Mutational analysis of the DNA mismatch repair genes, hMLH1 and hMSH2, in South African colorectal cancer patients

BY

Cecilia Maria Dorfling

Dissertation in fulfillment of the requirements for the degree of
Magister Scientiae (M.Sc.) in Human Genetics

In the
Department of Human Genetics and Developmental Biology
Faculty of Health Sciences, University of Pretoria
Pretoria

Supervisor: Professor E.J. van Rensburg

July 2001

ABSTRACT

Colorectal cancer (CRC) is one of the most common forms of neoplasia in Western populations but is uncommon in sub-Saharan Africa. In developing countries such as South Africa, differences in lifestyles and environment exist between the various population groups. These differences and the diverse patterns of cancer that exist, provide an ideal opportunity to study the pathogenesis of colorectal cancer. In South Africa, the incidence of CRC in black patients is approximately ten fold lower than that of white South African patients. The majority of black South African CRC patients presents with tumours without macroscopic polyps. Recently five genes involved in DNA mismatch repair (MMR) have been implicated in hereditary nonpolyposis colorectal cancer (HNPCC).

In this retrospective study, paraffin-embedded normal and tumour tissues from 109 black, and 110 Caucasian CRC patients were studied. To screen for the possible involvement of DNA mismatch repair genes, the presence of microsatellite instability (MSI) was investigated. In total 40 patients presented with MSI-H tumours, 27/109 (24,8%) tumours from black patients and 13/110 (11,8%) tumours from Caucasian patients. The proportion of MSI-H tumours from black patients attending Chris Hani Baragwanath Hospital (CHB) (12,2%; 5/41) and that of Caucasian patients is in accordance to published results on sporadic tumours. However the finding that 32,4% of black patients attending Kalafong and Pretoria Academic Hospitals, have tumours with MSI-H is much higher than is commonly reported in Western populations and is significantly higher than that of the Caucasian patients ($p = 0.002$; χ^2 –test). It has been observed that patients who present at CHB live mainly in urban Johannesburg/Soweto, in comparison to those seen at KPH who are mostly from peri-urban and rural areas.

Failure of PCR amplification, owing to the absence of high quality tissue, allowed 32 of the 40 MSI-H tumours to be fully screened for mutations in hMLH1 and hMSH2 using exon-by-exon PCR single strand conformation polymorphism (SSCP) analysis. Sixteen pathogenic mutations were found in 14 tumours, 10/22 (45%) from black patients and 4/10 (40%) from Caucasian patients. Five tumours presented with two mutations each, one is a compound heterozygote and the other four tumours are double heterozygotes. Ten of the sixteen mutations identified, are

novel. Five (5/32; 16%) of the pathogenic mutations are germline in origin, four (4/22; 18%) of which were detected in tumours from black patients. Thus HNPCC was diagnosed in ~0,93% (1/107) of Caucasian and 3,85% (4/104) of black patients via germline mutations. The frequency of recognised DNA repair gene mutations in black patients with HNPCC is four times higher than that in Caucasian patients with HNPCC. This is consistent with the notion that penetrance of HNPCC cancer is independent of environmental factors which is true as the frequency of HNPCC in a low incidence population (black South Africans) is much higher than that of a high incidence population (Caucasian South Africans). A missense mutation in hMSH2 (codon 127) was identified in three black patients. It is listed in the ICG-HNPCC database as a pathogenic mutation (in a Nigerian family). However, further investigation demonstrated that this is a polymorphic change exclusive to black Africans. Somatic mutations were detected in 6 (27%) tumours from black and 3 (30%) tumours from Caucasian patients.

In conclusion, the observed microsatellite instability and mutations in hMLH1 and hMSH2 thus clearly implicate the involvement of DNA mismatch repair genes in the pathogenesis of colorectal cancers of black and Caucasian South African patients. This study represents the first investigation of DNA mismatch repair genes in tumours from both population groups. It is also the first report of black South Africans with HNPCC.

OPSOMMING

Kolorektale kanker is een van die mees algemene tipe kankers in Westerse populasies, maar is minder algemeen in sub-Sahara Afrika. In ontwikkelende lande soos Suid Afrika, is daar verskille in leefstyle en omgewing tussen verskeie populasie groepe. Hierdie verskille asook die verskillende variasies van kanker wat bestaan, bied 'n ideale geleentheid om die patogenese van kolorektale kanker te bestudeer. Die voorkoms van kolorektale kanker in swart Suid Afrikaanse pasiënte is ongeveer tien keer laer as dié van blanke pasiënte. Die meerderheid van swart kolorektale kanker pasiënte het nie poliepe nie. Daar is al reeds vyf verkeerd-gepaarde DNA herstel (" DNA mismatch repair") gene geïdentifiseer wat betrokke is by oorgeërfde nie-poliep kolorektale kanker ("HNPCC").

In hierdie retrospektiewe studie van 109 swart en 110 blanke Suid Afrikaanse kolorektale pasiënte is parafien-ingebed normale en tumor weefsel ondersoek. Om die moontlike betrokkenheid van die verkeerd-gepaarde DNA herstel gene in kolorektale kanker te bepaal, is die voorkoms van mikro-satelliet onstabiliteit ondersoek. Veertig tumore het hoë mikro-satelliet onstabiliteit (MSI-H) getoon. Die 40 is saamgestel uit 27 (24,8%) tumore van swart pasiënte en 13 (11,8%) tumore van blanke pasiënte. Die verhouding van MSI-H tumore van blanke pasiënte sowel as dié van tumore van swart pasiënte wat behandel was by die Chris Hani Baragwanath Hospitaal (12,2%; 5/41) kom ooreen met gepubliseerde resulte. Daarteenoor is die bevinding dat 32,4 % van swart pasiënte wat by Kalafong en Pretoria Akademiese Hospitale behandel is, MSI-H tumore het, heelwat hoër as wat vir Westerse populasies geraporteer is, en dit is ook beduidend hoër as dié van die blanke pasiënte ($p = 0.002$; χ^2 –toets). Daar is opgemerk dat pasiënte wat by die Chris Hani Baragwanath Hospitaal behandel is grootliks in die stedelike Johannesburg / Soweto omgewings woon, terwyl die pasiënte wat behandel is by Kalafong en Pretoria Akademiese Hospitale meer van buite-stedelik en plattelandse omgewings afkomstig is.

Die afwesigheid van hoë kwaliteit weefsel het gelei tot 'n gebrekkige polimerase ketting reaksie (PKR) amplifisering. Daarom is slegs 32 van die 40 MSI-H tumore volledig nagegaan vir mutasies in die "mismatch repair" gene, hMLH1 en hMSH2, met behulp van die ekson-vir-ekson PKR enkel ketting konformasie polimorfisme

("SSCP") analise. Sestien siekte veroorsakende mutasies is gevind in 14 van die 32 MSI-H tumore. Die 14 tumore sluit 10/22 (45%) tumore van swart pasiënte en 4/10 (40%) tumore van blanke pasiënte in. Vyf tumore het elkeen twee mutasies. Een tumor is 'n saamgestelde heterosigoot en die ander vier is dubbel heterosigote. Tien van die sestien mutaties is nog nie vantevore gerapporteer nie. Vyf van die mutasies (5/32; 16%) is gevind in die kiemsellyn van pasiënte. Vier (4/22; 18%) was swart pasiënte. Dus is HNPCC in ~0,93% (1/107) blanke en 3,85% (4/104) swart pasiënte dmv kiemsellyn mutasies gediagnoseer. Die frekwensie van HNPCC in swart pasiënte is dus vier maal hoër as in blanke pasiënte. Dit ondersteun die gedagte dat die penetrasie van HNPCC kanker onafhanklik van omgewingsfaktore is. 'n Missense mutasie in hMSH2 (kodon 127) is geïdentifiseer in drie swart pasiënte. Die mutasie is gerapporteer op die ICG-HNPCC mutasie databasis as 'n siekteveroorsakende mutasie. Dit was gevind in 'n kolorektale familie van Nigirië. Verdere ondersoek het egter getoon dat dit 'n polimorfisme is wat slegs in die swart populasie voorkom. Somatiese mutasies is in 6 (27%) tumore van swart en 3 (30%) tumore van blanke pasiënte gevind.

Ten slotte, die mikrosatelliet onstabiliteit en die mutasies wat geïdentifiseer is dui daarop dat die DNA "mismatch repair" gene wel betrokke is by die patogenese van kolorektale kanker in swart en blanke Suid-Afrikaanse pasiënte. Hierdie is die eerste studie wat die betrokkenheid van DNA "mismatch repair" gene ondersoek in beide populasie groepe. Dit is ook die eerste aanmelding van Suid Afrikaanse swart persone met HNPCC.

TABLE OF CONTENTS

CHAPTER 3
MATERIALS AND METHODS

CHAPTER 4
RESULTS AND DISCUSSION

CHAPTER 5
CONCLUSION

LIST OF ABBREVIATIONS AND SYMBOLS

α	Alpha
A	Adenine (in DNA sequence)
ADP	Adenosine diphosphate
APC	Adenomatous Polyposis Coli
Arg	Arginine
ASIR	Age Standardized Incidence Rate
Asn	Asparagine
Asp	Aspartic acid
ATG	Methionine
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
β	Beta
bp	Base pair
BSA	Bovine serum albumin
χ^2	Chi-square test
°C	Degrees centigrade
C	Cytosine (in DNA sequence)
CHB	Chris Hani Baragwanath
Ci/mmol	Curies per millimole
cm	Centimeter
CRC	Colorectal cancer
CRC_	Identification of a black patient with colorectal cancer
CRW_	Identification of a Caucasian patient with colorectal cancer
C-terminal	Carboxy-terminal
Cys	Cysteine
D	Aspartic acid
dATP	Deoxyadenosine triphosphate
DCC	Deleted in Colon Cancer
del	Deletion
Df	Degrees of freedom

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
E	Glutamic acid
E coli	*Escherichia coli*
EDTA	Ethylenediamine tetra acetic acid
Exo1	Exonuclease I
F	Represents the forward primer sequence
F	Phenylalanine
FAP	Familial adenomatous polyposis
FS	Frameshift
γ	Gamma
G	Guanine (in DNA sequence)
G	Glycine (in amino acid sequence)
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
h	Human
H	Histidine
H&E	Haematoxylin and eosin
His	Histidine
hMLH	Human MutL homologue
hMSH	Human MutS homologue
HNPCC	Hereditary Nonpolyposis Colorectal Cancer
hPMS	Human PMS homologue
I	Isoleucine
ICG-HNPCC	International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer
IDLs	Insertion deletion loops
Ile	Isoleucine

ins	Insertion
IVS	Intervening sequence
K	Lysine
kb	Kilo base
KCl	Potassium chloride
KPA	Kalafong and Pretoria Academic
K-*ras*	Kirsten *ras* gene
L	Leucine
Leu	Leucine
LOH	Loss of heterozygosity
Lys	Lysine
µCi	MicroCurie
µg	Microgram
µg/ml	Microgram per millilitre
µl	Microlitres
M	Molar
M	Methionine (in amino acid sequence)
MBD4	Methyl-CpG binding thymine glycosylase gene
MDE	Mutation detection enhancement
Met	Methionine
$MgCl_2$	Magnesium chloride
min	Minute
MLH	MutL homologue
mM	Millimolar
mm	Millimetre
MSH	MutS Homologue
MSI	Microsatellite instability
MSI-H	Microsatellite instability -high
MSI-L	Microsatellite instability -low
MSS	Microsatellite stable
MutLα	Heterodimer consisting of hMLH1 and hPMS2
MutLβ	Heterodimer consisting of hMLH1 and hPMS1

MutSα	Heterodimer consisting of hMSH2 and hMSH6
MutSβ	Heterodimer consisting of hMSH2 and hMSH3
N	Asparagine
$NaBH_4$	Sodium borohydride
Na_2CO_3	Sodium carbonate
NaOH	Sodium hydroxide
ND	Not detected
NP 40	Nonidet P40
N-terminal	Amino-terminal
^{32}P	Phosphorus-32
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PF	PCR failed
Phe	Phenylalanine
pmole	Picomole
PMS	Post Meiotic Segregation
PTEN	Phosphatase and tensin homologue deleted on chromosome ten
Q	Glutamine
R	Represents the reverse primer sequence
R	Arginine
RER	Replication error
RPA	Replication protein A
^{35}S	Sulphur-35
S	Serine
S cerevisiae	*Saccharomyces cerevisiae*
Ser	Serine
SSCP	Single strand conformation polymorphism
T	Thymine

T_{ann}	Annealing temperature
ter	Termination
TGF-β RII	Transforming growth factor-β receptor type II gene
Tris	(Hydroxymethyl) aminomethane
Tween 20	Polyoxyethylensorbitanmonolaurat
Tyr	Tyrosine
U	Unit
V	Volt
V	Valine (in amino acid sequence)
Val	Valine
X	Stop codon
Y	Tyrosine

LIST OF FIGURES

LIST OF TABLES

page

ACKNOWLEDGEMENTS

My sincere appreciation to:

Prof EJ van Rensburg, my supervisor, for encouragement, assistance and guidance.

Prof I Segal (GIT Unit, Chris Hani Baragwanath Hospital, Wits), Dr M Hale and Mr J Stevens (Department Histopathology, Chris Hani Baragwanath Hospital) for supplying the paraffin tissue sections of tumours from patients attending the Chris Hani Baragwanath Hospital.

Prof L Dreyer and Dr N Angelo (Department of Anatomical Pathology, University of Pretoria), for their invaluable help with the pathology of the paraffin tissue sections and for supplying the paraffin tissue sections of tumours from patients attending Kalafong and Pretoria Academic Hospitals.

The Cancer Association of South Africa (CANSA) and the University of Pretoria, Faculty of Medicine Research Committee (NAVKOM) who provided funding for this project.

Marlene de la Rey, Tali Yawitch, Sonja Strümpher and Carina Schlebusch for encouragement, friendship and great times in the laboratory.

My family and especially Chris, for their unfailing moral support and constant encouragement.

CHAPTER 1
INTRODUCTION

1.1 INCIDENCE AND BIOLOGICAL ASPECTS OF COLORECTAL CANCER

1.1.1 Incidence

Colorectal cancer is one of the most common forms of neoplasia in Western countries (Dunlop, 1992), whereas it is uncommon in third world countries such as Egypt, (Soliman, *et al.*, 1998), Nigeria (Williams and Prince, 1975) and Southern Iran (Haghighi *et al.*, 1977). In the United States the age standardised incidence rate per 100 000 (ASIR) (1993-1997) is 39,1 and 27,6 for male and female Caucasians and 44,6 and 34,5 for male and female African Americans (Ries *et al.*, 2000). The rate for South African white males (24,7/100 000) and females (19,3/100 000) are lower (Sitas *et al.*, 1998), than rates in the United States. The ASIR for black South Africans however is approximately 10 times lower than that of white South Africans, 2,1/100 000 for black males and 1,6/100 000 for black females (Sitas *et al.*, 1998).

1.1.2 Adenoma-carcinoma sequence

There is evidence that the majority of carcinomas in the large intestine arise from preneoplastic lesions called adenomas or adenomatous polyps. The progression of an adenoma to a carcinoma, known as the adenoma-carcinoma sequence, is a slow process that may take between five to 10 years (Markowitz and Winawer, 1997).

There is a correlation between the risk of cancer and prevalence of adenomas (reviewed in Kronborg and Fenger, 1999). Third world countries such as Iran and several other regions in Africa such as Nigeria, where colorectal cancers are uncommon, have reported the rarity of adenomas (Haghighi *et al.*, 1977; Williams and Prince, 1975). The incidence of neoplastic polyps has also been found to be ten times more frequent in African American colorectal cancer patients whose lifestyles are more westernised, than in Africans from Nigeria (Williams *et al.*, 1975). Several studies on South African colorectal cancer patients have established that colorectal adenomas progressing to colorectal carcinoma are rare in black patients but not in Caucasian patients (Segal *et al.*, 1981). However, the incidence of polyps in the black patients is changing. During the period 1957-1968, only six adenomatous polyps were detected during surgery and in 14 000 autopsies, no polyps were found (Bremner and Ackerman, 1970). But polyps were found in eight of 205 cases during

the period 1969-1978 (Segal *et al.*, 1981) and in seven of 127 cases during the period 1990-1994 (Degiannis *et al.*, 1995). All of these studies were carried out at the Chris Hani Baragwanath Hospital in Johannesburg. A retrospective study performed on patients treated at the Pretoria Academic- and Kalafong Hospitals during the two-year periods 1986 to 1987, and 1996 to 1997, also showed a marked increase in the frequency of polyps in the black population, 9/43 (21%) in the period 1996-1997 compared to 0/15 (0%) during the period 1986-1987. The frequency of polyps in the Caucasian population also increased from 2/67 (3%) to 12/48 (25%) during the same periods (Angelo, 2000).

1.1.3 Risk factors

Age

Colorectal cancer incidence rates increase with age. It is uncommon in individuals younger than forty years and the incidence rates increase sharply after age 50 (Sandler, 1996). The mean age of rural South African patients with adenocarcinoma however is markedly lower (48,6 years) than that of urban European patients (66,4 years) (Jaskiewics *et al.*, 1998).

Diet

Studies have shown a correlation between dietary factors and colorectal cancers (reviewed by Sandler, 1996). Human studies on dietary factors are restricted by the obtainment of accurate data on dietary intake. It appears that fruit and vegetables have a protective effect, whereas high red meat and fat intake show a negative influence. One of the most controversial hypotheses is that fibre has a protective role. If fibre-containing food are benefial, it may not be the fibre itself, but the non-fibre elements in these foods, that are beneficial.

Medical history

Patients with a history of either adenomas or ulcerative colitis have an increased risk for colorectal cancer (Sandler, 1996).

Family history

Family history is an important risk factor for colorectal cancer. Individuals with a first-degree relative with colorectal cancer have an approximate twofold increased risk for developing this cancer (Little and Faivre, 1999).

Lifestyle

Physical inactivity, during work and leisure time, has repeatedly been shown to be a risk for colorectal cancer (Sandler, 1996).

1.1.4 Cancer in developing countries

Studying cancer epidemiology in migrant populations is important. It can be used to study the differences in lifestyle, environment and cancer incidence. Incidence rates of colorectal cancer among Asian residents in the Unites States increased dramatically (Flood, *et al.*, 2000). The initial lower risk of cancers such as colorectal cancer in migrants from the British Isles, southern Europe and east or Southeast Asia to New South Wales, Australia, converged towards the Australian-born level of incidence after the third decade of migration. These changes in incidence of cancer are attributed to differences in environment and lifestyles (McCredie *et al.*, 1999a; McCredie *et al.*, 1999b).

Differences in cancer incidence do not only exist between third world and industrialised countries, but also between rural and urban areas. Cancer registries from some countries have sub-divided incidence data into "rural" and "urban" areas of residence (Boyle *et al.*, 1985). These differences in incidence are becoming less obvious as urbanisation takes place in rural areas. The life expectancy of the world population has increased dramatically in the last century. It is attributed to the vast progress of science and technology that lead to substantial socio-economic changes (Magrath and Litvak, 1993). The incidence of colorectal cancer in for example Ghana has increased and it is predicted that it will become more common as the life expectancy of the population increases above 40-45 years (Naaeder and Archampong, 1994). Developing countries with their differences in lifestyle, environment and diverse patterns of cancer therefore provide ideal opportunities to study the pathogenesis of cancer (Magrath and Litvak, 1993).

1.2 MOLECULAR GENETICS OF COLORECTAL CANCER

1.2.1 Genetic model for colorectal cancer

During a normal cell cycle protein products of three different types of genes regulate cell growth, differentiation and cell death (Figure 1.1). The first two groups of genes, proto-oncogenes and tumour suppressor genes, have diverse functions in growth regulation, differentiation and apoptosis (Kinzler and Vogelstein, 1997). The third class of genes is involved in the repair of DNA damage. These DNA repair genes maintain genomic stability.

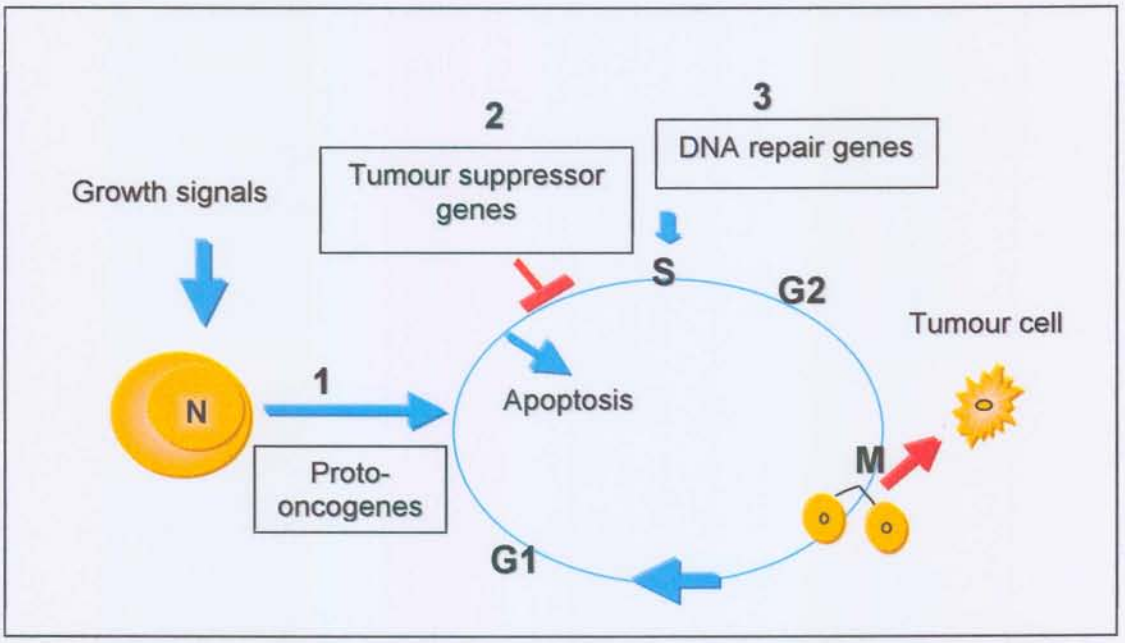


Figure 1.1 Genetic signals in cancer development. During a normal cell cycle protein products of three different types of genes regulate cell growth, differentiation and cell death.

Oncogenes

Proto-oncogenes regulate cell growth and differentiation. Only one copy of a proto-oncogene needs to be mutated to become oncogenic and it is a gain of function mutation (Gryfe *et al.*, 1997). This was the first class of genes found to be associated with cancer and more than 80 have been identified to date (Tumour suppressor and oncogene directory).

Tumour suppressor genes

Tumour suppressor genes inhibit cell growth or promote apoptosis (Kinzler and Vogelstein, 1997). The genetic mechanism of these genes was postulated by Knudson's two-hit model in retinoblastoma (Knudson, 1971). In contrast to oncogenes, both copies of a tumour suppressor gene need to be mutated/loss of

function to become oncogenic. Thus, individuals with a germline mutation need only one somatic mutation in any cell, whereas individuals with no germline mutation need two somatic mutations in one cell. The probability of two somatic mutations happening in one cell is low; therefore, individuals with a germline mutation have a much greater risk for developing cancer (Kinzler and Vogelstein, 1997).

Tumour suppressor genes can be divided into gatekeepers and landscapers, depending on their involvement in tumour initiation. Gatekeeper genes are said to be the traditional tumour suppressor genes that directly suppress the growth of tumours, whereas landscapers are only indirectly involved (Kinzler and Vogelstein, 1997; Kinzler and Vogelstein, 1998).

DNA repair genes

The third class of genes, codes for proteins that are involved in the repair of DNA damage. Different types of DNA damage can occur such as (reviewed in Coleman and Tsongalis, 1995), physical or chemical environmental mutagens that interact with the DNA and spontaneous reactions in the DNA. Physical damage includes ultraviolet light and ionising radiation (x-rays) that result in single- or double strand breaks. Various chemical agents can cause nucleotide base modifications and cross-link the DNA strands. Spontaneous alterations within the DNA can occur during DNA replication, when slippage of the polymerase results in an insertion or deletion of base pairs usually in repeat sequences known as microsatellites. Several types of DNA repair exist, nucleotide excision repair, base excision repair and DNA mismatch repair and a great number of genes involved in these repair pathways have been identified. Mutations in these genes result in deficient DNA repair, which gives rise to genetic disorders such as xeroderma pigmentosum and hereditary nonpolyposis colorectal cancer (HNPCC).

The DNA mismatch repair genes have been implicated in HNPCC. These genes are known as caretakers that maintain genome stability by repairing mismatches that occur during DNA replication. Like tumour suppressor genes, both copies of DNA mismatch repair genes need to be mutated. Inactivation of these genes does not directly initiate tumour formation but results in the accumulation of mutations throughout the genome at repeat sequences. Repeat sequences also occur in the

coding regions of tumour suppressor genes and once both copies of these genes are mutated tumour growth is initiated (Kinzler and Vogelstein, 1997). Therefore at least four separate mutations are required to initiate cancer.

Multistep process of carcinogenesis

Colorectal cancer is one of the best-characterized examples of the multistage nature of cancer as mutations in several oncogenes and tumour suppressor genes are necessary for the development of tumours (Fearon and Vogelstein, 1990). Two different molecular pathways exist for the two major colorectal cancer syndromes, familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC) (reviewed in Chung, 2000). The genes involved in FAP have been identified and a model for tumour progression is illustrated in figure 1.2. Early events include altered DNA methylation (Baylin *et al.*, 1991) and Adenomatous Polyposis Coli (APC) and β-catenin mutations, which is followed by the activation of K-*ras* and, the inactivation of p53 and the Deleted in Colon Cancer (DCC) gene. Fearon and Vogelstein (1990) suggested that the accumulation of mutations is more important than the order in which they occur.

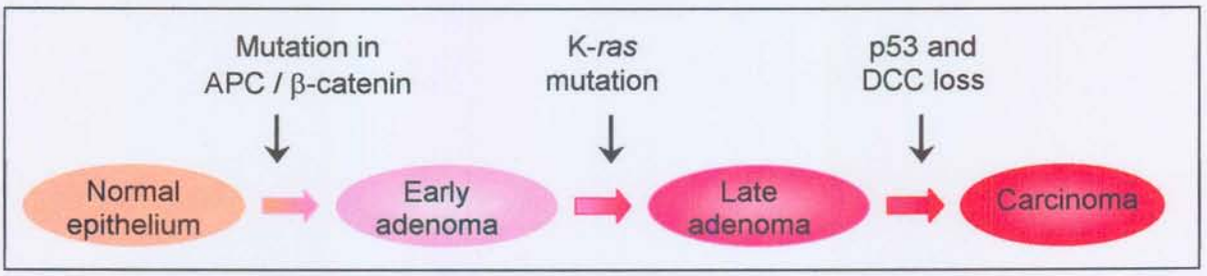


Figure 1.2 The APC pathway in colorectal carcinogenesis

The molecular pathway for HNPCC (Figure 1.3) has only been elucidated recently. It involves the loss of DNA mismatch repair, which results in microsatellite instability and the accumulation of mutations in tumour suppressor genes. These tumour suppressor genes include transforming growth factor-β receptor type II gene (TGF-β RII), BAX, methyl-CpG binding thymine glycosylase gene (MBD4), phosphatase and tensin homologue deleted on chromosome ten (PTEN) and AXIN2 (Markowitz *et al.*, 1995; Rampino *et al.*, 1997; Bader *et al.*, 1999; Guanti *et al.*, 2000; Sarraf *et al.*, 1999; Liu *et al.*, 2000). The AXIN2 gene interacts with APC and is inactivated by DNA mismatch repair deficiency. This provides the link between the APC and DNA mismatch repair pathways (Liu *et al.*, 2000). All of these tumour suppressor genes contain mononucleotide repeats in their coding regions that may become

mutated when the DNA mismatch repair genes are inactivated (Markowitz *et al.*, 1995; Rampino *et al.*, 1997; Bader *et al.*, 1999; Riccio *et al.*, 1999; Guanti *et al.*, 2000; Liu *et al.*, 2000).

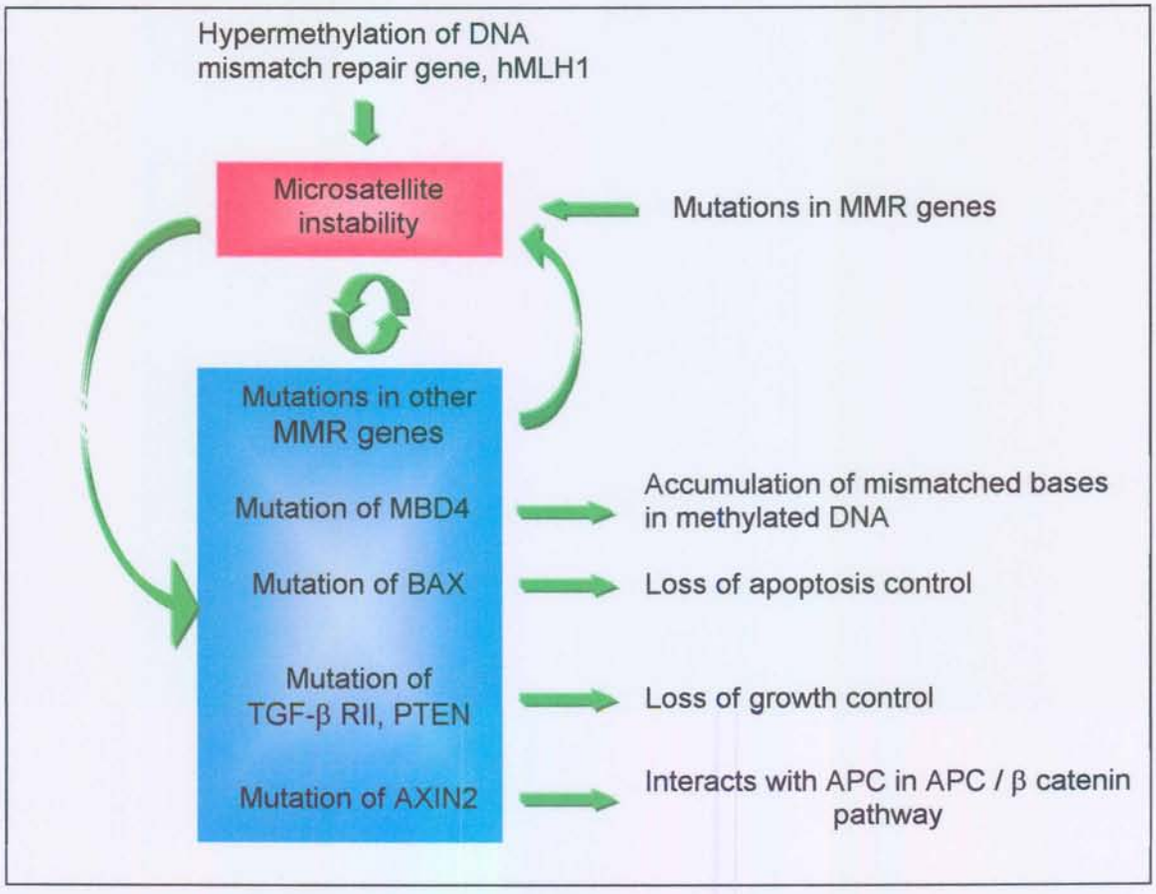


Figure 1.3 Molecular pathway of the DNA mismatch repair genes

1.2.2 Inherited and sporadic forms of colorectal cancer

Colorectal cancer can occur in both sporadic and hereditary forms. Ninety percent of all colon cancers diagnosed each year are sporadic cases, whereas the hereditary polyposis syndromes account for approximately one percent (Rustgi, 1994) and HNPCC has an occurrence of 2-13% (Houlston *et al.*, 1992; Lynch *et al.*, 1993). Sporadic age of onset is higher than familial forms where the age of onset is 40-50 years. Colorectal cancer can be classified into three main groups, those with large amounts of adenomatous polyps, those with very few or no polyps present and those with hamartomatous (non-adenomatous) polyps. Several syndromes with increased susceptibility to colorectal cancer have been identified in each of these three groups (Table 1.1) (Reviewed in Rustgi, 1994).

Polyposis syndromes

The polyposis syndromes are characterised by large numbers of colorectal polyps. Not all of these syndromes are associated with high colon cancer risk and the extra colonic features of the different syndromes are listed in table 1.1. Hamartomatous polyps are usually benign but adenomatous epithelium may develop from within the polyp where tumours can arise (Rustgi, 1994). The colon of patients with FAP may be coated with hundreds to thousands of polyps. One or more of these polyps will progress to adenocarcinoma in these patients. Left sided colon tumours (Figure 1.4) are most common in FAP (Bufill, 1990). Familial adenomatous polyposis is an autosomal dominant disorder and the gene responsible for this syndrome, adenomatous polyposis coli (APC) was identified in 1991 (Kinzler *et al.*, 1991; Groden *et al.*, 1991).

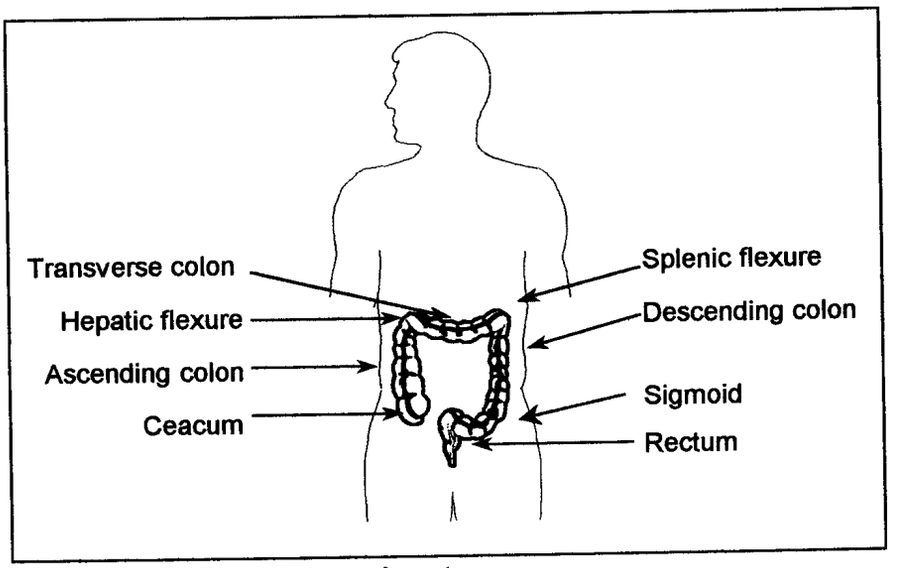


Figure 1.4 The colon and rectum

Nonpolyposis syndromes

In contrast to the polyposis syndromes, where the number of colonic tumours is high, HNPCC is characterized by a low number of tumours. It has an autosomal dominant mode of inheritance of susceptibility and tumours have a tendency to occur in the right colon (Liu *et al.*, 1995a; Lynch *et al.*, 1985). The right colon consists of the ceacum, ascending colon, hepatic flexure and the transverse colon (proximal to the splenic flexure), and the left colon consists of the splenic flexure, descending colon and sigmoid (Figure 1.4). The HNPCC syndrome, which is more common than FAP, can be divided into two groups, those with only colorectal tumours (Lynch syndrome I) and those with extra-colonic features such as endometrial, ovarian and upper gastrointestinal cancer (Lynch syndrome II) (Boland and Troncale, 1984). During

1993 the first gene implicated in HNPCC, hMSH2, was identified. It is a DNA mismatch repair gene (Leach *et al.*, 1993; Fishel *et al.*, 1993).

Table 1.1 The different colorectal cancer hereditary syndromes and the extracolonic features that may be present

DISORDER	EXTRACOLONIC FEATURES
Polyposis colorectal cancer syndromes	
Familial adenomatous polyposis (FAP)	Polyps elsewhere in gastrointestinal tract; retinal lesions; thyroid carcinoma (in females); brain tumours
Gardner's syndrome	As FAP plus; epidermoid cysts; benign bone tumour anywhere in the skeleton; abnormal dentition; desmoid tumours
Turcot's syndrome	Characterized by malignant brain tumours; focal nodular hyperplasia of the liver; café-au-lait patches; birthmarks and skin cancer
Attenuated adenomatous polyposis coli	Gastric polyps at the base of the glands
Hamartomatous polyposis syndromes	
Peutz-Jeghers syndrome	Small bowel hamartomas; pigmentation inside the mouth / cheek; breast, uterine, ovarian and testicular cancer; increased risk of gastrointestinal malignancies
Cowden's syndrome	Facial small raised spots; oral papilomas and horny overgrowth of the skin are prominent and distinctive features; fibromas; benign tumours composed of blood vessels/lymph vessels/ well differentiated fat cells; benign and malignant disease of breast and thyroid; nervous system abnormalities
Nonpolyposis colorectal cancer syndromes	
Hereditary nonpolyposis colorectal cancer (HNPCC)	Endometrial cancer; small bowel, ovarian, urinary tract transitional cell, upper gastrointestinal and pancreatic cancer
Muir-Torre syndrome	Endometrial cancer; small bowel, ovarian, urinary tract transitional cell, upper gastrointestinal and pancreatic cancer, sebaceous skin tumour

1.3 MOTIVATION FOR THIS STUDY

Colorectal cancer (CRC) is one of the most common forms of neoplasia in Western countries (Dunlop, 1992), whereas it is uncommon in third world countries such as Egypt, (Soliman, *et al.*, 1998), Nigeria (Williams and Prince, 1975) and Southern Iran (Haghighi *et al.*, 1977). In developing countries such as South Africa, differences in lifestyle and environment exist between the various population groups. In South Africa the incidence of CRC in black patients is approximately ten fold lower than that of white South African patients (Sitas *et al.*, 1998). There is evidence that the majority of carcinomas in the large intestine arise from pre-neoplastic lesions called adenomas or adenomatous polyps (Markowitz and Winawer, 1997) and there is a correlation between the risk of cancer and prevalence of adenomas (reviewed in Kronborg and Fenger, 1999). Several studies on South African CRC patients have established that colorectal adenomas progressing to colorectal carcinoma are rare in black patients but not in Caucasian patients (Segal *et al.*, 1981). However, the incidence of polyps in the black patients is changing. In studies carried out at Chris Hani Baragwanath Hospital and Pretoria Academic- and Kalafong Hospitals, it was found that the incidence of polyps in the black population is rising (Degiannis *et al.*, 1995; Angelo, 2000). There are essentially two types of CRC: those associated with multiple polyps and those with very few or no polyps. Both these types of cancer may occur in sporadic as well as hereditary forms i.e. familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC). Recently the DNA mismatch repair genes have been implicated in HNPCC. These genes are known as caretakers that maintain genome stability by repairing mismatches that occur during DNA replication.

Molecular genetic studies of cancer in developing countries provide ideal opportunities to study the pathogenesis of cancer because of the differences in lifestyle, environment and diverse patterns of cancer that exist (Magrath and Litvak, 1993). This study is a first step in a programme of investigation into the molecular genetic aetiology of colorectal cancer in black and Caucasian South African patients. Since the black patients appear to present mainly with nonpolyposis type tumours, this study will focus on the involvement of DNA mismatch repair genes in CRC in black and white South African patients.

CHAPTER 2
HEREDITARY NONPOLYPOSIS COLORECTAL CANCER AND DNA MISMATCH REPAIR

2.1 BACKGROUND

Warthin (1913) first described an "inherited cancer" which involved adenocarcinoma of the colon in a large family. This family became known as "Family G", and has been updated over the years and are still being followed. The observations of Warthin lead to the identification of many other families. In 1966, Lynch and co-workers referred to "Family N" as having the Cancer Family Syndrome (Lynch *et al.*, 1966) and in 1977, "Family R" was referred to as having hereditary nonpolyposis site-specific colorectal cancer (Lynch *et al.*, 1977). Several years later the terms Lynch syndrome I, for site-specific colorectal cancer, and Lynch syndrome II, for families that also present with extra colonic cancers, were introduced (Boland and Troncale, 1984). It was also during that time that the term hereditary nonpolyposis colorectal cancer (HNPCC) that included Lynch syndromes I and II, came into use (Lynch *et al.*, 1985). Studies indicated that most of the HNPCC colorectal cancers are mucinous, and poorly differentiated. Signet cell cancers are also common. The tumour spectrum of Lynch type II was uncertain (Lynch *et al.*, 1988). Initially the spectrum included cancer of the colon, endometrium and ovary. Later pancreatic cancer and cancer of the breast, brain, small bowel and even sometimes leukaemia were also included. In 1991, the International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC) set the Amsterdam Criteria to create uniformity in HNPCC studies (Vasen *et al.*, 1991). It specified that colorectal cancer must be diagnosed in at least three relatives, one of whom must be a first degree relative to the other two. They must belong to at least two successive generations and colorectal cancer must be diagnosed in at least one member of the family before the age of 50. Familial adenomatous polyposis must be excluded and all the tumours must be verified by pathological examination.

2.2 IDENTIFICATION OF HEREDITARY NONPOLYPOSIS COLORECTAL CANCER ASSOCIATED GENES

The first breakthrough in understanding the genetic basis of HNPCC came in May 1993 when an HNPCC locus was mapped to the short arm of chromosome 2 by genetic linkage analysis in two large HNPCC kindreds (Peltomaki *et al.*, 1993a). Linkage was obtained through the systematic search through the whole genome using highly informative microsatellite markers. The 346th tested marker, D2S123, showed linkage.

Tumour suppressor genes, that show loss of heterozygosity (LOH), have been associated with hereditary predisposition to cancer. If the germline of an individual is heterozygous for a marker and the tumour of that individual is homozygous, the tumour has lost heterozygosity indicating the presence of a tumour suppressor gene. Unexpected results in a search for loss of heterozygosity in CRC tumours indicated that neither sporadic nor familial CRC tumours showed LOH for 2p markers (Aaltonen *et al.*, 1993). Instead, microsatellite instability (MSI), also known as replication errors (RER), was identified. Microsatellite instability occurs when there is either a gain or a loss of nucleotides in short repeat sequences known as microsatellites. Microsatellite instability was identified in the majority of colorectal tumours of HNPCC patients and in 12-15% of colorectal tumours from sporadic patients (Aaltonen *et al.*, 1993; Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). This abnormality not only occurred at loci on chromosome 2p, but at multiple random loci throughout the genome, suggesting that this could be the result of an error during DNA replication or a deficiency of DNA error repair.

2.2.1 Linkage and cloning of hMSH2

At that time it was known that bacteria and yeast containing defective DNA mismatch repair genes manifest microsatellite instability (Levinson and Gutman, 1987; Strand *et al.*, 1993). In December 1993, a human gene homologous to one of the mismatch repair genes, MutS in bacteria and MSH2 (=Mut S homologue 2) in the yeast, was identified on the basis of sequence homology between species (Fishel *et al.*, 1993; Leach *et al.*, 1993). Germline mutations in hMSH2 were found that segregated with the disorder in kindreds that linked to chromosome 2 and both germline and somatic alterations of the gene occurred in patients, with or without classical HNPCC, whose tumours showed MSI (Leach *et al.*, 1993).

The hMSH2 gene

The hMSH2 gene is located on chromosome 2p22-21 (Fishel *et al.*, 1993) and it covers approximately 73 kilo bases (kb) genomic DNA excluding the promoter region (Kolodner *et al.*, 1994). The promoter region, about 4,4 kb long, contains CpG islands, a few elements involved in constitutive expression and Alu sequences, but no TATA-box. This is the characteristic structure for housekeeping genes. There are also two transcriptional start points and deletion analysis showed that less than 300 base pairs (bp) are crucial to initiate transcription (Iwahashi *et al.*, 1998). The cDNA is 2,8 kb long (Fishel *et al.*, 1993; Leach *et al.*, 1993) and the gene consists of 16 exons ranging from 100 to 279 bp (Kolodner *et al.*, 1994).

The open reading frame encodes a protein of 909 amino acids (Fishel *et al.*, 1993). The entire human protein has a 41% homology with *Saccharomyces cerevisiae* (*S cerevisiae)* MSH2. The most conserved region, the carboxy-terminal (C-terminal) region between amino acids 573-764, is 85% identical to the yeast protein (Fishel *et al.*, 1993; Palombo *et al.*, 1995). Structure based sequence alignment identified other conserved residues and domains that are important in DNA binding and Adenosine triphosphatase (ATPase) activity. The ATPase domain is located in the highly conserved C-terminal region (Obmolova *et al.*, 2000). The protein is expressed in a variety of tissues including thyroid, heart, smooth muscle and oesophageal and intestinal epithelia (Leach *et al.*, 1996). High levels of this protein are expressed in cells of the Lieberkühn that undergo rapid renewal in the colon and ileum (Wilson *et al.*, 1995).

MSH2 deficient mouse models

Mouse models deficient in MSH2 have been developed and used to study DNA mismatch repair and tumourigenesis. The MSH2 deficient mice are viable but differences in the survival rates and the presentation of tumours in the mice exist. Heterozygous mice ($MSH2^{+/-}$) showed no decrease of survival rate (<100 weeks) whereas the $MSH2^{-/-}$ mice died before one year of age (Reitmair *et al.*, 1996; De Wind *et al.*, 1998). The MSH2 deficient mice frequently develop lymphomas at an early age. Other types of cancer that occur include: tumours of the intestinal, lung, uterus and mammary gland and skin (Reitmair *et al.*, 1996; De Wind *et al.*, 1998).

Microsatellite instability was observed in cell lines derived from the MSH2 deficient mice (De Wind *et al.*, 1995; Reitmair *et al.*, 1996). Mismatch binding does not occur in embryonic stem cells derived from these mice and these cells have acquired tolerance to methylating agents. In addition, they do not require homologous DNA during recombination, suggesting that the MSH2 proteins are responsible for safeguarding the genome from unrestricted recombination (De Wind *et al.*, 1995).

2.2.2 Linkage and cloning of hMLH1

A second HNPCC locus, chromosome 3p, was mapped in November 1993 (Lindblom *et al.*, 1993) and soon after in March 1994, the second human DNA mismatch repair gene, hMLH1, was cloned. It is homologous to the MutL mismatch repair gene in bacteria and MLH1 (=Mut L homologue 1) in yeast. Germline mutations in hMLH1 were found that segregated with the disorder in kindreds that linked to chromosome 3 (Bronner *et al.*, 1994; Papadopoulos *et al.*, 1994).

The hMLH1 gene

The hMLH1 gene, located on chromosome 3p21.3-23 (Bronner *et al.*, 1994; Papadopoulos *et al.*, 1994) covers approximately 58-100 kb genomic DNA, excluding the promoter region (Kolodner *et al.*, 1995; Han *et al.*, 1995; Papadopoulos *et al.*, 1994). The coding region is 2,268 kb long (Bronner *et al.*, 1994; Papadopoulos *et al.*, 1994) and consists of 19 exons ranging from 43 to 371 bp (Kolodner *et al.*, 1995; Liu *et al.*, 1995a; Han *et al.*, 1995). The upstream region of the gene, -1 to –1295, was analysed and several transcription factor consensus binding sequences were identified. This region was found to be 56,4% G+C rich (Kane *et al.*, 1997).

The protein consists of 756 amino acids (Bronner *et al.*, 1994). It shares 34-41% identity with the yeast MLH1 protein (Bronner *et al.*, 1994; Papadopoulos *et al.*, 1994) and 55% similarity at amino-terminal region. Additionally the last 13 amino acids are identical between these two species (Bronner *et al.*, 1994). The protein is expressed in a variety of tissues including colon, lymphocytes, breast, lung, spleen, testis, prostate, thyroid, gall bladder and heart, in keeping with its housekeeping function (Papadopoulos *et al.*, 1994).

MLH1 deficient mouse models

Mouse models deficient in MLH1 have been generated (Edelmann *et al.*, 1996). The mice are viable, although $MLH1^{-/-}$ mice die at a younger age than $MLH1^{+/-}$ mice. The $MLH1^{-/-}$ mice were found to have normal reproductive behaviour but both females and males are infertile. Meiotic chromosomal studies showed that in the males, only leptotene to pachytene phases occur during meiotic prophase. The chromosomes at the pachytene stage show normal pairing but no chiasmata can be observed. This pachytene arrest results in abnormal spermatogenesis (no mature sperm) and therefore sterility. The sterile MLH1 deficient females show normal ovulation. During meiotic progression, only one polar body was observed, suggesting that meiosis II was never completed. It is therefore possible that ovarian meiosis is disrupted at the same stage as in testis (Edelmann *et al.*, 1996).

Different tumour spectrums exist between MLH1 deficient mice. $MLH1^{-/-}$ mice developed adenomas and carcinomas in the gastrointestinal tract at an earlier age than $MLH^{+/-}$ mice (Edelmann *et al.*, 1999) and the spectrum is much the same as that of MSH2 deficient mice. The majority of tumours that developed outside the gastrointestinal tract in $MLH1^{+/-}$ and $MLH1^{-/-}$ mice were lymphomas, and skin, lung and cervical squamous cell carcinomas occurred less often.

To understand the role of mismatch repair genes and the APC gene in tumour progression in colon cancer, a mouse model was developed with a mutation in the APC gene, APC1638N. These mice develop polyps in the colon and tumours in the small intestine, like patients with attenuated FAP. Mice were generated that are heterozygous for the APC1638N mutation and either homozygous or heterozygous for MLH1. They were all viable (Edelmann *et al.*, 1999). $MLH1^{-/-}APC1638N^{+/-}$ mice developed almost no tumours outside the GI tract. Adding the APC mutation to the $MLH1^{+/-}$ and $MLH1^{-/-}$ mice increased the incidence of tumours in the gastrointestinal tract with seven and 40-fold respectively. This shows that the APC mutation is an important early event in the onset of gastrointestinal cancer MLH1 deficient mice (Edelmann *et al.*, 1999).

2.2.3 New clinical criteria for hereditary nonpolyposis colorectal cancer

Since the Amsterdam Criteria was set, many investigators have criticised it, as it does not include families that present with extracolonic cancers nor the genetic analysis techniques available, for example MSI testing. More than 90% of colorectal cancers in HNPCC kindreds show microsatellite instability, but because of the low incidence of HNPCC, it is not cost effective to test all CRCs for MSI. The Bethesda Guidelines were therefore set at a National Cancer Institute Workshop on HNPCC, to identify tumours, which should be tested for MSI and will aid in the identification of HNPCC patients (Rodriguez-Bigas *et al.*, 1997). These guidelines included the testing of individuals presenting with cancer in families that meet the Amsterdam criteria, those individuals with two HNPCC-related cancers or individuals with CRC and a first-degree relative with CRC and/or HNPCC-related cancer and/or a colorectal adenoma. It also includes individuals (younger than 45 years) diagnosed with endometrial cancer or poorly differentiated right-sided colorectal cancers, and individuals (younger than 40 years) diagnosed with adenomas or signet-cell-type CRC.

Discussions within the International Collaborative Group led to the conclusion that any criteria should only include clinical analysis as genetic analysis are not accessible to all families and the techniques are not readily available in all countries. The Amsterdam Criteria II was therefore accepted (Vasen *et al.*, 1999). It is essentially the same as the first Amsterdam Criteria set in 1991, except that there must be three or more relatives with an HNPCC-related cancer, which includes colorectal, endometrial small bowel, urethra or renal pelvis cancer. A study comparing the different Amsterdam criteria and the Bethesda Guidelines in 70 families with known DNA mismatch repair gene mutations, found that the Bethesda Guidelines are the most sensitive clinical criteria to identify families with possible mutations in these genes (Syngal *et al.*, 2000). Apart from the clinical criteria, the criteria for microsatellite instability analysis also needed to be uniform. So many different microsatellite markers and number of different markers were used in all the studies that comparison of results was difficult. A National Cancer Institute workshop on microsatellite instability (MSI) for cancer detection and familial predisposition was held in 1998 (Boland *et al.*, 1998). A reference panel and an alternative list of recommended microsatellite markers to be used were set up. Microsatellite unstable tumours can now be divided into three groups. When 30% or more of the

markers analysed are unstable, it is called MSI-high (MSI-H), less than 30% is MSI-low (MSI-L), and no unstable markers is microsatellite stable (MSS). Since microsatellite instability is a result of DNA mismatch repair deficiency it can be used as a tool to identify colorectal cancer patients in which DNA mismatch repair genes might be mutated.

2.2.4 Mutations predisposing to hereditary nonpolyposis colorectal cancer

Intragenic mutations in human DNA mismatch repair genes

Mutations have been identified in hMLH1, hMSH2, hMSH6, hPMS1 and hPMS2, which are associated with HNPCC, atypical HNPCC and sporadic colorectal cancer (ICG-HNPCC Mutation database). The majority of the mutations occur in hMSH2 and hMLH1, presumably because of the redundant functions of hMSH6 and hPMS2. The mutations are scattered throughout the two genes, which means that each gene needs to be screened entirely. The type and amount of mutations occurring in hMSH2 and hMLH1 differ. In hMLH1 152 different mutations have been reported with missense (30%), frameshift (34%) and splice site mutations (24%) mutations being reported more often. Nonsense mutations (6%) and gross rearrangements (5%) occur less often. In hMSH2 a total of 116 different mutations have been reported with frameshift mutations (44%) occurring most often. The frequency of other types of mutations (gross rearrangements, splice site, missense and nonsense mutations) ranges from 9-19%. In both genes small deletions occur more often than small insertions. Cytosine (C) and guanine (G) are more often involved in base substitutions and ~23-29% of these substitutions occur at CpG sites. In general CpG dinucleotides are considered to be hotspots for mutations as spontaneous deamination of methyl-C to thymine (T) may occur (Cooper and Krawczak, 1996).

Several recurrent mutations occur of which the most common is the deletion of exon 16 of the hMLH1 gene. It was found that this is a founder mutation in Finland and it may be older than 40 generations (Moisio *et al.*, 1996). Another common recurrent mutation is the 3' splice site of exon 5 mutation (A->T at 943+3) in hMSH2 that leads to the deletion of the exon. This mutation may affect meiosis in humans as an increased frequency of disomy 13, 21, and XX and diploidy was identified in sperm of carriers of this mutation (Martin *et al.*, 2000). It is considered to be a hot spot region for mutations as 8% (4/52) of HNPCC families from Eastern England, 10% (3/29) of

North American HNPCC families and 50% (10/20) families of New Foundland have been reported to have the mutation (Liu *et al.*, 1994; Froggatt *et al.*, 1995; Froggatt *et al.*, 1999). The difference between a founder mutation and recurrent mutation, is that founder mutations occur on the same haplotype but recurrent mutations not. A common haplotype was found within Newfoundland but not in the other families (Froggatt *et al.*, 1999). Families from the United States of America, Denmark, Italy, Japan, Germany and Norway have been reported to present with this mutation (Möslein *et al.*, 1996; ICHG-HNPCC mutation database; Pensotti *et al.*, 1997; Viel *et al.*, 1997; Miyaki *et al.*, 1995). Another haplotype study investigated families from England, Italy, Hong Kong and Japan (Desai *et al.*, 2000). Several single nucleotide polymorphisms and microsatellite markers within or near MSH2 were used for haplotype analysis and no common haplotype could be identified. The authors speculate that this often *de novo* mutation could be caused by misalignment during replication or recombination caused by the adenine repeat $(A)_{26}$ of which the first is the mutated +3A. It is therefore the most common recurrent de novo mutation as it accounts s for 1% of all pathogenic MSH2 mutations. The risk for colorectal, endometrial and ovarian cancers was calculated for carriers of this mutation (Froggatt *et al.*, 1999). The risk for colorectal cancer was calculated for both sexes and males were found to be at a higher risk than females. For females there is a high risk for endometrial en premenopausal ovarian cancer.

These risks are however not specifically just for the exon 5 splice site mutation. A study looking at kindreds carrying various mutations in hMSH2 also found that males have a higher risk for colorectal cancer and that females have a higher risk for uterine cancer than colorectal cancer (Dunlop *et al.*, 1997). Another study compared the risk factors in families with either hMSH2 or hMLH1 mutations (Vasen *et al.*, 1996). Carriers of hMSH2 or hMLH1 gene mutations have the same lifetime risk for colorectal cancer and also a high risk for small bowel cancer. A significant higher risk for urinary tract, stomach and ovary cancer was identified in carriers of hMSH2 gene mutations. These families also have a higher, although not significant, risk for endometrial cancer than hMLH1 mutation carriers. It was also observed that a higher frequency of extracolonic cancer occurs in hMSH2 mutation carriers than in hMLH1 mutation carriers (Vasen *et al.*, 1996; Lin *et al.*, 1998a). These risk factors have great implications for screening and management programmes for families.

In approximately 30% of HNPCC cases, no mutations have been identified in the DNA mismatch repair genes. This suggests that other mechanisms of inactivation of these genes may be involved for example promoter methylation or other genes that still need to be identified may be involved (Lynch and de la Chapelle, 1999).

Epigenetic gene silencing

Genetic changes (alterations in the sequence of a gene) and epigenetic changes (do not change the structure of the gene, only functionally changes the gene) both play roles in tumourigenesis (Miyaki, 1998). An example of epigenetic change is gene silencing by methylation.

Two types of methylation can occur, hypomethylation and hypermethylation. Hypomethylation of CpG islands results in increased expression whereas hypermethylation results in decreased expression of genes (Miyaki 1998). CpG islands are G+C-rich regions, which are associated with the promoter regions of housekeeping genes (Antequera and Bird, 1993; Larsen *et al.*, 1992). Hypermethylation of the CpG islands in the hMLH1 promoter has been reported. Various transcription factor consensus binding sites are present in the upstream sequence (-1 to -1295) of the ATG (start codon) of the hMLH1 gene. To date there is no evidence to implicate any of the transcription factors in the transcription of this gene (Kane *et al.*, 1997). Methylation occurs frequently in sporadic tumours with microsatellite instability, but less often in MSS tumours and tumours with identified mutations in hMLH1 or hMSH2, and it is associated with loss of expression of hMLH1 (Kane *et al.*, 1997; Cunningham *et al.*, 1998; Herman *et al.*, 1998; Kuismanen *et al.*, 1999). It has also been found that there is a significantly lower prevalence of methylation in HNPCC tumours with identified hMLH1 mutations (Kuismanen *et al.*, 2000). In contrast, it was found that the normal mucosa of MSS tumours showed methylation but not the matching tumours (Kuismanen *et al.*, 1999). According to Kuismanen *et al.*, (1999), hypermethylation showed a correlation with increasing age and proximal location of the tumour in the bowel. Treatment of methylated cell lines with 5-aza-2'-deoxycytidine, an agent that demethylates DNA, restored expression of hMLH1 and DNA mismatch repair. Therefore, methylation plays a direct role in mediating the MSI+ phenotype and is a common mode of inactivation of mismatch repair in sporadic tumours (Herman *et al.*, 1998; Veigl *et al.*, 1998; Kane *et al.*, 1997). No hypermethylation has been

found in the promoter region of hMSH2 (Cunningham *et al.*, 1998; Herman *et al.*, 1998).

2.3 DNA MISMATCH REPAIR

2.3.1 Prokaryotic DNA mismatch repair

DNA mismatch repair was first elucidated in the prokaryote *Escherichia coli*. Four mutator genes have been identified in the Mutator HLS DNA mismatch repair system, MutS, MutL, MutH and MutU (protein product is DNA helicase II). The crystal structures of both MutS and MutL have been described (Ban and Yang, 1998; Lamers *et al.*, 2000; Obmolova *et al.*, 2000). MutS is an ATPase that forms homodimers. It recognises and binds to DNA mismatches and forms complexes with homodimers of MutL and MutH. MutH is an endonuclease that incises the unmethylated strand; where after excision of the mismatch occurs. DNA polymerase fills in the correct nucleotides and the repair is completed by DNA ligase (Modrich, 1991). A translocation model was proposed to explain how MutS utilises ATP during DNA mismatch repair. MutS binds to the mismatch and requires ATP hydrolysis to dissociate from the mismatch so that it can move along the DNA (Allen *et al.*, 1997). The identification and understanding of the different roles of the mutator genes in prokaryotes, facilitated in the identification of homologues of these genes in eukaryotic organisms like *S cerevisiae*.

2.3.2 DNA mismatch repair in yeast

Yeast MutS homologues

Six MutS-related genes, MSH1-6, have been identified in *S cerevisiae*. MSH1 recognizes DNA mismatches in the mitochondria and therefore plays an important role in the repair and maintenance of mitochondrial DNA (Chi *et al.*, 1994). The MSH4 and MSH5 proteins however are not involved in DNA repair, but form a heterodimeric structure, that is required for meiotic crossover and recombination (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995; Pochart *et al.*, 1997). The other three yeast homologues, MSH2, MSH3 and MSH6, are involved in mismatch recognition in nucleic DNA mismatch repair. Two different heterodimeric complexes can be formed between the proteins of these three genes i.e., MutSα consists of MSH2 and MSH6 and, MutSβ consists of MSH2 and MSH3 (Marsischky *et al.*, 1996).

Two pathways of MSH2-dependent mismatch repair exist, one that recognises and repair single base mispairs and requires MutSα, and the other which recognises insertion/deletion loops of 2-14 bp and requires either MutSα or MutSβ (Palombo *et al.*, 1995; Iaccarino *et al.*, 1996; Marsischky *et al.*, 1996; Johnson *et al.*, 1996; Alani *et al.*, 1996; Sia *et al.*, 1997; Umar *et al.*, 1998). There is thus a redundancy of MSH3 and MSH6, where if MSH3 is inactivated, mismatch repair can still be performed, but if MSH6 is inactivated, only insertion/deletion loops could be repaired (Marsischky *et al.*, 1996).

Yeast MutL homologues

Multiple homologues of MutL have been identified in *S cerevisiae*, four of which have roles in DNA mismatch repair. The MLH1 protein forms heterodimeric complexes with MLH2 (Wang *et al.*, 1999), MLH3 (Flores-Rozas *et al.*, 1998; Wang *et al.*, 1999) or PMS1 (Prolla *et al.*, 1994; Pang *et al.*, 1997). None of these four proteins interact with itself or with any of the other MutL homologs other than MLH1. The interaction regions between these proteins lie in their C-terminal regions (Wang *et al.*, 1999). The heterodimer MLH1-PMS1 forms a ternary complex with MSH2-MSH3 (Habraken *et al.*, 1997) or with MSH2-MSH6 (Habraken *et al.*, 1998). MLH2 mutants exhibit an increase in large deletions whereas MLH3 mutants exhibit an increase in two base pair frameshifts (Harfe *et al.*, 2000). It has been suggested that the MLH1-MLH3 complex forms a ternary complex with MSH2-MSH3 that repairs frameshifts (Flores-Rozas *et al.*, 1998). The heterodimeric complexes, MLH1-PMS1, MLH1-MLH2 and MLH1-MLH3, each also have distinct roles during meiosis. The MLH1-MLH3 complex promotes meiotic crossover. Deficient MLH1-PMS1 promotes post meiotic segregation, which is indicative of unrepaired heteroduplex DNA, and deficient MLH1-MLH2 elevates non-Mendelian segregation (Wang *et al.*, 1999).

2.3.3 DNA mismatch repair in humans

Human MutS homologues

Human homologues of the MSH2-6 proteins (Figure 2.1) have been reported to interact with nuclear DNA (Fishel *et al.*, 1993; Leach *et al.*, 1993; Watanabe *et al.*, 1996; Paquis-Flucklinger *et al.*, 1997; Winand *et al.*, 1998; Her *et al.*, 1998; Drummond *et al.*, 1995; Palombo *et al.*, 1995). The hMSH2 protein is expressed in a variety of tissues including thyroid, heart, smooth muscle and oesophageal and intestinal epithelia (Leach *et al.*, 1996). The expression distribution of hMSH3 and

hMSH6 proteins is the same as that of hMSH2 (Watanabe *et al.*, 1996; Wilson *et al.*, 1995; Acharya *et al.*, 1996). As in *S cerevisiae*, hMSH2, 3 (Watanabe *et al.*, 1996) and 6 (Drummond *et al.*, 1995; Palombo *et al.*, 1995; Marsischky *et al.*, 1996) proteins are involved in DNA mismatch repair. The hMSH4 protein is detected only in testis and ovary tissues (Paquis-Flucklinger *et al.*, 1997) and hMSH5 is expressed in various tissues but with higher levels in testis and thymus (Winand *et al.*, 1998; Her *et al.*, 1998). This is expected, since the yeast MSH4 and MSH5 proteins are required for meiotic crossover and recombination.

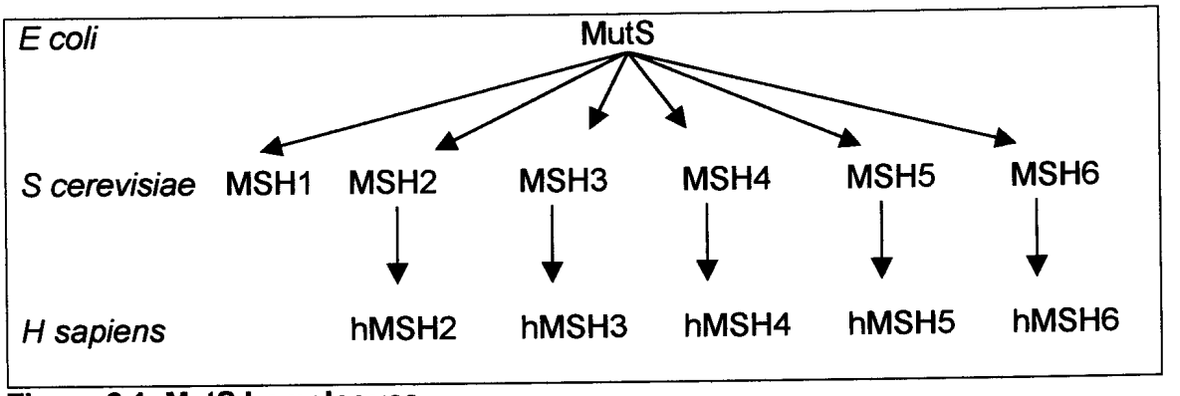


Figure 2.1 MutS homologues.

Human MutL homologues

Multiple human MutL homologues have been identified, three of which have been implicated in HNPCC, hMLH1 (Papadopoulos *et al.*, 1994; Bronner *et al.*, 1994), hPMS1 and hPMS2 (Nicolaides *et al.*, 1994). Recently another human homologue has been cloned, hMLH3, but to date it has not been implicated in HNPCC (Lipkin *et al.*, 2000). The nomenclature of the MutL homologues is illustrated in figure 2.2.

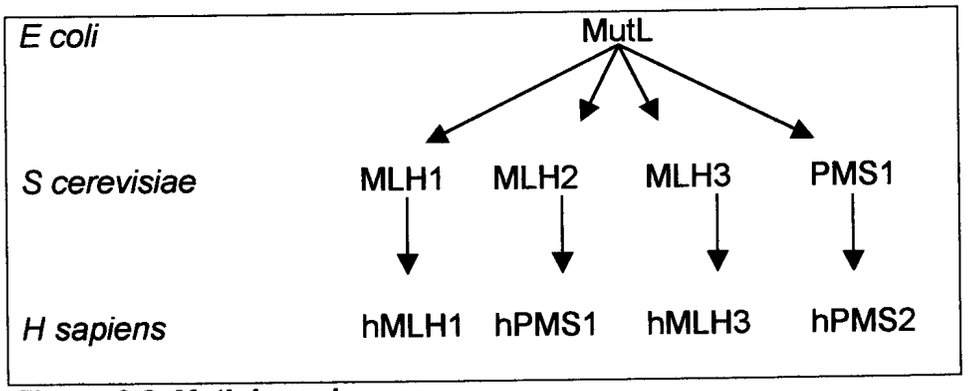


Figure 2.2 MutL homologues

The hMLH1 protein is most closely related to MLH1 (Papadopoulos *et al.*, 1994; Bronner *et al.*, 1994), and hMLH3 is most closely related to MLH3 (Lipkin *et al.*, 2000). But, hPMS1 is most closely related to MLH3 and/or MLH2 (Flores-Rozas and Kolodner, 1998), and hPMS2 is related to PMS1 (Nicolaides *et al.*, 1994). The hMLH1, hPMS2 (Li and Modrich, 1995) and hMLH3 (Lipkin *et al.*, 2000) proteins have been shown to be involved in DNA mismatch repair. The functions of the other homologues are unknown at this stage.

DNA mismatch recognition and repair

Mismatch recognition by MutS homologues

The hMSH2 protein is involved in DNA mismatch recognition. The hMSH2, hMSH3 and hMSH6 proteins form heterodimeric complexes (figure 2.2), MutSα (hMSH2/hMSH6) and MutSβ (hMSH2/hMSH3), which recognize mispairs in DNA (Palombo *et al.*, 1995; Drummond *et al.*, 1995; Acharya *et al.*, 1996). The hMSH2 protein has two regions of interaction, an amino-terminal region (amino acid 378-625) and a carboxy-terminal region (amino acid 875-934) that interact with hMSH3. The same regions are also able to interact with hMSH6 (Guerrette *et al.*, 1998). As in yeast the human MutSα binds to base-base mispairs as well as insertion deletion loops (Drummond *et al.*, 1995; Palombo *et al.*, 1996; Umar *et al.*, 1998) and human MutSβ has high binding affinity only for insertion deletion loops (Acharya *et al.*, 1996). Therefore just like the yeast proteins both hMSH3 and hMSH6 are partially redundant as MutSα functions in base-base mispairs and insertion deletion loops (IDLs), and MutSβ functions preferentially in IDLs (Johnson *et al.*, 1996; Acharya *et al.*, 1996; Marsischky *et al.*, 1996).

The residues important for ATPase activity of MutS homologues are conserved in prokaryotes and eukaryotes (Obmolova *et al.*, 2000). Mismatch binding of human MutSα is sensitive to ATP (Drummond *et al.*, 1995) and mutations in the ATP binding domains of hMSH2 and hMSH6 results in reduced ATPase activity. ATPase-deficient MutSα complexes retain the ability to bind to a mismatch but repair of the mismatch could not happen (Iaccarino *et al.*, 1998). Studies also indicated that human MutSα binds to a mismatch in the presence of adenosine diphosphate (ADP) and dissociates from the mismatch and move along the DNA in the presence of ATP (Gradia *et al.*, 1997; Blackwell *et al.*, 1998).

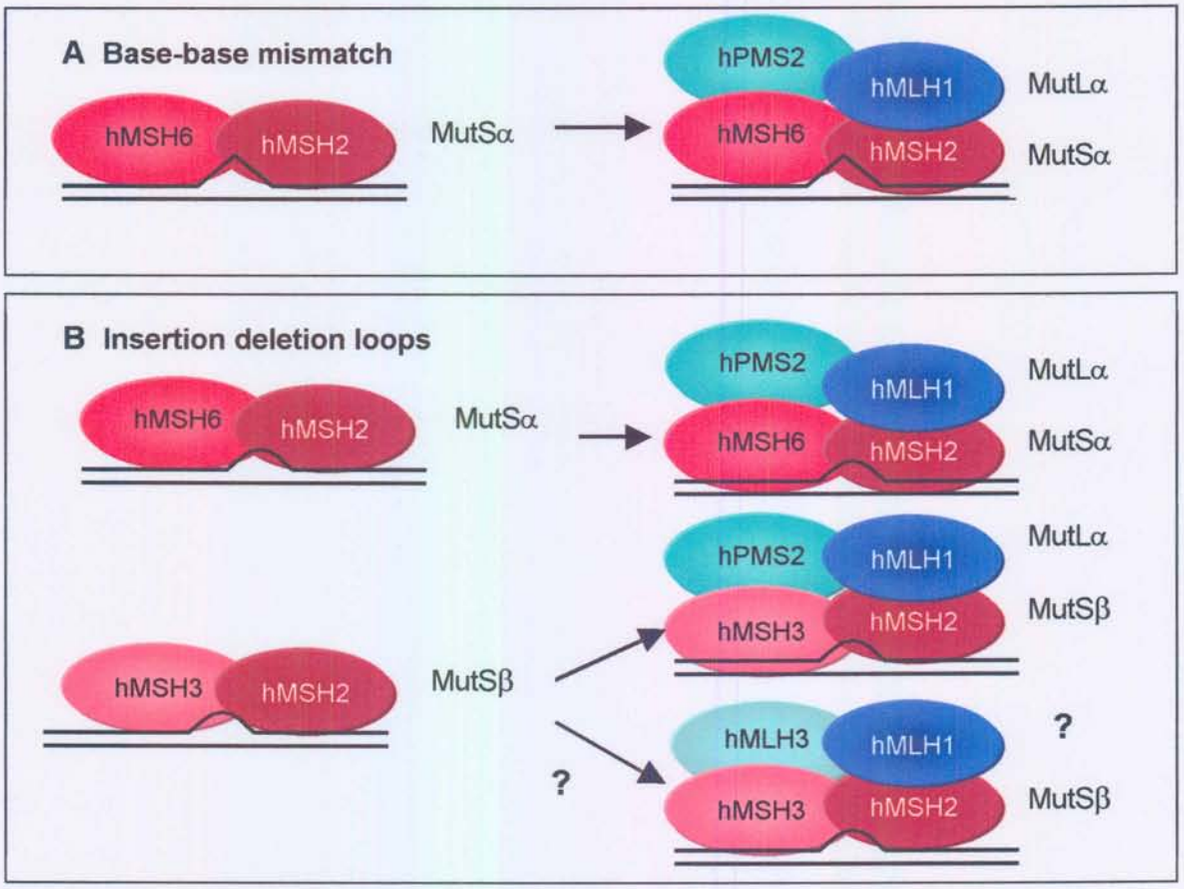


Figure 2.3 Complex formations of the human MutS and MutL proteins Base-base mispairs in A and insertion/ deletion loops in B. (Adapted from Lipkin *et al.*, 2000)

Role of MutL homologues in DNA mismatch repair

The hMLH1 and hPMS2 proteins form a heterodimeric complex, MutLα that restores stability in an hMLH1 deficient cell line that presents with base-base as well as IDL mispairs (Li and Modrich, 1995). The interaction region of these two proteins is located at the carboxy terminal region of hMLH1 between amino acids 506 and 675 (Guerette *et al.*, 1999). MutLα forms a complex with hMSH2 in co-immunoprecipitation experiments. The formation of this complex is mismatch dependant (Gu *et al.*, 1998). Recently it was demonstrated that hMLH1 and hPMS1 also forms a heterodimer, MutLβ, but no evidence could be found for the involvement of this heterodimer in DNA mismatch repair (Räschle *et al.*, 1999). It is speculated that hMLH3 has a less important role than hPMS2 and may participate only in the repair of insertion deletion loops (Figure 2.3) (Lipkin *et al.*, 2000). More research is needed in this part of eukaryotic DNA mismatch repair to explain the importance of the newly identified hMLH3 protein and the recruitment of other proteins in the repair process.

Excision and repair

The way in which the mismatch is corrected after formation of the MutS and MutL complex is still unknown in humans. No human homologue of MutH has been identified and the way in which the newly synthesized strand is recognized in eukaryotes is also still unknown. A possible mechanism for the repair of the mismatch may involve proliferating cell nuclear antigen (PCNA), which forms a mismatch repair initiation complex with hMSH2, hMLH1 and hPMS2. PCNA may also be involved in repair DNA re-synthesis (Gu *et al.*, 1998). Proteins involved in excision and resynthesis have been identified. They are Exonuclease I (Exo1) (Schmutte *et al.*, 1998; Tishkoff *et al.*, 1998; Wilson *et al.*, 1998), Replication Protein A (RPA) (Lin *et al.*, 1998) and DNA polymerase δ (Prelich *et al.*, 1987).

2.4 AIM

This study is a first step in a programme of investigation into the molecular genetic aetiology of colorectal cancer in black and Caucasian South African patients. Since the black patients appear to present mainly with nonpolyposis type tumours, this study will focus on the DNA mismatch repair genes, hMLH1 and hMSH2. This is a retrospective study in which paraffin-embedded tissues from colorectal cancer patients, black and Caucasian, attending Chris Hani Baragwanath and Kalafong and Pretoria Academic Hospitals will be analysed. All the samples will be screened for microsatellite instability. The samples that display microsatellite instability will be screened for mutations in the DNA mismatch repair genes hMLH1 and hMSH2 to determine the extent to which these genes are involved in the colorectal cancers and the nature of the mutations.

CHAPTER 3

MATERIALS AND METHODS

Unless otherwise stated, all chemicals and reagents used in this study were obtained from E Merck Darmstadt, BDH Laboratory Supplies Poole, Roche Molecular Biochemicals, Promega or United States Biochemicals and Genosys Biotechnologies Inc, USA, synthesized all primers.

3.1 PATIENTS

This is a retrospective study, in which paraffin-embedded tissues from South African colorectal cancer patients attending Chris Hani Baragwanath (CHB), Kalafong and Pretoria Academic (KPA) Hospitals were analysed. The Departments of Histopathology (South African Institute Medical Research) and Anatomical Pathology at the University of Pretoria collected these samples for routine diagnostic purposes. All patient identifiers were removed therefore the samples were anonymous.

The patients in this study attended the CHB Hospital during the five-year period 1990-1995 and KPA Hospitals during the ten-year period 1985-1995. Only patients of whom both normal and tumour formalin-fixed paraffin-embedded tissue sections were available, were included in this study. The clinical details of the patients are listed in Appendix A and B.

3.1.1 Black patients

One hundred and nine black patients with colorectal cancer (CRC) were included in this study. Forty-one of these patients attended CHB Hospital and 68 attended KPA Hospitals. The patients, 58 males and 51 females, ranged in age from 11-83 years and have a mean age of 52,11±16,47.

3.1.2 Caucasian patients

One hundred and ten Caucasian patients (CRW) attending CHB (2 patients), Kalafong and Pretoria Academic Hospitals (108 patients) were investigated. The patients comprised 37 males, 73 females with an age range of 23 - 88 years (mean age 65±13,2).

3.2 DNA EXTRACTION FROM PARAFFIN-EMBEDDED TISSUES

Micro-dissection of formalin-fixed paraffin-embedded normal and tumour tissue sections, using a haematoxylin and eosin stained slide as template, was carried out in collaboration with an anatomic pathologist. The micro-dissected tissues were transferred to a microfuge tube to which 200 µl of an extraction buffer (10 mM Tris, pH 8,3; 50 mM KCl; 1,5 mM $MgCl_2$; 0,45 % NP 40 and 0,45 % Tween 20) had been added previously. Thereafter proteinase K (Roche Molecular Biochemicals) was added to a final concentration of 200µg/ml. This was incubated at 56°C for 3 hours, the enzyme inactivated by boiling the mixture for 10 min, quenched on ice and spun down. The DNA suspended in the aqueous phase, was stored at 4°C.

3.3 MICROSATELLITE INSTABILITY ANALYSIS

3.3.1 Microsatellite markers

Eight microsatellite markers were used to evaluate each sample. Five dinucleotide markers located on chromosomes 2, 8, 11, 13, and 17 (Weissenbach *et al.*, 1992) and three mononucleotide markers, BAT25, BAT26, and BAT40 located in intron 16 of the c-*kit* oncogene, intron 5 of hMSH2 and intron 2 of 3-β-hydroxysteroid dehydrogenase gene were used (Parsons *et al.*, 1995). Six of the eight markers (D2S123, D13S175, D17S787, BAT25, BAT26 and BAT40) were recommended by the International Workshop on Microsatellite Instability and RER phenotypes in Cancer Detection and Familial Predisposition (Boland *et al.*, 1998). The polymerase chain reaction (PCR) primers for amplification and PCR product sizes are listed in tables 3.1 and 3.2 respectively.

3.3.2 PCR procedure

5'- End labelling of primers

One primer of each pair (Tables 3.1 and 3.2) was radio-isotopic labelled with [γ^{32}P] ATP (Seperations PTY, LTD) according to the method of Ausubel *et al.* (1995). Briefly, the reaction mixture with a total volume of 10 µl, contained 20 pmole primer, 10X kinase buffer, 10U T4 polynucleotide kinase (Roche Molecular Biochemicals) and 42 µCi [γ-^{32}P] ATP (7000Ci/mmol). The mixture was incubated at 37°C for 1 hour. Thereafter it was boiled for 5 min, quenched on ice, spun down and stored at -20°C, until required.

Table 3.1 Dinucleotide markers

DNA marker	Primer sequence 5'- 3' [a]		Alleles	PCR product size
D2S123	F[b]	AAACAGGATGCCTGCCTTTA	8	197-227 bp
	R	GGACTTTCCACCTATGGGAC		
D8S255	F[b]	TTTTGGAATTTCTAGCCTCC	6	107-129 bp
	R	TGAAACCCACAGATATTGGG		
D11S904	F[b]	ATGACAAGCAATCCTTGAGC	7	185-201 bp
	R	CTGTGTTATATCCCTAAAGTGGTGA		
D13S175	F	TATTGGATACTTGAATCTGCTG	7	101-113 bp
	R[b]	TGCATCACCTCACATAGGTTA		
D17S787	F	TGGGCTCAACTATATGAACC	9	138-166 bp
	R[b]	TTGATACCTTTTTGAAGGGG		

(a) F = forward primer sequence; R = reverse primer sequence
(b) [^{32}P] end labelled primer

Table 3.2 Mononucleotide markers

DNA marker	Primer sequence 5'- 3' [a]		PCR product size
BAT 25	F[b]	TCGCCTCCAAGAATGTAAGT	123 bp
	R	TCTGCATTTTAACTATGGCTC	
BAT 26	F[b]	TGACTACTTTTGACTTCAGCC	118 bp
	R	AACCATTCAACATTTTTAACCC	
BAT 40	F[b]	ATTAACTTCCTACACCACAAC	~150 bp
	R	GTAGAGCAAGACCACCTTG	

(a) F = forward primer sequence; R = reverse primer sequence
(b) [^{32}P] end labelled primer.

Multiplex DNA amplification of dinucleotide markers

Paired normal and tumour DNA were analysed for microsatellite instability in dinucleotide repeats by multiplex PCR with two sets of primers, set I contained primers for D2S123, D17S787, D13S175 and set II contained D11S904 and D8S255.

The multiplex PCR mixture contained 2µl DNA; 10 mM Tris (pH 8,3); 50 mM KCl; 1,5 mM $MgCl_2$; 0,25 mM of each dNTP; 2 µg bovine serum albumin (BSA); 4pmole of each of the primers; 0,5 U Taq Polymerase (Life Technologies) and 0,4 pmole end labelled primer in a total volume of 20 µl. An overlay of a drop of mineral oil prevented evaporation. PCR amplification was performed in a PTC-100™ Programmable Thermal Controller (MJ Research Inc.) with a denaturing step of 3 min at 94°C and then 35 cycles of 94°C (1 min), 58°C (1 min) and 72°C (1 min) with a final extension step of 7 min at 72°C.

DNA amplification of mononucleotide markers

The PCR was carried out in a total volume of 20µl with 2µl DNA; 10 mM Tris (pH 8,3); 50 mM KCl; 1,5 mM $MgCl_2$ (2 mM $MgCl_2$ for BAT 26); dNTPs (0,25 mM each); 2µg BSA; 4 pmole of both the forward and reverse primer; 0,5 U Taq polymerase (Life Technologies); 0,4 pmole of the end labelled primer and a drop of mineral oil that prevented evaporation. The reaction mixtures were denatured at 94°C for 3 min and then processed through 35 cycles of 94°C for 1 min, 50°C for 1 min, 72°C for 1 min (with the exception of BAT40 primers which were annealed at 52°C) and a final extension step of 7 min at 72°C.

3.3.3 Denaturing polyacrylamide gel electrophoresis

Ten microlitres of formamide loading buffer (95 % formamide; 12,5 mM EDTA, pH 8; 0,25 % bromophenol blue and 0,25 % xylene cyanole) was added to the PCR products, denatured for 5 minutes at 95°C and quenched on ice. Three microlitres were loaded onto a 6% polyacrylamide, 7 M urea gel.

The DNA fragments were size fractioned at 60 Watts for three hours (dinucleotide markers) or 90 minutes (mononucleotide markers). After electrophoresis the gels were transferred to 3MM Whatman paper, dried at 80°C under vacuum (Dry gel Sr, Slab Gel Dryer model SE 1160, Hoefer Scientific Instruments, San Francisco) for two

hours and exposed to medical X-ray film (Fuji RX-U) at -70°C using an intensifying screen.

3.3.4 Definition of microsatellite instability

The tumours in this study were scored as MSI-high (MSI-H) when, compared to normal tissue, three or more markers showed clear allele aberrations in the tumour, MSI-low (MSI-L) when one or two markers exhibit MSI and MS stable (MSS) when no instability occurred. This is in accordance with the recommendations of the International Workshop on Microsatellite Instability and RER phenotypes in Cancer Detection and Familial Predisposition (Boland *et al.*, 1998).

3.4 SINGLE STRAND CONFORMATION POLYMORPHISM ANALYSIS

Tumours that showed microsatellite instability (MSI-H) were screened for mutations in the hMLH1 and hMSH2 genes, using exon-by-exon PCR single strand conformation polymorphism (SSCP) analysis as described by Orita *et al.* (1989) with a few modifications. Both radioactive and silver staining detection methods were used.

3.4.1. PCR amplification of the hMLH1 exons

The 19 exons of hMLH1 and the exon-intron boundaries were amplified in 21 fragments. The primers were those designed by Kolodner *et al.* (1995) and Farrington *et al.* (1998). Due to a limited amount of DNA, two rounds of PCR were carried out. The first being a multiplex reaction with sets of primers (Table 3.3). However several exons (7, 9, 10, 12C and 17) were not multiplexed and exons 12B and 19 were amplified separately for silver staining. The primer sequences, PCR product sizes, T_{ann} and $MgCl_2$ concentrations for PCR amplification are all listed in table 3.4.

Table 3.3 hMLH1 primer sets for SSCP

	hMLH1
Set 1	Exons 13,14,16, 18
Set 2	Exons 1, 4, 5, 6
Set 3	Exons 11, 15
Set 4	Exons 2, 3, 8

Table 3.4 hMLH1 primers

Exon[a]	Primer sequence 5'-3' [b]	T_{ann}	[$MgCl_2$]	PCR product sizes
1	F AGGCACTGAGGTGATTGGC R TCGTAGCCCTTAAGTGAGC	55°C	2,0mM	229bp
2	F AATATGTACATTAGAGTAGTTG R CAGAGAAAGGTCCTGACTC	52°C	2,0mM	214bp
3	F AGAGATTTGGAAAATGAGTAAC R ACAATGTCATCACAGGAGG	52°C	2,0mM	207bp
4	F AACCTTTCCCTTTGGTGAGG R GATTACTCTGAGACCTAGGC	55°C	2,0mM	226bp
5	F GATTTTCTCTTTTCCCCTTGGG R CAAACAAAGCTTCAACAATTTAC	55°C	2,0mM	191bp
6	F GGGTTTTATTTTCAAGTACTTCTATG R CTCAGCAACTGTTCAATGTATGAGC	55°C	2,0mM	236bp
7	F CTAGTGTGTGTTTTTGGC R CATAACCTTATCTCCACC	54°C	2,0mM	184bp
8	F CTCAGCCATGAGACAATAAATCC R GGTTCCCAAATAATGTGATGG	52°C	2,0mM	217bp
9	F CAAAAGCTTCAGAATCTC R CTGTGGGTGTTTCCTGTGAGTGG	55°C	2,5mM	193bp
10	F CATGACTTTGTGTGAATGTACACC R GAGGAGAGCCTGATAGAACATCTG	55°C	1,0mM	293bp
11	F GGGCTTTTTCTCCCCCTCCC R AAAATCTGGGCTCTCACG	57°C	1,5mM	290bp
12	F TTAATACAGACTTTGCTACCAG R TTATTACAGAATAAAGGAGGTAG	50°C	1,5mM	428bp
12C	F CCATTTGGGGACCTGTAT R TATCCTCTGTGACAATGGC	57°C	3,0mM	306bp
12B	F TCTGAGCAAACCCCTGTC R CAGAATAAAGGAGGTAGGCTGTA	55°C	2,0mM	256bp
13	F TGCAACCCACAAAATTTGGC R CTTTCTCCATTTCCAAAACC	55°C	1,5mM	291bp
14	F TGGTGTCTCTAGTTCTGG R CATTGTTGTAGTAGCTCTGC	55°C	1,5mM	254bp
15	F CCCATTTGTCCCAACTGG R CGGTCAGTTGAAATGTCAG	57°C	1,5mM	181bp
16	F CATTTGGATGCTCCGTTAAAGC R CACCCGGCTGGAAATTTTATTTG	55°C	1,5mM	276bp
17	F GGAAAGGCACTGGAGAAATGGG R CCCTCCAGCACACATGCATGTACCG	55°C	1,0mM	228bp
18	F TAAGTAGTCTGTGATCTCCG R ATGTATGAGGTCCTGTCC	55°C	1,5mM	247bp
19	F GACACCAGTGTATGTTGG R GAGAAAGAAGAACACATCCC	57°C	1,5mM	269bp

(a)All primers were designed by Kolodner *et al.* (1995) except for exons 12B and C which were designed by Farrington *et al.* (1998)
(b)F = forward primer sequence; R = reverse primer sequence

Multiplex PCR amplification was carried out in 20µl reactions containing 2µl DNA, 10 mM Tris (pH 8,3); 50 mM KCl; [$MgCl_2$] (see table 3.4); 0,25 mM of each dNTP; 2 µg BSA; 4 pmole of each primer in the set and 0,5 U Taq DNA polymerase. The reaction mixtures were denatured at 94°C for 3 min and then processed through 35 cycles of 94°C for 1 min, T_{ann} for 1 min, 72°C for 1 min and a final extension step of 7 min at 72°C.

During the second round PCR only one primer pair were added at a time and with the exception of exons 12B and 19, both the forward and reverse primers were end labelled as described before. In the case of exons 12B and 19, these products were detected via silver staining. The second round PCR contained 2 µl of the first PCR product 10 mM Tris (pH 8,3); 50 mM KCl; [$MgCl_2$] (table 3.4); 0,25 mM of each dNTP; 2 µg BSA; 4 pmole of each primer; 0,5 U Taq DNA polymerase and 0,4 pmole of the forward and reverse end-labelled primers. PCR amplification was performed as before.

3.4.2 PCR amplification of the hMSH2 exons

The 16 exons of hMSH2 and the exon-intron boundaries were amplified in 16 fragments. The primers used for PCR amplification were those designed by Kolodner *et al.* (1994). As with hMLH1 two rounds of PCR amplification were carried out. The different sets of primers used in the multiplex first round PCR are listed in table 3.5 and the primer sequences, PCR product sizes and PCR amplification conditions are listed in table 3.6. Exons 5 and 11 were not multiplexed and exon 12 was amplified separately for silver staining. With the exception of exon 12, the second round PCR products were labelled with ^{32}P. The reactions were set up as for hMLH1.

Table 3.5 hMSH2 primer sets for SSCP

	hMSH2
Set I	Exons 1, 2, 3
Set II	Exons 4, 6, 7
Set III	Exons 8, 9, 10
Set IV	Exons 13, 14, 15, 16

Table 3.6 hMSH2 primers

Exon	Primer sequence 5'-3'[(a)]	T_{ann}	[$MgCl_2$]	PCR product sizes
1	F TCGCGCATTTTCTTCAACC R GTCCCTCCCCAGCACGC	55°C	1,5mM	285bp
2	F GAAGTCCAGCTAATACAGTGC R CTTCACATTTTTATTTTTCTACTC	55°C	1,5mM	286bp
3	F GCTTATAAAATTTTAAAGTATGTTC R CCTTTCCTAGGCCTGGAATCTCC	55°C	1,5mM	393bp
4	F TTCATTTTTGCTTTTCTTATTCC R ATATGACAGAAATATCCTTC	55°C	3,0mM	316bp
5	F CCAGTGGTATAGAAATCTTCG R CCAATCAACATTTTTAACCC	55°C	3,0mM	240bp
6	F GTTTTCACTAATGAGCTTGCC R GTGGTATAATCATGTGGG	55°C	3,0mM	251bp
7	F GACTTACGTGCTTAGTTG R GTATATATTGTATGAGTTGAAGG	55°C	3,0mM	327bp
8	F GATTTGTATTCTGTAAAATGAGATCG R GCCTTTGCTTTTTAAAAATAAC	55°C	3,0mM	222bp
9	F GTCTTTACCCATTATTTATAGG R GTATAGACAAAAGAATTATTCC	55°C	3,0mM	217bp
10	F GGTAGTAGGTATTTATGGAATAC R CATGTTAGAGCATTTAGGG	55°C	3,0mM	259bp
11	F CACATTGCTTCTAGTACAC R CCAGGTGACATTCAGAAC	55°C	1,5mM	198bp
12	F ATTCAGTATTCCTGTGTAC R CGTTACCCCCACAAAGC	55°C	1,5mM	327bp
13	F CGCGATTAATCATCAGTG R GGACAGAGACATACATTTCTATC	55°C	3,0mM	353bp
14	F TACCACATTTTATGTGATGG R GGGGTAGTAAGTTTCCC	55°C	3,0mM	352bp
15	F CTCTTCTCATGCTGTCCC R ATAGAGAAGCTAAGTTAAAC	55°C	3,0mM	261bp
16	F TAATTACTCATGGGACATTC R TACCTTCATTCCATTACTGG	55°C	3,0mM	230bp

(a)F = forward primer sequence; R = reverse primer sequence.

3.4.3 Mutation Detection Enhancement gel electrophoresis

With the exception of exon 12 of hMSH2 and exons 12B and 19 of hMLH1, which were analysed using SSCP and silver staining, all other exons were analysed via radioactive labelled SSCP.

Radioactive detection method

The final radio-active PCR product was diluted 1:10 with formamide loading buffer (95 % formamide; 12,5 mM EDTA, pH 8; 0,25 % bromophenol blue and 0,25 % xylene cyanole), denatured at 95°C for 5 min, quenched on ice and loaded (3µl) on a 0,5 X MDE (Mutation Detection Enhancement) gel (FMC Bioproducts).

The MDE gel was run in 0,6 X TBE buffer at 8 Watt in a cold room (4°C) for varying times (14-24 hours) depending on the PCR product sizes. After electrophoresis the gels were transferred to 3MM Whatman paper, dried at 80°C under vacuum for two hours and exposed to medical X-ray film (Fuji RX-U) at -70°C using an intensifying screen.

Silver staining detection

Exon 12 of hMSH2 and exons 12B and 19 of hMLH1 were analysed using silver staining analysis. Twenty microlitres of a 1:10 dilution was loaded on a 0.6 X MDE gel (20cmx20cmx1mm gel) using the Bio-Rad Protean II gel system. Electrophoresis was performed overnight (20 hours) at 100V at 10°C.

The gels were fixed in a solution of 10% absolute ethanol/ 0,5% acetic acid for 10 minutes, stained in 0,1% silver nitrate for 10 minutes and rinsed twice with distilled water. This was followed by a development step (1,5% NaOH / 0,01% NaBH4 / 0,148% formaldehyde) for 30 minutes. Thereafter stained gels were fixed with 0,75% NaCO3 (10 min) and treated with 10% glycerol for 10 minutes. Drying the gels under vacuum for two hours at 80°C completed this process.

3.5 SEQUENCING

Samples that displayed an abnormal SSCP pattern were sequenced using the Sequenase PCR Product Sequencing Kit (United States Biochemicals, Amersham Life Science) according to the manufacturers guidelines with slight modifications.

Samples were amplified in a new 20 µl PCR reaction using stock DNA. Prior to sequencing, five microlitres of this PCR product was treated separately with 10 U Exonuclease I and 2 U Shrimp Alkaline Phosphatase at 37°C for 15 min (both of these enzymes are active in the buffer used for PCR) and thereafter the enzymes were inactivated at 80°C for 15 min. Sequencing was performed with the Sequenase PCR Product Sequencing Kit using either the forward or reverse PCR primers, [^{35}S] dATP (1000Ci/mmol) (AEC Amersham PTY, LTD) and the conditions specified by the manufacturers with some modification. This entailed adding 1µl DMSO to the annealing mixture, resulting in a total volume of 11µl, which was added to the labelling mixture (without dithiothreitol solution).

Three microlitres of the sequencing reactions were loaded on 6% polyacrylamide/7M urea gels, electrophoresed at 60 Watts, dried at 80°C under a vacuum for two hours and exposed to medical X-ray film (Fuji RX-U). DNA sequence analysis was performed by comparison of published genomic sequences of hMLH1 (Kolodner *et al.*, 1995; Genome Database accession number 249617) and hMSH2 (Kolodner *et al.*, 1994; Genome Database accession number 203983) with those from colorectal cancer patients.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 CLINICAL FEATURES OF COLORECTAL CANCER PATIENTS

Clinical data including age at diagnosis, gender of the patient, tumour site, Dukes' tumour stage and Broders' classification grade were obtained for most of the patients and are presented in Appendix A and B. No Broders' classification could be obtained for the CHB patients. For clarification of the Dukes' and Broders' classifications see Appendix C.

4.1.1 Age at diagnosis

The age at diagnosis of only two black patients and one Caucasian patient is unknown. The black patients' mean age (52,11±16,47 years) is significantly ($p<0,0001$; df = 214; t-test) younger than that of the Caucasian patients' (65±13,2 years). Dividing the patients into three age groups, ≤35, 36-55 and >55, the differences between the ages of the two population groups become clear. Only 3 (2,8%) of 109 Caucasian patients (the age of one patient is unknown) are 35 years or younger, 20 (18,3%) are between 36 and 55 years and 86 (78,9%) are older than 55 years. Nineteen (17,8%) of 107 black patients (the ages of two patients are unknown) are 35 years or younger, 40 (37,4%) are between the ages of 36 and 55 years and 48 (44,9%) are older than 55 years.

4.1.2 Gender

The genders of all the patients were obtained. Fifty-eight of the 109 black patients were male (53%) and 51 (47%) female, but 37 of the 110 (34%) Caucasian patients were male and 73 (66%) female. The much higher total of Caucasian females than males is unusual, since the ASIR of males in South Africa is higher (24,7) than that of the females (19,3) (Sitas *et al.*, 1998).

4.1.3 Cancer site

Tumours located proximal to the splenic flexure were classified as right-sided and those distal to the splenic flexure as left-sided. The tumour sample collection of the black patients consisted of 38% (39/104) right-sided tumours, 30% (31/104) left-sided tumours and 32% (34/104) tumours in the rectum (Figure 4.1 A). Twenty-eight of 103 (27%) tumours of Caucasian patients occurred in the right colon, 29 (28%) in the left colon and 46 (45%) in the rectum (Figure 4.1 B). The location of five tumours from black patients and seven tumours from Caucasian patients is unknown. More tumours from black patients (38%) occurred in the right colon than tumours from Caucasian patients (27%).

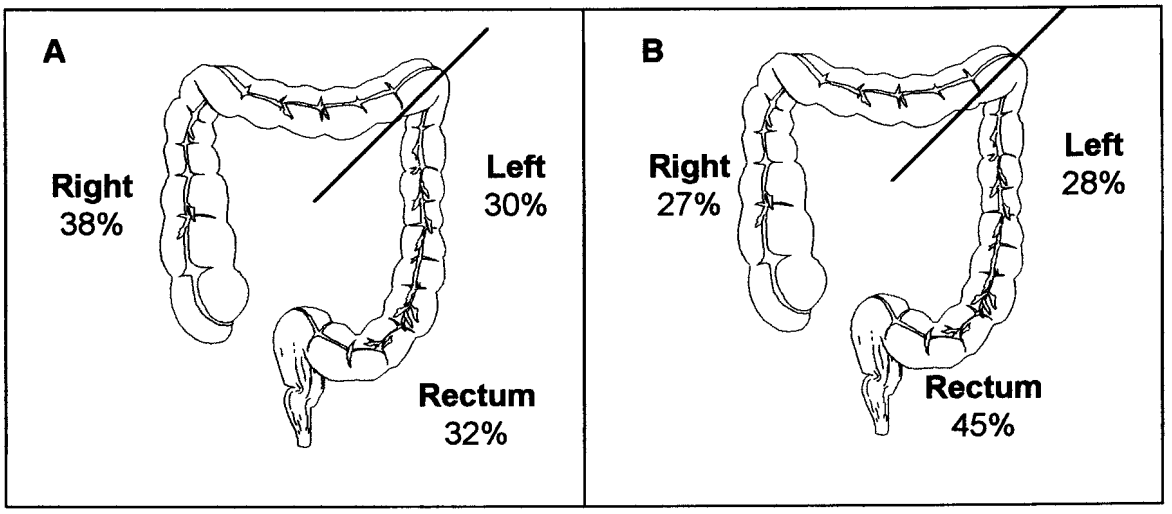


Figure 4.1 Cancer site in black (A) and Caucasian (B) patients

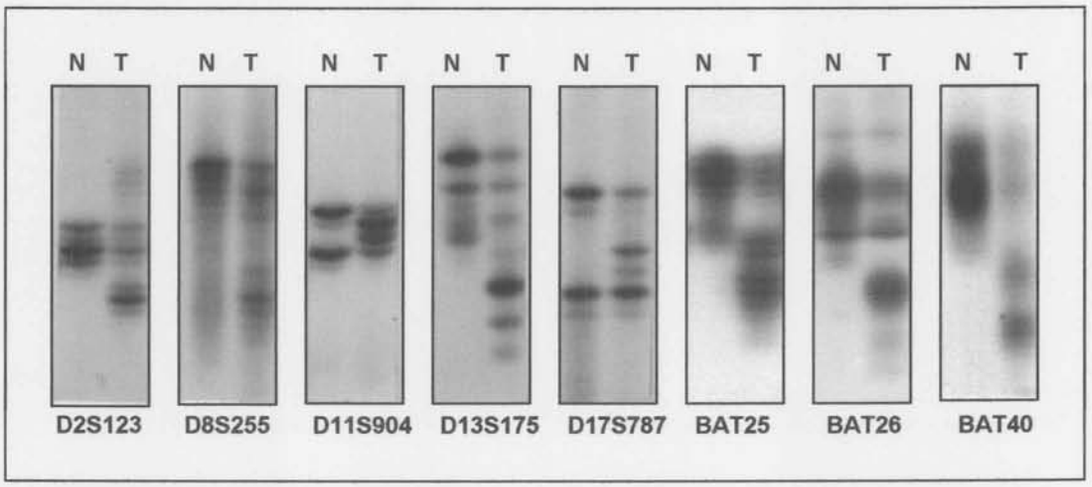
4.1.4 Grade and stage of tumours

The grade of differentiation of the tumours in this study was known for 64 black patients (59%) and 96 Caucasian patients (87%) with the majority of patients presenting with moderately differentiated tumours. The Dukes' classification of 78 (72%) tumours from black patients and 84 (76%) tumours from Caucasian patients was known with the majority presenting with invasion through the mucosa (Dukes' B) and with lymph node metastasis (Dukes' C).

4.2 ANALYSIS OF INSTABILITY

4.2.1 Microsatellite instability

Matched tumour and normal DNA from the patients were investigated for microsatellite instability (MSI) at eight loci. The instability presented as allelic shifts (BAT26; Figure 4.2) and in some cases, the presence of an extra allele was detected (D17S787; Figure 4.2).



N = Normal tissue; T = Tumour tissue

Figure 4.2 Representative autoradiographs showing microsatellite instability in the different markers

Tumours were classified (into three groups) according to the National Cancer Institute Workshop on Microsatellite Instability i.e.; high MSI (MSI-H), when three or more loci exhibited instability; low MSI (MSI-L) when one or two loci exhibited instability and stable MSI (MSS) when no instability was present. Thirty-seven black patients' tumours exhibited MSI, 27 were MSI-H and 10 MSI-L, whereas 34 Caucasian patients' tumours had MSI with 13 MSI-H and 21 MSI-L. Tumours exhibiting MSI-H and MSI-L are listed in tables 4.1 and 4.2.

Table 4.1 Microsatellite instability in tumours from black patients

Sample #	Markers								MSI
	D2S123	D8S255	D11S904	D13S175	D17S787	BAT25	BAT26	BAT40	status
CRC15	✓	✓	✓	-	✓	✓	✓	-	MSI-H
CRC19	✓	✓	-	✓	✓	✓	✓	✓	MSI-H
CRC22	✓	✓	-	✓	✓	✓	✓	✓	MSI-H
CRC33	✓	✓	✓	✓	-	✓	✓	-	MSI-H
CRC35	✓	-	✓	-	-	✓	✓	-	MSI-H
CRC36	✓	-	-	-	-	-	-	✓	MSI-L
CRC48	-	✓	-	✓	-	✓	-	✓	MSI-H
CRC49	✓	-	-	-	-	✓	✓	-	MSI-H
CRC50	✓	✓	-	-	-	-	✓	✓	MSI-H
CRC51	✓	-	✓	-	✓	✓	✓	-	MSI-H
CRC54	✓	-	-	-	-	-	✓	✓	MSI-H
CRC55	-	✓	-	-	-	-	✓	✓	MSI-H
CRC56	-	✓	✓	✓	✓	✓	✓	✓	MSI-H
CRC57	-	-	-	-	-	✓	✓	✓	MSI-H
CRC60	✓	✓	-	✓	-	✓	✓	PF	MSI-H
CRC61	✓	-	-	✓	-	✓	✓	✓	MSI-H
CRC64	✓	-	-	✓	✓	✓	-	PF	MSI-H
CRC72	✓	✓	✓	✓	✓	✓	✓	✓	MSI-H
CRC74	✓	✓	✓	✓	✓	✓	✓	✓	MSI-H
CRC76	-	✓	-	-	-	-	-	-	MSI-L
CRC77	✓	-	-	-	✓	✓	✓	✓	MSI-H
CRC79	-	-	-	-	-	-	-	✓	MSI-L
CRC81	-	-	✓	-	-	✓	✓	-	MSI-H
CRC83	✓	-	-	-	-	-	-	-	MSI-L
CRC85	✓	-	-	-	-	✓	✓	PF	MSI-H
CRC92	✓	-	✓	-	✓	✓	✓	✓	MSI-H
CRC94	-	✓	✓	✓	-	✓	✓	✓	MSI-H
CRC127	-	-	-	-	-	-	-	✓	MSI-L
CRC128	-	-	-	-	✓	-	-	-	MSI-L
CRC129	✓	✓	✓	✓	-	✓	✓	✓	MSI-H
CRC145	✓	-	✓	✓	✓	-	✓	✓	MSI-H
CRC146	-	-	-	✓	-	-	-	-	MSI-L
CRC147	-	-	✓	-	-	-	-	-	MSI-L
CRC148	-	-	✓	✓	✓	✓	✓	✓	MSI-H
CRC149	-	-	-	-	-	✓	-	-	MSI-L
CRC151	-	-	-	-	-	-	✓	-	MSI-L
CRC153	-	✓	✓	✓	✓	✓	✓	✓	MSI-H

✓ = Unstable; - = Stable; PF = PCR failed; CRC = black patient

Table 4.2 Microsatellite instability in tumours from Caucasian patients

Sample #	Markers								MSI status
	D2S123	D8S255	D11S904	D13S175	D17S787	BAT25	BAT26	BAT40	
CRW 1	-	-	-	-	-	✓	✓	✓	MSI-H
CRW16	-	-	✓	✓	-	-	-	-	MSI-L
CRW18	-	✓	✓	-	-	-	-	-	MSI-L
CRW20	✓	-	-	✓	-	-	✓	✓	MSI-H
CRW22	-	-	-	-	-	✓	-	-	MSI-L
CRW24	-	-	✓	-	-	-	-	-	MSI-L
CRW25	-	-	-	-	✓	-	-	-	MSI-L
CRW27	-	-	-	✓	-	-	-	-	MSI-L
CRW28	-	-	-	-	-	-	-	✓	MSI-L
CRW38	-	-	-	-	-	-	-	✓	MSI-L
CRW44	-	-	-	-	-	-	-	✓	MSI-L
CRW47	-	✓	✓	✓	✓	✓	✓	✓	MSI-H
CRW49	✓	✓	✓	✓	✓	✓	✓	✓	MSI-H
CRW56	-	-	-	-	-	✓	✓	✓	MSI-H
CRW57	PF	-	-	✓	-	✓	✓	✓	MSI-H
CRW58	-	-	✓	-	-	✓	✓	✓	MSI-H
CRW60	✓	-	✓	-	-	✓	✓	✓	MSI-H
CRW68	-	✓	-	-	-	-	-	-	MSI-L
CRW70	-	✓	-	-	-	-	-	-	MSI-L
CRW71	-	✓	-	-	-	-	-	-	MSI-L
CRW73	-	-	-	-	✓	-	-	-	MSI-L
CRW78	✓	-	✓	✓	✓	✓	✓	✓	MSI-H
CRW80	-	-	-	-	✓	-	-	-	MSI-L
CRW92	✓	✓	✓	-	-	✓	✓	✓	MSI-H
CRW96	✓	✓	✓	-	✓	✓	✓	✓	MSI-H
CRW99	✓	✓	-	-	-	✓	✓	✓	MSI-H
CRW101	-	-	-	-	-	-	-	✓	MSI-L
CRW105	-	✓	-	-	-	-	-	-	MSI-L
CRW108	✓	-	-	-	-	-	-	-	MSI-L
CRW110	✓	-	-	-	-	-	-	-	MSI-L
CRW123	-	-	-	✓	✓	✓	✓	-	MSI-H
CRW125	-	✓	-	-	-	-	-	-	MSI-L
CRW126	-	-	-	-	✓	-	-	-	MSI-L
CRW128	✓	-	-	-	-	-	-	-	MSI-L

✓ = Unstable; - = Stable; PF = PCR failed; CRW = Caucasian patient

In total 40 patients presented with MSI-H tumours. Of the 109 tumours from black patients 27 (24,8%) were found to have MSI-H, which is significantly different ($p = 0.021$; χ^2 –test) from the 11,8% (13/110) of MSI-H tumours from Caucasian patients (Table 4.3).

Table 4.3 Microsatellite instability status of tumours

	MSI-H	MSI-L	MSS	Total
Black patients	27 (24,8%)	10 (9,2%)	72 (66,1%)	109
	CHB 5 (12,2%) KPA 22 (32,4%)	1 (2,4%) 9 (13,2%)	35 (85,4%) 37 (54,4%)	41 68
Caucasian patients	13 (11,8%)	21 (19,1%)	76 (69,1%)	110

CHB = Chris Hani Baragwanath Hospital; KPA = Kalafong and Pretoria Academic Hospitals

The 11,8% microsatellite instability in the tumours of Caucasian patients is in accordance, with the literature on sporadic cases, which report 10-16% of instability (Kim *et al.*, 1994; Aaltonen *et al.*, 1998; Thibodeau *et al.*, 1998; Curran *et al.*, 2000; Abe *et al.*, 2000; Potočnik *et al.*, 2001). These research groups investigated 103-509 tumours from sporadic patients originating from the United States of America, Finland, Ireland, Japan and Slovenia, using 5-16 microsatellite markers and the published criteria for scoring MSI-H tumours (Boland *et al.*, 1998). Other groups investigating less than 100 patients reported a much broader range of instability, 13-29% (Ishimaro *et al.* 1995; Watatani *et al.*, 1996; Aaltonen 1993; Aaltonen *et al.*, 1994; Bocker *et al.*, 1996; Herfarth *et al.*, 1997).

In contrast to that of the Caucasian patients' tumours, the proportion of MSI-H in the black patients' tumours is double (24,8%). It is also higher than the instability reported in the larger studies. An interesting observation was made when the MSI-H tumours from the black patients were divided into two groups according to the hospital that was attended. An almost three-fold difference ($p = 0.033$; χ^2 –test) in the proportion of MSI-H tumours from CHB (5/41; 12,2 %) compared to KPH (22/68; 32,4 %) was observed. The finding that 12,2% of black patients from CHB displayed MSI-H is in keeping with other reports on sporadic cases. There is no significant difference between the proportion of MSI-H tumours found in the black patients from CHB and that of the Caucasian patients ($p=0,827$; χ^2 –test). However, the finding that 32,4% of black patients attending KPH have MSI-H is much higher than is

commonly reported and is significantly higher than that of the Caucasian patients (p=0,002; χ^2 –test). It has been observed that patients who present at CHB lived mainly in urban Johannesburg/ Soweto, in comparison to those seen at KPH who were mostly from peri-urban and rural areas.

4.2.2 Frequency of unstable loci

The frequency of instability observed in the eight loci differed between mononucleotide and dinucleotide repeats. Only two black patients, CRC72 and CRC74, and one Caucasian patient, CRW49, showed instability in all eight markers. As seen in figure 4.3 the dinucleotide markers, D8S255, D13S175, and D17S787 showed less than 50% instability in the Caucasian MSI-H patients, whereas in the black MSI-H patients only D17S787 showed less than 50% instability. The mononucleotide markers, BAT25 and BAT26, were most frequently associated with microsatellite instability in both Caucasian and black patients' MSI-H tumours. BAT26 is the only marker that shows instability specific to MSI-H tumours, but only in Caucasian patients.

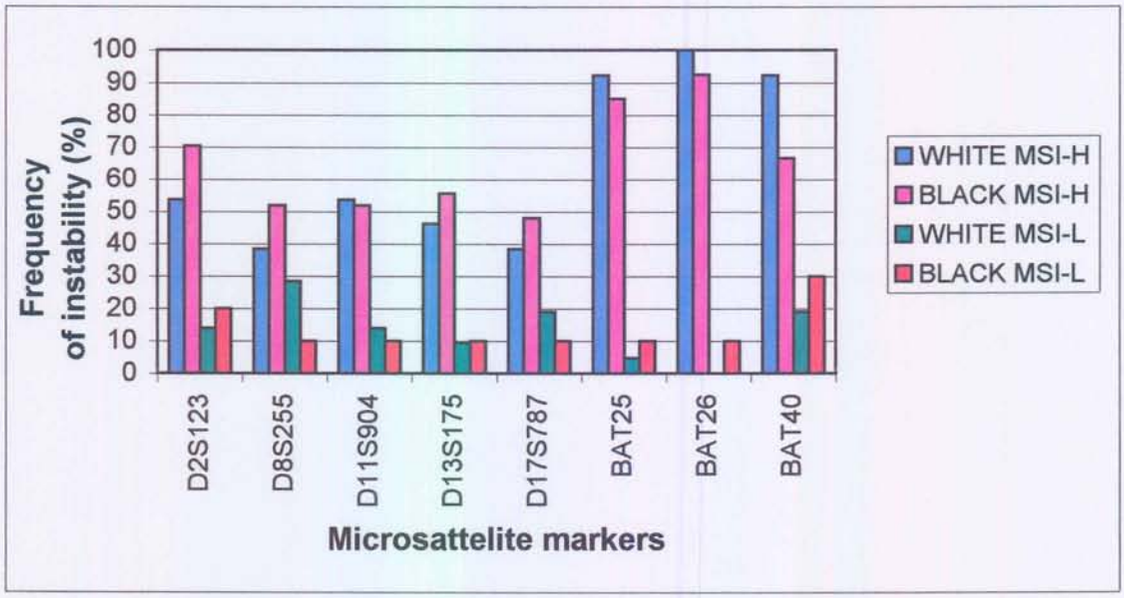
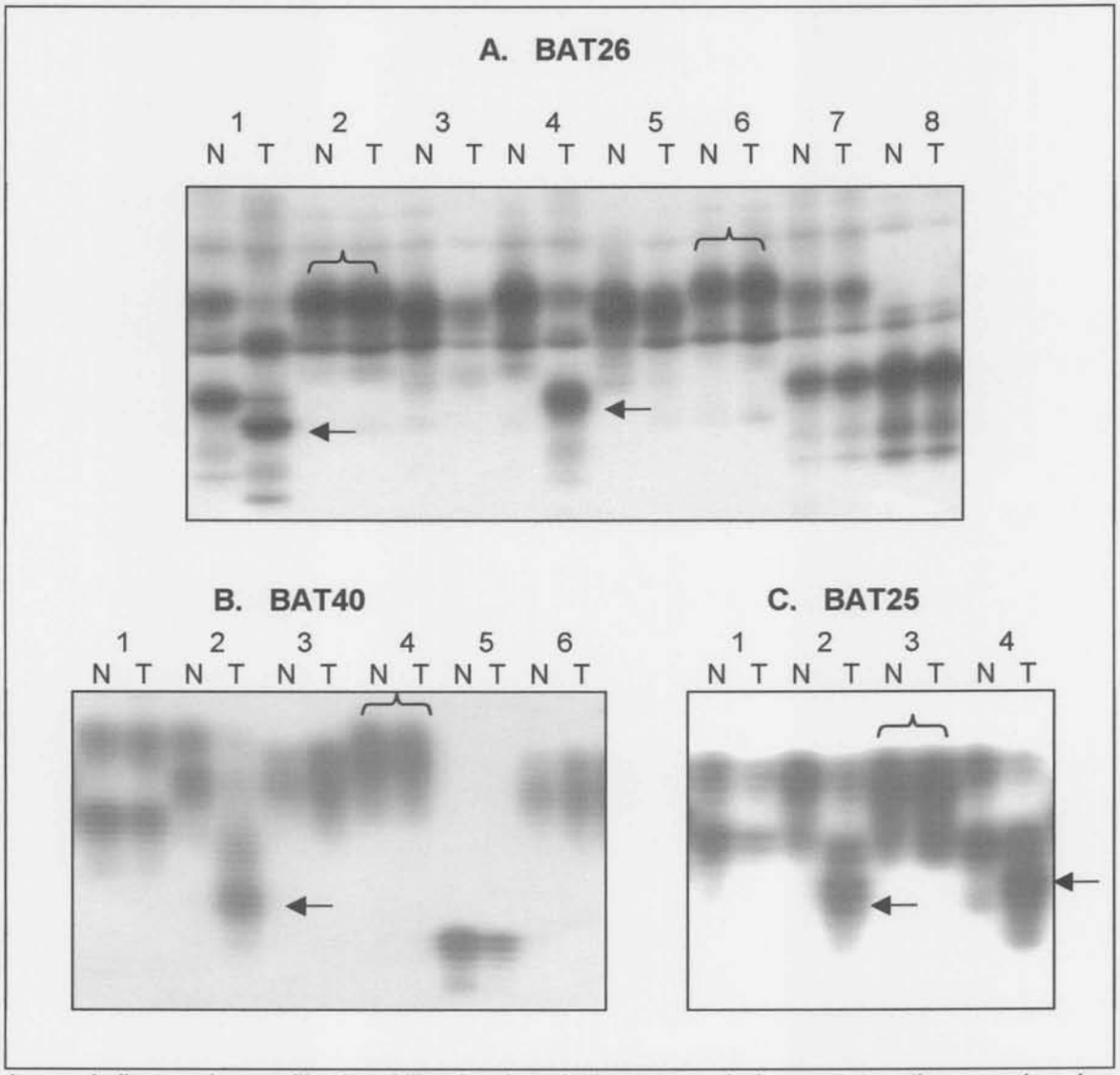


Figure 4.3 Instability displayed in the different microsatellite markers of MSI-H and MSI-L tumours

BAT25 and BAT26 have been described as quasimonomorphic, with a common allele (25 thymine and 26 adenosine repeats respectively) and normal size variation not exceeding two nucleotides in individuals from French (Zhou *et al.*, 1997; Hoang *et al.*, 1997) and Finnish origin (Aaltonen *et al.*, 1998). BAT40 has been described as polymorphic (Zhou *et al.*, 1997; Hoang *et al.*, 1997) with a large number of different

alleles. The Caucasian patients in this study presented with a common allele in both BAT25 and BAT26 with only one or two base pair variations, and were polymorphic at BAT40, as suggested in the literature (Zhou *et al.*, 1997; Hoang *et al.*, 1997; Aaltonen *et al.*, 1998). Some of the black patients however, were heterozygous in the germline at not only BAT40 but also BAT25 and BAT26, suggesting that all three markers are polymorphic in the black patients (figure 4.4). During 1999 two reports showed that African-Americans are heterozygous for BAT25, BAT26, and BAT40 (Pyatt *et al.*, 1999; Samowitz *et al.*, 1999) substantiating the results of this study.



Arrows indicate microsatellite instability; brackets indicate normal size repeats; other samples show polymorphic alleles.

Figure 4.4 Examples of the polymorphic mononucleotides in normal and tumour tissues of black patients

4.2.3 Clinico-pathologic features of patients with microsatellite unstable tumours

Age at diagnosis

When comparing the three groups of instability (MSI-H, MSI-L and MSS) no association with age was found for either the black ($p = 0.4355$; Kruskal-Wallis test) or the Caucasian ($p = 0.9826$; Kruskal-Wallis test) patients. However, in all three groups the black patients' mean ages (Table 4.4) is significantly ($p=0,0106$; $p=0,0142$; $p < 0,0001$; t-test) younger than that of the Caucasian patients (Table 4.5). The average age of the Caucasian patients with MSI-H tumours (63,15±15,61yrs) is similar to previously reported average ages (60-69 years) of patients presenting with MSI-H tumours (Kim *et al.*, 1994; Aaltonen *et al.*, 1998; Curran *et al.*, 2000; 2000; Potočnik *et al.*, 2001). However, the average age of black patients with MSI-H (48±17,19 years) is more than 10 years less than reported and it is closer to that reported for HNPCC patients, i.e. 40-45 years (Aaltonen *et al.*, 1994).

Gender

The MSI-H tumour phenotype is not associated with a specific gender (Kim *et al.*, 1994; Thibodeau *et al.*, 1998; Curran *et al.*, 2000; Potočnik *et al.*, 2001). In the present study more more tumours from black male patients (Table 4.4), 17/27 (63%), and Caucasian female patients (Table 4.5), 8/13 (67%) presented with MSI-H. Due to the small number of samples the statistical power for testing significance is too small. Larger groups of patients would give more power to the test and differences could become significant.

Cancer site

An association between tumour location (proximal to the splenic flexure) and MSI-H was observed; 17/27 tumours from black patients ($p = 0.0047$; χ^2 –test), and 8/12 tumours from Caucasian patients ($p = 0.0048$; χ^2 –test), presenting with MSI-H, were located proximal to the splenic flexure (Table 4.4 and Table 4.5). This prevalence of MSI-H tumours being located predominantly in the right colon has been reported many times (Kim *et al.*, 1994; Aaltonen *et al.*, 1998; Thibodeau *et al.*, 1998; Curran *et al.*, 2000; Potočnik *et al.*, 2001).

Grade and stage of tumours

In all three groups of microsatellite instability status, MSI-H, MSI-L and MSS, grade II differentiation and Dukes' stages C and D are most frequent in tumours from

both black and Caucasian patients (Table 4.4 and Table 4.5). Due to the small number of samples the statistical power for testing significance is too small.

Table 4.4 Clinicopathological features of black patients according to their miscrosatellite instability status[a]

	MSI-H (n=27)	MSI-L (n=10)	MSS (n=72)	Total (n=109)
Gender [b]				
Male	17 (63%)	4 (40%)	37 (51%)	58 (53%)
Female	10 (37%)	6 (60%)	35 (49%)	51 (47%)
Cancer site [c]				
Right	17 (63%)	4 (40%)	18 (27%)	39 (38%)
Left	7 (26%)	4 (40%)	20 (30%)	31 (30%)
Rectum	3 (11%)	2 (20%)	29 (43%)	34 (32%)
Tumour differentiation [d]				
Grade II	17 (85%)	8 (89%)	34 (97%)	59 (92%)
Grade III	3 (15%)	1 (11%)	1 (3%)	5 (8%)
Dukes' staging [e]				
Dukes' A	0	1 (10%)	1 (2%)	2 (3%)
Dukes' B	11 (48%)	4 (40%)	19 (42%)	34 (43%)
Dukes' C	12 (52%)	4 (40%)	22 (49%)	38 (49%)
Dukes' D	0	1 (10%)	3 (7%)	4 (5%)
Age range (years)	11-80	27-68	16-83	11-83
Mean age (years) [f]	48±17,19	50,5±13,39	53,93±16,49	52,11±16,47

[a] Since data were missing in some cases, the numbers shown do not always equal the totals in groups; [b] p=0,4008 (χ^2 –test); [c] p=0,0047 (χ^2 –test); [d,e] Values are too small for valid calculations; [f] p=0,4355 (Kruskal-Wallis test)

Table 4.5 Clinicopathological features of Caucasian patients according to their miscrosatellite instability status[a]

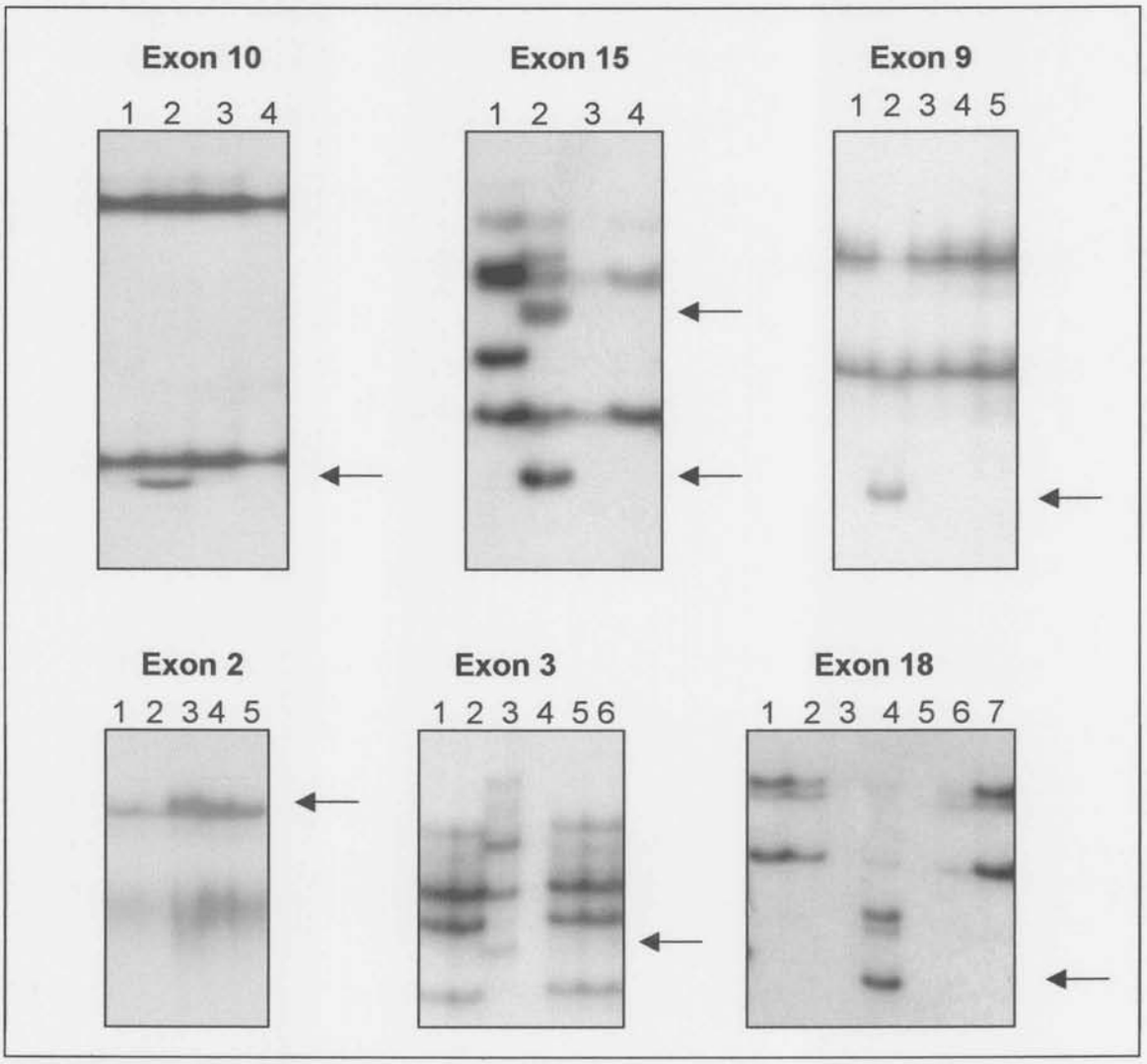
	MSI-H (n=13)	MSI-L (n=21)	MSS (n=76)	Total (n=110)
Gender [b]				
Male	5 (38%)	4 (19%)	28 (37%)	37 (34%)
Female	8 (62%)	17 (81%)	48 (63%)	73 (66%)
Cancer site [c]				
Right	8 (67%)	4 (22%)	16 (22%)	28 (27%)
Left	1 (8%)	5 (28%)	26 (36%)	29 (28%)
Rectum	3 (25%)	9 (50%)	31 (42%)	46 (45%)
Tumour differentiation [d]				
Grade II	7 (70%)	19 (95%)	63 (95%)	89 (93%)
Grade III	3(30%)	1 (5%)	3 (5%)	7 (7%)
Dukes' staging [e]				
Dukes' A	0	0	2 (3%)	2 (2%)
Dukes' B	4 (40%)	6 (40%)	25 (43%)	35 (42%)
Dukes' C	4 (40%)	8 (53%)	24 (41%)	36 (43%)
Dukes' D	2 (20%)	1 (7%)	7 (12%)	11 (13%)
Age range (years)	33-81	34-88	23-83	23-88
Mean age (years) [f]	63,15±15,61	65,24±15,24	65,25±12,3	65±13,2

[a] Since data were missing in some cases, the numbers shown do not always equal the totals in groups; [b] p=0,2883 (χ^2 –test); [c] p=0,0048 (χ^2 –test); [d,e] Values are too small for valid calculations; [f] p=0,9826 (Kruskal-Wallis test)

4.3 hMLH1 GENE ALTERATIONS

4.3.1 SSCP analysis

The 40 MSI-H patients were screened for mutations in the 19 coding regions of the hMLH1 gene using exon-by-exon PCR SSCP analysis. Seven tissue samples (CRC55, CRC56, CRC94, CRC129, CRC148, CRW47 and CRW57) however did not amplify adequately for SSCP analysis in all exons. Thus only 33 MSI-H patients were fully screened for hMLH1 mutations. Representative autoradiographs of some of the exons, in which aberrant bands were detected, are presented in figure 4.5. Those samples showing bands with altered mobility were sequenced.



Arrows indicate shifts

Figure 4.5 Representative examples of SSCP shifts detected in some of the exons of the hMLH1 gene

4.3.2 Pathogenic mutations in hMLH1

Twelve mutations were detected in eleven of the 33 MSI-H tumours that were fully screened for hMLH1 mutations. Ten of these mutations are pathogenic and two missense mutations are only putative pathogenic. These mutations included five frameshifts, three missense and two nonsense mutations, one amino acid deletion and one splice site mutation.

Frameshift mutations

The five frameshift mutations identified were all deletions of which two were germline (Figure 4.6). Both of the germline mutations are four base pair deletions that have not been reported before. One of these, 731del4, was detected in CRW60, a 51-year-old male in whom the tumour has invaded through the mucosa of the colon but no metastasis has occurred. The mutation is in exon 9 and it terminates the protein at codon 253 in the same exon. The second germline mutation was identified in a young male patient, CRC64 (25 years, cancer in the rectum). This mutation, 1685del4 in exon 15 results in a shortened protein that was terminated at codon 589 in exon 16.

Two somatic deletions involving a single nucleotide were identified in exon 10. Patient CRC77 has developed a moderately differentiated tumour with no metastasis (grade II, Dukes' B) in the left colon at the age of 41 years. Sequence analysis identified a one base pair, cytosine (C) deletion at nucleotide 813. This novel mutation caused a frameshift leading to a premature stop signal at codon 272. The second one base pair deletion in exon 10 was identified in the tumour of patient CRW49. She is a 75-year-old woman with cancer in the ceacum that has infiltrated the lymph nodes as well as the appendix. A run of five adenines (A_5) occurs at codons 286-287 in this exon. One of these base pairs was deleted which leads to a premature termination codon in exon 11. This mutation has not been reported before.

The largest deletion identified, removed 11 base pairs at nucleotide 1992 in exon 18, leading to the generation of a stop at codon 673. This novel somatic mutation was identified in a poorly differentiated tumour, located in the descending colon, of CRC145, a 50-year-old black male patient.

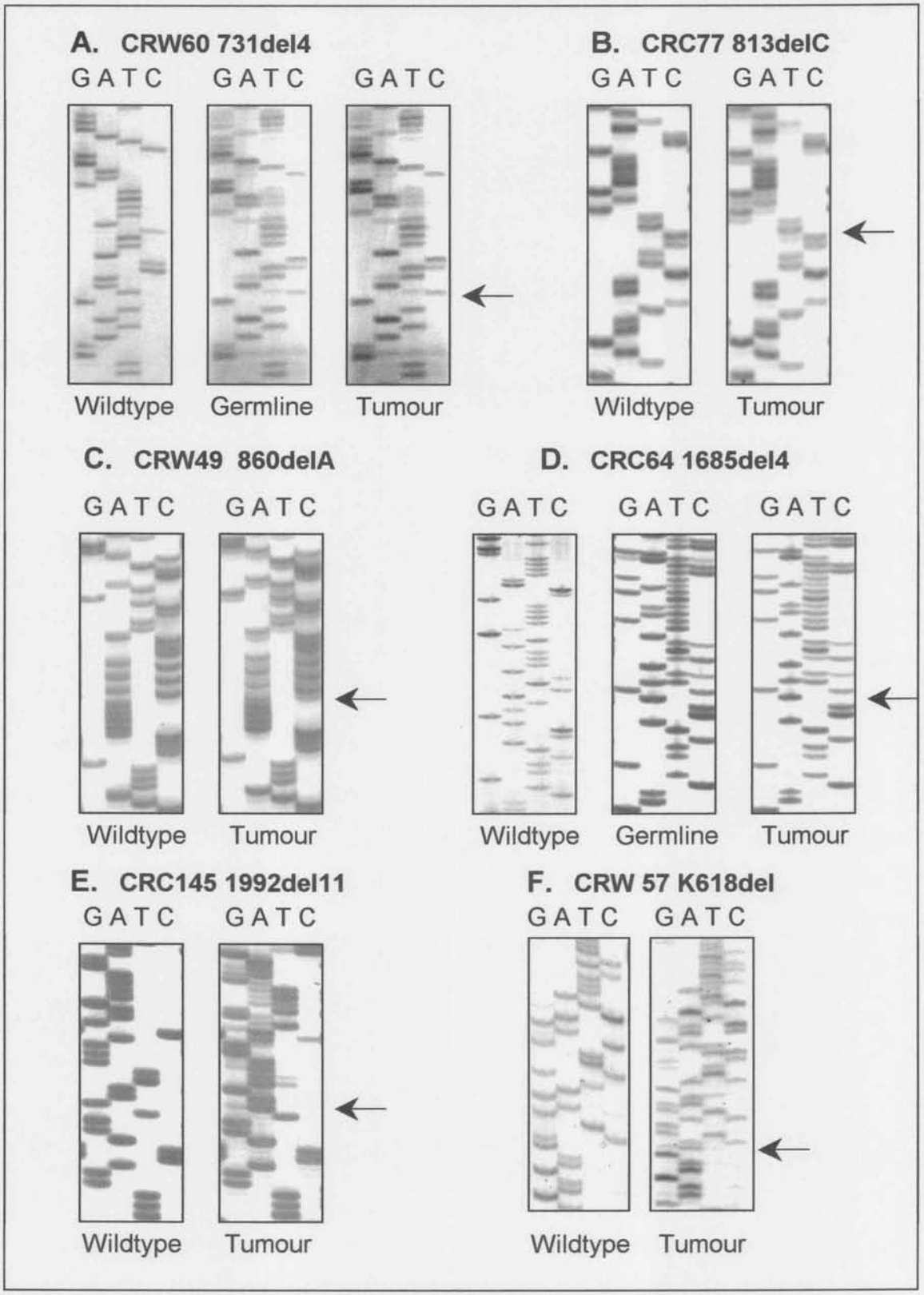


Figure 4.6 Sequence analysis of frameshift mutations and an in-frame deletion
A and C show the sequence of the anti-sense strand of exons 9 and 10. B, D, E and F depict the sense strand sequence of exons 10, 15 and 18. Arrows indicate the start of each deletion.

The deletions all occurred in the 3' half of the gene (exons 9-18). Two are located within the interaction region of hMLH1 with hPMS2 and terminates the protein within this region. The other three deletions terminate the protein before this region therefore all five frameshift mutations would prevent hMLH1 from forming a complex with hPMS2 and no DNA mismatch repair could take place.

Nonsense mutation

Two novel nonsense mutations were identified in hMLH1 (Figure 4.7). The first one, Q86X, was identified in a moderate to poorly differentiated adenocarcinoma of the ceacum of a 44 year-old male (CRC33) and is the only homozygous mutation identified in this study. It is a cytosine to thymine transition, at a CpG dinucleotide, located in exon 3. The second nonsense mutation, E172X, is the result of a guanine to thymine change at nucleotide 514. It was identified in CRC51, a 48-year-old black male that presented with a moderate to poorly differentiated Dukes' C adenocarcinoma in the ascending colon. Both of these mutations are located in the most conserved region of the gene (Han *et al.*, 1995) and will result in a protein without the interaction region of hMLH1 with hPMS2, therefore preventing DNA mismatch repair.

In-frame deletion

One of the previously described mutations identified in this study is the in-frame deletion of one of the three lysine residues at codons 616-618, in exon 16 (Figure 4.7). The mutation has been described ten times before (ICG-HNPCC database) but only in HNPCC families. This is the first somatic report of the mutation and it was identified in the tumour of CRW57, a 72 year-old Caucasian female that had a moderately differentiated tumour in the ceacum that had infiltrated the lymph nodes. Shimodaira *et al.* (1998) carried out functional analysis of this in-frame deletion, and found that it results in a non-functional protein. This is expected as the deletion occurs within the region of hMLH1 which interacts with hPMS2 and experiments have shown a >95% reduction in hMLH1 binding to hPMS2 when this mutation is present (Guerrette *et al.*, 1999).

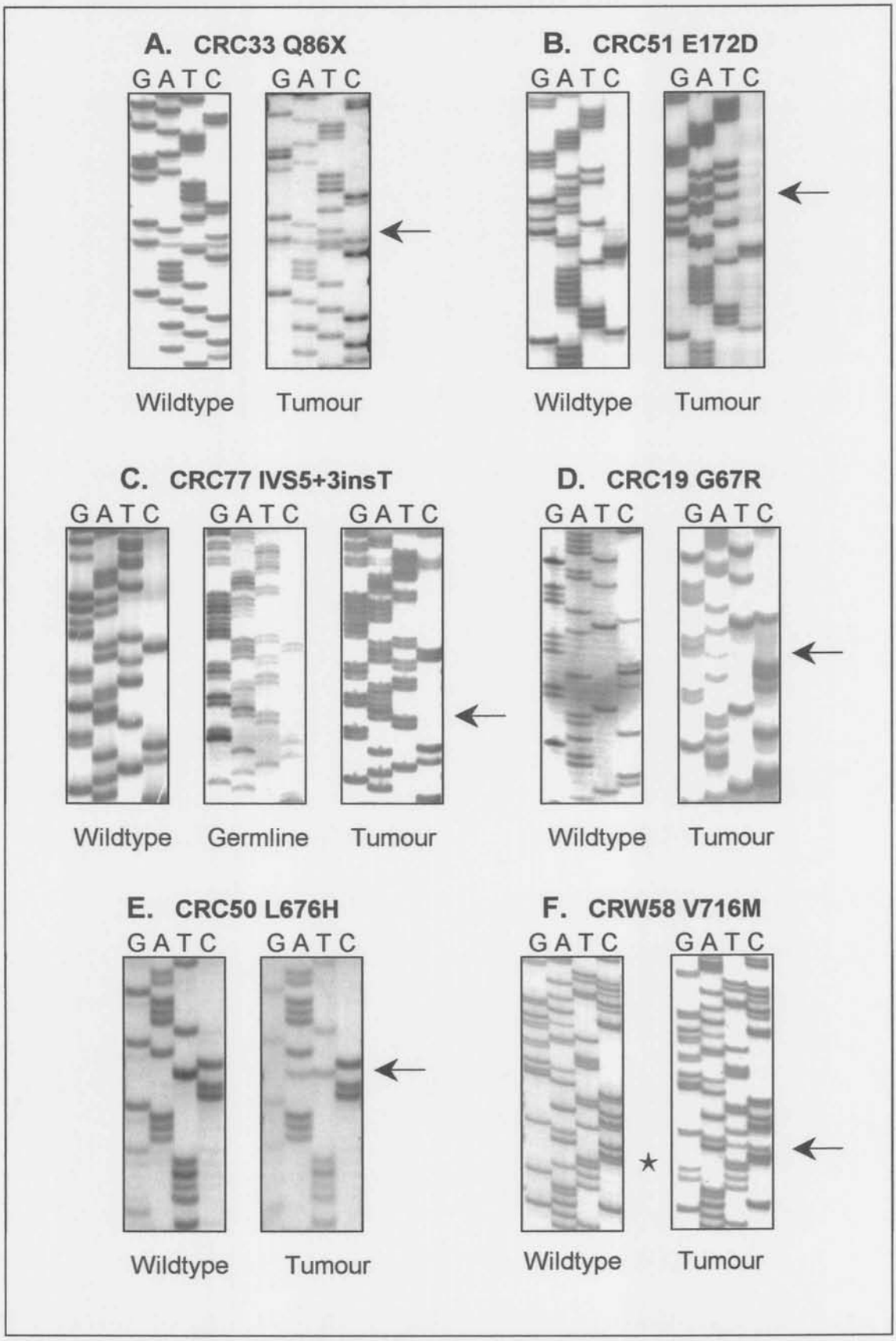


Figure 4.7 Sequence analysis of nonsense, splice site and missense mutations
A to E are shown as the sense strand and F as the anti sense strand. The arrows indicate the alteration. The star indicates the recognised polymorphism, H718Y.

Splice site mutation

The splicing variant in the tumour of CRC77 (41 year-old female) was an insertion T at position 3 of the splice donor site sequence in intron 5 (Figure 4.7). This novel germline mutation could possibly affect the correct splicing of this exon as the nucleotide in this position is an adenine 57% of the time and thymine has a frequency of only 2% at this position (Shapiro and Senapathy, 1987). If exon 5 is skipped as a result of this mutation it would be out of frame and would result in a premature stop at codon 159 in exon 6.

Missense mutations

One of the youngest patients with MSI-H is a 14-year-old girl (CRC19) with cancer of the ceacum. She carries a somatic missense mutation in exon 2, G67R, that has been described previously (ICG-HNPCC database), but this is the first report of the mutation as a somatic event (Figure 4.7). It is a single base substitution (CC**G**GGA→ CCAGGA) in a CpG dinucleotide. Functional studies showed that this non-conservative amino acid change, glycine to argenine, results in a non-functional hMLH1 protein (Shimodaira *et al.*, 1998).

A somatic novel missense mutation (L676H) was identified in CRC50, a 65-year-old black male patient, with moderately differentiated cancer of the ceacum (Figure 4.7). Cotton *et al.* (1998) has suggested criteria for the determination of the pathogenesis of missense mutations. This mutation does not occur in an evolutionary conserved part of the gene (exon 18), but it is a somatic alteration and a non-conservative amino acid change, leucine (non-polar) to histidine (basic). The ultimate test for the pathogenesis of the missense mutation would be functional studies, but since that could not be performed, the information at hand is sufficient to suggest that it may possibly be pathogenic.

A germline missense mutation in hMLH1, V716M, and a polymorphism H718Y, were identified in CRW58, a 63-year-old Caucasian female diagnosed with poorly differentiated adenocarcinoma (Figure 4.7). The mutation is a conservative amino acid change (nonpolar) and does not occur in an evolutionary conserved region or at the interaction region of the gene. To determine whether this is a polymorphic change, 100 Caucasian control chromosomes were analysed for this alteration. None of the control chromosomes were found to carry this mutation suggesting that

this is not a polymorphism (Cotton *et al.*, 1998). This mutation has also been identified in a Swiss family (Hutter *et al.*, 1998). It was described as a putative pathogenic mutation as it was identified in the germline of a 36-year-old patient who also carried a pathogenic nonsense germline mutation. Both parents of the patient presented with early CRC, but was not analysed. Analysis of the segregation of the mutation with the disease could therefore not be determined. Thus even though this mutation could not be identified in 100 control chromosomes, none of the available evidence suggests that it is pathogenic. It is therefore only a putative pathogenic mutation.

The L676H and V716M are both classified as putative pathogenic mutations and will not be included in any further analysis of pathogenic mutational data in this study.

Summary of pathogenic mutations identified in hMLH1

Ten pathogenic hMLH1 mutations were identified in 9/33 (27%) patients (Table 4.6). Patient CRC77 presented with two different hMLH1 mutations (germline and somatic).

Table 4.6 Pathogenic mutations detected in hMLH1

Patient	Exon / IVS	Codon	Nucleotide change (cDNA sequence)	Designation	Consequence
CRC 19	2	67	199 G→A	G67R [a]	Gly→Arg
CRC 33	3	86	256 C→T	Q86X	Gln→Stop
CRC 77	IVS 5	-	453+3 ins T	IVS5+3insT [a]*	Exon skipping
CRC 51	6	172	514 G→T	E172X	Glu→Stop
CRW 60	9	244	731 del GTTA	731del4 *	FS ter 253
CRC 77	10	271	813 del C	813delC	FS ter 272
CRW 49	10	287	860 del A	860delA	FS ter 296
CRC 64	15	562	1685 del AGAT	1685del4 *	FS ter 589
CRW 57	16	618	1852 del AAG	K618del [a]	Lys deletion
CRC 145	18	664	1992delGAATTGGGACG	1992del11	FS ter 673
CRC 50	18	676	2027 T→A	L676H[b]	Leu→His
CRW 58	19	716	2146 G→A	V716M [ab]*	Val→Met

Recommended nomenclature system (Antonarakis *et al.*, 1998); a = Previously described; b = Putative pathogenic mutation; FS ter = frameshift terminate at codon_; ins = insertion; del = deletion; IVS= intervening sequence; * = germline mutation

The 10 pathogenic mutations were detected in 3/11 (27%) Caucasian and 6/22 (27%) black patients' tumours. Seven of the ten mutations are novel. Three of the 10 were germline, but it is unknown whether they are familial or *de novo* alterations. Two putative pathogenic missense mutations were identified but are not included in any of these calculations.

Even though the number of hMLH1 mutations identified in this study is only a fraction of the total hMLH1 mutations reported to date, one similar feature exists when comparing the spectrum of mutations in the present study with the ICG-HNPCC database (Figure 4.8). Frameshift mutations are the most common type of mutation (50% [4/8] somatic frameshifts and 66% [2/3] germline frameshifts). The proportion of somatic nonsense mutations (29%; 2/7) is much higher than the reported 6% germline nonsense mutations in the database. More germline splice site mutations were identified in the present study (33%; 1/3) than reported. No somatic splice site mutations or germline nonsense and missense mutations were identified in the present study.

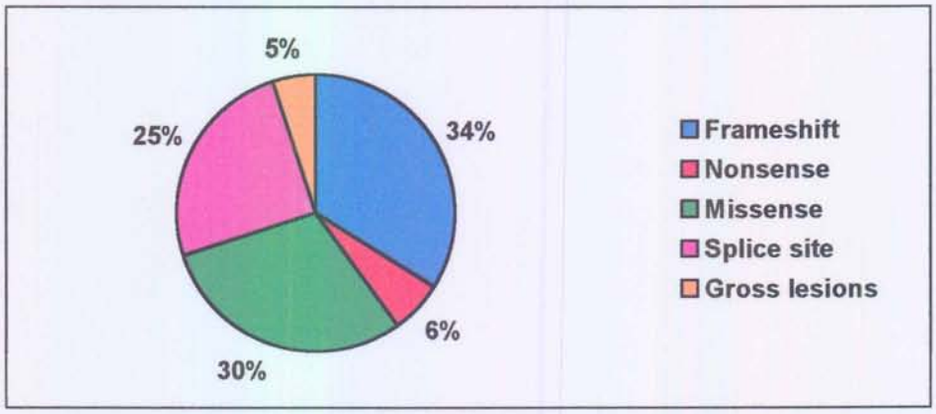


Figure 4.8 Spectrum of reported germline hMLH1 mutations (ICG-HNPCC database)

The mutations are distributed throughout the hMLH1 gene (Figure 4.9). Four of the ten pathogenic mutations in hMLH1 are located within the most conserved region, three are located in the interaction domain of hMLH1 with hPMS2, and the three frameshifts located in exons 9 and 10 will terminate the protein before the interaction domain. This shows that all of these mutations will affect the function of the hMLH1 protein.

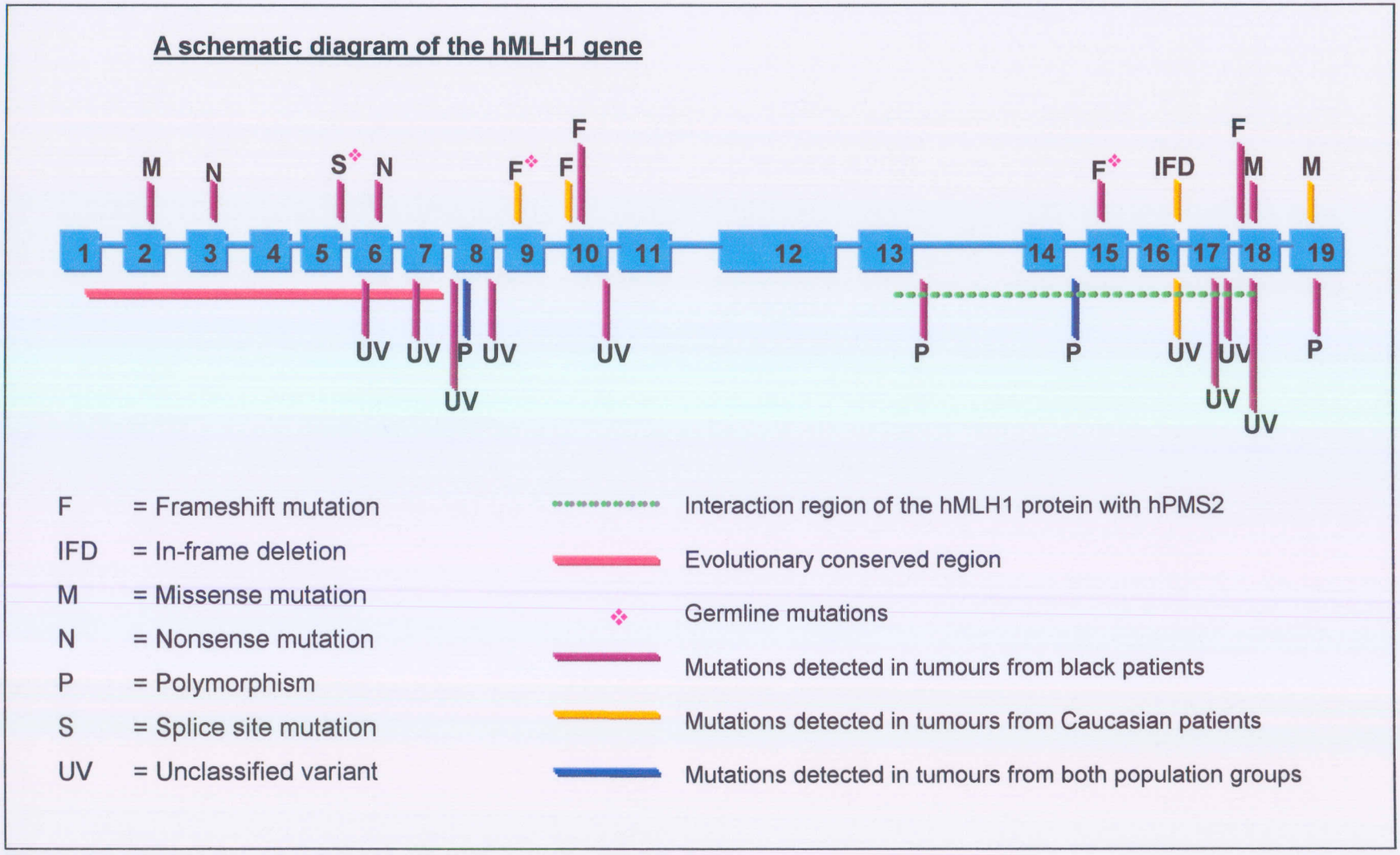


Figure 4.9 Distribution of sequence alterations in the hMLH1 gene

4.3.3 Polymorphisms in hMLH1

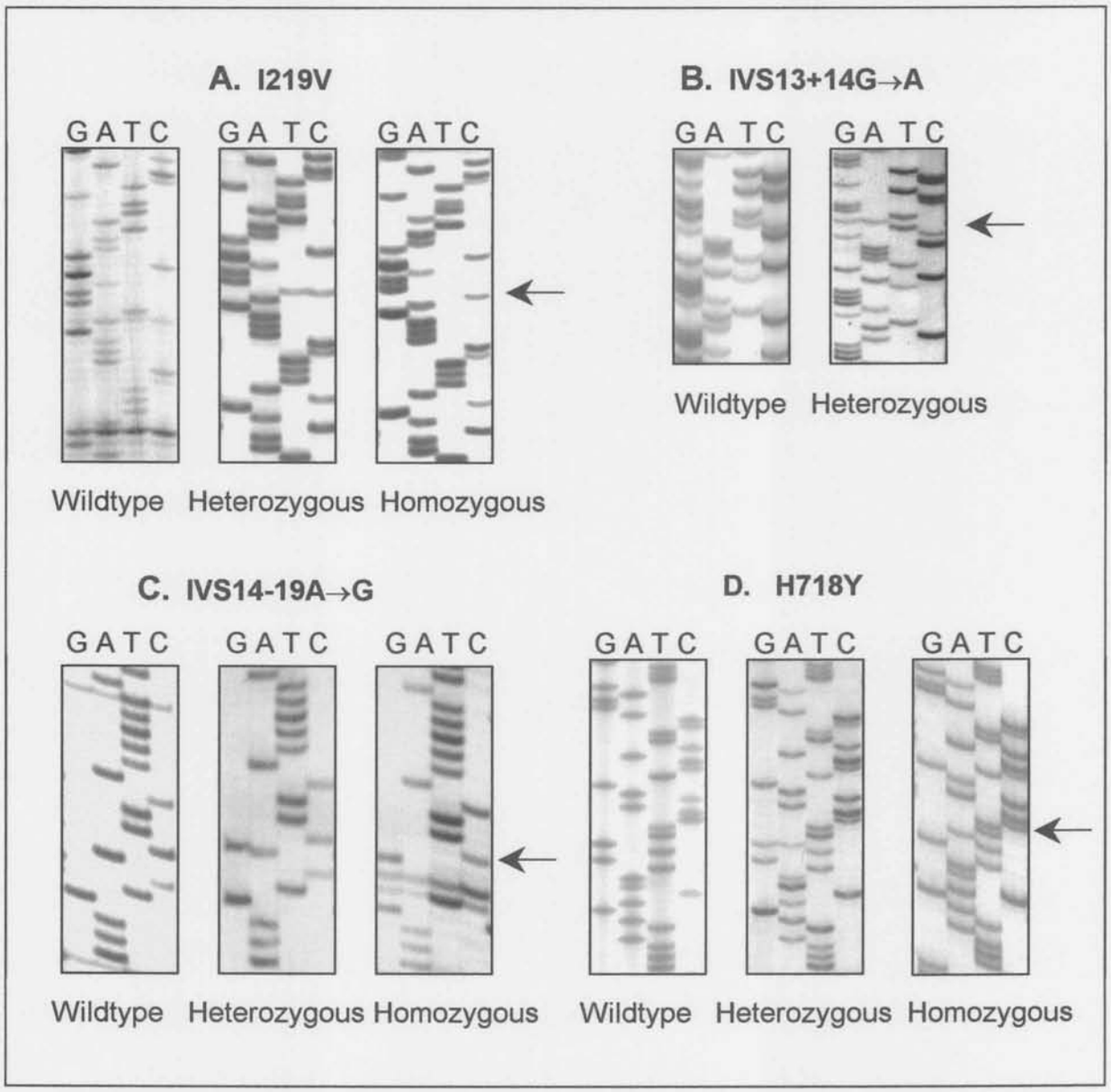
Four previously described polymorphisms were identified in the 33 MSI-H patients (Table 4.7 and Figure 4.10).

Table 4.7 Polymorphisms identified in hMLH1

Exon / IVS	Codon	Nucleotide Change cDNA sequence	Designation	Consequence
8	219	655 A→G	I219V	Ile→Val
IVS13	-	1558+14 G→A	IVS13+14G→A	-
IVS14	-	1668-19 A→G	IVS14-19A→G	-
19	718	2152 C→T	H718Y	His→Tyr

Recommended nomenclature system (Antonarakis *et al.*, 1998); IVS = intervening sequence

Nine of the black patients (eight heterozygous and one homozygous) presented with the sequence variant, H718Y, which has been reported three times before. It is a cytosine to thymine alteration at nucleotide 2152 in exon 19 and the resulting histidine to tyrosine amino acid change is non-conservative. Two groups investigating colorectal cancer patients identified this alteration (Farrington *et al.*, 1998; Weber *et al.*, 1999). Farrington *et al.* (1998) identified this missense mutation in 3 control individuals from the United States (two of whom were siblings) of whom the ethnic origin is unknown. Weber *et al.* (1999) investigated African American colorectal cancer patients and controls, and found three patients and one control to carry the alteration (allele frequency is 7.3% for both patients and controls). Kowalski *et al.* (1997) analysed hMLH1 and hMSH2 for mutations in microsatellite unstable endometrial cancers. The missense, H718Y, was identified in two of their cancer specimens, both from African American patients. Evaluation of control DNA revealed that the polymorphism is unique to African Americans with an allele frequency of 14%. To determine whether the polymorphism is also unique to black South Africans, 49 black South African controls (98 alleles) and 50 Caucasian South African controls (100 alleles) were analysed using two-lane sequencing. Nine black controls (seven heterozygous and two homozygous) and none of the Caucasian controls presented with the polymorphism. The allele frequency in the black control group was 11,2% and 12,8% for both patients and controls. This is similar to the reported allele frequencies.



Arrows indicate the sequence alterations.

Figure 4.10 Sequence analysis of polymorphisms identified in hMLH1
Sequences are shown in the antisense strand (A and D), and in the sense strand (B and C).

Interestingly only the black patients presented with the IVS13+14G→A polymorphism (3/27), although it has been described in Swedish families (Tannergård *et al.*, 1995). The IVS14-19A→G and I219V polymorphisms are common in both black and Caucasian patients and have been reported in kindreds from Germany, Switzerland, Sweden and Finland (ICG-HNPCC mutation database). Functional studies indicated that the I219V polymorphism has no effect on the function of the hMLH1 protein (Shimodaira, 1998).

4.3.4 Unclassified variants in hMLH1

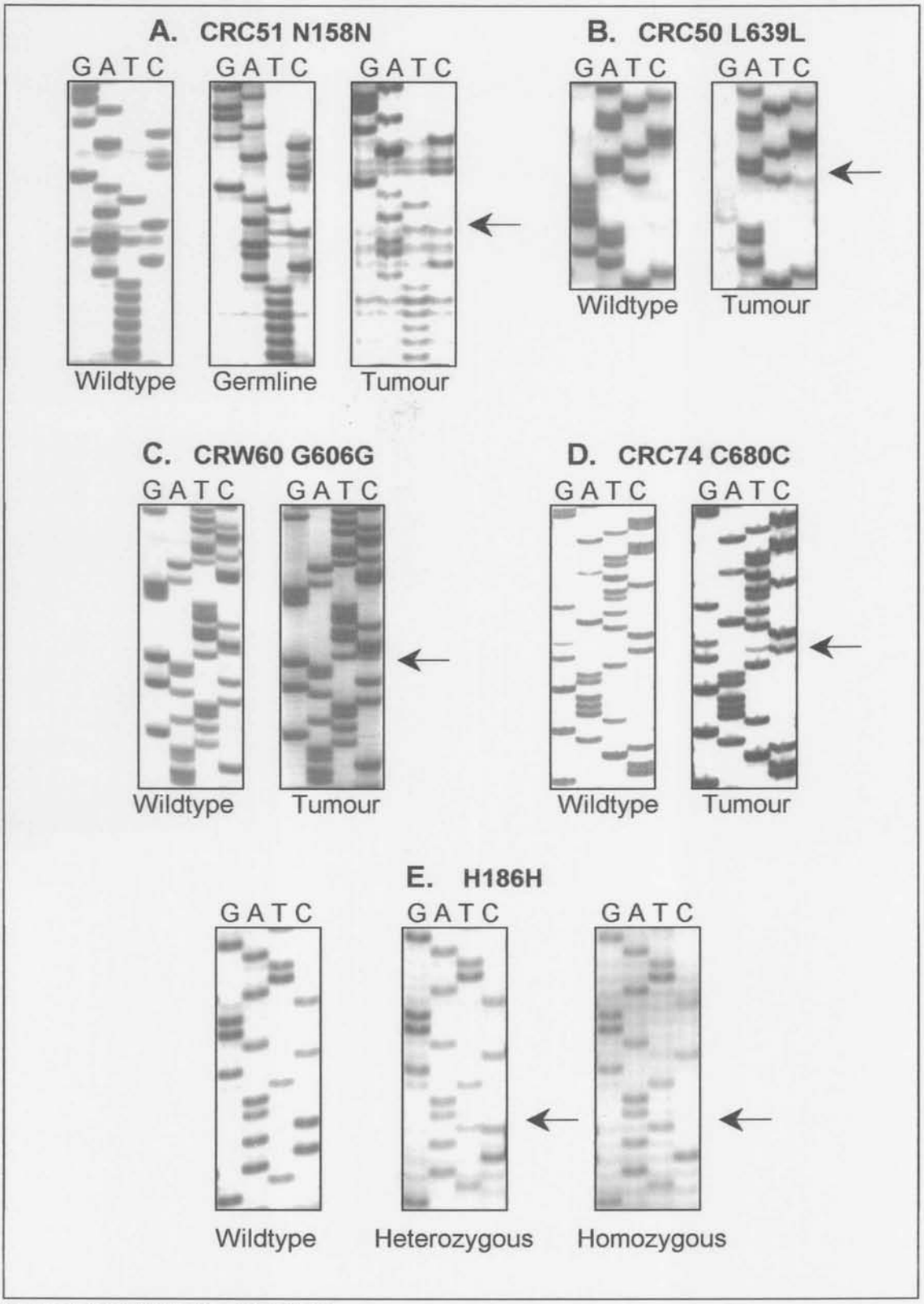
A number of sequence alterations that are probably not pathogenic but possibly polymorphic were detected. Nine of these unclassified variants (Table 4.8) were detected in hMLH1 of which five (in six tumours) are silent alterations (Figure 4.11) and four occurred in non-coding regions (Figure 4.12). Only one of the single base substitutions occurs in CpG dinucleotides, C680C (TGCGCT→TGTGCT). None of these sequence alterations are expected to be pathogenic and may possibly be polymorphic. Only a few of these unclassified variants were checked in the germline of the patients (N158N, IVS8+20delA and IVS10-21delTC). Tumours of patients CRC33 and CRC77 presented with the same silent mutation, H186H, but CRC33 showed a homozygous change whereas CRC77 a heterozygous change.

Point mutations in introns 7 and 17 were identified in patients CRC48 (IVS7-15C→T) and CRC60 (IVS17-39T→C). Two deletions, IVS10-21delTC and IVS8+20delA, were detected in patients CRC54 and CRC153 respectively. The IVS10-21delTC alteration was also identified in the germline of the patient. The alteration in CRC153, IVS8+20delA, was the only one found to be somatic, which means that it is not polymorphic. It is unknown if any of these intronic sequence alterations may have an effect on splicing.

Table 4.8 Unclassified variants identified in hMLH1

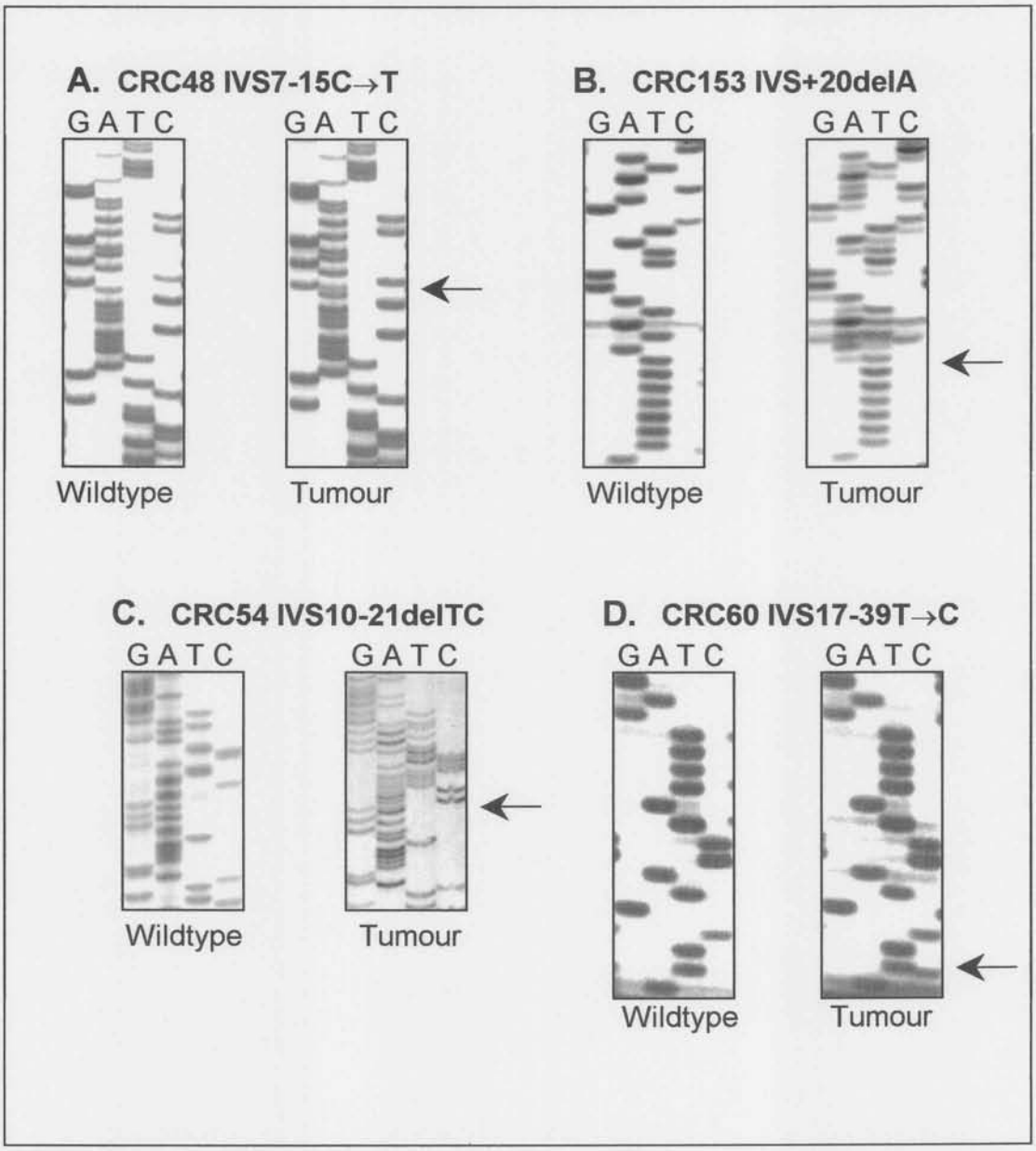
Patient	Exon / IVS	Codon	Nucleotide change (cDNA sequence)	Designation	Consequence
CRC 51	6	158	474 C→T	N158N	Asn→Asn
CRC 33,77	7	186	558 C→T	H186H	His→His
CRC 48	IVS7	-	589-15 C→T	IVS7-15C→T	-
CRC 153	IVS8	-	677+20 del A	IVS8+20delA	-
CRC 54	IVS10	-	885-21 del TC	IVS10-21delTC *	-
CRW 60	16	606	1818 A→G	G606G	Gly→Gly
CRC 50	17	639	1917 A→G	L639L	Leu→Leu
CRC 60	IVS17	-	1990-39 T→C	IVS17-39T→C	-
CRC 74	18	680	2038 C→T	C680C	Cys→Cys

Recommended nomenclature system (Antonarakis *et al.*, 1998); del = deletion; IVS = intervening sequence; * = germline mutation



The arrows indicate the alterations

Figure 4.11 Sequence analysis of the unclassified variants identified in the coding regions of the hMLH1 gene The sense strands (A, D and E) and antisense strands (B, C and F) were sequenced.



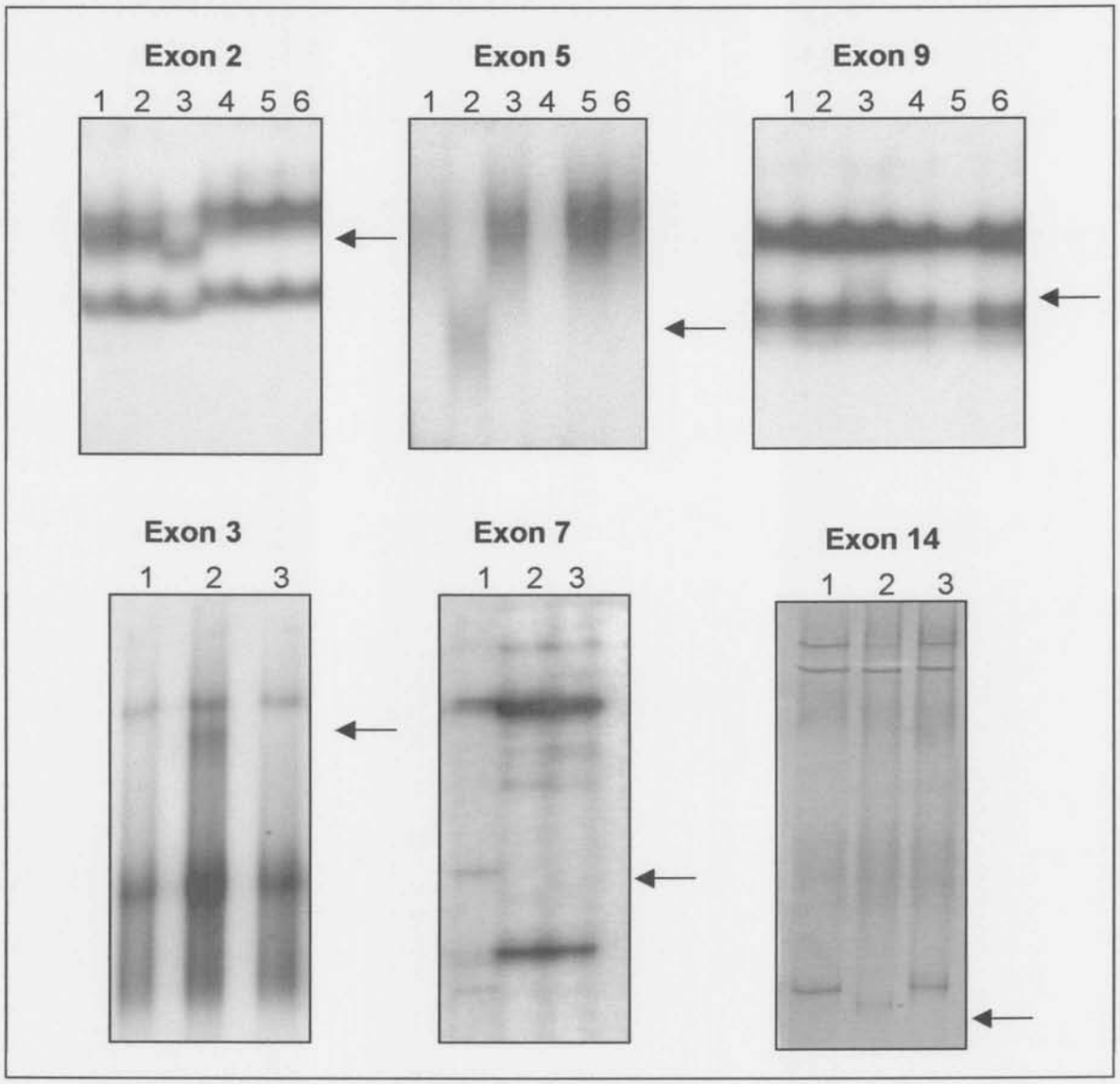
The arrows indicate the alterations

Figure 4.12 Sequence analysis of the unclassified variants identified in intronic sequences of hMLH1 Sequence alterations are shown in the sense strand (D) and in the antisense strand (A, B and C).

4.4 hMSH2 GENE ANALYSIS

4.4.1 SSCP analysis

The 40 MSI-H patients were screened for mutations in the 16 coding regions of the hMSH2 gene using exon-by-exon PCR SSCP analysis. Two tissue samples (CRC129 and CRW92) could not be amplified for exon 14. Thus only 38 MSI-H tumours were fully screened for hMSH2 mutations. Representative autoradiographs of some of the exons, in which aberrant bands were detected, are presented in figure 4.13. Those samples showing bands with altered mobility were sequenced.



Arrows indicate aberrant band migration in the different exons

Figure 4.13 Representative samples of aberrant band migration in some of the exons of the hMSH2 gene

4.4.2 Pathogenic mutations in hMSH2

Eight hMSH2 mutations were detected in the 38 MSI-H tumours that were fully screened for hMSH2 mutations. Six are pathogenic and includes three frameshift mutations; one nonsense and two splice site mutations. Two putative pathogenic missense mutations were identified.

Frameshift mutations

Three black male patients' tumours harboured frameshift mutations (Figure 4.14). All three of these mutations are deletions, two of which occurred in exon 2. One of these, 227delAG, was identified in the germline of patient CRC22, a 42-year-old with a moderately differentiated adenocarcinoma in the sigmoid and metastasis in the lymph nodes. This two base pair deletion creates a premature stop at codon 80 in exon 2. The mutation has been reported before in a Dutch family that does not fulfil the Amsterdam criteria, since the affected individuals occur only in one generation (Wijnen *et al.*, 1997). The second mutation identified in exon 2, is a five base pair deletion at nucleotide 261 that terminates the protein at codon 98 (in the same exon). This novel somatic mutation was identified in CRC50, who presented with a moderately differentiated adenocarcinoma of the ceacum at the age of 65. A one base pair deletion, 2333delG, was identified in the germline of CRC92 (47 years) who was diagnosed with a moderately differentiated adenocarcinoma of the rectum. This novel mutation in exon 14 terminates the protein at codon 811 in the same exon.

Nonsense mutation

The only nonsense mutation found in hMSH2 was detected in exon 7 of the tumour of CRW58. This female Caucasian patient was diagnosed with poorly differentiated adenocarcinoma of the rectum at the age of 63 years. This single base substitution occurs in a CpG dinucleotide. The mutation is a C→T transition (Figure 4.14) at nucleotide 1148 that introduces a premature termination codon at codon 383. The mutation has been described previously in a Swiss family that fulfil the Amsterdam criteria (Buerstedde *et al.*, 1995). However, this is the first time that this mutation has been observed as a somatic event. A germline putative pathogenic missense mutation, V716M, in hMLH1 has also been identified in this patient. But since this patient is 63 years old and together with all the other reasons discussed in section 4.3.2, the somatic nonsense mutation, R383X, is considered to be the pathogenic mutation in this patient.

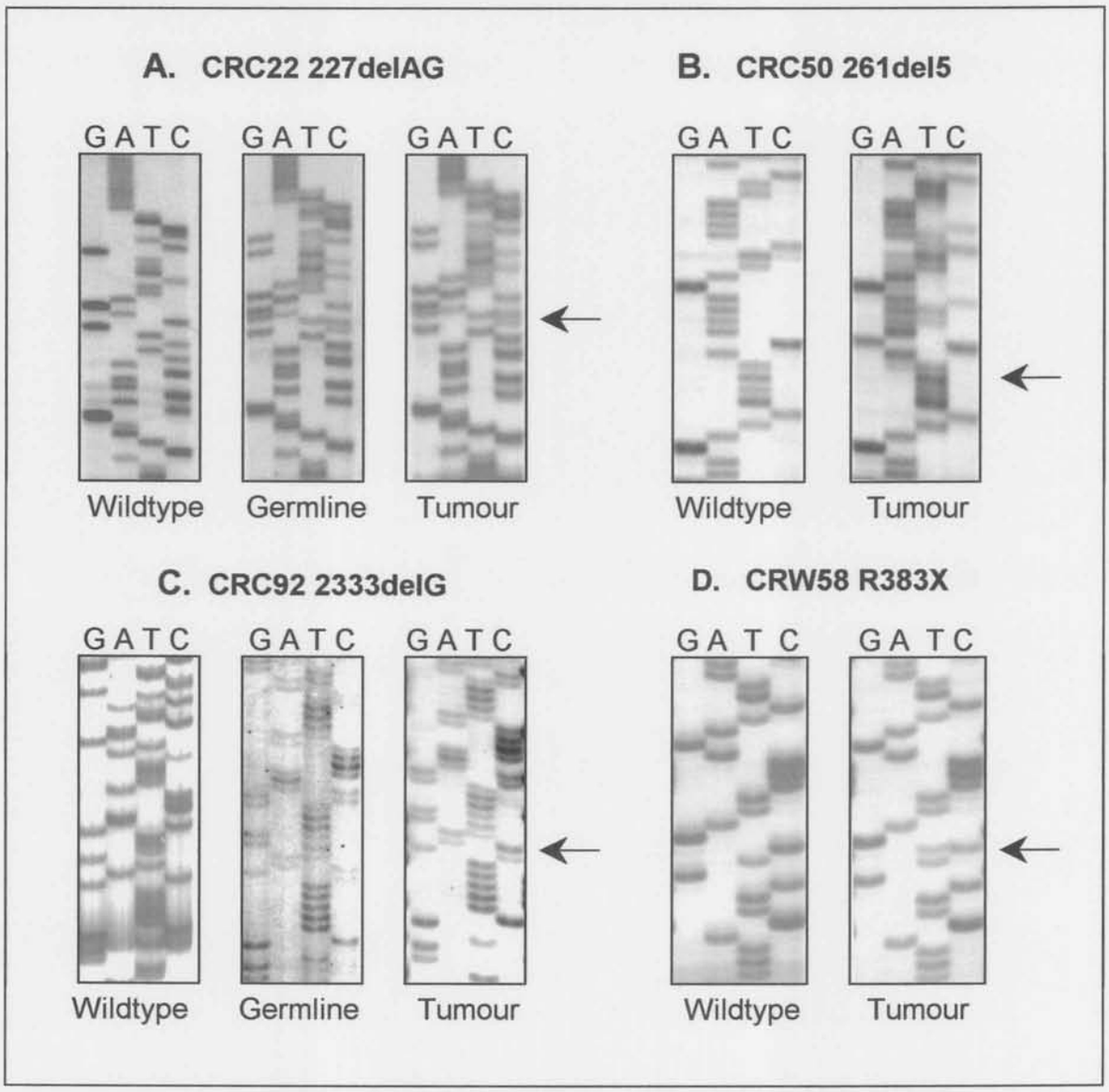


Figure 4.14 Sequence analysis of frameshift and nonsense mutations in hMSH2
Sequences are shown as the antisense strand in A and B, and the sense strand in C and D. In A, B and C the arrows indicate the start of the deletions and in D it indicates the nucleotide substitution

Splice site mutations

Two of the eight disease-causing mutations identified in hMSH2, occurred in the donor-splice site region of exon 5, causing the in-frame deletion (skipping) of exon 5. To date, the single most common hMSH2 mutation is the splice site mutation IVS5+3A→T (ICG-HNPCC Mutation database; also see paragraph 2.2.4) that was identified in the black patient, CRC48 (figure 4.15). However, this is the first time that this mutation has been observed as a somatic event in a colorectal cancer patient. Previously it was observed as a somatic mutation in an endometrial cancer patient (Desai *et al.*, 2000). It has been found that male carriers of this mutation are at higher risk of colorectal cancer, with a younger age at diagnosis than females

(38.8 versus 47.2 years) (Froggatt *et al.*, 1999). Interestingly, CRC48 is a 32 year-old black male with a moderately differentiated adenocarcinoma of the ceacum with metastasis in the lymph nodes. Detection of the IVS5+1G→A mutation in CRC19 (14 year-old female, grade II Dukes' B adenocarcinoma of the ceacum), lends further credence to the suggestion that this site is a hotspot for mutations. The guanine at position +1 is invariant (Shapiro and Senapathy, 1987).

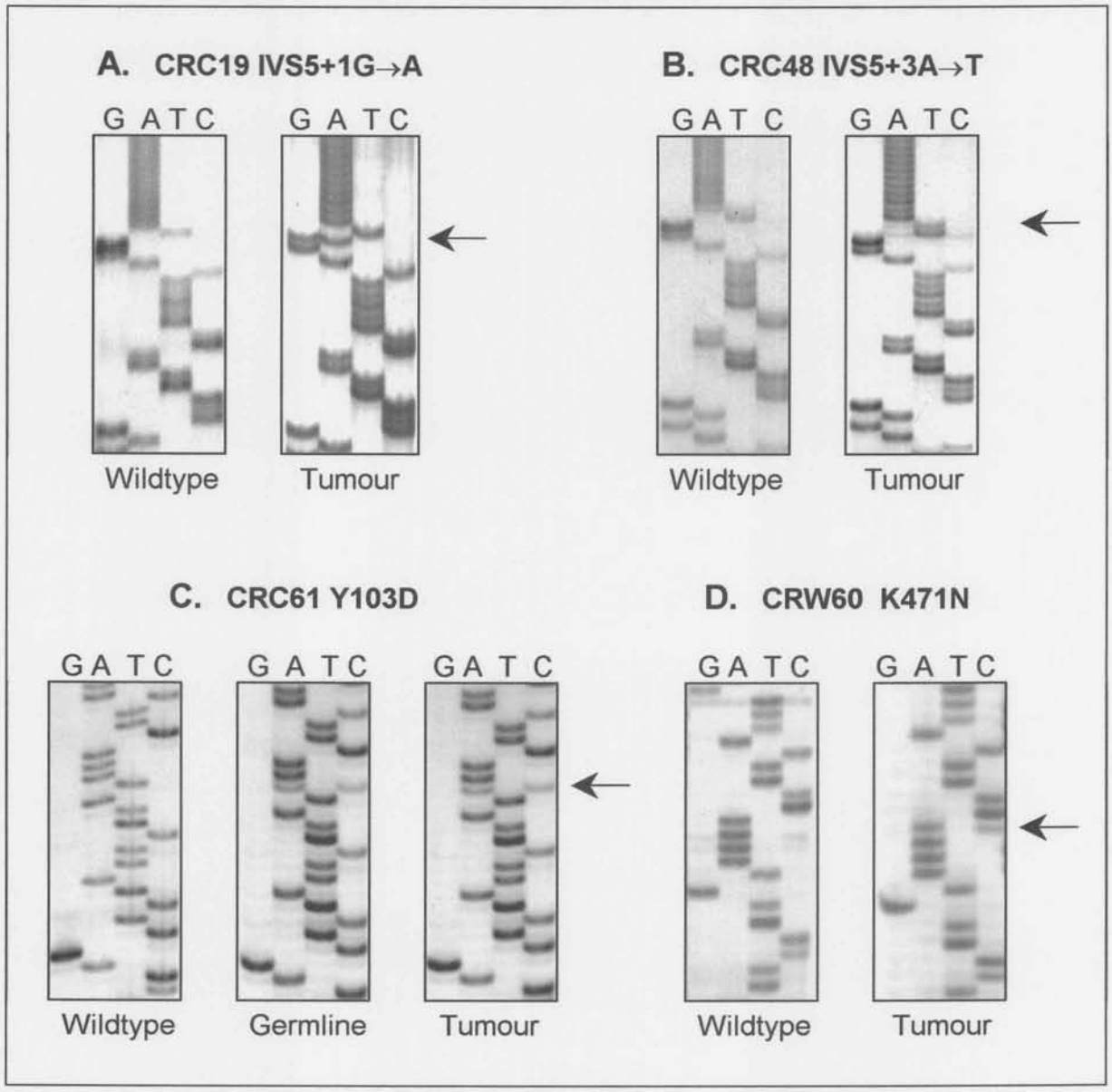


Figure 4.15 Sequence analysis of splice site and missense mutations in hMSH2
Arrows indicate the mutations in the sense strand A, B and D and the antisense strand in C.

Missense mutations

Two novel missense mutations were identified in hMSH2 (Figure 4.15) that may possibly be pathogenic. One, Y103D, was identified in the germline of CRC61, a black male patient that presented with a moderately differentiated adenocarcinoma of the colon in the area of the splenic flexure, at the age of 60 years. This is not a conservative amino acid but it is a non-conservative amino acid change, polar to acidic. This germline alteration was not identified in 50 control chromosomes analysed.

The second missense, K471N, is a somatic alteration identified in the tumour of CRW60, a 51-year-old male with moderately differentiated adenocarcinoma of the colon. This mutation could be the second hit, since a germline hMLH1 mutation has been identified in this patient. The K471N mutation is a somatic, non-conservative (Basic to polar) alteration that is located within the amino-terminal domain of the interaction region of hMSH2 with hMSH3 or hMSH6. Guerrette *et al.* (1998) investigated six missense mutations (that has been reported to co segregate with HNPCC) occurring in these interaction regions and found that they do not interfere with the interaction of these proteins hMSH2-hMSH3, hMSH2-hMSH6. Other functional studies therefore need to be carried out to determine whether the protein is still functional in the repair of mutations when missense mutations are present in these regions.

Both these mutations are putative pathogenic mutations and will not be included in any further analysis.

Summary of pathogenic hMSH2 mutations

In conclusion, six hMSH2 pathogenic mutations were detected in the 38 MSI-H patients (16%), 1/12 (8%) tumours of Caucasian patients and 5/26 (19%) tumours of black patients (Table 4.9). Two of the six mutations were germline mutations identified in black patients and three of the six are novel mutations. One of the previously described mutations identified, is the splice site mutation in intron 5, which is considered a mutational hotspot.

Table 4.9 hMSH2 pathogenic mutations

Patient	Exon/ IVS	Codon	Nucleotide change CDNA sequence	Designation	Consequence
CRC 22	2	76	227 del AG	227delAG [a] *	FS ter 80
CRC 50	2	87	261 del TTTTG	261del5	FS ter 98
CRC 61	2	103	307 T→G	Y103D *[b]	Tyr→Asp
CRC 19	IVS 5	-	942+1 G→A	IVS5+1G→A	Exon skipping
CRC 48	IVS 5	-	942+3 A→T	IVS5+3A→T [a]	Exon skipping
CRW 58	7	383	1148 C→T	R383X [a]	Arg→stop
CRW 60	9	471	1413 A→C	K471N[b]	Lys →Asn
CRC 92	14	778	2333 del G	2333delG *	FS ter 811

Recommended nomenclature system (Antonarakis *et al.*, 1998); a = Previously described; b = Putative pathogenic mutation; * = germline mutation; FS ter = Frameshift terminate at codon _; del = deletion; IVS = intervening sequence

Although no meaningful comparison can be made when the sample size is three, interestingly germline frameshift mutations were also the most common mutation in this study as was reported in the ICG-HNPCC database (Figure 4.16).

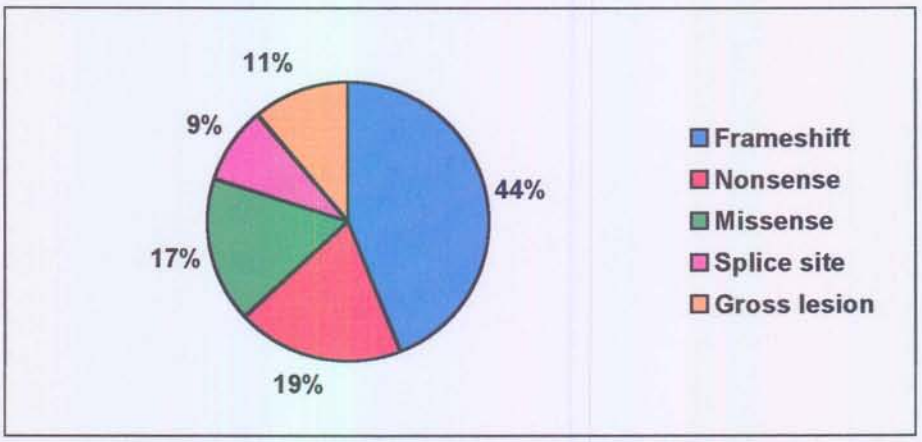


Figure 4.16 Spectrum of reported germline hMSH2 mutations (ICG-HNPCC database)

The mutations identified are distributed throughout the gene (Figure 4.17). Only one of the pathogenic mutations, 2333delG, is located in the evolutionary conserved region and the nonsense mutation, R383X, is located in the amino-terminal domain of the interaction region of hMSH2 with hMSH3 or hMSH6. The two frameshift mutations in exon 2 will terminate the protein before the conserved region or the interaction domains. It is clear that five of the six mutations may interfere with the interaction of the hMSH2 protein with either hMSH3 or hMSH6.

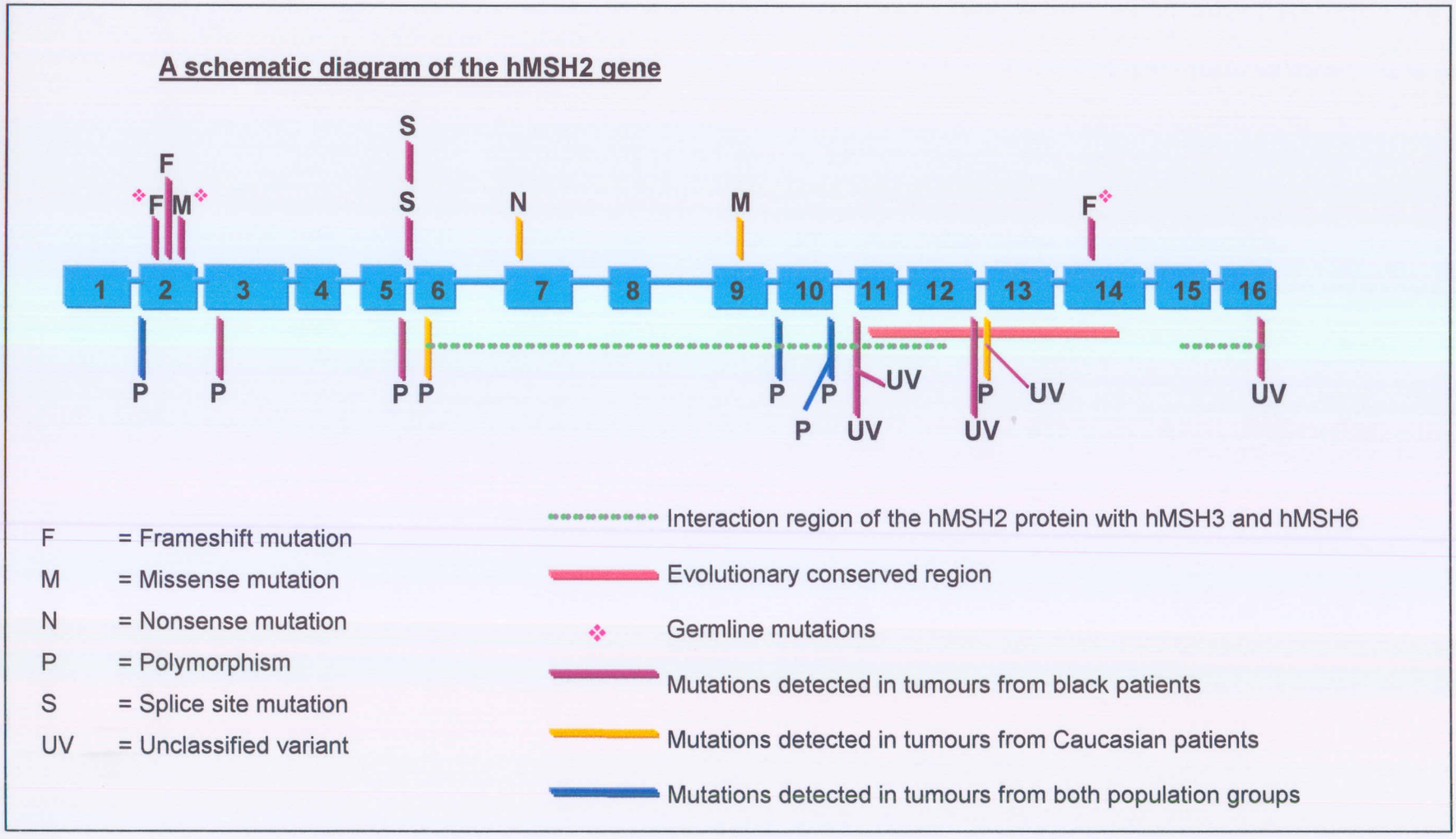


Figure 4.17 Distribution of sequence alterations in the hMSH2 gene

4.4.3 Polymorphisms in hMSH2

Seven previously described polymorphisms were identified in hMSH2 (Table 4.10). Six of the seven polymorphisms are point mutations, of which five are transitions and one is a transversion. The other previously described polymorphism is a missense mutation.

Table 4.10 Polymorphisms identified in hMSH2

Exon / IVS	Codon	Nucleotide Change cDNA sequence	Designation	Consequence
IVS 1	-	212 -4 poly T	212-4del$(T)_n$	-
3	127	380 A→G	N127S	Asn→Ser
6	322	965 G→A	G322D	Gly→Asp
IVS 5	-	942+27 poly A	942+27del$(A)_n$	-
IVS 9	-	1511-9 T→A	IVS9-9T→A	-
IVS 10	-	1661+6 T→C	IVS10+6T→C	-
IVS 10	-	1661+12 A→G	IVS10+12A→G	-
IVS 12	-	2006-6 T→C	IVS12-6T→C	-

Recommended nomenclature system (Antonarakis *et al.*, 1998); IVS = intervening sequence

A missense mutation, N127S, was identified in the germline of three of the 26 black MSI-H patients (CRC49, CRC57, CRC85). The adenine to guanine transition (figure 4.18) causes a conservative amino acid change. None of the 12 Caucasian MSI-H patients carried this alteration that has been described as a missense mutation in a Nigerian family (ICG-HNPCC mutation database). To investigate the possibility that this missense mutation is a polymorphism and that it is unique to the black population, 49 black controls and 50 Caucasian controls were screened for this missense using two-lane sequencing. Eight of the black controls (all heterozygous) and none of the Caucasian controls presented with the mutation. The allele frequency of the G allele was similar in the black patients (5,8%) and in the black controls (8,2%). Analysing these results with the chi-square test for 2x2 contingency table, it was found that the variant is not associated with colorectal cancer. This previously described missense mutation is therefore a polymorphism that appears to be unique to the black population.

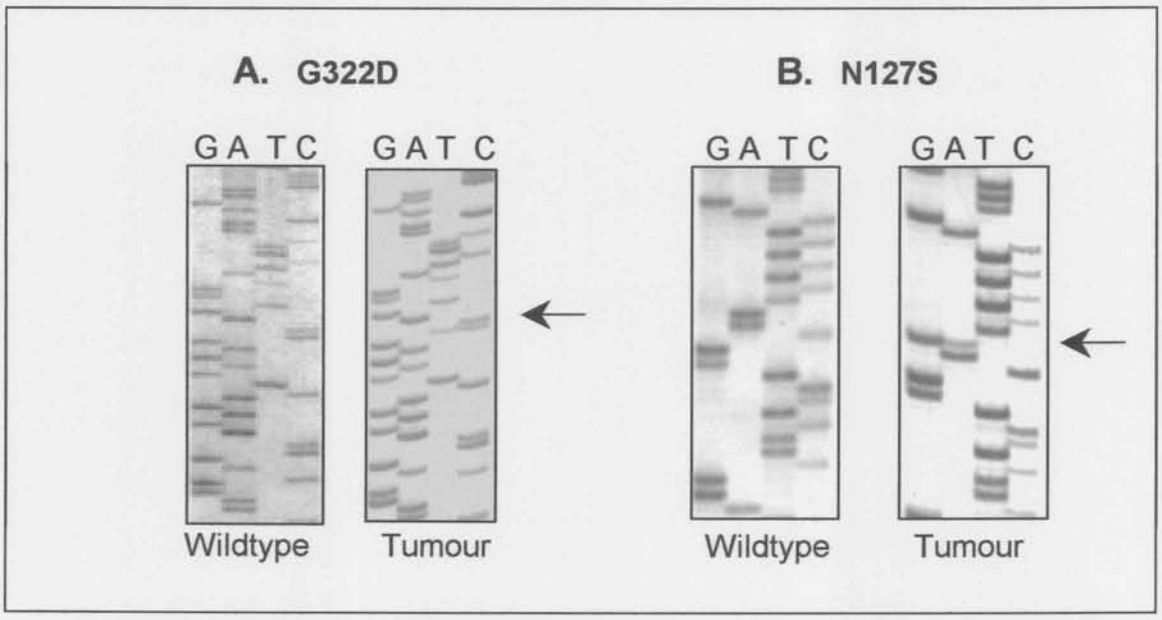


Figure 4.18 Sequence analysis of missense polymorphisms identified in hMSH2

The arrows indicate the G to A transition (antisense strand) in A and the A to G transition (sense strand) in B.

Another missense mutation, G322D, which has been described as a polymorphism, was identified in the germline of patient CRW58 (Figure 4.18). It is a single base substitution in a CpG dinucleotide. It is a non-conservative amino acid change (polar glycine to acidic aspartic acid) and the amino acid glycine is conserved in yeast and human MSH2 (Obmolova *et al.*, 2000). The variant has been reported in a family that fulfil the Amsterdam criteria (Maliaka *et al.*, 1996) and in sporadic colorectal cancer patients with an allele frequency of 1-6% (Froggatt *et al.*, 1996; Tomlinson *et al.*, 1997). To clarify the effect of this variant, functional studies was performed in yeast with the putative equivalent mutation G317D. It was found that the mutation partially inactivates mismatch repair. But it was also found that the human mutant interacts with hMSH6 and binds to mismatches. Further functional and population-based studies are required to determine the clinical relevance of this mutation. With the differences between human and yeast cells, it is also not certain that the yeast G317D alteration is indeed equivalent to human G322D (Drotschmann *et al.*, 1999).

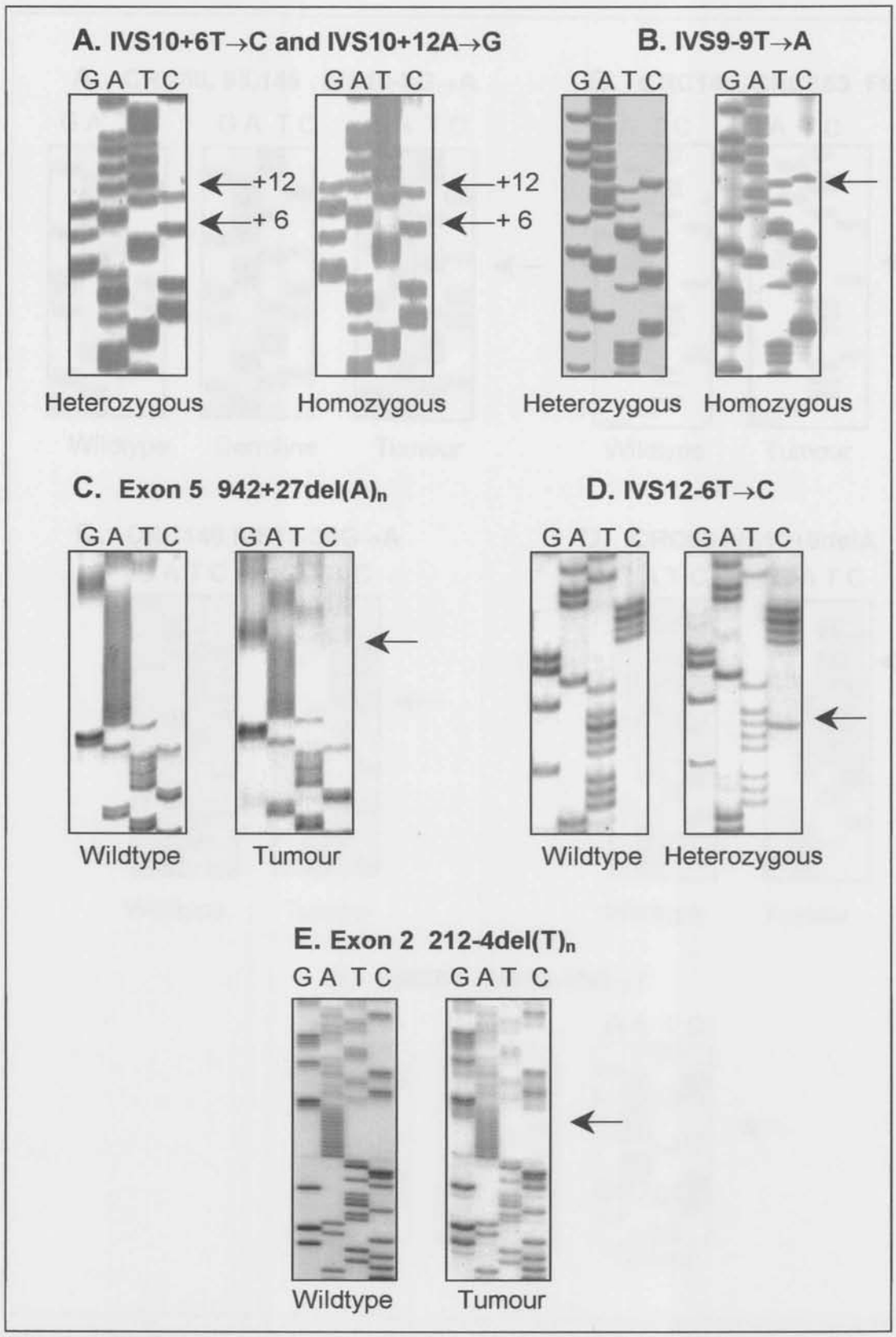


Figure 4.19 Sequence analysis of polymorphisms identified in the noncoding regions of hMSH2 Arrows indicate the alterations in the sense strands (A, C and D), and the antisense strands (B and E).

The polymorphism identified in intron 12, IVS12-6T→C, was detected in seven of the twelve white patients, but none of the black patients (Figure 4.19). The reported allele frequency is 93,7% and 6,3% for the T and C alleles respectively (in 45 patients and 26 controls, Farrington *et al.*, 1998), whereas the allele frequency in the MSI-H patients in this study is 75% and 25%. A common polymorphism occurs at a run of 12/13 thymidines, 4 base pairs upstream of exon 2 (Kolodner RD, personal communication). Most of the samples showed instability in this run of thymines where the tumours had deletions of one to five thymidines (Figure 4.19). This could be another target region for instability as the result of the inactivation of a DNA mismatch repair gene. The polymorphic region at the 5' end of exon 5 in the hMSH2 gene, the 26 adenosine base pairs, is the BAT26 mononucleotide used to screen for microsatellite instability. All of the MSI-H patients showed instability in that marker, therefore all of them have deletions of some degree in that poly A tract (Figure 4.19). Other previously identified polymorphisms were detected in introns 9 and 10 (Figure 4.19 and table 4.10) IVS9-9T→A (Børresen *et al.*, 1995; Farrington *et al.*, 1998), IVS10+6T→ C (Farrington *et al.*, 1998) and IVS10+12A→G (Wijnen *et al.*, 1994; Wahlberg *et al.*, 1997).

4.4.4 Unclassified variants in hMSH2

Five novel missense and intronic variants of unknown significance (Table 4.11 and Figure 4.20) were identified in hMSH2. Four variants occurred in intronic regions, of which three are point mutations, and one is a one base pair deletion. The variant in exon 16 is a silent mutation. All four point mutations are transitions.

Two variants were identified in intron 10, IVS10-9G→A and IVS10-10C→T. The IVS10-9G→A variant was detected in the tumour and germline of three black patients, CRC50, CRC85 and CRC145. The IVS10-10C→T variant was identified in the tumour and germline of one patient, CRC94. Both of these variants are suspected to be polymorphisms, but to confirm this, control chromosomes need to be analysed.

Two variants were identified in intron 12, IVS12-38G→A and IVS12+8delA. The guanine to adenine transition was identified in the tumour of CRC148. A run of six

adenines occurs at position +3 to +8 in intron 12. The tumour of patient CRC60 displayed the one base pair (adenine) deletion at 2005+8. It is not known which A is deleted, and whether it will influence correct splicing of the exon. Two patients (CRC145 and CRC153) displayed a heterozygous samesense point mutation, cytosine to thymine, at nucleotide 2766. Although it occurs in the carboxy-terminal region of the interaction region of hMSH2 with hMSH3 or hMSH6, it is not expected to be pathogenic.

Table 4.11 Unclassified variants identified in hMSH2

Patient	Exon / IVS	Codon	Nucleotide Change cDNA sequence	Designation	Consequence
CRC50, CRC85, CRC145	IVS 10	-	1662 - 9 G→A	IVS10-9G→A	-
CRC94	IVS10	-	1662 - 10 C→T	IVS10-10C→T	-
CRC148	IVS12	-	2006 - 38 G→A	IVS12-38G→A	-
CRC 60	IVS12	-	2005 + 8 del A	IVS12+8delA	-
CRC145, CRC153	16	922	nucl 2766 T→C	F922F	Phe→Phe

Recommended nomenclature system (Antonarakis *et al.*, 1998); del = deletion; IVS = intervening sequence

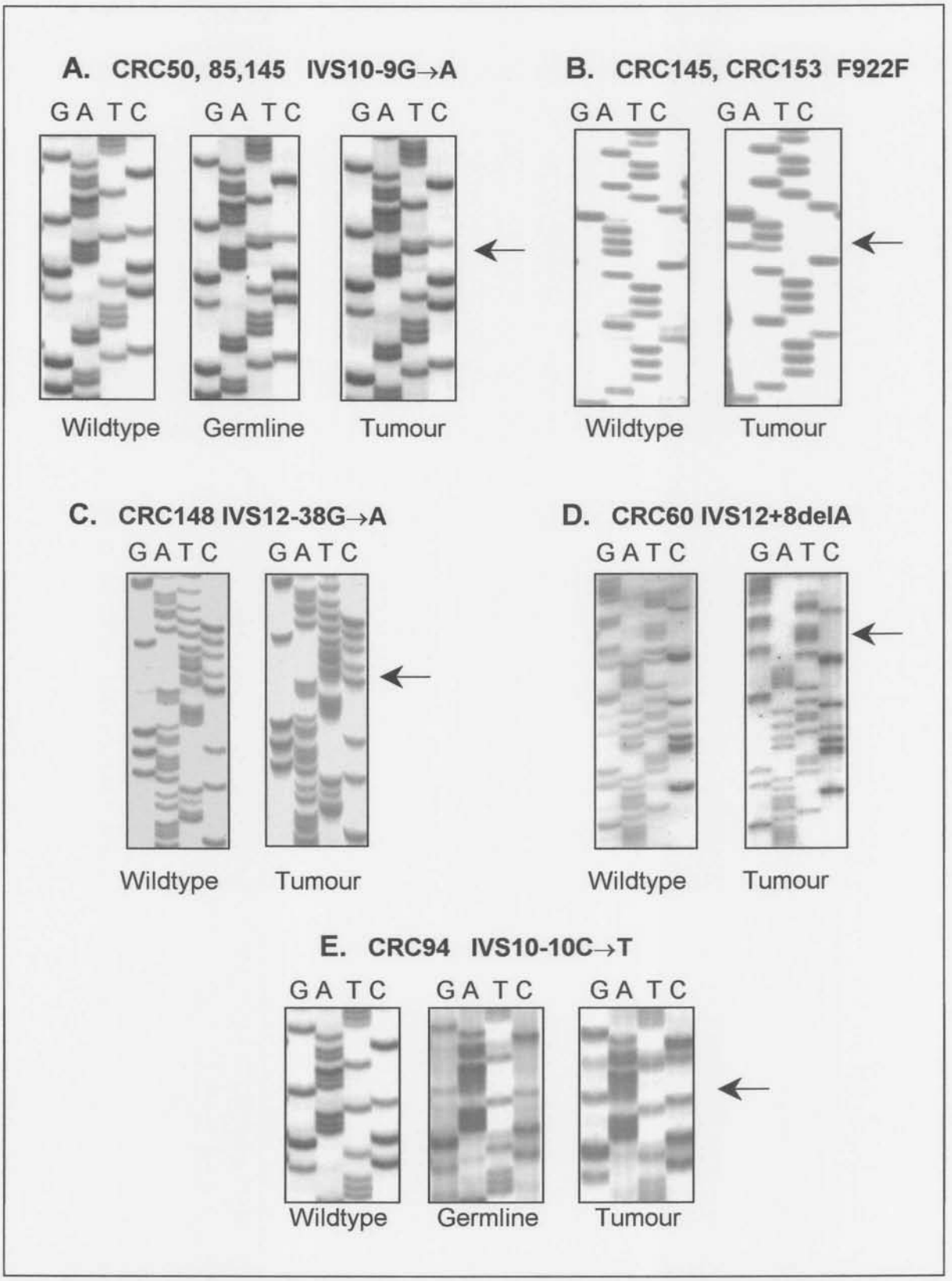


Figure 4.20 Sequence analysis of unclassified variants identified in hMSH2
The arrows indicate the sequence alterations in the sense strand (D) and antisense strands (A, B, C and E).

4.5 TUMOURS WITH COMPOUND HETEROZYGOSITY AND DOUBLE HETEROZYGOSITY

Five of the fifteen patients (Table 4.12) with pathogenic mutations presented with two different mutations each. Tumour CRC77 is a true compound heterozygote presenting with two different mutations in the hMLH1 gene, one of which is a germline mutation. These mutations may lead to a homozygous inactivation of the hMLH1 gene in this tumour. Herfarth *et al.* (1997) reported two tumours with two hMLH1 mutations each. One of the tumours was found to be hemizygous for the second mutation thus enabling the authors to prove that both alleles were affected in this tumour. A compound heterozygote for hMSH2 has also been reported (Børresen *et al.*, 1995).

Table 4.12 Patients with two disease-causing mutations

Patient	Gene	Designation
CRC77	hMLH1 hMLH1	813delC IVS5+3insT *
CRW60	hMLH1 hMSH2	731del4 * K471N ^{P}
CRC19	hMSH2 hMLH1	IVS5+1G→A G67R
CRC50	hMSH2 hMLH1	261del5 L676H ^{P}
CRW58	hMLH1 hMSH2	V716M ^{*P} R383X

*= germline mutation
^{P}= putative pathogenic mutation

Four tumours (CRW58, CRW60, CRC19 and CRC50) each have mutations in hMLH1 as well as hMSH2 and are therefore double heterozygotes (Table 4.12). In tumour CRW60 a pathogenic germline change in hMLH1 (frameshift) and a possible pathogenic somatic missense mutation in hMSH2 was identified. This is to a certain extent similar to CRC patients described by Möslein *et al.* (1996) and Liu *et al.* (1998). Two patients each had a germline pathogenic mutation (a frameshift in either hMLH1 or hMSH2) and a second germline missense mutation (in either hMLH1 or hMSH2) that could be pathogenic (Möslein *et al.*, 1996). Liu *et al.* (1998) reported a HNPCC tumour with a germline missense mutation (hMLH1) and a germline frameshift mutation (hMSH2). The frameshift mutation segregates with the disease. The tumour was found to express the hMLH1 protein but not the hMSH2 protein (Salahshor *et al.*, 2001). However, complete correlation between

hMLH1 germline mutations and the lack of hMLH1 protein was not detected and the consequence of the missense mutation is therefore not clear. In the case of CRW58, a putative pathogenic germline missense mutation (hMLH1) and a pathogenic somatic mutation (hMSH2) were identified.

In contrast to CRW60 and the patients described by Möslein *et al.* (1996) and Liu et al (1998), both the mutations in the tumours of the other two patients (CRC19 and CRC50) are somatically derived. Furthermore the mutations in CRC19 are both pathogenic, one removes the invariant G causing a splice site alteration and the other a missense mutations that has been shown to be pathogenic (Shimodaira *et al.*, 1998). In the case of CRC50 a disease causing frameshift mutation in combination with a possible pathogenic missense mutation was identified. Exactly how double heterozygotes contribute to the occurrence of the disease is not clear at this stage.

4.6 MSI-H PATIENTS WITH NO PATHOGENIC MUTATIONS

Fifty-five percent (12/22) of the black patients' tumours and 70% (7/10) of the white patients' tumours do not exhibit any pathogenic mutations in hMLH1 or hMSH2. Three of these tumours (CRC57, CRW1 and CRW56) showed microsatellite instability only in mononucleotide repeats. Since inactivating mutations in hMSH6 lead to instability of mononucleotide repeats (Papadopoulos *et al.*, 1995), it can be expected that such mutations are the cause of the exclusively mononucleotide instability in the tumours from CRC57, CRW1 and CRW56. Another explanation for the failure to detect pathogenic mutations is the sensitivity of the mutation detection method used, i.e. exon-by-exon PCR SSCP which has a sensitivity of 80% (Hayashi and Yandell, 1993). It is thus possible that some mutations have not been detected. Methylation of the hMLH1 promoter region is another inactivating mechanism that has been reported to occur (Kane *et al.*, 1997; Cunningham *et al.*, 1998; Herman *et al.*, 1998; Kuismanen *et al.*, 1999). This has not been investigated in the present study and may explain some of the cases. Patients without mutations in hMLH1 or hMSH2 or expected mutations in hMSH6 could possibly harbour mutations in other DNA mismatch repair genes that have already been identified or may be identified in future.

4.7 ANALYSIS OF COMBINED MUTATIONAL DATA

Due to the failure of PCR amplification only 32 of the 40 MSI-H tumours could be fully screened for mutations in hMLH1 and hMSH2. In total 16 pathogenic mutations were identified in 14 of the 32 (44%) MSI-H tumours. The 14 tumours with pathogenic mutations include 10 (10/22; 45%) from black and 4 (4/10; 40%) from Caucasian patients. Five (5/32; 16%) of these mutations are germline in origin, four (4/22; 18%) of which were detected in tumours from black patients. Although the prevalence of germline mutations in the black patients is 1,8 times higher than that of the Caucasian patients it is considered not significant (two-tailed p value =0,2077; Fisher's Exact test). Somatic mutations were detected in 6 (27 %) tumours from black patients and 3 (30 %) tumours from Caucasian patients (two-tailed p value =0,3274; Fisher's Exact test).

Comparison of these data to the results obtained in two other large studies (> 100 patients) that also investigated CRC tumours for microsatellite instability, hMSH2 and hMLH1 gene alterations show interesting similarities and differences. From Table 4.13 it is clear that the frequency of germline mutations of the patients presenting with MSI-H tumours in the Caucasian patients (1/10; 10%) and the black patients (5/22; ~23%) in the present study is similar (two-tailed p values range from 0.2502 to 1.000; Fisher's Exact test) to that of the reported studies (0 – 17%), with the black patients' being at the top end of the range. However, the frequency of pathogenic somatic mutations in both the black (6/22; ~ 27%) and Caucasian (3/10; 30%) patients is more than double that of the reported studies. Even though the frequencies of somatic mutations between the two South African populations do not differ significantly (two-tailed p value =0,3274; Fisher's Exact test), significant differences were observed between the black South African patients and the patients of Kuismanen *et al.* (2000) (two-tailed p value =0,0024; Fisher's Exact test) and Potočnik *et al.* (2001) (two-tailed p value =0,0125; Fisher's Exact test) which was not so for the Caucasian South African patients (two-tailed p value =0,1048 and 0,3640 respectively; Fisher's Exact test).

Thus HNPCC was diagnosed in ~ 0,93% of Caucasians and 3,85% of black patients via germline mutations in the present study of 107 Caucasian and 104 black CRC patients (p=0,349; χ^2 –test). The finding that 3,85% of the black patients have

HNPCC is 1,9 times higher than the maximum (1,96 %) reported by Aaltonen *et al.* (1998). The difference in frequency of HNPCC between the two population groups is also reflected by the average age at diagnosis, with the Caucasian patients' age being 65 years and that of the black patients' being 13 years younger at 52 years. The proportion of black patients younger than 45 years is also much more (33 %) than that of the Caucasian patients (10 %). This is what can be expected since the reported average age at diagnosis of HNPCC related CRC is 40 to 45 years (Marra and Boland, 1995).

Table 4.13 Comparison of data from previously reported studies with the present study

Study	Total patients	MSI-H	# Mutations	Type of pathogenic Mutation
1a Aaltonen *et al.*, 1998 1b Kuismanen *et al.*, 2000	509 -	63 (46)	10 4	10 germline 4 somatic
2 Potočnik *et al.*, 2001	345	35	10	6 germline 4 somatic
Present study (Black)	109(104)*	27(22)*	10	4 germline 6 somatic
Present study (Caucasian)	110(107)*	13(10)*	4	1 germline 3 somatic

Aaltonen *et al.* (1998) screened for germline mutations and Kuismanen et al (2000) for somatic mutations of the same series of patients. *Complete mutation screening could only be carried out for 32 of the 40 MSI-H tumours.

The 0,93% and 3,85% of HNPCC diagnosed represent the absolute minimum in this series of patients. Mutations can be missed because of limitations of the screening method used and because the other genes implicated in HNPCC, hPMS1, hPMS2 and hMSH6, have not been analysed. The finding that the frequency of HNPCC in black patients is four times higher than that of Caucasian patients is consistent with the notion that penetrance of HNPCC cancer is independent of environmental factors. According to Lynch and de la Chapelle (1999) if penetrance in HNPCC is not influenced by environmental factors then the proportion of HNPCC colorectal cancers will be much higher in a low incidence population compared to a high incidence population. In South Africa there is a 10-fold lower incidence of CRC in black compared to Caucasian populations. Therefore if the penetrance of HNPCC cancer is as high in blacks as in Caucasians (i.e. very low or no impact from environmental

factors) then the proportion of HNPCC colorectal cancers will be much higher in black than in Caucasian South Africans, which seems to be the case here.

Using this data one can also estimate the frequency of HNPCC in the general population. According to the National Cancer Registry (Sitas et al, 1998), the lifetime risk for CRC in Caucasians is 2,56% (2,94% male; 2,27% female) and for black South Africans is 0,22% (0,25% male; 0,2% female). Thus the risk of HNPCC type cancer in South African Caucasians is 0,93% of 2,56%, or ~ 1 in 4200; and for black South Africans it is 3,85% of 0,22%, or ~ 1 in 11800 subjects.

CHAPTER 5

CONCLUSION

Colorectal cancer (CRC) is one of the most common forms of neoplasia in Western populations (Dunlop, 1992) but is uncommon in sub-Saharan Africa (Soliman, *et al.*, 1998; Williams and Prince, 1975; Haghighi *et al.*, 1977). In developing countries such as South Africa, the differences in lifestyles and environment, as well as the diverse patterns of cancer that exist between the various population groups provides an ideal opportunity to study the pathogenesis of colorectal cancer (Magrath and Litvak, 1993). This study represents the first molecular genetic investigation into the role of DNA mismatch repair genes in putative sporadic cases of CRC from black and Caucasian South African patients.

To determine the possible involvement of the DNA mismatch repair genes, the presence of microsatellite instability was investigated in 109 black and 110 Caucasian patients. Twenty-seven tumours from black patients (24,8%) and 13 (11,8%) tumours from Caucasian patients were found to have MSI-H (p=0,021; χ^2 –test). The proportion of tumours with MSI-H from black patients attending Chris Hani Baragwanath Hospital (12,2%; 5/41) and those from Caucasian patients is in accordance to published results on sporadic tumours. However the finding that 32,4% (22/68) of black patients attending Kalafong and Pretoria Academic Hospitals, have tumours with MSI-H is much higher than is commonly reported in Western populations and is significantly higher than that of the Caucasian patients (p = 0.002; χ^2 –test). The patients, who attended CHB, live mainly in urban Johannesburg/Soweto, while those seen at Kalafong and Pretoria Academic Hospitals are mostly from peri-urban and rural areas. While determining the microsatellite status of the tumours, it was noticed that the two mononucleotide markers, BAT25 and BAT26, described as quasimonomorphic (Zhou *et al.*, 1997; Hoang *et al.*, 1997; Aaltonen *et al.*, 1998), were actually polymorphic in black patients. This finding was recently confirmed by two other studies (Pyatt et al, 1999; Samowitz et al, 1999).

Failure of PCR amplification, owing to the absence of high quality tissue in this retrospective study, allowed 32 of the 40 MSI-H tumours to be fully screened for

mutations in hMLH1 and hMSH2. In total 16 pathogenic mutations were identified, 10 of which are novel. These mutations occurred in 14 of the 32 (44%) tumours, one tumour was found to be a compound heterozygote and four tumours were double heterozygotes. Thus 45% of MSI-H tumours from black patients and 40% of tumours from Caucasian patients harbour mutations in either hMSH2 or hMLH1.

No pathogenic mutations were identified in 19/32 MSI-H tumours. Three of these tumours are expected to harbour mutations in the hMSH6 gene as microsatellite instability in these tumours was only found in mononucleotide markers. Methylation of the hMLH1 promoter region is another mechanism of inactivation of this gene. Some of the tumours without mutations may therefore harbour either methylation of the hMLH1 promoter region or mutations in other DNA mismatch repair genes that have already been identified or may be identified in future. The sensitivity of the mutation detection method is also less than 100% and may explain the lack of mutations in some of the tumours.

Five of the patients presenting with MSI-H tumours have pathogenic germline mutations. Four are from black patients (4/22 = 18%) and one from a Caucasian patient (1/10 = 10%). This frequency of germline mutations is similar to that of the reported studies (0-17%; two-tailed p values range from 0,2502 to 1,000; Fisher's Exact test) (Aaltonen *et al.*, 1998; Potočnik *et al.*, 2001). The frequency of pathogenic somatic mutations in both the black (6/22; ~27%) and Caucasian (3/10; 30%) patients is at least double that of the reported studies (8-11%) (Kuismanen et al, 2000; Potočnik *et al.*, 2001). However, only the frequency of somatic mutations in the black patients differ significantly from that of the reported studies (two-tailed p value =0,0024 and 0,0125; Fisher's Exact test). The missense mutation in hMSH2 (codon 127), listed in the ICG-HNPCC database as a pathogenic mutation (in a Nigerian family) and also detected in the current study, was shown to be a polymorphic change exclusive to black Africans.

HNPCC was diagnosed in ~0,93% (1/107) of Caucasian and 3,85% (4/104) of black patients via germline mutations (p=0,349; χ^2 –test). The finding that the frequency of HNPCC in black patients is four times higher than that of Caucasian patients is consistent with the notion that penetrance of HNPCC cancer is independent of environmental factors, according to Lynch and de la Chapelle (1999).

In summary, the observed microsatellite instability and mutations in the hMLH1 and hMSH2 genes thus clearly implicate the DNA mismatch repair genes in the pathogenesis of colorectal cancers in both black and Caucasian South African patients. This study represents the first investigation of DNA mismatch repair genes in tumours from both population groups. It is also the first report of black South Africans with HNPCC.

REFERENCES

Aaltonen LA, Peltomäki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, *et al.* Clues to the Pathogenesis of Familial Colorectal Cancer. *Science* 1993; **260**:812-816.

Aaltonen LA, Peltomäki P, Mecklin JP, Järvinen H, Jass JR, Green JS, *et al.* Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. *Cancer Res* 1994; **54**:1645-16458.

Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomaki P, *et al.* Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med* 1998;**338**(21):1481-1487.

Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Capelle, *et al.* Cancer risk in mutation carriers of DNA-mismatch-repair genes. *Int J Cancer* 1999;**81**:214-218.

Acharya S, Wilson T, Gradia S, Kane MF, Guerrette S, Marsischky GT, *et al.* hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc Natl Acad Sci USA* 1996; **93**:13629-13634.

Alani E. The *Saccharomyces cerevisiae* Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs. *Mol Cell Biol* 1996;**16**(10):5604-5615.

Allen DJ, Makhov A, Grilley M, Taylor J, Thresher R, Modrich P, *et al.* MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J* 1997;**16**(14):4467-4476.

Angelo N. Colorectal carcinoma - a new threat to Black patients? A retrospective analysis of colorectal carcinoma received by the Institute of Pathology, University of Pretoria [Dissertation]. Pretoria: University of Pretoria, 2000.

Antequera F, Bird A. Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci USA* 1993;**90**:11995-11999.

Antonarakis SE and the Nomenclature Working group. Recommendations for a nomenclature system for human gene mutations. *Human Mutations* 1998;**11**:1-3.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, *et al.* In Current Protocols on Molecular Biology. John Wiley & Sons, Inc 1995;**1**:3.10.1-3.10.5.

Bader S, Walker M, Hendrich B, Bird A, Bird C, Hooper M, *et al.* Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* 1999;**18**:8044-8047.

Ban C, Yang W. Crystal structure and ATPase activity of MutL: Implications for DNA repair and mutagenesis. *Cell* 1998;**95**:541-552.

Baylin Sb, Makos M, Wu JJ, Yen RW, de Bustros A, Vertino P, *et al.* Abnormal patterns of DNA methylation in human neoplasia – potential consequences for tumor progression. *Cancer Cells* 1991;**3**:383.

Blackwell LJ, Martik D, Bjornson KP, Bjornson ES, Modrich P. Nucleotide-promoted release of hMutSα from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism. *J Biol Chem* 1998;**273**(48):32055-32062.

Bocker T, Schlegel J, Kullmann F, Stumm G, Zirngibl H, Epplen JT, *et al.* Genomic instability in colorectal carcinomas: comparison of different evaluation methods and their biological significance. *J Path* 1996;**179**:15-19.

Boland CR, Troncale FJ. Familial colonic cancer without antecedent polyposis. *Ann Intern Med* 1984;**100**:700-701.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW *et al.* A national cancer institute workshop on microsatellite instability for cancer detection and familial predisposition: Development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 1998;**58**:5248-5257.

Børresen A-L, Lothe RA, Meling GI, Lystad S, Morrison P, Lipford J, *et al.* Somatic mutations in the hMSH2 gene in microsatellite unstable colorectal carcinoma. *Hum Mol Genet* 1995;**4**:2065.

Boyle P, Zaridze DG, Smans M. Descriptive epidemiology of colorectal cancer. *Int J Cancer* 1985;**36**:9-18.

Bremner CG, Ackerman LV. Polyps and carcinoma of the large bowel in the South African Bantu. *Cancer* 1970;**5**:991-999.

Broders AC. The grading of carcinoma. *Minnesota Med* 1925;**8**:726.

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, lescoe MK, *et al.* Mutation in the DNA mismatch repair gene homolog hMLH1 is associated with hereditary nonpolyposis colorectal cancer. *Nature* 1994;**368**:258-261.

Buerstedde J-M, Alday P, Torhorst J, Weber W, Scott R. Detection of new mutations in six out of 10 Swiss HNPCC families by genomic sequencing of the hMSH2 and hMLH1 genes. *J Med Genet* 1995;**32**:909-912.

Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med* 1990;**113**:779-788.

Chi N-W, Kolodner RD. Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches. *J Biol Chem* 1994;**269**(47):29984-29994.

Chung DC. The genetic basis of colorectal cancer: Insights into critical pathways of tumorigenesis. *Gastroenterol* 2000;**119**:854-865.

Coleman WB, Tsongalis GJ. Multiple mechanisms account for genomic instability and molecular mutation in neoplastic transformation. *Clin Chem* 1995;**41**(5):644-657.

Cooper DN, Krawczak M. The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. *Hum Genet* 1996;**85**:55-74.

Cotton RG, Scriver CR. Proof of "disease causing" mutation. *Hum Mutat* 1998;**12**:1-3.

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

Curran B, Lenehan K, Mulcahy H, Tighe O, Bennett MA, Kay EW, *et al.* Replication error phenotype, clinicopathological variables, and patient outcome in Dukes' B stage II (T3, No, M0) colorectal cancer. *Gut* 2000;**46**:200-204.

Degiannis E, Sliwa K, Levy R, Hale MJ, Saadia R. Clinicopathological trends in colorectal carcinoma in a black South African population. *Trop Gastroenterol* 1995;**16**(4):55-61.

Desai DA, Lockman JC, Chadwick RB, Gao X, Percesepe A, Evans DGR, *et al.* Recurrent germline mutation in MSH2 arises frequently de novo. *J Med Genet* 2000;**37**:646-652.

De Wind N, Dekker M, Berns A, Radman M, te Riele H. Inactivation of the mouse MSh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell* 1995;**82**:321-330.

De Wind N, Dekker M, Van Rossum A, Van der Valk M, Te Riele H. Mouse models for hereditary nonpolyposis colorectal cancer. *Cancer Res* 1998;**58**:248-255.

Drotschmann K, Clark AB, Kunkel TA. Mutator phenotypes of common polymorphisms and missense mutations in MSH2. *Curr Biol* 1999;**9**:907-910.

Drummond JT, Li G-M, Longley MJ, Modrich P. Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 1995;**268**:1909-1912.

Dunlop MG. Screening for large bowel neoplasms in individuals with a family history of colorectal cancer. *Br J Surg* 1992;**79**:488-494.

Dunlop MG, Farrington SM, Carothers AD, Wyllie AH, Sharp L, Burn J, *et al.* Cancer risk associated with germline DNA mismatch repair gene mutations. *Hum Mol Genet* 1997;**6**(1):105-110.

Edelmann W, Cohen PE, Kane M, Lau K, Morrow B, Bennett S, *et al.* Meiotic pachytene arrest in MLH1-deficient mice. *Cell* 1996;**85**:1125-1134.

Edelmann W, Yang K, Kuraguchi M, Heyer J, Lia M, Kneitz B, *et al.* Tumourigenesis in MLH1 and MLH1/APC1638N mutant mice. *Cancer Res* 1999;**59**:1301-1307.

Eppert K, Scherer SW, Ozcelik H, Pirone R, Hoodless P, Kim H, *et al.* Madr2 maps to 18q21 and encodes a TGFβ-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* 1996;**86**:543-552.

Farrington SM, Lin-Goerke J, Ling J, Wang Y, Burzak JD, Dunlop MG. Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. *Am J Hum Genet* 1998;**63**:749-759.

Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;**61**:759-767.

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

Flores-Rozas H, Kolodner RD. The *Saccharomyces cerevisiae MLH3* gene functions in MSH3-dependent suppression of frameshift mutations. *Proc Natl Acad Sci USA* 1998;**95**:12404-12409.

Flood DM, Weiss NS, Cook LS, Emerson JC, Schwartz SM, Potter JD. Colorectal cancer incidence in Asian migrants to the United States and their descendants. *Cancer Causes Control* 2000;**11**:403-411.

Froggatt NJ, Joyce JA, Davies R, Evans DGR, Ponder BAJ, Barton DE, *et al.* A frequent hMSH2 mutation in hereditary non-polyposis colon cancer syndrome. *Lancet* 1995;**345**:727.

Froggatt NJ, Joyce JA, Evans DGR, Lunt PW, Koch DJ, Ponder BAJ, *et al.* MSH2 sequence variations and inherited colorectal cancer susceptibility. *Eur J Cancer* 1996;**32A**:178.

Froggatt NJ, Green J, Brassett C, Evans DGR, Bishop DT, Kolodner R, *et al.* A common MSH2 mutation in English and North American Families: origin, phenotypic expression, and sex specific differences in colorectal cancer. *J Med Genet* 1999;**36**:97-102.

Goldberg PA, Madden MV, Harocopos C, Felix R, Westbrook C, Ramesar RS. 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* 1998;**41**:1250-1255.

Groden J, Thilveris A, Samowits W, Carlson M, Gelbert L, Albertson H, *et al.* Identification and characterization of the familial adenomatious polyposis coli gene. *Cell* 1991;**66**:589-600.

Gryfe R, Swallow C, Bapat B, Redston M, Gallinger S, Couture J. Molecular biology of colorectal cancer. *Curr Probl Cancer* 1997;**21**(5):233-300.

Gu L, Hong Y, McCulloch S, Watanabe H, Li G-M. ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. *Nucl Acids Res* 1998;**26**(5):1173-1178.

Guanti G, Resta N, Simone C, Cariola F, Demma I, Fiorente P, *et al.* Involvement of PTEN mutations in the genetic pathways of colorectal carcinogenesis. *Hum Mol Genet* 2000;**9**(2):283-287.

Guerrette S, Wilson T, Gradia S, Fishel R. Interactions of Human hMSH2 with hMSH3 and hMSH2 with hMSH6: Examination of mutations found in Hereditary Nonpolyposis Colorectal cancer. *Mol Cell Biol* 1998;**18**(11): 6616-6623.

Guerette S, Acharya S, Fishel R. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J Biol Chem* 1999;**274**(10):6336-6341.

Habraken Y, Sung P, Prakash L, Prakash S. Enhancement of MSH2-MSH3-mediated mismatch recognition by the yeast MLH1-PMS1 complex. *Curr Biol* 1997;**7**:790-793.

Habraken Y, Sung P, Prakash L, Prakash S. ATP-dependant assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes. *J Biol Chem* 1998;**273**(16):9837-9841.

Haghighi P, Nasr K, Mohallatee EA, Ghasseni H, Sadri S, Nabizadeh I, *et al.* Colorectal polyps and carcinoma in Southern Iran. *Cancer* 1977;**39**:274-278.

Han H-J, Maruyama M, Baba S, Park J-G, Nakamura Y. Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary nonpolyposis colorectal cancer (HNPCC). *Hum Mol Genet* 1995;**4**(2):237-242.

Harfe BD, Minesinger BK, Jinks-Robertson S. Discrete *in vivo* roles for the MutL homologs MLH2p and MLH3p in the removal of frameshift intermediates in budding yeast. *Curr Biol* 2000;**10**:145-148.

Hayashi K, Yandell DW. How sensitive is PCR-SSCP? *Hum Mutat* 1993;**2**:338-346.

Her C, Doggett NA. Cloning, structural characterization, and chromosomal localization of the human ortologue of *Saccharomyces cerevisiae* MSH5 gene. *Genomics* 1998;**52**:50-61.

Herfarth KKF, Kodner IJ, Whelan AJ, Ivanovich JL, Bracamontes JR, Wells SA Jr, *et al.* Mutations in MLH1 are more frequent than in MSH2 in sporadic colorectal cancers with microsatellite instability. *Genes Chromosom Cancer* 1997;**18**:42-49.

Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa J-PJ, *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci USA* 1998;**95**:6870-6875.

Hoang J-M, Cottu PH, Thuille B, Salmon RJ, Thomas G, Hamelin R. BAT-26, and indicator of the replication error phenotype in colorectal cancers and cell lines. *Cancer Res* 1997;**57**:300-303.

Hollingsworth NM, Ponte L, Halsey C. MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. *Genes Dev* 1995;**9**:1728-1739.

Houlston RS, Collins A, Slack J, Morten NE. Dominant genes for colorectal cancer are not rare. *Ann Hum Genet* 1992;**56**:99-103.

Human Gene Mutation Database, MLH1. Available at http://archive.uwcm.ac.uk/uwcm/mg/search/249617.html. Accessed: May 2001.

Human Gene Mutation Database, MSH2. Available at http://archive.uwcm.ac.uk/uwcm/mg/search/203983.html. Accessed: May 2001.

Human Gene Mutation Database, MSH6. Available at http://archive.uwcm.ac.uk/uwcm/mg/search/632803.html. Accessed: May 2001.

Human Gene Mutation Database, PMS1. Available at http://archive.uwcm.ac.uk/uwcm/mg/search/386403.html. Accessed: May 2001.

Human Gene Mutation Database, PMS2. Available at http://archive.uwcm.ac.uk/uwcm/mg/search/386406.html. Accessed: May 2001.

Hutter P, Couturier A, Membrez V, Joris F, Sappino A-P,Chappuis PO. Excess of hMLH1 germline mutations in Swiss families with hereditary non-polyposis colorectal cancer. *Int J Cancer* 1998;**78**:680-684.

Iaccarino I, Palombo F, Drummond J, Totty NF, Hsuan JJ, Modrich P, *et al.* MSH6, a *Saccharomyces cerevisiae* protein that binds to mismatches as a heterodimer with MSH2. *Curr Biol* 1996;**6**: 484-486.

Iaccarino I, Marra G, Palombo F, Jiricny J. hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSα. *EMBO J* 1998;**17**(9):2677-2686.

ICG-HNPCC (International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer) Mutation database. Available at http://www.nfdht.nl/database/mdbchoice.htm. Accessed: May, 2001.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 1993;**363**:558-561.

Ishimaru G, Adachi J, Shiseki M, Yamaguchi N, Muto T, Yokota J. Microsatellite instability in primary and metastatic colorectal cancers. *Int J Cancer* 1995;**64**(3): 153-157.

Iwahashi Y, Ito E, Yanagisawa Y, Akiyama Y, Yuasa Y, Onodera T, *et al.* Promoter analysis of the human mismatch repair gene hMSH2. *Gene* 1998;**213**:141-147.

Jaskiewics K, Lancaster E, Banach L, Karmolinski A. Proliferative activity of normal and neoplastic colonic mucosa in population groups with high and low risk for colorectal carcinoma. *Anticancer Res* 1998;**18**(6B):4641-4644.

Johnson RE, Kovvali GK, Prakash L, Prakash S. Requirement of the yeast MSH3 and MSH6 genes for MSH2-dependent genomic stability. *J Biol Chem.* 1996;**71**(13):7285-7288.

Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, *et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumours and mismatch repair-defective human tumour cell lines. *Cancer Res* 1997;**57**:808-811.

Kim H, Jen J, Vogelstein B, Hamilton SR. Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *Am J Pathol* 1994;**145**(1):148-156.

Kinzler KW, Nilbert MC, Su L-K, Vogelstein B, Bryan TM, Levy DM, *et al.* Identification of FAP locus genes from chromosome 5q21. *Science* 1991;**253**:661-664.

Kinzler KW, Vogelstein B. Gatekeepers and caretakers. *Nature* 1997;**386**:761-763.

Kinzler KW, Vogelstein B. Landscaping the cancer terrain. *Science* 1998;**280**:1036-1037.

Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971;**68**:820-823.

Kolodner RD, personal communication.

Kolodner RD, Hall NR, Lipford J, Kane MF, Rao MR, Morrison PT, *et al.* Structure of the human MSH2 locus and analysis of two Muir-Torre kindreds for MSH2 mutations. *Genomics* 1994;**24**(3):516-526.

Kolodner RD, Hall NR, Lipford J, Kane MF, Morrison PT, Finan PJ, *et al.* Structure of the human MLH1 locus and analysis of a large hereditary nonpolyposis colorectal carcinoma kindred for mlh1 mutations. *Cancer Res* 1995;**55**(2):242-248.

Kowalski LD, Mutch DG, Herzog TJ, Rader JS, Goodfellow PJ. Mutational analysis of MLH1 and MSH2 in 25 prospectively-acquired RER+ endometrial cancers. *Genes Chromosom Cancer* 1997;*18*:219-227.

Kronborg O, Fenger C. Clinical evidence for the adenoma-carcinoma sequence. *Eur J Cancer Prev* 1999;**8**:S73-S86.

Kuismanen SA, Holmberg MT, Salovaara R, Schweizer P, Aaltonen LA, De la Chapelle A, *et al.* Epigenetic phenotypes distinguish microsatellite-stable and -unstable colorectal cancers. *Proc Natl Acad Sci USA* 1999;**96**(22):12661-12666.

Kuismanen SA, Holmberg MT, Salovaara R, De la Chapelle A, Peltomäki P. Genetic and epigenetic modification of MLH1 accounts for a major share of microsatellite-unstable colorectal cancers. *Am J Pathol* 2000;**156**(5):1773-1779.

Lamers MH, Perrakis A, Enzlin JH, Winterwerp HHK, de Wind N, Sixma TK. The crystal structure of DNA mismatch repair protein MutS binding to a G:T mismatch. *Nature* 2000;**407**:711-717.

Larsen F, Gundersen G, Lopez R, Prydz H. CpG islands as gene markers in the human genome. *Genomics* 1992;**13**:1095-1107.

Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons, *et al.* Mutations of a must Homolog in Hereditary Nonpolyposis Colorectal Cancer. *Cell* 1993;**75**:1215-1225.

Leach FS, Polyak K, Burrell M, Johnson KA, Hill D, Dunlop MG, *et al.* Expression of the human mismatch repair gene hMSH2 in normal and neoplastic tissues. *Cancer Res* 1996;**56**:235-240.

Levinson G, Gutman GA. High Frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res* 1987;**15**:5323-5338.

Li G-M, Modrich P. Restoration of mismatch repair to nuclear extracts of H6 colorectal tumour cells by a heterodimer of human MutL homologues. *Proc Natl Acad Sci USA* 1995;**92**:1950-1954.

Lin KM, Shashidharan M, Ternent CA, Thorson AG, Blatchford GJ, Christensen MA, *et al.* Colorectal and extracolonic cancer variations in MLH1/MSH2 hereditary nonpolyposis colorectal cancer kindreds and the general population. *Dis Colon Rectum* 1998a;**41**:428-433.

Lin Y-L, Shivji MKK, Chen C, Kolodner R, Wood RD, Dutta A. The evolutionary conserved zinc finger motif in the largest subunit of human replication protein A is required for DNA replication and mismatch repair but not for nucleotide excision repair. *J Biol Chem* 1998b;**273**(3):1453-1461.

Lindblom A, Tannergård P, Werelius B, Nordenskjöld M. Genetic mapping of a second locus predisposing to hereditary nonpolyposis colorectal cancer. *Nature Genet* 1993;**5**(3):279-282.

Little J, Faivre J. Family history, metabolic gene polymorphism, diet and risk of colorectal cancer. *Eur J Cancer Prev* 1999;**8**:S61-S72.

Liu B, Parsons RE, Hamilton SR, Petersen GM, Lynch HT, Watson P, *et al.* hMSH2 mutations in hereditary nonpolyposis colorectal cancer kindreds. *Cancer Res* 1994;**54**:4590-4594.

Liu B, Nicolaides N, Markowitz S, Willson JKV, Parsons RE, Jen J, *et al.* Mismatch repair gene defects in sporadic cancers with microsatellite instability. *Nature Genet* 1995;**9**:48-55.

Liu T, Wahlberg S, Rubio C, Holmberg E, Grönberg H, Lindblom A. DGGE screening of mutations in mismatch repair genes (hMSH2 and hMLH1) in 34 Swedish families with colorectal cancer. *Clin Genet* 1998;**53**:131-135.

Liu W, Dong X, Mai M, Seelan RS, Taniguchi K, Krishnadath KK, *et al.* Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating β-catenin/TCF signalling. *Nat Genet* 2000;**26**:146-147.

Lipkin SM, Wang V, Jacoby R, Banerrjee-Basu S, Baxevanis AD, Lynch HT, *et al.* MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat Genet* 2000;**24**:27-35.

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

Lynch HT, Harris RE, Bardawil WA, Lynch PM, Guirqis HA, Swartz MJ, *et al.* Management of hereditary site-specific colon cancer. *Arch Surg* 1977;**112**:170-174.

Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* 1999;**36**:801-818.

Lynch HT, Kimberling WJ, Albano WA, Lynch JF, Biscone K, Schuelke GS, *et al.* Hereditary nonpolyposis colorectal cancer. Parts I and II. *Cancer* 1985;**56**:934-951.

Lynch HT, Lanspa SJ, Boman BM, Smyrk T, Watson P, Lynch JF, *et al.* Hereditary nonpolyposis colorectal cancer - Lynch syndromes I and II. *Gastroenterol Clin North Am* 1988;**17**(4):679-712.

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

Magrath I, Litvak J. Cancer in developing countries: Opportunity and challenge. *J Natl Cancer Inst* 1993;**85**:862-874.

Maliaka YK, Chudina AP, Belev NF, Alday P, Bochkov NP, Buerstedde J-M. CpG dinucleotides in the hMSH2 and hMLH1 genes are hotspots for HNPCC mutations. *Hum Genet* 1996;**97**:251-255.

Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, *et al.* Inactivation of the type II TGF-β receptor in colon cancer cells with microsatellite instability. *Science* 1995;**268**:1336-1338.

Markowitz AJ, Winawer SJ. Management of colorectal polyps. *CA Cancer J Clin* 1997;**47**(2):93-112.

Marra G, Boland CR. Hereditary Nonpolyposis colorectal cancer: the sydrome, the genes, and historical perspectives. *J Natl Cancer Inst* 1995;**87**:1114-1125.

Marsischky GT, Filosi N, Kane MF, Kolodner R. Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Dev* 1996;**10**:407-420.

Martin RH, Green J, Barclay L, Rademaker AW. Analysis of aneuploidy frequencies in sperm from patients with hereditary nonpolyposis colon cancer and an hMSH2 mutation. *Am J Hum Genet* 2000;**66**:1149-1152.

McCredie M, Williams S, Coates M. Cancer mortality in East and Southeast Asian migrants to New South Wales, Australia, 1975-1995. *Br J Cancer* 1999a;**79**:1277-1282.

McCredie M, Williams S, Coates M. Cancer mortality in migrants from the British Isles and continental Europe to New South Wales, Australia, 1975-1995. *Int J Cancer* 1999b;**83**:179-185.

Mecklin JP, Järvinen HJ. Tumor spectrum in cancer family syndrome (hereditary nonpolyposis colorectal cancer). *Cancer* 1991;**68**:1109-1112.

Miyaki M, Konishi M, Muraoka M, Kikuchi-Yanoshita R, Tanaka K, Iwama T *et al.* Germline mutations of hMSH2 and hMLH1 genes in Japanese families with hereditary nonpolyposis colorectal cancer (HNPCC): usefulness of DNA analysis for screening and diagnosing of HNPCC patients. *J Mol Med* 1995;**73**:515-520.

Miyaki M. Imprinting and colorectal cancer. *Nat Med* 1998;**14**(11):1236-1237.

Modrich P. Mechanisms and biological effects of mismatch repair. *Annu Rev Genet* 1991;**25**:229-253.

Moisio AL, Sistonen P, Weissenbach J, de la Chapelle A, Peltomäki P. Age and origin of two common MLH1 mutations predisposing to hereditary colon cancer. *Am J Hum Genet* 1996;**59**:1243-1251.

Moslein G, Tester DJ, Lindor NM, Honchel R, Cunningham JM, French AJ, *et al.* Microsatellite instability and mutation analysis of hMSH2 and hMLH1 in patients with sporadic, familial and hereditary colorectal cancer. *Hum Mol Genet* 1996;**5**:1245-1252.

Naaeder SB, Archampong EQ. Cancer of the colon and rectum in Ghana: a 5-year prospective study. *Br J Surg* 1994;**81**:456-459.

Nicolaides NC, Papadopoulos N, Liu B, Wei Y-F, Carter KC, Ruben SM, *et al.* Mutations of two PMS homologues in hereditary nonpolyposis colorectal cancer. *Nature* 1994;**371**:75-80.

Obmolova G, Ban C, Hsieh P, Yang W. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* 2000;**407**:703-710.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformation polymorphisms. *Proc Natl Acad Sci USA* 1989;**86**:2766-2770.

Palombo F, Gallinari P, Iaccarino I, Lettieri T, Hughes M, D'Arrigo A, *et al.* GTBP, a 160-Kilodalton protein essential for mismatch-binding activity in human cells. *Science* 1995;**268**:1912-1914.

Pang Q, Prolla TA, Liskay RM. Functional domains of the *Saccharomyces cerevisiae* MLH1p and PMS1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. *Mol Cell Biol* 1997;**17**(8):4465-4473.

Papadopoulos N, Nicolaides NC, Wei Y-F, Ruben SM, Carter KC, Rosen CA, *et al.* Mutation of a mutL homolog in hereditary nonpolyposis colorectal cancer. *Science* 1994;**263**:1625-1629.

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

Paquis-Flucklinger V, Santucci-Darmanin S, Paul R, Saunières A, Turc-Carel C, Desnuelle C. Cloning and expression analysis of a meiosis-specific mutS homolog: The human MSH4 gene. *Genomics* 1997;**44**:188-194.

Parsons R, Li G-M, Longley MJ, Modrich P, Liu B, Berk J, *et al.* Mismatch repair deficiency in phenotypically normal human cells. *Science* 1995;**268**:738-740.

Peltomäki PT, Aaltonen LA, Sistonen P, Pylkkanen L, Mecklin J-P, Järvinen H, *et al.* Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 1993;**260**:810-812.

Pensotti V, Radice P, Presciuttini S, Calistri D, Gazzoli I, Grimalt Perez AP, *et al.* Mean age of tumor onset in hereditary non-polyposis colorectal cancer (HNPCC) families correlates with the presence of mutations in DNA mismatch repair genes. *Genes Chrom and Cancer* 1997;**19**:135-142.

Pochart P, Woltering D, Hollingsworth NM. Conserved properties between functionally distinct MutS homologs in yeast. *J Biol Chem* 1997;**272**(48):30345-30349.

Potočnik U, Glavač D, Golouh R, Ravnik-Glavač M. Causes of microsatellite instability in colorectal tumors: implications for hereditary non-polyposis colorectal cancer screening. *Cancer Genet Cyogenet* 2000;**126**:85-96.

Prelich G, Tan CK, Kostura M, Mathews MB, So AG, Downey KM, *et al.* Functional identity of proliferating cell nuclear antigen and a DNA polymerase-delta auxiliary protein. *Nature* 1987;**326**:517-520.

Prolla TA, Pang Q, Alani E, Kolodner RD, Liskay RM. MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. *Science* 1994;**265**:10911093.

Pyatt R, Chadwick RB, Johnson CK, Adebamowo C, De la Chapelle A, Prior TW. Polymorphic variation at the BAT-25 and BAT-26 loci in individuals of African origin. *Am J Path* 1999;**155**:349-353.

Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, *et al.* Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* 1997;**275**:967-969.

Räschle M, Marra G, Nyström-Lahti M, Schär P, Jiricny J. Identification of hMutLβ, a heterodimer of hMLH1 and hPMS1. *J Biol Chem* 1999;**274**(45):32368-32375.

Reitmair AH, Redston M, Cai JC, Chuang TCY, Bjerknes M, Cheng H, *et al.* Spontaneous intestinal carcinomas and skin neoplasms in MSH2-defiecient mice. *Cancer Res* 1996;**56**:3842-3849.

Riccio A, Aaltonen LA, Godwin AK, Loukola A, Percesepe A, Salovaara R, *et al.* The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nat Genet* 1999;**23**:266-268.

Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Edwards BK (eds). SEER Cancer Statistics Review, 1993-1997, National Cancer Institute. Bethesda, MD, 2000.

Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan JPM, *et al.* A national cancer institute workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: Meeting highlights and Bethesda Guidelines. *J Natl Cancer Inst* 1997;**89**(23):1758-1762.

Ross-Macdonald P, Roeder GS. Mutation of a meiosis specific MutS homolog decreases crossing over but not mismatch correction. *Cell* 1994;**79**:1069-1080.

Rustgi AK. Hereditary gastrointestinal polyposis and nonpolyposis syndromes. *N Engl J Med* 1994;**331**(25):1694-1702.

Salahshor S, Koeble K, Rubio C, Lindblom A. Microsatellite instability and hMLH1 and hMSH2 expression analysis in familial and sporadic colorectal cancer. *Lab Invest* 2001;**81**:535-541.

Samowitz WS, Slattery ML, Potter JD, Leppert MF. BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. *Am J Pathol* 1999;**154**:1637-1641.

Sandler RS. Epidemiology and risk factors for colorectal cancer. *Gastroenterol Clin North Am* 1996;**25**(4):717-735.

Sarraf P, Mueller E, Smith WM, Wright HM, Kum JB, Aaltonen LA, *et al.* Loss-of-function mutations in PPARγ associated with human colon cancer. *Mol Cell* 1999;**3**:799-804.

Schmutte C, Marinescu RC, Sadoff MM, Guerette S, Overhauser J, Fishel R. Human exonuclease I interacts with the mismatch repair protein hMSH2. *Cancer Res* 1998;**58**:4537-4542.

Segal I, Cook SAR, Hamilton DG, Tom LO. Polyps and colorectal cancer in South African Blacks. *Gut* 1981;**22**:653-657.

Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987;**15**:7155-7174.

Shimodaira H, Filosi N, Shibata H, Suzuki T, Radice P, Kanamaru R *et al.* Functional analysis of human MLH1 mutations in *Saccharomyces cerevisiae*. *Nature Genet* 1998;**19**:384-389.

Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol Cell Biol* 1997;**17**:2851-2858.

Sitas F, Madhoo J, Wessie J. Cancer in South Africa, 1993-1995. National Cancer Registry of South Africa, South African Institute for Medical Research, Johannesburg, 1998.

Soliman AS, Bondy ML, Levin B, Hamza MR, Ismail K, Ismail S, *et al.* Colorectal cancer in Egyptian patients under 40 years of age. *Int J Cancer* 1997;**71**:26-30.

Soliman AS, Bondy ML, Guan Y, El-Badawi S, Mokhtar N, Bayomi S, *et al.* Reduced expression of mismatch repair genes in colorectal cancer patients in Egypt. *Int J Oncol* 1998;**12**:1315-1319.

Strand M, Prolla TA, Liskay RM, Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 1993;**365**:274-276.

Syngal S, Fox EA, Eng C, Kolodner RD, Garber JE. Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in MSH2 and MLH1. *J Med Genet* 2000;**37**:641-645.

Tannergård P, Lipford JR, Kolodner R, Frödin JE, Nordenskjöld, Lindblom A. Mutation screening in the hMLH1 gene in Swedish hereditary nonpolyposis colon cancer families. *Cancer Res* 1995;**55**:6092-6096.

Thibodeau SN, Bren G, Schaid D. Microsatellite instability in Cancer of the proximal colon. *Science* 1993;**260**:816-819.

Thibodeau SN, French AJ, Cunningham JM, Tester D, Burgart LJ, Roche PC, *et al.* Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. *Cancer Res* 1998;**58**:1713-1718.

Tishkoff DX, Amin NS, Viars CS, Arden KC, Kolodner R. Identification of a human gene encoding a homologue of *Saccharomyces cerevisiae* EXO1, and exonuclease implicated in mismatch repair and recombination. *Cancer Res* 1998;**58**:5027-5031.

Tomlinson IPM, Beck NE, Homfray T, Harocopos CJ, Bodmer WF. Germline HNPCC gene variants have little influence on the risk for sporadic colorectal cancer. *J Med Genet* 1997;**34**:39-42.

Tumour suppressor and oncogene directory available at: http://www.ncbi.nlm.nih.gov /CGAP/hTGI/tso/cgaptso.cgi. Accessed on 26/10/2000.

Turnbull RB, Kyle K, Watson FR, Spratt J. Cancer of the colon: the influence of the no-touch isolation technic on survival rates. *Ann Surg* 1967;**166**(3):420-427.

Umar A, Risinger JI, Glaab WE, Tindall KR, Barrett JC, Kunkel TA. Functional overlap in mismatch repair by human MSH3 and MSH6. Genetics 1998;**148**:1637-1646.

Vasen HFA, Mecklin J-P, Meera Khan P, Lynch HT. The international collaborative group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* 1991;**34**(5):424-425.

Vasen HFA, Wijnen JT, Menko FH, Kleibeuker JH, Taal BG, Griffioen G, *et al.* Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. *Gastroenterol* 1996;**110**:1020-1027.

Vasen HFA, Watson P, Mecklin J-P, Lynch HT and the ICG-HNPCC. New clinical criteria for Hereditary Nonpolyposis Colorectal Cancer (HNPCC, Lynch Syndrome) proposed by the International Collaborative Group on HNPCC. *Gastroenterology* 1999;**116**:1453-1456.

Veigl ML, Kasturi L, Olechnowicz J, Ma A, Lutterbaugh JD, Periyasamy S, *et al.* Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci USA* 1998;**95**:8698-8702.

Viel A, Genuardi M, Capozzi E, Leonardi F, Bellacosa A, Paravatou-Petsotas M, *et al.* Characterization of MSH2 and MLH1 mutations in Italian families with hereditary non-polyposis colorectal cancer. *Genes Chrom and Cancer* 1997;**18**:8-18.

Wahlberg SS, Nyström-Lahti M, Kane MF, Kolodner RD, Peltomäki P, Lindblom A. A low frequency of hMSH2 mutations in Swedish HNPCC families. *Int J Cancer* 1997;**74**:134-137.

Wang T-F Kleckner N, Hunter N. Functional specificity of MutL homologs in yeast: Evidence for three MLH1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. *Proc Natl Acad Sci USA* 1999;**6**(24):13914-13919.

Warthin AS. Hereditary with reference to carcinoma. *Arch Intern Med* 1913;**12**:546-555.

Watanabe A, Ikejima M, Suzuki N, Shimada T. Genomic organization and expression of the human MSH3 gene. *Genomics* 1996;**31**(3):311-318.

Watatani M, Yoshida T, Kuroda K, Ieda S, Yasutomi M. Allelic loss of chromosome 17p, mutation of the p3 gene, and microsatellite instability in right- and left-sided colorectal cancer. *Cancer* 1996;**77**:1688-1693 .

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

Weber TK, Chin H-M, Rodriguez-Bigas M, Keitzs B, Gilligan R, O'Malley L, *et al.* Novel hMLH1 and hMSh2 germline mutations in African Americans with Colorectal Cancer. *JAMA* 1999;**281**(24):2316-2320.

Weissenbach J, Gyapay G, Dip C, Vignal A, Morrissette J, Millasseau P, *et al.* A second-generation linkage map of the human genome. *Nature* 1992;**359**:794-801.

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

Wijnen J, Khan PM, Vasen H, Van der Klift H, Mulder A, Van Leeuwen-Cornelisse I, *et al.* Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. *Am J Hum Genet* 1997;**61**:329-335.

Williams AO, Prince DL. Intestinal polyps in the Nigerian African. *J Clin Pathol* 1975;**28**(5):367-371.

Williams AO, Chung EB, Agbata A, Jackson MA. Intestinal polyps in American Negroes and Nigerian Africans. *Br J Cancer* 1975;**31**(4):485-491.

Wilson DM III, Carney JP, Coleman MA, Adamson AW, Christensen M, Lamerdin JE. Hex1: a new human Rad2 nuclease family member with homology to yeast exonuclease 1. *Nucl Acids Res* 1998;**26**(16):3762-3768.

Wilson TM, Ewel A, Duguid JR, Eble JN, Lescoe MK, Fishel R, *et al.* Differential cellular expression of the human MSH2 repair enzyme in small and large intestine. *Cancer Res* 1995;**55**:5146-5150.

Winand NJ, Panzer JA, Kolodner RD. Cloning and characterization of the human and *Caenorhabditis elegans* homologs of the *Saccharomyces cerevisiae* MSH5 gene. *Genomics* 1998;**53**:69-80.

Zhou X-P, Hoang J-M, Cottu P, Thomas G, Hamelin R. Allelic profiles of mononucleotide repeat microsatellites in control individuals and in colorectal tumours with and without replication errors. *Oncogene* 1997;**15**:1713-1718.

APPENDICES

Appendix A
Patients attending Chris Hani Baragwanath Hospital

Black colorectal cancer patients				
Patient	**Age**	**Gender**	**Cancer site**	**Grade of differentiation / Tumour stage**
CRC 1	53	M	Ceacum	Duke C
CRC 2	48	F	Rectum	
CRC 3	72	F	Left colon	Duke B
CRC 7	73	M	Colon	Duke C
CRC 8	51	M	Rectum	Duke B
CRC 9	U	M	Rectum	
CRC 10	66	M	Descending colon	
CRC 11	72	M	Colon	Duke B
CRC 12	51	M	Ceacum	Duke B
CRC 15	49	M	Right colon	Duke B
CRC 16	74	M	Ascending colon	Duke B
CRC 17	56	M	Ceacum	
CRC 18	32	F	Rectum	
CRC 19	14	F	Right colon	
CRC 20	60	M	Rectum	
CRC 21	42	M	Ceacum	
CRC 22	42	M	Sigmoid	Duke C
CRC 23	39	F	Ceacum	
CRC 24	51	F	Sigmoid	Duke C
CRC 25	32	F	Ceacum	Duke C
CRC 26	73	F	Rectum	
CRC 27	72	M	Sigmoid	Duke C
CRC 28	43	M	Ceacum	
CRC 29	52	F	Rectum	
CRC 30	16	F	Hepatic flexure	Duke C
CRC 31	74	F	Left colon	Duke C
CRC 32	68	F	Left colon	Duke B
CRC 33	43	M	Ceacum	Duke C
CRC 34	66	M	Rectum	Duke C
CRC 35	33	M	Ceacum	Duke B
CRC 36	27	F	Ascending colon	Duke C
CRC 37	46	F	Sigmoid	Duke C
CRC 38	72	M	Colon	
CRC 39	22	M	Ceacum	
CRC 40	62	M	Descending colon	Duke C
CRC 41	51	F	Transverse colon	Duke C
CRC 43	76	F	Transverse colon	
CRC 45	31	M	Sigmoid	Duke B
CRC 46	60	F	Sigmoid	Duke B
CRC 47	40	F	Colon	Duke B
CRC 70	67	M	Colon	

M = Male; F = Female; CRC = black patient

White colorectal cancer patients				
Patient	**Age**	**Gender**	**Cancer site**	**Grade of differentiation / Tumour stage**
CRW 1	33	M	Ascending colon	
CRW 2	75	F	Colon	Duke C

M = Male; F = Female; CRW = Caucasian patient

Appendix B
Patients attending Kalafong and Pretoria Academic Hospitals

Black colorectal cancer patients				
Patient	**Age**	**Gender**	**Cancer site**	**Grade of differentiation / Tumour stage**
CRC 48	32	M	Ceacum	Gr II, Duke C
CRC 49	34	F	Hepatic flexure	Gr II, Duke C
CRC 50	65	M	Ceacum	Gr II, Duke B
CRC 51	48	M	Ascending colon	Gr III, Duke C
CRC 52	52	M	Ascending colon	Gr II, Duke C
CRC 53	63	F	Ceacum	Gr II, Duke D
CRC 54	59	M	Hepatic flexure	Gr II, Duke C
CRC 55	56	F	Ceacum	Gr III, Duke C
CRC 56	39	M	Ceacum	Gr II, Duke C
CRC 57	11	F	Hepatic flexure	Gr II
CRC 59	65	F	Sigmoid	Gr II, Duke D
CRC 60	80	F	Descending colon	Duke C
CRC 61	60	M	Splenic flexure	Gr II, Duke B
CRC 62	28	F	Sigmoid, rectum	Gr II, Duke B
CRC 63	19	M	Rectum	Duke C
CRC 64	25	M	Rectum	Gr II, Duke B
CRC 65	70	F	Rectum	Gr II, Duke A
CRC 66	63	F	Rectum	Gr II
CRC 67	30	F	Rectum	Gr II
CRC 68	70	F	Rectum	Gr II, Duke C
CRC 69	50	M	Rectum	Gr II, Duke C
CRC 71	67	M	Transverse colon	Gr II, Duke B
CRC 72	73	M	Transverse colon	Gr II, Duke C
CRC 73	29	F	Transverse colon	Gr III, Duke C
CRC 74	72	F	Transverse colon	Gr II, Duke C
CRC 75	62	F	Splenic flexure	Gr II, Duke B
CRC 76	66	M	Left colon	Gr II, Duke C
CRC 77	41	F	Left colon	Gr II, Duke B
CRC 78	24	M	Descending colon	
CRC 79	44	F	Descending colon	Gr II, Duke A
CRC 80	44	M	Splenic flexure	Gr II, Duke B
CRC 81	48	M	Descending colon	Gr III, Duke C
CRC 82	60	F	Sigmoid	Gr II
CRC 83	68	F	Sigmoid	Gr III, Duke C
CRC 84	37	F	Sigmoid	Gr II
CRC 85	57	F	Sigmoid	Gr II, Duke B
CRC 86	64	M	Rectum	Gr II, Duke B
CRC 87	44	M	Rectum	Gr II, Duke C
CRC 88	56	M	Rectum	Gr II, Duke C
CRC 89	68	M	Rectum	Gr II, Duke B
CRC 90	82	F	Rectum	Gr II, Duke C
CRC 91	41	M	Rectum	Gr II

Patient	Age	Gender	Cancer site	Grade of differentiation / Tumour stage
CRC 92	47	M	Rectum	Gr II
CRC 93	80	M	Rectum	Gr II, Duke B
CRC 94	41	F	Rectum	Gr II
CRC 95	49	M	Rectum	Gr II
CRC 96	52	F	Rectum	Gr II, Duke B
CRC 97	50	F	Rectum	Gr II
CRC 98	32	F	Rectum	Gr II
CRC 99	65	M	Rectum	Gr II, Duke C
CRC 100	64	M	Rectum	Gr II
CRC 123	42	F	Rectum	Gr II
CRC 124	72	M	Descending colon	Gr II, Duke C
CRC 125	50	M	Ascending colon	Gr II, Duke D
CRC 126	U	F	Descending colon	Gr II, Duke B
CRC 127	53	F	Sigmoid	Gr II, Duke D
CRC 128	34	M	Ceacum	Gr II, Duke B
CRC 129	63	M	Descending colon	Gr II, Duke B
CRC 144	46	F	Sigmoid	Gr II, Duke C
CRC 145	57	M	Ascending colon	Gr II, Duke B
CRC 146	49	M	Rectum	Gr II, Duke B
CRC 147	48	F	Rectum	Gr II, Duke C
CRC 148	68	F	Transverse colon	Duke B
CRC 149	52	M	Hepatic flexure	Gr II, Duke B
CRC 150	39	M	Rectum	Gr II
CRC 151	64	F	Ceacum	Gr II, Duke B
CRC 152	83	F	Ceacum	Gr II, Duke B
CRC 153	39	M	Transverse colon	Gr II, Duke B

M = Male; F = Female; CRC = black patient

White colorectal cancer patients				
Patient	**Age**	**Gender**	**Cancer site**	**Grade of differentiation / Tumour stage**
CRW 16	34	F	Colon	Gr II, Duke D
CRW 17	38	F	Sigmoid	
CRW 18	45	M	Rectum	Gr II
CRW 19	52	M	Sigmoid	Gr II, Duke D
CRW 20	36	M	Ceacum	Gr II, Duke C
CRW 21	52	M	Rectum	Gr II, Duke C
CRW 22	52	M	Sigmoid	Gr II, Duke B
CRW 23	23	F	Splenic Flexure	Gr II, Duke C
CRW 24	44	M	Rectum	Gr II, Duke C
CRW 25	72	F	Transverse colon	Gr II, Duke B
CRW 26	57	F	Left Colon	Gr II, Duke D
CRW 27	82	F	Rectum	Gr II, Duke C
CRW 28	71	F	Colon	Gr II, Duke C
CRW 29	69	M	Sigmoid	Gr II, Duke B
CRW 34	51	M	Ceacum	Gr II, Duke B

Patient	Age	Gender	Cancer site	Grade of differentiation / Tumour stage
CRW 35	43	M	Rectum	Gr II
CRW 36	55	F	Rectum	Gr II, Duke C
CRW 37	60	M	Splenic flexure	Gr II, Duke C
CRW 38	74	F	Right colon	Gr II, Duke C
CRW 39	76	M	Rectum	Gr II, Duke B
CRW 40	80	M	Sigmoid	Gr III, Duke B
CRW 41	66	F	Sigmoid	Gr II, Duke C
CRW 43	68	F	Rectum	Gr II, Duke B
CRW 44	73	M	Rectum	Gr II
CRW 45	80	F	Rectum	Gr II, Duke
CRW 46	37	F	Descending colon	Gr II, Duke D
CRW 47	81	F	Ceacum	Gr II, Duke C
CRW 48	73	F	Ceacum	Gr II, Duke B
CRW 49	75	F	Ceacum	Gr II, Duke D
CRW 50	70	F	Sigmoid	Gr II, Duke D
CRW 51	56	M	Rectum	Duke B
CRW 52	74	F	Rectum	Gr II, Duke C
CRW 53	62	M	Rectum	Gr III, Duke C
CRW 54	79	M	Rectum	Gr II, Duke C
CRW 55	77	F	Sigmoid	Gr II, Duke C
CRW 56	54	M	Sigmoid	Duke D
CRW 57	72	F	Ceacum	Gr II, Duke C
CRW 58	63	F	Rectum	Gr III, Duke B
CRW 59	64	F	Rectum	Gr II
CRW 60	51	M	Colon	Gr II, Duke B
CRW 61	68	M	Rectum	Gr II, Duke C
CRW 62	76	M	Transverse colon	Gr II, Duke B
CRW 63	58	F	Splenic Flexure	Gr II, Duke B
CRW 64	75	F	Sigmoid	Gr II, Duke B
CRW 65	51	F	Sigmoid	Gr II, Duke D
CRW 66	79	F	Transverse colon	Gr II, Duke C
CRW 67	73	M	Sigmoid	Gr II, Duke C
CRW 68	53	F	Sigmoid colon	Gr II, Duke B
CRW 69	63	M	Rectum	Gr II, Duke D
CRW 70	77	F	Rectum	Gr II
CRW 71	88	F	Rectum	Gr II, Duke B
CRW 73	64	F	Transverse colon	Gr II, Duke C
CRW 74	U	M	Rectum	Gr II, Duke C
CRW 75	62	F	Transverse colon	Gr II
CRW 76	77	M	Ceacum	Gr II, Duke B
CRW 77	58	F	Rectum	Gr II, Duke C
CRW 78	72	F	Ceacum	Duke C
CRW 80	80	F	Rectum	Gr II
CRW 81	74	F	Ceacum	Gr II
CRW 82	41	F	Sigmoid	Gr II
CRW 83	68	M	Rectum	Gr II
CRW 84	74	F	Sigmoid	Gr II

Patient	Age	Gender	Cancer site	Grade of differentiation / Tumour stage
CRW 85	79	F	Colon	Gr III, Duke C
CRW 86	77	F	Sigmoid colon	Gr II, Duke B
CRW 87	68	F	Splenic Flexure	Gr II, Duke B
CRW 88	80	F	Rectum	Duke A
CRW 89	75	F	Rectum	Gr II, Duke B
CRW 90	60	F	Rectum	Gr II, Duke B
CRW 91	61	F	Ceacum	
CRW 92	70	F	Ceacum	Gr III
CRW 93	67	M	Rectum	Gr II, Duke B
CRW 94	61	F	Ceacum	Gr II, Duke C
CRW 95	72	M	Ceacum	Gr II, Duke D
CRW 96	61	M	Hepatic flexure	Gr II, Duke B
CRW 97	82	F	Rectum	Duke B
CRW 98	76	F	Ceacum	Gr II, Duke C
CRW 99	73	F	Rectum	Gr III, Duke B
CRW 100	64	M	Left colon	Gr II, Duke A
CRW 101	72	F	Sigmoid	Gr III, Duke D
CRW 102	70	F	Rectum	
CRW 103	68	F	Sigmoid	
CRW 104	70	F	Sigmoid colon	Gr II, Duke C
CRW 105	36	F	Colon	
CRW 106	56	M	Ceacum	Gr II, Duke C
CRW 107	72	M	Transverse colon	Gr II, Duke B
CRW 108	67	F	Transverse colon	Gr II, Duke B
CRW 109	75	F	Splenic flexure	Gr II, Duke B
CRW 110	74	F	Rectum	Gr II, Duke C
CRW 111	43	F	Rectum	Gr II
CRW 112	78	F	Hepatic flexure	Gr II, Duke C
CRW 113	45	F	Colon	Gr II
CRW 114	68	F	Rectum	Gr II
CRW 115	80	M	Sigmoid	Gr II, Duke B
CRW 116	68	F	Sigmoid	Gr II, Duke B
CRW 117	65	M	Ceacum	Gr II
CRW 118	67	F	Ascending colon	Gr II, Duke C
CRW 119	57	F	Rectum	Gr II
CRW 120	83	F	Rectum	Gr II, Duke B
CRW 121	59	M	Rectum	Gr II, Duke C
CRW 122	77	F	Rectum	Gr II, Duke C
CRW 123	80	F	Rectum	Gr II
CRW 124	71	M	Rectum	
CRW 125	67	F	Rectum	Gr II, Duke C
CRW 126	69	F	Rectum	Gr II, Duke B
CRW 127	49	M	Rectum	
CRW 128	76	F	Rectum	Gr II, Duke C
CRW 129	59	F	Rectum	Gr II, Duke B
CRW 130	58	F	Rectum	Gr II, Duke B

M = Male; F = Female; CRW = Caucasian patient

Appendix C

Pathologic staging of colorectal cancer

Broders' classification (Broders, 1925)

This grades the degree of histological differentiation of the tumour.

Grade I = well-differentiated

Grade II = moderately differentiated

Grade III = poorly differentiated

Grade IV = totally anaplastic

Well-differentiated tumours grow slowly, while poorly differentiated tumours tend to grow fast.

Dukes' classification (Turnbull *et al.*, 1967)

Dukes' A = invasion into the mucosa

Dukes' B = invasion through the mucosa but still in the colon

Dukes' C = lymph node metastasis

Dukes' D = metastasis to other organs