	1µM
$MgCl_2$	2	10X	1X
dNTP	2	10X	1X
Taq	0.2	5U/µl	1U/µl
DNA	1		50ng
ddH_2O	10.8		
Total	20		

Ampli*Taq* Gold polymerase buffer (Applied Biosystems, USA),
Ampli*Taq* Gold DNA polymerase (Applied Biosystems, USA)
$MgCl_2$ (Applied Biosystems, USA
dNTPs (Invitrogen, USA)

5. The strip tubes /PCR tubes are transferred to an Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, USA)
6. The PCR conditions given in Table F.2 were used for every PCR reaction

Table F.7: PCR conditions for amplification of all reactions

Step	Temperature	Time	
Initial Denaturation	95°C	3 minutes	
Denaturation	95°C	1 minute	35 Cycles
Annealing	55°C	1 minute	
Extension	72°C	1 minute	
Final Extension	72°C	5 minutes	
Hold	4°C	∞	

F.4 Minelute 96 UF PCR Purification Kit Protocol

The Minelute 96 UF PCR purification kit (Qiagen, Germany) was used to purify higher numbers of samples due to the high throughput capabilities of the kit.

1. Place Minelute 96 UF PCR Purification plate on top of vacuum manifold
2. Pipette entire PCR sample into the wells of the plate
3. Apply a vacuum of approximately -800mbar until the wells are completely dry. Switch off vacuum source
4. Add 50µl deionized water to each well, apply vacuum (-800mbar) until the wells are completely dry. Switch off vacuum source
5. Carefully remove plate from the vacuum manifold
6. Tap the plate on a stack of clean absorbent paper to remove any liquid that might remain on bottom of plate
7. Add 20µl of elution buffer to each well
8. Dissolve purified DNA by pipetting samples up and down approximately 20 times
9. Recover purified PCR product by pipetting eluent out of each well and into a new 96 well plate or strip tubes

F.5 Nucleospin Extract II Clean-up Protocol:

The Nucleospin Extract II clean-up kit (Machery-Nagel, Germany) was used for individual samples and small batches

1. Adjust PCR product volume to 50µl with TE buffer
2. Mix 100µl of NT buffer with the 50µl of PCR product
3. Place column in collection tube and load entire sample onto column
4. Centrifuge for 1 minute at 14 000rpm. Discard flow-through and place column back into collection tube
5. Add 600µl NT3 buffer and centrifuge for 1 minute at 14 000rpm. Discard flow-through and place column back into collection tube
6. Centrifuge for 2 minutes at 14 000rpm to quantitatively remove residual NT3 buffer
7. Incubate samples at 70°C for 5 minutes to evaporate residual ethanol
8. Place column in clean Eppendorf tube
9. Add 20µl NE elution buffer and incubate for 1 minute at room temperature
10. Centrifuge at 14 000rpm for 1 minute, keep flow through and discard column

F.6 Sanger Sequencing

The Sanger sequencing protocol is given in Section 2.

The reagents used in the Sanger sequencing protocol are given in Table F.8 and the thermal cycling conditions are given in Table F.9.

Table F.8: Sanger Sequencing PCR Reaction mix

Reagent	Stock	Final	Volume (µl)
Cleaned PCR product			10µl
BigDye v3.1			4µl
BigDye Buffer	5x	0.5x	2µl
Primer	3.3µmol/µl	3.3 µmol	1µl
ddH_2O			3µl
Total			20µl

Table F.9: Cycle Sequencing PCR Protocol

Temperature	Time	Cycles
96°C	30 seconds	25 Cycles
50°C	15 seconds	
60°C	4 minutes	
10°C	∞	

F.7 DyeEx 96 Dye-Terminator Removal Protocol

The DyeEx 96 Dye-Terminator kit (Qiagen, Germany) was used to purify the Sanger sequencing products for large batches

1. Take DyeEx 96 plate out of bag and remove the tape sheets from the top and bottom of the plate
2. Place DyeEx 96 plate on top of collection plate and centrifuge for 3 minutes at 2 000rpm
3. Carefully place the DyeEx 96 plate on an appropriate elution plate with a suitable adaptor
4. Apply the sequencing samples in a volume of 20µl to the gel beds of each well. Ensure not to touch the sides of the wells and the gel surface with the pipette tip
5. Centrifuge for 3 minutes at 2 000rpm
6. Dry the samples in a concentrator 5301 (Eppendorf, Germany) and proceed according to instructions provided for DNA sequencer

F.8 DyeEx 2.0 Dye-Terminator Removal Modified Protocol

The DyeEx 2.0 Dye-Terminator kit (Qiagen, Germany) was used to purify Sanger sequencing products for individual and small batches

1. Vortex spin column to re-suspend gel and remove bubbles.
2. Loosen the column cap a quarter turn and snap off bottom closure of spin column. Place column into collection tube
3. Centrifuge for 3 minutes at 3 000rpm. Discard flow-through
4. Place spin column into clean Eppendorf tube

5. Load sequencing products onto gel bed. Take care not to pipette sample onto sides of column or touch the gel bed with pipette tip
6. Centrifuge for 3 minutes at 3000rpm. Discard spin column
7. Dry sample in a concentrator 5301 (Eppendorf, Germany) and proceed according to instructions provided for DNA sequencer

F.9 Protocol for Sanger Sequencing using the 3130xl Genetic Analyser

Preparing the Sample Sheet

1. Open Foundation Data Collection v3.0 software
2. To create a plate record:
    a. GA Instruments→ga3130xl→Plate Manager
    b. Click 'New' icon
    c. A new window will open to input information for the run:
        i. Name the plate so it can save results under that name
        ii. Can provide a description of that plate
        iii. Application is set to 'Sequencing Analysis'
        iv. Plate type is set to 96 well plate
        v. Owner name and operator are added
    d. Click 'OK'
3. The sample sheet will open
4. Add the sample names to the corresponding wells
5. Select following parameters into the sample sheet:
    a. Priority set to 100
    b. Results group is *Molecular_Genetics*
    c. Instrument protocol is set to *UltraSeq36_POP7_Z* (This is used for samples smaller than 500bp)
    d. Analysis protocol is set to *3130POP7_BDTv3-KB-Denovo_v5.2*

Starting the Sequencing Run

7. Place the 96-well plate in the 3130xl Genetic Analyser
8. Click on 'Run Scheduler' icon
    a. GA Instruments→ ga3130xl→3130xl→ Run Scheduler
9. Update the list to include the newly created sample sheets by clicking 'Find All'
10. Find the name of your plate
11. Highlight the name of your plate and then link it to the plate by clicking on the yellow onscreen plate that is positioned where your plate is (Slot A or B)
12. Can check you have linked the correct plate and runs by clicking 'Run View'
13. To start the run, click the green arrow in the top task bar and confirm that you want to start by clicking 'OK'

F.10 Retrieval of Data

Retrieval of Sanger Sequencing Data

1. Open the Sequencing Analysis v5.2 software
2. To add sample files to be analysed:
    a. File→ Add Samples
    b. Select folder containing sample files
    c. Click 'Add Selected Samples'
3. The sample manager window will open with all the files selected for analysis
4. Each BC box is already ticked for base calling. Click the green arrow in task bar to start base calling of the ticked samples
5. To view the electropherogram for a sample:
    a. Tick the box in the 'Show" column
    b. The electropherogram opens below the sample manager window
6. Save the analysis:
    a. File→ Save Samples
7. These saved results can be transferred to other storage media (such as a USB flash memory drive) and printed

Retrieval of the MLPA Run

1. Open the GeneMapper v4.0 Software
2. A new project is already open
3. To add sample files to be analysed:
    a. File→ Add Samples to Project
    b. Find the folder with the files
    c. Highlight the folder→ Click 'Add to list'
    d. The click 'Add and Analyse once all files selected into list
4. The 'Save Project As' window opens
    a. Name your project and click the ok button
5. The sample manager window will open and automatically analyse the samples
6. The Project will now be analysed automatically
    a. The "SQ" column should show a green square or a yellow triangle to pass analysis
    b. If a red circle is shown:
        i. Highlight the sample row and click 'Size Match Editor' (icon with three red peaks)
        ii. Adjust the sizes of the peaks as needed to fit the size standard
        iii. If changes made, highlight the sample with changed size standard
            1. Analysis→ Analyse Selected Samples
7. Highlight sample and click 'Display plots' (Icon with different coloured peaks)
8. Can choose which colour to view
9. To zoom plots:
    a. Move cursor onto edge of the plots so that magnifying glass appears
    b. Click and drag over the area to zoom into

c. Double click to return to full view

10. For peaks that were not sized:
    a. Click on peak to highlight
    b. Right click and select 'Add allele call'
11. Once all peaks are called, can prepare sheet for export
    a. Return to Project Main screen
    b. Click 'Genotypes' Tab
    c. Columns with each peak and size should appear
    d. Click File→ Export
        i. Export the file as a comma delimited file (.csv)
12. Download the 'P003-B1 HNPCC' Spread sheet from www.ngrl.org.uk
13. Open the exported .csv file
    a. Highlight all the text→ Copy→ Paste into a blank excel spread sheet
    b. The spread sheet should show each sample number, the name of the marker and the peak size of that marker
    c. Copy those three columns and paste them into the corresponding marker columns in the P003-B1 HNPCC spread sheet
14. Once all the markers and intensities have been pasted, the spread sheet can display the mathematical data as well as neat graphical data
    a. Click on the 'Results' tab to view graphical and mathematical data

F.11 Using the LOVD InSiGHT Database

1. Access www.insight-group.org on your internet browser
2. Select 'Mutations' Tab
3. Select the Gene database you wish to examine
4. Takes you to the main page for that gene
5. To view which variants have been reported:
    a. Click 'Variants' tab
    b. Can search for variants by exon number or using the variant nomenclature if you know the DNA, RNA or protein change
    c. If variant has been published, you can retrieve the publication
        i. In the 'Variant Reference' column
        ii. Click on hyperlinked name to go to PubMed where article abstract can be read
6. To determine the nomenclature of variants found in Sanger sequencing analysis
    a. In main page, click on 'coding DNA reference sequence'
    b. Displays the coding DNA sequence as taken from particular transcripts
    c. The sequence provides numbering so that you can determine the position of variants
    d. Intron sequences can be seen by clicking on each exon number
    e. Intron sequence also provides numbering to determine the position of variant

F.12 High Resolution Melt Procedure

1. DNA samples are diluted to 50ng/µl with TE buffer
2. HRM master mix is made up for each variant

Table F.10: HRM master mix that was used for all HRM reactions

Reagent	Concentration	Volume (µl)
SensiMix HRM	1X	12.5
EvaGreen dye		1
Primers	10µM	0.5
DNA template	50ng/µl	1
$MgCl_2$		1
ddH_2O		9µl
Total		25µl

3. Master mix is aliquoted into PCR tubes
4. Tubes are placed in the Illumina ECO (Illumina, USA)
5. ECO software program is opened on the PC connected to the Illumina ECO
6. HRM Analysis protocol is selected
7. Each tube position is labelled and the control samples are selected as controls
8. Conditions for the HRM are pre-set but the primer melting temperature must be adjusted (55°C for primers in this study).
9. PCR and HRM are initiated when the 'play' button in the ECO software program is pressed.

Appendix G: Patient Pedigrees

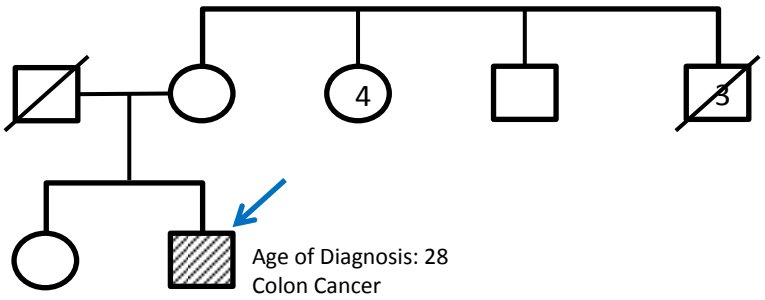


Figure G.1: The pedigree for black patient 205. This patient presented with colon cancer at 28 but has no known family history of the disease. This patient was found to have seven variants, two of which were not previously reported (*MLH1* c.1732-19T>A and *MSH2* c.-185C>A). The numbers indicate multiple females and deceased males

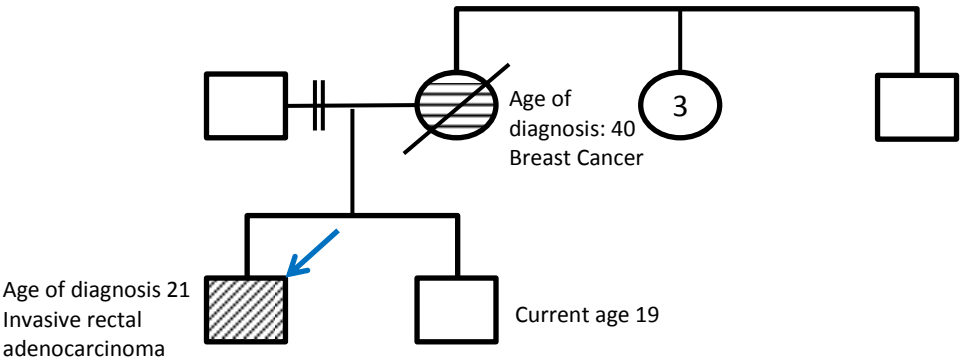


Figure G.2: Pedigree for black patient 235. Family history is does not fulfil the Amsterdam II criteria but patient developed rectal neoplasm at age 21. This patient was found to have five variants, four of which are known to be non-pathogenic and the fifth is a novel variant (*MSH2* c.1662-10C>T). The number '3' indicates three females

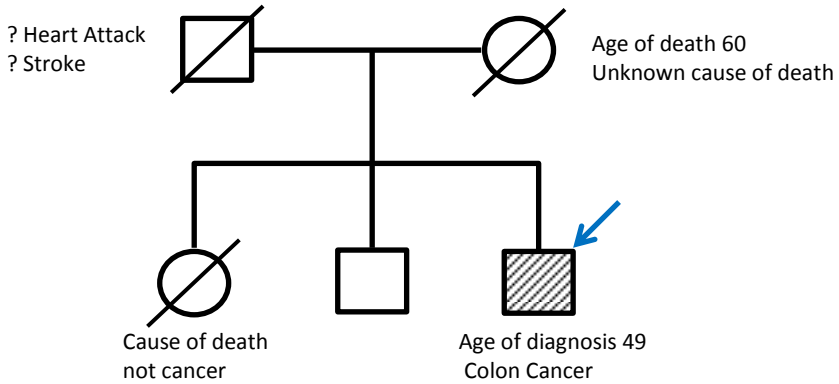


Figure G.3: Pedigree for black African patient 238. Four variant were detected in this patient, two of which are novel variants not previously reported.

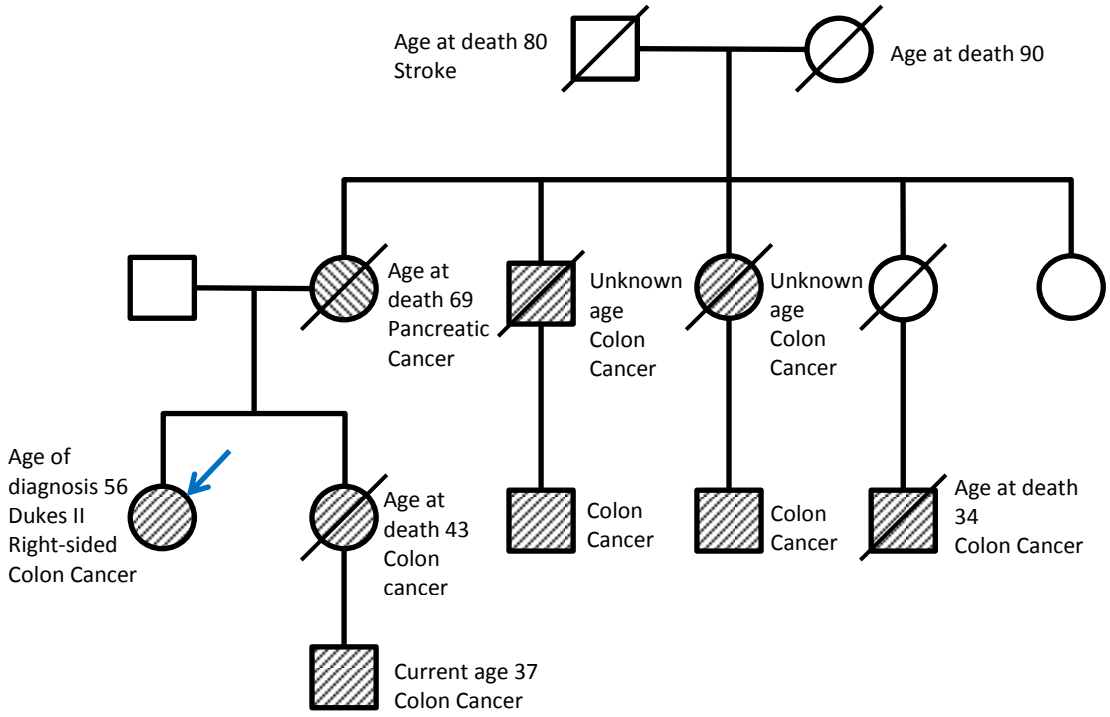


Figure G.4: Pedigree for patient 136. This white patient had a prominent family history of colon cancer and her mother had pancreatic cancer. The *MLH1* c.731_734delGTTA pathogenic mutation was found in this patient, which leads frame shift and a truncated protein product. The other family members must be screened for the presence of the mutation.

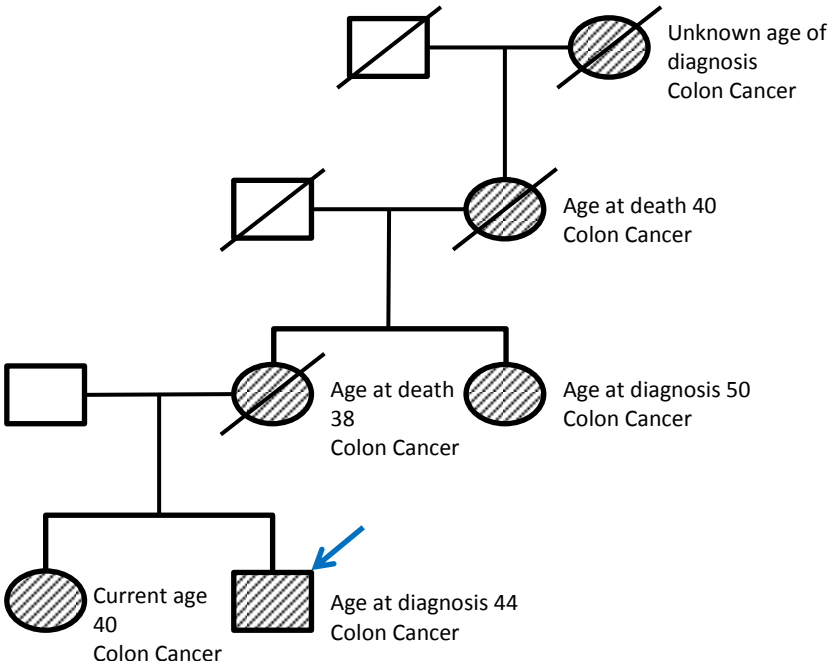


Figure G.5: The pedigree for patient 150 where the *MSH2* c.1811C>A variant was found. This variant was first found by the Wessex Regional Genetics Laboratory who predicted the variant to be pathogenic. In addition, SIFT and PolyPhen also predicted this missense mutation to have a deleterious effect. The patient's sister, who also developed colon cancer, was also investigated by the Wessex Regional Genetics Laboratory and found to have the same variant. This provides strong evidence that this variant is pathogenic

Appendix H: Suppliers of Equipment and Reagents

Table H.1: Table of Suppliers of Equipment and Reagents

Equipment	Supplier	Country
3130xl Genetic Analyser	Applied Biosystems	USA
GeneAmp 2720 Thermal Cycler	Applied Biosystems	USA
Foundation Data Collection v3.0	Applied Biosystems	USA
GeneMapper v4.0	Applied Biosystems	USA
Sequence Analysis v5.2	Applied Biosystems	USA
AmpliTaq Gold PCR Mix (Incl. buffer and $MgCl_2$)	Applied Biosystems	USA
BigDye Terminator v3.1 Sequencing Kit	Applied Biosystems	USA
Hi-Di Formamide	Applied Biosystems	USA
dNTP Mix	Invitrogen	USA
1kb+ DNA Ladder	Invitrogen	USA
DNAStar Seqman II	Lasergene	USA
Restriction Enzymes (HpyCH4III, RsaI & MluI)	New England Biolabs	USA
Restriction Enzymes (TaqI)	Thermo-Fisher Scientific	USA
Primers	Whitehead Scientific	RSA
Primers	Inqaba Biotec	RSA
SensiMix HRM Kit	Bioline	UK
Agarose Gel	Bioline	UK
Vacutec G:Box	Syngene	UK
GeneSnap v6.08	Syngene	UK
Ethidium Bromide	Sigma-Aldrich	UK
Minelute 96 UF Purification Kit	Qiagen	Germany
DyeEx 96 Kit	Qiagen	Germany
DyeEx 2.0 Kit	Qiagen	Germany
Nucloespin Extract II Column Clean-up Kit	Machery-Nagel	Germany
Concentrator 5301	Eppendorf	Germany
LightCycler 480	Roche	Switzerland

References:

Acharya, S., Wilson, T., Gradia, S., Kane, M.F., Guerrette, S., Marsischy, G.T., et al. 1996. hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc Natl Acad Sci*. 93(24): 13629-13634.

Agrawal, S., Bhupinderjit, A., Bhutani, M.S., Boardman, L., Nguyen, C., Romero, Y., et al. 2005. Colorectal cancer in African Americans. *American Journal of Gastroenterology*. 100: 515-523

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. 2002. Molecular Biology of the Cell 4th edition. Garland publishing. United States of America

Albertson, D.G. 2006. Gene amplification in cancer. *Trends in Genetics*. 22(8): 447-455

Anderson, W.F., Umar, A., Brawley,O.W. 2003. Colorectal cancer in black and white race. *Cancer and Metastasis Reviews*. 22: 67-82

Arnold, S., Buchanan, D.D., Barker, M., Jaskowski, L., Walsh, M.D., Birney, G., et al. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation and tumor characteristics. *Hum Mutat*. 30 (5): 757-770

Auclair, J., Buisine, M.P., Navarro, C., Ruano, E., Montmain, G., Desseigne, F., et al. 2006. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. *Hum Mutat*. 27 (2): 145-154

Bagnoli. S., Putignano, A.L., Melean, G., Baglioni, S., Sestini, R., Milla, M., et al. 2004. Susceptibility to refractory ulcerative colitis is associated with polymorphism in the hMLH1 mismatch repair gene. *Inflamm Bowel Dis*. 10(6): 705-708

Beiner, M. E., Rosen, B., Fyles, A., Harley, I., Pal, T., Siminovitch, K., et al. 2006. Endometrial cancer risk is associated with variants of the mismatch repair genes MLH1 and MSH2. *Cancer Epidemiol Biomarkers Prev*. 15: 1636-1640

Boland, C.R., Koi, M., Chang, D.K., Carethers, J.M. 2008. The biochemical basis of microsatellite instability and abnormal immunohistochemical and clinical behaviour in Lynch syndrome: from bench to bedside. *Familial Cancer*. 7:41-52

Boland, C.R., Goel, A. 2010. Microsatellite instability in colorectal cancer. *Gastroenterology*. 138: 2073-2087

Borresen, A.L., Lothe, R.A., Meling, G.I., Lystad, S., Morrison, P., Lipford, J. 1995. Somatic mutations in the hMSH2 gene in microsatellite unstable colorectal carcinomas. *Hum Mol Genet*. 4(11): 2065-2072

Botma, A., Nagengast, F.M., Braem, M.G.M., Hendriks, J.C.M., Kleibeuker, J.H., Vasen, H.F.A. et al. 2010. Body mass index increases risk of colorectal adenoma in men with Lynch syndrome: the GEOLynch cohort study. *Journal of Clinical Oncology*. 28 (28): 4346-4353

Brentnall, T.A., Rubin, C.E., Crispin, D.A, Stevens, A., Batchelor, R.H., Haggit, R.C. et al. 1995. A germline substitution in the human MSH2 gene is associated with high-grade dysplasia and cancer in ulcerative colitis. *Gastroenterology*. 109(1): 151-155

Campbell, M.C., Tishkoff, S.A. 2008. African genetic diversity: Implications for Human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*. 9: 403–433.

Campbell, P.T., Curtin, K., Ulrich, C.M., Samowitz, W.S., Bigler, J., Velicer, C.M., et al. 2009. Mismatch repair polymorphisms and risk of colon cancer, tumour microsatellite instability and interactions with lifestyle factors. *Gut*. 58 (5): 661-667

Capellá, G., Pera, G., Sala, N., Agudo, A., Rico, F., del Giudicce, G., et al. 2008. DNA repair polymorphisms and the risk of stomach adenocarcinoma and sever chronic gastritis in the EPIC-EURGAST study. *International Journal of Epidemiology*. 37: 1316-1325

Carrera-Bastos, P., Fontes-Villalba, M., O'Keefe, J.H., Lindeberg, S., Cordains, L. 2011. The western diet and lifestyle and diseases of civilization. *Research Reports in Clinical Cardiology*. 2011(2): 15-35

Chan, T.L., Chan, Y.W., Ho, J.W., Chan, C., Chan, A.S., Chan, E., et al. 2004. MSH2 c.1452-1455delAATG is a founder mutation and an important cause of hereditary nonpolyposis colorectal cancer in the Southern Chinese population. *Am J Hum Genet*. 74(5): 1035-1042

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

Chew, M.H., Koh, P.K., Ng, K.H., Eu, K.W. 2009. Improved survival in an Asian cohort of young colorectal cancer patients: an analysis of 523 patients from a single institution. *International Journal of Colorectal Disease*. 24(9): 1075-1083

Christensen, L.L., Madsen, B.E., Wikman, F.P., Wiuf, C, C., Koed, K., Tjønneland, A., et al. 2008. The association between genetic variants in *hMLH1* and *hMSH2* and the development of sporadic colorectal cancer in the Danish population. *BMC Medical Genetics*. 9:52

Clendenning, M., Baze, M.E., Sun, S., Walsh, K., Liyanarachchi, S., Fix, D., et al. 2008. Origins and prevalence of the American founder mutation of MSH2. *Cancer Res*. 68: 2145-2153

Croce, C.M. 2008. Oncogenes and cancer. *NEJM*. 358: 502-511

Cronjé, L., Becker, P.J., Paterson, A.C., Ramsay, M. 2009. Hereditary non-polyposis colorectal cancer is predicted to contribute towards colorectal cancer in young South African blacks. *South African Journal of Science*. 105: 68-72

Cunningham, D., Atkin, W., Lenz, H.J., Lynch, H.T., Minsky, B., Nordlinger, B., Starling, N. 2010. Colorectal cancer. *Lancet*. 375:1030-1047

da Silva, F.C., Valentin, M.D., Ferreira, F.O., Carraro, D.M., Rossi, B.M. 2009. Mismatch repair genes in Lynch syndrome: a review. *Sao Paulo Med J*. 127(1): 46-51

de la Chapelle, A., Wright, F.A. 1998. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc Natl Acad Sci USA*. 95: 12416-12423

de la Chapelle, A. 2004. Genetic predisposition to colorectal cancer. *Nature Reviews Cancer*. 4: 769-780

Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., et al. 2003. The case for early detection. *Nature Reviews Cancer*. 3: 1-10

Farrington, S.M., Lin-Goerke, J., Ling, J., Wang, Y., Burczak, J.D., Robbins, D.J., et al. 1998. Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. *American Journal of Human Genetics*. 63: 749-759

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin, D.M. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008*. Int J Cancer*. 127(12): 2893-2917

Fishel, R. 2001. The Selection for Mismatch Repair Defects in Hereditary Nonpolyposis Colorectal Cancer**:** Revising the Mutator Hypothesis**.** *Cancer Res*. 61: 7369-7374

Fletcher, O., Houlston, R.S. 2010. Architecture of inherited susceptibility to common cancer. *Nat Rev Cancer*. 10: 353-361

Fokkema, I.F., Taschner, P.E., Schaafsma, G.C., Celli, J., Laros, J.F., den Dunnen, J.T. 2011. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*. 32(5):557-63

Foulkes, W.D., Thiffault, I., Gruber, S.B., Horwitz, M., Hamel, N., Lee, C., et al. 2002. The founder mutation MSH2*1906G→C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. *American Journal of Human Genetics*. 71(6): 1395-1412

Frank, S.A. 2004. Genetic predisposition to cancer- insights from population genetics. *Nature reviews genetics*. 5:764-772

Fredriksson, H., Ikonen, T., Autio, V., Matikainen, M.P., Helin, H.J., Tammela, T.L.J., et al. 2006. Identification of germline MLH1 alterations in familial prostate cancer. *European Journal of Cancer*. 42: 2802-2806

Garber, J.E., Offit, K. 2005. Hereditary cancer predisposition syndromes. *Journal of Clinical Oncology*. 23(2):276-292

Gazzoli, I., Loda, M., Garber, J.E., Syngal, S., Kolodner, R.D. 2002. A hereditary nonpolyposis colorectal cancer carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. *Cancer Res*. 62: 3925-3928

Gille, J.J.P., Hogervorst, F.B.L., Pals, G., Wijnen, J.T.H., van Schooten, R.J., Dommering, C.J., et al. 2002. Genomic deletions of MSH2 and MLH1 in colorectal cancer families detected by a novel mutation detection approach. *British Journal of Cancer*. 87: 892-897

Goldberg, P.A., Madden, M.V., Haracopos, C., Felix, R., Westbrook, C., Ramesar, R.S. 1998. In a resource-poor country, mutation identification has the potential to reduce the cost of family management for hereditary nonpolyposis colorectal cancer. *Dis Colon Rectum*. 41(10): 1250-1253

Gryfe, R. 2009. Inherited colorectal cancer syndromes. *Clinics in Colon and Rectal Surgery*. 22(4): 198-208

Gundry, C.M., Vandersteen, J.G., Reed, G.H., Pryor, R.J., Chen, J., Wittwer, C.T. 2003. Amplicon melting analysis with labelled primers: A closed-tube method for differentiating homozygotes and heterozygotes. *Clinical Chemistry*. 49(3): 396-406

Hamilton, S.R., Liu, B., Parsons, R.E., Papadopoulos, N., Jen, J., Powell, S.M., et al. 1995. The molecular basis of Turcot's syndrome. *New England Journal of Medicine*. 332 (13): 839-847

Hesson, L.B., Hitchins, M.P., Ward, R.L. 2010. Epimutations and cancer predisposition: importance and mechanisms. *Current Opinions in Genetics and Development*. 20: 1-9

Hitchins, M.P., Ward, R.L. 2009. Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *Journal of Medical Genetics*. 46: 793-802

Hitchins, M.P., Owens, S.E., Kwok, C.T., Godsmark, G., Algar, U.F., Ramesar, R.S. 2011. Identification of new cases of early-onset colorectal cancer with an *MLH1* epimutation in an ethnically diverse South African cohort. *Clin Genet*. doi: 10.1111/j.1399-0004.2011.01660.x

Hudler, P., Voulk, K., Liovic, M., Repse, S., Juvan, R., Komel, R. 2004. Mutations in the *hMLH1* gene in Slovenian patients with gastric carcinoma. *Clin Genet*. 64: 405-411

Jasperson,K.W., Tuchy. T.M., Neklason, D.W., Burt, R.W. 2010. Hereditary and familial colon cancer. *Gastroenterology*. 138: 2044-2058

Järvinen, H.J., Renkonen-Sinisalo, L., Aktán-Collán, K., Peltomäki, P., Aaltonen, L.A., Mecklin, J.P. 2009. Ten years after mutation testing for Lynch Syndrome: Cancer incidence and outcomes in mutation-positive and mutation-negative family members. *Journal of Clinical Oncology*. 27(28): 4793-4797

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., Forman, D. 2011. Global cancer statistics. *CA: A Cancer Journal for Clinicians*. 61(2): 69-90

Jiricny, J., Nyström-Lahti, M. 2000. Mismatch repair defects in cancer. *Current Opinion in Genetics & Development*. 10(2): 157-161

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D. 2002a. The human genome browser at UCSC. Genome Res. 12(6): 996-1006

Kent, W.J. 2002b. BLAT- the BLAST-like alignment tool. *Genome Res*. 12(4): 656-664

Klug, W.S., Cummings, M.R., Spencer, C.A. 2006. Concepts of Genetics. Pearson Education International. United States of America

Knudson, A.G. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA*. 68(4): 820-823

Knudson, A.G. 2001. Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*. 1: 157-162

Kolodner, R.D., Hall, N.R., Lipford, J., Kane, M.F., Rao, M.R.S., Morrison, P., et al. 1994. Structure of the human MSH2 locus and analysis of two Muir-Torre kindreds for msh2 Mutations. *Genomics*. 24(3): 516-526

Kolodner, R.D., Hall, N.R., Lipford, J., Kane, M.F., Morrison, P.T., Finan, P.J., et al. 1995. Structure of the human MLH1 locus and analysis of a large hereditary nonpolyposis colorectal carcinoma kindred for MLH1 mutations. *Cancer Res*. 55: 242-248

Kowalski, L.D., Mutch, D.G., Herzog, T.J., Rader, J.S., Goodfellow, P.J. 1997. Mutational analysis of MLH1 and MSH2 in 25 prospectively-acquired RER+ endometrial cancers. *Genes Chromosomes Cancer*. 18(3): 219-227

Lagerstedt Robinson, K., Liu, T., Vandrovcova, J., Halvarsson, B., Clendennning, M., Frebourg, T., et al. 2007. Lynch syndrome (Hereditary nonpolyposis colorectal cancer) diagnostics. *Journal of the National Cancer Institute*. 99 (4): 291-299

Lebrun, C., Olschwang, S., Jeannin, S., Vandenbos, F., Sobol, H., Frenay, M. 2007. Turcot syndrome confirmed with molecular analysis. *European Journal of Neurology*. 14(4): 470-472

Ligtenberg, M.J.L., Kuiper, R.P., Chan, T.S., Goossens, M., Hebeda, K.M., Voorendt, M., et al., 2009. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nature Genetics*. 41 (1): 112-117

Liu, B., Nicolaides, N.C., Markowitz, S., Wilson, J.K., Parsons, R.E., Papadopolous, N., et al. 1995. Mismatch repair gene defects in sporadic colorectal cancer with microsatellite instability. *Nat Genet*. 9(1): 48-55

Lynch, H.T., Boland, C.R., Gong, G., Shaw, T.G., Lynch, P.M., Fodde, R., et al. 2006. Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. *European Journal of Human Genetics*. 14: 390-402

Marteau, J.B., Mohr, S., Pfister, M., Visvikis-Siest, S. 2005. Collection and storage of human blood cells for mRNA expression profiling: A 15-month stability study. *Clinical Chemistry*. 51(7): 1250-1252

Mitelman, F., Johansson, B., Mertens, F. 2007. The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*. 7(4): 233-245

Mullis, K.B., Faloona, F.A. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology*. 155: 335-350

National Cancer Registry (NCR). 2004. *Incidence of Histologically Diagnosed Cancer in South Africa, 1998-1999*. National Cancer Registry, South Africa

Ng, P.C., Henikoff, S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31(13): 3812-3814.

Nguyen, D.X., Bos, P.D., Massagué, J. 2009. Metastasis: from dissemination to organ-specific colonization. *Nature Reviews Cancer*. 9: 274-284

Nicolaides, N.C., Papadopoulos, N., Liu, B., Wei, Y.F., Carter, K.C., Ruben, S.M., et al. 1994. Mutations of two PMS homologues in hereditary non-polyposis colon cancer. *Nature*. 371: 75-80

Nystrom-Lahti, M., Parsons, R., Sistonen, P., Pylkkänen, L., Aaltonen, L.A., Leach, F.S., et al. 1994. Mismatch repair genes on Chromosomes 2p and 3p account for a major share of Heredetary Nonpolyposis Colorectal Cancer families evaluable by Linkage. *Am. J. Hum. Genet*. 55:659-665

Nystrom-Lahti, M., Wu, Y., Moisio, A.L., Hofstra, R.M., Osinga, J., Mecklin, J.P., et al. 1996. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. *Hum Mol Genet*. 5(6): 763-769

Ollberding, N.J., Nomura, A.M.Y., Wilkens, L.R., Henderson, B.E., Kolonel, L.N. 2011. Racial/ethnic differences in colorectal cancer risk: The multiethnic cohort study. *International Journal of Cancer*. 129 (8): 1899-1906

Ou, J., Niessen, R.C., Vonk, J., Westers, H., Hofstra, R.M.W., Sijmons, R.H. 2008. A database to support the interpretation of human mismatch repair gene variants. *Hum Mutat.* 29: 1337-1341

O'Brien, V., Brown, R. 2006. Signalling cell cycle arrest and cell death through the MMR system. *Carcinogenesis*. 27(4): 682-692

Park, S.J., Lee, K.A., Park, T.S., Kim, N.K., Song, J., Kim, B.Y., et al. 2008. A novel missense MSH2 gene mutation in a patient of a Korean family with hereditary nonpolyposis colorectal cancer. *Cancer Genetics and Cytogenetics*. 182: 169-139

Parmley, J.L., Hurst, L.D. 2007. How do synonymous mutations affect fitness? *Bioessays*. 29:515-519

Peltomäki, P. 2005. Lynch syndrome genes. *Familial Cancer*. 4: 227-232

Pineda, M., González, S., Lázaro, C., Blanco, I., Capellá, G. 2010. Detection of genetic alterations in hereditary colorectal cancer screening. *Mutation Research*. 693: 19-31

Ponti, G., de Leon, M.P. 2005. Muir-Torre syndrome. *Lancet Oncology*. 6: 980-987

Popat, S., Hubner, R., Houlson, R.S. 2005. Systematic review of microsatellite instability and colorectal cancer prognosis. *Journal of Clinical Oncology*. 23(3): 609-618

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

Rodriguez-Bigas, M.A., Boland, C.R., Hamilton, S.R., Henson, D.E., Jass, J.R., Khan, P.M., et al. 1997. A national cancer institute workshop on hereditary nonpolyposis colorectal cancer syndrome: Meeting highlights and Bethesda guidelines. *Journal of the National Cancer Institute*. 89 (23): 1758-1762

Rozen, S., Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist progammers. *Methods in Molecular Biology*. 132:365-386

Sanger, F., Nicklen, S., Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci.* 74(12): 5463-5467

Sargent, D.J., Marsoni, S., Monges, G., Thibodeau, S.N., Labianca, R., Hamilton, S.R. 2010. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *Journal of Clinical Oncology*. 28 (20): 3219-3226

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., Pals, G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res*. 30(12): e57

Shendure, J., Ji, H. 2008. Next-generation DNA sequencing. *Nature Biotechnology*. 26(10): 1135-1145

Shin, Y.K., Heo, S.C., Shin, J.H., Hong, S.H., Ku, J.L., Yoo, B.C. et al. 2004. Germline mutations in *MLH1, MSH2* and *MSH6* in Korean hereditary non-polyposis colorectal cancer families. *Hum Mutat*. 24(4): 351-358

Sjursen, W., Haukanes, B.I., Grindedal, E.M., Aarset, H., Stormorken, A., Engebretsen, L.F. 2010. Current clinical criteria for Lynch syndrome are not sensitive enough to identify MSH6 mutation carriers. *Journal of Medical Genetics*. 47 (9): 579-585

South, C.D., Hampel, H., Comeras, I., Westman, J.A., Frankel, W.L., de la Chapelle, A. 2008. The frequency of Muir-Torre syndrome among Lynch syndrome families*. J Natl Cancer Inst*. 100(4): 277-281

Stella, A., Surdo, N.C., Lastella, P., Barana, D., Oliani, C., Tibiletti, M.G., et al. 2007. Germline novel MSH2 deletions and a founder MSH2 deletion associated with anticipation effects in HNPCC. *Clin Genet*. 71(2): 130-139

Strachan, T., Read, A.P. 2004.Human Molecular Genetics 3rd edition. Garland publishing. United States of America

Stupart, D.A., Goldberg, P.A., Algar, U., Ramesar, R.S. 2009. Cancer risk in a cohort of subjects carrying a single mismatch repair gene mutation. *Familial Cancer*. 8 (4): 519-523

Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., Kondrashov, A.S., Bork, P. 2001. Prediction of deleterious human alleles. *Hum Mol Genet*. 10: 591-597

Takahashi, M., Shimodaira, H., Andreutti-Zaugg, C., Iggo, R., Kolodner, R.D., Ishioka, C. 2007. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. *Cancer Res*. 67: 4595-4604

Talseth-Palmer, B.A., McPhillips, M., Groombridge, C., Spigelman, A., Scott, R.J. 2010. MSH6 and PMS2 mutation positive Australian Lynch syndrome families: novel mutations, cancer risk and age of diagnosis of colorectal cancer. *Hereditary Cancer in Clinical Practice*. 8 (1): 5

Terdiman, J.P., Gum, J.R., Conrad, P.G., Miller, G.A., Weinberg, V., Crawley, S.C., et al. 2001. Efficient detection of hereditary nonpolyposis colorectal cancer gene carriers by screening for tumor microsatellite instability before germline genetic testing. *Gastroenterology*. 120: 21-30

Tomlinson, I.P.M., Beck, N.E., Homfay, T., Haracopos, C.J., Bodmer, W.F. 1997. Germline HNPCC gene variants have little influence on the risk for sporadic colorectal cancer. *Journal of Medical Genetics*. 34: 39-42

Tournier, I., Vezain, M., Martins, A., Charbonnier, F., Baert-Desurmont, S., Olschwang, S., et al. 2008. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat*. 29 (12): 1412-1424

Umar, A., Boland, C.R., Terdiman, J.P., Syngal., S., de la Chapelle, A., Rüschoff, J. et al. 2004a. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *Journal of the National Cancer Institute*. 96(4): 261-268

Umar, A., Risinger, J.I., Hawk, E.T., Barret, J.C. 2004b. Testing guidelines for hereditary non-polyposis colorectal cancer. *Nature Reviews Cancer*. 4: 153-158

Vasen H.F.A., Mecklin, J.P., Khan, P.M., Lynch, H.T. 1991. The international collaborative group on hereditary non-polyposis colorectal cancer (ICG-HNPCC). *Dis Colon Rectum*. 34(5): 424-425

Vasen, H.F.A., Watson, P., Mecklin, J.P., Lynch, H.T., ICG-HNPCC. 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the international collaborative group on HNPCC. *Gastroenterology*. 116: 1453-1456

Vasen, H.F.A., Möslein, G., Alonso, A., Bernstein, I., Bertario, L., Blanco, I., et al. 2007. Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *Journal of Medical Genetics*. 44(6): 353-362

Venables, J.P. 2004. Aberrant and alternative splicing in cancer. *Cancer Res*. 64: 7647-7654

Vogelstein, B., Kinzler, K.W. 2004. Cancer genes and the pathways they control. *Nature Medicine*. 10(8): 789-799

Wang, Y., Friedl, W., Lamberti, C., Nothen, M.M., Kruse, R., Propping, P. 1998. A novel missense mutation in the DNA mismatch repair gene hMLH1 present among East Asians but not among Europeans. *Hum Hered*. 48(2): 87-97

Wang, X.L., Yuan, Y., Zhang, S.Z., Cai, S.R., Huang, Y.Q., Jiang, Q., et al. 2006. Clinical and genetic characteristics of Chinese hereditary nonpolyposis colorectal cancer families. *World J Gastroenterol*. 12 (25): 4074-4077

Ward, R., Meldrum, C., Williams, R., Mokany, E., Scott, R., Turner, J., et al. 2002. Impact of microsatellite testing and mismatch repair protein expression on the clinical interpretation of genetic testing in hereditary non-polyposis colorectal cancer. *J Cancer Res Clin Oncol*. 128: 403-411

Watson, P., Ashwathnarayan, R., Lynch, H.T., Roy, H.K. 2004. Tobacco use and increased colorectal cancer risk in patients with hereditary nonpolyposis colorectal cancer (Lynch Syndrome). *Arch Intern Med*. 164 (22): 2429-2431

Watson, P., Vasen, H.F.A., Mecklin, J.P., Bernstein, I., Aarnio, M., Järvinen, H.J., et al. 2008. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *Int J Cancer*. 123 (2): 444-449

Weber, T.K., Conlon, W., Petrelli, N.J., Rodriguez-Bigas, M., Keitz, B., Pazik, J., et al. 1997. Genomic DNA-based hMSH2 and hMLH1 mutation screening in 32 eastern United States hereditary nonpolyposis colorectal cancer pedigrees. *Cancer Res*. 57: 3798-3803

Weber, T.K., Chin, H.M., Rodriguez-Bigas, M., Keitz, B., Gilligan, R., O'Malley, L., et al. 1999. Novel *hMLH1* and *hMSH2* germline mutations in African Americans with colorectal cancer. *JAMA*. 281 (24): 2316-2320

Wei, W., Liu, F., Liu, L., Li, Z., Zhang, X., Jiang, F., et al. 2011. Distinct mutations in *MLH1* and *MSH2* genes in hereditary non-polyposis colorectal cancer (HNPCC) families from China. *BMB Reports*. 44(5): 317-322

Wentink, M.Q., Räkers, M., Stupart, D.A., Algar, U., Ramesar, R., Goldberg, P.A. 2010. Incidence and histological features of colorectal cancer in the Northern Cape province, South Africa. *South African Journal of Surgery*. 48(4): 109-113

Wijnen, J.T., Fodde, R., Khan, P.M. 1994. DGGE polymorphism in intron 10 of MSH2, the HNPCC gene. *Hum Mol Genet*. 3(12):2268

Wijnen, J.T., van der Klift, H., Vasen, H., Khan, P.M., Menko, F., Tops, C., et al. 1998. *MSH2* genomic deletions are a frequent cause of HNPCC. *Nature Genetics*. 20(4): 326-328

Wittner, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G., Pryor, R.J. 2003. High-resolution genotyping by amplicons melting analysis using LCGreen. *Clinical Chemistry*. 49(6): 853-860

Woods, M.O., Williams, P., Careen, A., Edwards, L., Bartlett, S., McLaughlin, J., et al. 2007. A new variant database for mismatch repair genes associated with Lynch syndrome. *Hum Mut*. 28: 669-673

Woods, M.O., Younghusband, H.B., Parfrey, P.S., Gallinger, S., McLaughlin, J., Dicks, E. 2010. The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. *Gut*. 59: 1369-1377

Wu, Y., Nystrom-Lahti, M., Osinga, J., Looman, M.W.G., Peltomaki, P., Aaltonen, L.A., et al. 1997. MSH2 and MLH1 mutations in sporadic replication error-positive colorectal carcinoma as assessed by two-dimensional DNA electrophoresis. *Genes, Chromosomes and Cancer*. 18(4): 269-278

Zavodna, K., Bujalkova, M., Krivulcik, T., Alemayehu, A., Skorvaga, M., Marra, G., et al. 2006. Novel and recurrent germline alterations in the MLH1 and MSH2 genes identified in hereditary nonpolyposis colorectal cancer patients in Slovakia. *Neoplasma*. 53 (4): 269-276

Electronic References

Applied Biosystems. 2000. Ampli*taq* Gold with GeneAmp 10x PCR buffer. [Online]. Invitrogen. Available at http://tools.invitrogen.com/content/sfs/manuals/cms_041554.pdf

Accessed 07/04/2012

BLAST-Like Alignment Tool (BLAT). 2012. UCSC Genome Browser. Build hg19. Available from http:// genome.ucsc.edu/cgi-bin/hgBlat

Last accessed 20/06/2012

Database of Single Nucleotide Polymorphisms (dbSNP). 2012. National Center for Biotechnology Information, National library of medicine. dbSNP build 137. Available from http://www.ncbi.nlm.nih.gov/SNP

Last accessed 08/04/2012

GLOBOCAN v1.2. 2010. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. Lyon, France: International Agency for Research on Cancer; 2010. Available from: http://globocan.iarc.fr.

Last accessed on 18/08/2011

International Society for Gastrointestinal Hereditary Tumours (InSiGHT). 2011. Mutations. Available from: http://www.insight-group.org/mutations

Last accessed 29/08/2011

Leiden Open Variation Database (LOVD). 2012. Leiden University Medical Centre. LOVD v2.0. Available from http://www.LOVD.nl

Last Accessed 8/12/2011

NCR. 2011. Incidence of Histologically Diagnosed Cancer in South Africa, 2004 [pdf]. Available from: http://www.nioh.ac.za/assets/files/cancer%20%202004.pdf

Last accessed 4/6/2012

National Centre for Biotechnology Information (NCBI). 2011. NCBI. Available from: http://www.ncbi.nlm.nih.gov

Last accessed 08/04/2012

National Genetics Reference Laboratory (NGRL). 2009. National Genetics Reference Laboratory. Available from: http://www.ngrl.org.uk

Last accessed 24/11/2011

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., et al. 2001. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res*. 29(1): 308-311

Yuan, H.Y., Chiou, J.J., Tseng, W.H., Liu, C.H., Liu, C.K., Lin, Y.J., et al. 2006. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Research*. 34(suppl 2): W635-W641. Available from: fastsnp.ibms.sinica.edu.tw

Last accessed on 23/05/2012