Molecular pathology of the *hMSH2* mutator gene and its transcripts in patients with colorectal cancer in the West of Scotland

Abdoreza Davoodi- Semiromi

A thesis submitted to the Facullty of Medicine, University of Glasgow,
for the degree of Doctor of Philosophy
(Ph.D)

Duncan Guthrie Institute of Medical Genetics,
Yorkhill NHS Trust
Yorhill, Glasgow

ProQuest Number: 13815599

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 13815599

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

GLASGOW UNIVERSITY
LIBRARY
11084 (copy 1)

Declaration

I certify that this thesis does not contain material previously published or written by any other person except where referred to in the text and that the results in this thesis have not been submitted for any other degree or diploma.

Abdoreza Davoodi - Semiromi

TO THE GLORIOUS MEMORY OF MY BELOVED FATHER

HASHEM DAVOODI-SEMIROMI

(1910 - 1985)

Acknowledgements

I wish to express my deepest appreciation to the many people who have contributed toward this project. Among these are Professor J. M. Connor who made it possible for me to come to Glasgow and start this project in 1994, and also for his supervision, support, guidance and encouragement throughout the project. My continued gratitude must be extended to Dr. W. G. Lanyon, who patiently and thoughtfully supervised this project and helped me complete several scientific abstracts.

All the clinicians, especially Dr. R. Davidson who organised the collection of samples, and also the patients for their participation in this project, are to be thanked. A special thanks to Dr. C. Purdie from the Pathology Department at Glasgow Royal infirmary. My gratitude needs also to be extended to the many scientists from the molecular laboratory, especially Dr. A. Cooke. And thanks also to Dr. Sara Johnson, Dr. G. O' Neill and M. Bonyadi from the Developmental Genetics group and also to Dr. David Aitken , G. Graham from Biochemical Genetics group. I would like to thank the people from the P.G.R laboratory and in situ laboratory including Munis Dundar, Smita Purandare, Fawziah Mohammad, Alaa Elshafy, M. Mohaddes and S. R. Ghaffari. Many other members of the Duncan Guthrie Institute of Medical Genetics, have assisted me over the last four years in particular, Betty O' Hare, Ann Ross and Anne Theriault whose help never be forgotten.

Last but not least, my appreciation would like to be extended to my wife and colleague, Guity, who supported and encouraged me throughout the project and to my daughters Maryam and Yalda who suffered much over the last four years.

Table of contents

Chapter two: Materials and methods

List of Figures

List of Tables

List of abbreviations

A	Adenine
APC	Adenomatous polyposis colorectal carcinoma
APS	Ammonium persulphate
ARMS	Amplification refractory mutation system
ASO	Allele specific oligonucleotide hybridisation
bp	Base pair
BRCA1,2	Breast cancer gene 1, 2
C	Cytosine
^{0}C	Degree Celsius
CCM	Chemical cleavage analysis
cDNA	Complementary Deoxyribonucleic acid
CGH	Comparative genomic hybridisation
cm	Centimetre
CRC	Colorectal carcinoma
ddATP	Dideoxyadenisine triphosphate
ddCTP	Dideoxycytosine triphosphate
DCC	Deleted in colon cancer
ddGTP	Dideoxyguanosine triphosphate
ddTTP	Dideoxythymidine triphosphate
DGGE	Denaturing gradiant gel electrophoresis
ddNTPs	Dideoxyribonucleotide triphosphate
DEPC	Diethyl pyrocarbonate
DM	Myotonic dystophy
DMD	Duchenne muscular dystrophy
DNA	Deoxyribonucleic acid
dNTP	2'-deoxyribonucleoside triphosphate
DRPLA	Dentatorubral pallidoluysian ataxia
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
FAP	Familial adenomatous polyposis
FISH	Fluoresence *in situ* hybridisation
FRAX	Fragile X syndrome
G	Guanine
g	gram
HA	Heteroduplex analysis
HD	Huntington's disease
HNPCC	Hereditary nonpolyposis colorectal carcinoma
hMSH2	Human mut S homologous 2
hMLH1	Human mut L homologous 1
kb	Kilobase
kd	Kilo dalton
M	Molar
MJD	Machado-Joseph disease

µg	Microgram
µl	Microlitre
mg	Milligram
MI	Microsatellite instability
ml	Millilitre
µM	Micromolar
mM	Millimolar
mm	Millimetre
MMLV	Moloney Murine Leukaemia Virus
MOPS	3-(N-morpholino) propanesulfonic acid
mRNA	Messenger RNA
ng	Nanogram
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
pmol	Picomole
PMS1, 2	Post meiotic segregation 1, 2
PTT	Protein truncation test
RER	Replication error
RNA	Ribonucleic acid
r.p.m.	Revolutions per minute
RT-PCR	Reverse transcriptase polymerase chain reaction
SBMA	Spinal and bulbar muscular atrophy
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SSCP	Single strand conformation polymorphism
T	Thymidine
TBE	Tris-acetate ethylenediaminetetra acetic acid
TE	Tris- ethylenediaminetetra acetic acid
TEMED	N,N,N,N-tetramethylenthylenediamine
Tris	Tris (hydroxymethyl) aminomethane
UV	Ultraviolet
V	Vol
W	Watt

List of presentations

1. Clinical Genetics Society, Spring Conference London 1995; Davoodi, A., Lanyon, G., and J. M. Connor; **Microsatellite instability in hereditary nonpolyposis colorectal carcinoma (HNPCC).**

2. British Medical Genetics Conference, York 1995; Davoodi, A., Lanyon, G., and J. M. Connor; **A novel point mutation in the *hMSH2* gene revealed by the reverse transcriptase polymerase chain reaction (RT-PCR).**

3. British Human Genetics Conference, York 1996; Davoodi, A., Lanyon, G., R. Davidson and J. M. Connor; **Three novel mutations and two normal variations in the *hMSH2* gene in Scottish families with colorectal carcinoma.** Journal of Medical Genetics, Sept. 1996, Vol. 33, pp: sp58.

4. British Human Genetics Conference, York 1997; Davoodi, A., Lanyon, G., R. Davidson and J. M. Connor; **mRNA and DNA mutations at mismatch repair genes.** Journal of Medical Genetics, Sept. 1997, Vol. 34, pp: S79.

Note: The abstracts are bound at the back of this thesis.

Summary

Colorectal cancer (CRC) is the second or third most common cancer in Western countries, and its incidence is rising. Among the hereditary forms Familial Adenomatous Polyposis (FAP) is a rare, dominantly inherited disease which is caused by germ-line mutations of the Adenmatous Polyposis Colorectal Cancer (*APC)* gene. A second form of CRC that shows familial aggregation is Hereditary Non-Polyposis Colorectal Cancer (HNPCC).

The main aims of this study were: (i) to assess microsatellite instability in the tumour of patients with colorectal carcinoma and (ii) to identify germ-line mutations in two mismatch repair genes, using a variety of techniques.

To do this, in two groups of unselected patients, 30 pairs of normal/tumour DNAs and 47 RNA samples were made available for microsatellite instability studies and mutation analysis, respectively. Nine polymorphic markers were chosen for microsatellite instability analysis including: D13S160, D2S119, D8S282, D18S34, 635/636, D2S123, LPLCA- A, D2S378, LPLCA, B. Using this approach, 30% (9/30) of the tumours exhibited instability at one or more loci (RER+). There was no significant difference between age and sex in patients with or without microsatellite instability.

Using the reverse transcriptase polymerase chain reaction (RT-PCR), the entire coding sequence of the *hMSH2* gene was amplified in either two or three overlapping segments. By employing RNA and DNA based techniques a number of germ-line mutations were found in the *hMSH2* and *hLMH1* genes. The mutations included: one novel splice site mutation responsible for exon 15 skipping, two novel deletions of exons 2-6 and 2-8 from the mRNA transcripts, four different mRNA deletions at different sites in the *hMSH2* gene, one missense mutation at exon 6 and three intronic mutations at positions +9, -10 and +19 in intron 1, 11, and 14 respectively. A variation of the published sequence was also found at 3' end of the untranslated region in the *hMSH2* gene.

None of the families in this study were large enough for linkage analysis. In one HNPCC family, one splice site mutation was found at position +5 in intron 15. This resulted in exon 15 skipping from transcripts and produced a truncated hMSH2 protein. This germ-line mutation was present in 4 affected members of the family, but was not present in the healthy members of the family or in 177 other colorectal cancer patients. The individual who bears this germ-line mutation showed a RER+ phenotype in the tumour sample.

A patient whose tumour was RER+ had a deletion of exons 2 to 6 inclusive which resulted in an out-of- frame deletion of codons 71-359. This deletion created a premature termination codon within exon 7 at nucleotide position +3 from the splice site. No genomic mutations were found which could account for the exon skipping. No other member of the family was available for further investigation.

A patient with a tumour showing the RER+ phenotype had a deletion of exons 2 through exon 8 inclusive which resulted in an out-of- frame deletion of codons 71-466. This deletion also created a premature termination codon within exon 9 at position +12 from the splice site. No genomic alterations were found which could account for the exon skipping. This individual was the only member of the family available for investigation.

Four different partial mRNA deletions in different exons were found. Three of them produced premature termination codons. No alteration could be found by sequencing the genomic splice sites, or in the corresponding exons. One of the partial deletions was found in four unrelated individuals. The breakage points of the aberrant mRNAs were not located at the splice site junctions in any case. In order to clarify whether or not these partial mRNAs deletions are normal, RT-PCR from 20 healthy individuals was carried out. The whole coding sequences of the *hMSH2* gene were amplified in three overlapping fragments and no extra bands were observed. These results indicate that none of the mRNA deletions reported in this study are common normal variations.

A T to C transition was found in a short polypyrimidine tract in intron 12 at nucleotide position -6. Particularly, short poylypyrimidine tracts at the 3' end of the intron have been shown to effect mRNA splicing (Roscigno et al. 1993), but this mutation had no affect on the mRNA splicing process, since no short transcript was found. It was, however, found in 15% of the normal population when 86 normal chromosomes were tested.

A G to A transition was also found at codon 322 in exon 6 which changes glycine to asparatic acid. Since glycine is a non- polar neutral amino acid and asparatic acid is an acidic one, it is likely that the polarity of the protein will be changed. The frequency of this mutation in the normal population was 4%, when 62 normal chromosomes were analysed.

A C to G transversion was also found in the intronic region of exon 1 at nucleotide position +9. This mutation had no effect on RNA splicing, since no short transcript was found. The frequency of this mutation was 45% when 66 normal chromosomes were studied.

It is believed that the majority of *hMLH1* mutations cluster at the exonic region 15-16. These two exons were amplified by PCR followed by SSCP in patients as described ealier. On sequencing the genomic DNA, an A to G change was found at position -19 in intron 14. Further study showed this mutation in 44% of the normal population when 60 normal chromosomes were examined. No alteration was found in exon 16 of the *hMLH1* gene either by SSCP or by PCR, as described by Lahti et al. (1995). All of the point mutations which were found in this project were confirmed by enzymatic assays.

The results of this study provide further evidence that the *hMSH2* gene is involved in a proportion of patients with colorectal carcinoma in the West of Scotland, and show that frameshift mutations are the most common cause of the disease. This study also showed that the mutations in the *hMSH2* gene are scattered thoughout its entire coding sequence, without any sign of hotspots or common mutations. HNPCC is a heterogeneous disease, and only the simultaneous application of all available screening methods guarantees the highest probability of detecting all existing mutations.

CHAPTER ONE

INTRODUCTION

1. 1. Colorectal cancer

Cancer is a complex family of diseases, and carcinogenesis, the conversion of a normal cell into a cancer cell, is a complex multistep process. Clinically, cancer is a large group of diseases, perhaps a hundred or more, that vary in age of onset, rate of growth, state of cellular differentiation, diagnostic detectability, invasiveness, metastatic potential, response to treatment, and prognosis. In terms of molecular biology, however, cancer may represent a relatively small number of diseases caused by similar molecular defects in cell function, and resulting from similar alterations to a cell's genes. Ultimately, cancer is a disease of abnormal gene expression (Ruddon, 1995).

Colorectal cancer (CRC) is the second or third most common cancer in the Western countries with a cumulative life incidence of 5-6% and its incidence is rising. It is rare in the third world (Ruddon, 1995). Up to 90% of all CRCs are sporadic. Among the hereditary forms, Familial Adenomatous Polyposis (FAP) is a rare, dominantly inherited disease leading to cancer and is caused by germ-line mutations of the *APC* gene (Kinzler et al. 1991; Nishisho et al. 1991; Groden et al. 1991; Joslyn et al. 1991). Affected patients have several unusual phenotypic features, most notably the presence of thousands of benign tumours lining the entire large intestine. FAP accounts for about 1% of CRC cases in the western world (Jarvinen, 1992).

1. 1. 1. Hereditary nonpolyposis colorectal cancer (HNPCC)

A second form of CRC that shows familial aggregation is hereditary nonpolyposis colorectal cancer (HNPCC). It is more common than FAP, accounting for 4 to 13 % of all CRC in industrial nations (Lynch et al. 1993). It is impossible to reliably distinguish patients with this form of CRC from sporadic cases on physical examination alone, as neither have diffuse polyposis, or other unusual stigma.

In the past, the definition of HNPCC was exclusively based on pedigree structure and age at onset. The International Collaborative Group (ICG) on hereditary nonpolyposis colorectal cancer suggested three obligate criteria known as the **Amsterdam criteria** or **ICG-HNPCC** (Vasen et al. 1991) 1- At least three relatives with histologically verified colorectal cancers; one of them should be a first degree relative to the other two (familial adenomatous polyposis should be excluded). 2- At least two successive generations should be affected. 3- In one of the relatives, colorectal

cancers should be diagnosed when they are 50 years of age or less.

These criteria favour a typical form of HNPCC. For example, the requirement of two successive generations excludes families with low penetrance and *de novo* germ-line mutations. On the other hand, the definition ignores tumours other than colorectal, and thus potentially excludes families in which tumours in other organs are frequent. Finally, the age requirement obviously excludes families in which penetrance is delayed.

The high susceptibility to cancer often leads to the occurrence of more than one tumour at the same time (synchronous tumours), and to repeated occurrences (metachronous tumours). Further characteristics of HNPCC are : younger age at onset (average 40 - 45 years) than in sporadic CRC (over 60 years); more often right-sided tumours (approximately two-thirds) than in sporadic CRCs (approximately one-third), with no explanation thus far of the difference in sideness (Bufill, 1990); a generally unfavourable histology (poor differentiation) with a prognosis that is paradoxically better than in sporadic CRC.

Tumours in HNPCC, as in sporadic CRC, develop via a precancerous growth called an adenoma. Sporadic CRC can be prevented by colonscopic screening and removal of adenomas (Winawer et al. 1993); recently, the same has been shown for HNPCC (Jarvinen et al. 1995).

Much attention has been paid to the spectrum of tumours that occur in HNPCC. There is no explanation for the observations that some organs are, and others clearly are not, susceptible to cancer in this condition. Approximately two-thirds of the tumours seen in members of the HNPCC families under study were in the colorectum, with the rest in other organs. Among these, the endometrium, stomach, pancreas, bile ducts, kidney and ureters are most clearly involved, whereas other tumours in, for example, breast, sarcoma, skin, and lung may or may not arise as a result of the inherited susceptibility. Clarification of this point is obviously of major interest and can only be done by determining who has, and who does not have, the predisposing mismatch repair gene mutation that occurs in each family. In the large family reported by Warthin,(1913,1925,1931)endometrial and stomach cancer predominated, whereas later generations of the same family display a predominance of CRC.

The incidence of HNPCC estimates vary widely. The highest estimates could imply a frequency of 1 in 200 or higher (Houlston et al. 1992). A more commonly

quoted estimate, 1 in 2000, would still place HNPCC among the commonest heritable disorders, in the same order as cystic fibrosis and fragile X syndrome. Other estimates are lower, such as that made by Bodmer et al. (1994), who claimed that HNPCC is rarer than familial adenomatous polyposis in which the incidence is of the order of 1 in 8000 to 1 in 15000.

The great variation in estimates of incidence is due to the lack of objective diagnostic criteria other than pedigree structure. Most attempts to determine the incidence of HNPCC begin by pedigree analysis in order to determine the proportion of those who have HNPCC, in a given study population of colorectal cancer patients.

1. 1. 2. Tumourigenesis in colorectal cancer

Colorectal carcinomas have proven to be an excellent model system in which to study the role of hereditary factors, environmental carcinogenesis, oncogene activation, and tumour suppressor gene inactivation in the initiation and progression of human solid tumours (Fearon et al. 1990; 1992).Tumours of various stages of dysplasia and malignancy, from benign adenomas to invasive cancers, can be obtained surgically. There is a clear familial predisposition for some of the cases, and a variety of genetic mutational events have been identified in the disease tissue (Figure 1. 3). In addition, there is a clinically definable progression from adenoma to malignant neoplasia. It is important to note that although there is evidence for a preferred order of the genetic changes rather than their specific sequence that is important in the initiation and progression of the neoplastic process (Fearon et al. 1990, 1992). For example , more than 90% of the colon carcinomas studied had two or more genetic alterations, and 7% of early adenomas had more than one of the four component genetic alterations observed in the cancer cells [(Fearon et al. 1990) (see Figure 1.1)]. As early adenomas progressed to intermediate and late adenoma stages, the number of tumours with more than one genetic change went to 25% and 94% respectively. Some late-stage adenomas had all four of the common genetic mutations, indicating that probably more than these four events are required for progression to malignancy. Another interesting point is that most, if not all, colorectal cancers, appear to have similar mutational changes irrespective of hereditary factors or environmental exposure (Fearon et al. 1990).

The highly propensity of patients with the autosomal dominant familial adenomatous

polyposis syndrome (FAP) to develop colorectal carcinomas has provides an important clue to the predisposing genetic alterations that are involved in the initiation and progression of this type of cancer.

Affected individuals may develop hundreds or thousands of adenomatous polyps in their colons and rectums. Only a small percentage of these go on to become malignant, consistent with the multi- hit hypothesis of cancer. The inherited , affected gene present in *FAP* has been cloned and is called *APC* (adenomatous polyposis coli). The *APC* gene is mutated in the germ-line of *FAP* patients. Most often, these are point mutations or frameshift mutations, and, in a few cases, gene deletions. In any case, all the inherited mutations lead to inactivation of the *APC* gene, which is located on chromosome band 5q21. These data suggests that *APC* is a tumour suppressor gene.

That the *APC* gene is also involved in non-hereditary, sporadic cases of colorectal cancer is supported by the fact that allelic losses of 5q21 are seen in 35% to 60% of patients with no known familial predisposition. The data support the concept that one or both alleles of the *APC* gene are lost or inactivated at an early stage in the development of colorectal cancer, since 5q allelic losses are detected often in small, early adenomas (Vogelstein et al. 1988).

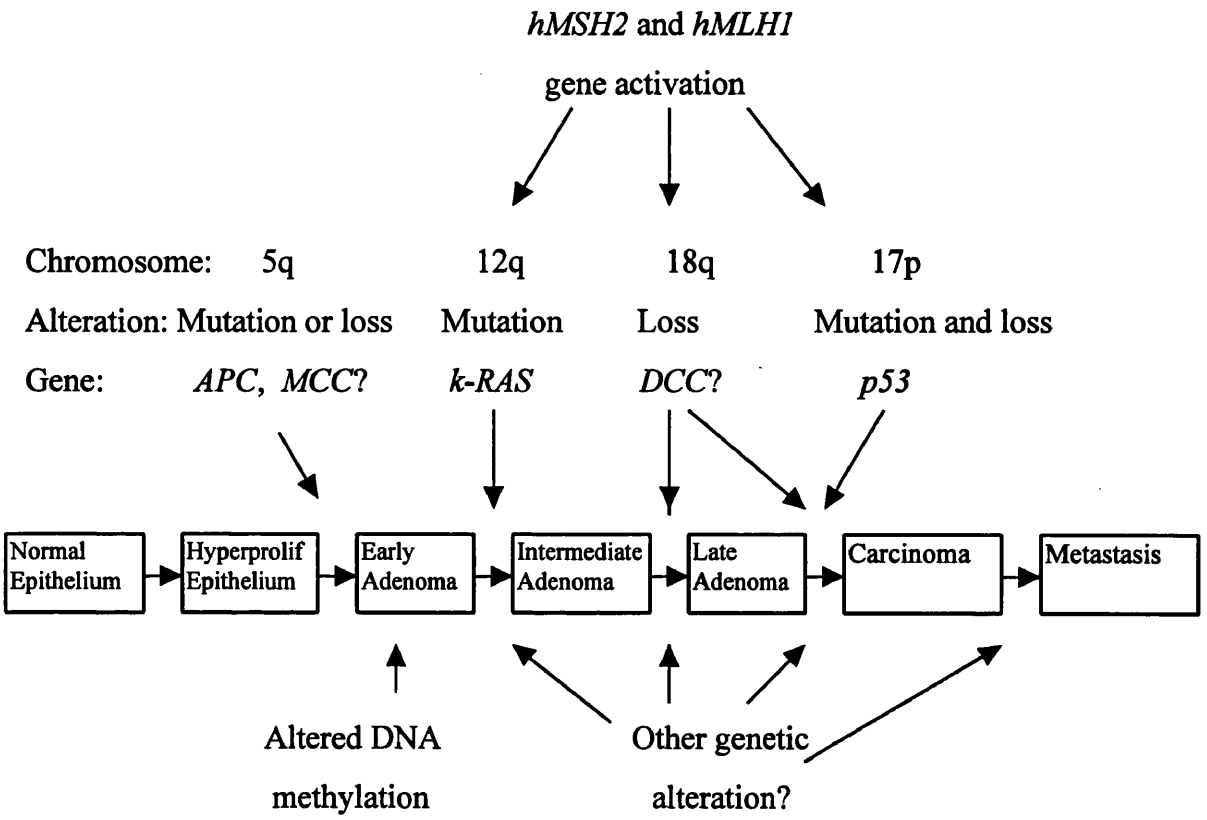


Figure 1. 1: A genetic model for colorectal tumorigenesis. Tumorigenesis proceeds through a series of genetic alterations involving oncogenes (*ras*) and tumour suppressor genes. The three stages of adenomas, in general , represent tumours of increasing size, dysplasia, and villous content (Fearon et al. 1990).

Another locus that maps to chromosome band 5q21 and is altered in colorectal cancer is called *MCC* (mutated in colorectal cancer). This *MCC* gene appears to be inactivated in about 15% of sporadic colorectal cancers (Kinzler et al. 1991). Loss or inactivation of the *APC* and *MCC* genes may be involved in increasing the rate of proliferation of the affected mucosal cells, thus increasing the chance for subsequent genetic abnormalities to occur.

Another early change that occurs in the genome of early adenomas is loss of methyl groups from DNA. About one-third of DNA sequences from adenomas show loss of methyl groups (Feinberg et al; 1988). Hypomethylation has been correlated with activation of silent genes and with alterations of chromosome condensation. The latter may lead to mitotic nondisjunction, resulting in gain or loss of chromosomes. Activation of genes involved in cell proliferation, such as *c-myc* and tyrosine kinase gene products, is also higher in colorectal cancer than in normal colonic mucosa. This would also contribute to the hyperproliferative state resulting from inactivation of tumour suppressor genes.

Activation of the *ras* proto-oncogenes has been implicated in a number of human cancers, including colorectal carcinoma. Mutations in the *k-ras* gene have been identified in about 50% of colonic adenomas larger than 1 cm in diameter, and in 50% of colorectal carcinomas, but in only about 10% of adenomas smaller than 1 cm (Vogelstein et al; 1988). Such *ras* mutations, usually in codons 12 or 13, tend to be observed in more dysplastic adenomas and are thought to be related to conversion of cells bearing these mutations into carcinomas.

Allelic losses-sometimes called loss of heterozygosity (LOH) because they are detected by RFLPs (restriction fragments length polymorphism)-of additional tumour suppressor genes appear to be a key to later stages in colorectal tumorigenesis. These losses usually involve only one of the two parental alleles, and the other allele may or

may not be inactivated by mutation. These lost genetic loci are thought to contain tumour suppressor genes. Subsequently, the first locus that appears to be lost in colorectal tumour progression is chromosome 18q. This region is deleted in 50% of late adenomas and more than 70% of carcinomas (Vogelstein et al. 1988). A candidate tumour suppressor gene located in the 18q band, *DCC* (deleted in colorectal carcinoma), encodes a protein with putative cell adhesion properties. Its expression is reduced or absent in colorectal carcinomas, suggesting its involvement in normal cell-cell or cell-extracellular matrix interactions required to maintain a normal state of differentiation.

The most common allelic deletion seen in colorectal cancer involves a large portion of the p region of chromosome 17. Loss of part of 17q is infrequently observed in adenomas at any stage, but it is observed in over 75% of colorectal carcinomas. Furthermore, in several patients in whom tumour progression was observed, 17q allelic losses correlated with progression of individual tumours from adenoma to carcinoma (Fearon et al. 1990; Vogestein et al. 1988). The common segment lost from 17q is now known to contain the p53 allele tumour suppressor gene (Baker et al. 1989). Loss and or mutation of p53 alleles has now been observed in a wide variety of human cancers. Point mutation of one p53 allele and loss of the remaining normal allele is a common finding in colorectal cancer, suggesting that the wild-type *p53* gene inhibits later events in colorectal carcinogenesis; when it is lost, this control is abrogated. In some tumours, however, an intermediate stage may occur; for example tumour cells may express one mutant and one wild- type p53 gene. Even in this situation, however, there is evidence that the mutant gene product, by binding to it or otherwise blocking the function of the normal p53 protein, has a so-called dominant negative effect (Fearon et al. 1990). Nevertheless, the data indicates that colorectal carcinomas with 17q allelic losses are commonly more aggressive than those without them (Kern et al. 1989).

Allelic losses in addition to those noted for chromosomes 5q, 18q and 17q have been observed in colorectal cancer. These include regions from chromosomes 1q, 4p, 6q, 8p, 9q, and 22q (Fearon et al. 1990; Fujiwara et al. 1993) One or more of these losses can be seen in 25% to 50% of patients. On the average, colorectal cancers contain four or five allelic losses. Patients with more than that number generally have a poorer prognosis. The majority of colorectal carcinoma have 17p and 18 deletions, suggesting that the suppressor genes contained in those bands are crucial to the carcinogenic

process. Other regions are lost in a more heterogeneous manner. This suggests that they may represent suppressor genes, any combination of which, if lost, could result in further clonal expansion and uncontrolled proliferation of the affected cells. This heterogeneity in allelic loss patterns could account for some of the observed differences in the biological properties and clinical course of individual patient's tumours.

As noted above, defects in DNA mismatch repair genes have also been detected in colon cancer. The first one discovered was the *hMSH2* gene, homologous to the bacterial *mutS* repair gene, found on human chromosome 2p. A second common mismatch repair gene defect has also been found . This gene is located on chromosome 3p21 and appears to be altered in 33% of HNPCC cases. The gene is called *hMLH1* and is homologous to *mutL* mismatch repair gene of *E.coli* . These data strongly suggest that , in addition to activation of oncogenes and inactivation of tumour suppressor genes, defects in DNA repair account for a major predisposing cause of human cancer.

In summary, the data indicate that although there is an overall sequence of genetic alterations in colorectal carcinogenesis, it is the progressive accumulation of the alterations that is the most consistent property of colorectal cancers.

1. 2. Mutator phenotype in colorectal cancer

The mutator genes are a group of genes which when mutated themselves can increase the frequency of other genetic mutations up to 100,000 fold above the wild type (Fowler et al. 1974). Mutator genes are well known in lower organisms like *E.coli* and *Salmonella.* The mutator phenotype refer to this high mutator strain of simple organism. Mutation in *mut*S (mut stands for mutator phenotype) and *mut*L genes in *E.coli* could increase the rate of mutation 100-fold above the wild type (Siegel et al. 1964; 1975). Mutation in some particular genes like *mut*D, in *E.coli* could increase the rate of mutation 1000-100,000-fold (Fowler et al. 1974). It is believed that replication error during DNA replication is the cause of mutator phenotype in the majority of the HNPCC tumours, and in a small number of sporadic cases. The finding of mutator phenotype in the majority of HNPCC tumours was unexpected, but not surprising, since it was studied in detail many years ago in lower organisms. This new mutator mechanism for cancer has been discovered by the detection of somatic mutations at simple repeated sequences, in the course of searching for the hereditary nonpolyposis colon cancer genes. These

mutator phenotypes are termed microsatellite instability (Thibodeau et al. 1993), ubiquitous somatic mutations (Ionov et al. 1993), genetic instability (Han et al. 1993), and replication error (Parsons et al. 1993). Microsatellite instability at simple repeat sequences can accumulate due to failures of the strand specific mismatch repair system to recognise errors due to slippage by strand misalignment (Streisinger et al. 1966) of these unstable sequences (Strand et al. 1993). Four different genes on chromosomes 2, 3 and 7 have been characterised, and defects of these genes lead to a genetic instability in which short sequences or microsatellites are characteristically altered in patients with colorectal carcinoma.

Genomic instability at short tandem repeat sequences reflects malfunction in the replication repair of DNA. Therefore, it is also referred to as the replication error (RER) phenomenon or mutator phenotype. The abnormality appears as an extra or absent allele in the tumour DNA when compared with normal DNA from the same person, indicating that a gain or loss of short repeat units has occurred. Such alterations reflect a malfunction in the replication or repair of DNA which persists throughout the lifetime of the tumour (Parson *et al*, 1993; Shibata *et al*, 1994). Microsatellite repeats are scattered throughout the human genome (it has been estimated that about 100,000 microsatellite repeats are scattered throughout the genome), about 85% of colorectal cancers from HNPCC patients shows microsatellite instability, and a majority of microsatellite loci are apparently involved (Aaltonen et al. 1993 & 1994). Thus, in HNPCC tumours, the total number of mutations at microsatellite loci alone could be as high as 85,000 per cell. Approximately 15% of apparently sporadic colorectal carcinomas exhibit microsatellite instability. RER+ colorectal tumours, whether representative of HNPCC or sporadic, are characterised by almost unique clinical and pathological features (Aaltonen et al. 1993; Ionov et al. 1993; Loth et al. 1993). The genetic background of mutator phenotype may not be the same in HNPCC and in sporadic cases. In HNPCC, microsatellite instability is believed to be due to a germ-line mutation plus a somatic event inactivating the second allele of the mismatch repair gene involved. Liu et al. (1995) found that only one in ten patients with RER+ sporadic colorectal tumour had a detectable germ-line mutation in a mismatch repair gene, suggesting that hereditary forms may not be very common in this group. Additionally, only three of seven sporadic tumour cell lines with microsatellite

instability had mutations in *MSH2, MLH1, PMS1*, or *PMS2*, implying that a significant fraction of sporadic RER+ cancers arise from mutations in other genes. It has been reported that a mutation in DNA polymerase may cause microsatellite instability in some sporadic tumours (da Costa et al. 1995). However, the mutator phenotype is not restricted to cancers belonging to the HNPCC tumour spectrum but occurs with variable frequencies in cancers not occurring in excess in HNPCC, such as lung cancer (Merlo et al. 1994; Shridhar et al. 1994).

1. 2. 1. Mutator phenotype occurs at an early stage of tumourigenesis

Repetitive sequences are short and specific sequences, present throughout the eukaryotic genomes, including both non-coding and coding regions. These are repeat sequences from mono - to hexa-nucleotides, which are dispersed throughout the eukaryotic genomic DNA. The length of repeats vary among individuals and also between species. As a result, they have shown great utility as genomic markers. In the human genome there are 50, 000 - 100, 000 interspersed (CA)n blocks, with the range of (n) being roughly 15- 30 (Hamada and Kakunaga 1982, Hamada et al. 1982, 1984a; Sun et al. 1984; Tautz and Renz 1984; Gross and Garrad 1986; Braaten et al. 1988). The function of the (CA)n repeats is unknown, but it has been proposed that they serve as hotspots for recombination (Slightom et al. 1980), and also some work has suggested that DNA can be exchanged between these duplicated sequences or participate in gene regulation. Additionally, areas of interest can be studied further though positional cloning (Botstein et al. 1980; Weber et al. 1989; Weber, 1990;).

Repeated DNA sequence families which do not include functional gene members are composed of arrays of tandem repeats, or of individual repeat units interspersed with other DNA sequences including: satellite DNA, minisatellite DNA, microsatellite DNA, interspersed repeats (Alu family) and the Kpn (L1) family. Expansion or deletion of alleles could create a microsatellite shift or novel allele on the autoradiogram as shown in Figure 1. 2.

Genomic instability is a landmark for some sporadic and hereditary cancers of the colon. Young et al. (1993) reported that genomic instability occurred in colorectal carcinomas, but it is rare in adenomas. They analysed 108 carcinomas and 46 adenomas with 4-15 microsatellite markers, and reported 6.5% MI in patients with colorectal carcinoma.

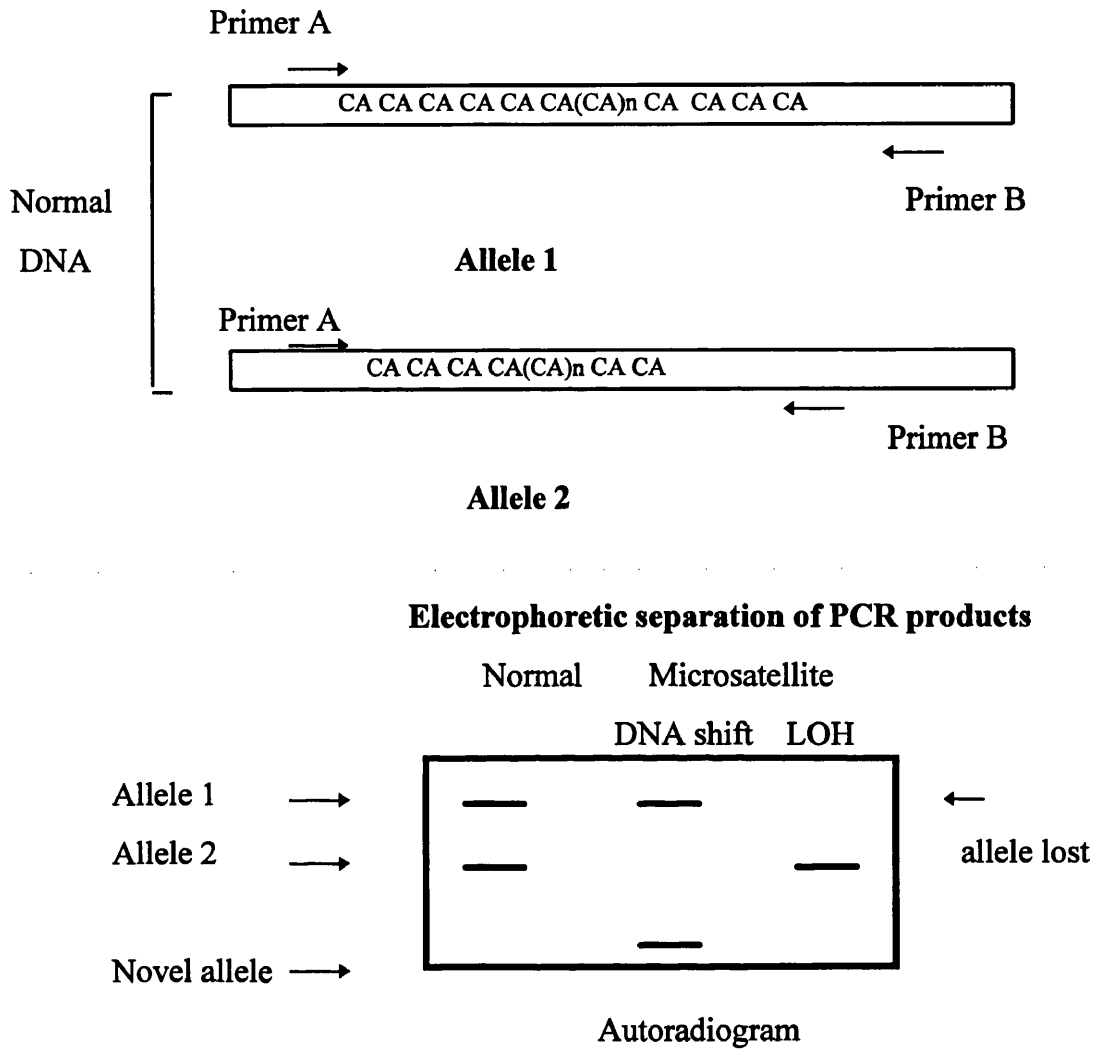


Figure 1. 2: Microsatellite alteration can be detected by PCR-based microsatellite analysis. The occurrence of an insertion or deletion mutation in the repetitive sequence alters the size of one allele and creates a microsatellite shift or novel allele on the autoradiogram. LOH: loss of heterozygosity.

Shibata et al. (1994) found several human tumour cell lines with up to 1,000-fold increases in mutation rates for endogenous microsatellite sequences, relative to normal or tumour cells without the mutator phenotype, and showed that genomic instability occurs very early in tumorigenesis. They showed MI in several single cell lines *in vitro* and *in vivo.* They examined three different adenomas, which were found adjacent to a carcinomas, and found MI in all of them. Their results demonstrated that MI occurs very early in colorectal carcinogenesis and continues to take place during tumour evolution and progression. However, it has been documented by many groups that MI

occurs very early in tumorigenesis (Aaltonen et al. 1994; Patel et al. 1994; Melo et al. 1994). Figure 1. 3 demonstrates genomic instability in early stage of tumorigenesis.

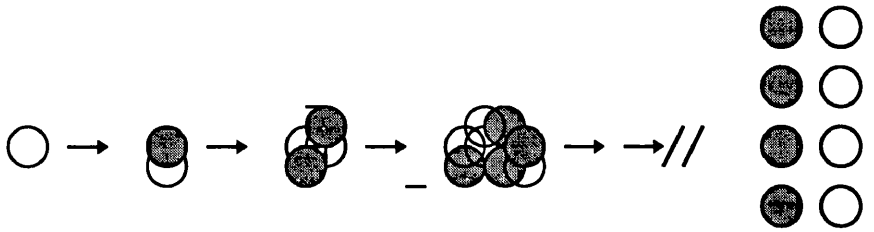


Figure 1. 3. Genomic instability occurs in the early stage of tumorigenesis. Open circles represent normal cells, shaded circles represent cells with unstable DNA. If genomic instability occurs at the two cell stage 50% of tumour cells have unstable DNA. In the case of mutation at later stages the proportion of cells with genomic instability is smaller.

1. 2. 2. Microsatellite instability in other diseases

Microsatellite instability is a common feature in HNPCC and in some of the sporadic CRC and related diseases such as endometrial cancer (Risinger et al. 1993; Burks et al. 1994; Peltomaki et al. 1993), gastric cancer (Strickler et al. 1994; Chong et al. 1994; Han et al. 1993; Mironov et al. 1994; Rhyu et al. 1994), pancreatic cancer (Han et al. 1993), prostatic cancer (Gao et al. 1994) and also in breast (Patel et al. 1994; Peltomaki et al. 1993), bladder cancer (Gonzalez et al. 1993; Dalbagni 1993), leukaemia (Cross et al. 1994), Muir-Torre syndrome [(Muir-Torre syndrome is characterised by the presence of at least one sebaceous tumour and visceral malignancy) (Honchel et al. 1994)], head and neck cancer, and also lung cancer (Mao et al. 1994), small cell lung cancer (Merlo et al. 1994), non-small cell lung carcinoma (Shridhar et al. 1994). MI has been also reported in primary neoplasms from HIV+ patients (Bedi et al. 1995). MI has not been reported in testicular cancer so far. Mutations within small repeated sequences in the TGF beta type II receptor have been found in several cell lines, and nude mouse xenografts of tumours of the microsatellite mutator phenotype (Benson et al. 1995; Markowitz et al. 1995). This finding suggests that the TGF beta receptor mutation is a common target for diverse mutator genes in colon cancer, and also aborted embryos (Kiaris et al. 1996). Recently, MI was reported within the coding sequence of the insulin like growth factor II receptor (IGFIIR) from patients with ulcerative colitis associated colorectal cancer (Souza et al. 1996). These alterations are most easily detected by analysis of microsatellite polymorphic markers, which seem particularly prone to

insertions or deletions in such tumours. In the hereditary form of colon cancer, MI is generally due to germ-line mutations in one of four mismatch repair (MMR) genes (Liu et al. 1995a). However, mutations in mismatch repair genes have so far been demonstrated in only a small percentage of sporadic cancers or cell lines derived from colorectal carcinomas (Liu et al. 1995; Borresen et al. 1995).

1. 3. The structure of the *hMSH2* and *hMLH1* genes

The sequences of the *hMSH2* cDNA were submitted to the gene bank by accession number U0405. The *hMSH2* gene spand about 73 kb in the genome and has a 2.8 kb transcript with 16 exons. The largest and smallest exons are exon 3 and exon 11 which are 279 bp and 98 bp, respectively. The largest and smallest introns are intron 8 and intron 12 which are 17 kb and 1.0 kb respectively (Figure 1. 4). The initiation codon (ATG) is located within exon one at nucleotide (nt) position 69, and the termination codon (TAG) is located in exon 16 at nt position 2892. The hMSH2 protein is one member of the mismatch repair system, and has a 100 kd molecular weight. The role of this protein is to distinguish a G-T mismatch in the heteroduplex. Comparison of the hMSH2 amino acid sequences with the other known *mutS* homologues has shown that it was most highly related to the *S. cerevisiae* MSH2 protein (Rennan and Kolodner; 1992), and that these two amino acids sequences are more highly related to each other than many pairs of bacterial sequences. The evolutionary relationship of the known mutS-related proteins indicates that MSH2 is a member of a group of *mutS* homologues that includes all of the bacterial homologues, and *S. cerevisiae msh1* and *msh2*. This group of proteins contains all of the members known to play a role in mismatch repair, and in which mutations of the respective genes have been found to cause a strong mutator phenotype in these respective organisms.

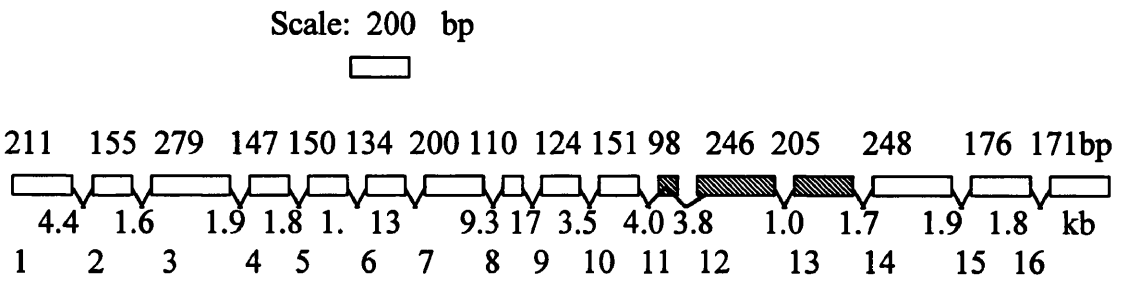


Figure 1. 4- Diagram of the organisation of the *hMSH2* locus. The boxes represent the individual exons from 1 to 16. The size of each exon is given at the top of each one (bp) and the size of each intron (kb) is given below the region between individual pairs of exons. Shaded boxes represented conserved region in the gene.

After cloning and mapping the first gene to the chromosome 2p, Papadopoulos et al. (1994) searched for additional homologous of bacterial and yeast mutator genes. The results of this study indicated that the *hMLH1* gene was located within band 3p21.3. This gene spand in about 85 kb in the genome and has a 2.7 kb transcript with 19 exons. The biggest exon has 473 bp (exon 12) and the shortest exon is exon 7 with 194 bp (Wijnen et al. 1996). The schematic diagram of this gene is presented in Figure 1.5.

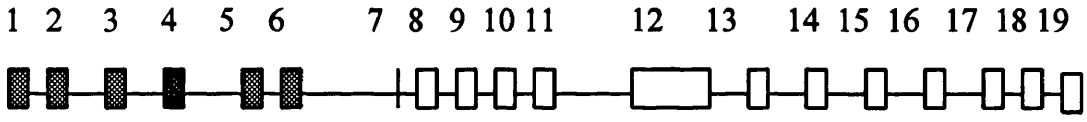


Figure 1. 5. Schematic presentation of the genomic organisation of human *hMLH1* gene. Exons and introns are presented by boxes and horizontal lines respectively. Shaded boxes represent the conserved region in the gene. The Figure is not to scale.

1. 4. DNA repair

Not all interactions of chemicals and irradiation with DNA produce mutations. In fact, all cells have efficient repair mechanisms that repair such lesions. DNA repair mechanisms include sets of enzymes that survey DNA for specific kinds of damage, remove the altered portion of DNA, then restore the correct nucleotide sequence. The important role of DNA repair in human cancer has been established by the finding that a number of inherited defects in DNA repair systems predispose individuals to getting cancer. Genetic instability is a component of the familial cancer phenotype. For example, patients with xeroderma pigmentosum, ataxia telangiectasia, and Bloom's syndrome have a genetic instability that results in a predisposition to cancer (Mckinmen et al. 1987., Cleaver et al. 1968., German et al. 1989). These syndromes are inherited as autosomal recessive diseases, but there is no reason that a dominant gene could not produce a similar result.

DNA repair is an important molecular defence system against agents that cause cancer, degenerative disease, and ageing. The faithful transmission of genetic information is paramount to the survival of a cell, an organism, and a species. There are at least three different ways in which mismatch nucleotides arise in DNA. First, physical damage to the DNA can give rise to mismatched bases. For example, the deamination of 5-methylcytosine creates thymine and therefore a G-T mispair (Friedberg, 1985). Secondly, misincorporation of nucleotides during DNA replication can yield mismatched base pairs, nucleotide insertions and deletions (Modrich,1991). Thirdly, genetic recombination produces regions of heteroduplexes resulting from the pairing of two different parental DNA sequences. Mismatched nucleotides produced by each of these mechanisms are known to be repaired by a specific enzyme system (Friedberg,1990; Modrich, 1991). The results of the first and third ways are DNA adducts, while the result of the second way is misincorporation of nucleotide in the DNA.

DNA repair in *Escherichia coli*

Since mutations are the primary cause of inheritable disease and cancer, they may also contribute to the ageing process, and it is important for individual cells to preserve the integrity of their genetic material. Study of the molecular systems responsible for maintenance of genetic integrity became available more than 20 years ago with the identification of *Escherichia coli* mutants that display a marked increase in the rate of spontaneous mutation. The genetic defects in this strain are known to inactivate mutation avoidance systems that are critical for the maintenance of genetic stability.

Several different DNA mismatch correction pathways have been identified in *E. coli.* These are important pathways of mutation correction. Enzyme systems that recognise and process mispaired bases have been identified in both prokaryotes and eukaryotes. *E. coli* possesses multiple mismatch repair pathways that are distinguished on the basis of mispair specificity and size of excision repair tracts associated with correction. The products of four genes, the mutH, mutL, mutS and mutU proteins are essential for *E.coli* methyl-directed mismatch repair, a system that ensures the precision of both chromosome replication and genetic recombination. The importance of this fidelity device has been demonstrated by the fact that inactivation of the corresponding human pathway is the primary cause of certain types of cancer.

1. 4. 1. Long - patch mismatch correction

The best defined mismatch repair pathway in *E. coli* is mut HLS pathway that promotes a long patch (approximately 2 kb) excision repair reaction that is dependent on the mutS and mutU (uvrD) gene products (Figure 1. 6). This is a methyl directed system and displays a broad specificity for different mispairs. Efficiency of correction efficiency depends on the nature of mismatch and can also be influenced by the sequence in which the mispair resides. This pathway recognises the state of methylation of each DNA strand at d(GATC) sequences and directs repair to the transiently undermethylated nascent DNA strand, using the methylated parental strand as a template. A heteroduplex bearing d(GATC) modification on both DNA strands is not a substrate for methyl-directed repair. Hemimethylated DNA is corrected on the unmethylated strand, with the modified strand serving as a template, and unmethylated molecules are corrected on one strand or the other. Mismatches corresponding to base- base mispair, insertion/ deletion of a few nucleotides in one strand of the helix are also subject to efficient methyl-directed repair *in vivo*. The importance of this is that methylated strand is recognised as the original sequence, thus mutations do not arise.

1. 4. 2. Initiation of mismatch correction

Ten different proteins and enzymes are involved in mismatch correction. These are products of the mutH, mutL, mutS genes, DNA helicase II, single strand DNA binding protein (SSB), exonuclease I, exonuclease VII, RecJ exonuclease, the replicative DNA polymerase III holoenzyme, and DNA ligase. Of the proteins involved in methyl-directed repair, only mutS is capable of specific interaction with a heteroduplex in the absence of other components. The multifunctional 97 kd product
of the muts gene binds to mispairs. The mismatch is recognised by this protein. MutS also has weak ATPase activity in the presence and absence of DNA and promotes the formation of alpha-shaped loop structures from heteroduplex DNA.

MutH protein interacts with d(GATC) sites in DNA that are hemimethylated on the adenine, and is responsible for incision on the unmethylated strand. Analysis of a new homogenous preparation of the 25 kd mutH protein revealed a single activity, an extremely weak, Mg^{2+}-dependent endonuclease that incises the unmethylated strand of hemimethylated DNA 5' to the G of d(GATC) sequences.

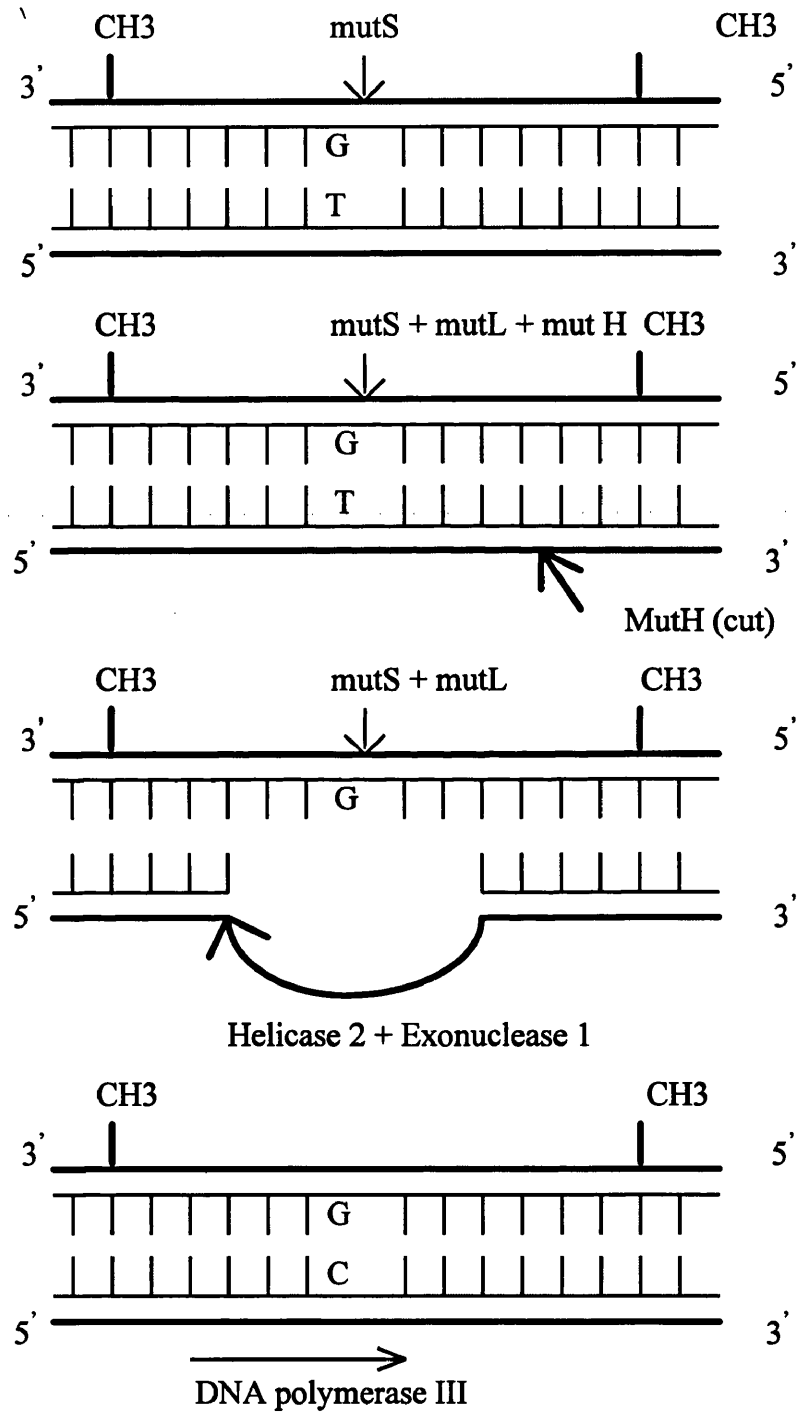


Figure 1. 6. Schematic diagram of the DNA mismatch repair system in the *E.coli*. Note that hemimethylation is on the template strand only.

No simple activity has been attributed to mutL, in contrast to mutS and mutH. MutS binding to a mismatch protects about 20 base pairs against DNaseI attack, but in the

presence of mutL and ATP, the protected region increases to about 100 base pairs. This effect is independent of the presence of a d(GATC) site within the heteroduplex. MutL is homodimer (70 kd)2 protein. Molecular interactions responsible for the mutL dependent effects observed in vitro have not been defined. One possibility is that mutL functions as a protein-protein interface between mutS and mutH.

1. 4. 3. The excision reaction

Three different proteins, mutS, mutL, and mutH act in a co-ordinated fashion to make a single strand nick at a hemimethylated d(GATC) sequence in the newly replicated strand. MutS binds to the mismatch and forms a complex with mutL that brings the unmethylated d(GATC) into close proximity to the mismatch, leading to cleavage by the mutH nuclease. Excision proceeds in either the 3' to 5' or the 5' to 3' direction from the nick to the mismatch. Several *in vivo* observations also suggest bi-directional action, and this has been confirmed by biochemical findings (Modrich, 1991). When the unmethylated d(GATC) sequences located 3' to the mismatch, a 3' to 5' hydrolytic activity is necessary. Exonuclease I is sufficient in this respect (Lahue et al. 1989; Lehman et al, 1964). However, if an unmethylated d(GATC)sequence resides 5' to the mismatch, excision is dependent on a 5' to 3' exonuclease VII or RecJ (Chase et al. 1974; Lovett et al. 1989).

1. 4. 4. Repair and replication fidelity

Now, it is clear that methyl- directed mismatch correction contributes in a major way to the fidelity of chromosome replication in *E.coli* (Modrich, 1991). Mutations in mutH, mutL, mutS or mutU result in a 100- 1000 fold increase in spontaneous mutability (Cox, 1976). Mutations leading to a deficiency in the Dam methylase, the enzyme responsible for d(GATC) methylation, also result in hypermutability (Marinus,1984; Marinus *et al*, 1974). Overproduction of this enzyme also leads to a mutator phenotype (Herman *et al*, 1981; Marinus *et al* , 1974). Mutations in *mutDS* have been attributed to the response of the methyl-directed pathway to DNA biosynthetic errors. The *mutDS* mutations, which affect the proof-reading exonuclease activity of DNA polymerase III holoenzyme, is one of the strongest mutators known, increasing spontaneous mutability by factors of 1,000 to 100,000 (Cox, 1973). This large increase in mutation rate is paradoxical, in view of the *in vitro* demonstration the

exonuclease editing increases fidelity of the polymerase III reaction by a factor of only 10- 200 (Fresht *et al*, 1983). It has been clarified that in addition to the associated deficiency in DNA polymerase III proof-reading activity, *mutDS* strains are also defective in methyl-directed mismatch correction (Schaaper, 1988; Damagnez et al. 1989). So, DNA polymeraseIII holoenzyme participates directly in methyl-directed mismatch repair.

1. 4. 5. Very short - patch mismatch repair

Mismatch- repair systems that are characterised by short excision repair tracts, about 10 nucleotides or less, have been discovered in both prokaryotes and eukaryotes. One of these short pathways is named VSP (Very Short Pathway), and was identified for the first time in prokaryotes by Lieb (1981). This repair system is able to correct a G to T mismatch to G-C. The existence of this system has been confirmed by heteroduplex transfection experiments (Zell *et al*, 1987), which have shown only G-T in this system being corrected to G-C.

5'- CTWGG 5'- CCWGG

3'- GG WCC 3'- GGWCC

W indicates A or T, the d(CCWGG) is the recognition site for the Dam DNA cytosine methyltransferase of *E.coli* , which modifies the internal cytosine residue within the sequence. These sites are hot spots for mutation. The VSP pathway functions as a mutation avoidance system to restore G-T mismatches, which are generated via 5-methylcytosine deamination, to G-C base pairs (Zell, 1987; Jones *et al*, 1987).

Biological experiments have implicated four proteins in the VSP reaction, i.e. mutL, mutS, DNA polymerase I and the product of the *vsr* gene, which is adjacent to that encoding the Dam methylase (Dzidic *et al*, 1989). Since mutL and mutS function in both methyl- directed and VSP repair, interaction of these proteins with a G-T mispair in the d(CCWGG) sequence context may serve to direct repair to the appropriate pathway (Modrich, 1991). Although no information on the mechanism of VSP repair is available, it has been observed that overproduction of VSP proteins partially overcomes the dependence of the reaction on mutL and mutS (DasGupta et al. 1982). This suggests that the VSP product may initiate the chemical steps involved in VSP correction, with mutL and mutS serving a regulatory function (Modrich, 1991).

1. 4. 6. Mut Y mismatch repair in *E. coli*

Both in vitro and *in vivo* experiments have demonstrated the presence of a second short- patch repair system in *E.coli.* The specificity of this system is for G-A mismatches, which are repaired to G-C base pairs (Lu et al. 1988). Transfection experiments also showed that the mutY mismatch repair system can repair A-C to G-C (Radicella *et al*, 1988). This pathway is independent of mutH, mutL, mutS and DNA helicaseII (Su et al. 1988), and does not require the presence of d(GATC) sites within heteroduplex (Au et al. 1988). A failure to process G-A mismatches to G-C base pairs can largely account for the mutator phenotype associated with loss of *mutY* function. The *mutY* gene produces a 36 kd protein in near homogeneous form and it has been demonstrated that it functions as an adenine glycosylase specific for G-A mispair (Au et al. 1989). The hypermutability associated with *mutY* mutations (Nghiem et al. 1988) and the associated defect in G-A to G-C mismatch repair suggests that this pathway corrects G-A biosynthetic errors in which the mismatched adenine is on the new DNA strand. Since G-A mispairs are also subject to methyl- directed correction, it is evident that *E. coli* can use multiple mechanisms to exclude this mismatch from newly synthesised DNA. This may reflect the multiple conformation states uniquely available to this mispair.

1. 4. 7. Rec F- dependent system in *E.coli*

This system was discovered by Fishel and Kolodner (Fishel and Kolodner, 1989; Fishel et al. 1986) as a short-patch repair system in *E. coli*. The *RecF* system is able to repair short insertions and deletions in heteroduplex DNA. The mechanism of this pathway is not clear yet, but it is independent of *mutH*, *mutL*, *mutS* and DNA helicase II (Fishel et al. 1986), and it functions on DNA containing symmetrically modified d(GATC) sites.

1. 5. Linkage analysis of HNPCC susceptibility genes

Early reports of the existence of human mismatch repair systems were published by Hare and Taylor (1985) and Glazer et al. (1987). These reports followed the demonstration that extracts of human cells contained mismatch binding proteins (Jiricny et al. 1988; Stephanson & Karran 1989). These data was supported by Holmes et al. (1990) and it was shown in a bi- directional nick directed fashion (Thomas et

al. 1991) of the *E.coli* system (Modrich 1991). The first evidence of the existence of mismatch repair deficient mutants came from the studies of Karran et al. (Branch et al. 1993), who demonstrated that two mutator cell lines resistant to methylating agents lacked a major mismatch binding activity. These finding were confirmed by experiments showing that the methylation tolerant human mutator cell line is deficient in mismatch repair *in vitro* (Kat et al. 1993). High mutating strains of *E.coli* and *Salmonella typhimurium* were originally isolated by Treffers et al. (1954) and Miyake et al. (1960). Another mutator strain in *E.coli* also discovered by Siegel and Bryson (1964). Thus, the occurrence of a high mutation rate in some strains of bacteria and yeast has been described a long time ago. A similar observation has been reported in *Drosophila* (Plough et al. 1941). Reenan and Kolodner (1992a, b) reported two *mutS* homologous genes, *msh1* and *msh2* in yeast, whose amino acid sequences are highly homologous to *mutS* in *E.coli.* The *msh1* and *msh2* are involved in mismatch repair in the mitochondria and nucleus, but it was shown that the disruption of *msh2* mutant in yeast elevated the rate of spontaneous mutation 70-100 fold over that of wild type. The similar results have been reported in bacteria for *mutS* (Siegel et al. 1963; Siegel et al. 1974; Cox et al. 1972) and also for *mutL* (Siegel et al. 1975). Thus, this group of genes is well known in bacteria and in yeast. Based on this knowledge, two different groups in 1993 used two different methods for mapping candidate gene(s) for CRC. When genomic instability was observed in the tumour of HNPCC kindreds, a link between mutator phenotype in *E.coli* and genomic instability in the tumour of HNPCC kindreds was immediately established. Therefore a search for a gene with predicted function was started. Vogelstein's group used a panel of human-mouse and human-rat hybrid cell line followed by the FISH technique (Leach et al. 1993). Kolodner's group (who found the *mutS* gene in yeast) used 360 bp PCR products from the *mutS* gene as a probe, and used both *S.cerevisiae* DNA and human cDNA as templates. The resulting 360 bp fragment was purified and a number of independent clones were obtained and sequenced (Fishel et al. 1993). The first HNPCC susceptibility locus was mapped and assigned to chromosome 2p by linkage analysis by Peltomaki et al. (1993). Thus in attempting to clone the gene, positional cloning strategies were combined with the search for a candidate gene with a DNA repair function. A human homologue of the *mut*S gene in bacteria (Modrich, 1991) and *msh2* gene in yeast (Reenan et al. 1992) was cloned (Fishel et al. 1993; Leach et al.

1993) and named *MSH2* (mut S Homologous). They chose two large kindreds with HNPCC, families C and J, for linkage analysis. The C kindred originated from North America and the J kindred from New Zealand. The mean ages of onset of CRC were 41.1 years (SD= 13.1) and 44.4 years (SD= 15.5), respectively. Extracolonic cancers were observed in both families. They began a systematic search through the whole genome with 345 informative microsatellite markers. Marker D2S123 [AFM093xh3] showed a highly significant two-point lod score of 6.39 at a recombination fraction of zero in family C and 1.45 in family J. There was no recombination between HNPCC and D2S123 in either family, suggesting very close linkage. With two flanking markers D2S119 (Distal to D2S123) and D2S136 (proximal to D2S123), recombination was observed in both families. Further analysis by using 96 markers from chromosome 2 placed locus D2S123 at a position 5 cM distal to D2S5, which has been mapped to 2p15-16 by in situ hybridisation, linkage and somatic cell hybrid analysis. The localisation of an HNPCC gene to chromosome 2 in kindreds from two different continents proved the existence of a genetically determined form of this disease. Aaltonen et al. (1993), studied an additional 14 smaller kindreds and linkage analysis excluded in 3 families by lod scores less than
-2.0, suggesting genetic heterogeneity.

To determine whether mutations of other mismatch repair genes could cause HNPCC Papadopoulos et al. (1994) searched for an additional homologues of bacterial and yeast mutator genes by scanning a database of human genes. In this search three genes were identified that had significant similarity to the yeast *mutL1* and was therefore termed *hMLH1* (mut L Homologous). The other genes had slightly greater similarity to the yeast *mutL* homologue *PMS1* (Post Meiotic Segregation) and were denoted *hPMS1* and *hPMS2* respectively.

The sequence of these three clones was used to design primers for PCR reaction assays with human genomic DNA as template. Somatic cell hybridisation studies localised the *hMLH1*, *hPMS1* and *hPMS2* genes to chromosomes 3, 2 and 7, respectively. The *hMLH1* gene was of particular interest because markers on chromosome 3p21 had been linked to HNPCC in several families. To determine the precise chromosomal location of *hMLH1*, a genomic DNA clone was used for fluorescence in situ hybridisation to human chromosome spreads. Of ~50 spreads analysed, 40 had a doublet signal on at least one

chromosome 3 homologue. Doublet signals were not detected on any other chromosomes. Detailed analysis of 17 individual spreads indicated that the *hMLH1* gene located within band 3p21.3. Detailed linkage analysis of HNPCC families showed that the responsible locus is centred at markers D3S1611 and D3S1277, and is delineated by a 1-cM interval bordered by markers D3S1298 and D3S1561. No recombinations were observed between markers D3S1611 or D3S1277 and cancer susceptibility. The *hMLH1* gene was <85kb and became an excellent candidate for the HNPCC gene on chromosome 3p21. The sequence of the hMLH1 gene was derived from cellular RNA, an open reading frame (ORF) began 42 nucleotides downstream of the 5' end of the cDNA, and contained 2268 nucleotides. Figure 1. 4 summarises relevant data on the four human DNA mismatch repair genes whose germ-line mutations have been implicated in HNPCC.

The discovery of sequence homologies between the *pms1* gene of *S.cerevisiae* and the *E.coli* mismatch repair gene *mutL* (Kramer et al. 1989), as well as between the *E.coli mutS* gene, the mouse *msh3*(Linton et al. 1989) and human *hMSH3*(Fujii and Shimada, 1989), implied that the mismatch repair systems of all organisms are closely related (below Figure). These latter findings were key to the identification of the human mismatch repair genes.

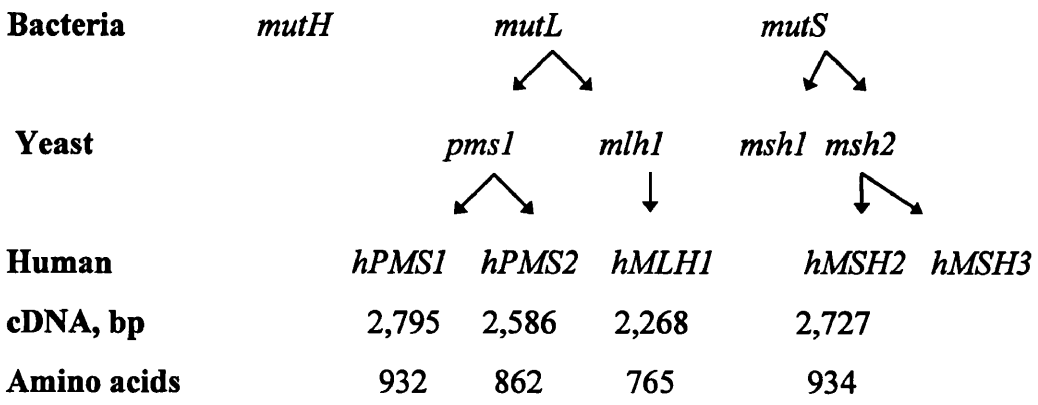


Figure 1. 7 Presentation of mismatch repair genes from bacteria to human, their cDNA length and the number of corresponding amino acids is shown.

1. 6. The molecular pathology of single gene disorders

A mutation is any change in the nucleotide sequence of a genome. The process of

mutation is mutagenesis and an organism showing an altered phenotype as the result of a mutation is a mutant. Not all mutations produce recognisable mutants, some mutations will occur in non-coding regions of the genome and have no detectable effect on the phenotype. These silent mutations are probably frequent in higher eukaryotes, where the larger part of the genome is located in the non-coding sequence. Mutations can occur at any stage of development and may give rise to patches of mutant tissue. The size of the patch depends upon the time during development at which the mutation occurred, but these somatic mutations do not affect the gametes and can not be passed on to the next generation. However, in the case of germ-line mutations in the sex cells, mutant gametes are produced and the mutation can be transmitted to the next generation. However, some environmental factors such as ionising radiation and chemicals can increase the rate of mutation. In some parts of the genome, due to presence of methylated cytosine residues, the rate of mutation is high and deamination of this base produces thymine (T), which is not recognisable by the DNA repair system.

Point mutations involve a single nucleotide pair, or sometimes a few adjacent nucleotide pairs, within the DNA. There are two types of point mutations which affect gene expression in different ways:

(i) **Base substitutions**: A particular base pair is replaced by a different base pair. These substitutions are known as transitions, where a purine (A, G) replaces for purine and a pyrimidine (C, T) replaces for pyrimidine, and transversion when a purine replaces for pyrimidine, and vice versa. When one base is changed by another, there are two possible choices for transversion, but only one for transition. The base thymine can undergo two possible transversions (T to G, or T to A) but only one chance for transition (T to C).

(ii) **Frame shift mutations**: Since proteins are encoded by a triplet of nucleotides the loss or insertions of one, two or four nucleotides in the coding region of a gene will alter the reading frame out of sequence. Therefore substitutions, deletions and insertions could be the cause of frame shift mutations.

The effect of mutations on the reduction of gene products will be described below:

1. 6. 1. Mutations which affect transcription

(i) Promotor mutation: It has been shown that several point mutations in the promotor region of a gene either within or adjacent to the promotor sequence could cause reduction

of a gene product. These mutations emphasises the importance of these highly conserved regions of DNA and confirm their proposed promotor function.

(ii) Inversion: This occurs when a region of DNA is back -to- front with respect to its normal orientation in the genome. In a patient with the phenotype of δβ thalassemia, in which there is no δ or β globin chain production, an inversion was found. The inversion involves a region of DNA between the δ and γ globin genes. There is also a small deletion at each end of the inversion.

(iii) Deletion: In most types of α^{o} thalassemia both α globin genes are lost. In α^{+} thalassemia there is a deletion involving chromosome 16 which leaves a single functional α gene. Duchenne muscular dystrophy (DMD) and haemophilia are further phenotypic examples of this type of mutation. The main cause for gene deletion is unequal crossing over, which occurs in meiosis or mitosis. There are many genes as member of multigene families such as colour vision gene family and also alpha and beta globin gene clusters. It is possible that two pairs of homologous chromosomes misalign at meiosis or mitosis. Unequal crossing over in this situation between mispair chromosomes or sister chromatids can lead to gene deletion in one chromosome and gene duplication in another chromosome. Duplication of several exons in familial hypercholesterolaemia and also unequal crossing over in Alu repeating on the short arm of the X chromosome and Y chromosome are further examples of this type of mutation (Lehrman et al. 1987; Rouyer et al. 1987).

(iv) Insertion: When one or many nucleotides are inserted at different positions along the gene, an abnormal protein results. It has been found that lipoprotein lipase deficiency (LPL) may result from gene insertions.

(v) Fusion genes: This type of mutation is a result of abnormal crossing over between two non-homologous chromosomes. For example, in colour vision and colour blindness, the genes for the red and green pigments lie in a tandem array on the X chromosome, on which there is a single red pigment gene and variable numbers of green pigment genes. Many of the different forms of the red/green colour blindness result from unequal crossing-over between the red and green pigments with the production of different numbers of fusion genes (Weatheral 1992).

1. 6. 2. Mutations which affect mRNA splicing

This type of mutation has been reported in many diseases such as retinoblastoma, neurofibromatosis type 1, colon cancer, acute intermittent porphyria, thalassemia, breast cancer and other disease. However, these types of mutations could be divided into the following groups:

(i) splice junction mutations, **(ii)** cryptic splice site mutations in intron and also in exon, **(iii)** consensus sequence mutations and **(iv)** mutations in poly A residues.

1. 6. 3. Mutations which affect translation

The ultimate product of a gene is protein, produced in the cytoplasm by a process called translation. The following mutations have an effect on this process:

(i) Mutations at initiation codon: Mutation at initiation codon could result in no product from the affected gene.

(ii) Nonsense mutations: A number of point mutations, deletions or insertions have been reported in different genes which, could make the transcription of a normal gene product impossible. All of above mutations are able to produce a premature termination codon, resulting in the synthesis of a truncated protein.

(iii) Frame shift mutations: These types of mutations are potentially pathologic, as they cause a shift in the translational reading frame of the corresponding proteins. By altering the reading frame downstream of the mutation, the polypeptide is often truncated due to the formation of a premature termination codon. This type of mutation has been reported in many genes.

(iv) Termination codon mutations: If point mutation occurs in the stop codon, a longer peptide will result. In this situation, a number of nucleotides which are normally located in the untranslated region would integrate in the coding sequence. A new termination codon, a few nucleotides down stream of the mutation, will be created.

1. 6. 4. Repetitive genomic mutations and disease

In the past five to six years a number of diseases have been characterised by the expansion of a trinucleotide repeat within the coding and non-coding sequences of the genes. Examples of this type of disease are presented in Table 1. 1. In some of these disease the trinucleotide repeats are within the coding sequence, for example: Huntington disease (HD), Spinal and bulbar muscular atrophy (SBMA), or Kennedy

Disease	Location of gene	Location of repeat	Repeat sequence	Normal length	Full length	Reference
Huntington disease	4p	Coding	(CAG)n	9-35	37-100	*
Kennedy disease	xq	Coding	(CAG)n	17-24	40-55	La Spada et al. 1991
Spino-Cerbellar ataxia 1(SCA1)	6p	Coding	(CAG)n	19-36	43-81	Orr et al. 1993
Dentatorubral pallidoluysian	12p	Coding	(CAG)n	7-23	49->75	Nagafuchi et al. 1994
Machado-Joseph disease (MJD,SCA3)	14q	Coding	(CAG)n	12-36	67->79	Kawaguchi et al. 1994
Fragile X site A (FRAXA)	Xq	5' UTR	(CGG)n	6-54	200->1000	Fu et al. 1991
Fragile X site E (FRAXE)	Xq	?	(CCG)n	6-25	>200	
Fragile X site F (FRAXF)	Xq	?	(GCC)n	6-29	>500	
Fragile 16 site A (FRA16A)	16q	?	(CCG)n	16-49	1000-2000	
Myotonic dystrophy (MD)	19q	3' UTR	(CTG)n	5-37	50-4000	Fu et al. 1992
Friedreich ataxia	9q	Intron	(GAA)n	7-22	200->900	Campuzano et al. 1996

Table 1.1: Presentation of different diseases caused by trinucleotide repeat expansions.

*: The Huntington's disease collaborative group, 1993

disease, Dentatorubral pallidoluysian (DRPLA), Machado-Joseph disease (MJD). In some, such as Myotonic dystrophy (DM), and Fragile X syndromes (FRAX), the repeats are located in the non-coding region of the gene.

In patients with MD there is a (CTG)n repeat in the 3' untranslated region of a gene encoding a protein kinase. This repeat is polymorphic and varies in size between 5

and 37 repeats in unaffected individuals, whereas in affected patients there are between 50 and 4000 CTGs. Relatively little is known about the precise mechanism of the mutation process although it has been reported that in FRAX syndrome full mutation does suppress the transcription process, resulting in the absence of the encoded protein (Mc Conike-Rosell et al. 1993) which is associated with mental retardation in affected individuals.

1. 7. Mutation detection methods

The identification of mutations in DNA is an important goal in research, and also in diagnostic laboratories. Based on the nature of mutations two major groups of techniques can be used: (i) identification of known mutations and (ii) identification of unknown mutations.

1. 7. 1. Identification of known mutations

The presence of known mutations can be searched using the following methods:

1. 7. 1. 1. Allele-specific oligonucleotides (ASO)

This technique is based on the hybridisation between target DNA and oligonucleotide probes. This method takes advantage of the difference in stability of hybrid probes with target DNA that may or may not match perfectly (Wallace et al. 1979). Two oligonucleotide probes with about 20 base pairs are required. One has normal sequences and the other one has mutant sequences. The normal probe hybridises to the normal DNA but not to the mutant DNA and this is true for the mutant probe. Both radiolabelled or fluorescent labelled probes can be used. This technique does not require electrophoresis but two normal and mutant probes are required. This technique has been applied to many diseases such as thalassemia and haemophilia.

1. 7. 1. 2. Amplification refractory mutation system (ARMS)

This technique was described for the first time by Newton et al. in 1989. It was based on the fact that PCR elongation requires a specific 3' end and DNA *Taq* polymerase lacks the 3' exonucleolytic proof reading activity. In practice, two independent PCR amplifications for each patient with a wild-type or mutant primer and a common primer should be performed. Internal primers are added in the same reaction as a control. By single PCR amplification a result is obtained. This is a very good method for detection of common mutations. By using different mutant primers many

mutations can be screened in a single PCR reaction.

1. 7. 1. 3. Artificial introduction of restriction sites

This technique is based on the fact that mutations can usually create or abolish sites for a particular restriction enzyme. Otherwise, a restriction site can be introduced to the 5' end of the primer. (Cohen and Levinson, 1988). In this method an artificial restriction site is introduced at the 5' end of the primer, with one base pair mismatched close to the site of mutation. The presence of the mutation will make the proper recognition site for a particular enzyme. Digestion of the PCR products followed by electrophoresis will identify the presence or absence of the mutation in DNA.

1. 7. 2. Identification of unknown mutations

The choice of the most appropriate screening technique is influenced by the size and structure of the gene and also the nature of the mutation in question. The spectrum of mutation ranges from visible chromosome rearrangements to single base substitutions. Therefore, the identification of unknown mutations is divided into two different groups;- (i) identification of large alterations and (ii) identification of single base change and/or small sequence alterations.

1. 7. 2. 1. Identification of large gene alterations

A large gene alteration is a mutation in which more than 500 bp of the gene is rearranged (Grompe, 1993). This type of mutation could be detected by the following methods:-

Molecular cytogenetic techniques:

In the last few years cytogenetic analysis has been improved by using fluorescent *in situ* hybridisation (FISH) techniques and also comparative genomic hybridisation (CGH). By using these techniques, micro-deletions or insertions and also aneuploidy could be detected. In brief, FISH labelled DNA probes are hybridised to the chromosome in interphase or to metaphase spreads. This technique is particularly suited for the detection of chromosomal rearrangements that can not be detected by conventional cytogenetics. CGH is a modified FISH technique in which differentially labelled test and reference DNAs are co-hybridised on normal target metaphase chromosomes to assay copy number differences (Kallionieemia et al. 1992).The smallest possible chromosome abnormality that CGH will detect has not yet been defined. This technique has the

potential to provide genome wide screening for complex chromosome abnormalities, particularly in solid tumours, in a single experiment and does not require metaphase spread from affected individuals.

Southern blot hybridisation

This technique was developed by E. Southern in 1975 and is based on DNA digestion by restriction enzymes followed by agarose gel electrophoresis and then hybridisation with radiolabelled probe. By using the appropriate enzyme and probe large alterations in the gene may be detected by the presence or absence of extra bands or alterations in band intensities, in the case of heterozygous status in autosomal dominant diseases. No detailed of knowledge of the structure of a gene is required. This technique has been used to detect trinucleotide repeat expansions in patients with myotonic dystrophy and also in Fragile X (Fu et al. 1991, Harly et al. 1992).

Pulsed field gel electrophoresis (PFGE)

This technique is in principle, very similar to Southern blotting. Gel electrophoresis in conventional Southern blotting can not separate DNA fragments more than 20 kb. Larger fragments, of 100-500 kb, can not pass through the pores of the gel. This method is based on the periodical change of the orientation of the electrophoretic field. Therefore, the large molecules must re-orientate and find a path through the gel matrix in other direction. The larger molecules require longer time to re-orientate than shorter ones. This allows the resolution of DNA fragments of 100-1000 kb in size, and even larger. The resultant gels can be blotted and hybridised to the appropriate probe. PFGE allows the identification of major deletions and also chromosomal rearrangement in a single experiment.

Multiplex PCR

The structure and sequence of a gene is required for this technique. In multiplex PCR, amplification of several segments (usually up to 5-6 segments) could be carried out in a single PCR reaction. Deletions are indicated by the absence of some of the bands in the multiplex pattern in homozygous status. In autosomal dominant diseases deletions are seen as 50% reduction of band intensities in a quantitative analysis of the

multiplex PCR reaction. Multiplex PCR has been used in screening for deletions in DMD in which 60% of cases represent deletions (Grompe, 1993).

1. 7. 2. 2. Identification of single base or small sequence alterations

These methods include: single strand conformational polymorphism (SSCP)(Orita et al. 1989), heteroduplex analysis (HA) (White et al. 1992), denaturing gradient gel electrophoresis (DGGE) (Fischer and Lerman, 1983; Myers et al. 1985), chemical cleavage of mismatch (CCM) (Cotton et al.1988), protein truncation test (PTT) (Roest et al. 1993) and direct DNA sequencing (Gyllensten, 1989).

Single strand conformational polymorphism (SSCP)

The SSCP is based on the migration of single stranded DNA (ssDNA) structure(Orita et al. 1989). ssDNA has a tendency to fold up and form complex structures. The electrophoresis mobility of such structures on the gels depends on their conformations, and also on their lengths. SSCP is a PCR based method, and after denaturing radiolabelled or nonradiolabelled PCR, products of a desired region of genomic DNA or RNA were loaded on nondenaturing polyacrylamide or MDE gels. Alteration in a DNA sample that differs from a standard DNA can be identified by a comparative mobility shift during gel electrophoresis, if the base change results in a change of conformation. This method is the most simple and popular technique in molecular biology. Its efficiency has an inverse correlation with the size of the PCR products.

Heteroduplex analysis (HA)

The HA is based on the fact that the mobility of heteroduplexes is less than homoduplexes on the polyacrylamide gel (White et al. 1992). The efficiency of this method for fragments less than 200 bp is very high and alteration such as insertions, deletions and single base substitutions are detectable. The gel for HA is exactly the same as SSCP.

Denaturing gradient gel electrophoresis (DGGE)

The DGGE is another PCR based method in which DNA migrates through a gradient gel. In this method the gradient of denaturant is increased from one side of the gel to the other. Migration is continuous until the heteroduplex DNA reaches a position in the gel where the strands melt and separate, after which the denatured DNA does not

migrate any further. This method is able to detect a single base pair difference between normal and mutant DNA.

Cleavage of mismatch analysis

These techniques are based on the cleavage of mismatches between target DNA and a probe. The three different forms of this group of techniques are: (i) chemical cleavage mismatch analysis (ii) Rnase A cleavage (iii) enzyme mismatch cleavage.

(i) Chemical cleavage mismatch analysis(CCM)

The CCM is based on certain chemicals which can react with single stranded bases of DNA to make them susceptible to cleavage by piperidine (Cotton and Cambell., 1989). In CCM, the radiolabelled probe from the normal individual will hybridise with test DNA and can be modified by hydroxylamine and osmium tetroxide. The modified base will cleaved by piperidine. After chemical treatment, the products are loaded on the denaturing polyacrylamide gels, followed by electrophoresis and autoradiography.

(ii) Rnase A cleavage

This technique is based on the fact that Rnase A is able to cleave a mismatch base in a single strand RNA probe. It was described for the first time by Myers in 1985. In this method labelled wild- type RNA probes hybridise to double stranded target DNA to form RNA/DNA duplexes. In the case of mutation the enzyme will recognise and cleave single -stranded RNA at the site of the mismatch. The cleaved product is separated by denaturing polyacrylamide gel electrophoresis and then by autoradiography. The efficiency of this method is about 50% (Grompe, 1993). By using sense and anti-sense strands of an RNA, each mutation has two chances of being detected by Rnase A cleavage.

(iii) Enzyme mismatch cleavage

The basis of this technique is the fact that bacteriophage resolvase is able to recognise and cleave mismatched nucleotides (Mashal et al. and Youil et al. 1995). This enzyme is able to cleave branched DNA and can cleave DNA at the site of mismatch. Wild-type radiolabelled DNA hybridises to target DNA and the site of mismatch in heteroduplex DNA is cleaved by resolvases. The cleaved products are separated on nondenaturing polyacrylamide gel, followed by electrophoresis. This technique has this advantage to predict the nature and the precise position of mutation and of being applicable to fragments of more than 1 kb. The sensitivity of this technique is about

90%, and it is able to detect all possible mismatches as well as deletion mutation.

Protein truncation test (PTT)

The PTT is one of the newest methods for detecting any forms of frame shift mutation resulting in premature terminations of translations (Roest et al. 1993). DNA or RNA can be used with this technique. In large exons, like the exon 15 in the *APC* gene or exon 11 in the *BRCA1* gene, the DNA can be used, otherwise the RNA should be converted to cDNA by RT-PCR and then used for PTT. The 5' site of the forward primer should carry a signal for transcription and translation. The cDNA or DNA is put into a coupled transcription-translation cell free system, which uses the T7 promoter to make mRNA, and the translation initiator to translate it. The protein product is then run on a SDS-PAGE, followed by autoradiography. Truncated protein results in a shorter protein, compared to the standard. The size of truncated protein reveals the position of the mutation.

Direct DNA sequencing

The direct DNA sequencing is the final step for all positive screening results. For some small genes, like p16 and the globlin genes, this could be used as a screening method. Double or single stranded PCR products could be used for this technique. The two main methods for direct DNA sequencing are based on the chemical method developed by Maxim and Gilbert (1977), and the enzymatic method which was developed by Sanger et al. (1987).

1. 8. Spectrum of mutations in mismatch repair genes

Five different mismatch repair genes that predispose to HNPCC have been identified. Germ-line mutations have been reported in the *hMSH2* and *hMLH1* genes. Mutations in *hPMS1* and *hPMS2* have been described in only one kindred each (Nicoladies et al. 1994). To date, no germ-line mutations have been described for a fifth DNA mismatch repair gene, *GTBP/p160* (Papadopoulos et al. 1995). However, since 1993 a total of 33 germ-line mutations including missense mutations (Leach et al. 1993; Liu et al. 1994; Moslein et al. 1996, Froggatt et al. 1996), nonsense mutations (Leach et al. 1993; Buerstedde et al. 1995; Mostein et al. 1996), small deletions (Kolodner et al. 1994; Wijnen et al. 1995; Buerstedde et al. 1955; Mostein et al. 1996), small insertions (Lazar et al. 1994; Wijnen et al. 1995), large deletions including one exon skipping

(Leach et al. 1993), and multiple exon skipping (Liu et al. 1994, 1995), and insertions (Liu et al. 1996), have been reported in the *hMSH2* gene. About 75% of these mutations are out-of-frame mutations resulting in truncated proteins. Only a small fraction of the mutations have been missense mutations. However, two/thirds of the mutations reported in this particular gene could cause truncated proteins. Although, some clusters of mutations are present in the exon 7 and exon 12, no common mutation or hotspot has been reported in this gene to date. A summary of mutations found in the *hMSH2* gene is presented in appendix 1. A high percentage of mutations in the *hMSH2* gene result in truncated proteins. This suggests that the protein truncation test may be the preferred method to screen for *hMSH2* defects. Missense mutations could be harder to interpret because not every change in amino acid sequences alters the function of the protein. HNPCC appears to follow the Knudson model (two hit model). According to this model, a heterozygous mutation in one of the mismatch repair genes in the germ-line gives a predisposition to cancer. Tumour development requires inactivation of the wild-type allele by somatic mutation or its loss.

Distribution of mutations in HNPCC and sporadic cases are different, only a small fraction of sporadic CRCs with RER+ phenotype carry a mutation in MMR gene (Figure 1.8). Leach et al. (1993) reported four mutations, one polymorphism and one somatic mutation. Two mutations causing truncated proteins are due to exon 5 skipping from transcripts, and the other a C to T transition that changes arginine to a stop codon. Kolodner et al. (1993) reported a mutation at position -6 in intron 12 in the *hMSH2* gene. In this paper, the sequence of the *hMSH2* was also reported, and due to this mutation it was concluded that this is the susceptible gene for colorectal cancer. Liu et al. (1994) reported missense mutations in exons 8 and 5. Exon 15 skipping from transcripts was reported as well as a deletion from nt 1981- 2073, due to a C to T transition that creates a new splice donor site within exon 12. A

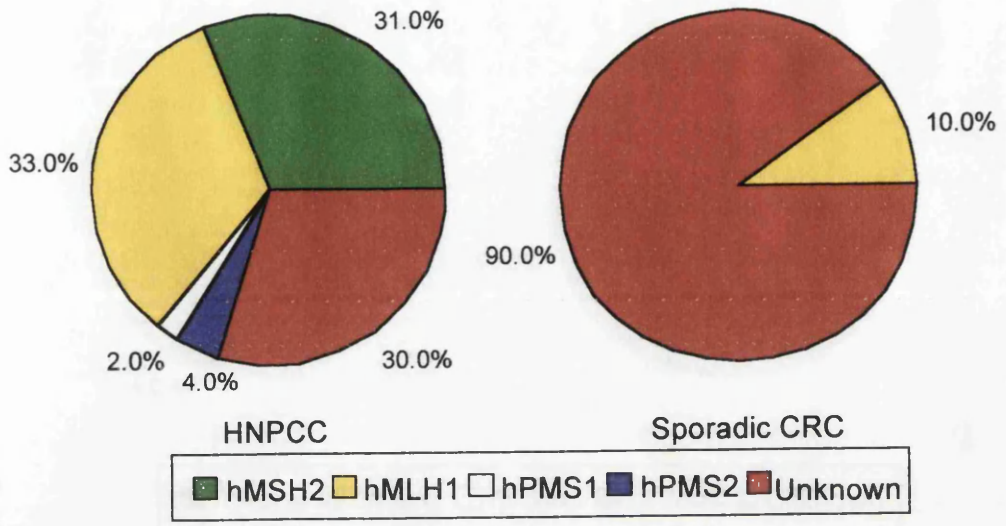


Figure 1. 8 Comparison of mutations in HNPCC and sporadic colorectal cancer.

deletion in exon 13 was also found, and another deletion from exon 8 to exon 15.

They were not able to find the cause of the last two mutations at the genomic level. However, five of six mutations reported by Liu et al. were frameshift mutations. Kolodner et al. (1994) reported a 2 bp deletion in exon 12 in a Muir-Torre syndrome. Liu et al. (1995) reported in colorectal cancer patients aged less than 35 years, missense and nonsense mutations, in exons 11 and four respectively. Froggatt et al. (1995) confirmed exon five skipping from transcripts by an enzymatic assay, and claimed that it is a common mutation in the American people and North England population. Wijnen et al. (1995) reported seven mutations, all of which are frameshift mutations. He also reported three exonic and one intronic polymorphisms. Buerstedde et al. (1995) reported one frame shift and one point mutations in exons 3 and 7 due to a two base pair deletion and a C to T substitution respectively. He also reported two polymorphisms in intron 10 at positions -8 and +10. Moslein et al. (1996) studied 6, 19 and 20 patients including sporadic, familial and HNPCC cases, respectively. She reported no mutation in sporadic cases, two mutations in familial (10%) and also five mutations in HNPCC (25%) cases. However, mutations were not found in 50% of HNPCC cases. Contribution of mismatch repair genes in HNPCC families is shown in Table 1.2 (Liu et al 1995). In the large study carried out by Liu et al. (1996) HNPCC mutations were not detected in 14 of 48 families, although in some HNPCC kindreds, genetic linkage to *hMSH2* or *hMLH1* could

be demonstrated. However, a deletion was found in the exon 7, and a 173 bp insertion in the same exon, both of these mutation were frameshift mutations. The cause of these mutations at genomic level could not be found. In this paper, a three bp deletion in exon 12 and exon 5 skipping from the transcript were also reported.

Table 1.2: Contribution of repair genes in HNPCC

Gene	Chromosome	Contribution to HNPCC
hMSH2	2P21-22	31%
hMLH1	3P21	33%
hPMS1	2q31-33	2%
hPMS2	7p22	4%

Two frameshift mutations due to deletion of CA in exon 10, and a TG insertion in exon 12, were reported by Nystrom- Lahti et al. (1996). Luce et al. (1995) detected truncated proteins for either *hMSH2* or *hMLH1* in 6 of 12 HNPCC kindreds, but reported the nature of the underlying genomic DNA defect for only 2 of them (this will be discussed later on). A C to T transition in codon 622 (exon 12) and the same alteration in codon 406 in the exon 7 was reported twice. Deletion of codon 596 in the exon 12 was reported three times. The reminder of the mutations were only reported once. Therefore, about 90% of mutations were reported only once, suggesting that mutations in this gene are population related. No germ-line mutations have been reported in exons one, two, five, eleven, and fifteen (see appendix 1). The pattern of mutations in *hMSH2* and *hMLH1* are compared to each other in Figure 1.9. As illustrated, the pattern of mutations in both member of the mismatch repair family looks the same, although frameshift mutations in *hMSH2* are slightly higher than *hMLH1*(84% in contrast of 60%). However, most of the mutations in these two genes predicted to cause truncated proteins, while substitution in *hMLH1* is more frequent than *hMSH2*.

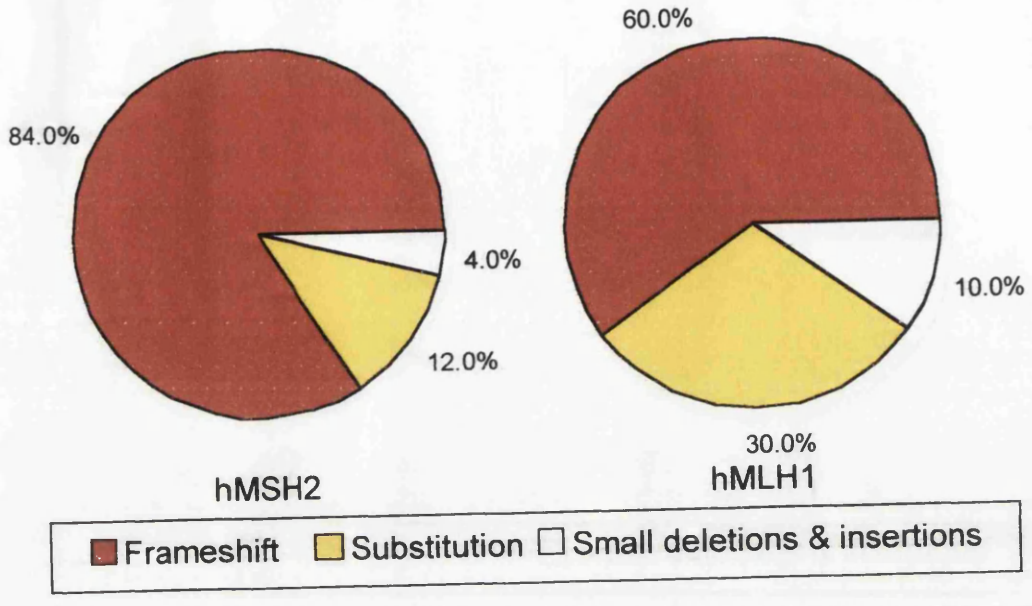


Figure 1.9. Comparison of germ-line mutations in two members of mismatch repair family.

Recurrence of mutations in *hMSH2*

The genetic heterogeneity of the disease, the complexity of the genes involved and the wide intragenic distribution of the germ-line mutations make a low chance for recurrence of mutations in mismatch repair genes. At the time of writing this section, not more than 80 germ-line mutations were reported for the two main genes which predispose to colorectal cancer and among them few germ-line mutations have been reported by different groups. These mutations are discribed below:

There is a cluster of mutations in exon 7 and 12 but there is no sign of a hotspot in this gene. A number of mutations have, however, been reported by different groups including: (i) exon 5 skipping from transcript due to A to G transition at position -3, resulting in-frame-deletion of exon 5 (Liu et al. 1995; Froggatt et al. 1995). As will be discusses later, it is believed that this is a founder mutation in the gene. (ii) three base pair deletion at codon 532 resulting in-frame-deletion of asparagine reported by Liu et al. 1996; Borresen et al. 1995 and Mary et al. 1994 (iii) a C to T transition at codon 621 resulting Arginine to stop codon reported by two groups, Liu et al. 1995; Kolodner et al. 1994. (iv) exon 15 skipping from transcript due to G to T transversion at position -1

(Wijnen et al. 1995; Liu et al. 1995) resulting in a premature termination codon within exon 16. Exon 2-6 deletion from transcript has been reported by Kohonen-Corish et al. 1996. Deletion of exons 2-6, and also exon 15 skipping from transcript due to a novel splice site mutation, which have been found in this study could be further examples of recurrence of mutations in this gene.

Absence of a common mutation in HNPCC's genes make it difficult for diagnostic laboratory to screen HNPCC families and their relatives who are at risk.

Founder mutations in the HNPCC families

Founder mutations have arisen once and subsequently passed on through successive generations to present time. To date, three founder mutations have already been detected in HNPCC families. Two of them are in the *hMLH1* and the third one is in the *hMSH2* gene. These mutations are: (i) a point mutation at splice acceptor site of exon 6 in the *hMLH*1 gene, resulting in the deletion of exon 5 from the transcript. This was found in five Finnish families (Lahati et al. 1995). (ii) Genomic deletion of exon 16 in the *hMLH1* gene, which was found in 14 Finnish families (Lahati et al. 1995) (iii) point mutation at splice donor site of exon 5 in the *hMSH2* gene, resulting in the deletion of exon 5 from the transcript (Leach et al. 1993; Froggatt et al. 1995). This mutation was found in three Anglo-Saxon families and has been reported from England. Each of the three mutations is expected to lead to truncation of the protein through the loss of one exon. It has been shown that in the Finnish population, exon 16 mutation of *MLH1* is an ancestral mutation, as most families carrying this mutation come from the same geographical region. The exon 6 mutation of *MLH1* has so far been found in five unrelated families. Their geographical origins are confined to a small region in southern Finland. Moisio et al. (1996) has suggested that the spread of mutation 1 started 16-43 generations (400-1,075 years) ago and mutation 2 some 5-21 generations (125-525 years) ago. The *MSH2* founder mutation that results in the skipping of exon 5 has been observed in three families in North America, all of whom report Anglo-Saxon heritage (Leach et al. 1993). It has also been reported in 12% of English kindreds (Froggatt et al. 1995) suggesting a common ancestor between North American and English HNPCC kindreds.

1. 9. The aims of this study

1. To identify germ-line mutations in the *hMSH2* genes of patients with colorectal cancer in the West of Scotland.

2. To assess the level of microsatellite instability in the tumours as a first screening step.

3. To characterise fully the mutations found, and determine if a genotype/phenotype correlation exists.

CHAPTER TWO

MATERIALS AND METHODS

2. 1. Patient selection

In this project, two groups of patients were identified: those undergoing surgery for colorectal cancer (who may or may not have a family history) and those with a family history of colorectal cancer who had been seen in a cancer genetics clinic. From 30 patients who were undergoing surgery, 32 normal and tumour tissues were available, including one patient with three tumours at different sites of her colon. There were 14 males and 16 females, age range 23-76 years, 47 RNA samples from some of the above patients and patients with family history of colon cancer, including 25 males and 22 females, age range 17-76 years, were also available for mutation analysis. None of the families in this study was large enough for linkage analysis (see appendix 2 for more information).

2. 2. DNA extraction

2. 2. 1. DNA extraction from peripheral blood

DNA was extracted from the whole fresh blood as described by Kunkel *et al.*, (1977).

2. 2. 2. DNA extraction from solid tissue

DNA extraction was carried out from tumour and normal solid tissues (Miller et el., 1988). A small amount of tissues were scaled (0.1 gr) and crushed with a mortar and pestle with 3 ml nuclei lysis buffer (10 mM Tris-Hcl, 400 mM NaCl and 2 mM Na2EDTA, pH 8.2) and 200µl SDS 10% and 100µl proteinase K (10 mg/ml). The mixture was transferred into a clean 20 ml universal tube and incubated in a water bath at 55^0C for overnight. The digested products mixed properly by vortexing for about one minute and centrifuged at 25000 rpm for 15 minutes. The supernatant was transferred into a clean universal tube. Next, one ml NaCl 6M, was added and mixed vigorously by vortexing and then centrifuged for 10 minutes at 2500 rpm. The supernatant was then transfered to a clean universal tube, and exactly two volumes of room temperature absolute ethanol or equal volume of propanol 2-ol was added and mixed gently by inversion. The DNA exploded out from the liquid phase. The DNA washed with 70% ethanol and dried for about 10 minutes and then was dissolved in appropriate volume of TE buffer (10 mM Tris-Hcl, 0.2 mM

Na2EDTA, pH 7.5) or autoclaved distilled water. The DNA was allowed to dissolve for 2 hours at 37° C before quantitating.

2. 2. 3. DNA extraction from paraffin embedded tissue

In some cases DNA extractions were carried from paraffin embedded normal or tumour tissues. The section must be dewaxed completely before digestion by proteinase K. A small amount of tissues (0.1 g) were cut off and weighed from paraffin block and used for DNA extraction. The tissue was placed into a 1.5 ml eppendorff tube and 1 ml xylene added to it, and vortexed vigorously for about three minutes. This step is important to dissolve paraffin by the xylene. The tube spun down in a bench top centrifuge (eppendorff centrifuge) for 3 minutes at full speed, at room temperature. The tissue was pelleted at the bottom of the tube and xylene removed by pipetting. The above step was repeated once more to dewax the paraffin properly, then the tissue was washed out by 1 ml of absolute ethanol and spun down for 3 minutes at full speed in a top bench centrifuge. The ethanol was poured down the sink. The above step was repeated once more. The tissue was dried in the same tube or in the plastic boat at room temperature for about 10 minutes. The tissues were chopped by scalpel, and then crushed by the mortar and pestle by 3 ml nuclei lysis buffer, 200µl 10% SDS and 100 µl proteinase K (10 mg/ml) as described in section 2.2.2. The rest of procedures were as described for DNA extraction from solid tissues. The only difference was that the incubation time for proteinase K digestion at 55^{0}C increased up to 72 hours.

2. 3. Polymerase chain reaction (PCR) of microsatellite sequences

Genetic instability was detected by comparing the results from tumour DNA with those from normal DNA. The sequence of markers, size, and anneealing temperature are presented in Table 2.1.

Table 2.1: The sequence of microsatellite markers and the condition of the PCR.

Name	Sequences	Expected size	Annealing T.
D18S34 (w1-2)	CAG AAA ATT CTC TCT GGC TA CTC ATG TTC CTG GCA AGA AT	110 bp	55^{0}C
635,636	TTG ACC TGA ATG CAC TGT GA TTC CAT ACC TGG GAA CGA GT	100 bp	55^0 C

D3S240 (W3-4)	CCA TGT CCC ATA TCT CTA CA TGA AAT CAC TGA TGA CAA TG	92 bp	57^0 C
LPLCA,A	GAG TAA AAG ATG ATC CAG TGT C CAA CCC AGA TTT GAG CTA CGT	171-175 bp	57^0 C
LPLCA,B	CAT GAA TTA GAA ATC CAG TGG G AGG TGA TGT CCC AGA GGA AC	106-134 bp	57^0 C
D2S123	AAA CAG GAT GCC TGC CTT TA GGA CTT TCC ACC TAT GGG AC	197-227 bp	55^0 C
D2S119*	CTT GGG GAA CAG AGG TCA TT GAG AAT CCC TCA ATT TCT TTG GA	214-232 bp	60^0 C
D3S1293	ACT CAC AGA GCC TTC ACA CAT GGA AAT AGA ACA GGG T	116-144 bp	53^0 C
D13S160	CGG GTG ATC TAA GGC TTC TA GGC AGA GAT ATG AGG CAA AA	250 bp	60^0 C
D8S282	GGG CAC AGG CAT GTG T GGC TGC ATT CTG AAA GGT TA	260-272 bp	58^0 C
D2S136	AGC TTG AGA CCT CTG TGT CC ATT CAG AAG AAA CAG TGA TGG	91-111 bp	53^0C
D3S1298	AGC TCT CAG TGC CAC CCC GAA AAA TCC CCT TGT GAA GCG	194-220 bp	53^0 C
D2S378	TGT GGG CTG GTC AGA TAT TC CGC TAG GAT CAC TAT GTT TTG C	203-217 bp	53^0 C

* The final $MgCl_2$ concentation for this marker was 3.0mM.

2. 4. RNA extraction

RNA extraction from fresh peripheral blood was carried out, using either acid guanidium-thiocyanate (Fluka) or Trizol solution (Gibco BRL).

2. 4. 1. RNA extraction from whole blood using acid guanidium - thiocyanate

RNA extraction from whole fresh blood was carried out using acid guanidium - thiocyanate - phenol - chloroform (Chomczynski et al. 1987).

RNA extraction from EDTA treated blood had a much better yield than heparinized blood.

Lymphoblasts from whole blood were separated using Histopaque 1077 (Sigma). Histopaque should keep at $+4^0$C, but half an hour before use should be left at room temperature to set up and then 5 ml was poured into a cleaned universal tube. Five ml EDTA treated fresh whole blood was gently poured on to it by pasteur pipette, and centrifuged at room temperature for 30 minutes at 1500 rpm. Three layers were observed, upper layer: Histopaque, middle layer: lymphocytes and at the bottom of the tube were red blood cells. Lymphocytes or buffy coat from the middle layer were transferred by a clean pasture pipette to another universal tube, and washed once with cold phosphate buffer sulphate (PBS), then centrifuged at 14000 rpm at $+4^0$C for 20 minutes. Lymphocytes pellets were resuspended in one ml of solution D (4M guanidium thiocyanate, 25 mM sodium citrate, 0.5 % sarkosyl, 0.1 M beta - mercaptoethanol). Next, they were transferred equally into two eppendorf tubes and the following solutions were added to each tube: 50 µl of 2M sodium acetate, 500 µl of phenol water saturated and 100 µl chloroform-isoamyl alcohol mixture (49:1). The tubes were very gently mixed by inversion, and then centrifuged at 14000 rpm at 4 C for 20 minutes in a top bench centrifuge. The top aqueous phase that which contained RNA was very carefully transfered into another 1.5 ml eppendorf tube. Extreme care should be taken in this stage, as the middle layer content DNA and could be cause DNA contamination. 500µl propanol-2ol or isopropanol was added to the aqueous layer and mixed very gently by inversion three times. The mixture was kept for at least one hour a -20^0C, or overnight and then spun down at $+4^0$C at 14000 rpm for 20 minutes. The RNA pellet was resuspand in 150 µl solution D. Then 150 λl isopropanol was added to the mixture and left at least for an hour at - 20^0C. Then, the tube spun down at 14000 rpm for 10 minutes. The RNA pellet was washed with 200 µl 70% ethanol and spun down at 14000 rpm for another 10 minutes. The RNA pellet was dried at room temperature for 10 minutes and dissolved in 30-50 µl DEPC treated water at 55^0C for 5 minutes and was then stored at -40^0 C for further analysis.

2. 5. Oligonucleotide design

The oligonucleotides are short sequences of nucleotides usually from 18 to 25 bp. One set of primers is required for amplification of a region from DNA. The

sequence of each primer is complementary to opposite strands of the region of DNA. The primers are designed so that each of them extends the synthesis of DNA towards the other. Thus primer "F" (F stands for forward primer) directs the synthesis of a strand of DNA that can then be primed by primer "R" (R stands for reverse primer) and *vice versa.* This results in the *de novo* synthesis of the region of DNA flanked by the two primers.

In this project, oligonucleotides were designed using the computer programme, OLIGO version 3.4 (© Wojciech Rychlik; Medprobe). OLIGO is a program which calculates free energy (ΔG), hybridisation temperature and secondary structure of DNA or RNA based on the highly accurate measurement of nearest neighbour ΔG values.

A good sequencing primer or hybridisation probe should meet the following criteria: (i) be highly specific for the intended target sequence and not base-pair to other regions within the template. (ii) be non self-complementary and not form dimers. (iii) form stable duplexes with a template under appropriate conditions.

All the oligonucleotides were synthesised on an ABI 391 automated DNA synthesiser. Ammonium hydroxide was used to elute the primers from the solid support and deprotection was carried out at 55^0C for 12-16 hours. In brief, one ml of ammonium was used for deprotection. In every 30 minute, 100 µl of ammonium was injected into the column of the primer. Another one ml of ammonium was used to wash any residue of oligonucleotides from the column. The 2 ml ammonium was collected into the 2 ml screw cap tube, and incubated for O/N at 55^0C. The tube was left on ice for 15 minute, and then 150 µl of aliquot of primer was taken and left in the fume hood for rapid evaporation. The concentration of oligonucleotides was measured by spectrophotometry and diluted as described previously. The sequences of the primers that have been used in this study are shown in the Table 2. 2.

Table 2.2. The sequences of the primers that have been used in this study for individual exon amplification in the *hMSH2* gene. The primers which have been designed by OLIGO for this project is marked by (*).

Exon	Sequence	References
1	GCT TAG TGG GTG TGG GGT C* GGG AAA GGA GCC GCG CCA*	
2	GAA GTC CAG CTA ATA CAG TGC CTT CAC ATT TTT ATT TTT CTA CTC	Kolodner et al., 1994
3	GCT TAT AAA ATT TTA AAG TAT GTT C GCC TTT CCT AGG CCT GGA ATC TCC	Kolodner et al., 1994
5	GGC ACC AGT GGC TAT AGG* GTT AAG GGC TCT GAC TGC*	
6	GCT TGC CAT TCT TTC TAT TT* CCT CTC CTC TAT TCT GTT CT*	
7	GAC TTA CGT GCT TAG TTG GTA TAT ATT GTA TGA GTT GAA GG	Kolodner et al., 1994
8	ATC TTT TTA TTT GTT TGT TT* CTA GTC CAT ACG TTA TAT GA*	
9	ATA GGA TTT TGT CAC TTT GTT* GAA TTA TTC CAA CCT CCA A*	
	GTC TTT ACC CAT TAT TTA TAG G GTA TAG ACA AAA GAA TTA TTC C	Kolodner et al., 1994
12	CCT GTT GAC AAT TTC TGT TTT TA CAA AGC CCA AAA ACC AGG	Kolodner et al., 1994
13	TAT GGG AGG TAA ATC AAC* TGA GGA TAG AAG CAG TTT*	Kolodner et al., 1994
	CGC GAT TAA TCA TCA GTG GGA CAG AGA CAT ACA TTT CTA TC	
14	TGT TAC CAC ATT TTA TGT GA TTT CCC ATT ACC AAG TTA C	Kolodner et al., 1994
15	TGC TGT CTC TTC TCA TGC TG* ACT TCT TTG CTG CTG GTT CC*	

In the course of this study, the whole coding sequence of the *hMSH2* was sequenced from different patients. The primers used for sequencing of the RT-PCR is shown in Table 2. 3, the primers designed by OLIGO as marked by (*).

Table 2.3. Primers used for sequencing of RT-PCR products in the *hMSH2* gene.

Name	Direction	Exon	Sequences
135*	S	1	ACC GGG GCG ACT TCT ATA CG
248*	S	1	CCA GGG GGT GAT CAA GTA CA
250*	AS	1	ATG TAC TTG ATC ACC CCC TGG G
398*	AS	2	CAC AAC ACT CTG CAG ATT CTT T
440*	AS	2	AGA AGC CTT ATA TGC CAA AT
SC2A*	S	3	GGC TTC TCC TGC CAA TCT CT
SC3A*	S	3	GGT TGG AGT TGG GTA TGT GG
576	S	3	CCA CAT ACC CAA CTC CAA CC
190	S	3	TCC TCA TCC AGA TTG GAC C
SC2B*	S	3	CTC CAG CAG TCT CTC CTC CG
SC4A*	S	4	TGA AAG GCA AAA AGG GAG AG
SC3B*	AS	4	CTC TCC CTT TTT GCC TTT CA
252	S	4	GAT GAA TAG TGC TGT ATT GCC
291	S	5	TGA CTA CTT TTG ACT TCA GCC
960*	AS	5	GGC TGA AGT CAA AAG TAG TCA
1170*	AS	7	AAA GTC TGC CTC AAT TCT GC
404	S	7	GTT ACC GAC TCT ATC AGG G
1351*	S	8	ACC AGA AAT TAT TGT TGG C
460	S	8	GGA TCA GGT GGA AAA CCA TG
1492*	AS	9	ATT CAC TGA GAT TAG GAT
513	S	10	CTG GAT TCC AGT GCA CAG
571	S	10	GAA GAA GCC CAG GAT GCC
1720*	AS	10	AAT TTA ACA CCA TTC TTC
1979*	S	12	TCC TAA TGA CGT ATA CTT TG
2166*	S	13	GCA GAA GTG TCC ATT GTG G
718	S	13	AAT TGA AAG GAG TCT CCA CG
2350*	S	14	GGG AAG AGG AAC TTC TAC CT
769	S	14	ATT GCA ACA AAG ATT GGT GC
810*	S	14	GCA TGT TTG CAA CCC ATT TT
811	S	14	AAC TAT GCT TTA TCA GGT GA
2551*	S	15	TTC ATG TTG CAG AGC TTG CT
812*	AS	15	AGC AAG CTC TGC AAC ATG AA
2625*	S	15	GGA ACC AGC AGC AAA GAA GT
2626*	AS	16	CAG GTT CCA CTT TGT TTA CG
2765*	S	16	CTG AAG TAA TAG CAA AGA AT

S: Sense, AS: Antisese

2. 6. PCR techniques

2. 6. 1. Long PCRs

There is no definition for long PCR products. In this project any PCR products more than 1000 bp were assumed to be long PCR products. As a rule, for each 1000 bp one minute is required for extension. Obviously, for amplification of very long target of DNA, many hours are required. On the other hand, *Taq* DNA polymerase is not efficient after 4 hours, therefore either a modification in the normal profile of PCR or a special *Taq* is required.

Normal *Taq* polymerase is not able to amplify a long PCR product of more than 3 Kb. Some commercial kits for long PCR products are available as GeneAmp XL PCR kit (Perkin Elemer), *Taqplus*™ DNA polymerase (Stratagene).

In this project, two fragments in RT-PCR were 1.8 Kb (fragment one) and 2.2 Kb (fragment two). Both of these fragments were amplified using normal *Taq* polymerase (Perkin Elemer). The only modification from the normal PCR profile was the increasing time in the extension for 150 seconds (two min and half).

The longest PCR products amplified in this project were 4.4 Kb (intron 15). In this PCR the extension time was 4 1/2 minutes, and the total volume of reactions was 20 µl. A mixture of normal *Taq* polymerase (Perkin Elemer) and another *Taqplus* DNA polymerase (Stratagene) was applied. A combination of normal *Taq* DNA polymerase (two ratios) (Perkin Elemer) and one ratio of *Taqplus* DNA polymeraase (Stratagene) was applied. The reaction buffer was the ordinary 10X Perkin Elemer buffer as described in 2.2.

2. 6. 2. Asymmetric PCR amplifications

Single stranded PCR products were generated by asymmetric PCR amplification as described by Gyllensten and Ehrlich (1988). Asymmetric PCR products were used for direct DNA sequencing. After amplification of double-stranded PCR products, the whole products (about 100 µl) were loaded into the low melting Nusieve agarose (FMC) gel. The expected bands were excised from the gel and incubated at -20^{0}C for O/N. Eight µl of the gel aliquot was used for asymmetric amplification, using limited lower or upper primer which was diluted 100 times from normal primer. Amplification in a total volume of 50 µl was performed for not less

than 40 cycles. The asymmetric PCR products were visualised by ethidium bromide staining following agarose gel electrophoresis, as will be described later in this chapter.

2. 6. 3. Reverse Transcriptase- Polymerase Chain Reaction (RT-PCR)

Reverse- transcriptase PCR (RT-PCR) amplifications of the *hMSH2* cDNA were performed on the total RNA from whole fresh blood of colorectal cancer patients. To synthesise the first strand of the cDNA, three approaches were applied. In the first approach specific primer was applied as a downstream primer, in the second one oligo (dT) and in the last one random hexamer. Random hexamer was the best for synthesising the first strand of cDNA.

The length of the whole coding sequencing of the *hMSH2* gene is 2.8 kb and has 16 exons. The whole coding sequence of the *hMSH2* gene was synthesised as either two or three overlapping fragments. The first fragment encompassed codon 1 to codon 628, and the second fragment encompassed from codon 250 to codon 934. In some cases it was not possible to do so due to contamination in the RNA or low quality of RNA. In such cases the cDNA was amplified in three overlapping fragments.

One microgram of total RNA plus one of the specific primer oligo (dT), or random hexamer were used as a downstream primer. The limitation of the specific primer was the specific annealing temperature, and it took time to optimise the RT-PCR. The disadvantage of oligo (dT) was its inability to amplify further than 1.5 kb away from the 3' end of cDNA. Random hexamer was found to have none of these disadvantages in my experiments.

RNA and the downstream primer were mixed in one DEPC treated eppendorf tube, and incubated at 65^0C for 10 minutes. After 10 minutes the tube was transferred onto the ice bucket, and the following solutions were added: five µl of the first strand buffer 5X (25 mM Tris-Hcl, pH8.3, 375 mMKCl, 15 mM MgCl2) 200 mM dNTPs (dATP, dCTP, dGTP, dTTP), 10 Mm DTT and 2µl (200 U/µl) murine malony leukaemia virus-reverse transcriptase (Gibco BRL). Short vortex and centrifugation were applied to collect all the contents at the bottom of the tube. The tube was incubated for 60 minutes at 40^0C. At the end of reverse- transcriptase 5-8 µl of the reverse transcript aliquots was used for PCR amplification using *Taq*

DNA polymerase Perkin elmer) in 30-50 µl total reactions. The cDNA was amplified during 35 cycles of PCR. Two large overlapping fragments were amplified, at 95^0C for one minute followed by 35 cycles thereafter at the same temperature for 30 seconds, 58^0C for one minute and 72^0C for 2 minutes and 30 seconds. The final extension was 72^0C for 10 minutes.

At the end of amplification 3 µl of PCR products was loaded onto 1-1.5% SeaKem agarose gels and stained with ethidium bromide, then visualised under U.V. light. The rest of PCR products were loaded into the 1.5% low melting agarose gel (Nusieve agarose) and expected products were excised from the gel. The excised band was transferred into the 1.5 ml eppendrof tube and kept at -20^0C for further analysis.

2. 7. Purification methods

Purification of PCR products depends on the size of the products. By purification of PCR products unincorporated dNTPs, primers and enzymes were removed. Different methods were applied for purification of PCR products including: GENECLEAN II® kit (BIO 101 inc.), 4 M ammonium acetate-isopropanol precipitation.

2. 7. 1. Purification of PCR products using the GENECLEAN II® kit

Purification of PCR products was carried out using a GENECLEAN II® kit for any products less than 500 bp according to manufacturers instructions. In brief, to one volume of PCR products, 3 volumes of NaI were added. The tube was mixed and placed at room temperature (rtm) for 15 minutes to allow binding of the DNA to the silica matrix. Every 1-2 min the tube was mixed by vortexing to ensure that the GLASSMILK stays suspended. Silica matrix and DNA were pelleted by centrifugation at 12000 rpm for 5 seconds. The supernatant was poured off into the sink. Another centrifugation was applied to collect the remainder of the NaI. The white pellet was usually firm enough to remaine intact when pouring out the liquid contents of the tube. The pellet was washed twice with cold NEW WASH. The supernatant was removed from the second wash, the tube was then spun down for a few seconds, and the last bit of liquid was removed. The pellet was resuspended in an appropriate volume of TE buffer or distilled water (DW). The tube was incubated for

5 min at 55^{0}C. Then a brief vortexing was applied. The DNA was pelted by centrifugation for 30 seconds. The supernatant containing the eluted DNA was carefully removed in a new tube. One µl of eluted DNA was loaded into 1% agarose gel to estimate the recovery of DNA. The gel was stained with ethidium bromide followed by electrophoresis, then visualised on a UV box. All the materials were provided within the kit.

2. 7. 2. DNA recovery from agarose gels using the GENECLEAN II® kit

The DNA band was excised from low melting Nusive agarose gel on the UV box in less than one minute, to minimise damage to DNA. The DNA band was placed into a prescaled 1.5 microcentrifuge tube. The tube containing the DNA band slice was weighed, and the weight of the DNA band was determined. Three volumes of NaI were added to agarose gel and then placed in the tube for 5 min at 55^{0}C. After a minute or two, the contents of the tube were mixed and returned to the water bath. After about five min, the agarose gel was dissociated completely. Five µl of GLASSMILK was added to the tube. The remaining steps are as described in section 2. 5. 6. 1.

2.7. 3. PCR purification using 4M ammonium acetate and isopropanol precipitation

The greatest advantage of this method is that all the materials are usually present in any molecular laboratory. There is no need to order any specific materials or reagents. As a result, this method is very cheap and efficient. All the asymmetric PCR products regardless of the size and all the double stranded PCR products less than 500 bp were cleaned using this method. In brief;

- to one volume of PCR products, one volume 4M ammonium acetate and 2 volumes isopropanol were added;
- the microcentrifuge tube was kept on ice for 10 min;
- the DNA was pelted after 15 min centrifugation at high speed on the top bench centrifuge;
- the supernatant was poured off and 400 µl of 70% ethanol was added to the pellet followed by 5 min centrifugation as before;
- the supernatant was poured down the sink and the pellet was dried for either 10

min at room temperature or 3 min at 55^0C on the thermocycle.

- the dried pellet was resuspended in an appropriate TE buffer of DW.

2. 7. 4. PCR purification using NICK^{TM} columns

This method was applied for cleaning the radiolabelled probe for Southern blotting (section 2. 6. 3) The probe was labelled as described in section 2. 5. 8. 1. The column was rinsed thoroughly with 400 µl 1X SSC. The probes (20-40 λl) were added to the top of the column followed by adding 400 µl of 1X SSC buffer. Unincorporated α- [^{32}p]dCTPs was permitted to drip down to the container provided and the labelled probes were collected. This has been achieved by careful monitoring(mini-monitor, series 900) of the drops which passed through the face of the monitor located at the bottom of the column. A further 400 µl of 1X SSC was added to the column, and the labelled probes were collected when the signals were at high peak. Two µl of cleaned labelled probe was added to 2 ml of scintillation solution, and incorporated α-[^{32}p] dCTP was measured by Liquid Scintillation Counter (WALLAC 1409).

2. 8. Labelling methods

Labelling methods were used for different techniques including CCM, Southern blotting, and labelling the DNA markers. One µCi of α- [^{32}p]- dCTP (Amersham) was used as a radioactive material for internal labelling of the PCR products, or δ- [^{32}p] dATP for the DNA size markers.

Two different approaches were applied for labelling of the PCR products: (i) internal labelling (ii) end labelling.

2. 8. 1. Internal labelling methods using *Taq* DNA polymerase or the Klenow enzyme

In this method, the PCR content was prepared as a normal way. 0.5 µl of the α- [^{32}p]- dCTP was added to the PCR content before adding *Taq* polymerase. The content, of the tube were mixed and collected at the bottom of the tube by brief centrifugation. The PCR content was layered by enough mineral oil. The tube was placed on the automatic thermocycle for amplification.

In this method, the radioactive materials were incorporated as one of the four essential dNTPs in the DNA.

A random Primed DNA Labelling Kit (Boehringer Mannheim Bio.) was used for internal labelling according to the manufacturer's instructions. In brief, PCR products were cleaned as described in section 2. 5. 6. Cleaned DNA (50 ng) was denatured for 5 minutes at 95^0C, and then kept on ice. In a 40 µl reaction the following materials were added:

(I) 6 µl dATP, dGTP, dTTP mixture (prepared by making a 1+ 1+ 1 mixture), (II) 4 µl reaction mixture, (III) 5 µl, 50 µCi[α^{32}p] dCTP, 3000 Ci/mmol (IV) 2 µl Klenow enzyme, (V) made up to 40 µl with DW, (VI) incubated for at least half an hour at 37^0C.

The reaction was stopped by placing the tube for 10 minutes at -20^0C or 15 minutes on the wet ice. All the probes for Southern blotting were labelled by the above methods.

2. 8. 2. End labelling method using T4- Kinase

This method could be used for 5' end labelling of the primers, DNA size markers or any other end labelling DNA fragments. In a total 10 µl reactions the following material were added; 1µl of 10X One-Phor- All Buffer Plus (Pharmacia), 50 ng DNA, 0.5-1.0 µl δ-[^{32}-p] dATP, 2 µl T4-kinase, adjusted to 10 µl by DW. The tube was incubated at 37^0C for 30 min and the reaction was stopped when the tube was placed at -20^0C for 10 minutes.

2. 9. Gel electrophoresis

2. 9. 1. Agarose gels

Agarose (Flugene) gels were used for checking the PCR products, southern blotting or for excising the band from the gel for asymmetric PCR. For different purpose different concentrations were used. One percent agarose gels were used for checking the PCR products. Small horizontal apparatus gel (GibCo BRL) was used for checking the PCR products.

2. 9. 2. Polyacrylamide gels

Normally, two different types of these gels were used: denaturing polyacrylamide gel and non-denaturing polyacrylamide gel.

There are many different denaturing systems but these using urea, formamide

and alkali as denaturants are the most common. In this research the urea was used in all the experiments as a denaturant. These types of gels were used either for sequencing or CCM. The difference between these two is the composition of the buffer, as will be described.

2. 9. 3. The DNA sequencing gels

The standard 40 cm long sequencing kit (Bio Rad) was used for sequencing. The thickness of the gel was 0.4 mm. The 8% gel was prepared as follows: in a clean beaker 22.4 g urea was scaled and to it 12 ml acrylamide (19:1), 3 ml 20X Glycerol Tolerant Gel Buffer (20X GTG: 216g Tris-base, 72g Taurine, 4g Na2EDTA.2H2O, the total volume adjust to one litter) were added . The total volumes was adjusted to 60 ml with distilled water. The gel was mixed for 5 minutes by stirrer, to it 200 µl APS 10% and 100 µl TEMED was added. The gel mixed properly and was poured into the kit.

2. 9. 3. 1. Preparation of denaturing polyacrylamide gels for sequencing

The front and back plates were cleaned by washing with the detergent and tap water, then dried using a clean tissue.

- The two plates were cleaned by ethanol and dried by cleaned tissue to remove any residual dirt.
- The back plate was siliconized in the fume hood using 3 ml silicon.
- The apparatus was assembled and sealed the bottom of the kit using either plasticine or 10 ml polyacrylamide sequencing gel.
- The gel, that had been prepared already, was poured by 50 ml syringe from the corner of the kit, and positioned the sample well former. The gel was left for about 60 minutes to be polymerised.
- The running buffer was 1X GTG and the gel was prewarmed at constant 70 W for 45 minutes. During the running the gel was kept warm in a constant 52-54 W. The temperature of the gel was between 45^0C-50^0C.

The same gel was used for CCM and microsatellite instability analysis, but TBE 10X (Tris- Base 108g, Boric acid 55g, Na_2EDTA 9.3g, the total volume adjusted to one litre by distilled water) was used for gel preparation. The running buffer was 1X TBE buffer.

2. 9. 3. 2. Preparation of non-denaturing polyacrylamide gels

The main difference between denaturing and non-denaturing polyacrylamide gel is the absence of denaturant in the non-denaturing gel. The concentration of the gel was 6%-8% in all the experiments. The concentration of the gel used depends on the size of the fragments. The recipes of the gels were as shown in the Table 2.4.

Table 2.4. The recipes of two different concentration of non-denaturing gel

	6%	8%
40% Acrylamide (19:1)	9	12
TBE 10%	6	6
Distilled water*	up to 60 ml	up to 60 ml

* To the above, 200 µl 10% APS and then 100 µl TEMED was added for polymerisation.

2. 9. 4. The Mutation Detection Enhancement (MDE^{Tm}) gel

MDE^{Tm} (Flowgene) is one type of acrylamide gel, but with better resolution. Analysis of SSCP products on MDE gels differs little from methods using conventional polyacrylamide. This gel was used in 0.5X for all the SSCP analysis with 10X TBE buffer. The gel ran at room temperature with 0.6% TBE running buffer. Whenever non-radioactive materials were used, the Protean II kit (Bio Rad) was used. In this case the gel ran at a constant 150 volts for the appropriate time, usually around 16-22 hours. The recipe of the MDE gel is presented in Table 2.5.

Table 2.5. Preparation of 0.5X MDE gel

Components	Amount
MDE gel solution (2X)	12.5
10X TBE	x
DW*	up to 40 ml

* To the final volume, 200 µl 10% APS and then 100 µl TEMED was added for polymerisation.

2. 9. 5. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Basically this is one type of denaturing acrylamide gel. SDS-PAGE is recommended for protein analysis. This type of gel was used for the protein truncation test (PTT). For the analysis of translation products it is recommended to

use SDS-PAGE, with a discontinuous buffer system. A 10% acrylamide resolving gel will give good separation of proteins in the range 20-100 kda in molecular weight. This gel has two different and distinct parts. The top part is stacking gel followed by separation gel, which could have a different concentration while the concentration of the stacking gel is usually 5% (figure 2.1). All the experiments were carried out in 0.75 mm thick gel in the protein II (Bio Rad). The formulation of SDS-page and stacking gel is presented in the Tables 2.6 and 2.7 respectively.

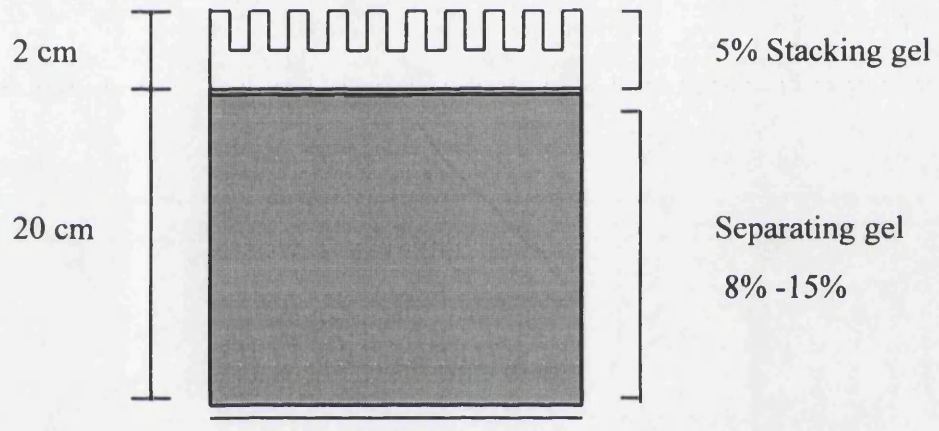


Figure 2.1. Schematic diagram of SDS-PAGE. The top part is always 5% stacking gel followed by different concentration of separating gel.

Table 2.6. The formulation of SDS-PAGE with different concentration.

Component	**8%**	**10%**	**12%**	**15%**
30% acrylamide	5.3	6.7	8.0	10.0
1.5 M Tris pH8.8	5.0	5.0	5.0	5.0
10% SDS	0.2	0.2	0.2	0.2
10% APS	0.2	0.2	0.2	0.2
TEMED	0.012	0.008	0.008	0.008
DW	9.3	7.9	6.6	4.6

Table 2.7. The formulation of 5% stacking gel

component	**Volume (ml)**
30% acrylamide	1.7

1M Tris pH6.8	1.25
10%SDS	0.1
10% APS	0.1
TEMED	0.01
DW	6.8

The percentage of acrylamide chosen will depend on the size of translated protein. In the Table 2.8 the effective separation range of SDS-PAGE is presented.

Table 2.8. The effective separation range of SDS-PAGE.

% Acrylamide	linear range of separation (Kda)
15	12-45
10	15-70
5	25-200

2. 9. 5. 1. Gradient SDS-PAGE

Toimprove resolution and characterise the small protein size the gradient SDS-PAGE was used (figure 2.2). The gradient gel maker has two distinctive gel chambers namely reservoir and mixing chambers. The two chambers are connected by a valve. A side outlet port from the mixing chamber should be connected to a tube for delivery of the gradient to an appropriate receptacle.

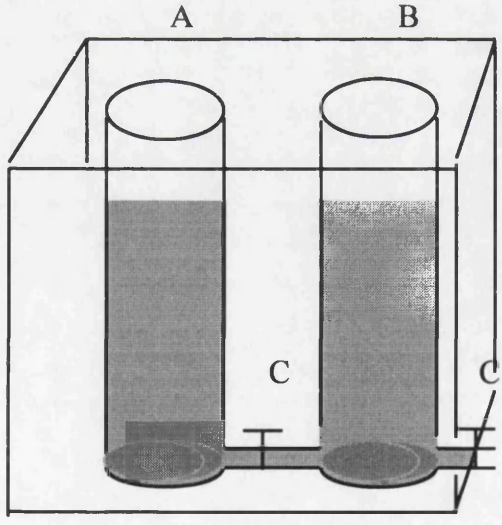


Figure 2.2. The gradient gel maker. A: Reservoir chamber, B: mixing chamber, and C: valve.

Equal volumes of two starting solutions, which define the limits of the gradient, are placed in the two chambers, with the valve between them closed. When the stopcock is opened, solution from the reservoir chamber enters the mixing solution, which is simultaneously withdrawn as the gradient is poured. The level of liquid in both

chambers falls at the same rate. The concentration of solute in the mixing chamber at any time is a linear function of the volume poured. If the gel is poured from the top, the more concentrated gel monomer solution is placed in the mixing chamber, and is introduced first into the gel mould, with gel of decreasing concentration added from above. However, the lowest gel concentration, corresponding to the largest pore size, ends up at the top of the sandwich.

2. 10. Mutation detection methods

2. 10. 1. Single strand conformational polymorphism analysis (SSCP)

Single strand conformational analysis is a method which typically involves the amplification by PCR of a discrete segment of genomic DNA in the presence of radiolabelled or non-radiolabelled nucleotides, melting of the PCR products, and analysis of the single strands on a non-denaturing polyacrylamide or MDE gels. Polymorphic differences in strand mobility result from the effects of primary sequence changes on the folded structure of a single DNA strand. The mobility of DNA fragments depends not only on the size of them but also on its sequences. The last characteristic of mobility is the base of SSCP technique. The primary sequence difference alter the intermolecular interactions that generate a three-dimensional folded structure. The molecules may thus move at a different rates through a non-denaturing polyacrylamide gel.

2. 10. 2. Chemical cleavage of mismatch (CCM) analysis

The chemical reactivity of matched T and C bases to osmium tetroxide and hydroxylamine were examined by Cotton and Campbell (Cotton and Campbell; 1989). They showed that mismatches in DNA can be detected by chemical reactivity using osmium tetroxide(detecting mismatched T) and hydroxlyamine (detected C). This technique is a PCR base method in which, target DNA will hybridise with a labelled probe to make a heteroduplex. Osmium tetroxide and hydroxylamine will modify respectively the T and C in the hetroduplex. Modified base(s) will be cleaved by piperidine. Electrophorsis and autoradigraphy will be applied after modification.

2. 10. 3. Southern blotting

Southern blotting carried out as described by Southern (1975). This technique

was applied with quick pre-hybridisation (Stratagene) solution and no depurination was employed.

2. 10. 4. Protein truncation test (PTT)

This technique was applied and optimised for the first time in the Duncan Guthrie Institute Of Medical Genetics for this project. PTT were applied for different genes namely *APC* (Powell *et al.*, 1993), ataxia- telangiectasia (Telatar *et al.*, 1996), DMD (Roest *et al.*, 1993) HNPCC (Liu *et al.*, 1994), and *BRCA1* (Plummer *et al.*, 1995).

PTT could based on DNA or RNA. However, the forward primer should bear at the 5' of its sequence the T7 promoter including a 29 bp for transcription followed by 7 bp for initiation. Two commercial kits were available for this method. The Amersham kit was more efficient than Promega kit, the following description is according to Amersham kit.

2. 10. 4. 1. Synthesising a forward primer containing the T7 promoter

If unpublished primers were used for PTT, it is recommended that optimisation carried out as usual and then the T7 promoter sequences are added to the 5' end of the primer. Next, a new long primer is synthesised for PTT. The synthesising of this long primer isthe same as for the ordinary one. No special reagent is required. The deprotection of the long primer is the same as the normal primer. The T7 promoter sequence that was added to the 5' end of the forward primer is presented below:

Transcription signals — Translation signals

GGA TCC TAA TAC GAC TCA CTA TAG GGA GA C CAC CAT GGC GGT

GCA GCC GAA GGA TG

2. 10. 4. 2. Amplification of target DNA

No special technique is required. Normal PCRs were carried out. The long PCR profiles were performed as described previously.

2. 10. 4. 3. Purification of the PCR products

Two methods were applied for cleaning the PCR products. Both the

GENECLEAN kit and 4M ammonium acetate- isoprppanol precipitation were efficient as described in section 2. 6. 6. 3. The pellets were dissolved in 5 µl Rnase free water, and 0.5 µl plus 2 µl loading mix were loaded onto 1% agarose gel to estimate the quantity of PCR products.

2. 10. 4. 4. Transcription-translation

Transcription: All experiments were carried out in 1/4 scale of the manufacturer's instructions to save material, and subsequently money. All the materials were provided in the kit. To avoid multiple defrosting and freezing, after opening a transcription vial, the contents of the tube were divided into several tubes. Each tube contained enough aliquot for four experiments. However, the transcriptions were carried in a water bath at 30^0C for 20 mins.

1. Transcription mix	2.0 µl
2. PCR products	0.5 µl
Total	**2.5 µl**

- **Translation**

Subsequently, while the tubes were on the bench, the following reagents were added to each tube:

3. Nuclease free water	1.8 µl
4. ^{35}S- Methionine	0.7 µl
5. Translation mix	7.5 µl
Total	**10.0 µl**

Translations were carried out a water bath at 30^0C for 60 mins. The tubes were placed at -20^0C to stop the reaction.

Five µl of translation products were mixed with 20 µl of loading mix (0.5 ml 1 M Tris pH 6.8, 0.8 ml glycerol, 1.6 ml 10% SDS, 0.4 ml 2-mercaptoethanol, 0.1 ml saturated pyroniny solution, 4.6 ml DW). The mixtures were denatured at 95^0C for 5 mins. At the same time, 5 µl of the standard protein size marker (rainbow, Amersham) was mixed to 15 µl of loading mix and was denatured in the same condition. All the tubes were kept on the bench.

SDS polyacrylamide gels were prepared as described in section 2. 7. 5. The 12 µl of

the sample and loading mix were loaded to each well. Electrophoresis was carried out at constant 200 volts. The electrophoresis was stopped when the dye reached the bottom of the gel. In some cases the time of running was altered according the necessity. The gels were fixed in 7% acetic acid for half an hour. Next, the gels were placed in Amplify (Amersham) for another 30 min. The gels were transferred on the 3 mm Whatmann papers and covered by Saran film, then dried for at least 30 mins at not more than 60^0C. The gels were exposed to the Kodak film and kept at - 20^0 C for O/N.

2. 10. 5. Direct DNA sequencing

All the sequencing in this project was carried out as described by Sanger (1997). Sequenase Version 2.0 DNA Sequencing Kit (USB) was used for direct DNA sequencing according to the manufacturer's instructions. Single stranded DNA (ssDNA) was generated by the asymmetric PCR method previously described. To improve understanding this method was classified into different parts:

2. 10. 5. 1. Annealing

Asymmetric PCR products were purified as described in section 2.7.3. The ssDNA was resuspended in 7 µl DW and to it 1 PM (1 PM/µl) of sequencing primer and 2 µl of sequenase buffer (5X) were added. Annealing was carried out at 65^0C for 2 min. The tubes were allowed to cool down to 34^0C over the period of 30 min. The tube was centrifuged briefly and was placed on the ice.

Meanwhile, for each sample 4 tubes were labelled for the 4 different termination mixtures. To each, 2.5 µl of dideoxytermination mix was added. The tubes were heated at 42C for 5 min.

2. 10. 5. 2. Labelling

To the tube from the annealing step 1µl 0.1M dithiothreitol (DTT), 2 µl labelling mix (1:5 diluted), 0.5µl [^{35}S] dATP (Amersham), and 2µl Sequenase T7 DNA polymerase (diluted as: sequenase DNA polymerase: Glycerol enzyme dilution buffer: Inorganic pyrophosphatase; 1: 6: 1, respectively) were added. Brief centrifugation was applied to collect the contents of the tube at the bottom of it.

2. 10. 5. 3. The termination reaction

To the each tube with 2.5 ul of dideoxytermination mix from annealing step, 3.5 µl of the aliquot from the tube from the stage labelling mixture was added. Amplifications were carried out at 42C for 5 min. The reaction terminated by adding 4 µl Stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF).

Gel electrophoresis carried out as previously described. All the sequencing gel ran at a constant 52-54 W. The running buffer was 1X glycerol tolerant buffer. The lengths of time for different gels were different, but when the bromophenol blue reached the bottom of the gel, sequences close to the primer were readable. The gels were transferred on the 3 mm Whatmann paper and covered with Saran film. Autoradiography was carried out as described before. Depending on the region of DNA to be sequenced in a particular experiment, the sequencing condition could be altered to provide maximum data of sequences close to the primer, or alternately those with a G-C rich region.

Two alterations were employed in the sequencing method: sequences very close to the primer and sequences with G-C rich region.

- There are two methods for the sequence very close to the primer. One is to use less nucleotide in the labelling step, and the other is to use Mn buffer. The last one was applied to solve this problem.

Mn buffer is a buffer solution of MnCl2 that can be added to normal sequencing reactions. A few nucleotides after primer were readable when Mn buffer was applied. To do this 1µl of Mn buffer was added to the labelling reaction prior to adding the enzyme. No other alteration was necessary. The other reagents were as usual.

- To eliminate compression in regions with G-C rich regions, dITP was substituted for dGTP. To use dITP, substitution was carried out and dITP substituted for dGTP labelling mixture. Also, the reaction time decreased to less than 5 min for both labelling and termination. All others aspects of the sequencing remained unchanged.

CHAPTER THREE

RESULTS

3. 1. Identification of microsatellite instability in tumour DNA

Thirty pairs of normal and tumour tissues were analysed from patients with CRC. All samples were analysed for genetic instability at a minimum of two, and a maximum of 9 microsatellite loci. In this study, genetic instability is defined as a gain or loss of one allele in at least one locus. The abnormality appears as extra alleles that typically are observed in tumour DNA when compared to normal DNA from the same individual. The results of MI are presented in the Table 3.1.

Patients	1	2	3	4	5	6	7	8	9	Age/Sex	Site
P1	-	ND	-	-	-	-	-	-	-	71/M	?
P2	-	-	-	-	-	-	-	-	-	26/F	?
P3	-	-	-	+	+	-	-	-	-	64/M	?
P4	-	ND	-	-	-	-	-	-	-	?/M	?
P5	-	-	-	-	+	-	-	-	-	?/F	?
P6	-	-	-	ND	-	-	-	-	-	?/F?	?
P7	-	-	-	-	-	ND	-	-	-	75/F	?
P8	-	-	-	-	-	ND	-	-	-	?/F	?
P9	-	-	-	-	-	-	-	-	-	?/M	?
P10	-	-	-	-	-	-	-	ND	-	65/F	?
P11	-	-	-	-	-	-	-	ND	-	32/M	pol.
P12	-	-	-	-	-	-	-	-	-	?/M	pol.
P13	-	-	-	-	-	-	-	-	-	?/F	?
P14	-	-	-	-	-	-	-	-	-	?/F	?
P15	-	-	-	-	-	-	-	-	-	71/M	R
P16	-	-	-	-	-	-	-	-	-	?/F	CE.
P18a	+	-	-	+	+	+	+	-	+	68/F	A
P18b	+	-	+	ND	+	+	+	-	+	=	T
P18c	+	-	+	ND	+	+	+	+	-	=	D
P21	N	ND	ND	ND	+	+	ND	ND	ND	35/F	?
P22	+	-	-	+	+	+	-	-	+	76/M	S
P23	-	-	-	-	-	ND	-	-	-	62/M	R
P24	-	-	-	-	-	ND	-	-	-	64/M	S
P25	-	-	-	-	-	-	-	-	ND	62/M	S
P26	ND	+	-	+	-	-	+	+	+	64/F	S
P31	ND	ND	ND	-	ND	ND	-	-	-	64/F	R
P49	ND	ND	-	-	+	-	-	ND	-	?/F	?
P50	ND	ND	-	-	-	ND	-	ND	-	33/F	?
P51	ND	ND	ND	ND	-	-	ND	ND	ND	?/M	?
P52	ND	ND	ND	ND	-	+	ND	ND	ND	?/M	?
P76	ND	ND	ND	-	ND	-	-	ND	-	23/M	?
P78	ND	ND	ND	ND	ND	ND	+	ND	+	42/F	?

Table 3.1. The results of MI in 30 unselected normal/ tumours DNA. The markers

are: 1: D2S160, 2: D2S119, 3: D8S282, 4: D18S34, 5: 635-636, 6: D2S123, 7: LPLCA, A, 8: D2S378, 9:LPLCA, B. ND: not done, pol: polyposis colon cancer. R: rectum, CE: cecum, A: ascending colon, T:- transverse colon, D: descending colon, S: sigmoid, ?: no information.

The average age of this group (MI+) was 58.16 years, (range 35-76 years) and the average age of MI$^-$ was 54.00, (range 23-75 years). No significant differences between the two groups were observed (t = 0.48, P = 0.64). The MI status and the sites of tumours are presented in Table 3.2.

MI status	A	T	D	S	Rec	Pol.	?	Total	%
MI+	2	1	1	1	-	-	6	11	34.4%
MI-	-	-	-	3	3	2	13	21	65.6%

Table 3.2. Location of CRC tumours. A: ascending colon; T: transverse colon; D: descending colon; S: sigmoid colon; Rec: rectum; Pol: polyposis colon cancer; ?: no information.

The group comprised 47% (14/30) males and 53% (16/30) females. No significant differences between females and males with or without MI were seen ($X^2 = 0.915$, df = 1, P > 0.5). Nine main different polymorphic markers were chosen for MI investigation. All the markers that were chosen in this study showed RER+ in at least one patient. The best marker was 635/636 on the chromosome 15q. In Figure 3.1 the nine markers and their RER are shown.

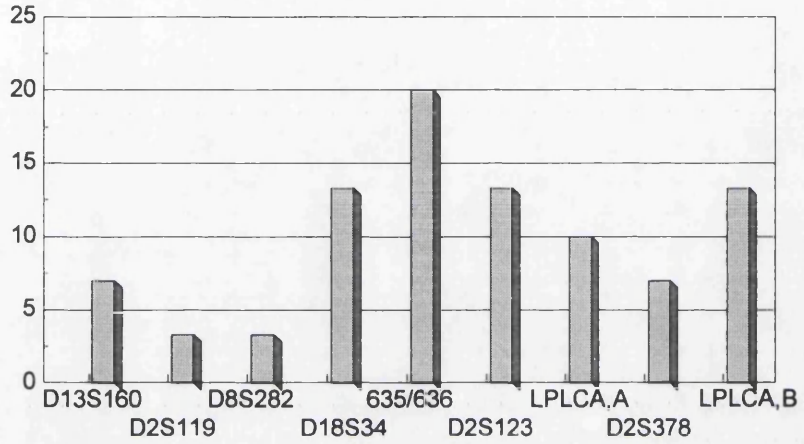
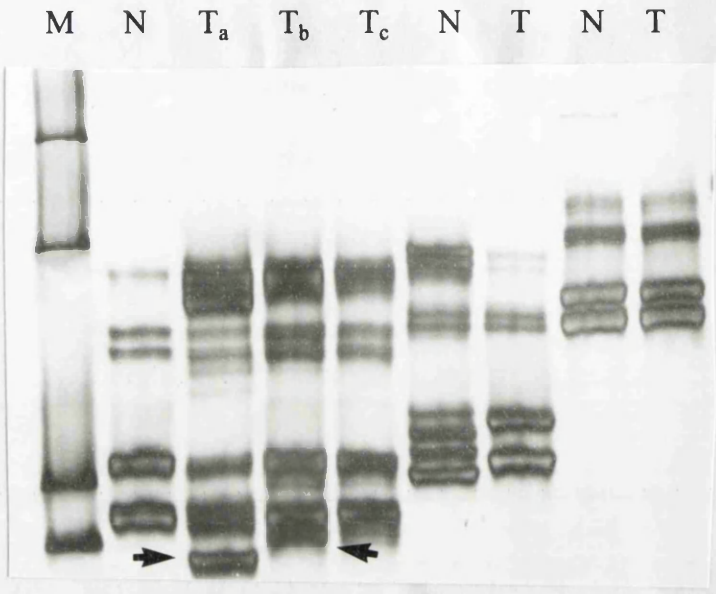
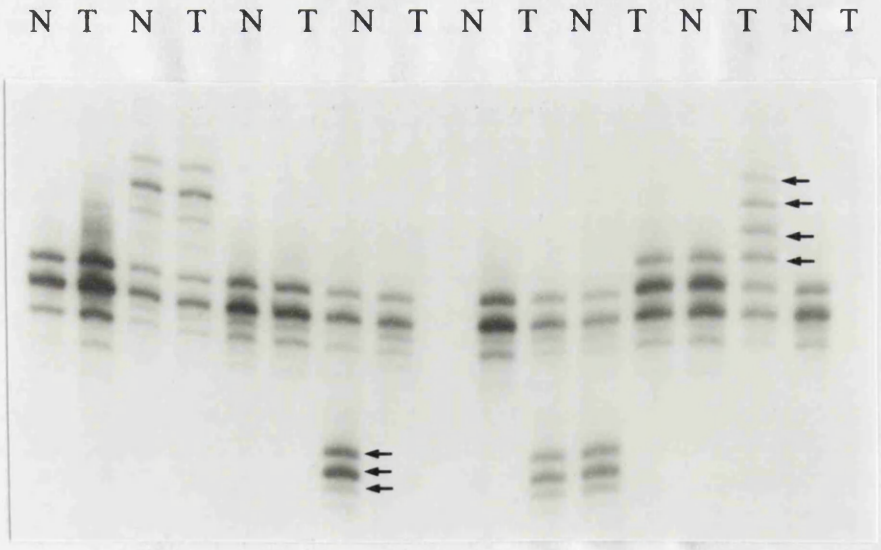


Figure 3.1. Demonstration of genomic instability for nine different markers. All of them showed MI in at least one patient. Vertical axis represents percentage of genomic instability.

MI could be detected by denaturing or non-denaturing acrylamide gel and also by radioactive or non radioactive PCR products. In Figure 3.2 A microsatellite instability in a CRC patient (Patient 18) is shown when non-denaturing acrylamide gel was used and stained by silver staining method. In this marker extra bands were present in two of three tumours (tumour A and B) but not in tumour C. In Figure 3.2 B an autoradiograph of MI has been shown for patients 22 and 26 when radioactive material was used. In Figure 3.3 A and B MI has been shown for patient 18 when markers 636/635 and D18S34 were used.



(A)



(B)

Figure 3.2. Microsatellite instability in colorectal carcinomas. Genomic DNA samples from paired normal and tumour tissues were subjected to PCR amplification using the dinucleotide repeat markers (A) D2S123, (B) LPLCA,B. Alterations in the electrophoretic mobility of PCR products from tumour were compared to normal tissues DNA and are shown by arrowheads. Silver staining was used as a means of detection in (A), while ^{32}P was used for radiolabel PCR amplification in (B). N: normal, T: tumour.

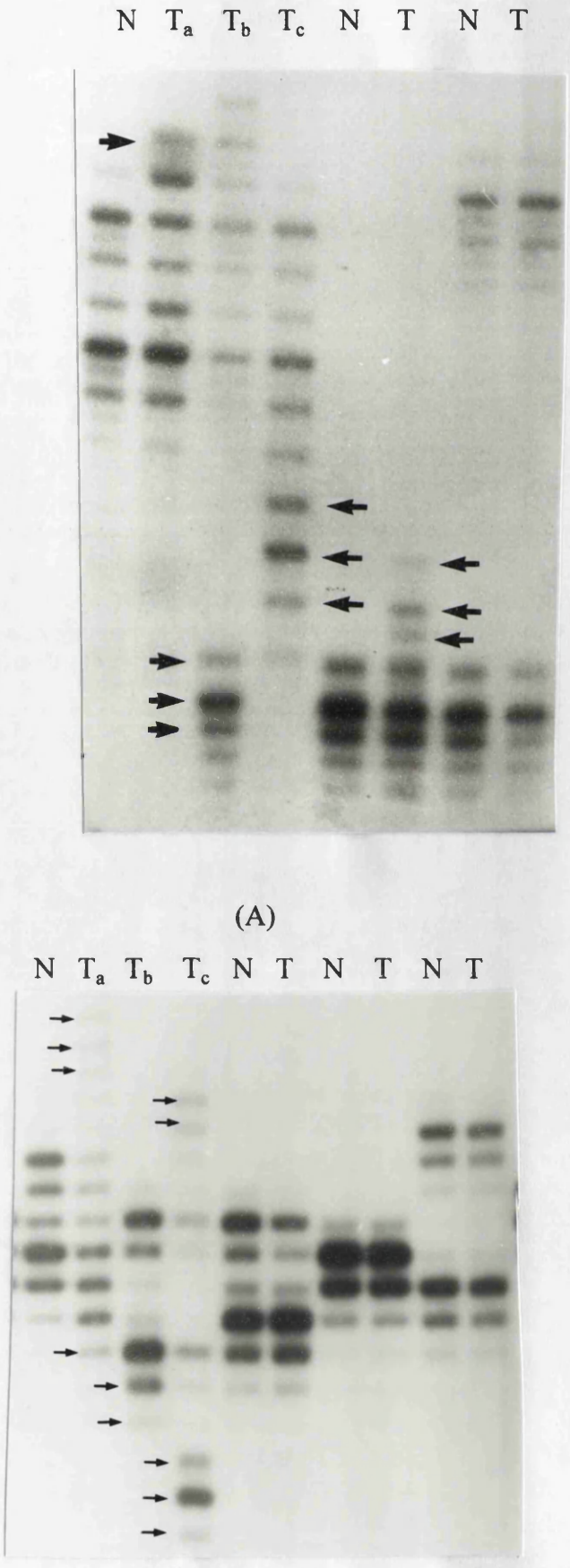


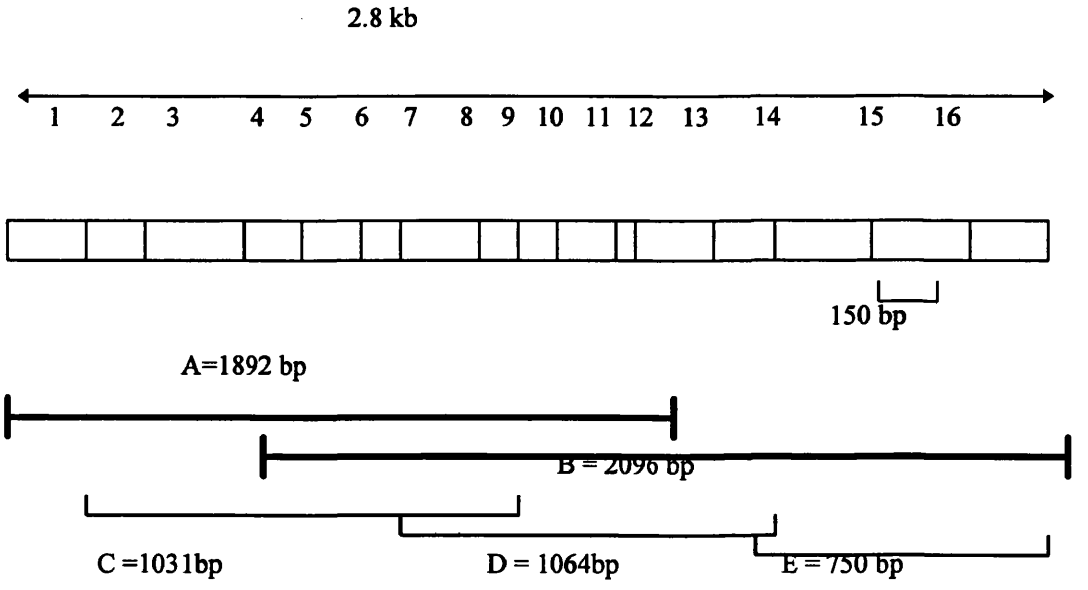
Figure 3.3. Analysis of genetic instability in paired normal (N) and tumour (T) DNA at locus 636/636 (A), and locus D18S34 (B). MI+ tumours show deviations from the normal pattern (*arrows*). N: normal, T: tumour.

In brief, microsatellite instability or replication error phenotype is a new molecular phenotype of a number of human cancers. In this part of study thirty unselected pairs of normal and tumour tissues were investigated for MI by using nine different polymorphic markers. Nine of 30 tumours were found to be MI+ in at least one marker (30%). There is no record for a majority of patients (25 of 30) (unknown cases) about the history of the disease and possible affected members in the family. There are two families with polyposis colon cancer and also two families with colorectal cancer and one family with Amsterdam criteria. MI+ was shown in six of 25 families (unknown cases) in at least one locus. Neither of the two families with polyposis colorectal cancers were MI+. Microsatellatie instability was observed in both families with colorectal cancer and also in the family with Amsterdam criteria.

3. 2. Strategy for mutation detection

3. 2.1. Amplification of whole coding sequence by RT-PCR

The entire coding sequence of the *hMSH2* gene was amplified in either two or three fragments (fragments A-E) (Lynch *et al.*, 1993; Liu *et al.*, 1994), range 633-2096 bp (Figure 3.4).



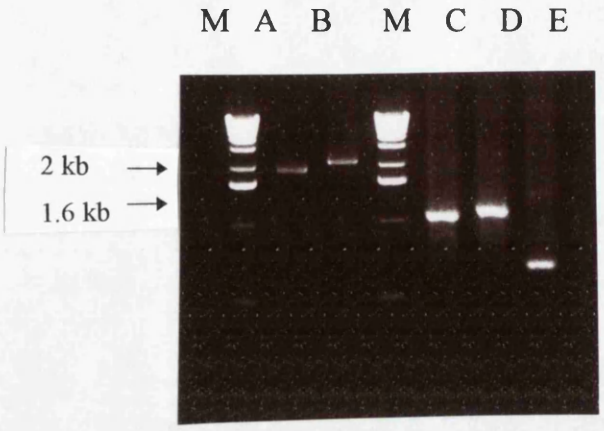


Figure 3. 4. Amplification of the entire coding sequence of the *hMSH2* gene in either two or three overlapping fragments. Fragment A encompasses from 5' UTR to exon 12 (1892 bp), fragment B encompasses from exon 4 through to 3' UTR (2096 bp), fragment C encompasses from exon 1 to exon 7 (1031 bp), fragment D from exon 6 through to exon 13 (1064 bp) and fragment E encompasses from exon 13 to exon 16 (750 bp). M:1 kb DNA marker.

3. 2. 2. Optimisation of protein truncation test

The protein truncation test (PTT) was applied as a tool for mutation screening for the first time at the Duncan Guthrie Institute of Medical Genetics in this project and then successfully has been utilized for mutation analysis in the *BRCA1* gene at the same department. This technique is based on the detection of truncated peptides after translation of the coding sequences amplified from a patient sample by PCR. To enable the translation of the PCR products, a tailed primer is used to introduce an RNA polymerase promoter, a translation initiation sequence, and an in -frame ATG triplet.

It has been shown that the PTT is able to detect frame shift mutations (Powell et al. 1993, Telater et al. 1996, Roset et al. 1993, Liu et al. 1994 and Plummer et al. 1995). Normally, RNA is used as a starting template in the PTT assay. However, if the structure of the gene of interest is favourable (containing large exons) DNA can also be used. The majority of genes do not contain such a favourable exon but rather are split into many small exons, leaving RNA as the only workable template.

Two commercial kits from Amersham and Promega were available for this technique and it was found that the first one works better than the last one. However, when

RNA was available from patients, RT-PCR was carried out and the products were electrophoresed on 1% agarose gel. In the case of any abnormal band, direct sequencing of the abnormal band was carried out to characterise the mutation. The RT-PCRs which did not show any extra bands were subjected to PTT. Any truncated products were characterised by direct sequencing and the alteration was confirmed by digestion of the PCR products by using appropriate restriction enzyme. The enzymatic digestion was also used for screening the other members of the family and for screening the normal population. All five forward primers in RT-PCRs contained T7 promoter and also an initiation signal in order to generate PCR products suitable for PTT analysis. PTT analysis was carried out as described in section 2. 10. 4. In brief, reverse transcriptase PCR products were purified by GENECLEAN II kit or by using 4M sodium acetate-ethanol and then washing with 75% ethanol as described in materials and methods. The pellet was dissolved in 5 μl DEPC treated water and then 0.5 μl of it was used for the transcription reaction including 2.0 μl transcription mix. The transcriptions were carried out in a pre-set water bath at 30C for 20 mins. The next step, translation reactions, was carried out by adding 1.8 μl nuclease free water, 0.7μl ^{35}S-methionine and also 7.5 μl translation mix. Translations were carried out in the same water bath and at the same temperature for one hour. The reactions were stopped by placing the tubes at -20C. The synthesised protein products were separated on 10-15% SDS-polyacrylamide and also on a different gradient of gel concentration. Electrophoresis was carried out at constant 200 volts and was stopped when the bromophenol blue reached at the bottom of the gel (5-7 hours). The Protean II cell kit from Bio Rad was used for electrophoresis. The gels were fixed for 30 mins and then soaked in Amplify (Enhancer) TM (Amersham) for half an hour and then dried at 60C for not less than 30 mins. The dried gel were exposed to X-ray film for at least 16 hrs and up to one week at -30C. The efficiency of this technique was improved by using concentrated PCR products and also using Amplify. However, this technique is not easy to use, particularly when RNA is the source for amplification. The presence of multiple bands due to initial translation makes it difficult to judge the real truncated products. However, this technique is able to detect pathologic mutations in two steps: PCR and then PTT. This method is very good for mutation analysis of large exons like exon 15 in the *APC* gene and also

exon 11 in the *BRCA1* gene which contains more than two-third of the *BRCA1* coding sequence. The most limiting factor for efficient use of this technique is the availability of RNA. Most diagnostic laboratories have usually stored DNA samples or tissue samples which do not allow isolation of high quality RNA. However, PTT is able to detect translation terminating mutations only, although in some respects this is a major advantage of the technique, it means that it does not detect all disease causing mutations.

3. 3. Identification of mutations in the *hMSH2* gene

In this part of study, 48 families were investigated for mutation analysis by RT-PCR followed by PTT and CCM. Of them, twenty-seven families had a history of CRC (only one of them had tumour, family 9806, MI+) and 21 families with un-known history of CRC. DNA from tumour was available from 14 families (3 RER+ , 11 RER$^-$) (see appendix 1).

Different members of a family have the same pedigree number and they were assumed as one family although in some cases more than one RNA sample was investigated for mutation analysis (families 12334, 12450, 12691 and 12780, each had two RNA samples). The families 11470, 11315, 11498 and 13894 had no RNA for mutation analysis.

3. 3.1. Identification of a novel splice site mutation in patient 21

Family history

The mother of this patient died in her 60s, due to bowel cancer. The proband was diagnosed at age 37 with colon cancer. Two of her brothers died of colon cancer and bladder cancer, and her older sister also had colon cancer. There were therefore five affected individuals in two generations.

Microsatellite instability

A small amount of tumour (less than 0.1 g)from patient 21 was available for microsatellite instability analysis, therefore MI was sought using marker loci at D2S123 and 635/636 (Figure 3.8).

Truncated protein

Amplification of exons 4 to 16 was carried out by RT-PCR as described in

materials and methods, and followed by PTT. In this analysis, lymphocyte RNA was used for RT-PCR, and the PCR product was used to generate polypeptide. When a shortened *hMSH2* polypeptide was identified in this assay, the sequence of the relevant region of the cDNA was determined to find the responsible mutation. However, this assay showed a truncated protein for patient 21 and her sister (patient 32) but not for a healthy member (Patient 28) of this family (fresh blood was available from these three individuals in this family). The result of PTT is shown in Figure 3. 6, and the truncated proteins are indicated by arrows.

Application of RT-PCR and CCM

Amplification of fragments E encompassing exons 14, 15 and 16 by the RT-PCR as a nested fragment within fragment B, followed by agarose gel electrophoresis showed a normal product (750 bp) and a shorter product (574 bp)(Figure 3. 7, A). Subsequently CCM was applied and revealed the presence of cleavage products in this segment (Figure 3. 7, B).

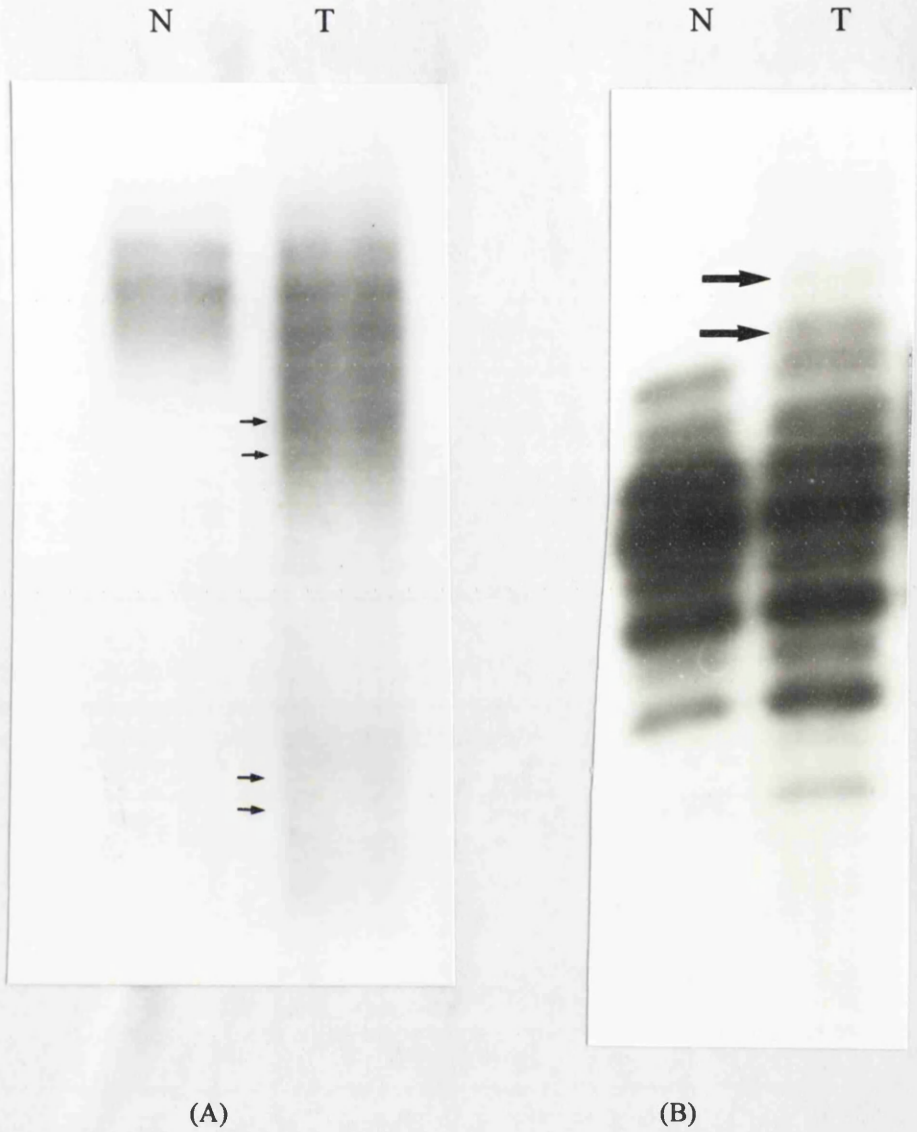


Figure 3.5. Genomic instability analysis in paired normal (N) and tumour (T) DNA at locus D2S123 (A) and 635/636 (B) for patient 21. Arrows show alteration in the electrophoretic mobility of PCR products from tumour compared to normal tissue DNA.

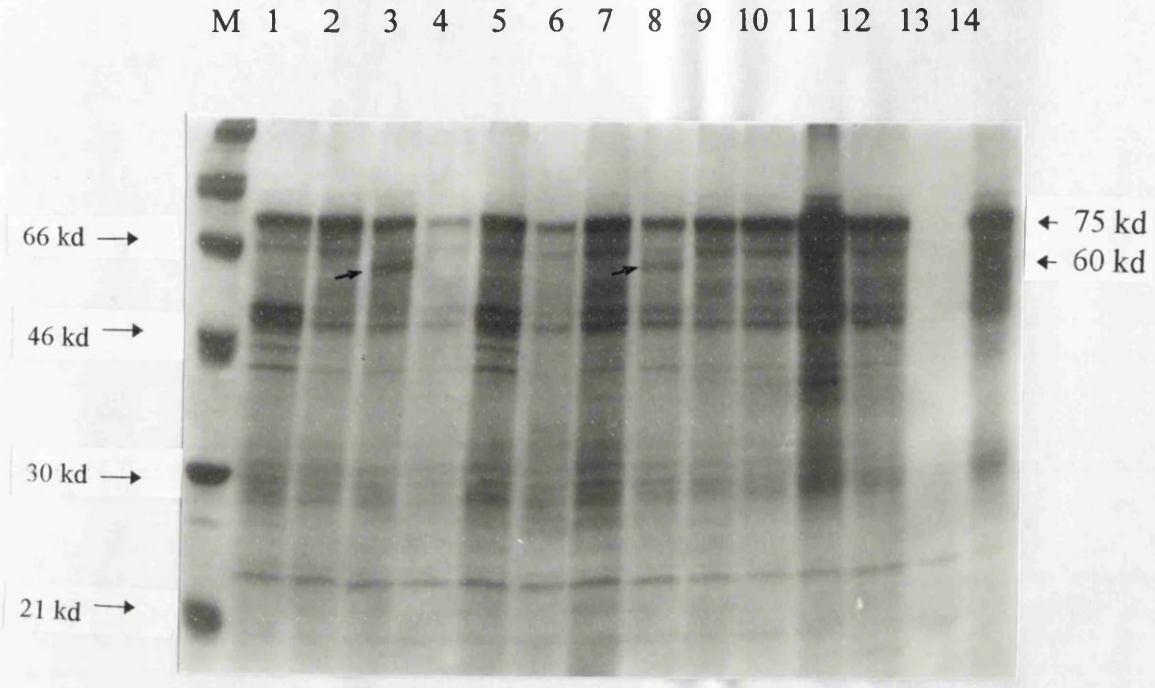


Figure 3. 6. Truncated protein in two members of one family, but not in the healthy members of the same family.

PTT assay on RT-PCR products from colorectal cancer patients. *In vitro* transcription and translation of RT-PCR products was carried out as described in "materials and methods", and the resultant labelled proteins were separated on a gradient sodium dodecyl sulphate polyacrylamide gel (see materials and methods). Arrows show normal products (75 kd) and truncated protein (60 kd) in Patient 21 and Patient 32 (lanes 3 and 8). M: protein standard size marker "rainbow" (Amersham).Lanes 1-2 normal, Lane 3: patient 21(affected member of the family), Lane 4 : normal, lane 5: P28(healthy member of family 21), lanes 6 - 7: normals, Lane 8: patient 32 (affected member of the family), Lane 9-12: normal controls, Lane 13: negative control, lane 14: positive control.

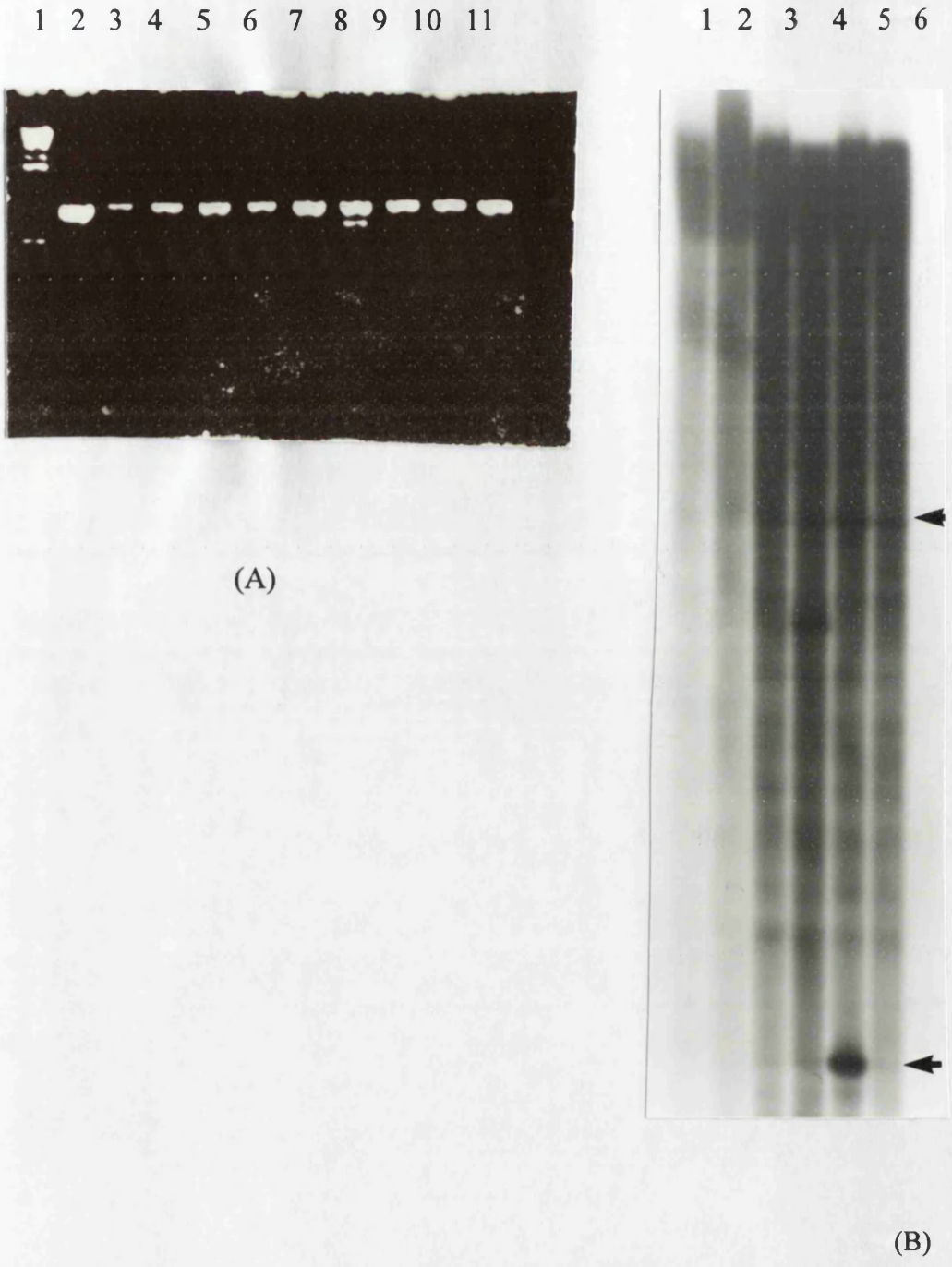


Figure 3. 7. A: Detection of short transcripts in the *hMSH2* gene by RT-PCR in Patient 21.
(A) Amplification of fragment E encompassing exons 13-16 by RT-PCR. The products were separated on 1% agarose gel and showed a normal product (750 bp) and a short products (574 bp). Lane 1: 1 kb standard DNA marker (Gibco BRL), Lanes 2 -7 normal control, lanes 9 -11: normal controls. Lane 8 shows additional product about 186 bp shorter corresponding to exon 15. B: CCM revealed a cleaved band in the segment E for Patient 21(shown by arrow). Lanes 1-3 and 6: normal controls; lane 4: positive control; lane 5: patient 21.

Exon skipping from m RNA transcripts

The CCM proved that the shorter PCR products are not artefacts. Application of direct DNA sequencing on RT-PCR products showed a mixed pattern of normal and mutant alleles, which was not easy to understand which exon has been skipped from the transcript (Figure 3. 8, A). Therefore, the RT-PCR products were loaded onto 1.5 % low melting Nusive agarose gel and the upper and lower bands corresponding to normal allele and mutant allele was excised from the gel. After re-amplification, direct sequencing revealed that the exon 15 was skipped from transcripts (Figure 3. 8, B). This is out of frame deletion of codons 820-878, creating a new termination codon 6 bp downstream of the splice site within exon 16.

A novel splice site mutation for exon 15

Exon 15 was amplified from genomic DNA, followed by direct sequencing in both normal control and Patient 21. On sequencing the genomic DNA, the splice donor site on intron 15 was found to have a G to C transversion at nucleotide position +5 (Figure 3. 9). This mutation was heterozygous in the tumours of the affected individuals in this family. Since the mother of this family had died it was not possible to show vertical transmission of this mutation.

This splice site mutation creates a restriction site for the enzyme, *Dde*I, and was used to identify the mutation in three other affected members of the family (Figure 3. 10). It was not found in the healthy member or 61 normal controls. This enzymatic assay was also used for screening of 116 CRC patients, but no alteration was found.

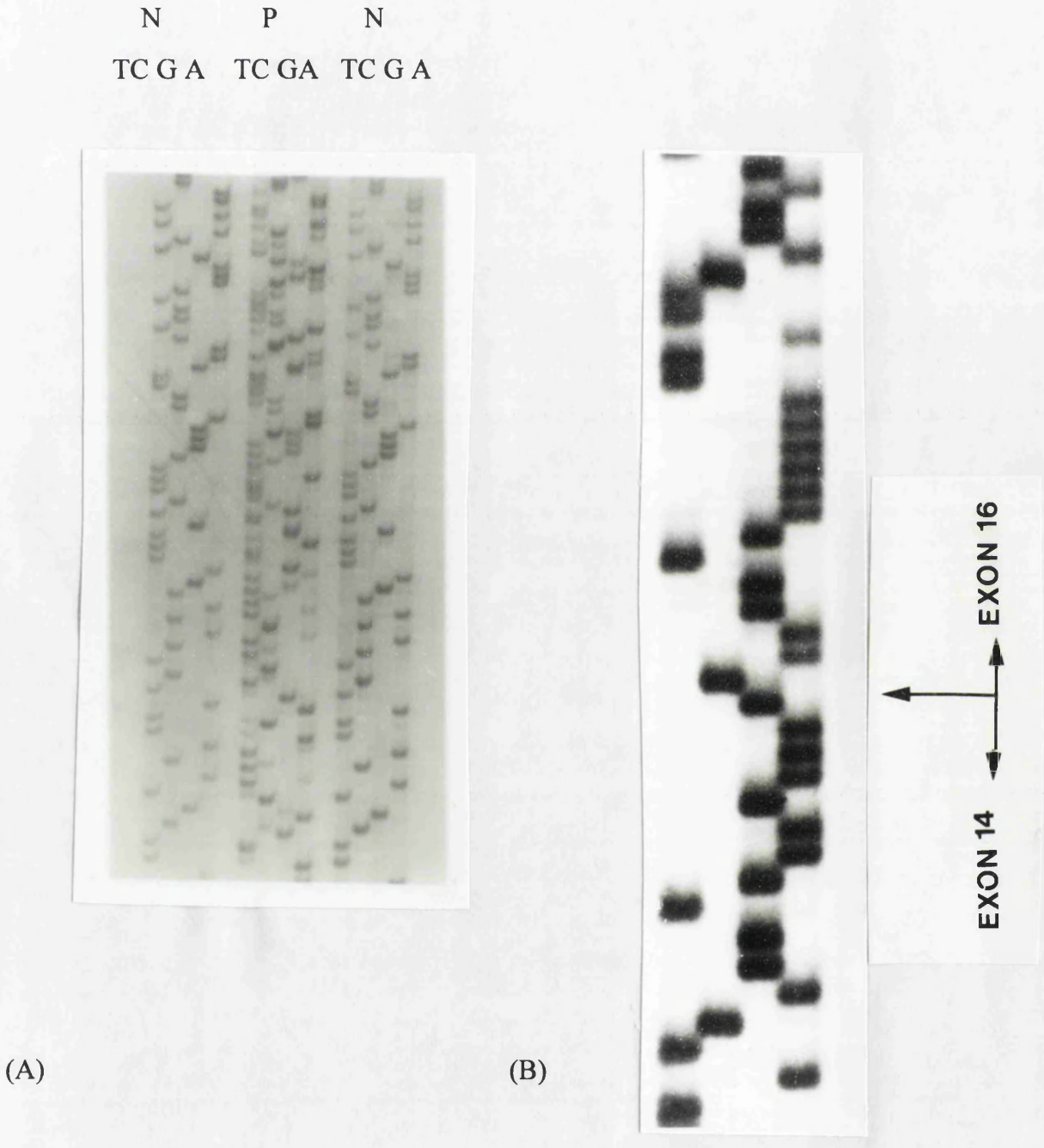


Figure 3. 8, A: Partial cDNA sequencing of mixed normal/mutant of patient 21 showed a parallel pattern corresponding to the normal allele and mutant allele. N: normal, P: patient. B: sequencing of the shorter RT-PCR products revealed that exon 15 is deleted from transcripts. Therefore, exon 14 joined directly to exon 16.

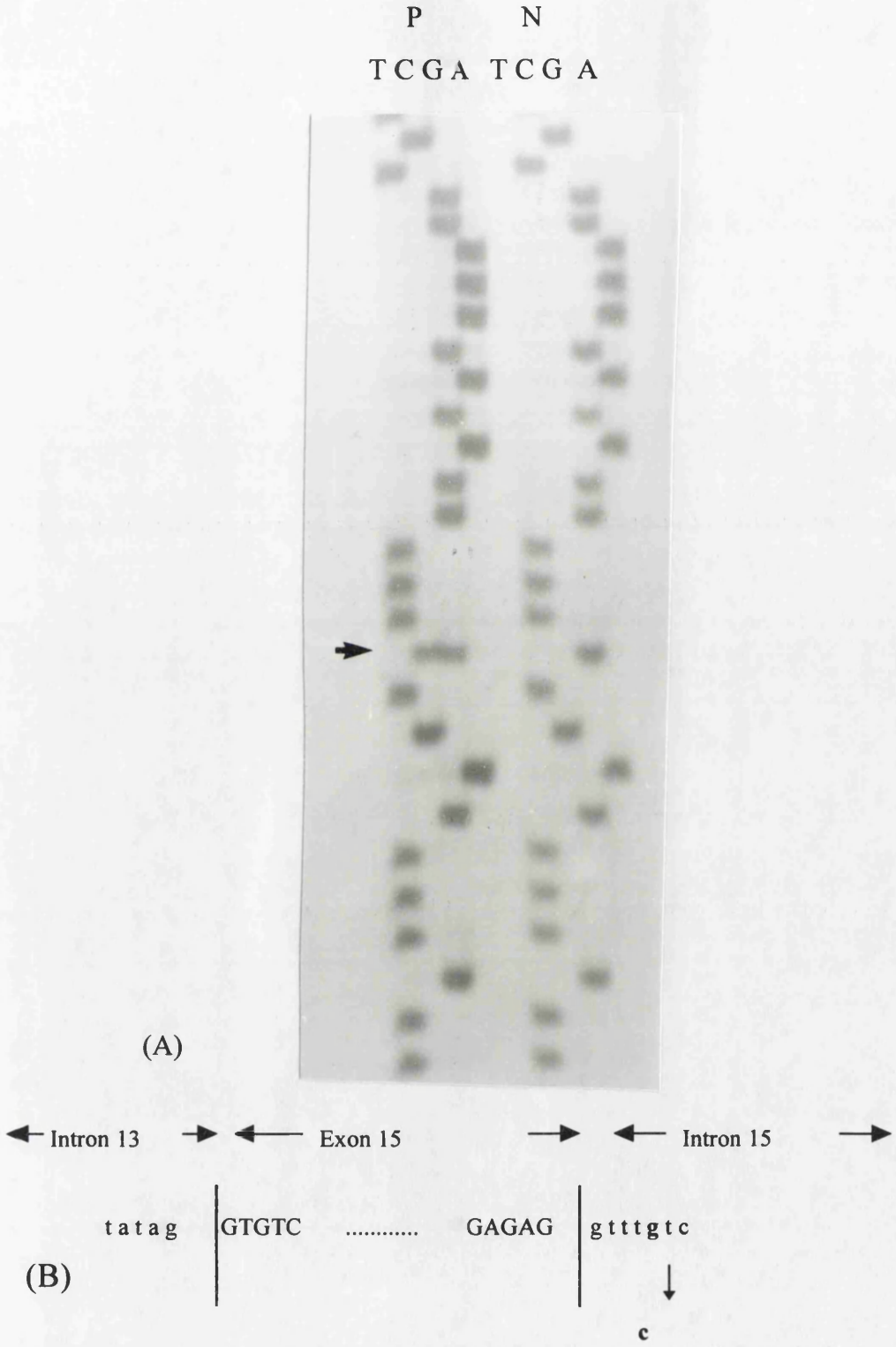
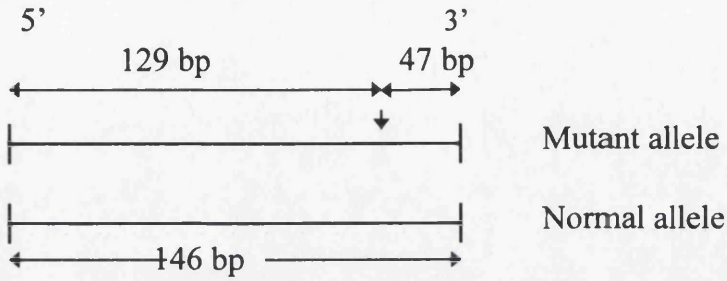


Figure 3.9 A: Direct DNA sequencing in patient 21 showed a new splice site mutation for exon 15 in nucleotide position +5. Altered nucleotide pointed by arrow and the exon/intron border is shown by a bar line. N: normal. P: patient. B: schematic presentation of exon/intron borders and the position of point mutation in patient 21 and four other affected members in her family. The exon/intron border is indicated by a vertical line.



Restriction map for the *Dde*I restriction enzyme in exon 15 and its 5' site of the *hMSH2* gene.

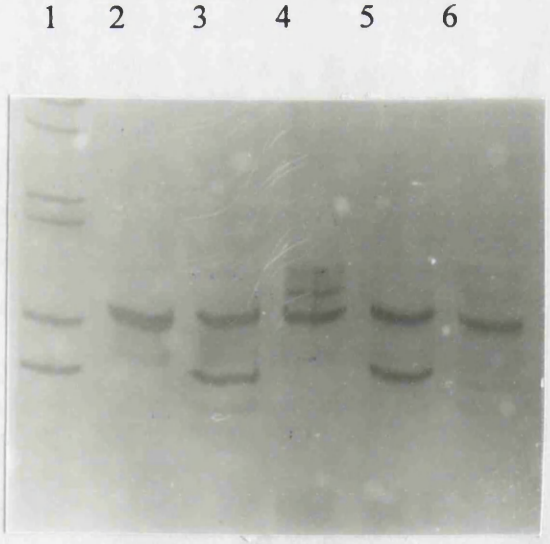


Figure 3. 10. Enzymatic assay with the enzyme *Dde*I.

A G to C transversion in intron 15 creates a restriction site for the enzyme *Dde*I. Application of this enzymatic assay revealed that all four affected individuals in the same generation bear the same mutation. This point mutation was not present in a healthy member or in 117 other controls. Lane 1: DNA size marker; lanes 2, 4, 6: normal controls; lanes 3 and 5: patients with splice site mutation.

Consensus value for donor splice site of exon 15

Shapiro and Senapathy (1987) described a method for potential consensus values at splice junction sequences. Based on this method, each of 8 nucleotides at the 5' splice junction site from position -2 through to position +6 have a particular score between 0 and 100 (see appendix 2). The final score for 5' splice site calculated by a formula is given below:

Score = 100(t - min t)/ (max t - min t)

where t is the total of percentage for eight nucleotides at 5' splice junction site and mint and max t are the minimum and maximum possible totals at 5' splice site.

The total score for normal and mutant sequence at 5' splice junction site is shown in table 3.3.

Table 3.3. Presentation of consensus values for normal and mutant sequences at donor site of exon 15.

Score	58	78	100	100	2	9	**84**	47	478
Normal	A	G	G	T	T	T	**G**	T	
Position	-2	-1	+1	+2	+3	+4	**+5**	+6	Total
Mutant	A	G	G	T	T	T	**C**	T	
Score	58	78	100	100	2	9	**6**	47	400

The min t and max t for each of the eight nucleotides is shown below:

Position	-2	-1	+1	+2	+3	+4	+5	+6	Total
Min t	13	4	0	0	2	8	5	15	47
Max t	58	78	100	100	57	71	84	47	595

Thus the consensus values for each normal and mutant sequences are given below:

Score = 100(t - min t)/ (max t - min t)

Normal score= 100 (478-47) / (595-47) = 78.36

Mutant score= 100 (400-47) / (597-47) = 64.18

Therefore the mutation at the 5' splice site of intron 15 reduces the consensus values by 14.18%, and may contribute to exon 15 skipping from the transcript as

demonstrated earlier.

3. 3. 2 Gross deletion in RNA revealed by RT- PCR

Patient 18 has three different tumours

This woman was operated on at the age of 68, due to of colon cancer. She had three different tumours each at different sites of her colon. One tumour was in the cecum (T1), the second one in the ascending colon (T2)and the last one in the transverse colon (T3).

MI in 8 of 9 markers in patient 18

Four specimens corresponding to three different tumours and one normal mucosa from patient 18 were studied for microsatellite instability. MI investigation with 9 different polymorphic markers showed RER+ in 8. Interestingly, tumours showed different patterns with different markers (Figure 3.11, A-B), see also table 3.1.

A Large RNA deletion encompasses exon 2 through to exon 6

Fragment A was amplified by RT-PCR as described before (materials and methods). This fragment was used as a target for re-amplification for a smaller fragment using a nested primer. Electrophoresis on 1% agarose gel on the smaller fragment showed two different bands, one corresponding to normal products and the other to shorter products. Subsequently, re-amplification of shorter products followed by direct sequencing was carried out. Direct sequencing on the normal allele showed a normal pattern, but exon 2 through to exon 6 was deleted from mutant allele (Figure 3. 12). This is an out-of-frame deletion of codons 71-359, creating a new termination codon at position +3 from the splice site in exon 7. No alterations in splice site junctions were found. Southern blotting of the genomic DNA, using a cDNA probe, after digestion with several restriction enzymes, produce no evidence of a deletion (Figure 3. 13).

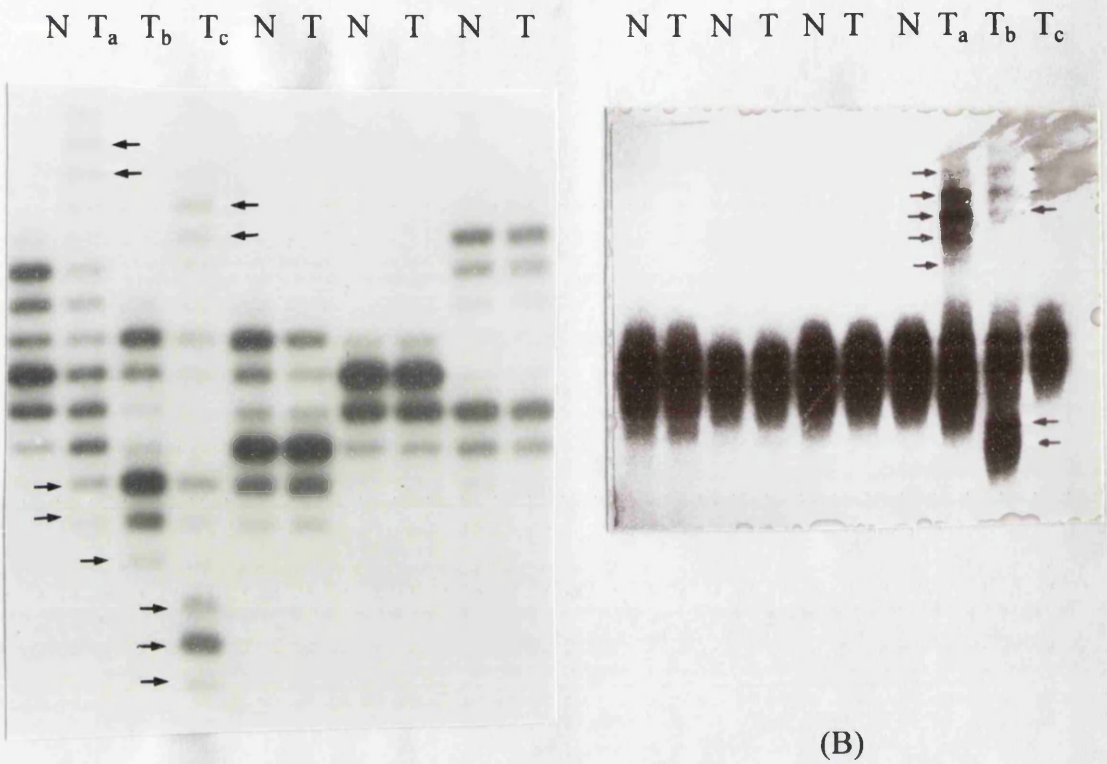


Figure 3. 11. Microsatellite instability with two different markers D2S123 (A) and LPLCA, B (B) for patient 18. This patient in MI point of view has been the best example in this research, as she showed RER+ with 8 of 9 markers. Interestingly, three different tumours in this particular patient showed different patterns with the same marker, suggesting different molecular pathology for different tumours. A: N: normal, Ta: cecum tumour, Tb: ascending tumour, Tc: transverse colon. Alteration in the tumour DNA is indicated by an arrow.

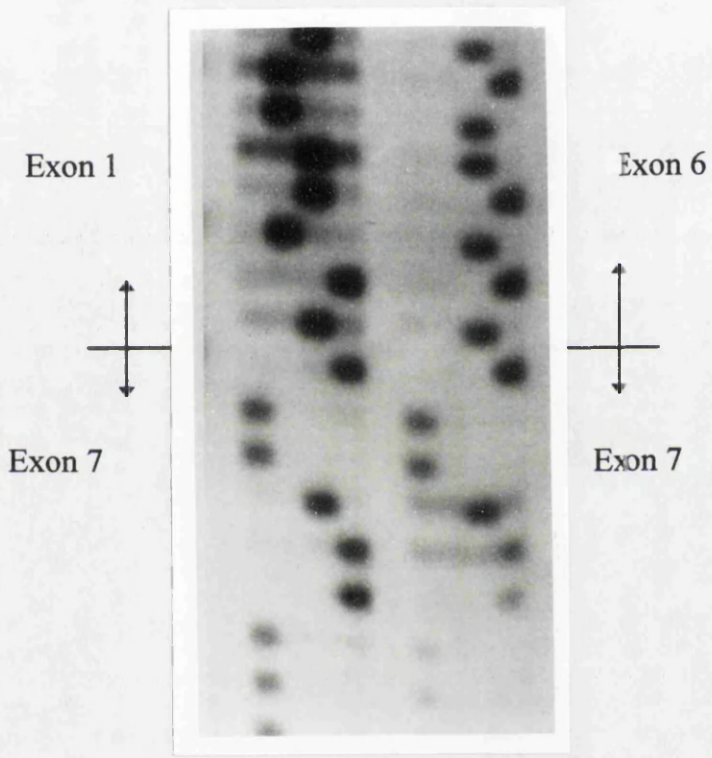


Figure 3. 12. Direct sequencing of RNA in patient 18.

Sequencing of mutant allele in patient 18 revealed a deletion from exon 2 through to exon 6 inclusive. N: normal, M: mutant.

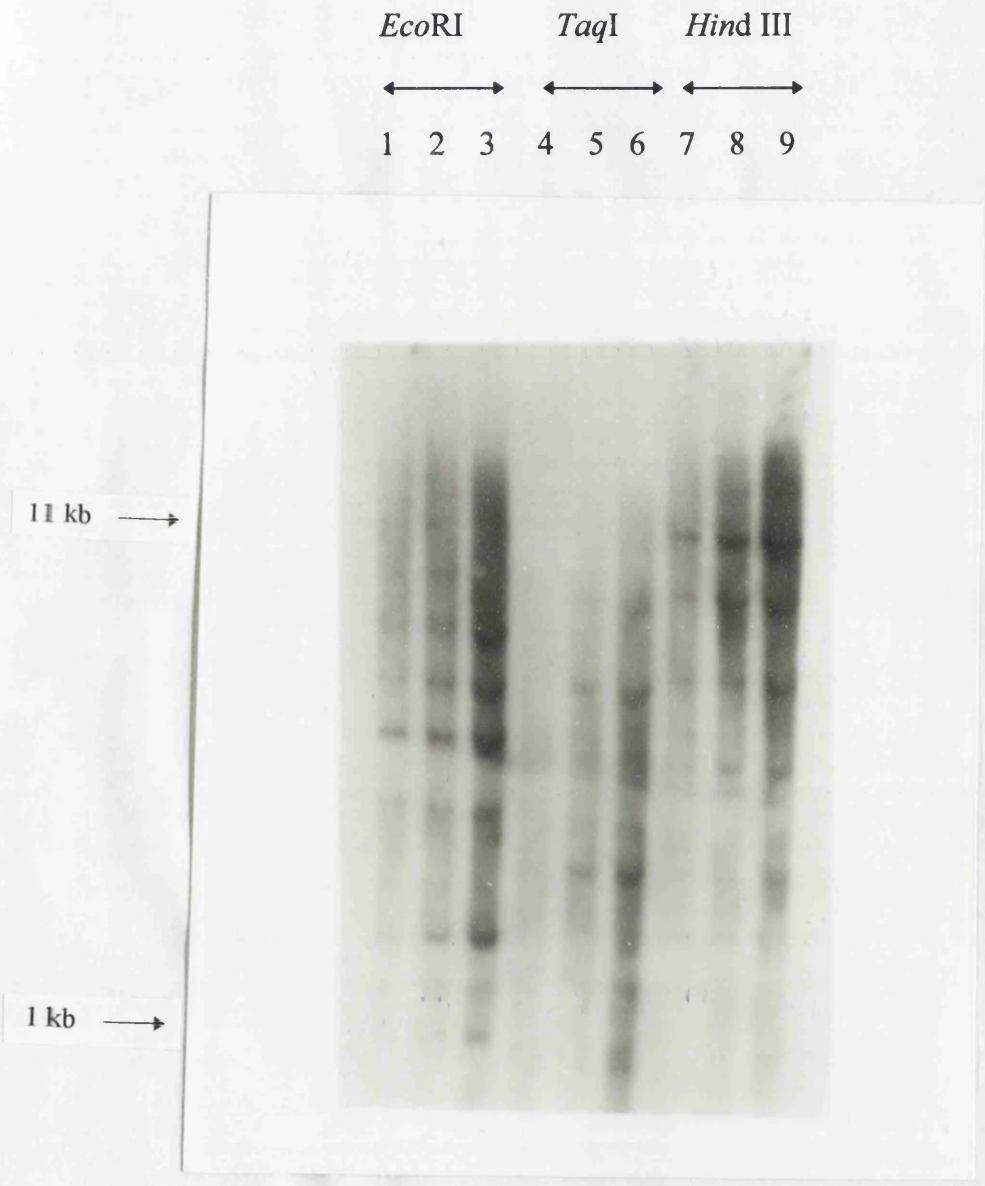


Figure 3.13. Southern blotting of genomic DNA in patient 18.

Southern blotting of the genomic DNA after digestion with three different restriction enzymes. The first three lanes were digested by *Eco*RI, the second three lanes were digested by *Taq*I and the last three lanes were digested by *Hind*III. Lanes 1, 2, 4, 5, 7 and 8 are normal controls. The lanes 3, 6 and 9 are patient 18.

3. 3. 3 . Multiple exon skipping in Patient 3

This man was operated on at the age 64, due to colon cancer. No more information about this patient and his family is available. MI studies were carried out with nine different polymorphic markers. MI was observed in two of nine markers including D18S34 and 635/636 (Figure 3.14).

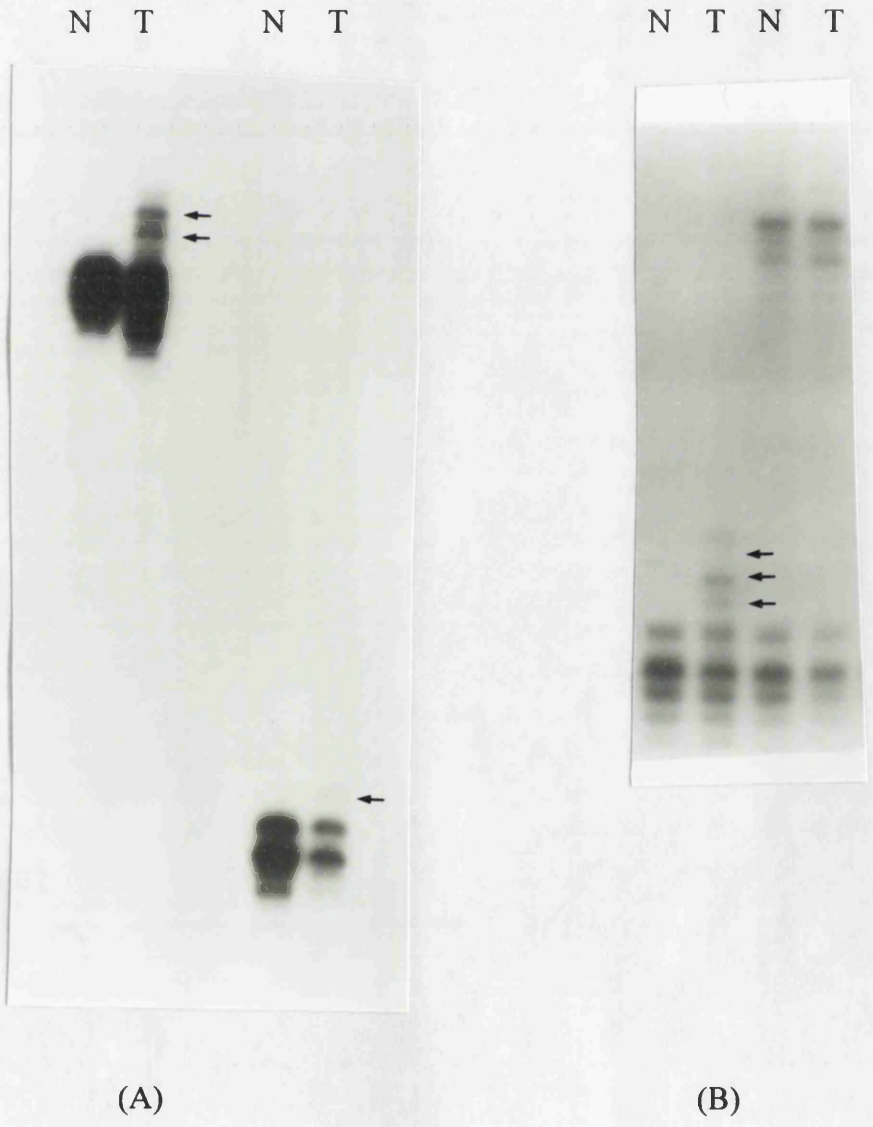


Figure 3. 14. Microsatellite instability in patient 3 using markers D13S34 (A) and 635/636 (B). Alterations in the tumour DNA are indicated by arrows.

Amplification of fragment A was carried out. The length of this fragment as mentioned before was 1892 bp. These long PCR products were used as target DNA for re-amplification. Re-amplification with two sets of primers as nested primers were carried out as shown in Figure 3.15. The downstream primer was the same for the two PCRs to amplify two different fragments in size, 1771 bp and 1733 bp, respectively. After gel electrophoresis short transcripts were present in both fragments, one band corresponding to normal allele and the other corresponding to mutant allele (Figure 3.16).

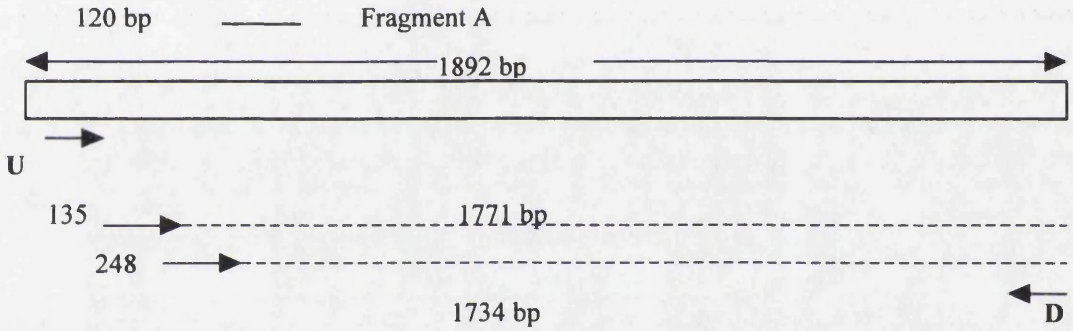


Figure 3.15. Schematic presentation of two different fragments for re-amplification of cDNA in patient 3. Primer D was used for both primers, 135 and 248, as a down stream primer.

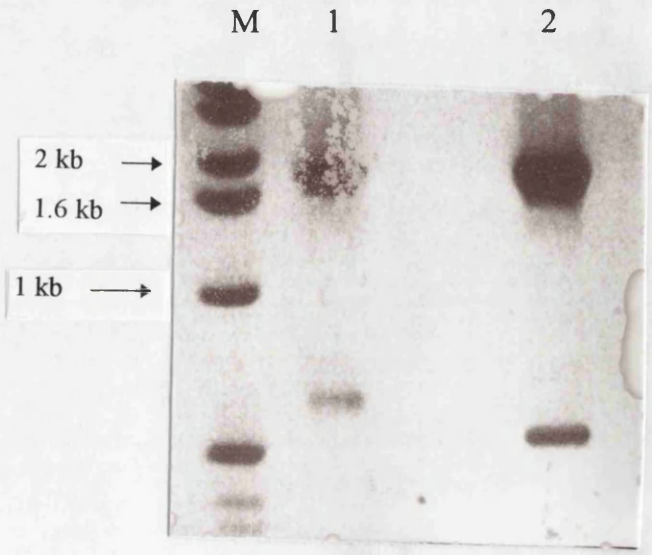


Figure 3. 16. RT-PCR products from patient 3 revealed a deletion from exon 2 through exon 8, when two different sets of primers were used as nested primers as shown in lanes 1 and 2. M: 1 kb DNA marker.

Re-amplification of the mutant allele in order to have enough targets for further analysis was carried out, followed by asymmetric PCR. Direct sequencing on the mutant allele demonstrated that exons 2-8 were deleted from transcripts, while the sequencing of the normal allele showed no alteration from normal sequence (Figure 3.17). This mutation is out of frame deletion of codons 71-466 and creates a new termination codon at position +12 from the splice site in exon 9. No other member of this family was available for our study.

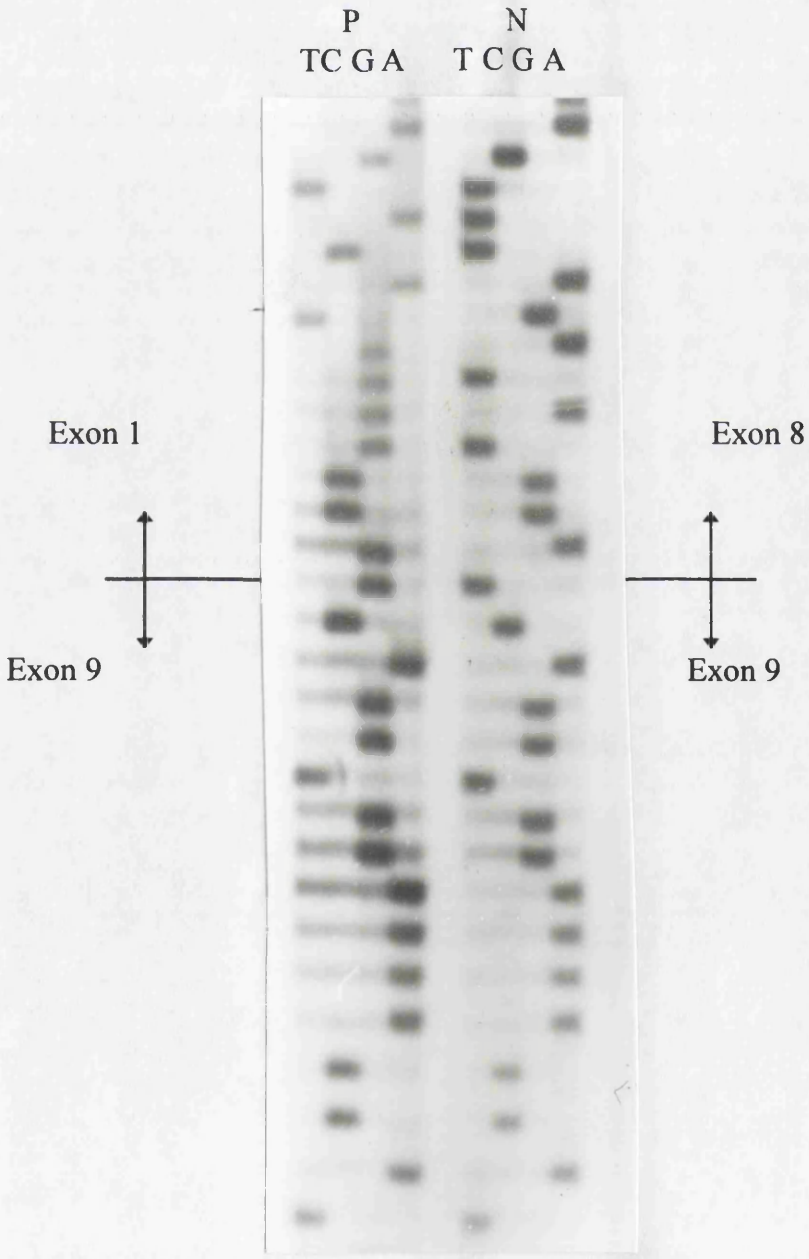


Figure 3. 17. Multiple exon skipping in patient 3.
Partial mRNA sequencing of normal control and mutant allele in patient 3 showing multiple exons skipping from exon 2 through to exon 8. P: patient, N: normal.

Genomic sequencing on all splice junction sites was carried out and no alteration from the normal sequence was found. However, Southern blotting carried out using different restriction enzyme was unable to produce evidence of a deletion.

3. 3. 4. Partial deletions in the *hMSH2* mRNA

In these patients no tumour DNAs were available for MI studies. None of them satisfied the Amsterdam criteria for HNPCC, but they have a positive family history of colon cancer and related disease.

Partial deletion including exon 9-12

This mutation was found for the first time in patient 17. In spite of the presence of five affected individuals in this family (including the proband), the HNPCC criteria are still not satisfied. This young patient was operated at age 17 because of colon cancer. Her mother died of colon cancer (the only first degree member), his uncle died at age 20 due to colon cancer and his aunt at age 58 because of the same disease. This is not sufficient to put this patient in HNPCC category. However, in the course of amplification of fragment D, encompassing exons 7-13, a short transcript was present for this individual but not for the rest of normal controls (Figure 3. 18A). Subsequently, re-amplification and asymmetric PCR were carried out on both mutant and normal alleles to produce single stranded DNA for sequencing. On sequencing it was demonstrated that the second part of exon 9 through to the first part of exon 12 was deleted from transcription (Figure 3.18, B). However, the same mutation was found in another patient with a positive family history of colon cancer but who did not satisfy the HNPCC criteria. All splice sites in exons 9 and 12 were intact, no alteration was found by genomic Southern blotting.

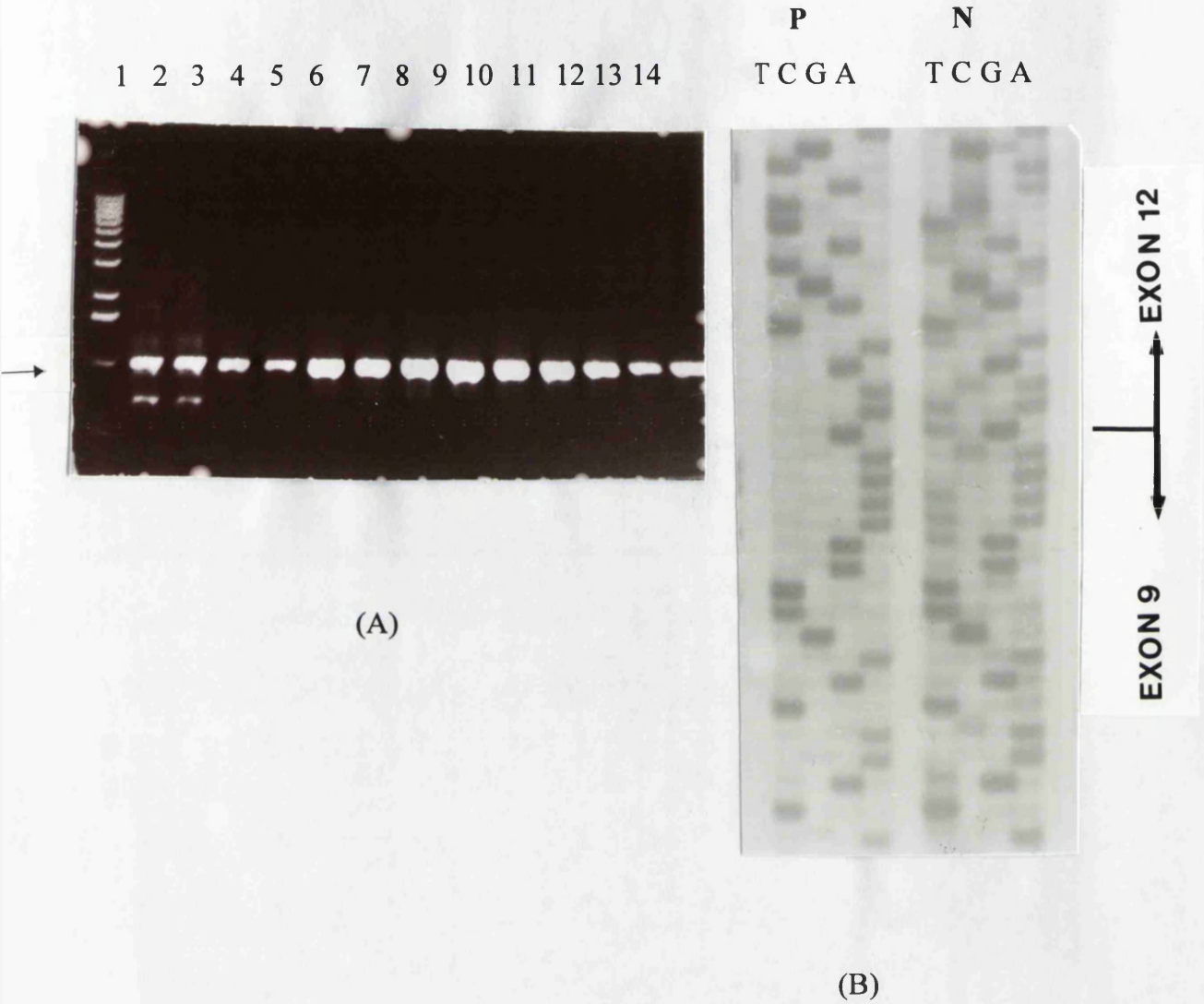


Figure 3. 18. A: RT-PCR products encompassing exons 7-13. Short transcripts were present for patients 17 and 30 but not in any other normal control. Lane 1: standard DNA size marker, Lane 2: patient 17, Lane 3: patient 30, Lane 3-14 : normal controls B: partial mRNA sequencing in patients 17 and 30 demonstrate deletion of exon 9-12. N: normal, P: patient.

Partial deletion in a patient with a positive family history of cancer

This patient (Patient 63) developed colon cancer in her 30s, while her parents were both healthy. Her grandfather (father's side) had stomach cancer in his 50s, her uncle (father's side) had throat cancer at age 50, and also her aunt (mother's side) had breast cancer.

Amplification of segment C in patient 63 was carried out. As described before, the products of large segments were used as a target for re-amplification with nested primers. Re-amplification of segment A with a nested primer showed two different fragments. one band corresponding to normal allele and the other shorter than normal. Sequencing of the normal band showed normal sequence, but sequencing of shorter products showed a partial deletion within exon 1 encompasses codons 8-49 (Figure 3. 19). This deletion is an out of frame deletion and create a new termination codon at nucleotide position +24 from intron 1. Subsequently, splice sites at 5' and 3' ends in exon 1 were sequenced and no alteration from the normal sequence was found. Genomic Southern blotting showed no evidence of deletion in DNA.

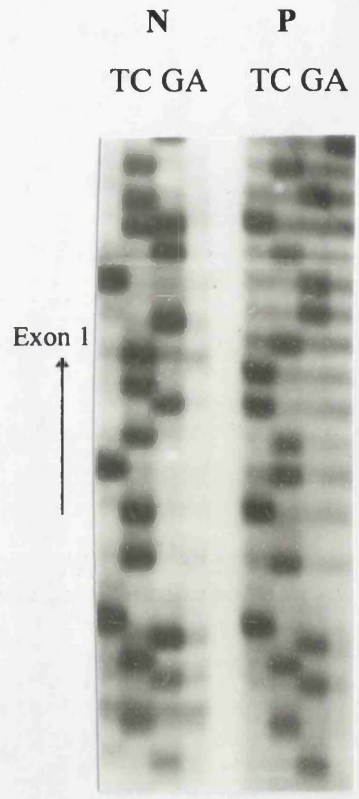


Figure 3. 19. Partial mRNA sequencing in patient 63. Sequencing of normal allele showed no alteration from published sequence, while sequencing of mutant allele documented a deletion within exon 1 encompassing codon 8 through to codon 49. This is an out of frame deletion and creates a premature termination codon within exon 1. N: normal, P: patient.

Partial deletion of exons 14-15 in patient 30

The bowel cancer is present on both maternal and paternal side of the proband. The proband had colon cancer at age 52, but bowel cancer and bowel problem were very frequently present in the family (five members in two generations). However, the Amsterdam criteria does not meet with this family. Amplification of fragment E, including exons 13-16, revealed a short transcript. Subsequently, direct sequencing of RT-PCR products showed a deletion from exon 14 to exon 15 (Figure 3.20), resulting from 380 bp deletion from transcripts. In order to determine the cause of this mutation, the splice site of both exons were sequenced and no alterations were found.

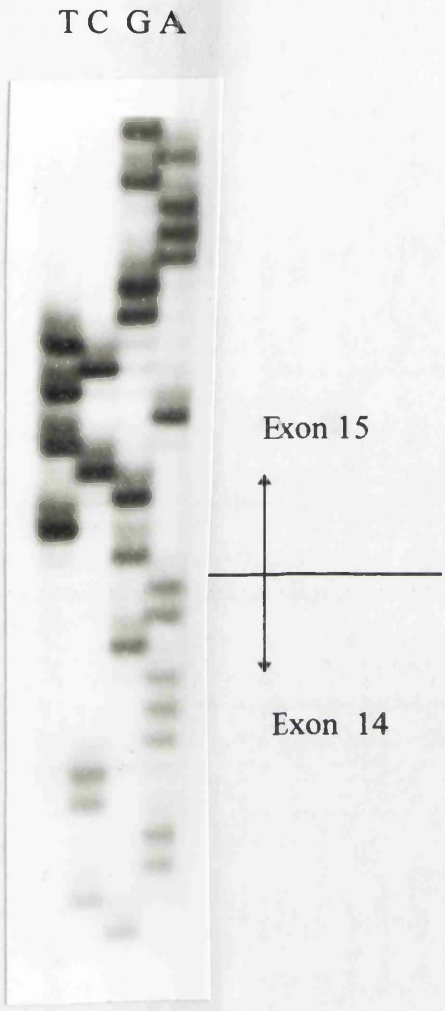


Figure 3. 20. Partial deletion in fragment E.

Partial mRNA sequencing in patient 30 revealed a deletion about 380 bp from exon 14 to exon 15. N: normal; P: patient.

Partial deletion of exons 4-6 in patient 73

Amplifications of fragment C for patient 73 showed two different bands. One of them corresponds to normal size, and the other was shorter than normal. On sequencing of shorter products a partial deletion was found from exon 4 through to exon 6 (Figure 3. 21), while the sequencing of the normal band showed normal sequences. Amplifications of individual exons (including exons 4 and 6) were carried out from genomic DNA. Subsequent direct DNA sequencing either in the coding sequence or in intronic flanking region of correspondence exons showed no alteration.

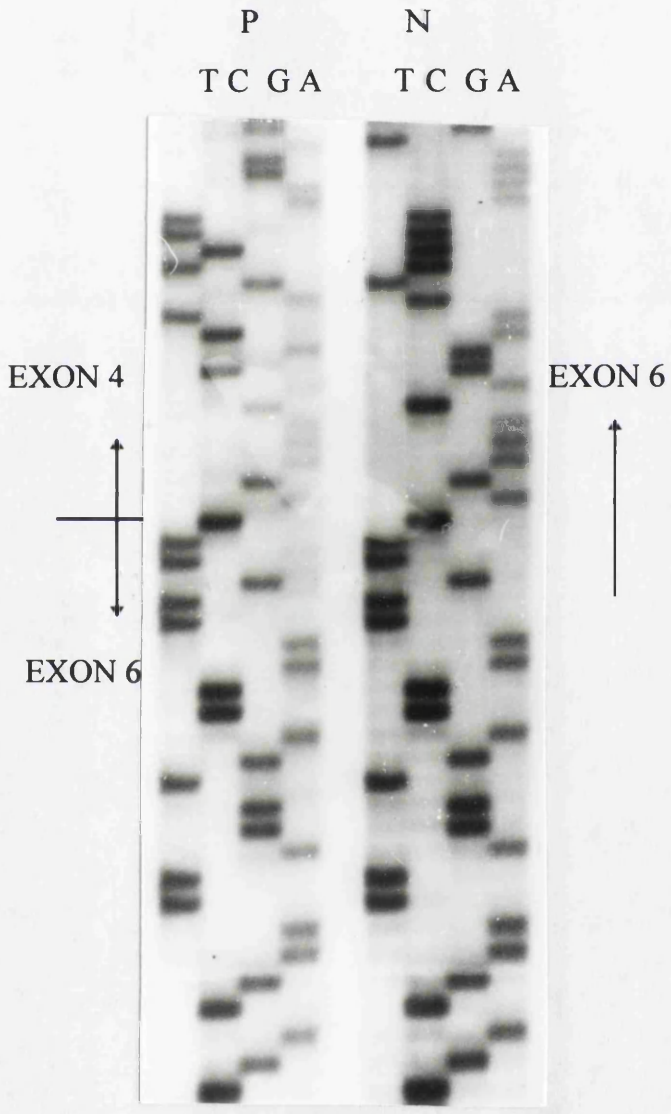


Figure 3. 21. Partial mRNA deletion for patient 73 from exon 4 - 6.
N: normal; P: patient.

The *hMSH2* transcript in the normal population

In order to find out if these mRNA deletions are normal variants in this population or not, RNA extraction was carried out from lymphocytes of 20 healthy individuals with the same method as was used for patient's RNA. RT-PCR was performed in two overlapping fragments. In the second round of PCR the whole

coding sequence of *hMSH2* gene was amplified by using three sets of nested primers. The profile and conditions of reverse transcriptase followed by PCR were the same as that used for patients. In the 20 healthy normal controls, no shorter products were observed, indicating the absence of these aberrant forms of mRNA in the normal population in the West of Scotland. Since some of aberrant mRNA mutations have been reported by different groups the possible frame shift mistake which could be made by *taq* DNA polymerase is being ruled out. In RNA splicing the branch point sequence, pyrimidine tract, 3' splice site and also 5' splice site are important determinants of splicing efficiency. In this study none of them was found to be abnormal. Another possibility is the presence of a gross genomic deletion as reported for example in Duchenne muscular dystrophy, but there is no documentation of this type of mutation in the mismatch repair genes. However, a 3.5 kb genomic deletion in the *hMLH1* gene has been reported as a cause of exon 16 missing from transcript. The possible site for these rare mRNA deletions could be somewhere within the introns which are not normally used for RNA splicing. However, the number of aberrant RNA splicing is not very low, although the mechanism of this particular phenomenon is not yet clear.

In summary, 47 families were investigated by a combination of different techniques including RT-PCR followed by PTT, CCM and sequencing. Splice site mutation was found in one case with MI+, two cases with multiple exon skipping and with MI+ phenotype, and four different partial mRNA deletion in different part of the gene (tumour was not available from these patients for MI analysis). Twelve patients with RER^- phenotype were tested and no alteration was found. No alteration was also found in patient 5 with RER+ phenotype. There was no RNA from patient 22 and also patient 26 with RER+ for mutation analysis with above techniques. None of the aberrant RNA was present in normal population when 20 healthy individual was tested.

3. 4. SSCP analysis of the *hMSH2* gene

A combination of DNA and RNA based methods could increase the efficiency of the mutation screening programme. Our strategy was based on RNA

based methods, but DNA has its own advantages in that it is more stable than RNA and it is very convenient. Exons 1, 2, 3, 5, 6, 7, 12, 13, 14, and 15 were however, chosen for SSCP analysis. In these exons mutations and polymorphisms have been reported by others. However, 59 DNA samples from 46 families were tested for genomic alteration by employing SSCP (family 21 with a germ-line splice site mutation was excluded from this part of study).

3. 4. 1. An infrequent variation at codon 322 in exon 6

Exon 6 was amplified by PCR, as described in materials and methods. The SSCP results from this exon showed a shift band in two patients, in comparison to other normal controls (Figure 3.22). Asymmetric PCR was carried out to produce single stranded DNA for direct sequencing. On sequencing of genomic DNA in exon 6 a G to A transition was found at codon 322 (Figure 3. 23). This mutation changes glycine to asparatic acid, and creates new sites for the following enzymes: (i) *Adh* I (ii) *Ple* I (iii) *Hinf* I. This transversion also abolished two sites for enzymes *Bsr* I and *Cvij* I. As the *Hinf* I was more popular than the others this enzyme was chosen to prove the presence of this mutation, and also for screening of the rest of the samples (Figure 3. 24) and also for screening 30 normal controls and it was present in three of them (5%).

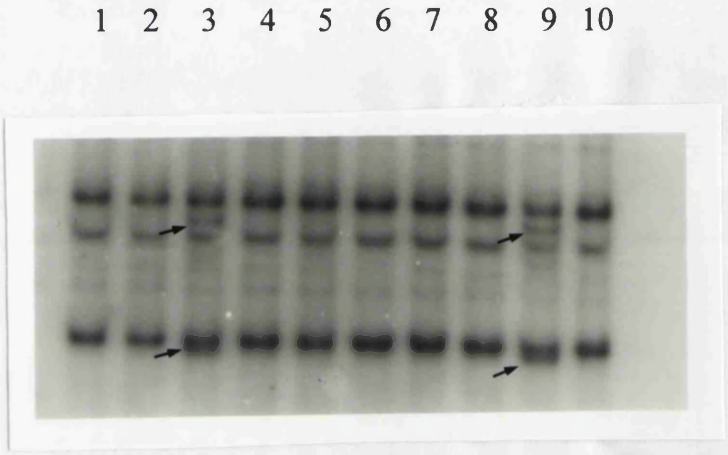


Figure 3. 22. Detection of mutation at the codon 322 by SSCP analysis.

The SSCP result of exon 6 showed shifted bands in two different individuals in comparison to normal controls. The shifted bands are indicated by arrows in lanes 3 and 9.

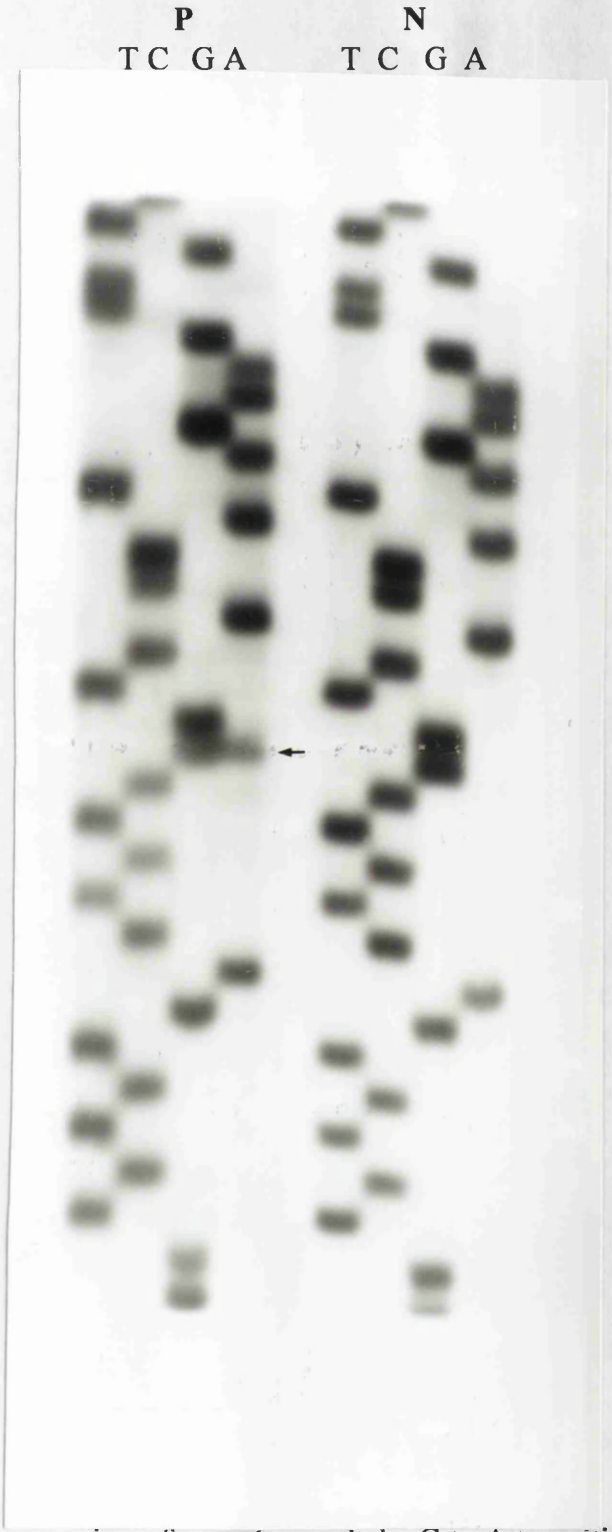
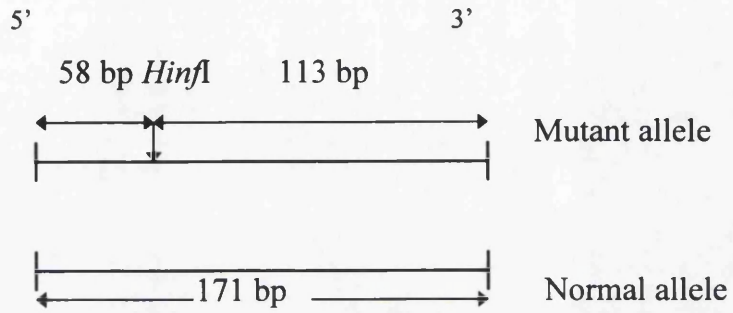


Figure 3. 23. Sequencing of exon 6 revealed a G to A transition.

On sequencing at genomic DNA, a G to A transition was found in two unrelated individuals but not in any other patients. This alteration changes glycine to asparatic acid. N: normal, P: patient



Restriction map for the enzyme *Hinf*I in exon 6 and its flanking region of the *hMSH2* gene.

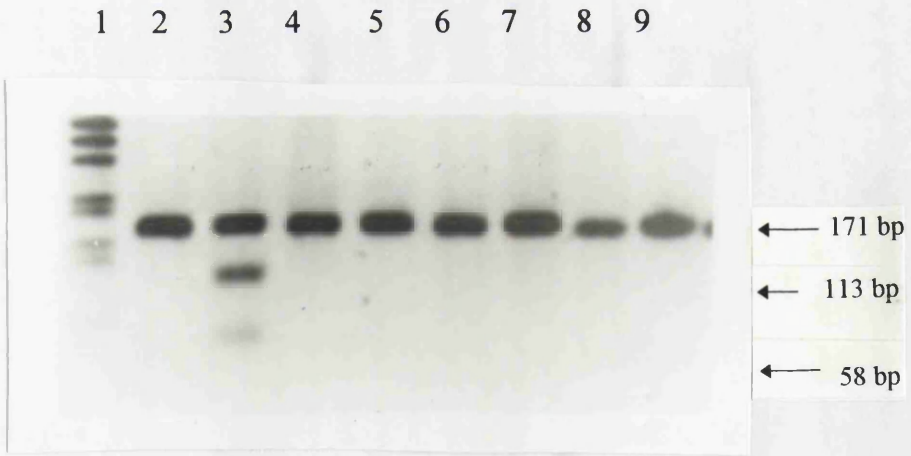


Figure 3. 24. PCR/RFLP analysis detecting point mutational changes in codon 322 of the *hMSH2* gene. *Hinf* I cleavage of mutant allele results in two fragments: 113 bp and 58 bp, while the length of normal allele is 171 bp. Lane 1: DNA size marker, Lanes 2, 4-9: normal controls, Lane 3: heterozygous individual.

3. 4. 2. A normal variation in the short polypyrimidine tract in intron 12

This is one of the first alterations in this particular gene, as reported by Fishel et al. (1993) when they published the sequence of the *hMSH2* mRNA. However, the results of SSCP in exon 13 revealed an alteration in many patients indicating a relatively high incidence in the region of the gene (Figure 3. 25). Direct DNA sequencing of exon 13 and flanking regions showed an alteration in short polypyrimidine tracts in intron 12, a T to C transition (Figure 3.26). This transition only creates a new site for the enzyme *Sfc*I. The presence of this alteration was searched for by an enzymatic assay using *Sfc*I, and it was found that 16% of patients bear this alteration (Figure 3. 27). This frequent mutation was also present in 15% of normal controls (86 chromosomes from normal individuals were tested). It seems that this alteration is not a harmful mutation. No alteration was found in the coding sequence of exon 13 in the splice junction site.

Figure 3. 25. The SSCP analysis of exon 13 and its flanking regions in the *hMSH2* gene. The exon 13 and flanking regions were amplified by PCR and the products were loaded on to the MDE gel. An alteration was found in many patients. The shifted bands are indicated by arrows in lanes 2-4.

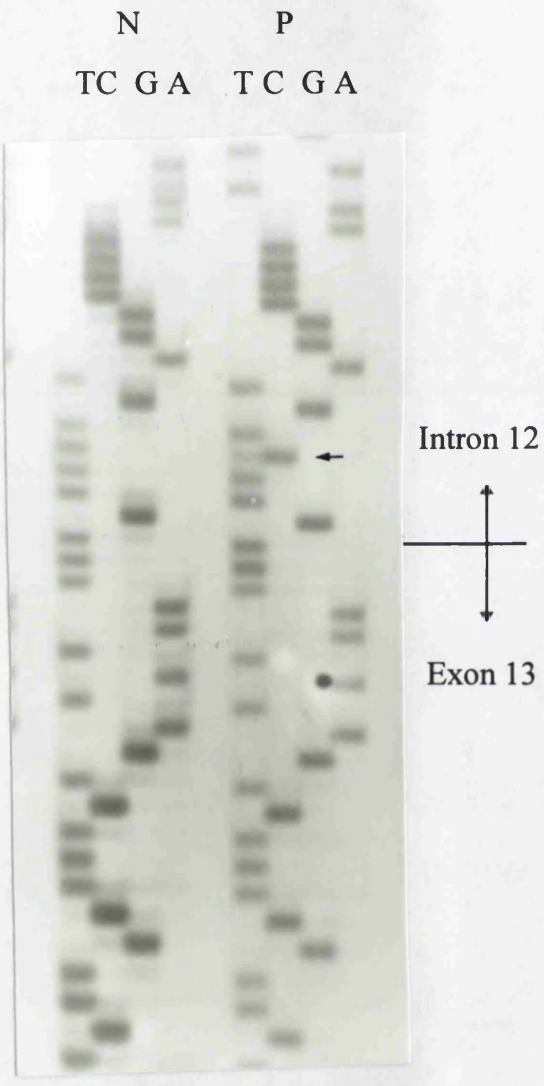
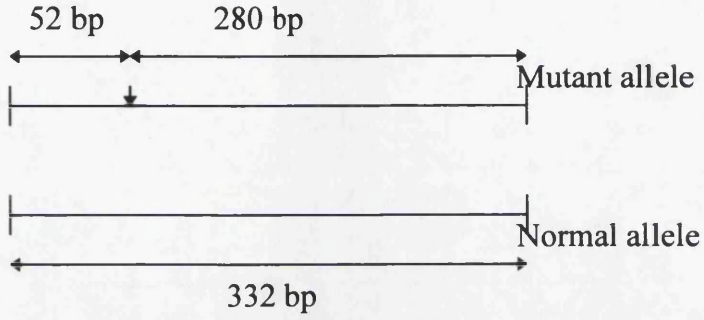


Figure 3. 26. Sequencing of the exon 13 and flanking regions revealed a T to C transition in short polypyrimidine tracts in the 3' end of intron 12 at position -6 from splice site. The mutation is indicated by an arrow and the splice site is indicated by a bar line. N: normal, P: patient.



Restriction map for the *Sfc*I in exon 13 and flanked region in the hMSH2 gene.

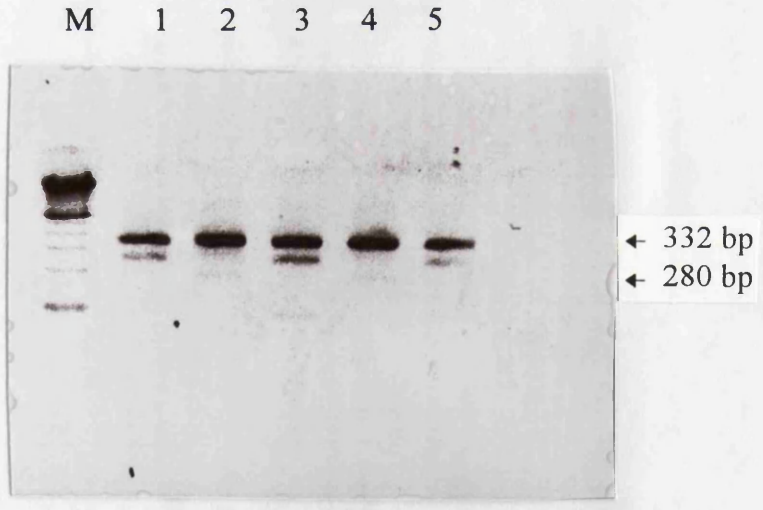


Figure 3. 27. PCR/RFLP analysis of exon 13 in the *hMSH2* gene.

Cleavage with *Sfc* I in the mutant allele produced two different fragments of 280 bp and 52 bp. It is very hard to see a 52 bp on the normal agarose gel. However, the size of normal allele was 332 bp and is indicated by an arrow. Lane M: DNA size marker, Lanes 1, 3 and 5: heterozygous patients, Lanes 2 and 4: normal controls.

3. 4. 3. A normal variation in intron 1

No mutation or even polymorphism has been reported in the coding sequence of exon one of the *hMSH2* gene. Therefore it was decided to screen this individual exon by SSCP analysis. Amplification of exon one and it's flanking region was carried out followed by SSCP analysis. The results of SSCP showed a polymorphism pattern in this exon (Figure 3. 28). The coding sequence of this exon has more than 60% Gs and Cs, therefore with ordinary sequencing protocols it was not possible to sequence this part of the gene. This problem was solved by using dITP for dGTP. However, on sequencing of the genomic DNA a C to G transversion was discovered at nucleotide position +9 in intron one (Figure 3.29). Notably, this variation was found and published later by a group from Edinburgh (Bubb *et al.*, 1996). The frequency of this polymorphism was found to be 45% when 66 normal chromosomes were tested.

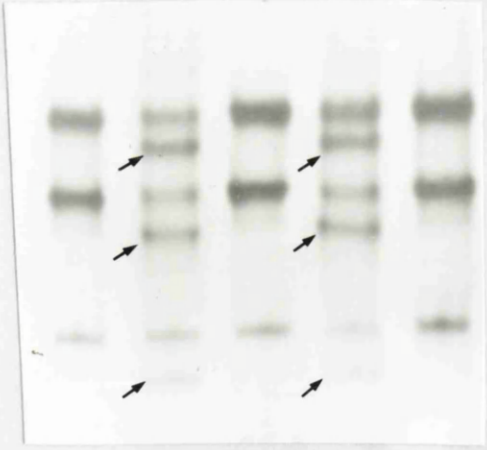


Figure 3. 28. The SSCP analysis of exon one in the *hMSH2* gene.
A polymorphism pattern was observed in the SSCP of exon one. The MDE gel was silver stained. The altered bands are shown by arrows.

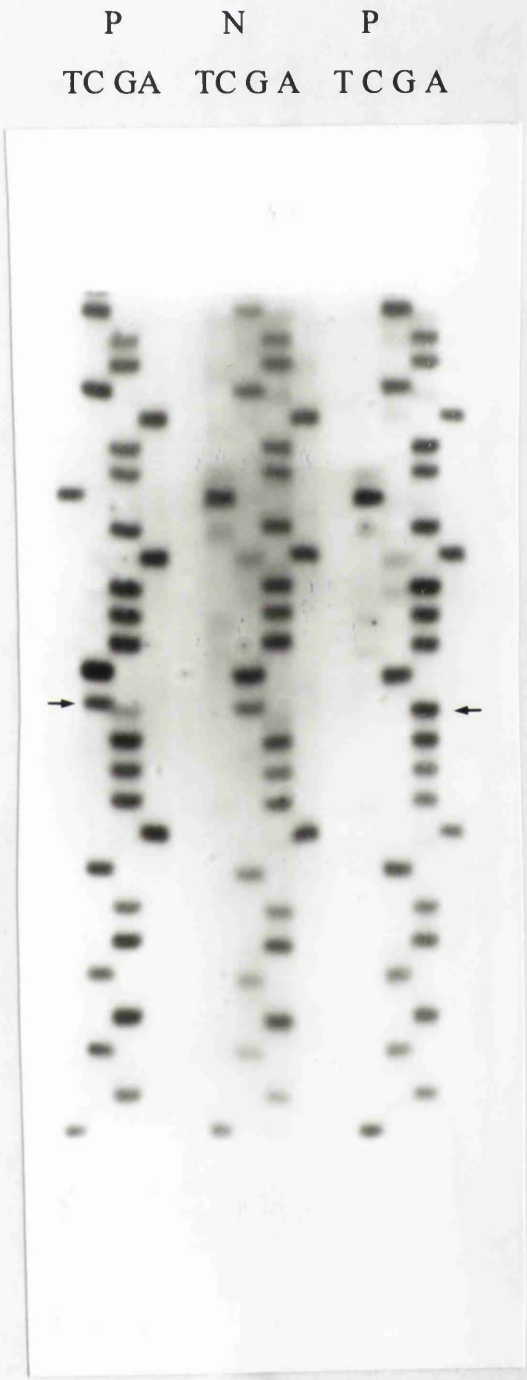


Figure 3. 29. On sequencing exon 1 and it's flanking region, a C to G transervsion was found in the nucleotide position +9 from splice site. The alteration is indicated by arrows. N: normal, P: patient.

3. 5. Variation from published sequence

In the course of mutation analysis in this project an alteration from the published sequence was observed in the *hMSH2* gene. The importance of this alteration in not clear yet, but it will present as a variation from the published sequence in the *hMSH2* gene. Interestingly, this alteration is located in the untranslated region (UTR) in the 3' end of this gene. However, a **G** instead of **T** was observed in nucleotide position -13 from termination codon (Figure 3. 30).

3. 6. Intronic mutations are frequently present in the *hMSH2* gene

Leach *et al.*, (1993) used two different sets of primers to amplify the conserved region of *hMSH2* from genomic DNA. They used one set of primers which were able to amplify a 1.4 kb fragment containing *hMSH2* codons 616-705. The entire sequence of intron 12 would amplify with this primer. This primer was used to amplify this region followed by CCM analysis in this project. Thirty-four DNA samples from 32 families including 22 unknown families, two HNPCC, three polyposis colorectal cancer, and five families with family history of cancer were screened by CCM. Cleaved bands were found in five individuals from four families, including two affected members of an HNPCC family (Patient 21 and patient 32 from family 8906), two unknown patients and one patient with CRC history. Some results of this search are presented in Figure 3. 31. It is likely that the cleaved bands are located in the non-coding sequences of the gene. However, none of these mutations were fully characterised by sequencing.

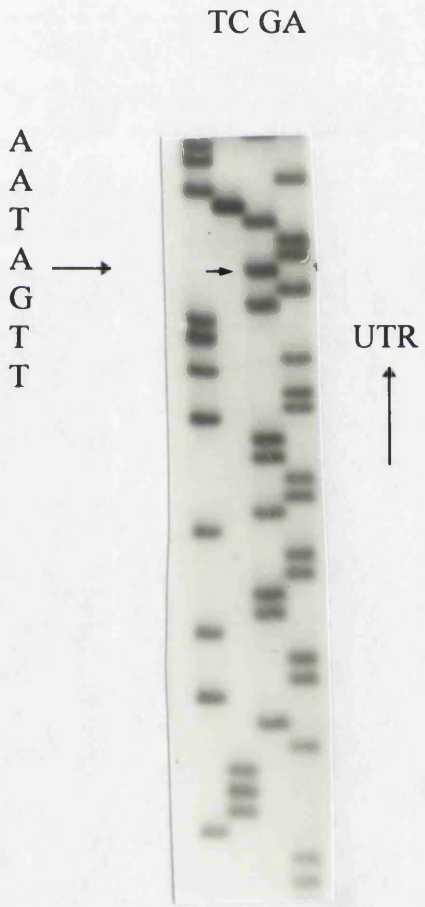


Figure 3. 30. A variation from the published sequence in the *hMSH2* gene.

A variation at the 3' end of the *hMSH2* gene was observed in the genomic DNA. At nucleotide position -13 from the termination codon, a **G** instead of **T** was observed. The nature of this alteration is not yet clear. The published sequence is given beside the photograph. The alteration is shown by an arrow. UTR: untranslated region, TGA: stop codon.

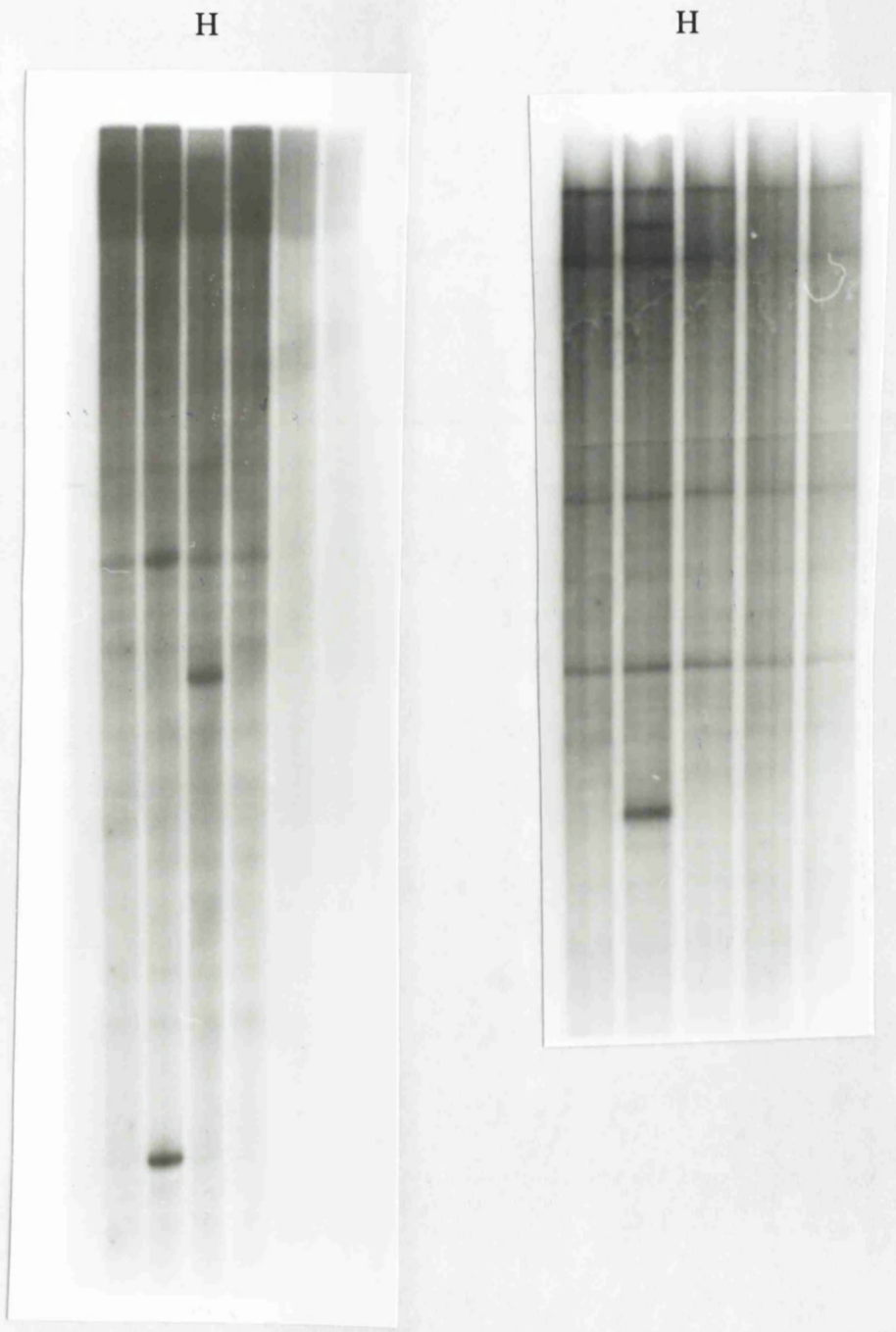


Figure 3. 31. The CCM analysis of the fragment containing *hMSH2* codons 616-705. The cleaved bands are indicated by arrows. H: hydroxylamine.

3. 7. Mutation in *hMLH1*

The *hMLH1* is a second gene predisposing to colorectal cancer and is located on the short arm of chromosome 3 (Lindblom *et al.*, 1993; Papadopoulos *et al.*, 1994). It is believed that the majority of *hMLH1* mutations cluster at the exonic region 15 and 16 (Wijnen *et al.*, 1996). These two individual exons were therefore amplified by the polymerase chain reaction, using primers in the flanking regions of the corresponding exons. Subsequently, SSCP analysis was carried out as described in chapter two. The SSCP results from exon 15 showed a band shift in many of the patients (Figure 3. 32 a). On sequencing the genomic DNA an A to G transition was found at position -19 in intron 14 (Figure 3.32 b). This alteration abolishes the restriction site for enzymes *TSP* 45 I and also *Mae* III in the mutant allele. The *TSP* 45 I was chosen to screen the normal population. By means of this enzymatic assay it was found that 44% of the normal population bears this mutation (Figure 3. 33).

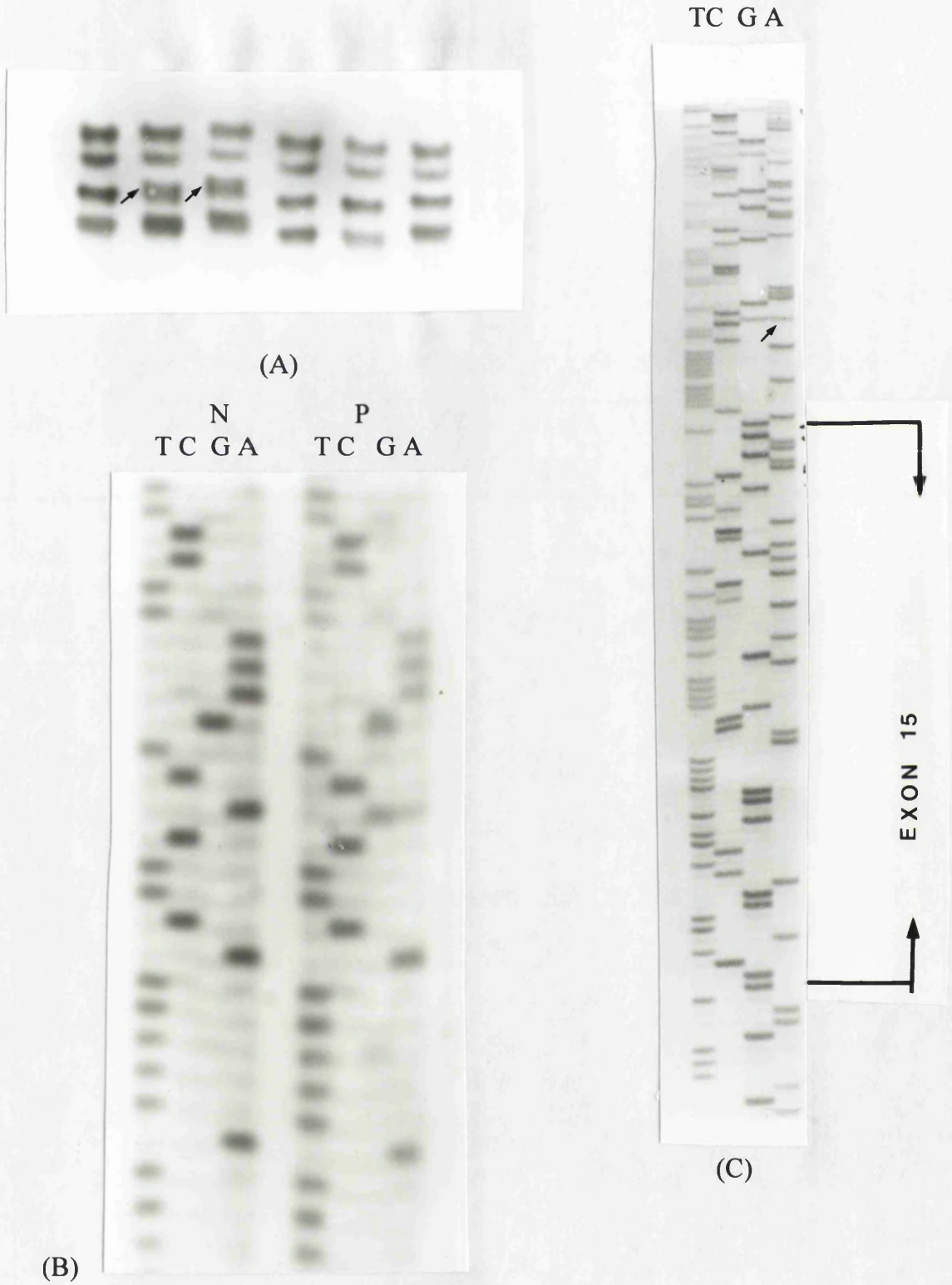
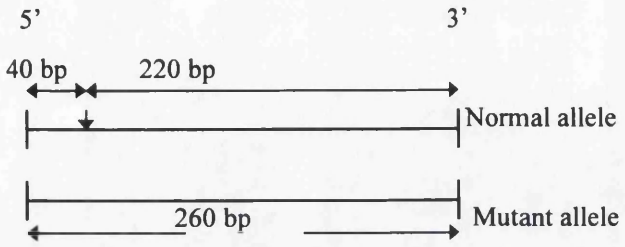


Figure 3. 32A. The SSCP result from exon 15 in the *hMLH1* gene. The alterations are indicated by arrows. (B) On sequencing of exon 15 and flanking region an A to G transition was found at position -19 in many patients and also in the normal population. The alteration is shown by an arrow. C: exon 15 and its flanking region. N: normal; P: patient.



Restriction map for the *TSP* 45 I enzyme in exon 15 and flanking region of the *hMLH1* gene.

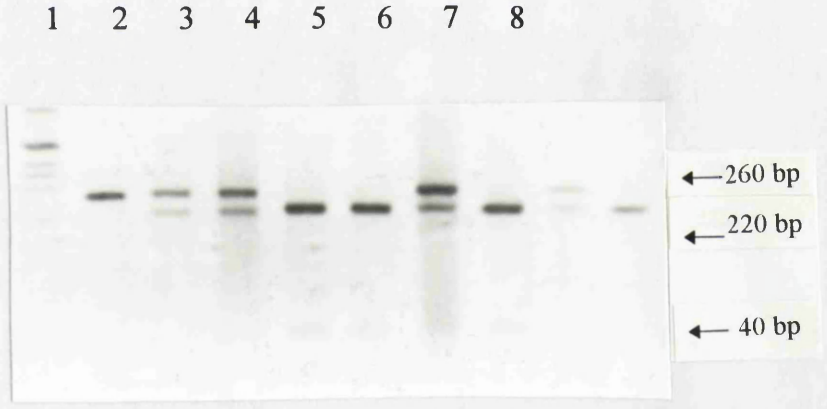


Figure 3. 33. PCR/RFLP analysis of exon 15 in the *hMLH1* gene.

Cleavage with *TSP* 45 I in the normal population revealed a high frequency of this polymorphism. Lane one : DNA ladder. Lane 2: undigested PCR products. Lanes 3, 4 and 7: heterozygous individuals. Lanes 5, 6 and 8: homozygous individuals.

In summary, 10 different exons including exons 1, 2, 3, 5, 6, 7, 12, 13, 14, and 15

from the *hMSH2* gene and also exons 15 and 16 from the *hMLH1* gene were tested by employing SSCP in 59 individuals from 46 families. However, three alterations including one rare normal variation in the coding sequence of the exon 6 and two polymorphisms in intronic regions of the *hMSH2* gene and also one polymorphism in the intronic region of exon 15 in the *hMLH1* gene were found. None of these variations could be the cause of disease as it was found in normal population and the frequency of them obtained by using appropriate restriction enzyme.

3. 8. Summary of mutations

The pathological and nonpathological mutations found in this study are presented in table 3. 3.

Kindreds	Gene	exons/introns affected	cDNA changes	Genomic changes	MI status	Protein alteration
21	*hMSH2*	15	exon 15 deleted	G/gtttgt to G/gtttct	+	Out of frame deletion
18	*hMSH2*	2-6	Exons 2-6 deleted	NI	+	Out of frame deletion
3	*hMSH2*	2-8	Exons 2-8 deleted	NI	+	Out of frame deletion
17, ...,	*hMSH2*	9-12	Exon 9-12 deleted	NI	NT	Out of frame deletion
63	*hMSH2*	1	Exon 1 deleted	NI	NT	Out of frame deletion
38,...	*hMSH2*	INV 1, +9	-	C to G	NT	No alteration
17,...	*hMSH2*	INV 12, -6	-	T to C	NT	No alteration
39, ...,	*hMSH2*	6	G to A	G to A	+/NT	Glycine to asparatic -acid
29	*hMSH2*	14-15	exons 14-15 deleted	NI	NT	Out of frame deletion
73	*hMSH2*	4-6	exons 4-6 deleted	NI	-	In frame deletion
5, 8, ...	*hMLH1*	INV 14, -19	-	A to G	-	No alteration

NI: not identified, NT: no tumour, INV: intron

Table 3. 3. A summary of mutations which are identified in this study.

CHAPTER FOUR

DISCUSSION

4. 1. Microsatellite analysis in CRC patients

Replication error or microsatellite instability is a new phenotype of human cancer and also some other diseases. The MI in these disease arises from alterations in normal regions of the genome consisting of short sequences of repeated DNA. Alterations in length of microsatellite sequences between normal DNA and tumour DNA occurs in about 90% of HNPCC cases and in about 15% of cases of non-familial colorectal cancer. MI is also found in sporadic endometrial, gastric cancer and lung cancer which is not associated with HNPCC.

MI was chosen as a first screening step in order to assess the level of MI in the tumour of patients with colorectal cancer in the West of Scotland. Thirty unselected pairs of normal and tumour tissues were tested for genomic instability at a minimum of two and a maximum of nine microsatellite loci. MI+ was found in nine patients in at least one marker and seven patients were found to be MI+ in at least two or more (up to eight) markers. There was no information about the history of disease and possible affected members of the family for a majority of cases (28 families). Of them, six families were found to be MI+ (21.5%) and a harmful mutation was found only in one family (17%). Both families with a history of colorectal cancer were found to be MI+ (patient 21 and 78) with 5 and 2 affected members in the families, respectively. A gemline mutation was found in one of them (patient 21). No RNA was available for mutation analysis from patient 78. There were no significant differences between the average age of patients with and without MI, additionally no significant differences were observed between males and females with or without MI. Although the connection between MI and defective MMR was established in human tumours from members of families with HNPCC by linkage studies and positional cloning. It is not possible to correlate MI phenotype and mismatch repair gene defect in this small collection of samples as there is not enough inoformation to evaluate this fully. However, germline mutations in four known

MMR genes have been causally linked to the vast majority of colorectal cancers with HNPCC and approximately 15% of sporadic exhibit MI phenotype. Liu et al. (1995) found that one in ten patients with MI+ sporadic colorectal tumour had a detectable germline mutation in a mismatch repair gene, suggesting that hereditary cases may not be very common in this group. Moreover, only three of seven sporadic cell lines with microsatellite instability had mutations in *MSH2*, *MLH1*, *PMS1*, or *PMS2*, implying that a significant fraction of sporadic RER+ cancers arise from mutations in other genes. Germline mutation in MMR genes seems to occur relatively frequently in young sporadic RER+ patients with CRC under 35 years of age as Liu et al. (1995) was able to find germline mutations in 5 patients (42%). It seems that young patients with RER+ phenotype have a high probability of carrying a germline mutation in one of the major MMR genes. It seems justified to propose that patients with RER+ sporadic CRC diagnosed under 50 years of age should be screened for mutations of at least two major mismatch repair genes, *hMSH2* and *hMLH1*.

4. 2. Mutation analysis

In this project mutational screening of *hMSH2*, and also two exons of the *hMLH1* gene was performed. The sites and also the effects, of mutations are listed in Table 3.3 (chapter 3) and will be discussed below:

A novel splice site mutation at position +5 in the intron 15 of the *hMSH2* gene was found in an HNPCC family. The mother and three of her offspring had colon cancer, while a fourth had cancer of the bladder. Microsatellite instability was therefore sought and found in one of those with colon cancer (Patient 21, age 37 at diagnosis), using marker loci at D2S123 (Aaltonene et al. 1993) and 635/636 (Thibodeau et al. 1993). Application of the protein truncation test showed one normal and one short peptide. Reverse transcriptase/PCR analysis of the region in the *hMSH2* mRNA, containing exons 14,15 and 16, revealed a short transcript.

Subsequent cDNA sequencing indicated the loss of exon 15. On sequencing the genomic DNA, the splice site donor sequence of intron 15 was found to have a G to C transversion at position +5. This mutation was heterozygous in the tumours of the affected individuals in this family. However, since the mother of this family had died, it was not possible to show the vertical transmission of the mutation.

This point mutation creates a restriction site for the enzyme, *Dde* I, and was used to identify the mutation in three other affected members of the family. It was not found in the healthy member or in 61 normal controls. The enzymatic assay was also used to screen a further 116 colorectal patients but no alteration was found. Liu et al.(1994) and Luce et al. (1995) also reported the skipping of exon 15, caused by a mutation at position +1 in intron 15. The deletion of exon 15 produces a translational frame shift and consequently a premature stop codon 6 base pairs downstream of the 5' end of the exon 16. Due to this mRNA deletion, codons 820-878 were deleted from the transcript resulting in a truncated peptide 58 amino acids shorter than normal. However, application of PTT for two affected members of this family showed truncated peptides. Another member of this family, with no history of colon cancer, had the same mutation but developed bladder cancer. Similar results have been reported by Liu et al. (1996). They reported two kindreds with a deletion of codon 618 in the *hMLH1* gene. One of them had a brain tumour characteristic of Turcot's syndrome, whereas the other had no extracolonic neoplasm. The explanation of our observation and that of Liu et al. for this particular case could be due to environmental conditions or other genetic factors.

The presence of one germ-line mutation in four affected members of one family could be evidence for inheritance of this mutation. It was not, however, possible to show the vertical transmission of this mutation in the family due to lack of pathological specimens from the mother of this family who also died of colon cancer when she was quite old. It is believed that RER phenotype is the

manifestation of mutation in MMR genes, and this finding could be another document to address this fact.

Shapiro and Senapathy (1987) described a method for potential consensus values at splice junction sequences. Based on this method, each of 8 nucleotides at the 5' splice junction site from position -2 through to position +6 have a particular score between 0 and 100. The final score for the 5' splice site was calculated (please see section 3.3.7) for this family and was found to be reduced from 78.36 in the normal allele to 64.18 in the mutant allele. Therefore the mutation at the 5' splice site of intron 15 reducesthe consensus value by 14.18%, and may contribute to exon 15 skipping from the transcript.

Two gross mRNA deletions were found in two different patients:
(i) mRNA deletion of exon 2-6 inclusive, (ii) mRNA deletion of exon 2-8 inclusive.

(i) This deletion was found in patient 18, whose microsatellite instability results showed replication errors in 8 of 9 markers. Interestingly, patient 18 had a tumour in three different sites of her colon, one of them in the ascending colon, the second one in the transverse colon and the third one in the descending colon. The results of microsatellite instability were different for different tumours, when examined by different markers as presented in Table 3. 1. As mentioned earlier, it is believed that the replication error in simple tandem repeat is due to DNA polymerase mistake or slippage during replication and this is the manifestation of defects in repair enzymes. Although the results from the three tumours seems to be different from each other, there are no significant differences between tumours in the proximal (ascending and transverse) and distal tumour (descending). Ta (ascending tumour) showed RER+ with 5 of 9 markers (55%), Tb (transverse tumour) showed RER+ with 6 of 8 markers (75%) and Tc (descending tumour) showed RER+ with 5 of 8 markers (62%).

The published results about the replication error and the site of the tumour in colorectal cancer and HNPCC are different. In general, approximately 70% of HNPCC colon cancers and 30% of sporadic tumours are located in the splenic flexure (Lynch and Smyrk, 1996). The importance of understanding this novel mechanism in tumorigenesis is that the RER+ is a clue to defects in repair genes. However, as shown in chapter three, the RT-PCR of patient 18 showed two different bands, one of them corresponds to the normal product, and the second one was shorter than normal. On sequencing of the normal band no alteration of the published sequence was found, but sequencing of shorter products showed a deletion from the 3' end of exon 1 to 5' end of exon 7 (exons 2-6 deleted). This is an out-of-frame deletion of codons 71-359, resulting in truncated proteins due to a new termination codon 3 bp downstream of splice site in the exon 7.

To define the nature of these defects at the genomic level, the regions surrounding affected exons, which seem to be responsible for alternative splicing, were amplified from the normal genomic DNA of corresponding individual. No alterations were found either in splice junction sites or branch point sequences in any of related exons. The 3' end of intron 1 has the longest poly-pyrimidine tract in this gene, and could potentially be a good site for alteration. The role of the poly-pyrimidine sequences in the lariat formation [pre-mRNA splicing is a two step process. In the first step, the 5' splice site is cleaved and, in a co-ordinated reaction, the 5' G residue of the intron is linked in a 2', 5'- phosphodiester linkage to an A residue near the 3' splice junction. In the second step, the 3' splice site is cleaved, the exons are joined and the intron is released as a "lariat"] and exon ligation has not been clearly established.

Mutations in the AG dinucleotide in human β-globin intron 1 decrease the efficiency of lariat formation, but do not abolish this reaction (Reed and Maniatis 1985; Ruskin and Green 1985). In contrast, deletion of the pyrimidine stretch in

human β-globin intron 1, or an adenovirus intron abolishes lariat formation (Frendewey and Keller., 1985; Reed and Maniatis., 1985; Ruskin and Green., 1985) and spliceosome assembly (Frendewey and Keller 1985; Binddereif and Green., 1986). This data suggested that pyrimidine tract is essential for lariat formation and that the AG dinucleotides increase the efficiency of this reaction. However, in spite of difficulties on sequencing of this region, due to slippage of DNA-polymerase, no alteration was found either in pyrimidine sequences or at AG dinucleotides sequence from normal published sequence. Southern analysis using different restriction enzymes (*Taq*I, *Eco*RI, *Pst*I) and different cDNA probes with various lengths were also unable to detect any genomic deletion. mRNA deletion from exons 2-6 inclusive has also been reported by another group from Australia (Kohonen-Corish et al. 1996), but no genomic abnormality or splice site mutations have been reported by them either.The absence of multiple exon skipping from transcripts is therefore likely to be due to mutations in regulatory regions (Liu et al. 1995) or a mutation within the intronic sequence (Vogelstein, personal communication) . The possible defects in any components which were discussed above could not be excluded. In a normal RNA splicing a balance between many factors are required. Similar results (multiple mRNA exon skipping) have been reported in this gene by other groups. For instance, deletions of exons 8-15 inclusive (frameshift mutation)(Liu et al. 1994) and many single exon skipping such as: deletion of exon 12 (Liu et al. 1996), exon 13 (Liu et al. 1995), exon 15 (Liu et al. 1996) for the *hMLH1* gene and also exon 7 and 13 (Liu et al. 1994, 1996) for the *hMSH2* gene could place in this category. However, exon skipping from transcripts in the *hMSH2* and *hMLH1* genes are not an unusual phenomenon in these two particular mismatch repair genes, and also in many other genes. Most of these deletions are frame shift mutations and could produce truncated proteins.

(ii) The second gross mRNA deletion was found in a patient with RER+ phenotype. Microsatellite instability was found in 2 of 9 markers (22%). The patient

was operated at the age of 64 because of colon cancer. Amplification of segment 1 of this patient showed two bands, one of them was as expected and the second one was shorter. On sequencing of the shorter products an mRNA deletion of exon 2 to exon 8 was found whereas no alteration was found at normal products from published sequence. This is an out-of-frame deletion of codons 71-359 creating a new termination codon, 3 bp downstream of splice site in the exon 9. No other member of this family was available for further study. This multiple exon skipping has been reported by another group from Australia who were also unable to find the cause of this mRNA deletion at genomic DNA (Kohnen-Corish et al. 1996). However, a significant number of mutations reported in MMR genes are due to defects in exons splicing (see Table 4.3).

Four different mRNA deletions were found in the course of mutation analysis in this project in several independent cases. One of these mRNA deletions was found in four unrelated cases (case C in Figure 4. 3). The molecular pathology of these mRNA deletions are disscused below:

In one patient a partial deletion was found within exon 1 encompassing codons 9-50. Due to this deletion 21 nt from 5' end of exon 1 and 66 nt from 3' of exon 1 were retained in the transcripts while the most sequences of exon 1 were deleted from transcripts. This is an out-of-frame deletion and created a new termination codon 22 bp upstream of 3' end of exon 1 (case A in Figure 4.7). On sequencing of genomic DNA at the 5'splice junction site of the intron 1 a C to G was found at position +8. Further investigation proved that this mutation was a normal variation, so it could not be assumed that this substitution was the cause of the deletion. No abnormalities were found by DNA blotting, using a cDNA probe or by sequencing of the coding sequence of the exon 1.

In another patient a partial mRNA deletion was found in which only 31 nt from the 5' end of exon 4 and 73 bp from the 5' end of exon 6 were retained in the transcripts. Most part of exon 4 and the entire coding sequence of exon 5 and 61 nt of

the 3' end of the exon 6 were deleted from the transcripts. This is an in frame deletion of the codons 228-340, producing a polypeptide with 115 amino acids less (case B in Figure 4.7). No abnormalities were found by sequencing the genomic DNA either within the exons or at the splice junction sites.

In four different and unrelated patients, a partial mRNA splicing was found from exon 9-12. RT-PCR products of corresponding patients showed two different bands, on sequencing of short transcripts it was found that 85 nt of the 5' end of exon 9 are retained in the transcripts while the exons 10,11 and 52 nt of the 3' end of the exon 12 were removed. It is notable that the 5' end of exon 10 and exon 11 and also exon 12 are located in the conserved region of the *hMSH2* gene. Due to this deletion codons 493-603 were deleted from transcripts, resulting in a new termination codon at the position +102 from the 3' end of exon 12. No alterations were found within exon 9-12, or in any of the splice junction sites at genomic DNA. DNA southern blotting with different restriction enzymes with cDNA probes and also genomic probes were unable to find any genomic re-arrangement (case C in Figure 4.7).

Another similar result was found in one patient in whom the 11 nt of the 5' end of the exon 14 and 20 nt from the 3' end of exon 15 were retained in the transcripts, while the most parts of exons 14 and 15 were deleted from transcripts. This is out-of-frame deletion of codons 742-872, creating a new termination codon 26 bp downstream from the 5' end of exon 16 (case D in Figure 4.7). However, no alterations were found on sequencing of the exons 14 and 15 and also in all splice junction sites.

Three of four partial mRNA deletions described here produced premature termination codons resulting in truncated proteins. In all of the cases mentioned above no alteration was found at either the 3' or 5' ends of the splice junction sites, and also at branch point sequences in intronic regions. However, mutation or deletion within the intron could not ruled out.

It is important to emphasise some of the common features of these mRNA deletions. Firstly, all of them have short repeats at breakpoints including 5 or 6 nucleotides (Figure 4. 7, bold cases). Secondly, the repeats terminated at **GA** (case B and D) or **GC**(case A and C). Thirdly, the short repeats in cases B and D have the same sequences (AAAGA), and finally case C has occurred in four unrelated patients.

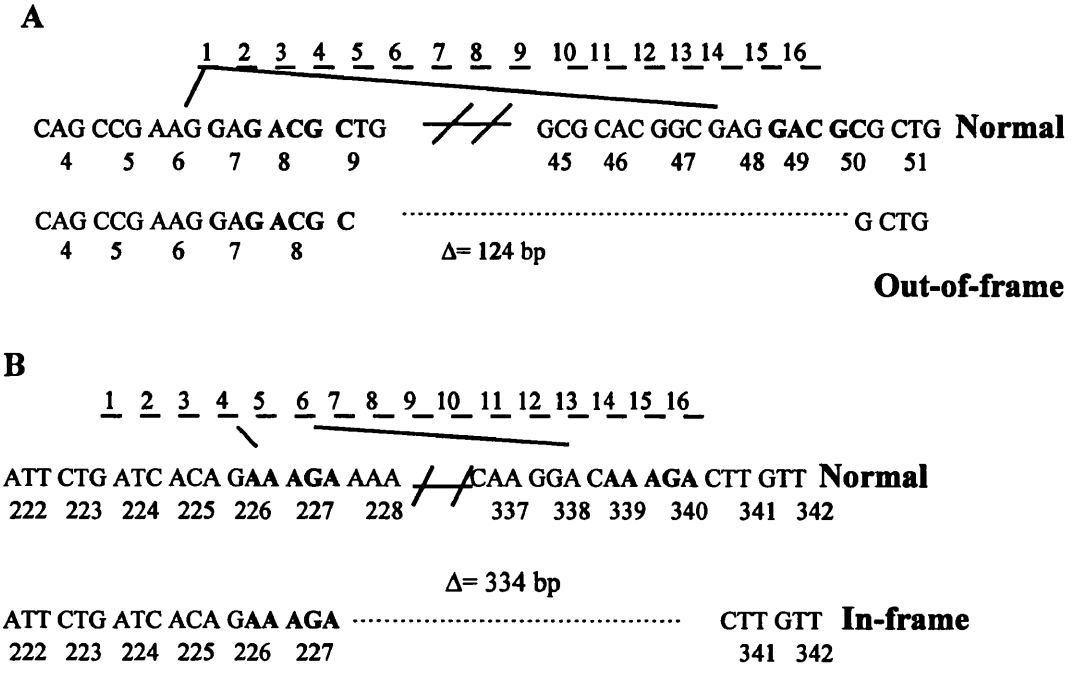
Possible explanations for this observation are discussed below:

Novel length alleles at dinucleotide marker loci and at tetranucleotide marker loci are known to be formed without exchange of flanking markers. This means that they are not generated by unequal crossing-over. Instead, as new mutant alleles have been observed to differ by a single repeat unit from the originating parental allele (Mahtani & Willard., 1993), the most likely mechanism to explain length variation is a form of exchange of sequence which commences by slipped strand mispairing. In several cases, the patients for deletions are marked by very short direct repeats. For example, the breakpoints in numerous pathological deletions of the mitochondrial genome occur at perfect or almost perfect short direct repeats. Of these, the most common is a deletion of 4977 bp which has been found in multiple patients with Kearn-Sayre syndrome. This deletion results in elimination of the intervening sequence between two perfect 13- bp repeats, and loss of the sequence of one of the repeats. Recombination does not appear to occur in the mitochondrial genome and Schoffner et al. (1989) have postulated that such deletions arise by a replication slippage mechanism, similar to that occurring at short tandem repeats. The ends of the duplicated sequences, like those of the common deletions, are often marked by short direct repeats, and the mechanisms of duplication and deletion appear to be closely related (Poulton & Holt., 1994).

Splice site recognition is especially problematic in mammals, in which genes are usually interrupted by multiple introns, often much longer than the exons, and splice site sequences are not highly conserved. The mammalian 5' splice site, branch

site and 3' splice site consensus sequences are /GURAGU,YNYURAC, and Y>10NYAG/, respectively (where / denotes a splice site, N, any nucleotide; R, purine; Y, pyrimidine; and underlining, the most highly conserved positions. As mentioned earlier in cases B and D(Figure 4.7) the repeats terminated at **GA** and in cases A and C at **GC** (Figure 4.7). On the other hand, the introns 2, 3, 8,12 and also 13 of the *hMSH2* gene were begun either with **GA**(introns 2 and 8) or **GC** (introns 3, 12 and 13), suggesting that these sequences (GA or GC) potentially could act as a cryptic splice site in splicing proccess.

Similar observation has been reported by Ali et al. (1992) they have reported the same mRNA deletions in the alpha subunit of G protein. The alpha subunit of G proteins play a vital role in signal transduction. They found three partial RNA deletions (not edge to edge). In the first deletion, exon 4 has joined to exon 10 resulting in in-frame-deletion of codons 89 to 259. In the second deletion, exon 8 joined to exon 10, resulting in in-frame-deletion of codons 215 to 248. Interestingly, a penta nucleotide (G **GTG** G) is present in both sites of breakage points, and they assumed that splicing could occurs between the amino acids 214 and 249 using the



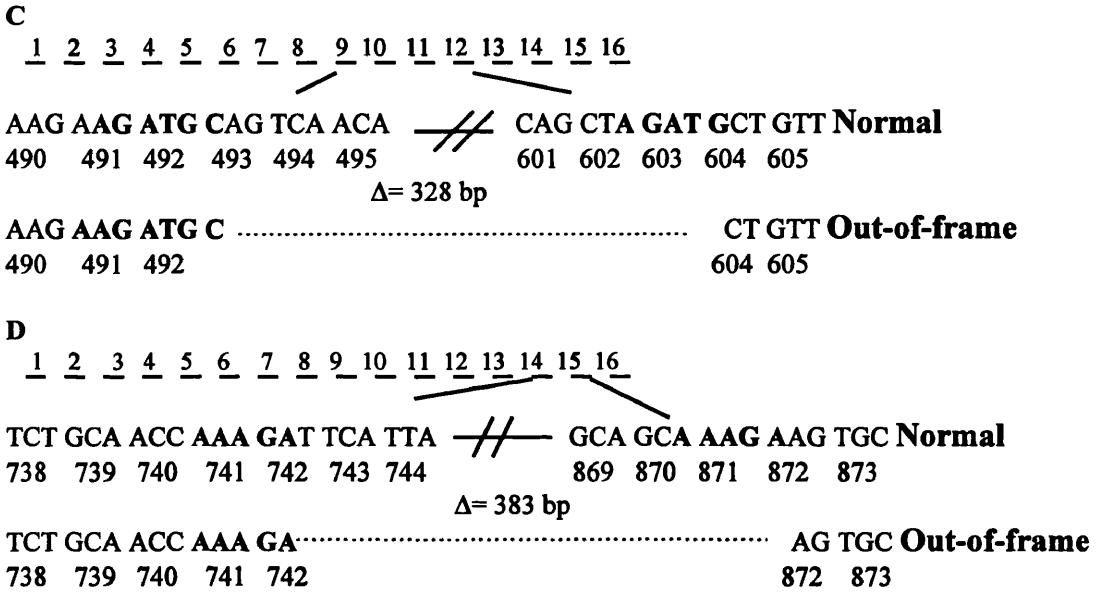


Figure 4. 3. Schematic presentation of abnormal splice junctions. Exons 1-16 of the normal transcript are indicated. The common repeat sequences before and after breakage point are presented in bold. Horizontal solid lines represent continuation of normal sequences. Dotted lines indicate the deletion of sequences between the indicated codons.

donor site GT (high lighted), and an unusual acceptor site TG (underlined). In the third deletion, exon 9 joined to exon 11, resulting in out of frame deletion of codons 241 to 281(second nucleotide of codon 241 joined to the first nucleotide of codon 281). In none of the deletions described by this group was genomic alteration involved.

The presence of partial mRNA deletion in another gene could be a further documentation of this type of rare mRNA missplicing. Alex Tipping et al. (Guy's

Hospital, London; personal communication 1997) had similar results in 6 of 8 cases. He described that the starting and ending of deletions was in the middle of exons and in none of the cases the deleted regions began with GT or GC and ended with AG. They have observed eight different repeats at the breakpoints but these transcripts only had one copy of the repeat while the full length of the cDNA carries only two copies. However, they were not able to find the cause of mRNA deletions in the genomic DNA. At least one other observation has been reported by a group from Australia (personal communication) and they were not able to find the cause of mRNA deletion at the genomic DNA. However, it is not yet clear as to what extent this aberrant mRNA is important in producing stable proteins in the cytoplasm and whether or not it is the cause of the disease.

Four point mutaions were found in different parts of the *hMSH2* gene. Three of them were found in the *hMSH2* gene, and the fourth one in the *hMLH1* gene. The most interesting of them will be discussed below:

Fishel et al. published the results and the sequence of the *hMSH2* cDNA gene in 1993. They also published details ofan intronic alteration at position -6 in intron 11 of the gene. They were only able to find this alteration in affected individuals and concluded that the candidate gene, *hMSH2,* is the proper gene for non-polyposis form of colorectal carcinoma. This alteration, a T to C transition, is located in the short polypyrimidine tract in the intron, in front of the exon encoding the most conserved region of the protein. The importance of the polypyrimidine tract at the 3' end of the intron in RNA splicing was discussed earlier. However, in this project, it was found that this mutation was present not only in a relatively large number of affected individuals but also in 15% of normal individuals (86 healthy chromosomes were studied). This simply showed that it could not be the cause of disease, but Fishel et al. (1993) concluded that the presence of this mutation in the tumour tissue, but not in normal tissue of the same individuals, provides a strong indication that this base change represents a mutation rather that a polymorphism. If the mutation analysis is

the basis for diagnostic laboratories, maximum care should be undertaken to avoid misdiagnosis. It is also possible that a polymorphism is reported as a mutation such as the above example and vice versa such as in the following example:

Huie et al. (1994) described one splice site mutation at position -13, a **T** to **G** transversion in intron 1 in a patient with adult onset of the autosomal recessive disorder glycogen storage disease type II (GSDII). The IVS1-13 **T** to **G** transversion in the polyprimidine tract in the acceptor splice site was found on one allele in over two thirds of adult onset GSDII patients studied (28/41), but was not present in 58 normal chromosomes. RT-PCR, followed by sequencing, confirmed the splicing of exon 2 in individuals with IVS1-13 **T** to **G** transversion. Raben et al. in 1996 described three mRNA splicing patterns for the same mutation which was described by Huie et al. in 1994, in adult lysosomal storage disease (glycogenosis type II). This intronic mutation creates three splice variants namely SV1, SV2 and SV3, in which exon 2 was partially or completely removed from transcripts. In SV1 a cryptic splice donor site, 36 nt downstream from exon 1 was used and exon 2 was spliced out; in SV2 exon 1 splice to exon 3, and in SV3, 60 nt of the 3' end of exon 2 are retained in the transcript while most of the exon 2 including the initiation codon are removed. Maximum care should be undertaken for proper interpretation of any alteration. It is possible for a polymorphism to be consider as a mutation and vice versa.

To date, one polymorphism has been reported in exon 6 of the *hMSH2* gene at codon 322 (Liu et al. 1995). According to this alteration G**G**C will alter to G**A**C which changes glycine to aspartic acid. However, since glycine is a non-polar amino acid and asparatic acid is an acidic one, it is likely that the polarity of the protein will be changed. This alteration was found in two unrelated and affected individuals. Further study of the normal healthy controls revealed that 4% of them bear this alteration (102 normal healthy chromosomes were studied). The presence of this transition in the healthy individuals put this alteration as a polymorphism rather than a harmful mutation. Liu et al. reported the frequency of this alteration to be 2%, but

they did not examine more than 30 individuals. They found this polymorphism to be heterozygous in the normal tissue of one patient, and homozygous in the tumour of the same patient. This result demonstrated that one germ-line mutation of the wild type allele was coupled with a second somatic mutation on the second allele in the tumour (second hit). However, due to the low frequency of this polymorphism (4%) it could be an infrequent variation.

It has been reported that the majority of mutations in the *hMLH1* gene are located in the region of exon 15 and 16. These two exons and their flanking regions were searched for any alterations. In intron 14, a G to A was found at position -19 relative to 5' end of exon 15. The presence of this alteration in a majority of normal individuals indicated that this alteration has no effect on the RNA splicing. Interestingly the frequency of this polymorphism is about 46% in the normal population (102 normal chromosome studied). Due to the very high frequency of this polymorphism it could be useful as a marker for linkage studies. Although one abnormal pattern of SCCP at exon 16 was present in two unrelated individuals, on sequencing of the corresponding exon no variation was found.

There are two common mutations in the Finnish population on the *hMLH1* gene. One of these two mutations is a 3.5 kb genomic deletion involving exon 16. This genomic deletion was searched by PCR using three primers as shown in Figure 4. 4 (Lahti et al. 1995). Primers (1) and (3) amplify a fragment of 634 bp specific for the mutant allele. In the normal allele the distance is about 4 kb and no product could be obtained by normal PCR. Primers (2) and (3) amplify a 475 bp fragment specific for the normal allele (Figure 4.1). However, this mutation was not present in all patients indicating differences between populations.

More than 90% of the gene consist of non-coding sequence. Although the structure of the peptides will be determined by the sequence of codons within the coding sequence it is clear that the non-coding sequence has its own importance. It

has been shown that intronic regions can have a suppresser effect on transcription. It is also clear that some sequences adjacent to the splice site exhibit a role in proper splicing and other sequences, particularly branch sites have an important role in intron removal. Some alterations to non-coding sequences may have no effect on splicing.

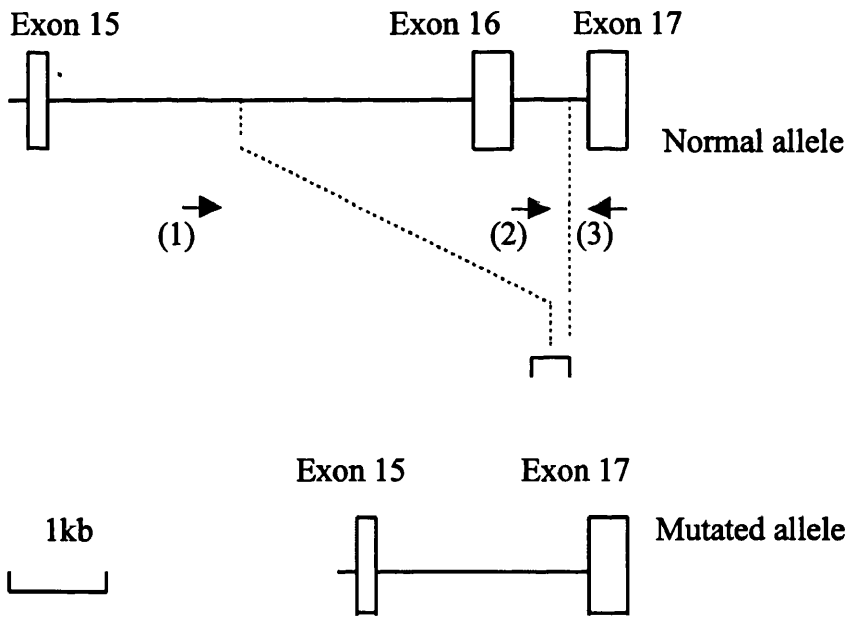


Figure 4. 1. Schematic presentation of primers 1, 2 and 3 and a 3.5 kb genomic deletion in the *hMLH1* gene. Open boxes denote exons 15, 16 and 17. Arrows show the site of primers.

4. 2. 2. Aberrant mRNA and its causes

The coding sequence regions of eukaryotic genes are interrupted by intervening sequences (introns). In RNA splicing two different main groups are

involved: *cis* acting factors (the factors that are within the gene in which the splicing is in progress e.g. 3' and 5' consensus sequences) and *trans* acting factors (the factors that are not within the gene in which the splicing is in progress e.g. ribonucleoproteins). The enzymatic removal of introns from pre-mRNA is preceded by the formation of a large ribonucleoprotein complex known as the **spliceosome**. The spliceosome is composed of five small nuclear RNAs(snRNA): U1,U2, U5, U4/U6 and numerous proteins that assemble together with pre-mRNA to facilitate the two reactions of splicing. Thus the selection of splice site, which ultimately determines the structure of spliced messenger RNAs and the proteins they encode, is accomplished by the assembly of the spliceosome at specific sites on a pre-mRNA splicing substrate. In splicing, the branch point, pyrimidine tract, 3' splice site, and 5' splice site are important determinants of splicing efficiency (Sterner et al.1993). It has also been shown that the length of the downstream intron plays a role in splice site selection (Dominiski et al. 1992). Dominiski and Kole have shown that short internal exons are ignored by the splicing machinery *in vitro* and *in vivo*, resulting in exon skipping. This phenomenon can be reversed by substitution of individual purines in the poly (Y) tract of the upstream intron with pyrimidines, suggesting an interplay between the tract and the downstream exon (Dominiski & Kole 1991). In addition to the elements mentioned above, both the sequence and the length of the internal exons affect splice site selection (Dominiski & Kole 1992).

The mechanism of splicing site selection in alternative splicing is poorly understood, and many details of this process remain to be elucidated. In several cases of alternatively spliced pre-mRNA, such as human immunodeficiency virus Rev protein, selection of splice sites appears to be regulated by specific *trans*-acting splicing factors (Dominski et al. 1992). However, specific factors may not always be involved, since alternative use of splice sites was shown to be affected by a change in the concentration of constitutive splicing factors (Ge et al. 1990; Harper et al. 1991; Krainer et al. 1990). The mechanism of splice site selection is complex and many

factors are involved in this process. Some patterns of alternative splicing will be discussed below:

The coding sequence regions of eukaryotic genes are split by intervening sequences or introns which must be removed from the hnRNA by RNA splicing. An unspliced RNA would be degraded within the nucleus or, if transported to the cytoplasm, would be unable to produce a functional protein. This is due to the presence of many interruptions of the protein coding region (termination codons) within the introns. Various cases of alternative RNA splicing have been described (Mckeown, 1992). In RNA splicing, a single gene is transcribed in many different tissues and the transcripts formed are processed differentially to yield various functional mRNAs in the different tissues. Transcription is controlled by different factors and the presence of more than one promoter and also 5' and 3' ends of RNA have very critical roles in splicing (Young et al. 1981). In many genes, the process of cleavage and polyadenylation occur at different positions within the primary transcripts and the resulting transcripts are then differentially spliced. The existence of tissue-specific splicing factors which regulate alternative splicing is also indicated by the existence of cases where transcripts with identical 5' and 3' ends are spliced differently in different tissues, and which therefore can not be explained by different usage of promoters or polyadenylations sites.

The 3' end of RNA has an important role in splicing. Splicing of calcitonine provides a good example of this point (Rosenfeld et al. 1984; Leff et al. 1988). The presence or absence of particular tissues-specific factors in the nervous system and thyroid gland determines the pattern of RNA splicing for calcitonine. Calcitonine is a peptide of 32 amino acids, and has four domains for calcium. Although the 3' end of this gene is different (the presence of two different polyadenylation sites), it is suggested that the existence of tissues-specific factors determines the pattern of RNA splicing. The interaction of tissues-specific factors at the site of binding could either

inhibit or promote the use of alternative splicing in different tissues, resulting in different transcripts in different tissues with particular role (s).

Nonsense codons could sometimes cause defects in pre-mRNA splicing. These defects include: (i) skipping of the nonsense-containing exon and, in some examples, one or more flanking exons. (ii) retention of the intron (iii) use of a cryptic splice site in the vicinity of the nonsense codon (Maquat 1995). On the other hand, nonsense codons could reduce mRNA half-life. For instance, human β globin mRNA that cause β^0 -thalassemia, a recessive haemolytic anaemia, appear to reduce the half- life of fully splice mRNA after the mRNA has been exported to the cytoplasm (Maquat et al. 1981; Lim et al. 1989; 1992; Lim and Maquat 1992). For example, the half life of β-globin mRNA in patients harbouring a frameshift induced nonsense codon within exon 2 was calculated to be 30 min where the half life of non-thalassemic β-globin mRNA was estimated to be $>$ 16 h (Maquat et al. 1981).

In splicing, different factors are involved and include branch site sequences, consensus intron/exon boundary sequences, exon length (Dominski and Kole, 1991), five different small nuclear ribonucleoproteins (Guthrie 1991) and many other components (Dominski and Kole, 1992; Robbenson et al. 1990; Fabrizio and Abelson 1990; Mcpheeters and Abelson, 1992). Each splice site is recognized independently by specific factors whose interactions provides exon justaposition for ligation in the correct order (Robbeson et al. 1990). Abnormal transcripts and scrambled exons, have been reported in the deleted colon cancer gene (DCC) (Nigro et al. 1991) in both normal and neoplastic cell types. Exons are joined but in an order different from that present in the genomic DNA. The mechanisms for generating such scrambled transcripts are not known (Nigro et al. 1991). Mutation analysis on the central domain of U6RNA has shown that mutations at nt 34-62 can partially or completely block the splicing process (Mcpheeters and Abelson, 1992). Mutation analysis on the central domain of U6RNA (one of the *trans* acting factors in RNA

splicing) has shown that mutations at nt 34-62 can partially or completely block the splicing process (Mcpheeters and Abelson, 1992). Successful exon recognition occurs when the balance of all of the components is sufficient to support the splicing process (Sterner and Berget, 1993). However, splicing is a complex process and such complexity is vastly increased by controlling elements that define tissues-specific alternative splicing patterns. It has been estimated that 100 or more gene products (about 2% of the yeast genome) are involved in splicing functions (Sharp, 1994), and therefore a mutation in any one of these could affect the whole process and perhaps be responsible for a mRNA deletion in which no abnormality could be found in genomic DNA.

A number of different mutations as described before have been reported in mismatch repair genes. There are some mRNA deletions including single and multiple exon skipping in which the authors were not able to determine the cause of mutations at the genomic DNA (Table 4. 2). All of these mutations could be potentially harmful, as they are out-of-frame deletions resulting in truncated protein. However, the number of published mutations in these two genes is not considerable when compared with published mutations in *BRCA1* and *BRCA2.* Aberrant RNA splicing is not restricted to mismatch repair genes, since it has been reported in many different diseases such as deficiency of medium-chain acyl-CoA dehydrogenase (Kelly et al. 1990), adenosine deaminase deficiency (ADA) (Akeson et al. 1987), argininosuccinate lyase gene (Abramson et al. 1990), Citrullinemia (Kobayashi et al. 1990) and also in alfa subunit of G protein (Unnisa Ali et al. 1992).

Kelly et al. have reported nine different aberrant RNAs in the medium-chain acyl-CoA dehydrogenase (MCAD) deficiency (1990), including single and multiple exons missing from the transcript. MCAD is a common inherited defect in energy metabolism and may be fatal. The incidence of this disease in Great Britain is 1 in 5000-10, 000 births. They found several single base changes and exons deletions involving many in multiple locations throughout the mRNA coding region. The gene

has 12 coding exons and the following exon deletions were found: (i) deletion of exon 2 (ii) deletion of exon 5 (iii) deletion of exon 2 and 5 (iv) deletion of exon 5 and retention of intron 3 (v) deletion of exon 2 through to exon 10 (vi) deletion of exon 2 and deletion of exon 7 through middle of exon 11 (vii) deletion of exon 7 through to the middle of exon 11(viii) deletion of exons 8-10 (ix) deletion of exon 2 and 5 as well as exon 8-10. Therefore, missplicing during RNA processing is not a result of a point mutation either in the donor or acceptor splicing sites (Kelly et al.1990).

Table 4. 2. Summary of published aberrant mRNA deletions in mismatch repair genes.

No.	Exon	Gene	Mutation	No. of kindreds	Predictated protein	Reference
1	2-6	*hMSH2*	Del. of exons 2-6	1	Out-of-frame deletion of exon 2-6	Kohonen -Corish et al., 199
2	3	*hMSH2*	Del. of exon 3	1	Out-of-frame deletion of exon 3	Froggatt et al., 1996
3	7	*hMSH2*	Del. of exon 7	1	Out-of-frame deletion of exon 7	Liu et al., 1996
4	8-15	*hMSH2*	Del. of exons 8-15	1	Out-of-frame deletion of exon 8-15	Liu et al., 1994
5	13	*hMSH2*	Del. of exon 13	2	Out-of-frame deletion of exon 13	Liu et al., 1994 Luce et al., 1995
6	3	*hMLH1*	Del. of exon 3	1	Out-of-frame deletion of exon 3	Viel et al., 1997
7	12	*hMLH1*	Del. of exon 12	2	Out-of-frame deletion of exon 12	Liu et al., 1994 Papadopoalos et al., 1994
8	13	*hMLH1*	Del. of exon 13	2	Out-of-frame deletion of exon 13	Liu et al., 1995 Luce et al., 1995
9	15	*hMLH1*	Del. of exon 15	1	Out-of-frame deletion of exon 15	Liu et al., 1996

It is notable that the MCAD mRNA missplicing occurs in normal tissues but at a lower frequency than that of the index patients. The cause of the high rate of missplicing in the patient is not yet clear (Kelly et al.1990). In another paper (Akeson et al. 1987) an exon 4 and also exon 4 and exon 7 skipping appeared from transcript in the adenosine deaminase deficiency but the cause of the missing exon in the

genomic DNA was not discovered. Genomic blotting with different restriction enzymes, and also DNA sequencing were done but no alteration was identified. Abramson et al. (1990) reported deletion of exon 7 in 5% of the mRNA of the *argininosuccinate lyase* gene from all tissue sources studied. As the length of intron 6 is 79 base pairs and it has been stated that the minimum size of an intron required for efficient splicing should be around 80 bases, the size could be the cause of exon 7 skipping from transcript. However, no alteration in splice sites or within the exon 6 and exon 8 have been found to be considered as the cause of exon 7 skipping from transcript (Abramson et al. 1990).

Nakai and Sakamoto (1994) constructed a database containing aberrant splicing mutations of 90 genes with 209 mutations. In this database, there are 91 cases of exon skipping including 48 mutations (53%) due to destruction of 5' splice site, 29 mutations (32%) due to destruction of 3' splice site and also 14 mutations (15%) with unknown aetiology. Nakai is dedicated to revising his database due to the increasing number of aberrant mRNA deletions without unknown eatilogy (personal communication). However, the number of aberrant spliced mRNAs in mismatch repair genes, in comparison to published mutations in these two genes, appears to be significant and therefore requires more investigation.

In summary, in the RNA splicing process two different groups are involved; *cis* and *trans* acting factors. Any abnormalities in any members of these two groups could be the cause of aberrant splicing. Among these factors there are some invisible factors e.g. GpC contents which are able to alter the secondary structure of the mRNA. However, mutations in introns more than 50 base pairs away from splice junction sites could also be the cause of exon skipping but can not be detected easily given a large average size of introns.

In conclusion, mutation detection in HNPCC is still problematic and difficult. To date, at least four different genes are known to be involved and there is no common mutation or hot spot in any of these genes, and mutations are distributed

throughout the entire coding sequence. The mutation screening of *hMSH2* and to a lesser extent *hMLH1,* was performed using four complementary methods. This study shows that the *hMSH2* gene is involved in a proportion of patients with colorectal cancer in the West of Scotland. There was no sign of hotspots or common mutations in the gene. HNPCC is a heterogeneous disease, and only by employing RNA and DNA based methods all existing mutations be detected.

Genotype-phenotype correlation

To establish a correlation between genotype and phenotype is one of the most important aim of any mutation screening study. Although it is still difficult to establish a proper link between genotype and phenotype in HNPCC.

At the beginning of this chapter it has been explained that there is a correlation between MI+ and germline mutations in HNPCC tumours and also in some proportion of sporadic cases of CRC. Indeed, MI is a reasonable tool for screening sporadic and familial cases of CRC in particular when patients are younger than 50 years of age. A tumour is composed of normal and malignant cells. It is possible MI could not be detect in some of the cases due to a low percentage of malignant cells in the piece of tumour from which DNA has been extracted. In the pathology department of Royal hospital, University of Glasgow, (Dr. Purdie, personal communication) a tumour has been defined as a mass of tissue in which there are more than 70% malignant cells and a normal tissue is a mass of cells in which there are more than 70% normal cells. However, it seems that in a mass of pure malignant cells either from HNPCC or sporadic tumours the range of MI+ phenotype might be more than the reported range in the publications (90% and 15% in HNPCC and sporadic cases, respectively). Therefore, due to low number of malignant cells in early stage of tumourigenesis some MI+ tumours, during MI assessment were missed.

In the clinic HNPCC can be classified as Lynch 1, Lynch 2 and Muir-Torre syndromes. Muir-Torre syndrome is thought to be a variant of Lynch syndrome and Kolodner et al. (1994) have been reported two germline mutations including a nonsense mutation and a frameshift mutation in exon 12 of *hMSH2* that are linked to inheritance of cancer susceptibility in these kindreds.

A germline mutation in the splice site of exon 5, resulting in exon 5 skipping from the transcript, has been reported by several groups from different parts of the glob and is claimed to be a common mutation in North-American and British populations (Froggatt et al. 1995).

Mutations in *hMSH2* and *hMLH1* can cause Lynch 1 and Lynch 2. It is believed that the h*MSH2* and h*MLH1* have roughly the same share in HNPCC (Nystrom-Lahti et al. 1996). h*MSH2* was reported to account for 29% (5/17) and h*MLH1* for 17% (3/17) of English HNPCC kindreds (Fraggett et al. 1996), while the proportion of kindreds attributable to mutations in h*MSH2* was 21% (7/34) in a Dutch series (Wijnen et al. 1995), and 34% (10/29) from North America (Liu et al. 1994). *hMLH1* was reported to account for 24% (8/34) of HNPCC kindreds from Japan (Han et al. 1995). It seems that the *hMLH1* gene is more mutable than *hMSH2*.

Mutation detection in HNPCC is still problematic (Kohonen-Corish et al. 1996) and difficult. To date, at least four different genes are involved. There is no common mutation or hot spot in any of these genes, and mutations are distributed through the entire coding sequence. It seems that the predisposing mutation varies from kindred to kindred, but a small number of recurrent mutations have been reported (referred to recurrence of mutation in this chapter).

Due to the heterogeneity of disease and lack of common mutations or hot spots in the MMR genes it seems that further study is required to make it possible to establish a link between genotype and phenotype.

Future work

The published mutations in MMR genes from global studies indicate that there is no germline mutation in about 30% of HNPCC patients with RER+ and also no germline mutation was found in a majority of sporadic CRC cases. Additionally, it has been shown that the frequency of germline mutation in CRC patients without Amsterdam criteria is quite low and recently, as expected, a germline mutation in *MSH6* has been reported in a family without Amsterdam criteria (Miyaki et al. 1997). This data suggests that there are additional CRC families with germline mutations in unscreened MMR genes and more mutation analysis of these families is required to clear up the molecular pathology of this heterogeneous disease. Any mutation analysis should cover at least two major MMR genes namely *MSH2* and *MLH1*. The whole coding sequence of both genes should be search for any defect within the coding sequence of the gene, although most mutations are unique and quite evenly distributed, with some clustering in *MSH2* exon 12 and *MLH1* exon 16.

REFERENCES

Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J-P., Jarvinen, H., Powell, S. M. Jen, J., Hamilton, S. R., Petersen, G. M., Kinzler. K. W., Vogelstein, B., Chapelle, A. de la.(1993) Clues to the pathogenesis of familial colorectal cancer. Science, 260: 812 - 816.

Aaltonen, L. A., Sankila, R., Mecklin, J. P., Jarvinen, H., Pukkala, E. (1994) A novel approach to estimate the proportion of hereditary nonpolyposis colorectal cancer burden. Cancer Detect. Prevent. 18: 57- 63.

Abramson, R. D., Barbosa, P., Kalmuck, K., O'Brien, W. E. (1991) Characterixation of the human Argininosuccinate Lyase gene and analysis of exon skipping. Genomics 10, 126-132.

Akeson, A. L., Wiginton D. A., Christopher States, J., Perme, C. M., Dusing, M. R., Hutton, J. J., (1987) Mutations in the human adenosine deaminase gene that affect protein structure and RNA splicing. Pro. Natl. Acad. Sci. USA 84, 5947-5951.

Ali, I. U., Reinhold, W., Salvador, C., Aguanno, S. (1992) Aberrant splicing of GS alfa transcript in transformed human astroglial and glioblastoma cell lines. Nuc. acid. Res. 20, 4263-4267.

Au, K. G., Cabrera, M., Miller, J. H., Modrich, P.(1988) *Escherichia coli* Mut Y gene product is required for specific A-G to C.G mismatch correction. Proc. Natl. Acad. Sci. USA 85:9163-9166.

Au, K. G., Clark, S., Miller, J. H., Modrich, P. (1989) The *Escherichia coli* mutY gene encodes an adenine glycosylase active on G-A mispairs. Proc. Natl. Acad. Sci.USA 86: 8877- 8881.

Aquillina, G., Hess, P., Branch, P., MacGeoch, C., Casciano, I., Karran, P., Bignami, M. (1994) A mismatch recognition defect in colon carcinoma confers DNA microsatellite instability and a mutator phenotype. Proc Natl Acad Sci, USA, 91, 8905-8909.

Baker, S. J., Fearon, E. R., Nigro, J. M., Hamilton, S. R., Preisinger, A. C., Jessup, J. M., Van Tuinen, P., Ledbetter, D. H., Baker, D. F., Nakamura, Y., White, R., Vogesltein, B.(1989) Chromosome 17 deletions and *p53* gene mutations in colorectal carcinomas. Science, 244, 217.

Bedi, G. C., Westera, W. H., Farzadegan, H., Pitha, P. M., Sidransky, D. (1995) Microsatellite instability in primary neoplastasms from HIV+ patients. Nature Medicine, 1, 65-68.

Benson, J. R., and Wells, K. (1995) Microsatellite instability and a TGFβ receptor: clues to a growth control pathway. BioEssays, 17, 1009-1012.

Berget, S, M. (1995) Exon recognition in vertebrate splicing. J Biol Chem 270:2411-2414.

Bindereif, A. and Green, M. R. (1986) Ribonucleoprotein complex formation during pre-mRNA splicing in vitro. Mol. Cell. Biol. 6:2582-2592.

Bubb, V. J., Curtis, L. J., Cunnigham, C., Dunlop, M. G., Carothers, A. D., Morris, R. S., White, S., Bird, C. C., Wyllie, A. H., (1996) Microstellite instability and the role of hMSH2 in sporadic colorectral cancer. Oncogene, 12, 2641-2649.

Bodmer, W., Bishop, T., and Karran, P. (1994) Genetics steps in colorectal cancer. Nat. Genet. 6: 217 - 219.

Boland, C, R. (1996) Setting microsatellites free. Nature Medicine 2:972-974.

Borresen, A. L., Lothe, R. A., Meling, G. I., Lystad, S., Morrison, P., Lipford, J., Kane, M. F., Rognum, T. O., Kolodner, R. (1995) Somatic mutations in the *hMSH2* gene in microsatellite unstable colorectal carcinoma. Human Molecular Genetics. 4, 2065-2072.

Botstein, D., White, R., Skolnick, M., and Davis, R. (1980) Construction of a genetic linkage MAP in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314-331.

Braaten, B., Madara, J. L., and Donowitz, M. (1988) Age-related loss of nongoblet crypt cells parallels decreased secretion in rabbit descending colon. American Journal of Physiology-Gastrointestinal and Liver Physiology, 1988, Vol.255, No.1, pp.18/1).

Branch, P., Aquilina, G., Bignami, M., Karren. P. (1993) Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage. Nature 362, 652-654.

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

Bufill, J. (1990) Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumour location. Ann. Intern. Med. 113: 779 - 788.

Burks, R. T., Kessis, T. D., Cho, K. R., Hedrick, L. (1994) Microsatellite in endimatrium carcinoma. Oncogene, 9: 1163 -1166.

Campuzano, V., Montermini, L., Molto, M. D., et al. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271, 1423, 1427.

Chase, J. W., and Richardson, C. C.(1974) Exonuclease VII of *Escherichia coli*: mechanism of action. J.Biol.Chem. 249: 4553 - 4561.

Chomczynski, P., and Sacchi, N. (1987). Single step method of RNA isolation by acid guonodium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162, 156- 159.

Chong, J. M., Fukayama, M., Hayashi, Y., Takizawa, T., Koike, M., Konoshi, M., KikuchiYanoshita, R. Miyaki, M. (1994) Microsatellite instability in the progression of gasteric carcinoma. Cancer Research, 54: 4595 - 4597.

Christiano, A. M., Hoffman, G. G., Chung-Honet, L. C., Lee, S. Cheng, W. Uitto, J., Greenspan, D. S. (1994) Structural organization og the human type VII collagene gene (COL7A1), composed of more exons than any previously characterizes gene. Genomics 21:169-179.

Cleaver, J. E. (1968) Defective repair replication. of DNA in Xeroderma pigmentosum. Nature 218, 652-656.

Cogan, J. D., Prince, M. A., Lekhakula, S., Bundey, S., Futrakul, A., McCarthy, E. M. S., Phillip, J. A. (1997) A novel mechanism of aberrant pre-mRNA splicing in humans. Hum, Mol, Genet, 6, 909-912.

Cohen, J. B., and Levinson, A. D. (1988) A point mutation in the last intron responsible for increased expression and transforming activity of the c-Ha-ras oncogene Nature 334, 119-124.

Cohen, J. H., and Levinson, A. D. (1988) Nature, 334. 119.

Cotton, R. G. H., and Campbell, R. D. (1989) Chemical reactivity of matched cytosin and thymine bases near mismatched and unmatched bases in a heteroduplex between DNA strands with multiple differences. Nucleic Acid Research17;4223-4233.

Cox, E. C. (1973) Mutator gene studies in *Escherichis coli*: the mutT gene. Genetics 73: 67- 80.

Cox, E. C. (1976) Bacterial mutator genes and the control of spontaneous mutation. Ann. Rev. Genet. 10,135-156.

Cross, N. C. P., Silly, H., Goldman, J. M., and Wada, C. (1994) Microsatellite instabillity in myeloid lekemias. Blood, 84:3236 - 3237.

Cruts, Marc., Hendriks, L., Van Broeckhoven, C. (1996) The presenilin genes: a new gene family involved in Alzheimer disease pathology. Hum. Mol. Genet. 5, 1449-1455.

da Costa, L. T., Liu, B., el Deiry., et al. (1995) Polymerase delta varients in RER colorectal tumours. Nature Genetics 9, 10-11.

Dalbagni, G., Presti, and J., Reuter, V. (1993) Genetics alterations in bladder cancer. Lancet, vol 342, pp: 469- 471.

Damagnez, V., Doutriaux, M. P., Radman, M.(1989) Saturation of mismatch repair in the mut S mutator strain of *Escherichia coli*. J. Bacteriol. 171: 4494 - 4497.

DasGupta, C., and Radding, C. M. (1982) Polar branch migration promoted by recA protein: Effect of mismatch base pairs. Proc. Natl. Acad. Sci. USA 79:762-766.

Dominski, I. Z., and Kole, R. (1991) Selection of splice sites in pre-mRNA, with short internal exons. Mol. Cell. Biol. 6075-6083.

Dominski, Z., and Kole, R. (1992) Cooperation of pre-mRNA sequence elements in splice site selection. Mol. Cell. Biol. 2108-2114.

Dzidic, S., and Radman, M. (1989) Genetic requirements for hyper - recombination by very short patch mismatch repair DNA polymerase I. Mol. Gen. Genet. 217:254-256.

Fabrizio, P., and Abelson, J. (1992) Mutational analysis of the yeast u2 snRNA suggests a structural similarity to the catalytic core of group I introns. Cell 71, 819-831.

Fearon, E. R., and Vogelstein, B. (1990) A genetic model for colorectal tumorogenesis. Cell, 61:759.

Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W. (1990) Identification of a chromosome for gene that is altered in colorectal cancers. Science, 247,49-56.

Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J., Ruppert, J. M., Hamilton, S. R., Preisinger, A. C., Thomas, G., Kinzler, K. W., Vogelstein , B. (1990) Identification of a chromosomal 18q gene that is altered in colorectal cancers. Science, 247:49-56.

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

Fearon, E. R., and Jones, P. A. (1992) Peogressing toward a molecular description of colorectal cancer development. FASEB J, 6:2786.

Feinberg, A.P.,Gehrke, C. W., Kuo, K. C., Ehrlich, M. (1988) Reduced genomic 5-methylcutosine content in human colonic neoplasia. Cancer Res.48, 1159.

Fisher, S. G., and Lerman, L. S.

(1983) DNA fragments differing by

single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. pro. Natl. Acad. Sci. USA 80, 1579-1583.

Fishel, R. A., Siegel, E. C., and Kolodner, R. (1986) Gene conversion in *Escherichia coli.* Resolution of heteroallelic mismatched nucleotides by co-repair. J. Mol. Biol. 188: 147- 157.

Fishel, R. A., and Kolodner, R. (1989) Gene conversion in *Escherichia coli*: the rec F pathway for resolution of heteroduplex DNA. j. Bacteriaol. 171:3046-3052.

Fishel, R., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. A. (1994a) The human mutator gene homolog MSH2 ans its association with hereditary nonpolyposis colon cancer. Cell 75: 1037 - 1048.

Fishel, R., Ewel, A., Suman, L., Lescoe, M. K., Griffith, J. (1994b) Binding of mismatch microsatellite DNA sequences by the human MSH2 protein. Science, 266, 1403-1405.

Fishel, R., Ewel A., and Lescoe, K. M., (1994c) Purified human msh2 protein binds to DNA containing mismatch nucleotides. Cancer Research, 54, 5539-5542.

Fowler, R. G., Degnen, G. E., and Cox, E. C. (1974) Mutationa specificity of a conditiona *Escherichia coli* mutator, mutD. Mol. Gen. Genet. 133, 179-191.

Frendewey, D., and Keller (1985) The stepwise assembly of a pre-mRNA splicing complex requires U sn RNPs and specific intron sequences. Cell, 42:355-367.

Fresht, A. R., Knill-Jones, J.W. (1983) Contribuation of 3' to 5' exonuclease activity of DNA polymerase III holoenzyme from *Escherichia coli* to specificty. J. Mol.Biol. 165:669-682.

Friedberg, E. C. (1985) DNA repair (New York:W. H. Freeman).

Friedberg, E. C. (1990) The enzymology of DNA repair. Mutat.Res. 236, 145 - 314.

Froggatt, N., Joyce, J, A., Davies, R., Evans, D, Ponder, B., Barton, D., Maher, E.,(1995) A frequent *hMSH2* mutaion in hereditary non-polyposis colon cancer syndrome. The Lancet, 345, 727.

Froggatt, N, J., Cecilia, B., Koch, D, H., Evans, D, G, R., Hodgson, S, V., Ponder, B, A, J., Maher, E, R (1996) Mutation screening of *MSH2* and *MLH1* mRNA in hereditary non-polyposis colon cancer syndrome. J Med Genet. 33: 726-730.

Fu, Y. H., Pizzuti, A., Fenwick, R., G., et al. (1992) An unstable triplet

repeat in a gene related to myotonic dystrophy. Science 260, 235-238.

Fu, Y. H., Friedman, D. L., Richards, S., et al. (1993) Decreased expression of myotonin protein kinase messenger RNA and protein in adult form of myotonic dystrophy. Science 260, 235-238.

Fujii. R., and Shimada, T. (1989) Isolation and characterzation of cDNA clones derived from the divergenetly transcribed gene in the region upstream from the human dihydrofolate reductase gene. J. Biol. Chem. 264, 10057-10064.

Fujiwara, Y., Emi, M., Ohata, H., Kato, Y., Nakajima, T., Mori, T., Nakamura, Y. (1993) Evidence for the presence of two tumor suppressor genes on chromosome 8p for colorectal carcinoma. Cancer Res. 53,1172.

Gagliardini, V., Fernandez, P-A., Lee, R. K. K., Drexler, H. C. A., Rotello, R. J., Fishman, M. C., Yuan, J. (1994) Prevention of vertebrate neuronal death by the crm A gene. Science, 263, 826.

Gao, C., Wu, N., Grignon, D. Zzchzrck, A., Liu, H., Salkowski, A., Li, G., Sakr, W., Sarkar, F., Potter, A. T., Chen, Y. Q., Honn, K. V. (1994) Oncogene, 9, 2999-3003.

Ge, H., and Manley, J. L. (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell 62, 25-34.

German, J., and Passarge, E. (1989) Bloom's syndrome. XII. Report from the registry for 1987. Clinical Genetics, Vol.35, No.1, 57-69.

Gilbert, W. (1978) News and views. Nature 271, 501.

Glazer, P. M., Sarker, S. N., Chisholm, G. E., Summers, W. C., (1987) DNA mismatch repair detected in human cell extracts. Mol. Cell. Biol. 7, 218-224.

Gonzalez, M. Z., Ruppert, M., Tokion, K., Tsai, Y. C., Spruck III, C. H., Miyao, N., Nichols, P. W., Hermann, G. G., Horn, T., Steven, K., Summerhayes, I. C., Sidransky, D., Jones, P. A. (1993) Microsatellite instability in bladder cancer. cancer Research, 53, 5620-5623.

Green, P. M., and Shapiro, L. J. (1987) Cell 49, 443-454.

Green, M, R. (1991) Biochemical mechanism of constitutive and regulated pre-mRNA splicing. Ann Rev Cell Biol 7:559-599.

Groden, J. Thliveris, A., Samowitz, W., Carlson, M., et al. (1991) Idemtification and characterization of the familial adenomatous polyposis coli gene. Cell 66, 589-660.

Grompe, M. (1993) The rapid detction of unknown mutations in nucleic acid. Nature Genet. 5, 111-117.

Gross, D. S., and Garrad W. T. (1986) The ubiquitious potential Z - forming sequenc of eucaryotes, (dT-dG)n . (dC-dA)n, is not detectable ib the genomes of eubacteria. Mol. Cell. Biol. 6: 3010 - 3013.

Guthrie, C. (1991) Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein. Science 253, 157-163.

Gyllensten, U. B., and Erlich, H. A. (1988) Generation of single- stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc. Natl. Acad. Sci. USA, 85, 7652-765.

Hamada, H. A., and Kakunaga, T. (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298: 396-398.

Hamada, H., M. G. Petrino, and Kakunaga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA 79: 6465 - 6469.

Hamada, H., M. G. Petrino, T. Kakunaga, M., Seidman, and Stollar, B. D. (1984a) Characteriziation of genomic poly (dT- dG). poly (dC-dA) sequences; structure, organization, and conformation. Mol. Cell.Biol. 4: 2610 - 2621.

Han, H-J., Yanagisawa, A., Kato, Y., Park, J-G., Nakamura, Y. (1993) Genetic instability in pancreatic cancer and poorly differentiated type of gasteric cancer. Cancer Resaerch, 53, 5087-5089.

Han, H-J., Maruyama, M., Baba, S., et al. (1995) Genomic structure of human mismatch repair gene, *hMLH1*, and its mutations analysis in patients with hereditary non-polyposis colorectal cancer (HNPCC). Hum. Mol. Genet. 4, 237.

Hare, J. T., and Taylor, J. H. (1985) One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repire. Pro Natl Acad Sci USA 82, 7350.

Harly, H. G., Brook, J. D., Rundle, S. A. (1992) Expasion of unstable DNA region and phenotypic variation in myotonic dystrophy. Nature 355, 545-546.

Herman, G. E., and Modrich, P. (1981) *Escherichia coli* K- 12 clones that overproduce dam methylase are hypermutable. J. Bcteriol. 145: 644-646.

Hinds, P. W., Mittnacht, S., Dulic, V. Arnold, A., Reed, S. I., Weinberg, R. A. (1992) Regulation of retinoblastoma protein functions by ectopic expression of human cyclins. Cell, 70, 993.

Holmes, J. Jr. Clar, S., and Modrich, P. (1990) Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines. Pro. Natl. Acad. Sci. USA 87, 5837-5841.

Honchel, R., Halling, K. C., Schaid, D. J., Pittelkow, M., Thibodeau, S. N. (1994) Microsatellite instability in Muir - Torre syndrome. Cancer Resaerch 54: 1159 1163.

Houlston, R. S., Collins, A., Slack, J., Morton, N, E. (1992) Dominant genes for colorectal cancer are not rare.Ann. Hum. Gent. 56: 99-103.

Huie, M. L., Chon, A. S., Tsujino, S., Shanke, S., Dimauro, S.,Engel, A. G., Hirschhorn, R. (1994) Aberrent splicing in adulat onset glycogene storage disease type II (GSDII) :molecualt identification of IVS1 (-13 T to G)mutation in a majority of patients and a novel IVS10 (+1GT to CT) mutation. Hum. Mol. Genet. 3, 2231-2235.

Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reaveal a new mechanisms for colonic carcinogenesis. Nature, Vol 363. pp: 558- 561.

Jarvinen, H. J. (1992) Epidemiology of familial adenomatous polyposis in Finland: Impact of family screening on the colorectal cancer rate and survival. Gut, Vol.33, No.3, pp.357-360.

Jarvinen, H. J., Mecklin, J-P., Sistonen, P. (1995) Screening reduces colorectal cancer rate in families with hereditary nonpolyposis colorectal cancer. Gastroenterology, 108,

Jones, M., Wagner, R., and Radman, M. (1987) Mismatch repair of deaminated 5-methylcytosine. J. Mol. Biol. 194: 155- 159.

Joslyn, G., Carlson, M., Thliveris, A., Albertsen, H., Gelbert, L., Samowitz, W., Groden, J., Stevens, J., Spirio, L., Robertson, M., Sargeant, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J. P., Warrington, J., McPherson, J., Wasmuth, J., Le, Paslier, D., et. al (1991) Identification of deletion mutations and three new genes at the familial polyposis locus. Cell, Vol.66, No.3, pp.601-613.

Kantor, J. A., Turner, P. H., Nienhuis, A. W. (1980) Beta thalassemia: mutaions which affect processing of the β -globin mRNA precursor. Cell 21:149-157.

Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gary, J. W., Waldman, F., Pinkel, D. (1992) Comparative genomic hybridisation for molecular cytogenetic analysis of solid tumours. Science, 258, 818-821.

Kat, A., Thilly, W. G., Fang, W. H., Longley, M. J., et al. (1993) An alkylation-tolerant, mutator human cell line is deficient in strand-specific mismatch repair. Pro. natl. Acad. Sci. USA, 90, 6424-6428.

Kawaguchi, Y., Okamoto, T., Taniwaki, M., et al. (1994) CAG expansion in a

novel gene for Machaso-Joseph disease at chromosome 14q32.1 Nature Genet. 8, 221-227.

Kelly, D. P., Whelan, A. J., Ogden, M. L., Alpers, R., Zhang, Z., Bellus, G., Gregersen, N., Dorland, L., Strauss, A. W. (1990) Molecualr characterzation of inherited medium-chain acyl-coA dehydrogenase deficiency. Pro. Natl. Acad. Sci. USA 87, 9236-9240.

Kern, S. E., Fearon, E. R., Tersmette, K. W. F., Enterline, J. P., Leppert, M., Nakamura, Y., White, R., Vogelstein, B., Hamilton, S. R. (1989) Clinical and pathologic associations with allelic loss in colorectal carcinoma. JAMA, 26, 3099.

Kiaris, H., Ergazaki, M., Segas, J., Spandidos, D. A. (1995) Detection of Epstein-Barr virus genome in squamous cell carcinomas of the larynx. International Journal of Biological Markers, 1995, Vol.10, No.4, pp.211-215.

Kinzler, K. W., Nilbert, M. C., Su, L-K., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K. J., Preisinger, A. C., Hedge, P., Mckechnie, D., Finniear, R., Markham, A., Groffen, J., Bogusi, M. S., Altshul, S. F., Horrii, A., Ando, H., Miyoshi,Y., Miki, Y., Nishisho, I., Nakamura, Y. (1991) Identification of FAP locus genes from chromosomes 5q21. Science, 253, 661.

Kinzler,K, W., Nilbert, M, C., Su, L-K., Vogelstein, B., Bryan , T, M., Levy, D, B., Smith, K, J., Preisinger, A, C., Hamilton, S. R., Hedge, P., Markham, A., Carlson, M., Joslyn, G., Groden, J., White, R., Miki, Y., Miyoshi, Y., Nishisho, I., Nakamura, Y., (1991) Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. Science, 251:1366-1370.

Knudson, A,G. (1971) Mutation and cancer: Statistical study of retinoblastoma.Pro. Nantl. Acad. Sci. USA, 68, 820-823.

Kobayashi, K., Jackson, M. J., Tick, D. B., O'Brien, W. E., Beaudet, A. L.., (1990) Heterogeneity of mutations in Argininosuccinate synthetase causing Human Citrullinemia. The journal of Biological Chemistry 265, 11361-11367.

Kohonen-Corish, Maija., Ross, V, L., Doe, W, F., Kool, D, A., Edkins, E., Faragher, I., Wijnen, J., Khan, P. M., Macrae,F., James, D., St John, B. (1996) RNA base mutation screening in hereditary nonpolyposis colorectal cancer. Am. J. Hum. Genet. 59, 818-824.

Kolodner, R. D., Hall, N, R., Lipford, J., Kane, M., Rao, M. R. S., Morrison, P., Wirth, L., Finan, P., Burn, J., Chapman, P., Earabino, C., Merchant, E., Bishops, T. (1994) Structure of the MSH2 locus and analysis of two Muir-Torre kindreds for msh2 mutations. Genomics, 24, 516-526.

Krainer, A. R., Conway, G. C., Kozak, D. (1990) Purification and characterzation of pre-mRNA splicing factor SF2 from Hela cells. Genes Dev. 4, 1158-1171.

Kunkel,L. M., Smith,K.D., Boyer,S.H., and Borgaonkar, D. S. (1977) Analysis of human Y chromosome specific reiterated DNA in chromosome variants. Proc Natl Acad Sci USA 68, 120- 124.

Lahue, R. S., Au, K. G., and Modrich, P. (1989) DNA mismatch correction in a defined system. Science 245: 160 - 164.

Lamond, A. I. (1991) Nuclear RNA processing, current opinion in cell biology. 3, 493-500.

Lavin, M., Shiloh, Y., Concannon, P., Good, RA., Gatti, RA. (1996) Ataxia-telangiectasia: Mutations in ATM cDNA detected by protein-truncation screening. American Journal of Human Genetics, , Vol.59, No.1, pp.40-44.

La Spada, A. R., Wilson, E. M., Lubahn, D. B., et al. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular. Nature 352, 77-79.

Lavin, M., Shiloh, Y., Concannon, P., Good, RA., Gatti, RA. (1996) Ataxia-telangiectasia: Mutations in ATM cDNA detected by protein-truncation screening. American Journal of Human Genetics, , Vol.59, No.1, pp.40-44.

Lazar, V., Grandjouan, S., Bognel, C., Couturier, D., Rougier, P., Bellet, D., Paillerets, B, B. (1994) Accumulation of multiple mutations in tumour suppressor genes during colorectal tumorigenesis in HNPCC patients. Human Molecular Genetics, 3, 2257-2260.

Leach, F. S., Nicolaides, N. C., Papadopoulos, N. Liu, B., Jen, J. (1993) Mutations of a MutS homologous in hereditary nonpolyposis colorectal cancer. Cell 75: 1215 - 1225.

Leff, S, E., Rosenfeld, M. G., Evans, R. M. (1986) Complex transcriptional unit: diversity in gene expression by alternative RNA processing. Annual Review of Biochemistry. 55, 1091-117.

Lehman, I. R., and Nussbaum, A. L. (1964) The deoxyribouncleases of *Escherichia coli* V on the specifity of exonuclease I (phosphodiesterase). J. Bio. Chem. 239, 2628-2636.

Lehrman, L. E., Goldstein, J. L., Russell, D. W., et al. (1985) Duplication of seven exons in the hypercholestrolaemia. Cell, 48, 827-835.

Lieb, M. (1981) A fine structure map of spontaneous repressor gene, including insertions of IS elements. Mol. Gen. Genet. 181:364-371.

Linton, J. P., Yen, J. Y., Selby, E., et al. (1989) Dual bidirectional promoters at the mouse dhfr locu cloning and characterization of two mRNA classes of the divergently transcribed Rep-1 gene. Mol. Biol. 9, 3058-3072.

Liu, B., Parsons, R. E., Hamilton, S. R., Petersen, G. M., Lynch, H. T., Watson, P., Markowitz, S., Willson, J. K. V., Green, J., Chapella, A. de la, Kinzler, W., Vogelstein, B. (1994) *hMSH2* mutations in hereditary nonpolyposis colorectal cancer kindreds. Cancer Research, 54, 4590-4594.

Liu, B., Farrington. S. M., Petersen, G. M., Hamilton, S. R., Parson, R., Papadopoulos, N., Fujiwara, T., Jen, J., Kinzler, K. W., Wylie A. H., Vogelestein, B., Dunlop, M. G. (1995a) Genetic instability occurs in the majority of young pateints with colorectal cancer.Nature Medicine, 1, 348 - 352.

Liu, B., Nicholaides, N. C., Markowitz, S., Willson, J. K. V., Parson, R. E., Jen, J., Papadopolous, N., Peltomaki, P., Chapelle, A, de la., Hamilton, S. R., Kinzler, K. W., Vogelstein, B. (1995b) Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instabilily. Nature Genetics, 9, 48-55.

Liu, B., Parsons, R., Papadopoulos, N., Nicolaides, N. C., Lynch, H. T., Watson, P., Jass, J. R., Dunlop, M., Wyllie, A., Jessup, J. M., Peltomaki, P., de la Chapelle, A., Hamilton, S. R., Vogestein, B., Kinzler, K. W., (1996) Analysis of mismatch repair gene in hereditary nonpolyposis colorectal cancer patients. Nat. Med. 2, 169-174.

Livingstone, L., White, A., Sprouse, J., Livanos, E., Jacks, T., Tstly, T. D. (1992) Altered cell cycle arrest and gene amplification potential accompany loss of wild- type p53. Cell, 70, 923.

Lim, S. K., Mullins, J. J., Chen, C. M., Gross, K. W., Maquat, L. E. (1989) Novel metabolism of several β- thalassemic β-globin RNAs in the erythroid tissues of thransgenic mice. EMBO J, 8, 2613-6219.

Lim, S. K., Sigmund, C.,Gross, K. W., Maquat, L, E. (1992) Nonsense codons in human β- globin mRNA results in the production of mRNA degradation products. Mol. Cell Biol, 12, 1149-1161.

Lovett, S. T., and Kolodner, R. D. (1989) Identification and purification of a single - strand DNA specific exonuclease encoded by the recJ gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 86:4639-4643.

Loth, A. R., Peltomaki, P., Meling, I. G., Aaltonen, A. L., Nystrom-Lahti, M., Pylkkanen, L., Heimdal, k., Andersen, I. T., Moller, P., Rognum, O. T., Fossa, D. S., Haldorsen, T., Langmark, F., Brogger, A., de la Chapelle, A., Borresen, A-L. (1993) Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. Cancer Resaerch, 53, 5849-5852.

Lu, A- L., and Chang, D. Y. (1988) A novel nucleotide excision repair for the conversion of an A/G mismatch to C/G base pair in E.coli. Cell 54:805- 812.

Luce, M. C., Marra, G., Chauhan, D., Laghi, L., Carethers, J. M., Cherian, S, P., Hawn, M., Binnie, C. G., Kam-Morgan, L. N. W., Cayouette, M. C., Koi, M., Boland, C. R. (1995) In vitro transcription/tranalation assay for the screening of *hMLH1* and *hMSH2* mutations in familial colon cancer. Gastroentrology 109, 1368-1374.

Lynch, H. T., Smyrk, T. C., Watson, P., Lanspa, S. J., Lynch, A. F., Lynch, P, M., Cavalieri, J., Boland, R. (1993) Genentics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an update review. Gastroentology, 104, 1535-1549.

Maliaka, Y. K., Chudina, A. P., Belew, N. F., et al. (1996) CpG dinucleotides in the *hMSH2* and *hMLH1* genes are hotspots for HNPCC mutations. Hum. Genet. 97, 251.

Mao, L., Lee, D. J., Tockman, M. S., Erozan, Y. S., Askin, F., Sidransky, D. (1994) Microsatellote alterations as clonal markers for the detection of human cancer. Pro. Natl. Acad. Sci. USA. 91: 9871-9875.

Mare, J. T., and Taylor, J. H. (1985) One role for DNA methylation in vertebrate cells in strand discrimination in mismatch repair. Pro. Natl. Acad. Sci. USA 82, 7350-7354.

Marinus, M. G., and Morris, N. R. (1974) Biological function for 6-methyladenine residues in the DNA of *Escherichia coli* K12. j. Mol. Biol. 85: 309-322.

Marinus, M. G., Poteete, A., Arraj, J. A. (1984) Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K- 12. Gene 28:123-125.

Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W., Vogelstein, B., Brattain, M., Willson, J. K. V. (1995) Inactivation of the type II TGFβ receptor in colon cancer cells with microsatellite instability. 268, 1336-1338.

Mary, J-L., Bishop, T., Kolodner, R., Lipford, J., Kane, M., Weber, W., Torhorst, J., Muller, H., Spycher, M., Scott, R, J. (1994) Mutational analysis of the *hMSH2* gene reveals a three base pair deletion in a family predisposed to colorectal cancer development. Human Molecular Genetics, 3, 2067-2069.

Mashal, R. D., Koontz, J., and Sklar, J., (1995) Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvase. Nature Genet. 9, 177-183.

Maxim, A, M., Gilbert, W. (1977) A new method for sequencing DNA.

Proc. Natl. Acad. Sci. USA, 74, 560-564.

Maquat, L. E., Kinniburgh, A. J., Beach, L. R., Honig, G. R., Lazerson, J., Ershler, W, B., Ross, J. (1980) Processing of human β-globin mRNA precursor to mRNA is defective in three patients with β^+ thalassemia. Pro Natl A cad Sci USA 77:4287-4291.

Maquat, L. E., Kinniburgh, A. J., Rachmilewitz, E. A., Ross, J. (1981) Unstable bet- globin mRNA - defeicent β^0 thalassemia. Cell, 27, 543-553.

Maquat, L. E. (1995) When cell stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. RNA, 1, 453-463.

Maquat, L. E. (1996) Defects in RNA splicing and the consequence of shortened translational reading frames Am. J. Hum. Genet. 59:279-286.

Mc Conike-Rosell, A., Lachiewicz, A. M., et al., (1993) Evidence that methylation of FMR-1 locus is responsible for variable phenotype expression of the fragile X syndrome. Am. J. Hum. Genet. 53, 800-809.

Mckeown, M. (1990) Regulation of alternative splicing, In genetic Engineering. ed.J. K. Setlow, 12: 139-81. London/New York:Plenum.

Mckimen, P. J. (1987) Hum. Genet. 175, 197.

Merlo, A., Mabry, M., Gabrielson, E., Vollmer, R., Baylin, S. B., Sidransky, D. (1994) Frequent microsatellite instability in primary small cell lung cancer. Cancer Research 54, 2098-2101.

Merlo, A., Mabry, M., Gabrielson, E., Vollmer, R., Baylin, S. B., and Sidransky, D. (1994) Frequent microsatellite instability in primary small cell lung cancer Cancer Res. 54:2098-2101.

Miller, S. A., Dykes, D. D., Polesky, H. F. (1988) A simple salting out procdure for extracting DNA from human nucleated cells.

Mironov, K. N., Aguelon, M. M., Potapova, I. G., Omori, Y., Gorbunov, V. O., Klimenkov, A. A., Yamasaki, H. (1994) Alteration of (CA)n DNA repeats and tumor suppressor genes in human gasteric cancer. Cancer Res, 54, 41-44.

Miyaki, M., Konishi, M., Tanaka, K., Kikichi-Yanoshita, R., Muraoka, M., Yasuno, Ma., Igari, T., Koike, M., Chiba, M., Mori, T. (1997) Germline mutation of mSH6 as the cause iof hereditary nonpolyposis colorectal cancer. Nat. Genet.Vol. 17, 271-272.

Modrich, P. (1991) Mechnisms and biological effects of mismatch repair. Annu. Rev. Genet. 25: 229 - 253.

Moore, M. J., Query, C. C., Sharp, P. A. (1993) Splicing of precursors to mRNA by the spliceosome. In: Gesteland R F, Atkins JF (eds) The RNA world. Cold Spring Harbor Laboratory Press, New York, pp 303-357.

Moselien, G., Tester, D. J., Lindor, N. M., Hobchel, R., Cunningham, J. M., French, A. J., Halling, K. C., Schwab, M., Goretzki, P., Thibodeau, S. N. (1996) Microsatelite instability and mutation analysis on *hMSH2* and *hMLH1* in patients with sporadic, familial and hereditary colorectal cancer. Hum. Mol. Genet. 5, 1245-1252.

Nagafuchi, S., Yanagisawa, H., Ohsaki, E., et al. (1994) Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA). Nature Genet. 8, 177-182.

Nakai, K., Sakamoto, H. (1994) Construction of a novel data base containing aberrant splicing mutations of a mammalian genes. Gene 141:171-177.

Nareisi, T. N., Schotz, M. C., Scott, J., Schoulders, C. C. (1993) Dinucleotide repeat polymorphisms at the lipoprotein lipase (LPL) locus.Hum. Genet.92,312-313.

Newton, C. R., Graham, A., Heptinstall, L. E., et al. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nuc. Acids. Res. 17, 2503-2516.

Nghiem, Y., Cabrera, M., Cupples, C. G., Miller, J. H. (1988) The Mut Y gene: A mutator locus in *Escherichia coli* that generates G-C to T-A transversions.Proc. Natl. Acad. Sci. USA 85:2709- 2713.

Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei Y-F., Carter, K, C., et al. (1994) Mutations of two PMS homologoues in hereditary nonpolyposis colon cancer. Nature 371: 75 - 80.

Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei,Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A. , Haseltine, W. A., Fleischmann, R. D, Fraser, C. .M., Adams, M. D., Venter,J. C., Dunlop, M. G., Hamilton, S. R., Petersen, G. M., De La-Chapelle, A., Vogelstein, B., Kinzler, K. W. (1994) Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature, Vol.371, No.6492, pp.75-80.

Nigro, J. M., Cho, K. R., Fearon, E. R., Kern, S. E., Ruppert, J. M., Oliner, J. D., Kinzler, K. W., Vogelstein, B. (1991) Scrambled exons. Cell 64, 606-613.

Nishisho, I., Nakamura,Y., Miyoshi, Y., Miki,Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., Hedge, P., Markham, A., Krush, A. J., Petersen, G., Hamilton, S. R., Nilbert, M. C., Levy, D. B., Bryan, T. M., Preisinger, A.C., Smith, K. J., et al. (1991) Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science 253, 665-669.

Nystrom-Lahti, M., Parsons, R., Sistonen, P., Pylkkanen, L., Aaltonen, L, A., Leach, F, S., Hamilton, S, R., Wtson, P., Bronson, E., Fusaro, R., Cavalieri, J., Lynch, J., Lanspa, S., Smyrk, T., Lynch, P., Drouhard, T., Kinzler, K, W., Vogelstein, B., Lynch, H, T., Chapelle, A, de la., Peltomaki, P. (1994) Miamatch repair genes on chromosomes 2p and 3p account for a major share of hereditary nonpolyposis colorectal cancer families evaluable by linkage. Am. J. Hum. Genet. 55: 659-665.

Nystrom-Lahti, M., Kristo, P., Nicolaides, N, C., Chang, S-Y., Aaltonen, L, A., Moisio, A-L., Jarvinen, H., Mecklin, J-P., Kinzler, K., Vogelstein, B., Chapelle, A, de la.,Peltomaki, P. (1995) Founding mutations and alu-mediated recombination in hereditary colon cancer. Nature Medicine, 1, 1203-1206.

Orr, H. T., Chung, M., Banfi, S., et al. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type I. Nature Genet. 4, 221-226.

Orita, M., Suzuki, Y., Hayashi, K. (1989) A rapid and sensitive detection of point mutations and genetic polymorphism using polymerase chain reaction. Genomic 5, 874-879.

Papadopoulos, N., Nicolaides, N. C., Wei, Y- F., Ruben, S. M., Carter, K. C., et al. (1994) Mutation of a *mutL* homolog is associated with hereditary colon cancer. Science 263: 1825-1829.

Papadopoulos, N., Nicolaides, N. C., Wei, Y- F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D., Venter, J. C., Hamilton, S. R., Petersen, G. M., Watson, P., Lynch, H. T., Peltomaki, P., Meckline, J-P., Chapelle, A. de la, Kinzler, K. W., Vogelstein, B. (1994) Mutation of a mutL homolog in hereditary colon cancer. Science, 263, 1625.

Parsons, R., Li, G. M., Longley, M. J., Fang, W. H., Papadopoulos, N., Jen, J., Chapelle, A., Kinzler, K., Vogelstein, B., Modrich, P. (1993) Hypermutability and mismatch repair deficiency in RER+ tumor cells.Cell 75, 1227 1236.

Patel, U., Grundfestnroniatowki, S., Gupta, M., Banerjee, S. (1994) Microsatellite at five chromosomes in primary breast tumors. Proc. Nat. Acad. Sci. USA 91:9871-9875.

Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J-P. (1993) Genetic mapping of a locus predisposing to human colorectal cancer. Science, 260: 810 - 812.

Peltomaki, P., Aaltonen, L, A., Sistonen, P., Pylkkanen, L., Mecklin, J-P., Jarvinen, H., Green, J, S., Jass, J, R., Weber, J, L., Leach, F, S., Petersen, G, M., Hamilton, S, R., Chapelle, A. de la, Vogelstein, B. (1993) Genetic mapping of a

locus predisposing to Human colorectal cancer. Science, 260, 810-812.

Peltomaki, P., Lothe, R. A., Aalatonen, A., Pylkanen, L., Nystrom-Lahti, M., Seruca, R., David, L., Holm, R., Ryberg, D., Haugen, A., Brogger, A., Borresen, A-l., de la Chapella, A. (1993) Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. Cancer Research, 53, 5853-5855.

Plough, H. H. (1941) Spontaneous mutability in Drosophila. Cold Spring Harbor Symp. Quant. Biol. 9, 127-37.

Plummer, S. J., Anton,-Culver, H., Webster, L., Nobel, B., Liao, S., Kennedy, A., Belinson, j., Casey, G., (1995) Detection of BRCA1 mutations by protein truncation test. Hum. Mol. Genet. 4, 1989-1991.

Ponz, L. D., Sassatelli, R., Benatti, P. Roncucci, L. (1993) Identification of hereditary nonpolyposis colorectal cancer in the general population: the 6 years experince of a population- based registary. Cancer 71: 3493 - 3501.

Poulton, J., Holt, I. J.(1994) Mitochondrial DNA: Does more lead to less? Nature Genetics, Vol.8, No.4, 313-315.

Powell, S. M., Petersen, G. M., Krush, A. J. et al. (1993) Molecular diagnosis of familial adenomatous polyposis. New England J of Medicine 329, 1982-1987.

Raben, N., Nichols, R. C., Martiniuk, F., Piotz, P. H., (1996) A model of mRNA splicing in adults lysosomal storage disease (glycogenosis type II). Hum. Mol.Gene.5, 995-1000.

Radicella, J. P., Clark, E. A., Fox, M. S. (1988) Some mismatch repair activity in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 85: 9674-9678.

Reed, R. and Maniatis, T. (1985) Intron sequences involved in lariat formation during pre-mRNA splicing. Cell, 41:95-105.

Reenan, R. A. G., Kolodner, R. D. (1992a) Isolation and characterisation of two Sacharomyces cerevisiae gene encoding homologous of the bacterial Hex A and Mut S mismatch repair proteins. Genetics 132: 963 - 973.

Rhyu, M. G., Park, W. S., Meltzer, S. J. (1994) Microsatellite instability occurs frequently in human gastric carcinoma. Oncogene, Vol.9, No.1, pp.29-32.

Risinger, J. I., Berchuck, A., Kohler, M. F., Watson, P., Lech, H. T., Boyed, J., (1993) Genetic instability in endometrial carcinoma. Cancer Research 21: 5100 - 5103.

Robberson, B. L., Cote, G. J., Berget, S. M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons.Mol. Cell. Biol. 10:84-94.

Roest et al. (1993) Hum. Mol. Genet.2, 1719-1721.

Rosenfeld, M. G., Amara, S. G., Evans, P. M. (1984) Alternative RNA processing: determining neuronal phenotype. Science, 225, 1315-20.

Rouyer, F., Simmler, M. C., Page, D. C., et al. (1987) A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination. Cell, 51, 417.

Ruddon, R. W. (1995) Cancer biology. Oxford University press, third edition.

Ruskin, B., and Green, M, R. (1985) Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. Nature, 317:732-734.

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

Sankila, R., Aaltonen, L. A., Jarvinen, H. J., Mecklin, J-P. (1996) Better survival rates in patients with MLH1-associated hereditary non-poyposis colorectal cancer. Gastroentrology, 110:682-687.

Schaaper, R. M. (1988) Mechnisms od mutagenesis in the *Escherchia coli* mutator mut D5: role of DNA mismatch repair. Proc. Natl. Acad. Sci. USA 85:8126-8130.

Schaaper, R. M. (1993) Base selection, proofreading and mismatch repair during DNA replication in Escheichia coli. J. Biol. Chem. 268:23762-23765.

Shapiro, M., and Senapathy, P. (1987) RNA splicing junctions of different classed of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acid Res. 19, 7155-7174.

Siegel, E. C., and Bryson, V. (1964) Selection of resistance strains of Escherichis coli by anti biotics and antibacterial agents: role of normal and mutator strains. Antimicrob. Agents Chemother. 1963, 629-34.

Siegel, E. C., and Kamel, F. (1974) Reversion of frameshift mutations by mutator genes in *Escherichia coli*. J. Bacteriol. 117, 994-1001.

Siegel, E. C., and Ivers, J. J. (1975) mut-25, a mutation to mutator link to purA in *Escherichia coli*. J. Bacteriol 121, 524-529.

Sharp, P. A. (1994) Split genes and RNA splicing. Cell 77, 805-815.

Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S. and Perucho, M. (1994) Genomic instability in repeated sequences is an early somatic event in colorectal tumourigenesis that persists after transformation. Nature Genetics 6, 273-281.

Smith, C. W.J., and Nadal-Ginard, B. (1989) Mutually exclusive splicing of

alpha tropomysine exons enforced by an unusual lairnat branch point location: implication for constitutive splicing. Cell 56:749-758.

Southern, E, M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis.J. Mol. Biol. 98,503-517.

Stephenson, C., and Karran, P. (1989) Selective binding to DNA base pair mismatches by proteins from human cells. J. Biol. Chem. 264, 21177-21182.

Sterner, D. A., and Berget, S. M. (1993) In vitro recognition of a vertebrate mini-exon as an exon- intron-exon unit.Molecular and Cellular Biology 13, 2677-2687.

Strand, M., Prolla, T. A., Liskay, R. M., Petes, T. D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274 - 276.

Strickler, G. J., Zheng, J., Shu, Q., Burgart, J. L., Alberts, R. S., Shibata, D. (1994) P53 mutations and microsatellite instability in sporadic gasteric cancer: when guardians fail. Cancer Res. 54, 4750-4755.

Shridhar, V., Siegfried, J., Hunt, J., Mar Alonso, M., del., and Smith, D, I. (1994) Genetic instability of microsatellite sequences in many non- small cell lung carcinomas Cancer Res. 54:2084.

Su, S-S., and Modrich, P. (1986) *Escherichia coli* mut S- encoded protein binds to mismatched DNA base pairs. Proc. Natl. Acad. Sci. USA 83:5057-5061.

Sun, L., K. E. Paulson, C. W. Schmid, L. Kadyk, and L. Leinwand. (1984) Non - Alu family interspersed in human DNA and their transcriptional activity. Nucleic Acids Research. 12: 2669 -2690.

Tannergard, P., Lipford, J. R., Kolodner, R., et al. (1995) Mutation screening in the *hMLH1* gene in Swedish hereditary nonpolyposis colon cancer families. Cancer Research 55, 6092.

Tautz, D., and Renz M. (1984) Simple sequences are ubiquotous repetitive components of eulkaruotic genomes. Nucleic Acids Research 12:4127 - 4138.

Thibodeau, S. N., Bren, G., and Schaid, D. (1993) Microsatellite instability in cancer of the proximal colon. Science, 260, 816-819.

Thomas, D. C., Roberts, J. D., Kunkel, T. A. (1991) Heteroduples repair in extracts of human Hela cells. J. Biol. Chem. 266, 3744-3751.

Treffers, H. P., Spinelli, V., Belser, N. O. (1954) A factor (or mutator gene) influencing mutation rates in *Escherichia coli*. Pro. Natl. Acad. Sci. USA 40, 1064-71.

Uchida, T., Wada, C., Wang, C., Egawa, S., Ohtani, H., Koshiba, K. (1994) Genomic instability of microsatellite repeats and mutations of H- K-, and N- ras, and p53 genes in renal cell carcinoma. Cancer Res, 54, 3682-3685.

Umar, A. et al. (1994) Defective mismatch repair in extracts of colorectal and endometrium cancer cell line exhibiting microsatellite instability. J. Biol. Chem. 269,14367 - 14370.

Vasen, H. F. A. (1994) What is non-polyposis colon cancer (HNPCC)? Anticancer Research, 14, 1613-1615.

Vasen, H. F. A., Wijnen, J. Th., Menko, F. H. (1996) Cancer risk in families with HNPCC by mutation analysis. Gastroentrology, 110: 1020-1027.

Vassen, H. F. A., Mecklin, J, P., Meera-Khan, P., Lynch, H, T. (1991) The international collaborative on hereditary non-polyposis colon cancer (ICG-HNPCC). Diaease of the colon and rectum, 34, 424-425.

Vogelstein, B., Fearon, E. R., Hamilton., Kern, E., Preisinger, A. C., Leppert, M., Nakamura., White, R., Smits, A. M. M., Bos, J. L. (1988). Genetic alteration during colorectal tumor development. N Engl. Med. 319:525.

Wallace, R. B., Shaffer, J., Murphy, R. F., et al. (1979) Hybridisation of synthetic oligonnucleotides tp Phix 174 DNA: the effect of single base pair mismatch. Nuc. Acids Res. 6, 3543-3557.

Weber, J. L,. May, P. E. (1989) Abundant calss of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum.Genet. 44, 388-396.

Weber, J. L,. (1990) Informativeness of human (dC-dA)n. (dG-dT)n polymorphisms. Genomics 7,524-530.

Warthin, A. S. (1913) Hereditary with reference to carcinoma. Arch. Intern. Med. 12:546- 555.

Warthin, A. S. (1925) The further study of a cancer family. J Cancer Res. Clin. Oncol. 9, 279-286.

Warthin, A. S. (1931) Heredity of carcinoma in man. Ann. Intern. Med. 4, 681-696.

Weatherall, D. J. (1992) The molecular of single gene disorders-The new genetics and clinical practice. Thirs edition Oxford University Press.

Wijnen, J., Fodde, R., Khan, P. M. (1994) DGGE polymorphism in intron 10 of MSH2, the HNPCC gene. Human Molecular Genetics, 3, 2268.

Wijnen, J., Vasen, H., Khan, P. M., Menko, F. H., Klift, H., Leeuwen, C., Broek, M., Leeuwen-Cornelisse, L., Nagengast, F., Meijers-Heijboer, A., Lindhout, D., Griffioen, G., Cats, A., Kleibeuker, J., Varesco, L., Bertario, L., Bisgaard, M., Mohr, J., Fodde, R. (1995).Seven new mutations in *hMSH2*, an HNPCC gene, identified by denaturing gradient-gel eletctrophoresis. Am. Hum. Genet. 56, 1060-1066.

Wijnen, J., Meera Khan, P., Vasen, H., et al. (1996) Majority of *hMLH1* mutations responsible for hereditary nonpolyposis colorectal cancer cluster at the exonic region 15-16. Am. J. Hum. Genet. 58, 300.

Winawer, S. J. (1993) Colorectal cancer screening comes of age. New England Journal of Medicine, 1993, Vol.328, No.19, pp.1416-1417.

White, M. B., Carvalho, M., Derse, D., et al. (1992). Detecting single base substitutions as heteroduplex polymorphisms. Genomic 12, 301-306.

Youil, R., Kemper, B. W., Cotton, R. G. H. (1995) Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. Pro. Natl. Acad. Sci. USA 92, 87-91.

Young, R, A., Hagenbuchle, O., Schibler, U. (1981) A single mouse α- amylase gene specific two different tissue specific mRNAs. Cell, 23, 452-458.

Young, J., Leggett, B., Gustafson, C., Ward, M., Searle, J., Thomas, L. (1993) Genomic instability occurs in colorectal carcinoma but not in adenomas. Human Mutation 2: 351 - 354.

Zell, R., Fritz, H. J. (1987) DNA mismatch repair in *Escherichia coli* counteracting the hydrolytic deamination of 5-methylcytosine residues. EMBO J. 6: 1809 - 1815.