BIOINFORMATICS IDENTIFICATION OF CODING MICROSATELLITES THAT ARE MUTATED IN MISMATCH REPAIR DEFICIENT COLORECTAL CANCERS

BY

JANE PARK

A thesis submitted in conformity with the requirements
for the degree of Masters of Science
Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto

BIOINFORMATICS IDENTIFICATION OF CODING MICROSATELLITES THAT ARE MUTATED IN MISMATCH REPAIR DEFICIENT COLORECTAL CANCERS

Masters of Science Degree 2001

Jane Park

Department of Laboratory Medicine and Pathobiology
University of Toronto

ABSTRACT

High frequency microsatellite instability (MSI-H) colorectal carcinomas (CRCs) display MMR deficiency and account for about 15% of human colorectal cancers. Gene inactivation via mononucleotide tract frameshift mutations is characteristic of the pathogenesis of MSI-H colorectal carcinomas. Frameshift mutations have only been described in selected candidate genes including *TGFβRII*, *IGFIIR*, *BAX* and others [1-5]. We developed "Kangaroo", a program capable of highly specialized searches in nucleotide and protein sequence databases. Utilizing Kangaroo, we identified genes with coding polyadenine tracts that had functional significance in cell cycle regulation. Our results revealed widespread novel mutations in mismatch repair deficient colorectal cancers, while Kangaroo provided a method for genome scanning to identify candidate mutation hot spots in evolution and human disease.

ACKNOWLEDGEMENTS

As with everything we do in life, we have our ups and our downs; this degree has been no exception. I never would have made it through without the fanastic people in both the Redston and Gallinger labatories.

As members of the Redston lab, we endured turbulent times over the past year but I think it united us. Laura, thank-you for being such an amazingly helpful and kind resource in the lab. You have the patience of a god, as we found out with our experiences with a particular path fellow. Nando, even with your gruff exterior we all know that you are all heart. You thought me the ropes and I benefited from your expertise. Mags, you are a bad influence but I love ya for it. Your sense of humour, your resilence and your sense of adventure are admirable. Your presence in the lab made the days bearable, thanks. To the past members, Fidel, Travis, Tony and Judit, my experience would not have been the same without all of you. Oh yeah, can't forget my summer student of 2000, Kyle, who was always entertaining and helpful.

Hyeja, I will miss our morning coffee talks, you have been a good "uhnee" to me and taught me so much about Korean culture. Colleen, your cool, relaxed perspective balanced the cosmos in the lab atmosphere. Chrisitine, my twin, our paths have crossed before and these last 2 years gave us some time to really get to know each other. We made it through together and I'm grateful that we had each other for support and comradery. I know our paths will cross again, hopefully not to far in the future.

Mark, I know you have done the best you could in an unpleasant situation. You gave me independence with my project and I guess you prepared me for self-directed learning, for which I am thankful. I am thankful for the opportunity for me to be a part of your lab and your research.

To my dear friends Leyla and Maria. I love you both, you two are the best friends that one could ever ask for. I would be lost without my network of great friends and family. Moms, Dad and James, thanks for your continuing support and encouragement, all of you have gone out of your way so I could get an education.

Joseph, my love, you have been such an integral part of my completing this degree. As painful as it was to write this thesis, those days, months were tolerable due to your persistence, your guidance, your helpful conversations, and of course your presence.

I will sorely miss the company of the people I have met throughout the duration of this degree. All the great people on the 9th floor and especially those in 989, thanks for making my time here memorable.

See ya,

Jane Park

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS

A	Adenine
ACF	Aberrant Crypt Foci
APAF-1	Apoptotic Protease Activating Factor 1
APC	Adenomatous Polyposis Coli
ASN.1	Abstract Syntax Notation
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine Triphosphate
ATR	Ataxia Telaniectasia and Rad3-related
AXIN	Axil/Conductin homologue
BAX	Bcl2-Associated X
BCL	B-cell lymphoma
BLM	Bloom Syndrome
BRCA1	Breast and Ovarian Cancer 1
BRCA2	Breast and Ovarian Cancer 2
C	Cytosine
CARD	Caspase Recruitment Domain
CBF2	CCAAT-Box Binding Factor 2
CDC	Cell Division Cycle
CDK	CDC Dependent Kinase
cDNA	Complementary Deoxyribonucleic Acid
CDK4	Cyclin-Dependent Kinase –4
C-MYC	Avian Myelocytomatosis Viral Oncogene homologue
CDS	Coding DNA Sequence

CHK1	Checkpoint Kinase-1
CIMP	CpG Island Methylator Phenotype
CRC	Colorectal Carcinoma
CSI	Chromosomal Instability Pathway
CTNNB1	Catenin, β 1
DCC	Deleted in Colon Cancer
DNA	Deoxyribonucleic Acid
DNA-PK	DNA-dependent Protein Kinase
DP-1	Dimerization Partner 1
ERCC5	Excision Repair Complementing Defect in Chinese Hamster 5
FADD	Fas-Associated Protein with Death Domain
FAK	Focal Adhesion Kinase
FAP	Familial Adenomatous Polyposis
FISH	Fluorescent in situ Hybridization
G	Guanine
GI	Gastrointestinal
GRK4	G-protein Coupled Receptor Kinase 4
GSK3β	Glycogen Synthase Kinase 3β
GTBP	G/T Base Pair
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
IDL	Insertion/Deletion Loop
IGFIIR	Insulin-like Growth Factor Receptor Type II
KRAS	Kirsten Rat Sarcoma 2 Viral Oncogene Homologue
LEF	Lymphoid-Enhanced Factor

LOH	Loss of Heterozygosity
MCR	Mutation Cluster Region
MBD4	Methyl-CpG-Binding Domain –4
MDM2	Mouse Double Minute homologue
MED1	Methyl-CpG-Binding Endonuclease –1
MLH1	MutL homologue 2
MMAC1	Mutated in Multiple Advanced Cancers
MMP	Microsatellite Mutator Phenotype
MMR	Mismatch Repair
MSH2	MutS homologue 2
MSI	Microsatellite Instability
MSI-H	High Frequency Microsatellite Instability
MSI-L	Low Frequency Microsatellite Instability
MSS	Microsatellite Stable
MutLα	MLH1-PMS2 heterodimer
MutSα	MSH2-MSH6 heterodimer
MutSβ	MSH2-MSH3 heterodimer
NCBI	National Center for Biotechnology Information
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PARP	Poly (ADP-ribose) Polymerase
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
Pol3	Polymerase 3

PTEN	Phosphatase and Tensin homologue Deleted on Chromosome Ten
RB	Retinoblastoma gene
RECQL	RecQ Family DNA helicase gene
RER	Replication Error
RIZ	Retinoblastoma Protein-Binding Zinc Finger
RNA	Ribonucelic Acid
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SMAD	Human homolog of Drosophila Mad gene
STK11	Serine/Threonine Kinase -11
T	Thymidine
SSCP	Single-strand Conformation Polymorphism
STRs	Short Tandem Repeats
TP53	Tumor Protein 53
TCF	T-Cell Factor
TGF-β	Tumour Growth Factor β
TGF-β RII	TGF-β type II Receptor
TNM	Tumour Invasion, Lymph Node Metastases, Distant Organ Metastases
TSG	Tumour Suppressor Gene
WNT	Human Wingless homolog
WRN	Werners Syndrome

Chapter One

Review of the Literature

1.1 Cancer

Cancer is the most common genetic disease: one in three people in the western world develop cancer, and one in five die from it (www.cancer.ca). In recent years, the incidence rates of cancer have steadily risen possibly due to the increase of carcinogens in the environment or perhaps as the result of improved screening programs and diagnostics. Regardless of the reason, an enormous cancer burden in society implores the need for a cure. It is the hope that through our understanding of the etiology and pathogenesis of cancer, we will be able to offer the public reliable early detection tests and improved novel therapies with less toxicity tailored for the patients' specific cancer. With the dramatic recent advances in cancer biology research, for the first time, a cure for cancer seems attainable.

Normal cells transform into malignant cells through a multi-step process. Gradually through the cells malignant progression, external damage is undetected due to loss of internal controls which ultimately affects the cells' ability to control cell growth in terms of arrest, repair, and programmed cell death (PCD) [6]. Cancer cells have a clear growth advantage over surrounding normal cells. They gain access to a unique supply of nutrients and oxygen by promoting angiogenesis, a process that encourages the growth of new blood vessels and recruitment of existing ones [7]. Additionally, the telomerase enzyme is reactivated to maintain stable telomeres during repeated cycles of cell proliferation [8]. Cancer cells gain this advantage through mutations in proto-oncogenes and tumour suppressor genes and through their ability to adapt to their changing environment. This process takes many years to occur. This perhaps explains why there is a lag period between the initiation of a cancer and its detection [9].

Categorization of cancers at the tissue level is based on the cell of origin from which the malignancy developed. Sarcomas are derived from soft tissue and bone, leukemias from blood and carcinomas from epithelial tissues. Cancer usually develops from a benign, well-

differentiated neoplasm [10]{Jukes, Osawa, et al. 1987 JUKES1987 /id}, which has a structure resembling that of normal tissue. Carcinoma *in situ* then ensues. Through time and progression, the neoplasm reaches an invasive stage (cancer) where it attains the ability to penetrate the basement membrane and invade the underlying and surrounding tissues. These cells can also migrate through blood or lymphatic vessels to distant organs to form new colonies of cancer by a process called metastasis.

The development and progression of a cancer is indeed a complex and dynamic process. It involves, many genes, cell-cell interactions, and the interaction between cancer cells and their surrounding environment. In order to fully understand cancer biology, we should persevere in exploring every gene or protein, whatever its structure or putative function, as a possible candidate for its role in cancer.

1.2 Clonal Expansion and Natural Selection

The mutation rate in human cells is approximately 1.4×10^{-10} mutations per base per cell generation, which can account for 2 to 3 mutations in a single cell throughout its lifetime [11]. Cancer cells must therefore acquire, an increased mutation rate, or the ability to select mutant cells with a growth advantage over its neighbours, that would expand and pass on its advantageous and neoplastic mutations to its clonal progeny. These cells must also acquire the progressive ability to escape host mechanisms that regulate cellular proliferation. The most powerful evolutionary force is natural selection, which works not only at the level of whole organisms but also at the level of its component cell (Figure 1.2-1) [12].

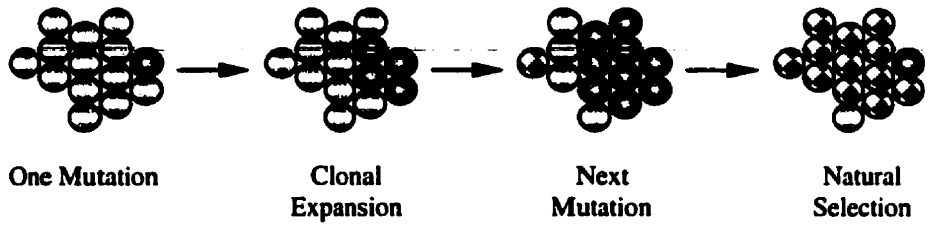


Figure 1.2-1 *Clonal expansion and natural selection of individual tumour cells.*

There are two ways of making a series of successive mutations more likely; some mutations enhance cell proliferation (activating oncogenes/inactivating tumour suppressor genes) creating an expanded target population of cells for the next mutation and other mutations affect the stability of the entire genome (gatekeeper/caretaker genes), thus increasing the overall mutation rate [13]. Additional mutations in the relevant target genes and consequent waves of clonal expansion produce cells that invade surrounding tissues and metastasize.

Tumour development as explained by clonal expansion can result in new cellular equilibriums in the tumour in contrast to the idea that all malignant growth occurs in an exponential fashion. This may explain both benign tumours and long lag phases in tumour growth [14].

1.3 Mutator Phenotype

According to the Mutator phenotype (MMP) hypothesis, the mutation rate early in the evolution of a tumour must be much greater than the mutation rate in normal somatic cells [15]. Initial mutation in a key enzyme involved in DNA replication or DNA repair would adversely affect the fidelity or efficiency of these processes. Thus heterozygote carriers of defective MMR genes (hMSH2/hMLH1) have intact MMR function but acquiring a subsequent mutation that compromises the repair system increases their intrinsic mutation rates. As expected, the tumours of HNPCC patients exhibit markedly elevated mutation rates increasing the likelihood for

mutations in key growth regulatory genes, thereby providing a selective growth advantage. Early mutations that affect genomic stability can produce the mutator phenotype.

The mutator phenotype can increase the probability that cells will gain the ability to adapt to the changing environment of the tumour thereby enabling the tumour cell population to progress along the path toward metastasis. During tumour development, the number of cells in a tumour may be severely constrained by factors such as nutrient and/or blood supply. Selection may be weak in these situations and only a slow increase in the numbers of any genotype will occur. The next mutation that confers a selective advantage is relatively likely to occur in a cell with a raised mutation rate. Through clonal expansion and selection of cells exhibiting the mutator phenotype, tumour proliferation flourishes.

During the final stages of tumour progression, there might be selection against cells exhibiting a mutator phenotype since a deleterious or lethal mutation may be much more likely than an advantageous mutation. An accumulated mutational load might induce apoptosis and slow the growth of this tumour.

1.4 Cancer Genes

It is understood that cancer is a genetic disease that arises from genetic alterations of important functional genes. These alterations can turn on proto-oncogenes or inactivate tumour suppressor genes. Either process would result in a proliferation signal leading to malignant potential of the altered cell.

1.4.1 Oncogenes

Within the genome, there are normal genes known as proto-oncogenes that encode for proteins involved in cell growth and differentiation. When proto-oncogenes are modified through point

mutations (*K-RAS*), gene amplifications (*C-MYC*), or chromosomal translocations (*BCR/ABL*), they become oncogenes. Such changes give cells potential for neoplastic transformation [16]. Approximately 100 dominant oncogenes have been identified at present. These oncogenes are classified according to their function: growth factors, growth factor receptors, signal transducers, cell cycle regulators, and transcription factors [16]. A single mutant allele may affect the phenotype of the cell in a dominant negative manner.

1.4.2 Tumour Suppressor Genes

Tumour suppressor genes (TSGs) function normally as cell growth inhibitors but due to loss of function mutations, the cell loses this important maintenance mechanism. Loss of TSG function was first identified by loss of heterozygosity (LOH) studies, wherein large segments of chromosomes containing the TSGs are translocated or lost. According to Knudson who based his work on the retinoblastoma gene suggested that both alleles must be inactivated for cancer formation to proceed. This is known as the "Two-Hit" Hypothesis. This predisposition is dominantly inherited, while oncogenesis is recessive [17]. The second event might be a new intragenic mutation, gene deletion, chromosomal loss, or somatic recombination [18]

Currently, there have been approximately 30 TSGs that have been identified. Multiple mechanisms exist in the genetic cascade of cancer. One key alteration is the activation of any or all of the various mechanisms that limit the proliferative autonomy of the cell. Some changes suppress proliferation by inhibiting transcription (*RB1*, *P16*, *VHL*), while others effect differentiation (*NF1*, *APC*). Finally, some trigger apoptosis (*TP53*, *WT1*, *PTEN*) and others restrict angiogenic and metastatic capabilities.

1.4.2.1 Gatekeeper Genes

Gatekeeper genes control the initiation of a neoplasm in a specific tissue. Early tumourigenesis of a particular neoplasm depends on the disruption of this gatekeeper gene that under normal circumstances keep the cells in check. *APC* is one such gatekeeper gene. Its loss is a prerequisite for cancer formation in the colon [19]. When *APC* is intact, oncogenes such as *KRAS, C-MYC* and inactive *TP53* are not capable of initiating the cancer cascade. Gatekeeper genes directly regulate the growth of tumours by inhibiting their growth or by promoting their death. The function of these genes is rate limiting for tumour growth since both alleles must be inactivated for tumour development to occur. Individuals who inherit one damaged copy of the gatekeeper gene will not develop cancer. A "second hit" must occur to the normal allele in order to predispose an individual to further mutation which ultimately results in cancer. Since these genes are rate limiting they tend to be frequently mutated in cancers through somatic and germline mutations.

1.4.2.2 Caretaker Genes

The function of caretaker genes is to minimize the rate of genetic alterations or mutations during normal cell division and during carcinogenesis. The DNA mismatch repair system of genes is one such example. These genes are involved in maintaining the integrity of the genome during cell replication. Loss of function of both alleles makes the cell prone to making errors during replication. Inactivation of this repair system leads to genetic instability and indirectly promotes growth by causing increased mutations and thus accelerates the development of cancers.

In dominantly inherited diseases, such as HNPCC, one mutant allele of a caretaker gene (hMSH2/hMLH1) is inherited. In other cases, both alleles when inherited are mutated and thus subject the individual to cancer susceptibility earlier in life, eg. Xeroderma pigmentosum [20]. Targets in cells with defective caretakers are the gatekeeper TSG, other TSG and oncogenes.

Somatic mutations of these genes are rarely found as initiator events in tumours arising in the general population because such a mutation would still need to be followed by several other mutations in order for tumour to develop, i.e. BLM [21], ATM [22]. The inactivation of these genes act to prime the environment for further aberrations. Failure of any of the cellular caretakers involved in maintaining replication fidelity will result in a mutator phenotype resulting in genomic instability.

1.5 Colorectal Cancer

Colorectal cancer (CRC) is the second most common diagnosed cancer in both men and women in Canada. It is estimated that 17,200 new cases will be diagnosed in the year 2001 and approximately 6,400 deaths will result from this disease. Colorectal cancer is generally a malignancy associated with the elderly, with a mean age at diagnosis of 73 years (www.cancer.ca). Most people (80%) present with sporadic disease with no apparent inherited genetic defects. The others (20%) have some family history of colorectal cancer, with only 5-6% of cases attributable to known genetic mutations. The overall incidence of colorectal cancer has been shown to gradually decrease in the last 10 years. The reason is likely a result of aggressive screening programs and early endoscopic intervention. As with other cancers, the development of colorectal carcinomas has both genetic and non-genetic factors influencing its progression.

1.5.1 Pathological Characterization of CRC

Over 95% of colorectal cancers are carcinomas and about 95% of these are adenocarcinomas [5]. Virtually all of these lesions arise from polyps. Their development takes many years to occur and they evolve into several morphological patterns. Tumours in the proximal colon take on polypoid and exophytic characteristics and obstruction is uncommon. In the distal colon, lesions

usually are annular and constricting in nature. Interestingly, right-sided cancers typically resemble tumours found in HNPCC patients. Histologically, these tumours are poorly differentiated, contain signet ring cells, and are necrotic. They tend to exhibit an immune response characteristic of a Crohn's like lymphoid reaction with peritumoural lymphocytes and infiltrating lymphocytes [23].

1.5.2 The Adenoma-Carcinoma Sequence

The genetic pathway leading to colorectal cancer has been extensively studied and the progression from a benign adenoma to a carcinoma is characterized by a series of well-defined histological stages. This progression may take up to 10 to 15 years [10] and is known as the adenoma-carcinoma sequence [24]{KINZLER1996A}. Kinzler and Vogelstein demonstrated that colorectal tumour initiation and progression requires at least seven different somatic changes[25]. The earliest detectable indication of the possible development of a neoplasia in the colonic epithelium is the presence of ACF, aberrant crypt foci [26-30]. Adenomatous polyps are benign growths that may undergo malignant transformation. These polyps are classified into 3 histological types: tubular, tubulovillous, and villous. Mutations in the Adenomatous Polyposis Coli (*APC*) gatekeeper gene are proposed to occur early during the development of polyps [31]. During the adenomatous stage, the tumour acquires *K-RAS* mutations [32] and during the transition to malignancy, *TP53* mutations and deletions on chromosome 18q are observed. Not every tumour needs to acquire every mutation, nor do the mutations always need to occur in a specific order. Cytogenetic analysis of CRCs has revealed that tumour cells undergo loss or gain of parts of chromosomes. The most frequently observed losses involve chromosomal regions 5q, 17p and 18q [24]

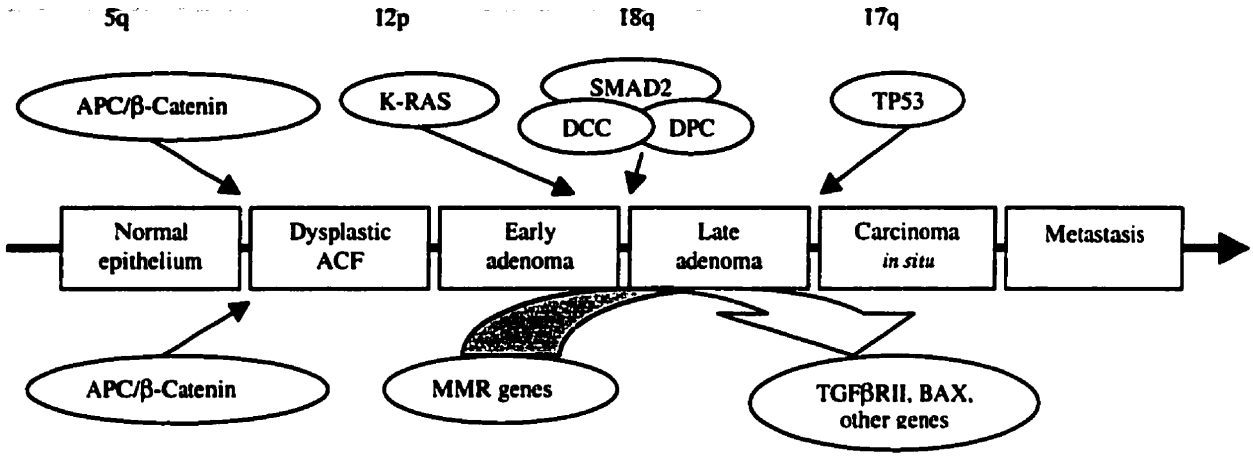


Figure 1.5.2-2 *The adenoma-carcinoma sequence. The critical genes are illustrated, the events above the sequence are typical events in the chromosomal instability pathway, whereas below the sequence represents the typical genetic events associated with microsatellite instability*[25,33]

In addition, other genes may be involved in this progression but have yet to be identified. At the present it is unknown how non-genetic factors such as environment, diet and smoking affect these genes and their mutability.

1.6 Molecular Genomic Instability Pathways of Colorectal Carcinogenesis

It is now widely accepted that different types of genomic instability may result in distinct molecular pathways of carcinogenesis [34,35]. Two destabilizing pathways have emerged leading to the acquisition of the "mutator phenotype" in colorectal cancers. The first is the chromosomal instability pathway, which is characterized by allelic losses, aberrant chromosomal amplifications, and translocations [36]. The second is the microsatellite instability pathway, which is characterized by the instability of short, repetitive DNA sequences (microsatellites). The genes responsible for this instability are the human DNA mismatch repair (MMR) genes. Mutations in these genes lead to the propagation of single base-pair mismatches, resulting in an increased somatic mutation rate.

Mutations in genes that interact with each other appear to be mutually exclusive within an individual cancer. One such example is the RB1 gene [37]. It encodes a protein that has a critical role in the normal progression of cells through the cell cycle [38]. RB1 is inactivated by hyperphosphorylation by cyclin D1. Cancers with cyclin D1 overexpression do not have inactivation of RB1, suggesting that these two genetic events are redundant in disrupting the RB1 pathway [39]. This is also seen in colorectal cancers, where β-Catenin (CTNNB1) mutations are only observed in cancers with intact *APC* [40]. The fact that only a single gene mutation is observed within each functional pathway highlights the importance of these regulatory pathways in tumourigenesis and the frequency with which somatic genetic events may target their individual components.

1.7 Chromosomal Instability Pathway

The hallmark of chromosomal instability (CSI) is clinically manifested by aneuploidy, and frequent LOH in tumour specimens. These changes are likely due to alterations in mitotic checkpoint genes such as BUB1 [41] in some cases. It has been hypothesized that methylation abnormalities are intrinsically and directly involved in the generation of CSI [42]. Demethylation is associated with chromosomal aberrations, including mitotic dysfunction and translocation [43]. Frequent rearrangements are seen in these cancer cells resulting from chromosomal breaks. The most common numerical chromosome aberrations in CRCs, in order of decreasing frequency are: loss of chromosome 18, gain of chromosome 7, loss of the Y chromosome in males, gain of chromosome 20, loss of chromosome 17, gain of chromosome 13 and loss of chromosomes 14 and 22 [44]. Mutations in tumour suppressor genes are common, as are somatic oncogene mutations or amplification.

Instability of chromosomal number can result from *TP53* mutations, which alters karyotypic replication of the centrosome, generating extra spindles and creating changes in ploidy but also with individual losses and gains [45].

1.7.1 Familial Adenomatous Polyposis and the *APC* Gene

Germline mutations in the Adenomatous Polyposis Coli gene are responsible for Familial Adenomatous Polyposis (FAP) [46]. FAP is a dominantly inherited syndrome characterized by the development of hundreds of colonic polyps in afflicted individuals at a very young age. A subset of these polyps invariably progress to cancer. In addition to colonic cancer, these individuals can also develop small intestinal adenomas, gastric adenomas, desmoid tumours, and thyroid neoplasms. The identification of this gene has led to a greater understanding of the molecular genetics involved in the very early events of polyp formation and the progression to neoplasia. Its inactivation occurs very early in the adenoma-carcinoma sequence. Inactivated *APC* can be found in aberrant crypt foci (ACF) [47,48]. There is increasing evidence that *APC* is a major gatekeeper within colonic epithelial cells and that inactivation of this gene leads to uncontrolled cellular proliferation. Biallelic inactivation of *APC* is initiated through somatic mutations in sporadic tumours or initially with germ-line mutations in familial cases on a single allele [49]. The second allele is either inactivated through methylation or through somatic loss leading to the formation of adenomas. *APC* has many putative roles in the cell [50]. Its role in neoplasia is thought to be the result of its control of the WNT signalling pathway via regulation of β-catenin levels. *APC* mutations usually result in protein truncation and subsequent loss of function. *APC* inactivation causes β-catenin stabilization leading to the activation of the T-Cell Factor (TCF)/Lymphoid Enhancer Factor (lef) transcription pathway. Activation of this pathway results in the up-regulation of c-myc, cyclin D1, Matrix Metalloprotease matrilysin, uPA, and Cox2.

APC, therefore functions as a tumour suppressor gene and inactivation leads to a constitutively active proliferative signal. The *APC*/β-catenin signalling pathway is important in the initiation of carcinogenesis irrespective of MSI or CSI status.

1.7.2 Other Genetic Targets in Chromosomal Instability Pathway

1.7.2.1 *Kirsten RAS*

Transducer signals from extracellular growth factors can regulate the cell cycle and therefore proliferation. One example is the *KRAS* gene. Mutations in *KRAS* have been identified in approximately 50% of CRCs. Most of these are point mutation in codons 12, 13 and 61[48]. These mutations are also seen in 80-90% of nondysplastic or hyperplastic ACF but only in 57% of dysplastic or adenomatous aberrant crypt foci [30]. Even though *KRAS* mutation is an early event, it is not entirely associated with tumourigenesis. In addition these mutations lead to a constitutively activated *KRAS* gene which results in the up-regulation of DNA methyltransferase, cyclin D1, and gastrin [47].

1.7.2.2 *TP53*

Loss of heterozygosity at chromosome 17p and missense mutations in the *TP53* tumour suppressor gene on the remaining chromosome 17 are found in more than 80% of CRCs and represent a late event in the adenoma-carcinomas sequence [24,51]. Most of these mutations localize to exons 5-8 in the *TP53* gene. The gene product is a sequence specific transcriptional activator that plays diverse roles in the regulation of the cell cycle and apoptosis. *TP53* acts as a checkpoint control protein that determines cellular fate upon DNA damage and is often referred to as the gatekeeper of the genome. *TP53* can delay the G1/S transition thus allowing for repair of DNA damage [52]. It can alternatively trigger apoptosis in response to DNA damage most

probably when the lesions are too extensive and DNA repair fails. Loss of *TP53* tumour suppressor activity results in cells tolerating DNA damage, which eventually leads to cancer formation by giving these cells a growth advantage through the cells inability to eliminate damaged cells by apoptosis.

1.7.2.3 18q

Cytogenetic analysis has revealed frequent loss of the long arm of chromosome 18 (18q) in colorectal cancer. It is now known that approximately 50% of late adenomas and in 70% of CRCs have loss of 18q [24]. Several putative TSGs, *DCC*, *SMAD4* and *SMAD2* have been mapped to 18q21 [53]. *DCC* is a putative TSG that encodes for a cell surface receptor for the ligand netrin. Several lines of evidence have pointed to the importance of *DCC* in colorectal cancer. First, in addition to frequent loss of 18q in colon cancer, decreased *DCC* protein has been found in colon cancer cell lines. Second, fusion of normal 18q or the *DCC* gene itself in CRC cell lines suppresses tumourigenicity [54,55]. Third, decreased DCC protein expression has been associated with poor survival in patients with colorectal cancer [56].

The *SMAD* genes play an important role in the transforming growth factor beta (TGF-β) signal transduction pathway [57]. TGF-β is one of the more potent inhibitors of cell growth. Tumours can become resistant to TGF-β either by loss of function of one of the two receptors, TGF-βRI, TGF-βRII, which interact with the TGF-β ligand [1] or by mutations in other downstream genes e.g. members of the SMAD family, involved in the signalling cascade. Mutational analysis has revealed that *SMAD4* and *SMAD2* mutations only occur in 16% and 6% respectively in primary colorectal cancer [58,59]. *SMAD4* is thought to play a late role in carcinogenesis since metastatic colorectal tumours are more likely to have *SMAD4* mutations compared with primary tumours [60].

1.8 Microsatellite Instability Pathway

Approximately 15% of all CRCs display high frequency microsatellite instability (MSI-H) (alterations at >40% of the loci tested). Microsatellite instability (MSI), also known as replication error (RER), is defined as a change in length of a microsatellite either through an insertion or deletion of repeated units within a tumour when compared with normal tissue. Mono-, di-, tri-, and tetranucleotide markers can be used for studying MSI. Tumours with MSI are commonly diploid and a low percentage of cases display allelic loss [61]. Only 10% of these MSI-H CRCs are due to inherited defective MMR genes. The other 90% of the cases are sporadic resulting from the probable suppression by methylation suppression of the promoter of the hMLH1 gene.

1.8.1 Microsatellite DNA

Microsatellite DNA sequences are short tandem-repeats that are randomly distributed, accounting for approximately 1% of the human genome [62]. Typically these tandem sequences consist of DNA repeats of six base pairs or fewer and the total length of the stretch is less than 100 base pairs. In humans, the most common repeat sequences are $(A)_n/(T)_n$ and $(CA)_n/(GT)_n$, which are characteristically located within non-coding DNA sequences. The polymorphisms seen in microsatellites are unique to the individual and thus can serve a role in forensics through DNA fingerprinting.

1.8.2 Criteria for Determining Microsatellite Instability

In 1997, "The International Workshop on Microsatellite Instability and RER Phenotypes in Cancer Detection and Familial Predisposition" proposed a panel of five microsatellite markers to

be used in MSI analysis [5]. For the purpose of providing uniformity, two mononucleotide repeats (BAT25 and BAT26) and three dinucleotide repeats (D2S123, D5S346, and D17S250) were recommended. Using this reference panel, tumours having instability in two or more markers are defined as MSI-H (high instability), tumours having instability in one marker requires further testing with up to five additional markers (BAT-40, BAT-RII, D18S58, D18S69, and D17S787) are screened. If <40% of all markers exhibit MSI, the tumour is defined as MSI-L (low instability), and tumours where none of the markers exhibit MSI are defined as MSS (microsatellite stable).

The mononucleotide marker BAT26 has some advantages over other markers in MSI analysis. It is extremely sensitive [63-66] and shows negligible size variation between both alleles within one individual and between different individuals [63]. Several studies support the use of BAT26 on tumour DNA alone, without matching normal DNA [63,64,67]. However, a germline polymorphism in the BAT26 locus has been detected [65,68], although rare, emphasizing the need for the matching normal DNA to avoid misclassification. Instability at this marker is only informative when alleles are shifted more than 3 base pairs.

MSI-H colorectal cancer appears to follow a different path from MSS. MSI tumours show infrequent LOH, infrequent *KRAS* and *TP53* mutations whereas; MSS tumours frequently show mutations of these genes and losses of 5q, 17p, 18q [69,34]. The methylation of genes is becoming recognized as a method of inactivation and suppression of genes thought to be important in tumourigenesis. This mechanism of methylation needs to be elucidated to further our understanding of the etiology and pathogenesis of colorectal carcinomas [70].

1.8.3 Hereditary Non-Polyposis Colorectal Cancer (HNPCC)

HNPCC is the second most common inherited form of colorectal cancer accounting for approximately 2-4% of all colorectal cancer in the Western world. Individuals afflicted with disease present with cancer at a mean age of 45 and have first-degree relatives with colorectal cancer, HNPCC or other variants. These patients also have a better long-term prognosis and have a lower prevalence of metastatic lesions at the time of diagnosis. HNPCC is a cancer syndrome and therefore predisposes afflicted individuals to other cancers such as endometrial, gastric, renal, ureteral, and ovarian cancer. The hallmark of this disease is a defect in the DNA mismatch repair system. The DNA of patients with HNPCC has a typical phenotype known as microsatellite instability. The frequency of MSI in HNPCC is >90% while it is only about 15% in sporadic colon cancer [63]. There appears to be a global instability in the replication and repair of microsatellites throughout the genome. Instability of these microsatellites in the coding region of key genes (oncogenes and suppressor genes) can in theory predispose cells to tumourigenesis.

The link showing the involvement of MMR genes in HNPCC was shown with the discovery of the *MSH2* gene in HNPCC kindreds. It is now believed that six MMR genes (*hMSH2*, *hMLH1*, *hMSH3*, *hMSH6*, *hPMS1*, and *hPMS2*) are important in HNPCC. hMSH2 and hMLH1 play prominent roles in this syndrome. A high prevalence of mutations in hMSH2 (50%) and hMLH1 (30%) has been detected in these individuals. Usually, these individuals only possess a mutation in one of the MMR alleles retaining the normal allele. This is sufficient for the daily operation of the MMR system. It is when the remaining normal allele is altered that tumour formation is initiated through the subsequent mutations that inevitably occur. The end result is the greater accumulation of mutations in other key regulatory cell cycle control genes (*KRAS*, *TP53*, *TGFBRII*) that accelerate tumour progression. Thus, the mutations in the MMR genes promote a positive feedback loop that enhances this hypermutability.

Approximately 20% of HNPCC cases lack identifiable mutation in MMR genes [71] suggesting either that there are mutations in other possible MMR-related genes can predispose to the development of tumour with MSI [72] or that the sensitivity of the currently utilized methods for mutation detection is not precise enough to reveal the remaining aberrations in these patients.

Table 1.8.3-1. *Human mismatch repair genes.*

MMR Gene	Chromosomal Location	Length	Protein Size
hMSH2	2p15-16	16 exons	106 kD
hMLH1	3p21	19 exons	85 kD
hMSH3	5q11-12	24 exons	160 kD
hMSH6	2p16		160 kD
hPMS2	7q22	15exons	96 kD
hPMS1	2q31-33		

1.8.4 Microsatellite Instability in Colorectal Cancer

Microsatellite instability is found in more than 90% of hereditary non-polyposis colorectal cancer, [64,73,74] but it is also seen in approximately 15% of sporadic colon cancers [75]. The presence of MSI is thought to occur very early in carcinogenesis. This phenotype is present in ACF, which are microscopic lesions that have been postulated to precede the development of adenomas and are considered the earliest premalignant lesions in colon cancer [26,27,75]. Whether the mutational events of the genes, *APC*, *KRAS*, and *TP53* are specific for tumours of a particular MSI status, is unsettled (Table 1.8.4-2).

Table 1.8.4-2. *Mutation frequencies of target genes in the adenoma to carcinoma sequence.*

Gene	MSI-H CRCs	MSS CRCs	Reference
APC	21%	43%	69,76
KRAS	7%	22-31%	77,78
TP53	0-40%	33-67%	69,77,78

It has been debated when during the adenoma to carcinoma sequence MSI actually appears. It may be an early event {SHIBATA2001}, after *APC* mutation, or at the adenoma-carcinoma boundary [79-81]. They all may be correct depending on the type of colorectal cancer. Konishi *et al.*, (1996) found that in HNPCC patients, MSI is an early event and in sporadic cases, MSI is an event in the later stages of tumourigenesis [69]. Stoler *et al.*, (1999) through their investigations revealed that MSI is not a late event but also occurs early (evidence for genomic instability as a cause rather than an effect), even in sporadic tumours [82].

1.8.5 Microsatellite Instability in Hereditary Non-Polyposis Colorectal Carcinomas

About 10% of all MSI-H CRCs are due to inherited defects in MMR genes and this accounts for 70% of the autosomal dominant cancer syndrome HNPCC [83]. The majority of germline mutations responsible for this disease are present in hMSH2 and hMLH1 and only rarely in hPMS2 and hMSH6 [84,85]. Most studies focus on detecting aberrant hMSH2 and hMLH1 using standard mutation methods but this is inefficient. However the use of hMSH2 and hMLH1 immunohistochemistry has been shown to be an efficient alternative [86].

In hereditary cases, the gene mutations are present in the germline, and thus are present in every cell in the body. These genetic defects are passed from generation to generation resulting in a higher frequency of cancers in these families. Due to this predisposition, kindreds of patients with hereditary colorectal syndromes should be screened for germline mutations in the MMR genes as a measure of early detection.

Table 1.8.5-3. *Mutation frequencies of genes in hereditary and sporadic cases of colorectal cancers.*

Gene	Repeat	Hereditary CRCs	Sporadic CRCs
TGFβRII	$(A)_{10}$	78-83%	82-90%
IGF2R	$(G)_8$	13%	9%
BAX	$(G)_8$	52-55%	13-50%
hMSH3	$(A)_8$	50-58%	39-46%
hMSH6	$(C)_8$	33%	28-36%
Caspase 5	$(A)_8$		62%

1.8.6 Microsatellite Instability in Sporadic Disease

The remaining cases of MSI-H CRCs are sporadic accounting for approximately 13-15% of all CRCs. In sporadic cases of cancer, gene mutations are induced by exposure to carcinogens and gene silencing due to epigenetic mechanisms. The defect occurs in a somatic cell, and will only be passed on to this cell's descendants. This is a stochastic event and therefore the frequency of cancer in these families is not increased. In addition to HNPCC, defects in MMR have been associated with sporadic colorectal, endometrial, and gastric carcinomas. Unlike the HNPCC cases, the majority of these sporadic carcinomas do not have identifiable mutations in either *hMLH1* or *hMSH2*, but rather epigenetic transcriptional silencing has been linked to their MSI phenotype. *hMLH1* methylation is usually responsible for the majority of instability seen in sporadic cancers [87-89].

1.8.7 DNA Mismatch Repair

The DNA mismatch repair system is known to play two major roles in the cell. First, it repairs errors made during DNA replication, or as a result of some environmental insult to the DNA or its precursors. Second, it is involved in the processing of recombination intermediates to yield new configurations of genetic markers [83]. DNA mismatch repair genes encode for enzymes that survey the newly replicated DNA for errors and repairs all mismatched bases. Defects in MMR genes result in replication errors and genetic instability [90].

This highly efficient mismatch repair system can address heteroduplexes containing extrahelical nucleotides [91]. It is extremely important especially in humans for this system to be intact due to the presence of numerous tracts of repeated nucleotide motifs in our genome. Mismatches occur during DNA replication at mononucleotide repeats when the primer and template strands slip relative to one another. The result is either an insertion if the primer strand slips back, or a deletion if the primer (nascent) slips forward [90].

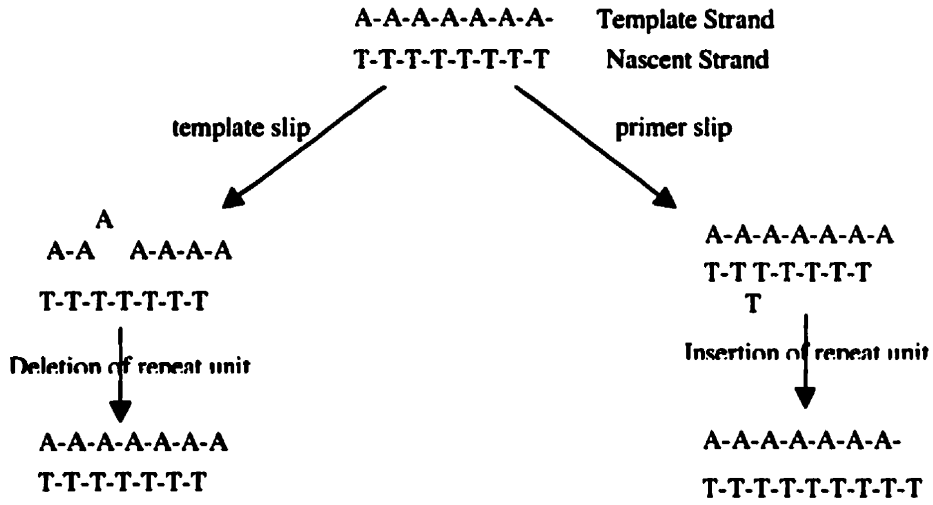


Figure 1.8.7-3. *Consequences of slippages in DNA replication with deficient mismatch repair.*

A malfunctioning mismatch repair system would result in the transmission of the resulting insertion/deletion loops (IDLs) that result in frameshift mutations. When a subset of MMR genes are mutated in tumour cells, it is suggested that these cells acquire a selective growth advantage [92].

The human mismatch repair machinery is complex. Mismatches and insertion/deletion loops (IDLs) are recognized by MutS homologs, hMutSα and hMutSβ. Both of these homologs contain hMSH2, which heterodimerizes with hMSH6 (GTBP) in hMutSα [93,94] or with hMSH3 in hMutSβ [95,96]. Studies have shown that hMutSα preferentially recognizes single base mismatches and loops of 1 base, whereas recognition of IDLs containing 2 to 8 base pairs are primarily

mediated by hMutSβ [97,72,98]. There appears to be some redundancy in the function of the MutS homologs in that they both can recognize IDLs of 1 base [99,100]. The MutL homolog consisting of hMLH1 and hPMS2 [101] binds the MutS heterodimers that are bound to mismatched DNA. In yeast, mlh1 can form complexes with two other MutL homologs, MLH2 and MLH3 [102,103]. The mammalian homolog of the yeast MLH3 has been cloned and the microsatellite instability associated with expression of a dominant-negative MLH3 protein is consistent with its role in MMR [102]. The predicted hMLH1 interaction domain indicates that hMLH3 might replace hPMS2 in the mismatch repair complex. Very few germline or somatic mutations have been found in *hPMS2*, and *hPMS2-/-* mice do not display colon cancer susceptibility [104,105]. However, *hMLH3* might have a unique role in mismatch repair, since it does not show great similarity to *hPMS2*. Raschle *et al.*, (1999) & Leung *et al.*, (2000) have also identified another complex (hMutLβ) formed by hMLH1 and hPMS1 [106,107]. Research shows that there is redundancy in DNA repair functions within the cell. Therefore, it may be required that more than one MMR gene must be inactivated to in order to observe high frequency microsatellite instability [108].

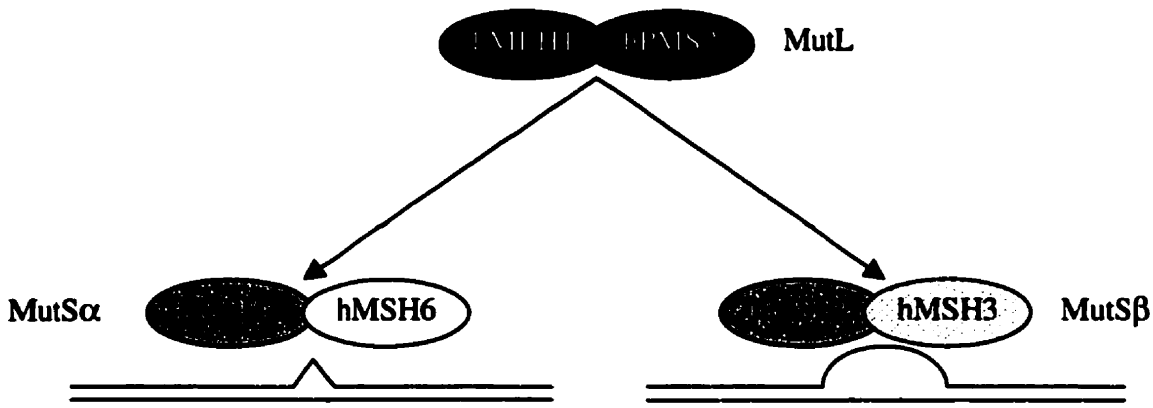


Figure 1.8.7-4. *Schematic of the MutL and MutS homologs in the human mismatch repair system. The MutSα heterodimers recognizes one mispairs and small IDLs, whereas the MutSβ heterodimers recognizes the larger IDLs.*

Umar suggest that since hMSH2 interacts with replication processivity factor PCNA at a step preceding DNA resynthesis, the termini of the Okazaki fragments might be used to determine strand specificity [109]. It could be the result of interactions between hMSH2 and PCNA or it could be the use of termini of Okazaki fragment as a strand discrimination signal [109]. A novel DNA mismatch repair gene *MBD*4, also known as *MED*1, has recently been cloned [110], and was proposed to be a functional homolog of MutH. It forms a complex with hMLH1, binding to only methylated-CpG containing DNA, thus suggesting a possible role as a mediator of methylation-based strand-specificity in eukaryotic mismatch repair [110].

Proliferating cell nuclear antigen (PCNA) encircles DNA and tethers DNA polymerase to the template during DNA replication. It has been speculated that the PCNA interaction might allow the MMR machinery to use the position of the replication complex to determine which strand should serve as a template during repair [111], or might trigger involvement of the 3' to 5' exonuclease activities of the DNA polymerases in mismatch removal [112,113]. It has been postulated that PCNA guides the MMR complexes to free termini in newly replicated DNA strands [114,115].

The human mismatch repair system has been linked directly to signaling to DNA damage-induced apoptotic response. A "hydrolysis-independent sliding clamp" (SC) has been proposed by Fishel *et al.*(unpublished work). In the SC model, the MMR proteins are signaling molecules that are direct sensors that promote DNA repair or incite apoptosis. Using ADP-ATP nucleotide exchange, the ADP bound MutS heterodimers bind the DNA. The bound ADP becomes ATP, which transforms the structure of MutS so that it forms a clamp around the DNA. Multiple signaling clamps bind near the mismatch, to a reach a threshold number, which may explain the functional redundancy seen in the MutS heterodimers, MutSα, and MutSβ. If the MMR proteins do in fact act as direct sensors to downstream effectors of apoptosis signaling,

then the need for a p53 response is eliminated. Thus, the selection of these mismatch deficient cells is based on their resistance to DNA damage-induced apoptosis.

1.9 Alternative Pathways

The adenoma-carcinoma sequence may not be the only feasible pathway for the progression of CRCs. It is possible that cells acquire all the genetic mutations necessary for malignant behaviour without forming a visible adenomatous polyp [116,117]. This also suggests that genomic instability is an early event in the multi-step process of carcinogenesis that occurs in a pre-neoplastic lesion. Studies have shown that MSI-H adenomas and carcinomas have similar degrees of genetic diversity [32] and equally long periods of progression have been proposed [118]. Adenomas may arise after long periods of occult progression rather than simply being the starting point for carcinomas [118].

1.9.1 CpG Island Methylator Phenotype

A phenotype described by widespread gene methylation has been defined as the CpG island methylator phenotype (CIMP) [70]. It was postulated by Toyota *et al.*, (1999) that colorectal cancers could be better classified according to their CIMP status rather than according to MSI or LOH status. The hypermethylation of CpG islands within promoter regions leads to silencing of gene expression in the absence of coding region mutations [70]. CpG islands are short sequences rich in the CpG dinucleotide and can be found in the 5' region of about half of all human genes. Methylated CpGs are recognized by proteins that recruit histone deacetylases, leading to stable transcriptional repression [119]. This is most commonly seen in physiological conditions such as X chromosome inactivation and genomic imprinting [120]. Aberrant methylation of CpG islands has

also been detected in genetic diseases such as the Fragile X syndrome, in aging cells [121,122]; and in cancer [123].

There is global hypomethylation in benign adenomas and malignant colon cancers compared to in normal tissues [124]. This global hypomethylation occurs in the setting of localized hypermethylation [125]. Hypermethylation has been proposed as a mechanism leading to the inactivation of tumour suppressor genes. The causes for aberrant methylation are unclear. Aging could play a role in this process since the majority of CpG islands methylated in colon cancer have been shown to be methylated in a subset of normal colonic cells during the aging process [122].

Approximately half of the tumour suppressor genes that have been shown to have germ-line mutations in familial cancers are aberrantly methylated in sporadic cancers [70]. Aberrant methylation of TSGs, RB, VHL, p16, MLH1, and BRCA1 have been reported [126]. Methylation of hMLH1, as well as INK4A^{p16} has been observed in 28-55% of colon tumours [127,128].

It is still unclear where methylation lies in the sequence of tumourigenesis [42]. There is evidence that suggests CIMP precedes methylation of hMLH1. First, CIMP is detected in approximately half of all colonic adenomas but not all of these tumours have hMLH1 methylation, and MSI is rarely seen. Second, CIMP is not simply caused by MMR defects because MSI is absent in more than half of the CIMP+ cases. In sporadic CRCs, the majority of cases with MSI may be caused by CIMP followed by methylation of hMLH1 and other MMR genes.

However, based on the evidence by Esteller *et al.*, (2000), in their research in promoter methylation of various genes (p16^{INK4a}, p15^{INK4b}, p14^{ARF}, p73, APC, BRCA1, hMLH1, GSTP1, MGMT, CDH1, TIMP3, and DAPK) in 15 different tumour types, they suggest that methylation is a common event and is present in all tumour types. This would imply that methylation is not

predictive of a particular tumour phenotype (CIMP) and thus could not be used to categorize tumours regardless of MSI status

1.10 Replication Errors

Spontaneous point mutations in mammalian cells emerge as a function of time rather than as a function of the number of cellular division cycles [129]. DNA replication is strictly regulated and produces <1 error per 10^{10} nucleotides synthesized [130]. The fidelity of base insertion by replication polymerases is less effective, making one mistake for every 10^4 to 10^6 nucleotides incorporated. The overall fidelity is attained through the combined effort of DNA polymerase proofreading and further editing by the post-replication mismatch repair system [108].

DNA synthetic errors are initiated by one of two very different events: insertion of an incorrect deoxyribonucleoside monophosphate or template-primer slippage. Misinsertion results in a base substitution mutation whereas deletion and addition errors can result from template-primer slippage during polymerization. Mutations observed both in microsatellites and in genes are consistent with errors due to slippage by DNA polymerase during the course of DNA synthesis. This is seen in MSI-H cancers as described above.

The driving force for mutation accumulation is the production of alterations in nucleotide sequence that exceeds the cellular capacity for repair. Alternatively, the rate of DNA repair can be reduced, which leads to persistent damage that causes mutations during subsequent rounds of replication. Acting as another positive feedback regulator for the emergence of a mutator phenotype, DNA polymerases may be targeted directly for mutation or indirectly by mutating a molecular switch, that causes the cell to use a less accurate polymerase (pol α) over a more accurate one [131]. Thus, polymerase alpha performs a greater amount of replication, the chances of acquiring replication errors increases. Through evaluating the sequence of polymerase α

(GI35567) a polyadenine repeat $(A)_8$ at nucleotide position the 3167 to 3174 was revealed, therefore presenting itself as a target in MSI cancers. Also, it has been found that a specific mutation in the DNA polymerase epsilon causes +1 frameshift mutations in yeast, a mutational event seen typically in MSI-H tumours [132]. It seems likely that performing a mutational analysis of polymerase genes in specific cancer cells may elucidate the presence of error-prone DNA polymerases. The observance that DNA polymerase β is mutated in 5 of 6 cases of colon cancer is evidence of this speculation [133]. Additionally, mutations in DNA polymerase delta are present in colon cancer cell lines, further contributing to the accumulation of mutations in these cancers by compromising the fidelity of DNA replication [134].

1.10.1 Hypermutability of Mononucleotide Repeats

Mononucleotide repeats are slippery sequences for DNA replication and for transcription [135,136]. It could be predicted that evolution would select against long stretches of adenines within protein coding sequences [137]. It is hypothesized that nature selects against long stretches of mononucleotide tracts in protein-coding sequences to avoid the possible propagation of an aberrant insertion/deletion. In one study, 150 Lys-Lys-Lys motifs in 128 human proteins from the SWISS-PROT database were analyzed. It was found that the observed frequency (0.020) was much less than the predicted frequency of (0.067) finding the triple lysine repeat. Evolution has elected to avoid long stretches of adenines as a means to code for Lys-Lys-Lys [136].

1.10.2 Somatic Genetic Targets of MSI CRCs

1.10.2.1 *TGF-β Pathway*

Typically, MSI-positive tumours acquire somatic deletions and insertions in simple repeated sequences. Several genes involved in tumourigenesis contain mononucleotide repeats in their

coding regions. One such gene is the *Transforming Growth Factor-β Receptor type II gene (TGF-βRII)*. TGF-β is a potent inhibitor of colonic epithelial cell growth [1]. About 90% of the DNA mismatch repair-deficient colorectal carcinomas display mutations in the poly-A tract and a short GT repeat present in the coding sequence of the *TGF-β RII* gene [1,138-140]. Also somatic deletions have been seen in the *Insulin-like Growth Factor Receptor type II (IGFIIR)* (13%) [2]. *IGFIIR* is required for the extracellular activation of *TGF-β*, and targeting of IGF by microsatellite instability may represent another step at which disruption of the *TGF-β* growth control pathway occurs.

1.10.2.2 APC/β-Catenin/TCF Pathway

As mentioned above, *APC* is termed the gatekeeper of colonic epithelium. Germline mutations of this gene result in FAP and somatic mutations (~80% of sporadic CRCs) also appear to be responsible for the initiation of colorectal cancers in the general population [48,141]. The APC protein interacts with GSK3-β and AXIN to form a complex that is essential in downregulating β-Catenin [142,143,144]. *APC* mutations lead to the accumulation of intracellular β-Catenin, which results in increased transcription at the TCF/LEF promoter leading to the up-regulation of genes such as *C-MYC*, *CYCLIN D1* and other essential genes. In colorectal cancers, mutations in the regulatory regions of the β-Catenin gene (*CTNNB1*) have been found exclusively in approximately 50% of colorectal tumours without *APC* mutations. Recently Liu *et al.*, have shown *AXIN2* mutations in MSI CRCs (11 of 45) with intact APC protein and without *CTNNB1* mutations [145]. Thus, these mutations are mutually exclusive in CRCs. Frameshift mutations have also been described in the $(A)_9$ tract of the TCF4 gene. 40% of the MSI CRCs exhibited mutations distal to the functional domains, in a region that encodes for alternatively spliced

transcripts. Therefore, the functional consequence of these mutations remains unclear.

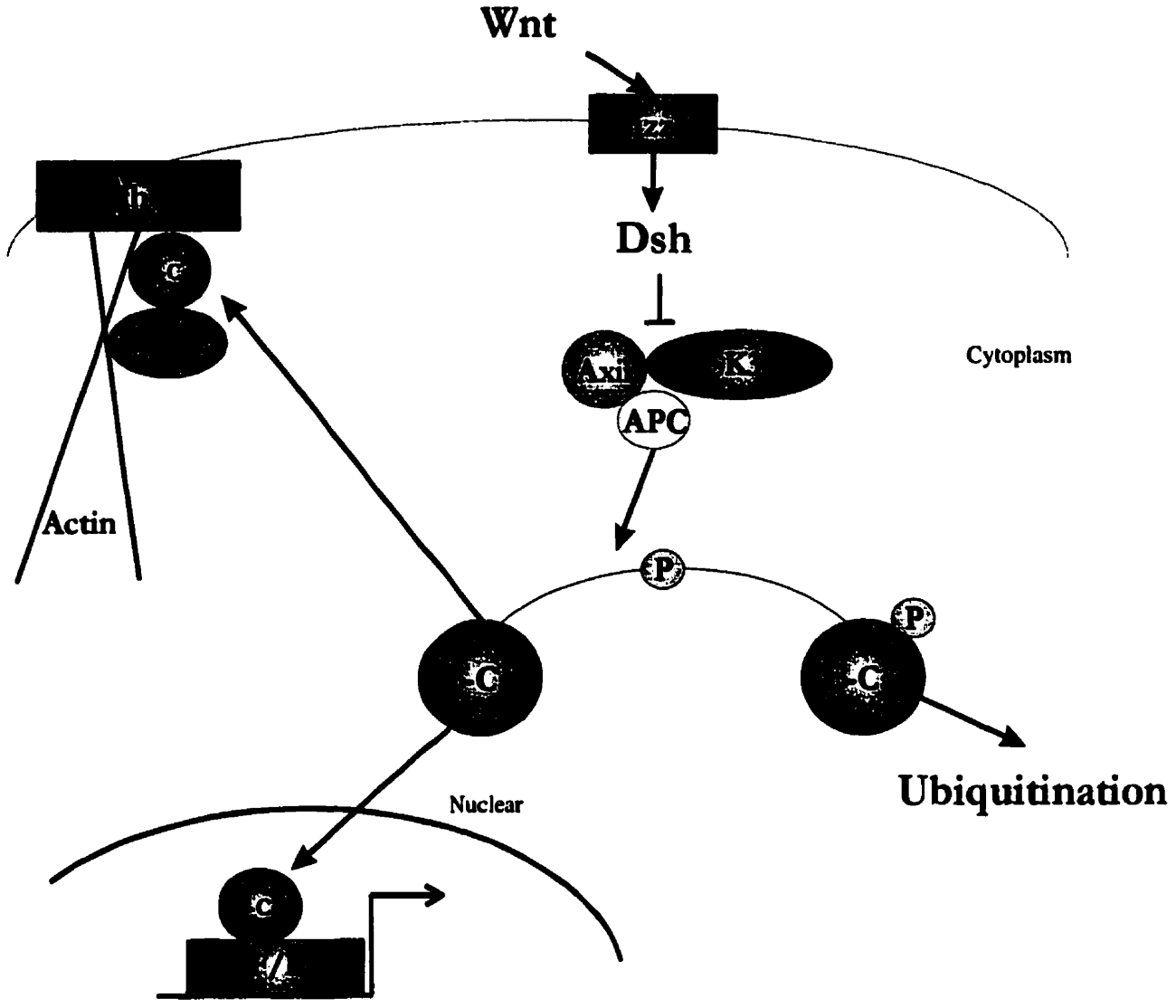


Figure 1.10.2.2-5. *Schematic of Wnt signalling pathway.*

1.10.2.3 DNA MMR genes

Moreover, the mismatch repair genes *hMSH3* and *hMSH6* frequently show somatic deletions [146]{1}. Accumulation of somatic mutations in these mismatch repair genes is believed to provide a positive feedback loop such that these tumours can enhance their own phenotype. The target for mutations in *hMSH3* is an $(A)_8$ repeat, whereas *hMSH6* has a $(C)_8$ repeat. Both genes have

confirmed roles in human mismatch repair, and germline *hMSH6* mutations are associated with a minor subset of HNPCC cases [147-149]. The recently identified MED1 (MBD4) gene is thought to be a homolog of the MMR genes and is mutated at the $(A)_{10}$ in 10 of 23 (43%) MSI CRCs and cell lines tested [110].

1.10.2.4 Other Genes

In the quest to find other relevant genes in MSI tumourigenesis, somatic mutations in other genes containing mononucleotide repeats have been identified. These genes include: *BAX*, an apoptosis regulator, *PTEN, RIZ, caspase 5*, and *hRAD50*.

BAX is a proapoptotic gene that is mutated with some specificity in MSI CRCs. Approximately 50% of all MSI CRCs have acquired a frameshift mutation due to insertion/deletion at the $(G)_8$ repeat in the coding region of *BAX*. Fewer mutations exist in the *caspase 5* gene that contains a $(A)_8$.

PTEN is a phosphatase and tensin homologue deleted from chromosome 10. *PTEN* is also known as *MMAC1*, mutated in multiple advanced cancers or *TEP1*, (TGF-β regulated and epithelial cell-enriched phosphatase 1). PTEN degrades phosphatidylinositol (3,4,5)-triphosphate [PtdIns(3,4,5)P_3], which is a regulator of PKB/Akt. The loss of PTEN indirectly regulates PKB through the accumulation of PtdIns(3,4,5)P_3 [150]. PTEN interacts directly with focal adhesion kinase (FAK) and rescues its tyrosine phosphorylation. Expression of PTEN was shown to inhibit cell migration, integrin-mediated cell spreading, and formations of focal adhesion [151]. Thus, the loss of PTEN could enable the cell take on malignant characteristics such as metastasis. This gene has been found to be mutated at the $(A)_6$ and $(A)_8$ repetitive tracts in 19% of CRCs and cell lines with MSI.

$(A)_8$ and $(A)_9$ mutations of the Retinoblastoma Protein-Binding Zinc Finger (*RIZ*) gene was identified in 25%-38 % of MSI CRCs [152-154] of MSI CRCs, most of these mutations were found to be heterozygous. Mutations in the $(A)_9$ tract in the *hRAD50* gene was seen in 13 of 39 MSI colon cancers[155]. hRAD50 forms a complex with hMRE11 and MBS1 and functions in homologous recombination, activation of cell cycle checkpoint, nonhomologous end joining, meiotic recombination and telomere maintenance [156,157].

The relative importance of these frameshift mutations is relatively unknown. The presence of mutations in these genes does not establish a role in colorectal carcinogenesis. The National Cancer Institute workshop on microsatellite instability made the following recommendations for establishing the role of exonic microsatellite mutations [5]

1) a high frequency of inactivation in MSI-H cancer;
2) biallelic inactivation by simultaneous alteration of the other allele's repeat tract , point mutation, or loss;
3) involvement of the target gene in a bona fide growth suppressor pathway;
4) inactivation of the same gene, or another gene within the same pathway; and
5) functional suppressor studies in *in vitro* or *in vivo* models.

To date, *TGF-βRII* is the only candidate gene to fulfil most of these requirements.

The majority of colorectal cancers follow the chromosomal instability pathway towards tumourigenesis, and the several genetic targets have been identified at the various stages of progression. However, in the smaller subset of tumours with microsatellite instability, the progression of the tumours may differ within individuals since countless genes are affected by the deficient mismatch repair system. With only one identified gene (*TGF-βRII*) with a definite role in MSI pathogenesis, exhaustive searches are required to identify additional progression targets to further our understanding of this pathway.

1.10 Thesis Overview

1. To identify apoptosis regulatory genes with polyadenine repeats and characterize their mutation frequencies against a panel of high frequency microsatellite instability (MSI-H) primary colorectal cancers as well in a panel of colon cancer cell lines (Chapter Two).
2. To develop, in collaboration, a bioinformatics approach to systematically identify genes containing a specific sequence in the coding region of the human genome from annotated human sequences (Chapter Three)
3. To utilize this bioinformatics approach to investigate codon usage and codon bias in the human genome (Chapter Three).
4. To implement this bioinformatics approach to identify genes with polyadenine repeats that have functional significance in cell cycle control and to characterize mutation frequencies of the selected genes against the same panel of samples listed in Aim 1 (Chapter Four).

Chapter Two

Apoptosis Regulatory Genes in Colorectal Carcinomas with Microsatellite Instability

Chapter Two

Apoptotic Regulatory Genes in Colorectal Carcinomas with Microsatellite Instability

2.1. Background

Apoptosis is critical for tissue morphogenesis, homeostasis, development of immunity, host defense mechanisms, and elimination of damaged and potentially neoplastic cells [158]. Defective apoptosis may cause or exacerbate a variety of human diseases, including Alzheimer's, Huntington's Chorea, autoimmune diseases, and cancer [159]. A major form of human cancer is the epithelial derived carcinomas. They normally undergo apoptosis when they detach from their surrounding substrates and when they are subjected to cellular insults. Therefore, suppression of apoptosis can be an important mechanism in promoting carcinoma development [160].

In MSI-H cancers, deregulation of apoptosis would seem to be particularly important in order to tolerate the abundance of mutations that are acquired in association with the profound defect in DNA repair. Several genes with roles in the apoptotic pathway are mutated specifically in cancers that display microsatellite instability. The pro-apoptotic gene *BAX*, is one such example. Bax belongs to the Bcl-2 family of proteins and is a key player in apoptosis. Programmed cell death is promoted by Bax-induced formation of ion-permeable channels that disrupt the mitochondrial membrane barrier [161]. Bax also contributes to the apoptotic response by binding to antiapoptotic proteins of the Bcl-2 family via its BH3 domain to inhibit their function [162]. More than half of gastrointestinal MSI-H tumours contain *BAX* mutations, whereas these mutations are rarely found in MSS tumours. Bax and Bcl-2 are antagonists and their interactions regulate apoptosis and other apoptotic stimuli within a cell [163].

BAX mutations are found in approximately 50% of MSI-H CRCs [3,35,77,164-166]. Our mutational analysis in a panel of young patients with MSI-H CRCs concurs with the published findings. These inactivating mutations occur at a hotspot, $(G)_8$ in the coding sequence of *BAX*

and may explain the lack of p53 mutations in MSI-H CRCs. The exact mechanism of p53-mediated apoptosis is largely unknown. Bax is most likely involved, and through the disruption of the mitochondrial membrane, it releases cytochrome c. Cytochrome c activates the Apaf-1 protein, which in turn activates the enzyme procaspase 9. *APAF-1* also has a mononucleotide tract within its coding region and is likely to be a target for inactivation in MSI-H cancers. Although the functional significance is uncertain, *APAF-1* has been described to have somatic frameshift mutations at an $(A)_9$ repeat albeit in small number (4/30, 13%) of MSI-H CRCs [167]. Earlier this year, Soengas et al, established that *APAF-1* is inactivated in metastatic melanomas, which lead to defects in the execution of apoptosis. Interestingly, they found that *APAF-1* was inactivated through an epigenetic mechanism that silences the gene through methylation probably at an enhancer or insulator region of the *APAF-1* gene [119]. Consequences of *APAF-1* mutations result in its inability to heterodimerize with caspase 9 through their respective CARDs (Caspase Recruitment Domain) and subsequent downstream signalling to the effector caspase (caspase 3) is obliterated [168].

Caspases are a family of cysteine proteases that selectively cleave protein substrates involved in maintaining genomic integrity, thereby assuring the orderly execution of apoptosis [169]. Initiator caspases, 8 & 9, initiate the proteolytic cascade leading to the cascade of cleavage and activation of downstream caspases 3&6; these (downstream/execution) cleave selected target proteins. Caspases disable normal DNA repair processes in order to prevent counterproductive events from occurring simultaneously. This is achieved by inactivating at least two key proteins involved in the homeostatic maintenance of genomic integrity; *PARP* (poly(ADP-ribose) polymerase), and *DNA-PK* [170]. Caspase 3 is responsible for cleaving the above two proteins as well as *ATM*; the gene mutated in Ataxia Telangiectasia, *RAD51*; the recombination homolog, and *BLM*; the gene defective in Bloom's Syndrome [171].

Exploring the possibility of the deregulation of these downstream apoptotic genes in MSI-H tumourigenesis requires scanning the coding sequences for hypermutable repetitive sequences. The catalytic subunit of *DNA-PK* contains two such repeats, an $(A)_{10}$ and an $(A)_8$, and *BLM*, an $(A)_9$. *ATM*, *DNA-PK*, and *ATR* (Ataxia Telangiectasia and Rad3-related gene), that contains an $(A)_9$ tract in its protein encoding sequence all belong to the PI3-kinase family. Members of this family mediate phosphorylation of protein substrates, and are implicated in DNA damage sensing pathways, DNA repair, and cell cycle control [172]{FLAGGS1997}. With the exception of *BLM* that has been found to be mutated in 2 of 11 MSI-H gastrointestinal tumours, *DNA-PK* and *ATR* have not yet been implicated in MSI-H cancers.

Early in the initiation of this project, *BCL10* (B-cell CLL/lymphoma 10) was cloned and described as a cancer-associated gene with pro-apoptotic activity. *BCL10* encodes a protein that possesses at least two important functional domains, an amino-terminal CARD involved in NF-κB activation, and a carboxy-terminal non-CARD domain involved in activation of caspase-9 [173-177]. Ruland *et al.*, (2001) showed that one-third of Bcl10 -/- mouse embryos developed exencephaly, leading to embryonic lethality. The surviving Bcl10 -/- mice were severely immunodeficient and interestingly, $Bcl10^{-/-}$ cells retained susceptibility to various apoptotic stimuli in vivo and in vitro [178]. While wildtype *BCL10* suppresses transformation, mutant forms of *BCL10* identified in gastric MALT lymphomas lose this property, and often acquire gain-of-function transforming activity [173,177]. In addition to its involvement in MALT lymphomas, *BCL10* was also reported to be mutated in a variety of human neoplasms, including colorectal carcinoma [177,179,180]. Prompted by the identification of a single base frameshift mutation in the $(A)_8$ in a cell line with known mismatch repair deficiency [177], we decided to explore the possibility of selective mutations in the two repetitive tracts, $(A)_8$ and $(T)_7$ in *BCL10*.

Another reported apoptotic regulatory gene with mutations in coding repetitive tracts is *caspase 5* [166], which is mutated in 62% of MSI-H CRCs. This is an extremely high mutation frequency that has not been investigated by other groups as of yet.

Mutations may involve a host of other genes related to apoptosis regulation. The presence of hypermutable repetitive DNA sequences in these genes makes them highly susceptible to mutation in MSI-H CRCs. We screened several of these apoptosis regulatory genes (*BAX, DNAPK, ATR, APAF-1, BLM* and *BCL10*) for frameshift mutations present at their polyadenine repeats. We have found that the need to deregulate the apoptotic pathway in MSI-H tumours is substantiated by somatic frameshift mutation of at least one of the apoptosis regulatory genes in 65% of our panel of young MSI-H CRC patients.

2.2. Material and Methods

2.2.1. Tissue Samples

All of the DNA samples used in this project were obtained from patients (<50 years of age) with resected colorectal carcinomas that were identified through the Ontario Cancer Registry as part of an ongoing population-based study [181]. Paraffin embedded tissue samples were obtained from the respective treating hospitals. A blinded histopathological review of each case was performed to locate regions of high neoplastic cellularity (>50%). Tissue from these regions was microdissected and DNA was extracted as described previously [182]. Briefly, tissue was scraped from 2-3 unstained 10-μm slides into 50-100 μl of lysis buffer [10 mM Tris-Cl, pH 7, 100 mM KCl, 2.5 mM $MgCl_2$ and 0.45% Tween 20]. Following a 10-minute incubation at 95°C, the tissue samples were subjected to proteinase K (20 mg/ml, 15-35 μl) digestion overnight at 65°C.

2.2.2 Cell Lines

Initially this project started with a total of 13 colorectal carcinoma cell lines, which were obtained from the American Type Culture Collection (Manassas, VA), including 6 MSI-H cell lines (SW48, LS174, LS411, LoVo, HCT-116, and DLD-1) and 7 microsatellite stable (MSS) cell lines (HT-29, SW480, SW620, SW837, Colo320HSR, LS513 and LS1034). In addition, one MSI-H endometrial cell line (HEC1A) was included in our panel. As the course of the project progressed, additional cell lines were added to our panel. HCT-8, an MSI-H colorectal cell line and several MSS cell lines (SW403, SW948, SW1116, SW1417, LS123, LS1034, SKCO-1, Colo201, CaCO-2, and T84). DNA was extracted from the cell lines using DNeasy Tissue Kit (Qiagen, Mississauga, ON), according to the manufacturer's instructions.

2.2.3 Microsatellite Instability Testing.

Microsatellite instability (MSI) was tested in the primary colorectal carcinomas by PCR of at least 5 of the reference panel loci outlined in the National Cancer Institute Workshop on Microsatellite Instability [183]. Colorectal cancers were designated as MSI-H if two or more of the mononucleotide/dinucleotide repeat markers had MSI, and MSS if none displayed MSI. If only one of the markers revealed MSI, then up to five additional reference panel loci were tested to determine if the case was MSI-H (≥40% of loci having MSI) or had low frequency microsatellite instability (MSI-L; ≤30 % of loci having MSI). The loci used in our study were BAT-25, BAT-26, D2S123, D5S346, D17S250, BAT-40, BAT-RII, D18S58, D18S69, and D17S787, with PCR conditions as described previously [182,183]. The overall results of the MSI testing in the series of primary colorectal cancers have been published previously [181].

2.2.4 BAT-26 Analysis of Cell Lines

The MSI status of the colorectal carcinoma cell lines was confirmed by PCR analysis of the BAT26 locus due to the high sensitivity of this marker. Primers for BAT-26 have been previously published and are as follows: Forward, 5'-TGACTACTTTTGACTTCAGCC-3', and Reverse,5'-AACCATTCAACATTTTTAACCC-3'. 1μl of genomic DNA extracted from the cell lines was combined with 10x PCR buffer, 4.5mM $MgCl_2$, 0.4 mM deoxynucleotide triphosphates, 0.6 μM of each the forward and reverse primers with 1 unit of Taq Polymerase (Life Technologies, Gibco BRL, Burlington, ON) in a 15 μl PCR volume. PCR cycling conditions were 2 minutes at 94°C, followed by 35 cycles of 15 seconds at 94°C, 15 seconds at the annealing temperature of 56°C followed by 20 seconds at 72°C (DNA Engine, model PTC-200; MJ Research, Watertown, MA). This marker is readily analyzed in high-resolution agarose gels (SuperGels, Helixx Technologies, Inc. Toronto, ON) run in 2x TAE buffer. 5 μl of 6x loading dye was added to the PCR products, and these were run on the SuperGels for 65 minutes at 200 Volts. The results were visualized by staining with ethidium bromide for 5 minutes, followed by ultraviolet light illumination.

2.2.5 *APAF-1, BLM, DNA-PK,* and *ATR* Mutation Screening

Primers were designed to amplify the region containing the hypermutable tract so as to produce a PCR product <150 base pairs (see Appendix A for primer sequences). A shorter sequence is desirable in order to increase efficiency of amplifying DNA from archival tissues embedded in paraffin blocks. The reverse primer was end-labelled in a final volume of 10μl; 0.3μM of the reverse primers was combined with 60μCi of [γ- ^{33}P]ATP (Easytides®, NEN - US, Boston, MA) and 5.88 units of FPLC*pure*™ Polynucleotide Kinase (Amersham Pharmacia Biotech, Baie d'Urfe, Quebec). The reaction was incubated at 37°C for 1 hour and denatured at 90°C for 2 minutes. In a 15 μl PCR reaction, 2 μl of genomic DNA from primary colorectal carcinomas or 1

μl of DNA from the cell lines was combined with 10x PCR buffer, 1.5 mM $MgCl_2$, 0.13 mM deoxynucleotide triphosphates, 0.4 μM of each forward and reverse primer and 1 unit of Platinum Taq polymerase (Life Technologies, Gibco BRL, Burlington, ON). PCR cycling conditions were 2 minutes at 94°C, followed by 35 cycles of 15 seconds at 94°C, 15 seconds at the respective annealing temperatures (see Appendix B for specific temperatures) and 20 seconds at 72°C (DNA Engine, model PTC-200; MJ Research, Watertown, MA). After PCR, 7.5 μl of denaturing formamide dye was added to each tube. Samples were physically denatured by heat at 94°C for 4 minutes and immediately cooled on ice prior to loading on a denaturing 7% polyacrylamide gel. The gel was transferred onto 3mm-Whatman paper, dried and exposed to Kodak Biomax film (Rochester, NY).

2.2.6 *BCL10* Mutation Screening

From the original series of 607 colorectal cancer patients, a panel of 322 colorectal carcinomas was selected for study, including all 102 MSI-H cancers, all 20 MSI-L cancers, and 200 randomly selected consecutive MSS cancers. To amplify the repeated mononucleotides, specific primers were designed as follows: for the $(A)_8$, P2F, 5'-CTGAGAGACATTTTGATCATCTACG-3' and P2R, 5'-GGGCTGGAAAATTGTTAGACTACTT-3'; for the $(T)_7$, P3F, 5'-TGAAAAACTGAGGGCATCCACTGT-3' and P3R, 5'-CTTCTCTGAATTTGCCTGTTCTAG-3'. The PCR conditions are the same as above, and the specific annealing temperatures are 55°C and 57°C (for the $(A)_8$ and $(T)_7$ primer sets, respectively).

To screen for other coding region mutations, 7 sets of primers were used to amplify and sequence the entire *BCL10* gene, including all intron/exon junctions. The primer sequences for

exon 1 are published [177]. The 6 remaining primer pairs were designed to produce overlapping fragments and are as follows: P2.1F, 5'-AAGACTGCCAACTAATAGTCACG-3' and P2.1R, 5'-AGATGATCAAAATGTCTCTCAGC-3'; P2.2F, 5'-TACGTGTATACCTGTGTGAG-3' and P2.2R, 5'-TGGGTTTTCCTGTAAGTAGTC-3'; P2.3F, 5'-TCAAGTAGAAAAAGGGCTGG-3' and P2.3R, 5'-GCATTATTACATTTAAATTAGCTC-3'; P3.1F, ACAAGTCACAAGATGGACAGTG-3'and P3.1R, AGTGGATGCCCTCAGTTTTTCAG-3'; P3.2F, TCTCCAGATCAAATTCAGATGAG-3'and P3.2R, 5'-CCCAGGTCTGGGAAGTGTAGT-3'; P3.3F, 5'-GTTCTAGAAGTAGGCAGAACTG-3' and P3.3R, 5'-GTCATTGTCGTGAAACAGTACG-3'. All sequenced fragments were less than 200 base pairs in length.

2.2.7 Putative Mutant Sequencing

All tumours with putative mutations in a specific gene (*APAF-1, BLM, DNAPK, ATR*) were amplified by PCR as described above, with the exception of radiolabeled reverse primer. These PCR products were run on 2% agarose gels to achieve optimum separation of the actual product from non-specific bands. The products were then gel purified using the Concert™ Rapid PCR Purification System (Life Technologies, Gibco BRL, Burlington, ON, Cat. No. 11458-0213). The sequencing reaction was performed using the reverse primer with the Thermo Sequenase radiolabeled terminator cycle sequencing kit (Amersham Pharmacia Biotech, Cleveland, OH) according to the manufacturer's instructions.

2.3 Results

2.3.1 BAT26 Analysis of Cell lines

The results of the BAT-26 mutational analysis concur with the published data on the MSI status of the cell lines in our panel (Table 2.3.1-4). Shifts due to instability are easily visualized as seen

in Figure 2.3.1-6. The positive results for BAT-26 are, SW48, HEC1A (endometrial), HCT116, DLD-1, LS174, and LS411. No product was produced for the known MSI-H cell line, LoVo. This was expected since this cell line is known to be homozygously deleted for exons 4 through to 8 of *hMSH2*. BAT-26, is located within 5th intron of *hMSH2* explains the absence of amplification of the BAT-26 locus.

2.3.2 *BAX, DNAPK, BLM, ATR,* and *APAF-1*

BAX mutation analysis was performed in our laboratory prior to the initiation of this project, and thus only 46 of the 102 MSI-H CRCs were screened for mutations at the $(G)_8$ mononucleotide repeat. None of the cell lines were available at the time of this mutation analysis. For all of these tumours, the corresponding normal tissues were screened as well to investigate the possibility of germline *BAX* mutations in these patients. In total, 20 mutations were observed, 12 of which were one base pair (bp) deletions, 3 2bp-deletions, and 5 1bp-insertions in 18 primary CRCs. Thus, *BAX* is mutated in 18 of 46 (39%) MSI-H CRCs. These frameshift mutations are all up-stream of the Bcl-2 domain in *BAX* (as seen in Figure 2.3.2-7).

The two members of the PI3 kinase family, *DNA-PK* and *ATR* were screened for mutations at polyadenine tracts in our complete panel of 102 MSI-H CRCs. *DNA-PKcs* (catalytic subunit) has two repeats; an $(A)_{10}$ repeat at position 487, and an $(A)_8$ at 10807 (Figure 2.3.2-8). The *ATR* gene also contains an $(A)_{10}$ repeat at its amino-terminal, at position 2311 (Figure 2.3.2-8). 33 of 99 (33%) MSI-H CRCs screened has mutations at the $(A)_{10}$ tract in *DNA-PK* and only 8 out of 98 (8%) that had mutations in the $(A)_8$ repeat. A total of 41 alterations were observed in 39 MSI-H CRCs. Of these, 35 were single base deletions and 6 were single base insertions. Only one tumour (R4) had mutations in both polyadenine tracts and one tumour (HC200) showed evidence of bi-allelic inactivation at the $(A)_{10}$ repeat and was confirmed through manual sequencing. 2 of the 8 (25%) MSI-H cell lines, HEC1A and LS411 displayed 1-bp deletions.

Mutations in the *ATR* gene were found at a slightly lower mutation frequency, 29% (28 of 97) of MSI-H CRCs exhibited frameshifts at the $(A)_{10}$ run. Of these 28 CRCs, 13 (46%) also had mutations in the other PI3-kinase member, *DNA-PK*. The majority of the mutations were single base deletions (27) with only one 1-bp insertion. The same two MSI-H cell lines that harboured the mutations at the $(A)_{10}$ repeat in *DNA-PK* also contained mutations at the *ATR* $(A)_{10}$ repeat. Biallelic inactivation was observed in 2 tumours (HC172, HC323). Biallelic inactivation is usually difficult to determine due to the heterogeneity of the archival DNA sample. Although areas of high neoplastic cellularity were microdissected, normal cells inevitably contaminated the extraction. Consequently, bi-allelic inactivations may go undetected in mutational analysis of these primary tumours.

The *BLM* helicase gene exhibited 21 single base deletions and 1 insertion at its $(A)_9$ repetitive tract, for a total mutation frequency of 23% (22 of 97). 2 MSI-H cell lines also exhibited the 512delA deletion, the mutated cell lines were, LoVo and LS411. At a much lower frequency, *APAF-1* exhibited 4, 1-bp deletions at its $(A)_8$ repeat out of 90 MSI-H CRCs screened. No mutations were found in the 21 cell lines, 8 of which displayed MSI.

All these mutations were confirmed through manual sequencing. The mutation frequencies for the apoptotic regulatory genes examined are represented in Figure 2.3.2-9 and are also tabulated in Table 2.3.2-5.

Table 2.3.1-4. *Tabulation of mutational analysis of the cell lines. BAT-26 analysis was used to confirm microsatellite instability status. A positive result is recorded when a shift (expansion) is greater than 3 base pairs in size.*

Cell line	MMR Deficiency	MSI Status	BAT26	*DNA-PK*	*ATR*	*BLM*	*APAF-1*
LoVo	*hMSH2and hMSH6*	+	NA	-	-	+	-
SW48	*hMLH1*-hypermethlyation	+	+	-	-	-	-
LS174		+	+	-	-	-	-
LS411		+	+	+	+	-	-
HCT-116	*hMLH1*	+	+	-	-	-	-
DLD-1	HMSH6	+	+	-	-	-	-
HCT-8	hMSH6	+	+				-
HEC1A	*hPMS2* and *hMSH6*	+	+	+	+	-	-
HT-29			-	-	-	-	-
LS123			-				
LS513			-	-	-	-	-
LS1034			-	-	-	-	-
SW403			-				
SW480			-	-	-	-	-
SW620			-	-	-	-	-
SW837			-	-	DNP*	-	-
SW948			-	-	-	-	-
SW1116			-				
SW1417			-				-
SW1463			-				
COLO201			-				-
COLO320HSR			-	-	-	-	-
CACO2			-				-
T84			-				-
SKCO-1			-				-

* DNP- Did not amplify by PCR

Figure 2.3.1-6 *BAT-26 analysis of the initial 16 cell lines. These were run on a 6% denaturing polyacrylamide gel for 2 hours at 60 watts. The reverse Bat-26 primer was radiolabelled with ^{33}P using the same method as described above for mutational analysis of the various genes. The PCR products were run in the following order: Lane 1, SW48;Lane 2, SW480; Lane 3, SW620; Lane 4, SW837; Lane 5, SW948; Lane 6, SW1417; Lane 7, HEC1A; Lane 8, LoVo; Lane 9, Colo320HSR; Lane 10, HT-29; Lane 11, HCT-116; Lane 12, DLD-1; Lane 13, LS174; Lane 14, LS411; Lane 15, LS513; Lane 16, LS1034. Lanes 1, 7, 11, 12, 13, and 14 exhibit positive (+) results for instability at this locus. LoVo, which was run on Lane 8, did not produce a PCR product.*

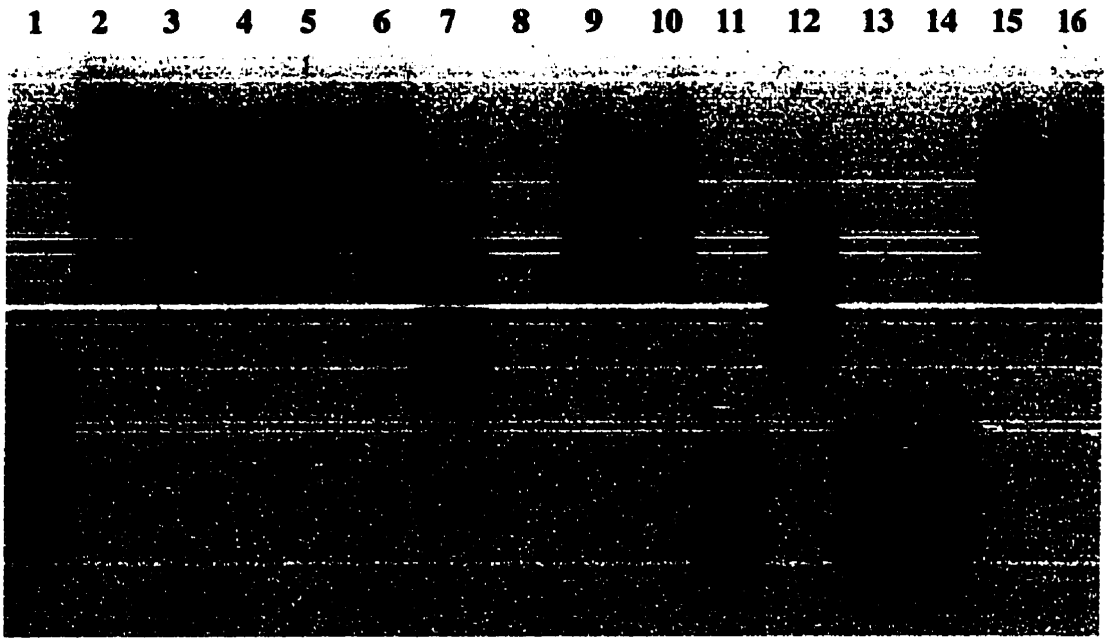


Figure 2.3.2-7 *Schematic diagrams of the BAX, BLM and APAF-1 genes. Functional domains are noted at their positions within the coding sequence of the genes. The location of the mononucleotide repeats screened within these genes are also noted in relation to functional domains.*

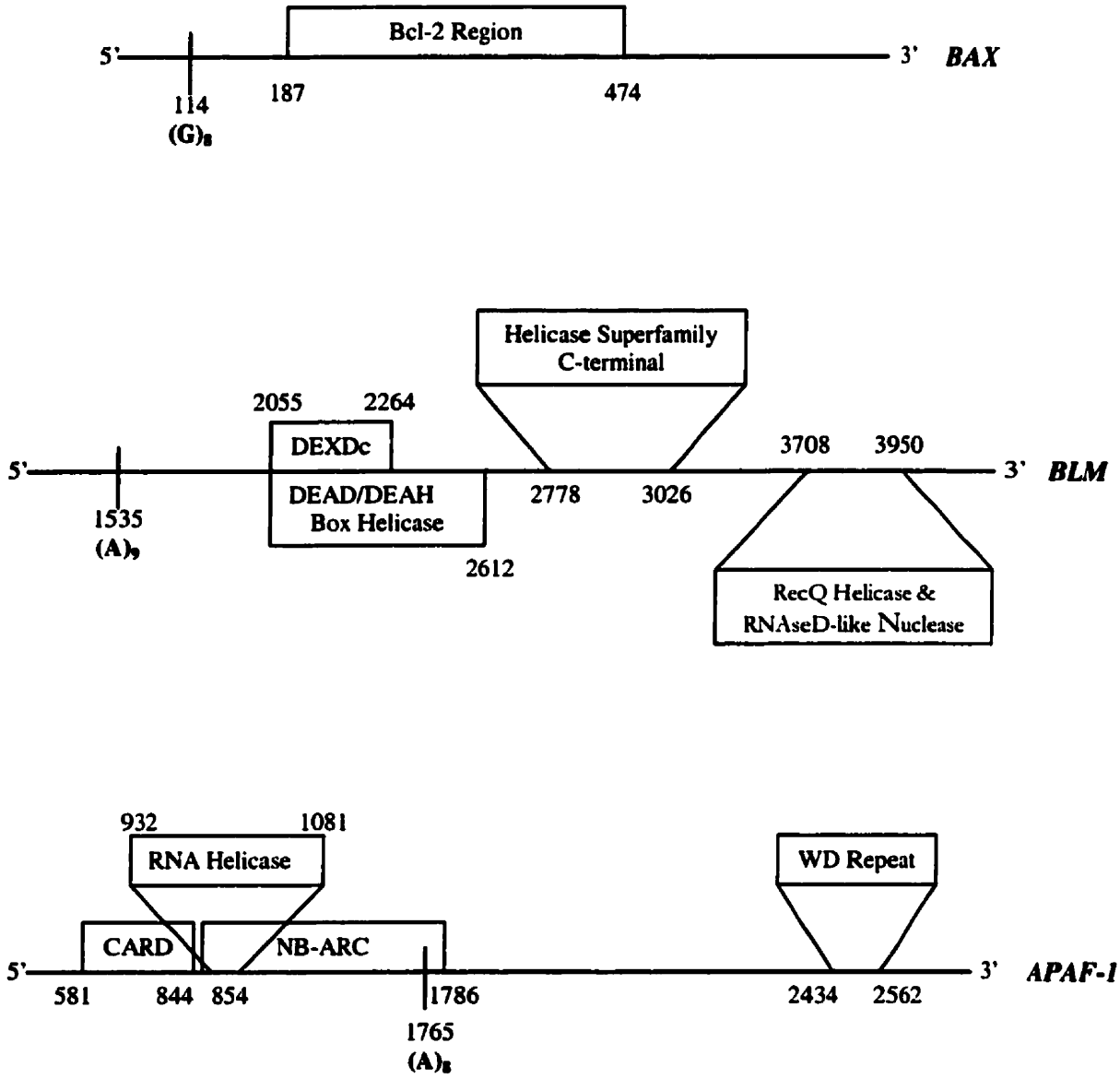


Figure 2.3.2-8. *Schematic diagram of the PI3-kinase family genes, DNA-PK and ATR. These two genes share four functional domains, FAT, PI3-PI4 kinase, the PI3-Kinase catalytic regions, and the FATC domain. In both genes the $(A)_{10}$ repeat is upstream from the domains and frameshift mutations at this hotspot would produce a truncated protein product without any functional domains. Mutations at the DNA-PK $(A)_8$ tract would produce a product with the FAT domain only.*

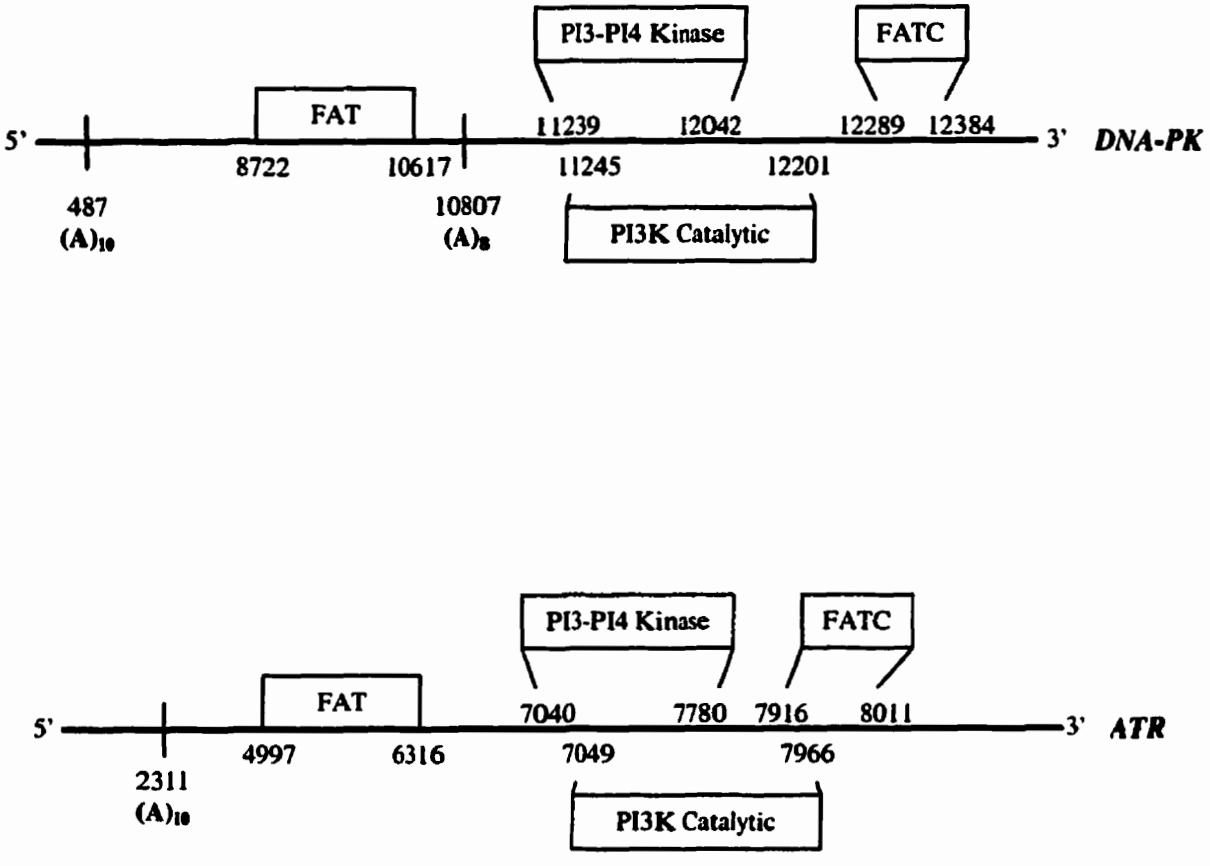
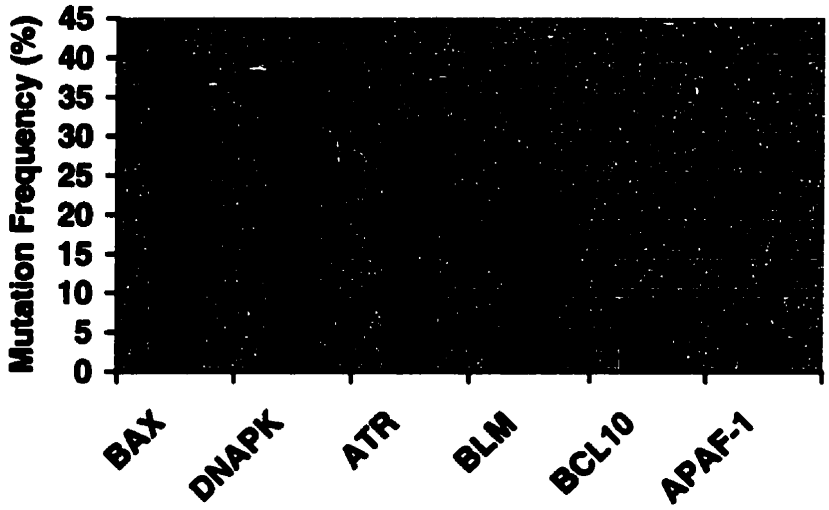


Table 2.3.2-5. *Categorization of the somatic frameshift mutations in the various CDS mononucleotide repeats of the apoptotic regulatory genes, BAX, DNA-PK, ATR, BLM, and APAF-1 in primary colorectal carcinomas.*

Gene	Repeat	Deletions	Insertions	Number of mutated CRCs	Number Tested	Mutation Frequency (%)
BAX	$(G)_8$	15	5	18	46	39
DNA-PK	$(A)_{10}$	31	2	33	99	
	$(A)_8$	4	4	8	98	
				39	**100**	**39**
ATR	$(A)_{10}$	27	1	28	97	29
BLM	$(A)_9$	21	1	22	97	23
APAF-1	$(A)_8$	4	0	4	90	4.4

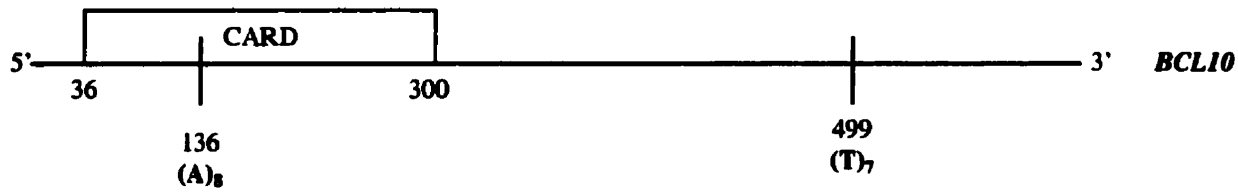
Figure 2.3.2-9. *Mutation Frequencies of the apoptotic regulatory genes at their respective mononucleotide repeats.*



2.3.3 *BCL10*

A panel of primary colorectal carcinomas and colorectal carcinoma cell lines of known microsatellite instability status were screened by PCR in order to determine the frequency of frameshift mutations at repeated sequences in the coding region of *BCL10*. In total, 10 of the 108 MSI-H (9%) colorectal cancers and MSI-H cell lines had band shift alterations in one of the two mononucleotide sequences in *BCL10* (Figure 2.3.3-10). This instability was more frequent at the $(A)_8$ mononucleotide tract (8/108) than at the $(T)_7$ tract (2/108) (Table 2.3.3-6). In contrast, no alterations were identified in any of the 227 MSS primary colorectal carcinomas or MSS cell lines tested ($p<0.001$; Fisher's Exact Test).

Sequencing the mononucleotide tracts in the cases with band shifts revealed that single base frameshift deletion mutations accounted for the majority (9/10) of alterations (Table 2.3.3-7). In the 9 MSI-H primary colorectal carcinomas with mutations, 7 had single base deletions within the polyadenine $(A)_8$ tract in exon 2 (136delA), and 2 had deletions of a thymidine within the run of polythymidines $(T)_7$ in the third coding exon (499delT). The final alteration was in one of the 6 MSI-H cell lines, LoVo, which displayed an insertion mutation as published previously (136insA)[177,186].



To determine whether *BCL10* mutations follow the two-hit inactivation paradigm of other human tumor suppressor genes, we searched for the presence of bi-allelic inactivation during tumorigenesis. In the colorectal carcinomas with *BCL10* mutations, there were no tumours with mutations in both the $(A)_8$ and $(T)_7$ tracts. In 8 of the 9 primary colorectal carcinomas the neoplastic cellularity of the microdissected sample was sufficient (at least 70%) to determine whether there was loss of heterozygosity of the *BCL10* locus. In all 8 of these cases, as well as in LoVo, the mutated allele and the wild type allele were present in approximately equal proportions, indicating that there was no allelic deletion of wild type *BCL10*. In order to identify other intragenic mutations, and further evaluate the possibility of a "second hit" alteration in *BCL10*, we sequenced the complete coding region in all of the cases with frameshift alterations. Three alterations were detected (Table 2.3.3-7), and all were found to correspond to previously published polymorphisms [187,173].

Table 2.3.3-6. *Frequency of BCL10 alterations in primary colorectal carcinomas and cell lines*[1].

Microsatellite Instability Status	Number Tested	$(A)_8$ Alterations	$(T)_7$ Alterations	Total Alterations
Primary colorectal carcinoma:				
MSS	200	0	0	0
MSI-L	20	0	0	0
MSI-H	102	7	2	9[2]
Colorectal carcinoma cell lines:				
MSS	7	0	0	0
MSI-H	6	1	0	1

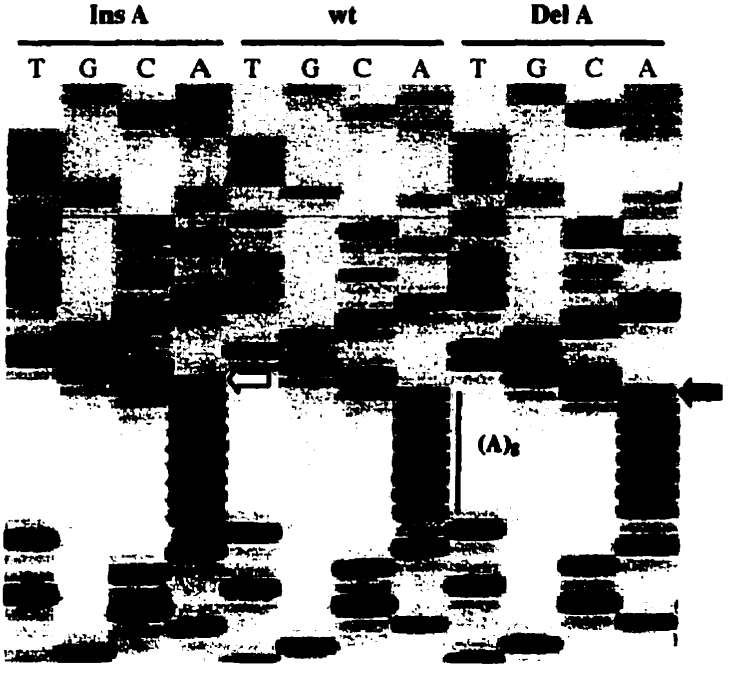
[1]MSS, microsatellite stable; MSI-L, low frequency MSI; MSI-H, high frequency MSI.
[2]MSS/MSI-L *versus* MSI-H, $P<0.001$.

Table 2.3.3-7. *BCL10 sequence alterations in primary colorectal carcinomas and cell lines.*

Exon	Codon	Nucleotide Alteration[3]	Number of Carcinomas	Effect on BCL10
Mononucleotide tract mutations[1]:				
2	46	136delA	7	Truncation, predicted length 68 amino acids
2	46	136insA	1	Truncation, predicted length 48 amino acids
3	167	499delT	2	Truncation, predicted length 171 amino acids
Other coding region sequence variants[2]:				
1	8	24 G→C	1	Silent polymorphism
3	213	638 G→A	1	Silent polymorphism
Intronic variants[2]:				
Intron 1		+11 C→G	3	Silent polymorphism

[1]All 323 primary colorectal carcinomas and 13 cell lines were tested.
[2]Only the 9 carcinomas and 1 cell line with frameshift mutations were tested.
[3]Nucleotide numbering begins at the *BCL10* initiation codon, RefSeq accession number NM_003921.

Figure 2.3.3-10 *Sequencing confirmation of BCL10 mutations. Reverse primer sequence of exon 2, containing the A(8) mononucleotide repeat. Two mutations are shown: left, an insertion of an adenine (136insA) in exon 2 (Ins A, cell line LoVo), and right, a deletion of an adenine (136delA) in exon 3 (Del A). Mutations are indicated with arrows. Wild-type sequence is also shown (wt).*



2.4 Discussion

2.4.1 General Discussion

Normal cells will self-initiate programmed cell death in response to adverse environmental, and intrinsic conditions. This process, also known as apoptosis, is a cellular protective mechanism to control unfavourable cell proliferation. During the development of a malignancy, cells lose this ability to commit suicide when their integrity is compromised. The activation of the apoptotic pathway involves the inactivation of specific proteins important in cell proliferation, genome stability and the activation of downstream targets of apoptosis. *TP53* has been dubbed the guardian of the genome, attributed by its role as a tumour suppressor gene (gatekeeper) in controlling the cell cycle. p53 can arrest the cell cycle in order for damaged DNA to be repaired, and if the damage incurred is irreparable, p53 drives the cell down the apoptotic pathway [188,189]. Most cancers carry *TP53* mutations, which are responsible for the uncontrolled proliferation of tumours. Interestingly, tumours that display microsatellite instability have fewer *TP53* mutations than their chromosomal instability counterparts [184]. The reason for this distinction may be dependent on the fact that *TP53* has no mononucleotide CDS repeats longer than 5 bases. The identification of other substitutes (apoptosis regulatory genes) for the loss of *TP53* with longer mononucleotide repeats may be the critical targets for apoptosis in MSI-H tumours. One such example would be *APAF-1*. Caspase 9 and its cofactor Apaf-1 can be essential downstream effectors of the p53 during apoptosis, such that inactivation of either *APAF-1* or *CASPASE 9* substitutes for the loss of *TP53* [119].

2.4.2 *APAF-1*

APAF-1 is a 130-kD protein that participates in the cytochrome c-dependent activation of caspase-3 [190]. The amino-terminal houses the CARD domain, while the COOH-terminal region

of *APAF-1* comprises multiple WD repeats, which are proposed to mediate protein-protein interactions. WD-repeat-containing proteins are those that contain 4 or more copies of the WD-repeat (tryptophan-aspartate repeat), a sequence motif approximately 31 amino acids long, that encodes a structural repeat. Deletion of the WD40 repeats (2434-2562) makes *APAF-1* constitutively active and gives it the ability to process procaspase-9 independent of cytochrome c and dATP.

Caspases play an integral role as initiators and effectors of the apoptotic signal. Caspases are cysteine proteases that have an unusual and absolute requirement for cleavage after asparatic acid. Four amino acids amino terminal to the cleavage site are also recognized and are required for efficient catalysis [169]. The Death domain, Death Effector Domain (DED), and Caspase Recruitment Domains (CARD) are all involved in protein-protein interactions involved in the caspase arm of the apoptotic pathway [191]. There are two known caspase activation pathways in the apoptotic cascade. Activation of procaspase-8 requires association with its cofactor FADD (Fas-associated protein with death domain) through the DED and procaspase-9 activation, that involves a complex with the cofactor Apaf-1, through the CARD.

Given the importance of Apaf-1 in the activation of apoptosis, the tract of nine adenines starting at position 1765 is a likely target for alteration in MSI-H tumours. The PCR amplification for the $(A)_9$ in *APAF-1* was not as successful compared to the other genes. Only 90 of the 102 produced informative results. A surprisingly low mutation frequency of 4% (4/90) was detected as 1-bp deletions. Soengas *et al.* (2001) proposed an alternative method for the inactivation of *APAF-1* earlier this year. These researchers showed that metastatic melanomas often lose *APAF-1*. *APAF-1* -negative melanomas were chemoresistant and were unable to execute a typical apoptotic program in response to p53 activation. Restoring physiologic levels of Apaf-1 through gene transfer or 5aza2dC treatment markedly enhanced chemosensitivity and rescued the apoptotic defects associated with *APAF-1* loss. Soengas *et al.* (2001) concluded that

APAF-1 is inactivated in metastatic melanomas, leading to defects in the execution of apoptotic cell death. Interestingly, *TP53* is rarely mutated in melanomas, an event also seen in MSI-H cancers. Thus, the most commonly targeted gene in the deregulation of apoptosis (*TP53*) is left unscathed in these particular cancers that have decided to target other apoptotic regulatory genes. Perhaps MSI-H cancers like that of melanomas have inactivated the *APAF-1* not by a frameshift mutation in the repeat tract but by methylation.

Although the promoter region of *APAF-1* contains CpG islands, these methylation targets are not per se altered in this gene. The methylation is inferred to occur at an enhancer or insulator region. *APAF-1* joins a growing list of cancer genes that can be inactivated by epigenetic mechanisms.

2.4.3 *DNA-PK*

DNA-dependent protein kinase (DNA-PK) is a nuclear protein serine/threonine kinase present in a wide variety of eukaryotic species. One of its striking features is that it must be bound to DNA to express its catalytic properties. It can be fractionated into 2 components: a large polypeptide corresponding to the catalytic subunit, and an autoimmune antigen, Ku. On its own, the catalytic subunit of DNA-PK is inactive and relies on the other DNA-PK component to direct it to the DNA and trigger its kinase activity. One physiologic function for DNA-PK may be to modulate transcription, since it has been shown to phosphorylate several transcription factors in vitro [192]. Other functions described for DNA-PK include, the mediation of DNA double strand break repair and V(D)J recombination. In mice, mutations produce the SCID phenotype along with a predisposition to T-cell lymphomas (suggesting a tumour suppressor role in normal cells)[193,194]. DNA-PK has additionally been proposed as a DNA damage sensor along with poly (ADP-ribose) polymerase (PARP), due to its ability to bind and be activated by DNA strand break. The role of DNA-PK in apoptosis is indicated by the targeted cleavage of a DEVD site in the

catalytic subunit to ensure destabilization of the genome as the cell prepares for death. Furthermore DNA-PK phosphorylates p53 at ser15 and ser37, and is implicated as a necessary upstream regulator of p53 DNA-damage response [195]. This phosphorylation impairs the ability of the p53 antagonist, MDM2 from binding leading to inhibition of p53-dependent transactivation. Jimenez *et al.*, (1999) found conflicting results, they demonstrated that the p53 response is fully functional in primary mouse embryonic fibroblasts lacking DNA-PK [196]. Regardless, *DNA-PK* is an appealing target for mutagenesis in MSI-H cancer progression. *DNA-PK* has two CDS repeats; an $(A)_{10}$ repeat beginning at codon 162, and an $(A)_8$ repeat at codon 3601 within the PI3-K homology region of the catalytic subunit. The $(A)_{10}$ is at the amino terminal of the gene at position 487 proximal to all of the functional domains (Figure 2.3.2-8). We have identified *DNA-PK* mutations in 39/100 (39%) of our primary MSI-H CRCs. 33/41 (80.5%) of the total alterations observed occurred at the $(A)_{10}$ hotspot. This biased mutation frequency at the longer repeat tract is to be expected, since increased replication slippages occur at longer lengths of mononucleotide runs. While the majority of mutations at the $(A)_{10}$ tract are 1-bp deletions (31/33), mutations at the $(A)_8$ tract is split equally between insertions and deletions (4 each). The $(A)_9$ and the $(A)_{11}$ mutants produced by frameshifts at the longer polyadenine repeat would result in truncated protein products of 170 and 175 amino acids, respectively. The $(A)_7$ and $(A)_9$ containing mutants created at the eight adenine tract would produce proteins, 3651 and 3606 amino acids in length. Regardless of the truncated protein size, the protein itself has not retained enough functional domains to control any of the crucial pathways.

One tumour (R4) possesses mutations in both of the polyadenine tracts. It is unknown whether this is evidence of biallelic inactivation or double mutations on a single allele. Another tumour (HC200) displayed biallelic inactivation at the $(A)_{10}$ mutation site and this was confirmed by the absence of the wild-type allele when this segment of *DNA-PK* was manually sequenced.

To our knowledge, these are the first *DNA-PK* mutations reported in human colorectal cancer with MSI-H. Recently, *DNA-PK* knockout mice were described as having dysplastic aberrant crypt foci (ACF) [197]. The significance of our findings suggests that DNA-PK may further destabilize the genome and promote apoptosis in a pathway independent of p53. This DNA-PK inactivation most probably prevents signalling from or repair of the degraded genomic DNA that is produced during the latter steps of apoptosis.

The second PI3-K family member that we screened was the Ataxia-Telangiectasia and Rad 3- related gene, *ATR*.

2.4.4 *ATR*

ATM and *ATR* are related and conserved proteins that are the central components of the DNA damage response [198]. *ATM* is the gene mutated in patients with Ataxia Telangiectasia and they are defective in several responses to IR including G1 arrest, reduction in DNA synthesis and G2 arrest [199,200]. Atm controls the initial phosphorylation of several key proteins such as p53, Mdm2, BRCA1, chk2, and Nbs1 in response to DNA damage. Atr also phosphorylates p53 at serine residue 20, which is important in stabilization of the p53 protein [201,202]. The expression of a dominant negative Atr sensitizes mammalian cells to all forms of DNA damage and diminishes the G2/M checkpoint response induced by γ-radiation [203]. Knockout mice have revealed interesting results concerning the functions of Atm and Atr. $ATM^{-/-}$ mice are viable and display growth retardation and infertility, while $ATR^{-/-}$ mice die early in embryogenesis. The $ATR^{-/-}$ blastocyst cells die in culture with a phenotype resembling mitotic catastrophe [204]. This suggests a role for Atr in possibly monitoring DNA replication. Atr has also been shown to control downstream DNA damage responses by phosphorylating checkpoint kinase 1 (Chk1), whereas Atm may target both checkpoint kinases 1 and 2 [205,206].

The loss of the *ATM* gene product results in increased genomic instability and is associated with an increase in cancer. It is likely that the loss of *ATR* would similarly destabilize the genome and lead to an increased risk of cancer. *ATR* has not been linked to a specific disorder and somatic mutations in the gene have not yet been reported. Our panel of 102 MSI-H CRCs was screened for mutations in the $(A)_{10}$ repeat at nucleotide position 2311. This repetitive element is at the amino terminal and is positioned upstream of all four of *ATR's* functional domains, FAT, PI3-PI4 kinase, PI3K catalytic region, and the FATC (Figure 2.3.2-8). Only 97 of the 102 primary tumours amplified with a mutation frequency of 29% (28/97). The resultant frameshift mutations lead to a truncated inactive Atr protein product, the $(A)_9$, would produce a 777 amino acid protein and the $(A)_{11}$, a 775 amino acid protein, considerably smaller than the full-length wild-type protein coded by 2645 amino acids. Following with *DNA-PK*, these results are the first to be described in the microsatellite instability progression of colorectal tumourigenesis.

New evidence has been proposed that since Atm, Atr and DNA-PK are all capable of phosphorylating p53 that they may signal different but partially overlapping types of DNA damage to a common p53 effector pathway [200].

2.4.5 *BLM*

Another proposed sensor of DNA damage is BRCA1, which is part of a large complex named BASC (BRCA1-associated genome surveillance complex) [207]. This complex is comprised of *ATM*, the *Nbs-Mre11-RAD50* complex; mismatch repair proteins (*hMSH2/6* and *hMLH1*), and the Bloom's helicase (*BLM*). This is a hypothetical model, where each of these proteins has the ability to recognize aberrant DNA structures and could relay the signal to Atm and BRCA1. The Bloom's syndrome gene (*BLM*) was shown to be cleaved at a consensus cleavage site, TEVD ($BLM^{412\text{-}415}$) by caspase 3 during apoptosis [21]. The kinetics of BLM apoptotic cleavage is highly

similar to that of DNA-PK, Atm and PARP. Cleavage and loss of localization very likely obliterates the in vivo function of Blm.

The *BLM* gene was the first identified gene responsible for the single-gene disease, the Bloom Syndrome. Patients with Bloom Syndrome are characterized by a high incidence of cancer and genomic instability. Mutations causing Bloom Syndrome delete or alter helicase motifs and may disable the 3'-5' helicase activity. The Bloom syndrome gene product is a 1417 amino acid, with 3'-5' DNA helicase and ATPase functions [171]. It is also a member of the RecQ family of helicases that have the DEXH box motif. RecQ is a member of the E.coli recF recombination pathway of genes in which mutations abolish the conjugational recombination proficiency and ultraviolet resistance of a mutant strain [208,209]. BLM co-localizes with RAD51 to nuclear matrix in discrete nuclear foci [171]. The potential function of BLM in DNA repair suggests that its cleavage and redistribution may aid nuclear disassembly and prevent the complex in which it resides from participating in the repair of fragmented DNA. Cleavage of another RECQ-helicase, WRN was not cleaved likely because it lacks consensus cleavage sites for caspases.

BLM helicase and hMLH1 may function cooperatively in maintaining genomic stability independent of the DNA MMR system [210]. BLM may also recognize loops of di- and tri-nucleotide repeats generated by DNA replication due to polymerase slippage or by single stranded annealing events occurring during gene conversion [211,212]. It was suggested that BLM might untangle these loops for MMR complexes to process them efficiently. Hence, the BLM is thought presently to function in the maintenance of the genome through its interactions with other proteins that participate in DNA replication and repair [209].

The presence of a mononucleotide repeat in the coding sequence of *BLM* has thus received attention from those of us who are in pursuit of genes with functional significance in MSI-H tumours. Calin *et al.* (1998) found somatic mutations at the $(A)_9$ sequence in 2 of 11 (18%) MSI-H CRCs [213]. It should be noted that the MSI status was not determined by

implementing the markers established by NCI. Our panel of MSI-H CRCs was also screened at *BLM* [512-515] for frameshift mutations. Twenty-one, 512delA mutations and one 512insA mutation were identified in 97 primary MSI-H CRCs (22/97, 23%). Two cell lines, LoVo and LS411 also carried the 1-bp deletion. The deletion mutations result in pre-mature termination of the *BLM* gene producing a 529 amino acid protein void of all functional domains (Figure 2.3.2-8). In addition, the insertion $(A)_{10}$ variant encodes for a 515 amino acid protein that lacks functionality. The cleavage of BLM during apoptosis is mediated by caspase 3, which also cleaves, APC, DNA-PK, and Atm [214]. Thus, other substrates of caspase 3 may already be targeted for cleavage before and in addition to BLM. The TEVD cleavage site is to the amino-terminal of the $(A)_9$ site at codon 512. Thus, if in MSI-H tumours, BLM is not preferentially cleaved over the other substrates, the frameshift mutations at the hotspot may account for the inactivation of *BLM* in this subset of tumours.

The presence of an excessive number of mutations in the coding and non-coding sequences of *BLM* patients [215,216], our findings and those of others [213], points to *BLM* as a potential link between the two pathways of genetic instability. The co-existance of both chromosomal and microsatellite instability pathways has only been described in cell lines, LoVo and V1394 [42].

2.4.6 *BAX*

The only gene previously described to have a significant role in apoptosis in MSI-H tumours is *BAX*. Homozygous (or hemizygous) frameshift insertion or deletion mutations in *BAX* were found in multiple primary colorectal cancers as well as colorectal cancer cell lines. The $(G)_8$ tract in *BAX* [38-41] has been found to include 1-bp insertions and 1 and 2-bp deletions in 21 of 41 (51%) of MSI-H CRCs [3]. Other studies have shown that the mutation frequency in *BAX* is between 32-62% [35,164,166,217]. The resulting frameshift was thought to interfere with the suppressor role of the

wildtype *BAX* gene. The search for mutations in this mononucleotide sequence revealed 20 insertions and deletions in 18 of 46 (39%) of our panel of MSI-H CRCs. The truncated protein products range from 58-72 amino acids in length.

Once the MSI-H phenotype is manifested, mutations at mononucleotide sequences in the genes described here would be more likely to occur than other frameshift or missense mutations in *TP53*. In cancers that progress down the MSI-H pathway, the generation of thousands of DNA mismatches during every replication of each MSI-H tumour cell may trigger the p53-mediated apoptotic response to DNA damage. But the response would be futile because the chain leading to apoptosis is broken in several downstream links. Therefore, we can speculate that frameshift mutations in these genes eliminate the selective pressure for *TP53* mutations during colorectal tumourigenesis.

2.4.7 *BCL10*

Of interest we screened *BCL10*, which was at the time a gene with putative functions in apoptosis. We screened for frameshift mutations at the $(A)_8$ and $(T)_7$ mononucleotide tracts of *BCL10* in a panel of primary colorectal carcinomas and colorectal carcinoma cell lines of known microsatellite instability status and found that mutations were relatively uncommon, and entirely limited to MSI-H tumours. The instability was more frequent at the $(A)_8$ mononucleotide tract than at the $(T)_7$ tract, which is consistent with the dependent relationship between tract length and relative instability observed in mismatch repair deficient cancers [1,5].

The predominance of single base deletion mutations is similar to the spectrum of mutations at other repeated sequences in MSI-H colorectal carcinomas [183,218]. While almost 25% of *BCL10* mutations reported previously are present in one of the two mononucleotide tracts [219,220], single base insertions predominate, and this spectrum of alterations is more consistent with the type of repetitive sequence frameshifts that occur in the presence of intact DNA

mismatch repair [221]. In addition, most of the mononucleotide tract mutations reported previously are invariably present in only a fraction of subclones [173,219,220], while the mutations we identified in MSI-H colorectal carcinomas were readily detected in bulk DNA, consistent with the presence of a clonal alteration in all neoplastic cells.

There was no evidence to suggest that any of the *BCL10* alterations identified in our study were bi-allelic. While most of the *BCL10* mutations described previously are also mono-allelic, there is some experimental evidence suggesting that these alterations could have functional importance in tumourigenesis [219]. The BCL10 protein is 233 amino acids in length and has two domains responsible for its NF-κB activation and pro-apoptotic function, the CARD and the non-CARD C-terminus [174,222-225]. Unlike the CARD found in APAF-1, the BCL10 CARD does not mediate binding to caspase-9 in the progression to cell death [222]. However, it does induce self-oligomerization, which is important for activation of NF-κB [223], a transcriptional regulator of cellular responses to stress, inflammation, injury, and apoptosis [214,224,225]. The non-CARD domain of the C-terminus engages with caspase-9, promoting autoproteolysis and zymogen activation [222], which mediates apoptosis by cleaving specific proteins involved in critical cellular processes and cellular structural integrity [169].

The mutations identified in our study predict two different types of BCL10 protein truncation, either within the CARD in exon 2, or distal to the CARD in exon 3 (Table 2.3.3-7). Both types of protein truncations are predicted to disrupt the normal function of BCL10 [220]. The 136delA and 136insA mutations would most likely be loss of function mutants where the activation of both NF-κB and caspase-9 are lost. The 499delT mutations occur distal to the CARD, and while they are predicted to lose caspase-9 activation, they retain NF-κB activation. This retained NF-κB activation may explain the transformation enhancement associated with some BCL10 mutants [219]. However, the predominance of mutations in the CARD domain in our

series does not support the notion that maintenance of NF-κB activation is functionally important in colorectal tumourigenesis.

Although *BCL10* was originally reported to be mutated in a variety of human neoplasms [219], several investigators have been unable to confirm the presence of mutations in non-hematological human cancers [179,226,227]. To explain these discrepancies, Willis *et al.* attributed their high *BCL10* mutation frequency to their use of cDNA rather than genomic DNA in mutation screening. This may have resulted in detection of aberrations caused by posttranscriptional modifications of RNA that were not necessarily templated in DNA [228]. The biological significance of aberrations detected by this approach remains uncertain, and similar *BCL10* alterations have been described in cDNA subclones from peripheral blood leukocytes [229]. Amongst the published mutation surveys in human tumours, two studies have failed to identify *BCL10* mutations in a total of 64 primary colorectal carcinomas [179,230]. Although the colorectal carcinomas in these latter studies are not classified by MSI status, it is likely that the number of MSI-H cases included in these series is too few to expect to identify *BCL10* mononucleotide tract mutations. In a larger study [231], no *BCL10* mutations were identified in a series of 132 primary colorectal carcinomas and 8 cell lines, including 49 that were known to have microsatellite instability. The reasons for the apparent discrepancy between these findings and our results are not clear. Although the details of case selection and the definitions for microsatellite instability are not given in Stone *et al.*, all of our cases were obtained from a population-based study and met international criteria for high frequency microsatellite instability. Furthermore, the mutation screening method we utilized was specifically designed to identify frameshift mutations within the mononucleotide tracts. Similar to our results, two recent studies have reported *BCL10* mononucleotide tract frameshift mutations in MSI-H colorectal carcinomas [167,180]. In combination with the findings of our study, the results of these other investigations support the contention that *BCL10* mutations are almost never identified in MSS

colorectal carcinomas. Aside from the frameshift mutations reported in MSI-H primary colorectal carcinomas, and the mutation reported in the MSI-H cell line LoVo, the only other colorectal carcinoma mutation reported to date is an intronic single base deletion predicted to yield a splice aberration in a cell line of unspecified MSI status [219].

The mutation frequency of the $(A)_8$ tract in *BCL10* (7%) might be considered relatively low in comparison with other MSI-H target genes that harbour eight mononucleotides. For instance, *BAX* $(G)_8$, *IGFIIR* $(G)_8$, *hMSH3* $(A)_8$, and *hMSH6* $(C)_8$ have been reported to be mutated in 26-43% of MSI-H carcinomas [35,77]. While these differences could be due in part to variability in the sequence-specific susceptibility to frameshift mutations in the presence of DNA mismatch repair deficiency, it is also possible that the biologic selection for *BCL10* mutations, is not as strong as it is for these other alterations.

BCL10 may be one of several pro-apoptotic genes (including *BAX*) that are specifically inactivated in MSI-H colorectal carcinomas via instability of coding sequence repetitive DNA tracts. Mutations in other genes such as *caspase 5* [166] and *E2F-4* [232,233] have been described but the significance and their roles in regulating apoptosis is unclear [234].

The escape of apoptosis may be facilitated by the accumulation of heterozygous mutations in multiple genes whose products play partially redundant and partially synergistic roles at different points in the apoptotic signalling network. Perucho and colleagues postulated an accumulative haploinsufficiency mechanism [3,153,165,166,166,167,235]. Due to their exacerbated mutator phenotype, these tumour cells elude not only the assertion that the occurrence of biallelic mutations is a very rare event in tumourigenesis [165,166] but also the premise that biallelism is a requirement for mutation functionality. Yet there is evidence of aberrant functionality of monoallelic inactivations, as seen by Yin *et al.* (1997) where heterozygous *BAX* knockout mice elicit an alteration in cell growth [236]. Most of the alterations described here are mono-allelic with the exception of a few tumours with *DNA-PK*, *ATR* and *BAX* biallelic inactivations. The

functional importance of mono-allelic gene inactivation in human tumours is difficult to determine. Although biallelic alterations provide strong support for the biologic importance of inactivating mutations in putative tumour suppressor genes, it is often difficult to document this occurrence in human cancers. With the exception of *TGFBRII* [1], biallelic inactivation is infrequently reported for many of the coding region repeated sequences mutated in colorectal carcinomas [77].

With the discovery of new apoptosis related genes and with the unravelling of other DNA repair systems, it will become clearer, what real functions the known genes perform in relation to the different genomic instability pathways. Chen *et al.*, (2000) proposed that the mismatch repair pathway as well as a second repair pathway might be involved in the response of mismatch repair-deficient cells [237]. Failure of the MMR system would activate the second intact repair system to initiate the apoptotic response. This is a viable hypothesis, one that will be supported or refuted in the near future.

Chapter Three

Kangaroo: A Bioinformatics Approach for the Systematic Retrieval of Genes with Microsatellites

Chapter Three

Kangaroo: A Bioinformatics approach for the systematic retrieval of genes with microsatellites

3.1 Background

Scanning the various databases containing human genome sequence information for a specific sequence of interest is exhausting and extremely time consuming. No available search programme enables the user to search for low complexity sequences such as short mononucleotide repeats exclusively in human annotated coding regions. The problem with using powerful programmes, for example BLAST is that it is based on a heuristic algorithm. In other words, it is based on trial and error. The given query sequence is divided into smaller "words", and the program then searches the database (NCBI GenBank) for records that contain these "words". When more of these "words" align with a given sequence, the resultant e-value decreases. The lower the e-value, the lower the probability of identifying the sequence by chance. When low complexity queries are entered, the program considers it background noise; there is merely not enough information for the programme to process. BLAST is extremely useful for finding homologies for specific sequences and one can try setting a very low e-value in order to retrieve records that are highly similar to the query sequence but this is inefficient use of a highly capable tool. Another problem with BLAST and programmes similar to it, is that the search limits cannot be specific for human annotated coding sequences. Intron/exon boundaries can be inferred using programmes such as GRAIL, Genie, and NetGene. Unfortunately, the accuracy of the defining boundaries has not been proven biochemically for human genes. Thus, a bioinformatics approach that encompasses these parameters is essential.

3.1.1 Kangaroo

A search algorithm, Kangaroo was developed at the Samuel Lunenfeld Research Institute. Kangaroo retrieves through SeqHound (in-house database, mirrored after NCBI GenBank) the biological information stored in NCBI GenBank Flatfiles. This biological information is stored in Abstract Syntax Notation 1 (ASN.1) from which annotations such as open reading frames (ORFs), sequence motifs, homologies, and restriction sites can be extracted. Our web-based application (http://bioinfo.mshri.on.ca), Kangaroo allows the user to enter a search query using Regular Expression patterns or simple text strings to search 1) Entire DNA 2) Amino Acids, or 3) Annotated Coding DNA. The query can incorporate more than one pattern using Regular Expression connectors and can be used to search 10 different organism genomes. Kangaroo has no limitations on the size and complexity of the sequence. It only reports exact matches to the query without the records with substitutions or similarities. Thus, Kangaroo accurately reports the occurrence of the query within the given organism genome. Each and every record is scanned exhaustively. When performing a search that retrieves many hits, even though the complete search may take 15-20 minutes, results are displayed on screen within seconds. The results are hyperlinked to its respective SeqHound record, which like the NCBI Entrez page contains links to related articles and lists the various annotations including the raw sequence. Using this multi-faceted program, genes containing specific sequences within their coding sequences can be identified.

3.1.2 Microsatellites

Microsatellites are short tandem repeats, which are simple in sequence (1-4 bp) and are found randomly distributed throughout the human genome [62]. Of the mononucleotide repeats, runs of A and T are very common, accounting for 0.3% of the nuclear genome. Runs of G and C are much more rare. While the majority of microsatellites are usually found in non-coding, intronic

regions, coding microsatellites do exist. Triplet repeats are not infrequent in coding DNA and are sites that are prone to pathologic expansions [238]. Abnormal expansions of the $(CAG)_n$ repeat in genes have been identified and are linked to diseases such as Huntington's, Kennedy, Spino-cerebellar ataxia (SCA1), Dentatorubralpallisoluysian atrophy (DRPLA), and Machado-Joseph disease (MJD, SCA3) [62].

In addition to triplet repeats, contraction and expansion of mono-, and dinucleotides often occur in tumours with microsatellite instability (MSI). In normal cells, mononucleotide repeats are targets for slippage during DNA replication. The frequency of DNA polymerase slippage events varies directly with mononucleotide run length [239-241]. If misalignments are not repaired by 3' exonuclease proofreading activity in normal cells, the DNA mismatch repair complex of genes will rescue the cell from the propagation of somatic mutations. Mutations that arise are the result of the insertion/deletion of one repeat unit in the majority of cases resulting in frameshift mutations. Thus, genes with mononucleotide repetitive elements are susceptible to mutation in MSI tumourigenesis. We were interested in determining exactly how many human genes contained coding mononucleotide tracts of 6 base pairs and greater. We suspect that since slippage events occur more frequently as the length of the mononucleotide increases, there would be selection against long mononucleotide tracts. This negative selection can occur at the amino acid level, where several different codon combinations can transcribe synonymous triplet codons. For example, AAA and AAG code for the amino acid lysine. Three consecutive lysine residues can be coded by any combination of AAA and AAG. If selection exists against long stretches of mononucleotide tracts, then codons that minimize the runs of mononucleotides would be favoured. By means of Kangaroo multi-faceted features, we determined the number of genes with mononucleotide tracts (≥6 bp in length) and also the frequency of codon usage in the human genome.

3.2 Material and Methods

3.2.1 Search Algorithm

Kangaroo was written entirely in C computer programming language using NCBI toolkit (Ostell,J. v. 7.0 (2000) available at ftp://ncbi.nlm.nih.gov/toolbox/ncbi_tools) and developed on a dual Pentium II processor Linux machine. The web-based application runs on a four processor Sun Solaris server. Kangaroo is available at http://bioinfo.mshri.on.ca. All GenBank records are retrieved from our own in-house SeqHound database[242], which mirrors NCBI latest GenBank release (v.123.0 Apr.2001), the NCBI taxonomy database and PDB [243]. The search algorithm is based on Regular Expression functions and is part of the NCBI C toolkit ftp://ftp.ncbi.nlm.nih.gov/toolbox. The program supports most common Regular Expression metacharacters. All searches were performed exclusively on human records from the pri (primate sequence entries) GenBank division. Coding region information was derived from these human sequence annotations as entered in the GenBank flatfile by the individual record submitters and stored in our permanent in-house database.

3.2.2 Regular Expression[i]

"Regular Expression" is a set of characters that represent one or more search strings. To find if a certain pattern is present within a given record such as DNA or protein we construct a regular expression that represents that pattern. For example, the pattern "GGATGA" represents the DNA sequence "GGATGA" and no other sequence. The regular expression "GGG[AT]GGG" represents both GGGAGGG and GGGTGGG sequences. As you can see from these examples some regular expression characters match only one character (i.e. G represents only Guanine) while others can match much more than one character. Using relatively small number of symbols one can specify many different patterns to search for in a single search. The symbols that represent more than one symbol are called metacharacters and they are used to specify a range of

characters. The scope of symbols that each metacharacter represents depends on the position and the context that these characters are used. The tables (Table 3.2.2-8–10) below show some sample regular expressions and the use of metacharacters to specify more than one string.

Table 3.2.2-8. *Regular expression metacharacters*

IUPAC	
Symbol	Name
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil
W	A or T
R	A or G
K	G or T
Y	C or T
S	C or G
M	A or C
B	C, G, or T
H	A, C, or T
D	A, G, or T
V	A, C, or G
N	A, C, G, or T

Table 3.2.2-9. *Regular Expression Symbols*

Regular Expression Symbols	
Symbol	Meaning
.	Matches any character
[..]	Matches any character listed
[^...]	Matches any character not listed
|	Alternation
(...)	Grouping
^	Matches must be at beginning of sequence
$	Matches must be at the end of sequence

Table 3.2.2-10. *Regular Expression Patterns.*

Pattern	Match
CG(AA|TT)GC	CGAAGC or CGTTGC
PXXP	Two Prolines separated by any two amino acids
^MSE	All peptide records that begin with MSE
[ST]X[VIL]$	PDZ binding sites. All matches must be at the end of a sequence.
AAAAAAAAN+AAAAAAAA	All records that contain two A(8) tracks separated by any number of bases.
[^C]CCCCCCCC[^C]	All records that contain only C(8).

3.2.3 Mononucleotide Search

For each of the four nucleotides, adenine (A), guanine (G), cytosine (C) and thymine (T) Kangaroo searches were performed starting at $(N)_6$. In order to eliminate retrieving duplicates, the query was limited to return only the entries that contained the specific run of nucleotides restricted on either side to exclude those sequences with additional identical nucleotides adjacent to the particular input. These restrictions were established using Regular Expression metacharacters. For example, if all the annotated files with a DNA sequence of a consecutive run of only eight adenines, "AAAAAAAA", then the ends of this sequence must be restricted to any other base other than A. This can be accomplished by using the metacharacter, [^A] (corresponds to "not A", see Table 3) in a query sequence, [^A]AAAAAAAA[^A]. These mononucleotide

searches were performed individually[ii] on Kangaroo for each nucleotide and for each length, from 6 base pairs to 13 base pairs. The searches were also designed to indicate the subset of files that had the query sequence "in frame", that is AAA CGA, not GGA AAC. In order to make the program more time efficient, the human annotated CDS was retrieved and stored as a separate in-house database.

3.2.4 Codon Usage Search

To observe the frequency with which synonymous codons are combined to create a stretch of three homogeneous amino acids in the human genome, we implemented one of Kangaroo's multiple features. Since we are able to limit our searches a specific nucleotide "in-frame" we eliminated the possibility of duplication of overlapping sequences. For the four homogeneous codons, AAA, CCC, GGG, TTT, the amino acids produced are Lysine, Proline, Glycine, and Phenylalanine. Redundancy within the human nuclear amino acid code, results in amino acids that can be coded by more than one codon. Thus, Lysine is coded by AAA and AAG; Proline by CCC, CCA, CCG, and CCT; Glycine by GGG, GGA, GGC, GGT; and Phenylalanine by TTT and TTC. These codons can then arranged to form a three amino acid stretch of synonymous codons. The number of combinations that can produce the triple amino acid repeats are; 8 for the amino acids that are coded by only two codons, lysine (KKK) and phenylalanine (FFF); and 64 for the proline (PPP) and glycine (GGG). The frequency of each codon predicted from the codon usage for human protein-coding sequences is taken from the studies of Nakamura *et al.*[244] (Table 3.2.4-11). The differences between the predicted frequencies of the different nucleotide sequences and the actual frequencies were analyzed with a χ^2 test.

3.3 Results

Mononucleotide microsatellites are predominately located in the non-coding regions of the human genome. However, there are >40,000 mononucleotide microsatellites, 6 to 13 base pairs in length in annotated human CDS. Some of these may fall within the same gene but for the most part, there are tens of thousands of genes with hypermutable tracts susceptible to mutation. Polyadenine tracts are expressed more (sometimes several fold) than any other mononucleotide repeat. As anticipated, as the length of the mononucleotide repeat increases, the frequency with which they appear decreases.

To accurately obtain the frequency with which a specific three-codon block is used for coding a triple repeat of the synonymous amino acids, we implemented Kangaroo's coding region search functionality. Using the prediction frequencies in Table 3.2.4-12, the probability of all the triple synonymous codon combinations were calculated and are shown in Table 3.2.4-13. Kangaroo is able to quantify the number of times a specific codon combination appears in the human CDS. These results are displayed in the "Hits in frame" column. By totalling the number of times all the possible codon combinations are found for a given triplet amino acid block, the expected number of hits for a given combination can be deciphered. The observed over expected values for each of the four classes of codons clearly shows that codon stretches that contain segments of eight or more mononucleotide repeats are underrepresented with significant Chi Squared values and where $p<0.0001$ in the human (ORFs/CDS).

Table 3.2.4-11. *Total number of records containing these specific mononucleotide sequences within the annotated human coding region.*

Repeats	Adenine	Cytosine	Guanine	Thymine
6	14119	5724	11116	7495
7	3728	1155	1697	972
8	1012	181	199	162
9	257	77	9	22
10	67	19	2	8
11	20	4	1	9
12	2	8	1	1
13	3	2	0	1

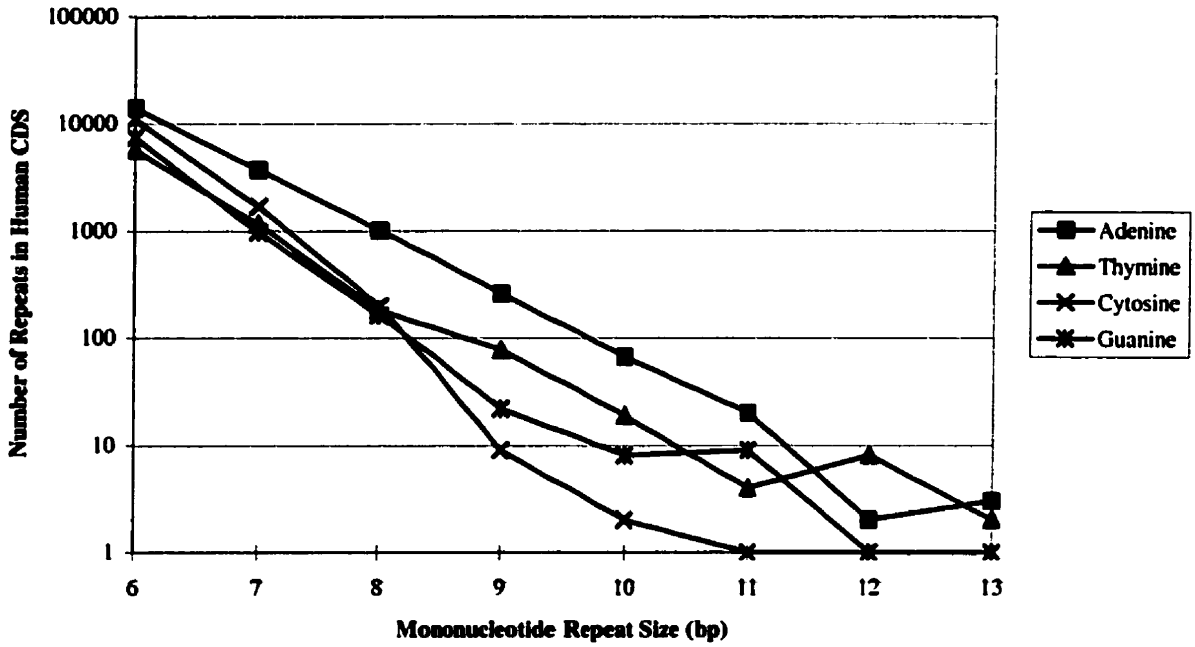


Figure 3.2.4-11. *The semi-log plot of the mononucleotide results in Table 3.2.4-11.*

3.4 Discussion

The biological basis of codon choice is not well understood in the human genome. Codon preferences have been discussed form two perspectives; translational accuracy and efficiency; selective and non-selective substitution biases. Factors that may influence codon bias include

codon context, global and local G + C content, and mutational biases [245]. In human genes, there are other factors to consider such as effects of DNA polymerase and repair systems, methylation, CpG Islands [246], tissue or organelle specificity, mRNA stability, transcriptional rate, and evolutionary age.

Codon usage can be biased in different species. For example, the amino acid lysine has two codons, AAA and AAG. Some organisms, such as *Lactobacillus acidophilus*, use the two codons equally. Others show extreme preferences: *Streptomyces venezuelae* uses AAA only for 2.2% of the time, whereas *Buchneria aphidicola* uses it for 91% of lysine residues [247]. In *Homo sapiens*, AAA is used almost equally with AAG at 42%[244]. Degeneracy allows for synonymous substitutions that do not change the protein in which it codes. Synonymous mutations were initially assumed to be effectively neutral [248] but it was apparent upon compilation of the first coding region sequences that the different synonymous codons were not used equally. Preferred codons correspond to the most abundant tRNA for each amino acid in some cases, the local chromosomal base composition can influence codon bias [249]. However, the extent of codon bias and the rate of synonymous substitutions have been described as a negative relationship [250]. Natural selection of synonymous mutations has been detected in bacteria, yeast, *Drosophilia*, nematode, plants and for the chloroplast genome [251]. In mammals, however, evidence for codon selection remains unclear.

We have shown that the use of synonymous codons that code for three identical amino acids to produce a long stretch of mononucleotides is unfavourable as seen by the low observed over expected values (Table 3.2.4-13). Studies by Laken *et al.*[135] and Linton *et al.*[136] and have also observed similar results. Through the characterization of a novel *APC* gene mutation, Laken et al. observed that a transversion of a thymine (T) to an adenine (A) in the 15th exon generated a stretch of eight consecutive As. They found that this newly synthesized repeat was susceptible to further somatic mutations as seen in the 25 % of the tumours tested that

contained nine As within the genomic DNA. This one base pair insertion creates a frameshift and leads to a truncated, dysfunctional APC protein.

Linton *et al.* observed through their studies of the apolipoprotein B gene, that the deletion of a cytosine (C) within an exon resulted in the creation of a repetitive tract of eight As [136]. This mutation, similar to most somatic frameshift mutations leads to the premature termination of the apoliopoprotein B gene resulting in a truncated apo-B protein. Interestingly, approximately 10% of the altered reading frame alleles produced a full-length functional apo-B. These alleles were restored by the transcriptional insertion of an extra adenine into the stretch of adenines.

Homonucleotide repeats are known locations of DNA polymerase slippage events during replication [239] however, the results presented by Linton *et al.*, suggest that reading frame restoration was due to the transcriptional insertion of an extra adenine into the tract of polyadenines. Chamberlin and Berg first described slippage of RNA polymerase during transcription in 1962 [252]. More recently, Wagner *et al.*, (1990) described transcription slippages by E.Coli RNA polymerase during RNA elongation at runs of 10 or more adenines and thymines, resulting in the addition of untemplated thymine or adenine residues and restoration of the proper reading frame to out-of-frame lacZ constructs[253]. Although the extra adenine corrected the frameshift mutation, Linton *et al.* hypothesized that the same mechanism could introduce frameshift mutations whenever long stretches of adenines occurred in protein-coding sequences.

Along the same line of thought as our hypothesis, Linton predicted that short, 2-5 stretches of As would be used to code three consecutive lysine residues overwhelmingly over the codon combinations that would result in stretches with eight or nine As. They studied 150 Lys-Lys-Lys motifs and found that the shorter string of adenines were at a higher-than-predicted frequency, whereas sequences with eight or nine consecutive As were observed at a much lower-than-predicted frequency ($p<0.004$ by χ^2)[iii]. We found similar results for the three lysines, with

our Chi-Square analysis revealing a p-value of <0.001. Our study was more comprehensive than that of Linton *et al.* By the means of implementing Kangaroo's functionality, we were able to analyze not only the Lys-Lys-Lys (KKK) motifs but also all of the four possible triplet amino acid blocks that could be encoded by codons leading to stretches of eight or nine mononucleotides in the human genome. For example, there are 8265 KKK motifs in our annotated human CDS database, and these were used to classify the relative frequencies with which the different codon combinations were employed. Furthermore, for the three other searches we classified the 1097 FFF motifs, 14347 GGG motifs and 16552 PPP motifs as displayed in Table 3.2.4-13.

We have found that there seems to be selection of an unknown nature against codon combinations or even homogeneous codons, AAA, CCC, GGG, or TTT to reduce the probability of slippage at long stretches of mononucleotide repeats creating somatic mutations during DNA replication and RNA transcription. This is also supported by the quantification of mononucleotide sequences in human annotated coding regions. As, seen in Table 3.2.4-11., the number of repeats of 6 base pairs is reasonable, but as the length of the tract increases, the number of repeats dramatically decreases.

Perhaps it is through evolutionary changes, for example that the codon, AAG is observed more frequently and the codon, AAA is minimized due to its potential to produce hypermutable sequences of mononucleotides. Amino acid coding in mitochondrial DNA has reassigned the codon, AAA that according to the canonical code should encode for lysine, codes for asparagine (N), along with AAT and AAC in mitochondria. It may be extreme mutational pressure towards increased genomic G+C content that mitochondria have eliminated the A-rich codon (AAA) entirely, in favour of G-ending codon (AAG) with equivalent function [254].

Table 3.2.4-12. *Predicted frequencies of the individual codons as shown by the studies of Nakamura et al. for human CDS in the year 2000.*

Amino Acid	Codon	Predicted Frequency
Lysine (K)	AAA	0.42
	AAG	0.58
Phenylalanine (F)	TTT	0.45
	TTC	0.55
Proline (P)	CCC	0.33
	CCA	0.27
	CCG	0.11
	CCT	0.28
Glycine (G)	GGG	0.25
	GGA	0.25
	GGC	0.34
	GGT	0.16

Table 3.2.4-13. *Codon usage for three-residue amino acid repeats coded by homonucleotides*

Codon Combination[a]	Amino Acid	Observed[d]	Expected[e]	Observed/Expected
AAA-AAA-AAR[b]	Lysine	475	1458	0.326[f]
AAR-AAR-AAR[c]	Lysine	7790	6807	1.14
TTT-TTT-TTY[b]	Phenylalanine	106	222	0.477[f]
TTY-TTY-TTY[c]	Phenylalanine	991	875	1.13
GGG-GGG-GGN[b]	Glycine	115	897	0.128[f]
GGN-GGN-GGN[c]	Glycine	14232	13450	1.06
CCC-CCC-CCN[b]	Proline	90	1784	0.0504[f]
CCN-CCN-CCN[c]	Proline	16462	14276	1.15

[a]R= A or G, Y= C or T, and N= A,C,T, or G.
[b]Codon combinations that produce a mononucleotide repeat of eight or more base pairs.
[c]All possible codon combinations that code for a given amino acid subtracting the set of eight or more base pair repeats in group [b].
[d]In-frame totals derived from searches utilizing Kangaroo.

[e]Expected values were calculated as the total number observed times the probability for each codon combination. The codon probability was obtained from http://www.kazusa.or.jp/codon/[244] (Table 5), where the probability of observing a given three-codon combination is the product of the three individual codon probabilities.
[f]$p<0.001$, Chi square, mononucleotide repeats of 8 or more base pairs versus all other codon combinations.

[i] This section is from the Kangaroo web interface
[ii] These searches were performed by D.Betel, a M.Sc. student under the supervisor of Dr. C.Hogue at the Samuel Lunenfeld Research Institute.
[iii] The statistics, (Chi-Square) were performed with the assistance of Laurent Briollais.

Chapter Four

Utilization of Kangaroo to Identify Possible Candidate Genes in High Frequency Microsatellite Instability Colorectal Carcinomas

Chapter Four

Utilization of Kangaroo to Identify Possible Candidate Genes in High Frequency Microsatellite Instability Colorectal Carcinomas

4.1 Background

It has been reported that poly A/T repeats are more stable than poly C/G [255], suggesting that the stability of the C/G repeats frameshift intermediate due to hydrogen bonding enables the mismatch to be tolerated long enough so as to be far enough away from the extending 3' end and escape the action of the exonuclease [256]. If the integrity of the cells genome has been compromised, as in mismatch deficient (MMR-) tumours, a $(N)_8$ repeat is 3000 times more susceptible to mutation compared to tumours that are MMR competent (MMR+) [239]. Thus, these mononucleotide tracts are hypermutable in tumours that display microsatellite instability. As evidence shows, for example, the gene, *TGFβRII* is mutated in approximately 80% of MSI tumours at an $(A)_{10}$ tract. The identification of genes containing coding mononucleotide microsatellites may be valuable in discovering candidate genes important in the carcinogenesis of MSI tumours. Using Kangaroo we are able to isolate and quantify genes containing mononucleotide repeats. These candidate genes may be important in understanding the underlying molecular genetic events leading to MSI cancers.

Repeats of any given size and type should have relatively stable mutation frequencies in intronic DNA. Coding sequences with higher mutation frequencies must have been biologically selected for (growth inhibitory genes), while those with lower frequencies were likely selected against (growth stimulatory genes)

For the purpose of this project, Kangaroo enables us to search for polyadenine tracts, 8 base pairs in length or greater in human coding sequences exclusively. Repeats longer than 7 base pairs in length have been shown to be more mutable than shorter repeats [239]. Genes were investigated through surveying the literature to reveal functions of the gene and its interacting

partners (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed). Selection of genes was based on the specific gene's importance in cell cycle control (*CDC25C, CDC7, CHK1, FAP-1, MITOSIN, MIG-2*), signal transduction (*GRK4, RACK7, AXIN2, DP-2, PP-1 inhibitor*), implication in disease (*RECQL1, ERCC5, BLYM*), DNA replication (*DNA ligase III, CBF2*), and some miscellaneous and putative genes (*VRK2, NP220*), with no apparent implication in tumourigenesis. As well, the known mismatch repair genes (*hMSH3, hMSH6*, and *hPMS2*) with mononucleotide repeats were screened to determine the mutation frequencies in our tumour series.

In order to improve efficiency of finding genes with significant mutation frequencies in primary colorectal carcinomas, we initiated a rapid screening regime of 13 genes only in the colon cancer derived cell lines (including LoVo, an endometrial derived, MSI-H cell line). The rationale behind this effort including the thought, that if the 8 MSI-H cell lines exhibits a significant mutation frequency, then it is more likely that the primary CRCs will also carry similar mutations. Thus, only those genes with substantial mutation frequencies would then be screened in the primary tumours, of which there is a finite supply. In addition, to the MSI-H cell lines, MSS cell lines complete this panel thus possibly enabling us to find other genes that may help to differentiate even further the two destabilizing pathways.

It has become apparent that cell cycle checkpoints function in association with DNA damage recognition and repair pathways. Thus, genes in one pathway can affect multiple pathways leading to apoptosis if critical controls are still intact or to the uncontrollable promotion of cancer.

Progression through cell cycle is mediated at checkpoints. The cell cycle occurs despite continuous assault on the genome from carcinogens and other environmental hazards leading to instability of DNA. Checkpoints constantly monitor the genome for accuracy and the orderly events in the cell cycle namely (DNA replication, mitosis and integrity of DNA). Activation of a

specific checkpoint halts progression until either the problem is corrected or the cell decides to undergo programmed cell death or simply live with the problem, the major checkpoints being G_1/S and G_2/M. The three cellular components involved in genome transmission are the DNA itself, the spindle, and the spindle pole. Checkpoints are intracellular signal transduction pathways; damage or errors must be sensed and the signal transduced to effectors, which then regulate cell cycle progression. Damage sensors are important in initiating the damage response; to date these sensors are poorly understood due to the lack of their identification. A possible candidate might be Ataxia Telangiectasia Mutated (*ATM*) or ATM-Rad3 related (*ATR*) genes, which both recognize DNA directly, suggesting a role in the sensing pathway. If the damage that the DNA incurs involves a crucial function the checkpoints arrest the cell cycle progression and activates the DNA repair system in the cell. Thus, the regulation of the DNA repair system and DNA damage checkpoints are thus intertwined. Checkpoints are not essential for proliferation and thus can be lost through mutation. The loss of one or more checkpoints is a feature seen in many cancers [257,258]. Tumours that display chromosomal instability usually involve the loss of a checkpoint due to mitotic spindle defects [259,260]. On the other hand tumours that progress through the microsatellite instability pathway may lose their ability to activate their checkpoints through the loss of one or several genes. Thus, the loss of mismatch repair in these cells increases cell survival and the concomitant decrease in the level of apoptosis [261].

Other means of deregulating cellular processes are through the desensitization of the cell to external cues by aberrant expression of positive regulators, such as cyclins, or by the loss of negative regulators, such as the CDK (CDC-dependent kinase) inhibitors. The family of proteins known as CDK inhibitors includes the tumour-suppressor gene *p16*, as well as other proteins such as p21, p27 and p15. Loss of p27 expression has shown to predict poor prognosis in certain colorectal cancer patients[262]. In colorectal cancers, the cyclins, D2 and E have been shown to be amplified. Overexpression of cyclin D1 is implicated in breast cancers and other tumours but not

in colorectal cancers. However, recently Kong *et al.*, have found that a cyclin D1 polymorphism is linked with early onset of colorectal patients [263], which may be important with respect to microsatellite unstable tumours, which present themselves at an earlier stage of life.

To fully understand the molecular pathogenesis of MSI colorectal cancers, we must first determine the genes that are mutated, suppressed or over-expressed. By identifying genes with frameshift mutations at mononucleotide repeats in this study we are determining the importance of the loss of particular protein functions in tumourigenesis.

4.2 Materials and Methods

4.2.1 Selection of Candidate Genes

Kangaroo retrieved 1361 records that contain consecutive runs of 8 or greater adenines. Kangaroo is an invaluable tool, yet human manipulation is still required to select genes of interest for specific projects. For this project, genes with functional significance were selected according to their role in cell cycle control, proliferation and apoptosis. Functions were investigated by searching NCBI GenBank nucleotide records (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide) and published articles pertaining the relevant information on the individual genes. When possible, genes with known intron/exon junctions were chosen to alleviate initial primer design problems that may be encountered. However, due to errors in sequencing by the original submitters of particular sequences, incorrect fragments were amplified. In total, 21 genes were screened in the primary CRCs. Of these, only 20 were screened against the cell line DNA (*NP220* excluded). In addition, 20 mononucleotide repeats (13 genes) were screened exclusively against the cell line panel.

4.2.2 Mutational Analysis

Once the genes were selected according to functionality, primers were designed to include the mononucleotide repeat. Primers were designed to be less than 150 base pairs in length in order to increase the efficiency of amplifying archival paraffin-embedded tissues. Occasionally, the web-based programme, Primer 3 (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) was used to design primers for difficult segments of DNA. Subsequently, these fragments were analyzed with an NCBI tool, BLAST, to look for homologies to the specific segment to increase the probability of amplifying only the sequence of interest (see Appendix A for primer sequences). The PCR methods and detection methods for mutational analysis are the same as described in Chapter 2, section 2.2.5.

4.2.3 Sequencing of Putative Mutations

See Chapter 2, section 2.2.7

4.2.4 Tissue Samples

See Chapter 2, section 2.2.1

4.2.5 Cell lines

See Chapter 2, section 2.2.2

4.3 Results

4.3.1 Mutational Analysis in Complete Panel

In Table 4.3.1-14, the results for the 21 genes successfully screened against the complete panel of 101 primary tumours and cell lines available are listed. Others have screened the mismatch

repair genes, *hMSH3*, *hMSH6*, and *hPMS2* and we show here, comparable results. The polyadenine tracts in *hMSH3* and *hPMS2* were mutated by a frequency of 42% and 5%, respectively in the tumours. The homocytosine repeat in *hMSH6* was found to be mutated in 29/100 (29%) in the tumours and is mutated in 4/8 of the MSI-H cell lines. This frequency is higher than for *hMSH3* and *hPMS2* that are mutated in 2/8 and 1/8, respectively.

Other genes, besides the MMR genes that exhibit a moderate to significant mutation frequency are *AXIN2* (15%), *CDC25C* (11%), *RECQL1* (12%), *ERCC5* (10%), *CBF2* (16%), *RACK7* (19%), and *GRK4* (21%). MSI-H cell lines also exhibit these mutations with the exception of *ERCC5* and *RACK7*, where the former, the sample size of the cell lines is too low to detect a mutation frequency of 10%. Whereas with *RACK7*, not a MSI-H cell line but a MSS cell line (1/9) is mutated. *CBF2* showed mutations in both the MSI-H (3/8) and MSS (1/14) cell lines.

3/8 MSI-H cell lines had mutations in any of the 4 mononucleotide repeats in *AXIN2*, which we screened simultaneously, G7, A6, C6 and C5. This was the only gene screened with multiple smaller repeat sequences. The MSI-H cell line, LS411, upon sequencing exhibited a Leu-to-Arg, CTG-CGG at residue 662 that falls just upstream from the Dax and Dix regions. 14/93 primary tumours had *AXIN2* mutations, of the 2 tumours with multiple histological types, 1 had a subtype that carries the *AXIN2* mutation, indicating either a late event mutation, or selective/disadvantageous mutation. 1 tumour, HC387 was removed due to insufficient evidence of sequencing data. Therefore, there are 14 mutations in a total of 13 primary tumours. Only one tumour (HC186) displayed bi-allelic inactivation, most of the mutations occurred within the G7 tract of *AXIN2*, 10/14, with 3/14 in the C6 and 1/14 in the A6 tract.

The remaining genes have mutation frequencies that are less than 5%; these include, *CHK1*, *CDC7*, *DNA lig III*, *DP-2*, *BLYM*, *FAP-1*, *PP-1 Inh*, *VRK2*, *MIG-2*, *MITOSIN*, and *NP220*. Within this group of genes with low mutation frequencies in primary tumours, there are

a few that are mutated in at least one MSI-H cell line. *DP-2*, *BLYM*, and *FAP-1* are mutated at their $(A)_8$ repetitive element in 1/8 MSI-H cell lines and none of the MSS cell lines. Also within this group, *CHK1* with a mutation frequency of 4% is mutated in a MSS cell line (1/9).

Overall, the mutation frequencies in the cell lines were similar to those in the primary tumours, of the 11 genes that were ≤5% in primary tumours there were 3 genes with mutations in 1/7 MSI-H cell lines, which accounts for ≤5%. In this subset of gene, we would expect to see the occasional mutation in 1/7 cell lines. Thus, the cell lines show the same trend as the primaries.

4.3.2 Mutational Analysis in Cell Lines

Twenty-two mononucleotide repeats in 14 genes were screened against a cell line panel of 8 MSI-H cell lines and 16 MSS cell lines (Table 4.3.2-15). Of the 4 repetitive elements in the Caspase 8-associated protein 2 gene (*CASP8AP2*), the A_8 tract is the only one that is mutated in one MSI-H cell line (LoVo). The A_8 in p72 is mutated in the endometrial derived cell line, HEC1A. The DNA recombination/repair gene, *RAD50* has two polyadenine tracts of 8 and 9 base pairs in length. While, the A_8 is not mutated in any of the eight MSI-H cell lines, the A_9 tract is mutated in LoVo and HCT-116. The transcription factor, *TFE3*, which has a G_8 repeat, is mutated in the MSI-H cell line, SW48. Frameshift mutations were not detected in the other mononucleotide repeats in the remaining 10 genes in either the MSI-H or MSS cell lines.

4.3.3 Intratumour Heterogeneity

When areas of a specific tumour sample display distinct pathological characteristics, each defining area was microdissected separately. When a gene is mutated in various stages of tumour development visible in a sample, this indicates that the mutation might be an early event and through clonal progression, all the dysplastic, tumour cells carry this same mutation. If only a single area exhibits the mutation, then, either the mutation is exclusive to that particular clone, or

it may be a later event and the other clones will acquire that mutation as well or, the clone is being selected against by carrying the mutation and will be eliminated. In our sample group, we had a selected number of tumours in which multiple areas were sampled. Of these, 6 harbour a mutation in *GRK4* and 3 of these 6 represented tumours where more than one tumour sample was mutated. The MMR genes, *hMSH3* 4 tumours out of 8 tumours with *hMSH3* mutations have this mutation in more than one tumour tissue, indicating an early or advantageous event. *hMSH6*, is also seen to be mutated in 2/5 tumours with multiple tumour areas. This is also seen in 1/2 tumours with multiple samples with *RECQL1*, 1/3 with *RACK7* mutations and in all the tumours (4/4) with mutations in *CBF2*. One of the two tumours with mutations in *CDC25*, *hPMS2*, and *MITOSIN* are each mutated in one tumour with multiple samples and in those tumours, more than one sample area contains the frameshift mutation in those specific genes (Table 4.3.3-16).

4.3.4 **Intronic Polyadenine Repeats**

Genes with intronic polyadenine repeats with lengths of 7 to 10 consecutive adenines were screened against the complete panel of cell lines and primary tumours. The shortest repeat, A_7 is not mutated in any of the cell lines nor tumours. The A_8, A_9, and A_{10} demonstrated higher but expected mutation frequencies of 24%, 12%, and 15%, respectively. The A_{10} tract in the intron of *BLFZ1* was the only one with mutations in the cell lines. HCT-8, DLD-1, LS174 are the MSI-H cell lines with an insertion/deletion mutation in the intronic A_{10} (Table 4.3.4-17).

Table 4.3.1-14. *Summary of mutation frequencies of genes screened against our complete panel of primary MSI-H tumours and cell lines.*

Gene	GenBank Accession Number	Function	Repeat	Cell lines		Primary MSI-H Tumours	
				MSI-H	MSS		
hMSH3	NM_002439	*Mismatch repair*	A_8	2/8	0/15	42/101	42 %
hMSH6	NM_000179	*Mismatch repair*	C_8	4/8	0/15	29/100	29 %
hPMS2	NM_000535	*Mismatch repair*	A_8	1/8	0/15	5/99	5 %
AXIN2	NM_004655	*Wnt signalling pathway*	G_7, A_6, C_6, C_5	3/8	0/15	14/93	15 %
CHK1	NM_001274	*Checkpoint protein*	A_9,A_7	0/7	1/9	3/81	4 %
CDC25C	NM_001790	*Tyrosine phosphatase, G2/M phase regulator*	A_8	2/7	0/9	10/93	11 %
CDC7	NM_003503	*G1/S phase regulator*	A_9	0/7	0/9	0/99	0
RECQL1	NM_002907	*RecQ helicase involved in DNA repair*	A_9	1/7	0/9	11/93	12 %
ERCC5	NM_000123	*Cut and patch repair*	A_9	0/7	0/9	8/80	10 %
DNA LigIII	NM_013975	*Ligates DNA during replication/repair*	A_8	0/8	0/13	1/85	1 %
DP-2	NM_006286	*Complexes to E2F to regulate cell cycle*	A_8	1/7	0/9	2/94	2 %
BLYM	NM_005179	*Burkitt's lymphoma transforming gene*	A_8	1/7	0/9	5/96	5 %
CBF2	NM_005760	*CCAAT box binding protein*	A_9	3/8	1/14	14/91	16 %
RACK7	NM_012408	*Protein kinase c binding protein*	A_8	0/7	1/9	17/91	19 %
FAP-1	NM_006264	*Fas inhibitor*	A_8	1/8	0/14	0/92	0
GRK4	HSGRKG08	*Desensitize G protein-coupled receptors by phosphorylating activated receptors*	A_9	2/8	0/14	19/91	21 %
PP-1inh	NM_006741	*Protein Phosphatase Inhibitor*	A_8	0/7	0/9	2/97	2 %
VRK2	NM_006296	*Serine/Threonine kinase*	A_8	0/8	0/14	2/92	2 %
MIG-2	Z24725	*Mitogen activated*	A_8	0/7	0/9	1/77	1 %
MITOSIN	NM_016343	*Mitotic-phase progression*	A_8	0/8	0/14	3/93	3 %
NP220	NM_014497	*DNA binding nuclear protein*	A_8			2/71	3 %

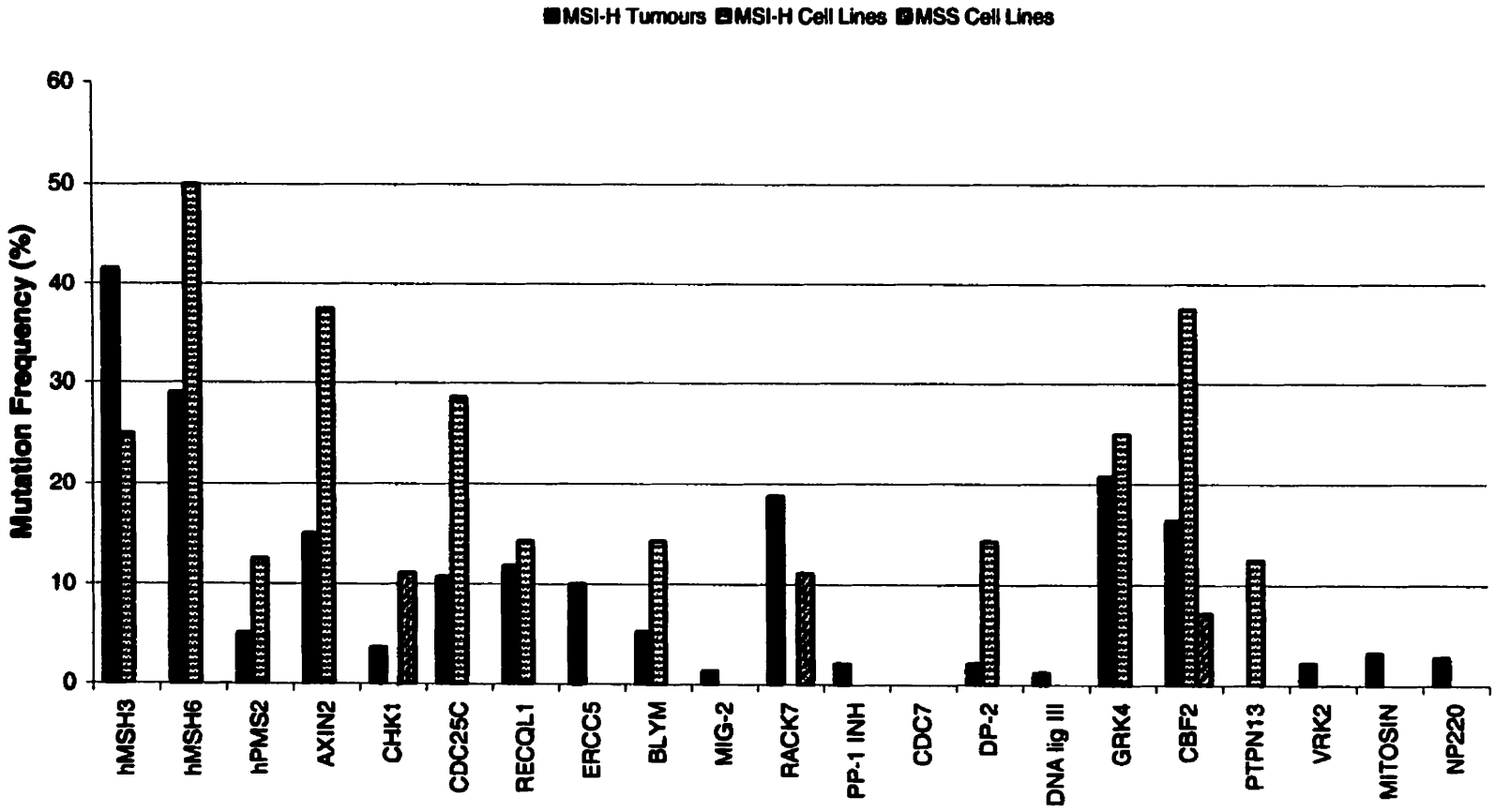


Figure 4.3.1-12. *Graphical representation of results tabulated in Table 4.3.1-14*

Table 4.3.2-15. *Summary of mutation frequencies of genes screened only against the cell lines.*

Gene	GenBank Accession Number	Function	Repeat	Cell Lines	
				MSI-H	MSS
Death associated protein 6 (DAXX)	NM_001350	*Enhances Fas-mediated apoptosis*	A_8	0/8	0/16
Silencer of Death Domain (SODD/BAG4)	NM_004874	*Apoptosis*	A_7	0/8	0/16
			A_8	0/8	0/16
MAP kinase kinase kinase (MTK1)	NM_005922	*Regulates the p38 and JNK pathways*	A_8	0/8	0/16
Caspase 8 associated protein 2 (CASP8AP2/RIP25)	NM_012115	*FLASH homolog*	$A_{7.1}$	0/8	0/16
			$A_{7.2}$	0/8	0/16
			A_8	**1/8**	0/16
			A_9	0/8	0/16
BASS1	AF139897	*Putative apoptosis inhibitor*	T_7	0/8	0/16
			A_8	0/8	0/16
TC10	NM_012249	*Rho-family GTPase*	A_8	0/8	0/16
MOK (also RAGE)	NM_014226	*Renal tumour antigen*	A_9	0/8	0/16
GNEF	NM_016340	*Guanine nucleotide exchange factor*	A_7	0/8	0/16
			A_9	0/8	0/16
p72	NM_006386	*Transcriptional/translational*	A_8	**1/7**	0/16
RAD50	NM_005732	*DNA recombination/repair protein*	A_8	0/8	0/16
			A_9	**2/8**	0/16
HOXA11	NM_005523	*Transcription regulation*	A_9	0/8	0/16
REV1	NM_016316	*Damage bypass replication*	A_7	0/8	0/16
SHC transforming protein 1	NM_003029	*Signal transduction*	G_8	0/8	0/16
TFE3	NM_006521	*Transcription factor for Immunoglobulin Heavy-chain enhancer 3*	G_8	**1/8**	0/16

Table 4.3.3-16. *Summary of gene mutations in tumours with multiple areas with different histological characteristics.*

Gene	Tumours with Heterogeneous areas with at least one mutation	Tumours with mutation in majority of areas
GRK4	6	3
CDC25C	1	1
BLYM	1	1
hMSH3	8	4
hMSH6	5	2
CBF2	4	4
MITOSIN	1	1
RECQL1	2	1
RACK7	3	1

Table 4.3.4-17. *Mutation Frequencies of polyadenine repeats found within introns.*

Gene	Position	Repeat	Cell Lines		Primary MSI-H Tumours
			MSI-H	MSS	
BLFZ1	Intron 2	A_7	0/8	0/17	0/83
RhAG- Rh glycoprotein gene	Intron 8	A_8	0/8	0/17	24/101
Beta Defensin 1	Intron 1	A_9	0/7	0/7	12/99
BLFZ1	Intron 2	A_{10}	3/8	0/17	13/84

4.4 Discussion

4.4.1 Genes screened in cell lines

Following our original thought that genes in apoptosis genes must be disturbed in order for cells with defective mismatch repair to propagate without committing suicide, we chose genes with functional significance in apoptosis regulation. There are several apoptosis pathways of which the CD95-FADD-caspase 8 pathway is one, which have physiological significance. In addition, there are other speculative pathways such as, CD95-RIP-RAIDD-caspase2 [264], CD95-DAXX-ASK1-JNK-unknown caspase, and CD95-FLASH-unknown caspase or adaptor [265]. From our Kangaroo searches, we identified and investigated the following genes; Death-Associated protein 6, *DAXX*; Silencer of Death Domain, *SODD* [266]; Caspase 8-Associated protein 2 (*CASP8AP2*); *BASS1*; and the PDZ domain containing guanine nucleotide exchange factor, GNEF.

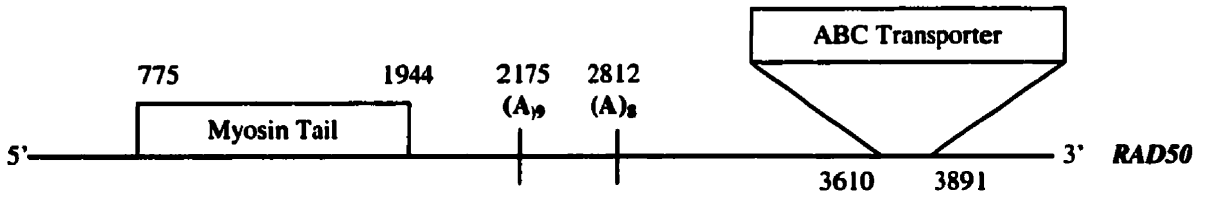
Human *DAXX* encodes a 740-amino acid polypeptide containing a nuclear localization signal. Functional analyses demonstrated that Daxx binds to the Fas death domain and enhances Fas-mediated apoptosis [267]. The authors suggested that DAXX and FADD define 2 distinct apoptotic pathways downstream of Fas. The DAXX-ASK1 connection provides a mechanism for caspase-independent activation of JNK by FAS and perhaps by other stimuli [268]. We were not able to detect any frameshift mutation in either the MSI-H or the MSS cell lines, indicating that the A_8 tract in *DAXX* is not a target for mutation in colorectal tumourigenesis and perhaps intact *DAXX* is a suppressor of apoptosis as suggested by the study in mice where *DAXX* when knocked out enhances apoptosis rather suppressing it [269].

GNEF was chosen to study due to its association to and functional similarity to Ras. It contains two repeats, an A_7 and an A_9, neither of which were found to be disrupted in our panel of cell lines. Negative results were also seen for the putative apoptosis inhibitor, *BASS1*. As well, the two polyadenine repeats, A_7 and A_8 found in *SODD*, also known as Bcl2-Associated

athanogene 4, *BAG4* were intact. Enhanced expression of *SODD* is seen in pancreatic cancers [270] but no association has been illustrated in colorectal cancers.

The caspase 8-associated protein 2 (*CASP8AP2*) is also known as *RIP25* (FLASH homolog). The RIP stands for receptor-interacting protein, and is an adaptor protein with a death domain [271]. This gene is involved in Il-2 signalling and is suspected to play an important role in downstream events from Caspase 8 in apoptosis. This gene contains two A_7, one A_8, and one A_9 repeat. Our results reveal a single base deletion at the A_8 tract in the MSI-H cell line, LoVo. This result may be used to support the task of scanning this particular gene in the panel of MSI-H primary tumours.

Two genes were screened against our panel that have functions involved in DNA repair,



REV1 and *RAD50*. The REV1 encodes a DNA polymerase with deoxycytidyl transferase activity [272] and is homologous to the E.Coli DinB and UmuC proteins. No mutations in this A_7 repeat were found. However, the *RAD50* gene, that contains an A_8 and an A_9 repeat is mutated in the A_9 tract of the MSI-H cell lines, LoVo and HCT-116. The single adenine deletion in the mononucleotide tract results in protein truncation and thus, loss of the ABC transporter domain. Similar results have also been shown in 31% of MSI-H gastrointestinal cancers [155]. Recently, the nuclear complex containing RAD50 and MRE11 and nibrin or p95 has been implicated in DNA damage detection, activation of cell cycle checkpoints [273], and DSB repair as part of the BASC, BRCA1-Associated Surveillance Complex [207]. The functional revelations and striking results by us and others, makes *RAD50*, a potential target in the MSI pathway.

Other genes, consist of signal transduction genes; the MAP kinase kinase kinase 1 gene (*MTK1*); SHC transforming protein 1; MOK protein kinase; and the Rho-family member, TC10: a transcription factor; TFE3: a RNA helicase, p72; and a transcriptional regulator; HOXA11.

MTK1 activates and phosphorylates MAPKK, which in turn activates and phosphorylates MAPK. MAPK kinase activates a specific MAPK. There are different types of MAPKs some of which are activated by environmental stresses, JNK MAPKs, and others such as the ERK MAPKs by mitogenic stresses (osmotic shock, UV irradiation, wound stress, and inflammatory factors) [274,275]. The MAP kinase superfamily is an important group of kinases and thus, *MTK1* and *MOK*, another member of this family were screened for mutations. No frameshift mutations were observed at the polyadenine repeats in the coding regions of these genes, predicting that the function of these genes may be essential not only for propagation of the normal but also aberrant cells.

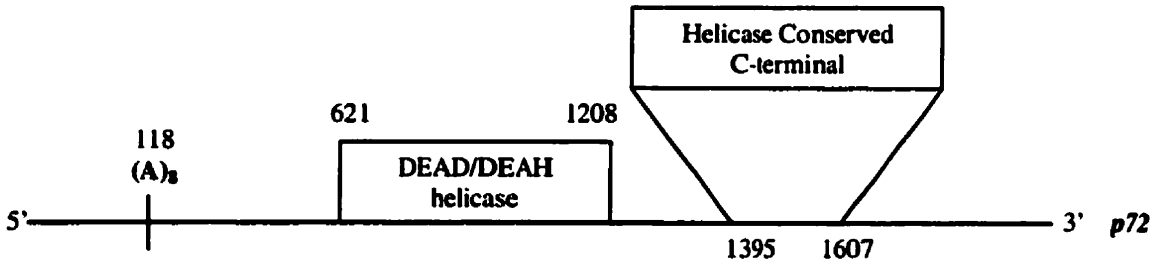
Although *SHC transforming protein 1* did not have the typical polyadenine tract, it was chosen to study due to its quintessential role in signal transduction [276]. It does however contain a G_8 repeat, which along with the one found in TFE3 were the only two deviations from the >8 polyadenine tract, merely for interest purposes. Only one MSI-H cell line, SW48 displayed a mutation in the *TFE3* gene. The TFE3, transcription factor for immunoglobulin heavy-chain enhancer 3 is located on Xp11.22 [277], a region that is involved in translocations found in several malignancies [277-279].

The *TC10* gene is a Rho (Ras superfamily) family GTPase that is highly similar to CDC42[280]. Activated forms of both proteins stimulate transcription mediated by nuclear factor kappaB, serum response factor, and the cyclin D1 promoter, they also activate c-Jun NH(2)-terminal kinase [281]. This family has been shown to regulate a variety of apparently diverse cellular processes such as actin cytoskeletal organization, mitogen-activated protein kinase

(MAPK) cascades, cell cycle progression and transformation. Thus, perhaps due to the fact that it could be involved in so many different processes, it is not mutated in any of our cell lines.

The transcriptional regulator, HoxAll, (homeo box 11, hox11) is primarily involved in forearm morphogenesis; the decision behind choosing this gene was the very fact that it was involved in embryogenesis and not tumourigenesis [282]. The rational behind choosing this gene is study a gene that would not typically be involved in the progression from normal cell to a cancerous one. Unfortunately, this gene is not one of such genes. There were no insertion/deletions detected at the A_9 repetitive element.

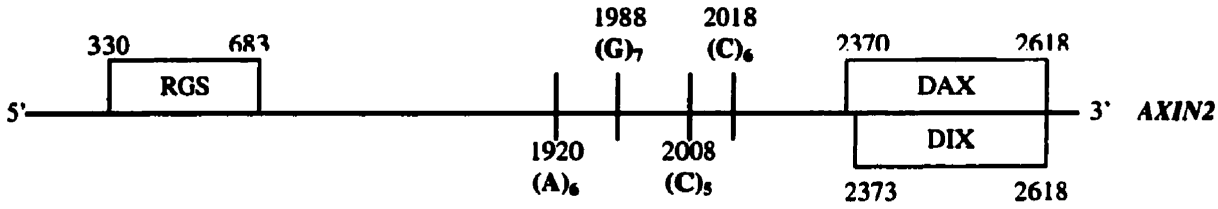
The only other gene with a detectable mutation in any of the MSI-H cell lines is p72, which codes for a DEAD-Box protein. These RNA-dependent ATPases, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. They are implicated in a number of cellular processes involving alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly. Some members of this family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division [283]. A one base pair deletion was detected in the polyadenine repeat located at residue 118, upstream from both the functional and conserved domains in the cell line, HEC1A.



4.4.2 Genes screened in both Primary CRCs and Cell Lines

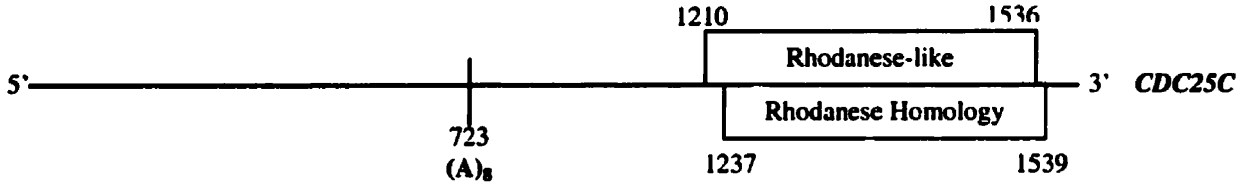
In the progression of colon cancer from adenoma to carcinoma, the Adenomatous Polyposis Coli (*APC*) gene is mutated as an early event. The *APC* gene functions in Wnt/Wingless signal transduction. Inactivating APC mutations are observed in approximately 70% of colon cancer patients with MMR deficiency [284]. This mutation leads to the stability of Beta-catenin, which activates the transcription of genes downstream of the TCF/LEF promoter. Other genes in the Wnt/Wingless pathway have been implicated in colon cancer. *Beta Catenin (CTNNB1), GSK3β,* and *AXIN2* have been described as being mutated in varying mutation frequencies. Activating CTNNB1mutations occur almost exclusively in MSI-H CRCs. When tested in 100/102 of our panel, 18% were found to have mutations[i]. Axin2 is thought to play an important role in the regulation of the stability of beta-catenin in the Wnt signalling pathway. In mouse, conductin (axin) organizes a multiprotein complex of APC (adenomatous polyposis of the colon), beta-catenin, glycogen synthase kinase 3-beta, and conductin, which leads to the degradation of beta-catenin. This deregulation of beta-catenin is an important event in the genesis of a number of malignancies. The *AXIN2* gene has been mapped to 17q23-q24 [285], a region that shows frequent loss of heterozygosity in breast cancer, neuroblastoma, and other tumours. Our data presented here shows that *AXIN2* mutations are present in 14/93 tumours of our panel, with only one tumour that has both *CTNNB1* and *AXIN2* mutations. These results reveal that *CTNNB1* and *AXIN2* may not be totally mutually exclusive, but there seems to be a trend indicating that these genes may have functional significance in alternative pathways. Lui et al, 2000 showed that *AXIN2* was mutated in 11 of 45 CRC with defective MMR. They showed that the mutations stabilize beta-catenin and activate beta-catenin/T-cell factor signalling [145]. Preliminary data suggested that *AXIN2* is overexpressed in CRC. Webster *et al.* (2000) found two sequence variants causing amino acid substitutions in four colon cancer cell lines, a Ser-to-Leu at residue

215 in LS513 and a Leu-to-Met at residue 396 in HCT-8, HCT-15, and DLD-1. They chose to study further the L396M mutation since it fell within the RGS (regulators of G-protein signalling) region that was shown to interact with glycogen synthase kinase 3β [286]. We were not able to detect these mutations since we were looking for frameshift mutations specifically at clustered in exon 7 containing the four mononucleotide repeats. However, we found in another MSI-H cell line, LS411, in the region of our interest, a Leu-to-Arg at residue 662, upstream from the dax region (found in Dishevelled, and Axin) and the dix regions. Additionally, one tumour displayed biallelic inactivation of this gene. *GSK3β* also contains a T_6 repeat within the serine/threonine protein kinase domain and analysis of this repeat would be essential in understanding the role of these genes in colon cancer. This gene has been screened in colon cancers that were not stratified into MSI status [286].



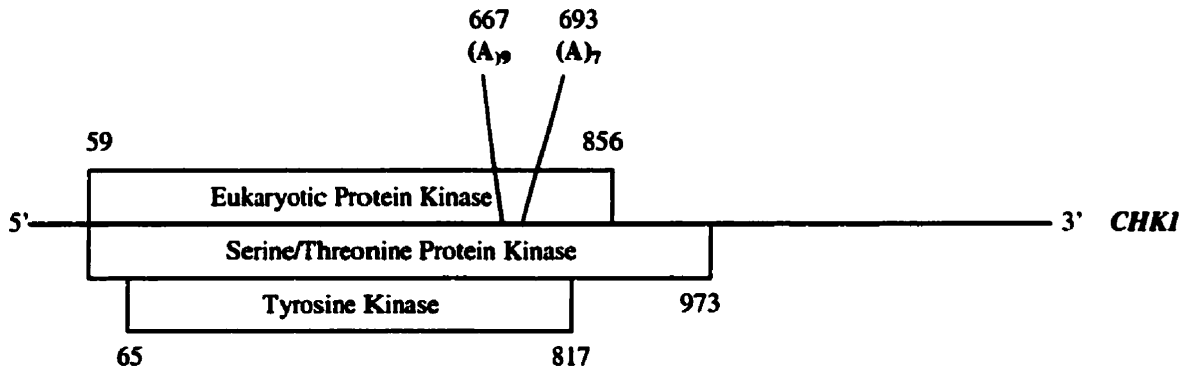
The *CDC25C* gene is highly conserved during evolution and it plays a key role in the regulation of cell division. The encoded protein is a tyrosine phosphatase and belongs to the cdc25 phosphatase family [287,288]. The cdc25C phosphatase dephosphorylates cdc2 kinase which then in complex with cyclin B can catalyze transition from the G(2) phase to mitosis. It is also thought to suppress p53-induced growth arrest by binding to it through a site in the cdc25C promoter. In addition, through a segment containing three CCAAT-boxes, cdc25C transcription is deregulated by p53. This transcriptional repression by p53 leads to reduced levels of active cdc2 kinase contributing to G(2) arrest and G(2)/M checkpoint control [289]. The protein encoded by checkpoint kinase 1 gene (Chk1) phosphorylates cdc25c on serine-216 in response to DNA

damage, thus preventing mitotic entry [290]. This phosphorylation of ser-216 creates a binding site for 14-3-3 protein and inhibits the function of the phosphatase by escorting the protein out of the nucleus into the cytoplasm [291]. *CDC25C* contains a mononucleotide repeat (A_8), a susceptible target for mutation in MSI cancers, it is found distal to serine 216 at lysine 241. In our panel of MSI-H tumours, we found 10/93 frameshift mutations resulting in probable protein loss due to truncation. The consequences of losing cdc25C would be similar to the response directed by DNA damage, both result in the loss of the phosphatase function resulting in G2/M arrest in these cells.



A study of the cdc25 isoforms in colorectal carcinomas revealed that the various isoforms are differentially regulated and may participate in the development of these tumours. They found a correlation between increased cdc25B mRNA levels and the relapse-free, overall, and cancer-related survival of the patients [292].

Although detection of DNA damage and the subsequent initiation of the cell cycle checkpoint are not fully understood, as suggested earlier, *ATM* and *ATR* are possible damage detectors due to their direct DNA binding ability. Atm [206] and Atr [205] are upstream regulators of



Chk1, and thus are important in this checkpoint as well. Chk1 also phosphorylates cdc25A and cdc25B, and since the ser-216 is conserved among the isoforms, it was suggested that Chk1 might regulate other DNA damage checkpoints, such as those controlling the G1/S phase transition [291]. *CHK1* also contains a coding region mononucleotide repeat, an A_9, we screened this gene and found a low mutation frequency of 4%, (3/81%). For this particular gene we had a low amplification rate wherein only 81/102 primary tumours amplified using these gene primers. The mutations in *CHK1* and *CDC25C* are mutually exclusive in our series.

Canman *et al.* (2001) suggest that the loss of checkpoint results in mitotic catastrophe and that the cooperation of *ATR* and *CHK1* are essential in preventing such a crisis [293]. Mitotic catastrophe is an apoptosis-like process that begins in prophase, after the dissolution of the nuclear membrane and is associated with the entry of cdc2 and cyclin b1 into the nucleus. Chromatin condensation and micronucleation and is distinguished from PCD by the lack of DNA degradation. Although mutations in the similar *ATM* confers susceptibility to cancer, homozygous deletions of either *ATR* or *CHK1* results in embryonic lethality in mice [204,205,294,295]. Mitotic catastrophe is also seen when 14-3-3σ alleles are inactivated in colorectal cancer cells. 14-3-3 normally sequesters cyclin b1 and cdc2 in cytoplasm and inhibits mitosis following DNA damage [296]. Thus, cell cycle arrest is established initially but the cells died when they entered into mitosis. From the consequential loss of *ATR* or *CHK1*, 29% of our sample population may display malignancies of a lower grade attributable to mitotic catastrophe and not apoptosis.

In addition to cdc25C, Atr and Chk1 also interact with p53 in various ways to affect the G2/M checkpoint [189]. Atr phosphorylates p53 at SER15, a site that prevents it from binding to its repressor, MDM2 [188]. P53 is also phosphorylated on SER15 by Chk1 as well serines, 20 and 37 [297].

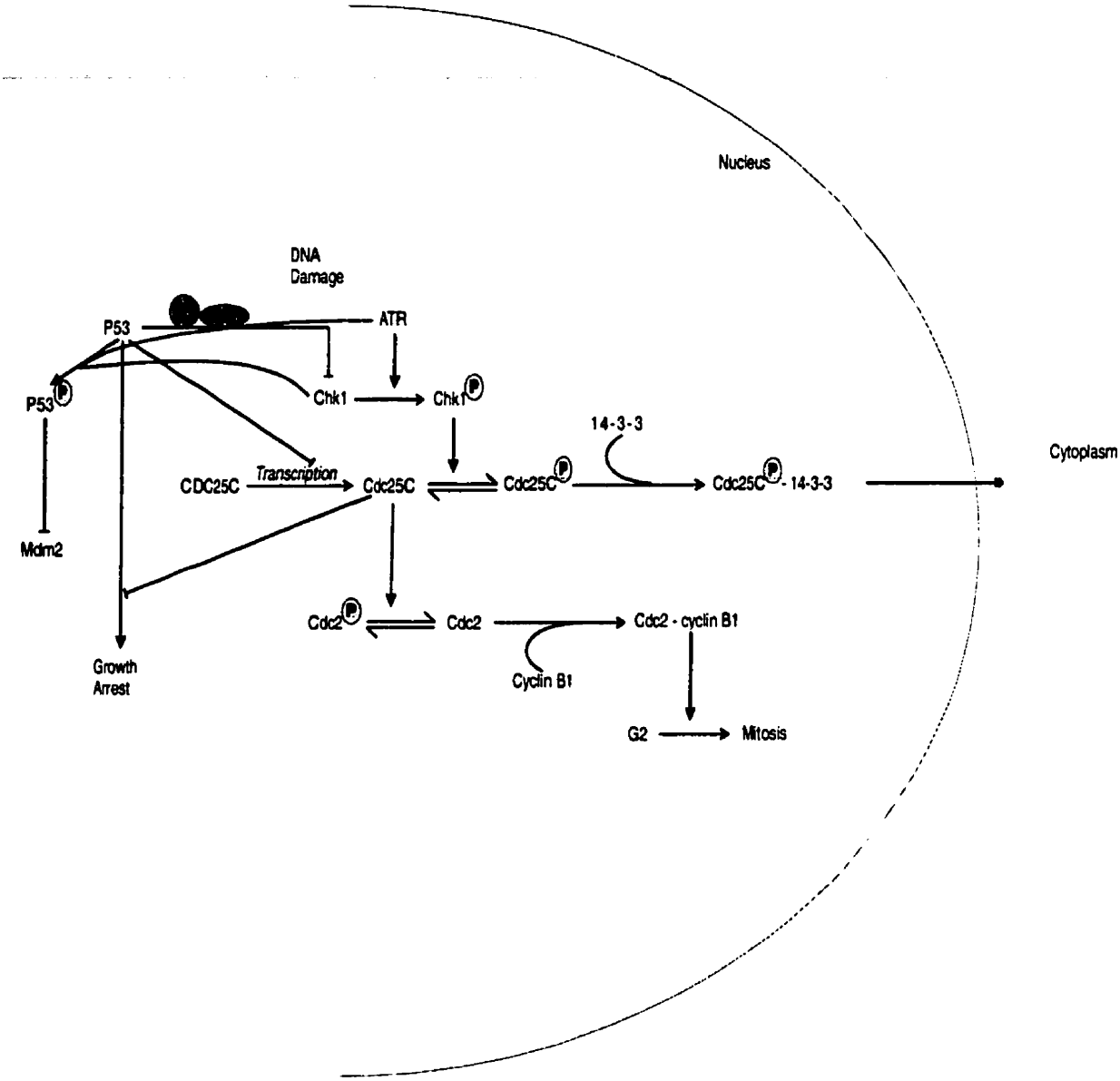


Figure 4.2.2-13. *Possible pathways outlining the participation of CDC25C in processes such as growth arrest, cellular proliferation, and apoptosis.*

In summary, all three genes were screened in 76 tumours. 29 tumours have a mutation in either one of the three genes, *CHK1*, *CDC25C* and *ATR*. 2 tumours (HC369, R88) have both *ATR* and *CDC25C* mutations; another 2 tumours (HC265, HC323) have both *ATR* and *CHK1*. 17 tumours have only *ATR* mutations, 1 has a *CHK1*, and 7 have single *CDC25C* mutations.

Another gene within our selection of genes is suggested to play an important role in mitotic-phase progression. It is the gene *MITOSIN*, a centromeric protein (CENPF) that is expressed throughout S, G2, and M phases of the cell cycle but is absent in G1 [298]. This gene was identified by direct binding to purified retinoblastoma protein [299]. Rb is an important protein in cell cycle regulation and thus interacting proteins are also given importance by association. *MITOSIN* showed a low mutation frequency of 3% (3/93) demonstrating that *MITOSIN* might not be involved in the progression of MSI colorectal cancers.

DP-2 is an E2F dimerization partner (DP). E2F functions as a tumour suppressor gene [300] and transcription factor that is the critical determinant of the G1/S phase transition during the mammalian cell cycle. It activates the transcription of a group of genes that encode proteins necessary for DNA replication, and is directly regulated by the action of the Rb protein, and indirectly through the action of the G1 cyclins and associated kinases [301]. Rb inhibits progression from G1 to S phase of the cell cycle, thus if DP-2 cannot heterodimerize with E2F, E2F may not be able to interact with Rb to arrest cells in G1. In determining if this was the case in our MSI-H tumours, the segment of the gene containing the A_8 repeat was examined for insertion/deletion mutations. Only 2 out of 94 tumours were mutated.

Through the years, research unveils novel functions of known genes, as is the case with Cdc7 kinase (*CDC7*). In a CDK-cyclin-like complex, Cdc7 requires its substrate, Dbf4. This complex is required for the initiation of DNA replication in eukaryotes and has been conserved in evolution. Other observations suggest that Cdc7-Dbf4 also plays another, less well characterized, role in checkpoint function and in the maintenance of genomic integrity [302]. The overexpression of the Cdc7-Dbf4 in neoplastic cells and tumours suggests that it may be an important early biomarker during cancer progression [303]. The genomic integrity of tumours with microsatellite instability is inherently comprised. The mutational analyses of the A9 repeat in *CDC7* showed that it was not targeted for mutation in our series of cancers and cell lines (0%).

The overexpression seen in other tumours indicates that perhaps Cdc7 is intact but the availability of Dbf4 is in fact deregulated.

CBF2 encodes for a CCAAT-box binding protein 2 and contributes to gene-specific transcriptional activation. If perhaps, CBF2 were mutated, the loss of this protein may inactivate important downstream genes of those that contain CCAAT-boxes. We found *CBF2* mutations in all of our subgroups, 14/91 (16%) in the MSI-H tumours, 3 out of 8 MSI-H cell lines (LS411, HCT-116, HCT-8) and also in one MSS cell line (SW948). The presence of this mutation in the MSS cell line suggests that perhaps this mutation is not specific to tumours with MSI phenotype but may occur in colorectal cancers sporadically.

GRK4 is a member of the G protein-coupled receptor kinases, a family of serine/threonine kinases that appear to play a role in receptor desensitization. They phosphorylate the activated form of the corresponding receptor, increasing its ability to interact with arrestins. Binding of the arrestins to the receptors then blocks further activation of the respective signal transduction pathways. Mutational analysis of the polyadenine (A_9) repeat within the coding region of *GRK4* revealed a 21% mutation frequency in our panel of primary tumours (19/91) and mutations as well in 2/8 MSI-H cell lines (LS411, HCT-116). Biallelic inactivation was detected in 2 tumours and also in the cell line, HCT-116. Further functional studies to investigate the loss of GRK4 in MSI-H cells would likely shed light on the exact consequence of these single base deletions with the mononucleotide element.

RECQL1 belongs to the RecQ family of DNA helicases along with its other member, the Bloom syndrome gene (*BLM*) described in Chapter 2 [304]. As mentioned earlier, DNA helicases are enzymes involved in various types of DNA repair, including mismatch repair, nucleotide excision repair, and direct repair. Mutations at the A_9 element are found in 11/93 (11%) of tumours and also in HCT-116. 3 tumours had both *BLM* and *RECQL1* mutations, the relevance of this data is unknown without further studies.

Protein kinase C (PKC) phosphorylates a variety of target proteins, which control growth and cellular differentiation. Thus *RACK7*, which encodes for a PKC binding protein [305], was a convincing candidate for mutational analysis. We found *RACK7* mutations in 17 out of 91 (19%) MSI-H tumours tested. This mutation frequency is one of the highest detected in this project yet none of the MSI-H cell lines were mutated, which is unusual. One the MSS cell lines (SW948) showed a frameshift mutation. This result indicates that further mutational analysis of MSS primary tumours might reveal a significant finding linking *RACK7* to colorectal cancer tumourigenesis. The analysis of corresponding normal tissue from each patient would verify this hypothesis.

The last group of genes screened, either are mutated at a very low frequency, too low to be detected in our small sample of MSI-H cell lines or are not mutated at all. The significance of these findings indicates that the mononucleotide tracts in these specific genes are not targeted for mutation; in other words, the loss of the resulting proteins is not advantageous to the abnormally proliferating cell or it is inactivated through another mechanism.

Within this group of genes, there are two that are implicated in other diseases, *ERCC5* and *BLYM*. *ERCC5* is identical to *XPG*, Xeroderma Pigmentosum G gene. It has been demonstrated that ERCC5 is a single-strand specific DNA endonuclease, which plays the catalytic role in nucleotide excision repair. The XPG nuclease (ERCC5) acts on the single-stranded region created as a result of the combined action of the XPB helicase and XPD helicase at the DNA damage site. The XPG endonuclease cleaves the damaged DNA strand 3-prime to the lesion during nucleotide excision repair [306]. This gene has been mapped to chromosome 13q33, a region lost in prostate cancers [307], however, studies reveal that ERCC5 is not the target gene within this locus [308]. 8/80 single base deletion mutations were observed in our series, further investigation into those individual patient tumours is required to lay importance on the intact ERCC5 protein.

The *BLYM* gene is mutated in 5 out of 96 (5%) tumours. *BLYM* is a transforming gene that is activated in Burkitt's lymphoma cells [309]. It is hypothesized that it may be involved in different stages of progression to neoplasia [310-312]. Through its sequence homology to transferrin it was also suggested that the transforming gene products might function via a pathway related to transferrin [313].

Human NP220 binds to a double-stranded DNA fragments by recognizing clusters of cytidines. NP220s are a novel type of nucleoplasmic protein with multiple domains [314]. A mutation frequency of 3% was observed, whereas the Mitogen-inducible gene (*MIG2*) was only mutated in 1%. The serine/threonine kinase, VRK2 was also found not to be a target in the microsatellite instability pathway as it was only mutated in 2% of cases.

Although the function of a particular gene may indicate involvement in the microsatellite instability pathway, this is not always the situation. DNA ligase III is thought to be involved in DNA repair and recombination. Through it's unique C terminal domain, it interacts with the DNA strand break repair protein, XRCC1 [315]. The $(A)_8$ repetitive element falls within the poly(ADP-ribose) polymerase and DNA ligase zinc finger domain, thus creating a opportunity to inactivate this encoded protein, and further promote genomic instability; however, our results indicate that DNA ligase III is only mutated in 1%.

FAP-1 is a tyrosine phosphatase that interacts with the cytosolic negative regulatory domain of Fas to block its apoptotic function. Fas is expressed in colonic epithelial cells and is also expressed in colon carcinomas and data suggests that its pathway is functional when Fas is expressed at high levels [316]. FAP-1 is a negative regulator of Fas-mediated apoptosis in human cancer cells and in connection with an additional signal-transducing molecule can possibly completely suppress Fas-mediated apoptosis [317]. Also, FAP-1 regulates the activity of both initiator caspases 3 and 8. Our failure to find any mutations in the polyadenine (A_8) repeat raises several questions. Is intact FAP-1 required for protection of the malignant cell? Is FAP-1

overexpressed in more aggressive cancers? What is the other signal-transducing molecule and is it the absence of this molecule that permits Fas-mediated apoptosis in microsatellite unstable tumour cells? Hence, the role of Fas signaling in the regulation of apoptosis in colon carcinoma cells and its role in influencing the response to treatment with chemotherapeutic agents is an avenue that should be further explored [317].

The mismatch repair genes have been screened by other groups and with varying mutation frequencies observed in MSI-H cancers. We found that our mutation frequencies of 42% and 29% in *hMSH3* and *hMSH6*, respectively, fell within the published ranges of 37-46% for *hMSH3* and 18-36% for hMSH6 [35,164,166,235,318]. Very few somatic mutations in *hPMS2* have been reported and we have similar results, with a 5% mutation frequency [85,319-321].

The quest for determining the genetic targets of microsatellite instability tumours can be simplified in the sense that genes with repetitive elements within the coding regions are suspect. However, the magnitude of the number of genes with these potential targets makes this a considerable project with both human and computer manipulations. The subgroup of genes investigated in this project is a small part of a larger more comprehensive study in mapping the genomic instability seen in colorectal cancers with microsatellite instability. Although measures are taken to verify the accuracy of the submitted sequences into the public databases, errors are inevitable, and time is spent characterizing genes with unforeseen exon-intron boundary problems leading to several false-positives.

The increase in slippage and decrease in proofreading as a function of increasing length of repetitive sequence leaves DNA mismatch repair as the major guardian against genome instability in repetitive DNA sequences. Thus, in tumours with mutations in the mismatch repair genes, microsatellites, and especially long stretches of mononucleotides are to say the least, at-risk-motifs (ARMS) [322]. Most microsatellites occur in non-coding and intronic regions, where these slippages would frequently occur without repair, since there should be no functional

consequence. The mutation frequencies in the intronic repeats did not display a relationship with the repeat length. The longer coding homonucleotide repeats are subjected to lowered proofreading efficiency of the frameshift intermediates [323-325]. The genes with stretches of nine adenines or longer in our subgroup of genes were mutated at a higher frequency in most cases. Within homonucleotide repeats, deletion repeats can be (but are not always higher) than addition rates, by factors of about 10-100-fold [272]. Our results exhibit similar trends, where the majority of mutations are single base deletions, followed by single base insertions and two-base pair IDLs.

Reasons why some genes that have repetitive elements do not exhibit these typical frameshift mutations;

1) They have shorter mononucleotide repeat tracts that are susceptible but not within the parameters of our study.
2) They have di-, trinucleotide elements that are susceptible, but are not within the parameters of our study.
3) Sequence context dependant repair by specific MMR genes, not *hMLH1* or *hMSH2*.
4) Functionality. It is advantageous to the cell to keep these specific genes intact.
5) Redundancy. Other targets in the same molecular pathway are targeted, negating the need for inactivation of the particular gene.

Although the tumours are microdissected, inevitably contaminating normal cells are included when genomic DNA is extracted. However, cell lines are usually derived from a clonal population, which should exhibit more homogeneity than the primary tumours. Screening genes initially in cell lines appeared to be an efficient way to quickly identify, susceptible genes with a high frequency of mutations. We found when there is a significant mutation frequency in the MSI tumours that the MSI cell lines are indicative of that result. Still our panel of MSI cell lines is limited, thus, increasing the number of these cell lines, including others with different

mutations, or different combinations of mutations in MMR genes would complement this screening strategy.

The pronounced genomic instability may increase susceptibility to apoptosis as a result of an accumulation of mutations in genes that are required for cell growth. In addition the inactivation of certain checkpoint proteins may initiate mitotic catastrophe. Genes involved in cell cycle controls and in particular checkpoints are becoming increasingly valuable in understanding how malignant cells evade them and continue to proliferate. These cells may initially be arrested in a damage checkpoint but they eventually override the arrest and re-enter the cell cycle, despite that they have not repaired the damage that elicited the arrest [326,322]. This adaptation may promote tumourigenesis by allowing genomic instability to occur in the absence of repair.

In our discussion of the effects of insertion/deletion mutations that cause frameshift mutations, we assumed that the resulting protein is inactive, due to the loss of functional domains. Woerner *et al.* 2001 describes possible situations where the mutant protein might exert trans-dominant functions on the remaining wild-type proteins. Also, these mutant proteins may interact with other proteins, thereby representing gain of function mutations instead of presuming that the frameshift mutations produce inactive proteins [327]. Therefore, further cDNA studies and functional studies are required to discover what happens to the truncated product of these genes with the insertion/deletion mutations.

---

[i] These CTNNB1 mutational analyses were performed by L.Primdahl[342].

Chapter Five

Conclusions

Colorectal cancer is the second leading cause of cancer in Canada, and attempts to reduce the burden of this disease have been a major force of health care providers and policy makers nationwide. The progression of colorectal cancer is one of the best understood of human neoplasms, with major targets identified such as *APC*, *KRAS*, and *TP53*. However, much more about this malignancy is left to be discovered. Two genomic destabilizing pathways have been described and are generally accepted; the Chromosomal Instability and the Microsatellite Instability pathways. The latter is defined by its deficiency in mismatch repair, due to the loss of one or more mismatch repair genes. Promoter methylation of hMLH1 and somatic inactivation of hMSH2 account for the majority of deficiencies seen in sporadic MSI-H colorectal cancers. Characteristic of this pathway is the hypermutability of microsatellites, especially stretches of mononucleotides. Transforming growth factor Beta-receptor II (*TGFβRII*), is the only gene thus far, whose mutations at its A_{10} tract are significant and specific for MSI colorectal tumours. Other genes with mononucleotide repeats are susceptible to mutation and have been previously described [3,35,77,153-155,164,166,167,219,235,328-332].

The increased susceptibility of mononucleotide repeats is due to slippages that occur during normal DNA replication. Given that tumours with deficient mismatch repair cannot detect/repair these insertion/deletion mutations, a frameshift mutation is propagated. In these MSI-H tumours, many genes are mutated and within a given individual, thousands of mutations may exist. Thus, the challenge of identifying the genes that are important in the distinctive pathogenesis of microsatellite instability is one that involves both breadth and depth. As a first line approach, a genome-wide search for genes with mononucleotide repeats enables the

researchers to narrow their focus by identifying possible candidate genes with susceptible targets of mutation.

We developed Kangaroo, an irreplaceable search tool in identifying genes with these susceptible mononucleotide repeats. Although Kangaroo can retrieve these results, further stratifications are required to ensure that the repetitive element, or searched sequence is completely within the coding region of the DNA and to verify the exon/intron boundaries. The latter is important when designing primers; ideally, the primers should also be within the same exonic region as the targeted segment. Kangaroo has eliminated the tedious process of researching individual entries because the results are displayed with hyperlinks to the relevant SeqHound (mirrored after NCBI GenBank Flatfile) (http://www.ncbi.nlm.nih.gov/) page. Perhaps, in order to reduce the number of genes from which novel genetic targets in carcinogenesis are chosen, the results obtained from Kangaroo could be filtered again, by comparing the results against other databases of known tumour suppressor genes and oncogenes. Kangaroo is a relatively simplistic program that has the power to perform many tasks with impressive results when it is properly manipulated.

Kangaroo's capabilities were tested through our searches to investigate human codon usage and to look for codon bias. Although the frequency of the single homogeneous (AAA, CCC, GGG, TTT) codon usage was not largely significant, the lack of these codons in succession was striking. If the trend towards increased GC content is supposed as the reason behind this bias, then codons with higher GC content are selected and rather than a passive system, where amino acid composition is unaffected, the trend could result in the alteration of the amino acids [247,333]. This would then imply that every codon and every amino acid follows a single trend determined by the overall compositional property of the genome. Kreitman suggests otherwise, where codon preference applies to every gene in a genome (thus, a passive system), yet the evolution of each amino acid and each protein is highly context-dependent [248,334].

Another hypothesis, describes the synonymous codon use between genes is varied by the differences in the efficiency of mismatch repair [335]. There is evidence that mismatches are repaired with different probabilities and biases depending on the nucleotides and the length of the mispair [336,337]. However, it was found that mismatch repair efficiency could only attribute a small fraction to synonymous codon usage [335].

Our results seem to suggest that there is an evolutionary thrust against runs of mononucleotides, possibly to avoid unnecessary insertion/deletion mutations. Perhaps by selection for error-minimization, alternative usage of synonymous codons which is invisible at the protein level is implemented without changing the amino acid [338]. Thus, subtleties at the codons that may prevent slippage-induced mutations do not affect the grand scheme, namely the amino acids and resulting protein as suggested by Kreitman. Studies by Linton *et al.*, also revealed a selection bias against long stretches of mononucleotides [136]. Additionally, their studies show evidence for slippage not only occurring during DNA replication but also during RNA transcription, resulting in insertion/deletions and insertions, respectively. If we consider that slippages can occur at different stages of the cell cycle, then it is conceivable that some of the deletion mutations caused by slippages during DNA replication can be reversed by insertions during RNA transcription.

Our project brings forth putative tumour suppressor genes that have sustained frameshift mutations probably induced during DNA replication. These findings form the basis for further investigations into the functional role of these alterations. According to the principles set out by Boland et al, 1998, the guidelines for determining tumour suppressor genes (see Background for list) are quite stringent and none of the novel genes screened in this project abide by those set standards. It is entirely possible that as with most of the other genes published to date, the mutated genes we identified are bystander genes, affected by the MSI phenotype of the tumour cell. However, it is also difficult to argue against the possible role of at least a subset of these

targets in MSI-H colorectal carcinogenesis. Some researchers suggest that an increase in the number of gene mutations decreases the latency period for tumourigenesis, indicating a quantitative effect, which is either additive or synergistic in nature [339]. There has been further evidence that perhaps, tumour suppressor genes are dosage dependent, that one hit of particular TSG are enough for genetic instability. The loss of the second allele may be an effect, rather than the cause of tumour growth [339]. In addition, the presence of wild type and mutated alleles does not necessarily exclude biallelic inactivation of putative TSGs, since genomic imprinting or promoter methylation of the wild type alleles might also contribute to inactivation of the second allele, as is seen in hMLH1. Functional studies with one or both of the alleles of the putative TSG knock out of MSI cell lines would reveal the effects of those cells to specific stimuli, such as growth stimulants/inhibitors, and apoptotic promoters/inhibitors. Through the examination of each stage of the adenoma-carcinoma sequence, the emergence of gene mutation can be used to decipher, if the mutation is an early or late event. An aberration occurring as an early event is of greater consequence than that of a later event. Also, different heterogeneous areas in a single tumour can determine successful clonal propagation. With all these combined efforts we aspire to build an intricate genetic model of MSI-H colorectal carcinogenesis.

Recently, Woerner *et al.*, (2001) also performed a systematic approach in order to find coding mono- and dinucleotide microsatellites in humans. They too utilized certain filters as to eliminate repeat tracts in pseudogenes, vector sequences, and homopolymeric nucleotide stretches at the most 5' or 3' ends of sequences. Exon/intron boundaries were determined by MALIGN analysis, and finally the true repeats were verified by direct sequencing of repeat candidates from genomic DNA. As mentioned in chapter 3, these alignment programs are inaccurate and we believe that the method, by which Kangaroo extracts only human annotated sequences from the original flatfile, is a more reliable approach in determining exon/intron

boundaries. Their results are on a small samples population with unknown biases, yet with similarity to our own results.

Most of the genes we screened were mutated in less than 10% in our MSI-H tumours. That is 14 of 24 genes (excluding the MMR genes), where the mononucleotide sequences are not susceptible to mutation. This could be due to a Passive system, whereby cells constitutively need these specific genes for basic cell function and proliferation, whether the cell be normal, or abnormal, in order to prevent death. On the other hand, in an Active system, cells would induce a DNA repair system like the mismatch repair system to correct the genes of crucial proteins required for propagation. There are still gaps in our knowledge of the mismatch repair system; are the MMR genes non-specific? Does the MMR machinery pass once, twice, or is there another back-up detection system? How do the MutS homologues decipher between the template and nascent strand in order to retain genomic integrity? From an evolutionary standpoint, have genes critical for cell survival acquired strategies against mutation or if mutation occurs, is it more efficiently repaired? We know that these repetitive elements cannot be mutated less frequently but can they be more efficiently repaired due to the context of the sequence? Perhaps by creating more stable loops, they are easily detected or do not allow the MMR system to bypass it.

The continuous assault on the genome affects the cell in innumerable ways given that carcinogenesis can occur along different pathways, with areas of redundancy. Thus, there is a problem in being able to discriminate within individual tumours, the true mutation profile; mutations with significant, but non-essential effects on tumour growth, and mutations with negligible effects only [340]. Irrelevant significant mutations can be determined by looking at the mRNA and protein level response. When a frameshift mutation occurs, a truncated protein emerges, but is this truncation enough to lose all protein functions, or is some or even the complete function retained? cDNA studies from mRNA from the tissue of the organ of interest,

colon would indicate if the protein is even expressed in the colon. If it is only slightly expressed or not at all, then a gene that is highly mutated but not expressed in the colon would be immaterial. It is difficult to disclose the effects of an individual gene mutation in a MSI tumour due the multitude of possible genetic alterations that the tumour has already acquired. Thus, the frequency of mutation of a specific gene to cause cellular changes is difficult, almost impossible to determine. For diagnosis purposes, a substantial mutation frequency that occurs early in tumour progression and that is specific for a patient population to provide precise and accurate results.

The cell cycle checkpoints surface as probably the most important in regulation of the cell cycle. Although MSI cells have a defective MMR system, the checkpoints should detect the damage and arrest the cell. When it is determined that the damage cannot be rectified, the checkpoint protein must initiate programmed cell death. In spite of this, MSI tumours have found a way to evade checkpoints by mutation of critical genes that arrest the cell cycle or that promote apoptosis. Our results implicate several genes involved in cell cycle checkpoints, *CHK1*, *CDC25C*, *ATR*, *DNA-PK* and the recQ helicases, *BLM* and *RECQL1*. The recQ helicases are a part of a sensor for DNA damage in S-phase as a member of BASC (BRCA1-Associated Surveillance Complex) [207]. RecQ helicases appear to associate with the replication machinery and to recruit signal-transducing kinases (perhaps DNA-PK) to the site of the lesion [208].

How is it that patients with MSI colorectal cancers have better prognosis and increased survival than those with MSS tumours? There is still debate about whether MSI patients respond better to chemotherapy than CSI patients [341] or whether chemotherapy has any effect at all. In the near future advanced studies will likely resolve this issue. Due to the heterogeneity of MSI tumours that acquire numerous mutations during tumourigenesis, chemotherapy is unlikely to succeed. A specific chemotherapeutic agent may be highly selective and toxic for MSI cell lines but the *in vivo* tumour might already contain-or rapidly develop- a population of cells that will be

resistant to the agent [90]. Gene replacement would also be unlikely to be effective since the cell now has an intact MMR system but carries the burden of pre-existing mutations. However, this treatment may slow down the tumourigenic potential of MSI tumours. This ultimately, is our goal: if not to cure, then to be able to treat the disease more effectively by tailoring the treatment. It is hoped that data from studies such as those described here will eventually contribute to the success of this research endeavour. Finally, further investigations into finding correlations between the mutation data and the corresponding clinical data may provide additional insight into whether there is a quantitative effect on individual tumours and whether sub-groups exist within the MSI pathway to advance our current knowledge of colorectal cancer.

APPENDIX A
Primer Sequences

Table A-1. *Primer sequences and Database Accession numbers Apoptotic Regulatory genes in Chapter 2.*

Gene	Genbank Accession	OMIM	Chr. Position		5' primer sequence (5'→3')		3' primer sequence (5'→3')
DNAPK –A8	NM_006904	600899	8q11	10771	GATGTAAGAGCTGAACTAGC	10827	GCTGCATACATTCTTTCATAC
A10				461	GTAAATTCTATGGAGAACTTGC	528	CACCCAATAATCCTAGGAGC
ATR	NM_001184	601215	3q22-24	2278	CTTCTGTCTGCAAGCCATTC	2322	GCAAGTTTTACTGGACTAGG
APAF1	NM_013229	602233	12q23		GTTACTTTTTTCCCTGTATTTAGAAAC		TATTCTCTGACCATCCTCAG
BLM	NM_000057	210900	15q26.1	1467	CCTTTGTAAGTAGCAACTGGGC	1582	TGTTTATTCTGATCTTTCACAGCAG
BCL10- T7	NM_03921	603517	1p22	432	TGAAAAACTGAGGGCATCCACTGT	529	CTAGAACAGGCAAATTCAGAGAAG
A8				101	CTGAGAGACATTTTGATCATCTACG	191	AAGTAGTCTAACAATTTTCCAGCCC
BAX	NM_007527	600040	19q13.3		ATCCAGGATCGAGCAGGGCG		ACTCGCTCAGCTTCTTGGTG

Table A-2. *Primer sequences of intronic repeats,*

Gene	5' primer sequence (5'→3')	3' primer sequence (5'→3')
FIP2	TGAGCCCAGGAGTTCAAGAC	TACAGGTGTGTGCCAATGTG
BLFZ1	TGGAGAGGAAAAGGCTGAAC	ATCCATCCATCTTTGCTCTTG
Beta Defensin-1	TCTGGGATACTGGGAGAAGG	ACATCGGGCTGAAATTATGG
Rh GLYCOPROTEIN	ACGAGGTCAGGTCAGGAGTTCGAGA	GGGCTACCATGTCCACCTAA
BLFZ1	GATGGCTCTTTGGTTTGGTG	GCTTTCTGGAATGGTTGGTG

Table A-3. *Primer Sequences and Database Information for the genes screened in primary colorectal tumours and cell lines in Chapter 4.*

Gene	Genbank Accession	OMIM	Chr. Position		5' primer sequence (5'→3')		3' primer sequence (5'→3')
MSH6	NM_000179	600678	2p16	3196	TATAGTCGAGGGGGTGATGG	3270	GATGGCGTGATCCTTTAAGC
MSH3	NM_002439	600887	5q11-12		AATCAAGCTGGATGATGCTG		CACTCCCACAATGCCAATAA
PMS2	NM_000535	600259	7p22	1219	AAGCAGGATCAATCCCCTTC	1346	AGAGGGCTCCTTCTTGGTTC
ERCC5 (XPG)	NM_000123	133530	13q33	2678	AGAATGGTGGCATGAAGCTC	2775	ACAGCTGGGTTAGGAAAGCC
BLYM	NM_005179	164830	1p32	39	ACTGCACTCCATACTGCATGAC	106	ACCCTTCAGGTCACTGAGAAG
CHK1 (A9,A7)	NM_001274	603078	11q24	146	TTTAGAATTGCCATGGGACC	174	CCAAGAAATCGGTACTCTTTCAC
DP2	NM_006286	602160	3q23	378	TGCTTTAAATGTGCTAATGGC	483	TCCGCCTCTGCTTCTCTATC
MIG-2	Z24725			515	TATCAGACACCCCGAAGAAC	609	GGCCCCTCTAATTCAAGTGC
RACK7 (PRKCBP1)	NM_012408		20q12	257	CGAGCCAGAAGACACAGAGG	360	GTGACGTGCTTTTCAGCTCC
TF-34	NM_003423	603972	19p13-p12	404	CCCAGAGCAAAGTATTTCAATG	514	GGTTAAAAGCTTTGCCACATTC
PROTEIN PHOSPHATASE 1 INHIBITOR	NM_006741			363	AGAAGTGGAGTCAAGGCTGG	455	GGTTGAGGGTTCTTTTGTGC
CCAAT BOX (CBF2)	NM_005760		2p	1340	AAATCAAATGGCTCTGTCCC	1465	GGGTATGCCCTATTCACACC
TC10	NM_012249		2	533	GATTGAAGACTGTTTTTGATGAGG	607	TAAACAACAGTTTATACATCTTGATCC
DNA LIGASE III	NM_013975	600940	17q11.2-12	121	TCAGAGTCTGGGGGTGATATG	213	CCTTCCAGCTCTGTGAGGTC
VRK2	NM_006296	602169	2p16-15	978	ACAAAAGGCTGCAACAAAGC	1115	GCACTTTCCATGTTGCACAG
MITOSIN	NM_016343	600236	1q32-41	5871	GAGAAACCAGCTTCGTGGAG	5987	CCTGAATGCAATGGAGACAC
PROTEIN TYROSINE PHOSPHATASE, TYPE 13 (PTPN13/ FAP-1)	NM_006264	600267	4q21.3	825	CCAGTTCAAAACTAGTGGCCC	909	TCAGCTGTACAAAGCAAGTCC
NP220	NM_014497		2p13	3867	TGGGGATGAGAAGACAGTGG	3979	TCTTTTCAGCCTTCTCTGTTCC
RECQL1	NM_002907	600537	12p12-11	83	TACGGAAAGGCAACAAGAGC	195	TCTTCTTTATTCCAAGCGGC
CDC25C	NM_001790	157680	5q31	650	GCCAGAGAACTTGAACAGGC	743	CCTTCCTGAGCTTTCCTTGG
CDC7	NM_003503	603311	1p22	10	TCTCAATTGGCTTGTGATGG	93	TGCAAGTTTAAAATTCTGCTCG
G PROTEIN-COUPLED RECEPTOR KINASE (GRK4)	HSGRKG08		4p16.3	601	GTTTGCGCCTGTCAAGTGCG	677	GAGCCATAGCTTCACCTTTCC
AXIN2	NM_004655	604025	17q23-24	Intron	AACCCAGTTTCTTTCCTTCT	2027	GGTCCTGGGTGAACAGGTG
β-CATENIN	NM_001904	116806	3p21-22		AGTCACTGGCAGCAACAGTC		TCTTCCTCAGGATTGCCTT

Table A-4. *Primer sequences for genes screened exclusively in cell lines.*

Gene	Genbank Accession	OMIM	Chr. Position		5' primer sequence (5'→3')		3' primer sequence (5'→3')
DAXX	NM_001350	603186	6p21.3	2166	AGTGTGGCCACACAATGCG	2204	CTAATCAGAGTCTGAGAGC
Silencer of Death Domain-8A (BAG4)	NM_004874	603884		1321	GCTGTTTGTAAGATTCAGGC	1377	TCGACCTTTGTTCTAAATCC
BAG4-7A				1093	CTTCTGAAGAATGTGTACC	1158	AAGATACTGGACCTTCTCCAGC
Caspase 8 associated protein 2 (CASP8AP2/ RIP25) -7.1	AF164678		6	2749	GACAAATGTACAGAAGCAG	2869	GCACATCAGCTGACACTCTACG
CASP8AP2-7.2				3761	TGGAAGTACACTGTCCAAGC	3805	TGCATCCTCTATGCTGCTTCC
CASP8AP2-7.3				5392	ACCACAGAATCTCCCAG	5441	ACAATTTGATCCTGGCTCTG
CASP8AP2-8.0				3669	CTCCAGTATTAACACTGTAAAGC	3738	AGCGCTCTTCGGATTGAGAG
CASP8AP2-9.0				3336	GAATGGCATAGTTGATCG	3395	TGGTCATCTACAAACTTCC
BASS1-7.0	AF139897			571	TTCGGATGAAGAGGAAGAAG	650	TGCTGCAGCAATATTCAG
BASS1-8.0				913	TGATCTTGAAGCCCATGAGC	991	TGGTGTGAGTTGGCTGATGG
Cyclin E binding protein (CEB)	NM_016323			440	GCTGATTCTCTCATCAGATGG	549	GAGTGCAAGAGAATGGTAATCTCC
MTK1	NM_005922	602425	6	439	CGAGCAGCTCTTAGAACAAC	546	TGAGGTAAGGCTTATTGAGATCC
SHC	NM_003029	600560	1q21		GGGATGAGGAGGAGGAAGAG		TCCCGAAGCCTCATGTCTAC
MOK	NM_014226	605762	14q32		CGAGGGAGAAGATACCCATTA		GCCAAAGTCCCCTAATTTCAG
RAD50	HSRAD50	604040	5q31	2744	AGAGCAGGTAAGCCCTTTGG	2864	TCCTGTGCTATTTTGTTGCTTG
				2128	AACTGCGACTTGCTCCAGAT	2195	CAAGTCCCAGCATTTCATCA
p72	NM_006386			51	ATTTGGAGCAAGCGGTGGTG	130	TTGGGGAGCTCACTCAAATC
Guanine Nucleotide Exchange Factor 1 (GNEF) – 7.0	NM_016340			1619	GGTGACCCTGCTATGACTCG	1686	GCAGCACAGGCAATATTCAA
GNEF-9.0				2021	GGTGTTCCTCATATTCCCAAAA	2084	CAATATCTCCTGGCACATGCT
REV1	NM_016316	606134	2q11.1	3194	GGCAGCAGTGAAAGAAAAGAA	3261	GGCAGAGTTTTTGCAGGACT

APPENDIX B
PCR ANNEALING TEMPERATURES

Table B-5. *The annealing temperatures for the sets of primers for the genes screened for mutational analysis in this project.*

Gene	Repeat	Annealing Temperature (°C)
ATR		56
AXIN2		60
BAG4	A8	52
	A7	52
BASS1		54
BCL10	A8	57
	T7	57
BLFZ1	A7-INTRON	52
BLFZ1	A10-INTRON	54
BLM		54
BLYM		56
CASP8AP2	7.1	52
	7.3	52
	8.0	50
	9.0	52
CBF2		56
CDC7		50
CDC25		54
CEB		46
CHK1		48
DAXX		52
DNA-PK	A8	56
	A10	57
DP-2		55
ERCC5		54
FAP-1		58
FIP2	A8-INTRON	52
GNEF	A7	54
	A9	54
GRK4		56
HOXA11		50
MIG-2		52
MITOSIN		54
MOK		54
MSH3		50
MSH6		50
MTK1		52
NP220		60
p72		54
PP-1 INH		54
PMS2		52
RACK7		56
RAD50	A8	52
	A9	52
RECQL1		58
REV1		54
RhAG	A10-INTRON	50
SHC		52
TC10		54
TF-34		59

TFE3		54
VRK2		52
β-DEFENSIN	A8-INTRON	52

Reference List

1. Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W., & Vogelstein, B. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability [see comments]. *Science* **268**, 1336-1338 (1995).

2. Souza, R. F., Appel, R., Yin, J., Wang, S., Smolinski, K. N., Abraham, J. M., Zou, T. T., Shi, Y. Q., Lei, J., Cottrell, J., Cymes, K., Biden, K., Simms, L., Leggett, B., Lynch, P. M., Frazier, M., Powell, S. M., Harpaz, N., Sugimura, H., Young, J., & Meltzer, S. J. Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. *Nat.Genet.* **14**, 255-257 (1996).

3. Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J. C., & Perucho, M. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**, 967-969 (1997).

4. Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D. E., Ruggeri, B. A., Klein-Szanto, A. J., Testa, J. R., Neri, G., & Bellacosa, A. The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nat.Genet.* **23**, 266-268 (1999).

5. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N., & Srivastava, S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248-5257 (1998).

6. Rudin, C. M. & Thompson, C. B. Apoptosis and Cancer in *The Genetic Basis of Human Cancer* 193-204. McGraw-Hill Companies Inc., (1998).

7. Folkman, J. Tumor angiogensis: role in regulation of tumor growth. *Symp.Soc.Dev.Biol.* **30**, 43-52 (1974).

8. de Lange, T. & Jacks, T. For better or worse? Telomerase inhibition and cancer. *Cell* **98**, 273-275 (1999).

9. Loeb, L. A. A mutator phenotype in cancer. *Cancer Res.* **61** , 3230-3239 (2001).

10. Muto, T., Bussey, H. J., & Morson, B. C. The evolution of cancer of the colon and rectum. *Cancer* **36**, 2251-2270 (1975).

11. Jackson, A. L. & Loeb, L. A. The mutation rate and cancer. *Genetics* **148**, 1483-1490 (1998).

12. Tomlinson, I. & Bodmer, W. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat.Med.* **5**, 11-12 (1999).

13. Cahill, D. P., Kinzler, K. W., Vogelstein, B., & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* **9**, M57-M60 (1999).

14. Loeb, K. R. & Loeb, L. A. Significance of multiple mutations in cancer. *Carcinogenesis* **21**, 379-385 (2000).

15. Perucho, M. Cancer of the microsatellite mutator phenotype. *Biol.Chem.* **377**, 675-684 (1996).

16. Park, M. Oncogenes in *The Genetic Basis of Human Cancer* 205-228. McGraw-Hill Companies Inc., (1998).

17. Knudson, A. G., Jr. Mutation and cancer in man. *Cancer* **39**, 1882-1886 (1977).

18. Knudson, A. G., Jr. Retinoblastoma: a prototypic hereditary neoplasm. *Semin.Oncol.* **5**, 57-60 (1978).

19. Morin, P. J., Vogelstein, B., & Kinzler, K. W. Apoptosis and APC in colorectal tumorigenesis. *Proc.Natl.Acad.Sci.U.S.A* **93**, 7950-7954 (1996).

20. Berneburg, M. & Lehmann, A. R. Xeroderma pigmentosum and related disorders: defects in DNA repair and transcription. *Adv.Genet.* **43**, 71-102 (2001).

21. Bischof, O., Galande, S., Farzaneh, F., Kohwi-Shigematsu, T., & Campisi, J. Selective cleavage of BLM, the bloom syndrome protein, during apoptotic cell death. *J.Biol.Chem.* **276**, 12068-12075 (2001).

22. Savitsky, K., Sfez, S., Tagle, D. A., Ziv, Y., Sartiel, A., Collins, F. S., Shiloh, Y., & Rotman, G. The complete sequence of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species. *Hum.Mol.Genet.* **4**, 2025-2032 (1995).

23. Redston, M. Carcinogenesis in the gi tract: from morphology to genetics and back again. *Mod.Pathol.* **14**, 236-245 (2001).

24. Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M., & Bos, J. L. Genetic alterations during colorectal-tumor development. *N.Engl.J.Med.* **319**, 525-532 (1988).

25. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159-170 (1996).

26. Heinen, C. D., Shivapurkar, N., Tang, Z., Groden, J., & Alabaster, O. Microsatellite instability in aberrant crypt foci from human colons. *Cancer Res.* **56**, 5339-5341 (1996).

27. Pretlow, T. P., ORiordan, M. A., Pretlow, T. G., & Stellato, T. A. Aberrant crypts in human colonic mucosa: putative preneoplastic lesions. *J.Cell Biochem.Suppl* **16G**, 55-62 (1992).

28. Roncucci, L. Early events in human colorectal carcinogenesis. Aberrant crypts and microadenoma. *Ital.J.Gastroenterol.* **24**, 498-501 (1992).

29. Shpitz, B., Hay, K., Medline, A., Bruce, W. R., Bull, S. B., Gallinger, S., & Stern, H. Natural history of aberrant crypt foci. A surgical approach. *Dis.Colon Rectum* **39**, 763-767 (1996).

30. Takayama, T., Katsuki, S., Takahashi, Y., Ohi, M., Nojiri, S., Sakamaki, S., Kato, J., Kogawa, K., Miyake, H., & Niitsu, Y. Aberrant crypt foci of the colon as precursors of adenoma and cancer. *N.Engl.J.Med.* **339**, 1277-1284 (1998).

31. Powell, S. M., Zilz, N., Beazer-Barclay, Y., Bryan, T. M., Hamilton, S. R., Thibodeau, S. N., Vogelstein, B., & Kinzler, K. W. APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235-237 (1992).

32. Shibata, D., Schaeffer, J., Li, Z. H., Capella, G., & Perucho, M. Genetic heterogeneity of the c-K-ras locus in colorectal adenomas but not in adenocarcinomas. *J.Natl.Cancer Inst.* **85**, 1058-1063 (1993).

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

34. Jass, J. R., Biden, K. G., Cummings, M. C., Simms, L. A., Walsh, M., Schoch, E., Meltzer, S. J., Wright, C., Searle, J., Young, J., & Leggett, B. A. Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. *J.Clin.Pathol.* **52**, 455-460 (1999).

35. Johannsdottir, J. T., Jonasson, J. G., Bergthorsson, J. T., Amundadottir, L. T., Magnusson, J., Egilsson, V., & Ingvarsson, S. The effect of mismatch repair deficiency on tumourigenesis; microsatellite

instability affecting genes containing short repeated sequences. *Int.J.Oncol.2000.Jan.;16.(1.):133.-9.* **16**, 133-139 (0 AD/1).

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

37. Haber, D. A. & Fearon, E. R. The promise of cancer genetics. *Lancet* **351 Suppl 2**, SII1-SII8 (1998).

38. Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323-330 (1995).

39. Sherr, C. J. Cancer cell cycles. *Science* **274**, 1672-1677 (1996).

40. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., & Kinzler, K. W. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**, 1787-1790 (1997).

41. Cahill, D. P., Lengauer, C., Yu, J., Riggins, G. J., Willson, J. K., Markowitz, S. D., Kinzler, K. W., & Vogelstein, B. Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300-303 (1998).

42. Lengauer, C., Kinzler, K. W., & Vogelstein, B. DNA methylation and genetic instability in colorectal cancer cells. *Proc.Natl.Acad.Sci.U.S.A* **94**, 2545-2550 (1997).

43. Bergman, Y. & Mostoslavsky, R. DNA demethylation: turning genes on. *Biol.Chem.* **379**, 401-407 (1998).

44. Bomme, L., Bardi, G., Pandis, N., Fenger, C., Kronborg, O., & Heim, S. Cytogenetic analysis of colorectal adenomas: karyotypic comparisons of synchronous tumors. *Cancer Genet.Cytogenet.* **106**, 66-71 (1998).

45. Fukasawa, K., Choi, T., Kuriyama, R., Rulong, S., & Vande Woude, G. F. Abnormal centrosome amplification in the absence of p53. *Science* **271**, 1744-1747 (1996).

46. Nagase, H. & Nakamura, Y. Mutations of the APC (adenomatous polyposis coli) gene. *Hum.Mutat.* **2**, 425-434 (1993).

47. Chung, D. C. The genetic basis of colorectal cancer: insights into critical pathways of tumorigenesis. *Gastroenterology* **119**, 854-865 (2000).

48. Smith, A. J., Stern, H. S., Penner, M., Hay, K., Mitri, A., Bapat, B. V., & Gallinger, S. Somatic APC and K-ras codon 12 mutations in aberrant crypt foci from human colons. *Cancer Res.* **54**, 5527-5530 (1994).

49. Rowan, A. J., Lamlum, H., Ilyas, M., Wheeler, J., Straub, J., Papadopoulou, A., Bicknell, D., Bodmer, W. F., & Tomlinson, I. P. APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits". *Proc.Natl.Acad.Sci.U.S.A* **97**, 3352-3357 (2000).

50. Fearnhead, N. S., Britton, M. P., & Bodmer, W. F. The abc of apc. *Hum.Mol.Genet.* **10**, 721-733 (2001).

51. Baker, S. J., Preisinger, A. C., Jessup, J. M., Paraskeva, C., Markowitz, S., Willson, J. K., Hamilton, S., & Vogelstein, B. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res.* **50**, 7717-7722 (1990).

52. Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B., & Craig, R. W. Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* **51**, 6304-6311 (1991).

53. Thiagalingam, S., Lengauer, C., Leach, F. S., Schutte, M., Hahn, S. A., Overhauser, J., Willson, J. K., Markowitz, S., Hamilton, S. R., Kern, S. E., Kinzler, K. W., & Vogelstein, B. Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nat.Genet.* **13**, 343-346 (1996).

54. Tanaka, K., Oshimura, M., Kikuchi, R., Seki, M., Hayashi, T., & Miyaki, M. Suppression of tumorigenicity in human colon carcinoma cells by introduction of normal chromosome 5 or 18. *Nature* **349**, 340-342 (1991).

55. Goyette, M. C., Cho, K., Fasching, C. L., Levy, D. B., Kinzler, K. W., Paraskeva, C., Vogelstein, B., & Stanbridge, E. J. Progression of colorectal cancer is associated with multiple tumor suppressor gene defects but inhibition of tumorigenicity is accomplished by correction of any single defect via chromosome transfer. *Mol.Cell Biol.* **12**, 1387-1395 (1992).

56. Shibata, D., Reale, M. A., Lavin, P., Silverman, M., Fearon, E. R., Steele, G., Jr., Jessup, J. M., Loda, M., & Summerhayes, I. C. The DCC protein and prognosis in colorectal cancer. *N.Engl.J.Med.* **335**, 1727-1732 (1996).

57. Massague, J. TGF-beta signal transduction. *Annu.Rev.Biochem.* **67**, 753-791 (1998).

58. Takagi, Y., Kohmura, H., Futamura, M., Kida, H., Tanemura, H., Shimokawa, K., & Saji, S. Somatic alterations of the DPC4 gene in human colorectal cancers in vivo. *Gastroenterology* **111**, 1369-1372 (1996).

59. Eppert, K., Scherer, S. W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, H., Tsui, L. C., Bapat, B., Gallinger, S., Andrulis, I. L., Thomsen, G. H., Wrana, J. L., & Attisano, L. MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**, 543-552 (1996).

60. Miyaki, M., Iijima, T., Konishi, M., Sakai, K., Ishii, A., Yasuno, M., Hishima, T., Koike, M., Shitara, N., Iwama, T., Utsunomiya, J., Kuroki, T., & Mori, T. Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene* **18**, 3098-3103 (1999).

61. Lengauer, C., Kinzler, K. W., & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643-649 (1998).

62. Strachan, T. & Read, A. P. *Human Molecular Genetics*Bios Scientific Publishers Limited,(1996).

63. Hoang, J. M., Cottu, P. H., Thuille, B., Salmon, R. J., Thomas, G., & Hamelin, R. BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines. *Cancer Res.* **57**, 300-303 (1997).

64. Aaltonen, L. A., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomaki, P., Chadwick, R. B., Kaariainen, H., Eskelinen, M., Jarvinen, H., Mecklin, J. P., & de la, C. A. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N.Engl.J.Med.* **338**, 1481-1487 (1998).

65. Samowitz, W. S., Slattery, M. L., Potter, J. D., & Leppert, M. F. BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. *Am.J.Pathol.* **154**, 1637-1641 (1999).

66. Gryfe, R., Kim, H., Hsieh, E. T., Aronson, M. D., Holowaty, E. J., Bull, S. B., Redston, M., & Gallinger, S. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer [see comments]. *N.Engl.J.Med.2000.Jan.13.;342.(2.):69.-77.* **342**, 69-77 (0 AD/1/13).

67. Zhou, X. P., Hoang, J. M., Li, Y. J., Seruca, R., Carneiro, F., Sobrinho-Simoes, M., Lothe, R. A., Gleeson, C. M., Russell, S. E., Muzeau, F., Flejou, J. F., Hoang-Xuan, K., Lidereau, R., Thomas, G., & Hamelin, R. Determination of the replication error phenotype in human tumors without the requirement for matching normal DNA by analysis of mononucleotide repeat microsatellites. *Genes Chromosomes Cancer* **21**, 101-107 (1998).

68. Pyatt, R., Chadwick, R. B., Johnson, C. K., Adebamowo, C., de la, C. A., & Prior, T. W. Polymorphic variation at the BAT-25 and BAT-26 loci in individuals of African origin. Implications for microsatellite instability testing. *Am.J.Pathol.* **155**, 349-353 (1999).

69. Konishi, M., Kikuchi-Yanoshita, R., Tanaka, K., Muraoka, M., Onda, A., Okumura, Y., Kishi, N., Iwama, T., Mori, T., Koike, M., Ushio, K., Chiba, M., Nomizu, S., Konishi, F., Utsunomiya, J., & Miyaki, M. Molecular nature of colon tumors in hereditary nonpolyposis colon cancer, familial polyposis, and sporadic colon cancer. *Gastroenterology* **111**, 307-317 (1996).

70. Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., & Issa, J. P. CpG island methylator phenotype in colorectal cancer. *Proc.Natl.Acad.Sci.U.S.A* **96**, 8681-8686 (1999).

71. Peltomaki, P. & Vasen, H. F. Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology* **113**, 1146-1158 (1997).

72. Buermeyer, A. B., Deschenes, S. M., Baker, S. M., & Liskay, R. M. Mammalian DNA mismatch repair. *Annu.Rev.Genet.* **33**, 533-564 (1999).

73. Peltomaki, P. & de la, C. A. Mutations predisposing to hereditary nonpolyposis colorectal cancer. *Adv.Cancer Res.* **71**, 93-119 (1997).

74. Lindblom, A., Tannergard, P., Werelius, B., & Nordenskjold, M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat.Genet.* **5**, 279-282 (1993).

75. Pedroni, M., Sala, E., Scarselli, A., Borghi, F., Menigatti, M., Benatti, P., Percesepe, A., Rossi, G., Foroni, M., Losi, L., Di Gregorio, C., De Pol, A., Nascimbeni, R., Di Betta, E., Salerni, B., de Leon, M. P., & Roncucci, L. Microsatellite instability and mismatch-repair protein expression in hereditary and sporadic colorectal carcinogenesis. *Cancer Res.* **61**, 896-899 (2001).

76. Fodde, R., Kuipers, J., Rosenberg, C., Smits, R., Kielman, M., Gaspar, C., van Es, J. H., Breukel, C., Wiegant, J., Giles, R. H., & Clevers, H. Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat.Cell Biol.* **3**, 433-438 (2001).

77. Fujiwara, T., Stolker, J. M., Watanabe, T., Rashid, A., Longo, P., Eshleman, J. R., Booker, S., Lynch, H. T., Jass, J. R., Green, J. S., Kim, H., Jen, J., Vogelstein, B., & Hamilton, S. R. Accumulated clonal genetic alterations in familial and sporadic colorectal carcinomas with widespread instability in microsatellite sequences. *Am.J.Pathol.* **153**, 1063-1078 (1998).

78. Olschwang, S., Hamelin, R., Laurent-Puig, P., Thuille, B., De Rycke, Y., Li, Y. J., Muzeau, F., Girodet, J., Salmon, R. J., & Thomas, G. Alternative genetic pathways in colorectal carcinogenesis. *Proc.Natl.Acad.Sci.U.S.A* **94**, 12122-12127 (1997).

79. Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S., & Perucho, M. Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat.Genet.* **6**, 273-281 (1994).

80. Homfray, T. F., Cottrell, S. E., Ilyas, M., Rowan, A., Talbot, I. C., Bodmer, W. F., & Tomlinson, I. P. Defects in mismatch repair occur after APC mutations in the pathogenesis of sporadic colorectal tumours. *Hum.Mutat.* **11**, 114-120 (1998).

81. Young, J., Leggett, B., Gustafson, C., Ward, M., Searle, J., Thomas, L., Buttenshaw, R., & Chenevix-Trench, G. Genomic instability occurs in colorectal carcinomas but not in adenomas. *Hum.Mutat.* **2**, 351-354 (1993).

82. Stoler, D. L., Chen, N., Basik, M., Kahlenberg, M. S., Rodriguez-Bigas, M. A., Petrelli, N. J., & Anderson, G. R. The onset and extent of genomic instability in sporadic colorectal tumor progression [see comments]. *Proc.Natl.Acad.Sci.U.S.A.* **96**, 15121-15126 (1999).

83. Kolodner, R. D. Mismatch repair: mechanisms and relationship to cancer susceptibility. *Trends Biochem.Sci.* **20**, 397-401 (1995).

84. Ku, J. L., Yoon, K. A., Kim, D. Y., & Park, J. G. Mutations in hMSH6 alone are not sufficient to cause the microsatellite instability in colorectal cancer cell lines. *Eur.J.Cancer* **35**, 1724-1729 (1999).

85. Nicolaides, N. C., Littman, S. J., Modrich, P., Kinzler, K. W., & Vogelstein, B. A naturally occurring hPMS2 mutation can confer a dominant negative mutator phenotype. *Mol.Cell Biol.* **18**, 1635-1641 (1998).

86. Marcus, V. A., Madlensky, L., Gryfe, R., Kim, H., So, K., Millar, A., Temple, L. K., Hsieh, E., Hiruki, T., Narod, S., Bapat, B. V., Gallinger, S., & Redston, M. Immunohistochemistry for hMLH1 and hMSH2: a practical test for DNA mismatch repair-deficient tumors. *Am.J.Surg.Pathol.* **23**, 1248-1255 (1999).

87. Kane, M. F., Loda, M., Gaida, G. M., Lipman, J., Mishra, R., Goldman, H., Jessup, J. M., & Kolodner, R. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res.* **57**, 808-811 (1997).

88. Veigl, M. L., Kasturi, L., Olechnowicz, J., Ma, A. H., Lutterbaugh, J. D., Periyasamy, S., Li, G. M., Drummond, J., Modrich, P. L., Sedwick, W. D., & Markowitz, S. D. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc.Natl.Acad.Sci.U.S.A* **95**, 8698-8702 (1998).

89. Esteller, M., Levine, R., Baylin, S. B., Ellenson, L. H., & Herman, J. G. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene* **17**, 2413-2417 (1998).

90. Jiricny, J. & Nystrom-Lahti, M. Mismatch repair defects in cancer. *Curr.Opin.Genet.Dev.* **10**, 157-161 (2000).

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

92. Esteller, M., Corn, P. G., Baylin, S. B., & Herman, J. G. A gene hypermethylation profile of human cancer. *Cancer Res.* **61**, 3225-3229 (2001).

93. Drummond, J. T., Li, G. M., Longley, M. J., & Modrich, P. Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* **268**, 1909-1912 (1995).

94. Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J. J., & Jiricny, J. GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. *Science* **268**, 1912-1914 (1995).

95. Fujii, H. & Shimada, T. Isolation and characterization of cDNA clones derived from the divergently transcribed gene in the region upstream from the human dihydrofolate reductase gene. *J.Biol.Chem.* **264**, 10057-10064 (1989).

96. Linton, J. P., Yen, J. Y., Selby, E., Chen, Z., Chinsky, J. M., Liu, K., Kellems, R. E., & Crouse, G. F. Dual bidirectional promoters at the mouse dhfr locus: cloning and characterization of two mRNA classes of the divergently transcribed Rep-1 gene. *Mol.Cell Biol.* **9**, 3058-3072 (1989).

97. Risinger, J. I., Umar, A., Boyd, J., Berchuck, A., Kunkel, T. A., & Barrett, J. C. Mutation of MSH3 in endometrial cancer and evidence for its functional role in heteroduplex repair. *Nat.Genet.* **14**, 102-105 (1996).

98. Marra, G. & Schar, P. Recognition of DNA alterations by the mismatch repair system. *Biochem.J.* **338**, 1-13 (1999).

99. Fishel, R. & Wilson, T. MutS homologs in mammalian cells. *Curr.Opin.Genet.Dev.* **7**, 105-113 (1997).

100. Umar, A., Risinger, J. I., Glaab, W. E., Tindall, K. R., Barrett, J. C., & Kunkel, T. A. Functional overlap in mismatch repair by human MSH3 and MSH6. *Genetics* **148**, 1637-1646 (1998).

101. Li, G. M. & Modrich, P. Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. *Proc.Natl.Acad.Sci.U.S.A* **92**, 1950-1954 (1995).

102. Lipkin, S. M., Wang, V., Jacoby, R., Banerjee-Basu, S., Baxevanis, A. D., Lynch, H. T., Elliott, R. M., & Collins, F. S. MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat.Genet.2000.Jan.;24.(1.):27.-35.* **24**, 27-35 .

103. Harfe, B. D. & Jinks-Robertson, S. Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in Saccharomyces cerevisiae. *Genetics* **156**, 571-578 (2000).

104. Prolla, T. A. DNA mismatch repair and cancer. *Curr.Opin.Cell Biol.* **10**, 311-316 (1998).

105. Baker, S. M., Harris, A. C., Tsao, J. L., Flath, T. J., Bronner, C. E., Gordon, M., Shibata, D., & Liskay, R. M. Enhanced intestinal adenomatous polyp formation in Pms2-/-;Min mice. *Cancer Res.* **58**, 1087-1089 (1998).

106. Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P., & Jiricny, J. Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. *J.Biol.Chem.* **274**, 32368-32375 (1999).

107. Leung, W. K., Kim, J. J., Wu, L., Sepulveda, J. L., & Sepulveda, A. R. Identification of a second MutL DNA mismatch repair complex (hPMS1 and hMLH1) in human epithelial cells. *J.Biol.Chem.* **275**, 15728-15732 (2000).

108. Aquilina, G. & Bignami, M. Mismatch repair in correction of replication errors and processing of DNA damage. *J.Cell Physiol* **187**, 145-154 (2001).

109. Umar, A. & Kunkel, T. A. DNA-replication fidelity, mismatch repair and genome instability in cancer cells. *Eur.J.Biochem.* **238**, 297-307 (1996).

110. Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., Golemis, E. A., Genuardi, M., & Neri, G. MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc.Natl.Acad.Sci.U.S.A* **96**, 3969-3974 (1999).

111. Johnson, R. E., Kovvali, G. K., Guzder, S. N., Amin, N. S., Holm, C., Habraken, Y., Sung, P., Prakash, L., & Prakash, S. Evidence for involvement of yeast proliferating cell nuclear antigen in DNA mismatch repair. *J.Biol.Chem.* **271**, 27987-27990 (1996).

112. Gradia, S., Subramanian, D., Wilson, T., Acharya, S., Makhov, A., Griffith, J., & Fishel, R. hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Mol.Cell* **3**, 255-261 (1999).

113. Tran, H. T., Gordenin, D. A., & Resnick, M. A. The 3'-->5' exonucleases of DNA polymerases delta and epsilon and the 5'-->3' exonuclease Exo1 have major roles in postreplication mutation avoidance in Saccharomyces cerevisiae. *Mol.Cell Biol.* **19**, 2000-2007 (1999).

114. Kleczkowska, H. E., Marra, G., Lettieri, T., & Jiricny, J. hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev.* **15**, 724-736 (2001).

115. Umar, A., Buermeyer, A. B., Simon, J. A., Thomas, D. C., Clark, A. B., Liskay, R. M., & Kunkel, T. A. Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* **87**, 65-73 (1996).

116. Lynch, H. T., Smyrk, T., McGinn, T., Lanspa, S., Cavalieri, J., Lynch, J., Slominski-Castor, S., Cayouette, M. C., Priluck, I., & Luce, M. C. Attenuated familial adenomatous polyposis (AFAP). A phenotypically and genotypically distinctive variant of FAP. *Cancer* **76**, 2427-2433 (1995).

117. Stoler, D. L., Chen, N., Basik, M., Kahlenberg, M. S., Rodriguez-Bigas, M. A., Petrelli, N. J., & Anderson, G. R. The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proc.Natl.Acad.Sci.U.S.A* **96**, 15121-15126 (1999).

118. Tsao, J. L., Yatabe, Y., Salovaara, R., Jarvinen, H. J., Mecklin, J. P., Aaltonen, L. A., Tavare, S., & Shibata, D. Genetic reconstruction of individual colorectal tumor histories. *Proc.Natl.Acad.Sci.U.S.A* **97**, 1236-1241 (2000).

119. Soengas, M. S., Capodieci, P., Polsky, D., Mora, J., Esteller, M., Opitz-Araya, X., McCombie, R., Herman, J. G., Gerald, W. L., Lazebnik, Y. A., Cordon-Cardo, C., & Lowe, S. W. Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature* **409**, 207-211 (2001).

120. Latham, K. E. X chromosome imprinting and inactivation in the early mammalian embryo. *Trends Genet.* **12**, 134-138 (1996).

121. Ahuja, N., Li, Q., Mohan, A. L., Baylin, S. B., & Issa, J. P. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer Res.* **58** , 5489-5494 (1998).

122. Issa, J. P., Ottaviano, Y. L., Celano, P., Hamilton, S. R., Davidson, N. E., & Baylin, S. B. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat.Genet.* **7**, 536-540 (1994).

123. Baylin, S. B., Esteller, M., Rountree, M. R., Bachman, K. E., Schuebel, K., & Herman, J. G. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum.Mol.Genet.* **10**, 687-692 (2001).

124. Goelz, S. E., Vogelstein, B., Hamilton, S. R., & Feinberg, A. P. Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* **228**, 187-190 (1985).

125. Jones, P. A. & Laird, P. W. Cancer epigenetics comes of age. *Nat.Genet.* **21**, 163-167 (1999).

126. Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M., & Issa, J. P. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv.Cancer Res.* **72**, 141-196 (1998).

127. Guan, R. J., Fu, Y., Holt, P. R., & Pardee, A. B. Association of K-ras mutations with p16 methylation in human colon cancer. *Gastroenterology* **116**, 1063-1071 (1999).

128. Liang, J. T., Chang, K. J., Chen, J. C., Lee, C. C., Cheng, Y. M., Hsu, H. C., Wu, M. S., Wang, S. M., Lin, J. T., & Cheng, A. L. Hypermethylation of the p16 gene in sporadic T3N0M0 stage colorectal cancers: association with DNA replication error and shorter survival. *Oncology* **57**, 149-156 (1999).

129. Strauss, B. S. The origin of point mutations in human tumor cells. *Cancer Res.* **52**, 249-253 (1992).

130. Kunkel, T. A. DNA replication fidelity. *J.Biol.Chem.* **267**, 18251-18254 (1992).

131. Hubscher, U., Nasheuer, H. P., & Syvaoja, J. E. Eukaryotic DNA polymerases, a growing family. *Trends Biochem.Sci.* **25**, 143-147 (2000).

132. Kirchner, J. M., Tran, H., & Resnick, M. A. A DNA polymerase epsilon mutant that specifically causes +1 frameshift mutations within homonucleotide runs in yeast. *Genetics* **155**, 1623-1632 (2000).

133. Wang, L., Patel, U., Ghosh, L., & Banerjee, S. DNA polymerase beta mutations in human colorectal cancer. *Cancer Res.* **52**, 4824-4827 (1992).

134. Flohr, T., Dai, J. C., Buttner, J., Popanda, O., Hagmuller, E., & Thielmann, H. W. Detection of mutations in the DNA polymerase delta gene of human sporadic colorectal cancers and colon cancer cell lines. *Int.J.Cancer* **80**, 919-929 (1999).

135. Laken, S. J., Petersen, G. M., Gruber, S. B., Oddoux, C., Ostrer, H., Giardiello, F. M., Hamilton, S. R., Hampel, H., Markowitz, A., Klimstra, D., Jhanwar, S., Winawer, S., Offit, K., Luce, M. C., Kinzler, K. W., & Vogelstein, B. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat.Genet.* **17**, 79-83 (1997).

136. Linton, M. F., Raabe, M., Pierotti, V., & Young, S. G. Reading-frame restoration by transcriptional slippage at long stretches of adenine residues in mammalian cells. *J.Biol.Chem.* **272**, 14127-14132 (1997).

137. Raabe, M., Linton, M. F., & Young, S. G. Long runs of adenines and human mutations. *Am.J.Med.Genet.* **76**, 101-102 (1998).

138. Lu, S. L., Akiyama, Y., Nagasaki, H., Saitoh, K., & Yuasa, Y. Mutations of the transforming growth factor-beta type II receptor gene and genomic instability in hereditary nonpolyposis colorectal cancer. *Biochem.Biophys.Res.Commun.* **216**, 452-457 (1995).

139. Parsons, R., Myeroff, L. L., Liu, B., Willson, J. K., Markowitz, S. D., Kinzler, K. W., & Vogelstein, B. Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.* **55**, 5548-5550 (1995).

140. Akiyama, Y., Iwanaga, R., Saitoh, K., Shiba, K., Ushio, K., Ikeda, E., Iwama, T., Nomizu, T., & Yuasa, Y. Transforming growth factor beta type II receptor gene mutations in adenomas from hereditary nonpolyposis colorectal cancer. *Gastroenterology* **112**, 33-39 (1997).

141. Jen, J., Powell, S. M., Papadopoulos, N., Smith, K. J., Hamilton, S. R., Vogelstein, B., & Kinzler, K. W. Molecular determinants of dysplasia in colorectal lesions. *Cancer Res.* **54**, 5523-5526 (1994).

142. Ikeda, S., Kishida, S., Yamamoto, H., Murai, H., Koyama, S., & Kikuchi, A. Axin, a negative regulator of the Wnt signaling pathway, forms a complex with GSK-3beta and beta-catenin and promotes GSK-3beta-dependent phosphorylation of beta-catenin. *EMBO J.* **17**, 1371-1384 (1998).

143. Behrens, J., Jerchow, B. A., Wurtele, M., Grimm, J., Asbrand, C., Wirtz, R., Kuhl, M., Wedlich, D., & Birchmeier, W. Functional interaction of an axin homolog, conductin, with beta-catenin, APC, and GSK3beta. *Science* **280**, 596-599 (1998).

144. Kishida, S., Yamamoto, H., Ikeda, S., Kishida, M., Sakamoto, I., Koyama, S., & Kikuchi, A. Axin, a negative regulator of the wnt signaling pathway, directly interacts with adenomatous polyposis coli and regulates the stabilization of beta-catenin. *J.Biol.Chem.* **273**, 10823-10826 (1998).

145. Liu, W., Dong, X., Mai, M., Seelan, R. S., Taniguchi, K., Krishnadath, K. K., Halling, K. C., Cunningham, J. M., Boardman, L. A., Qian, C., Christensen, E., Schmidt, S. S., Roche, P. C., Smith, D. I., & Thibodeau, S. N. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling. *Nat.Genet.* **26**, 146-147 (2000).

146. Malkhosyan, S., Rampino, N., Yamamoto, H., & Perucho, M. Frameshift mutator mutations. *Nature* **382**, 499-500 (1996).

147. Akiyama, Y., Sato, H., Yamada, T., Nagasaki, H., Tsuchiya, A., Abe, R., & Yuasa, Y. Germ-line mutation of the hMSH6/GTBP gene in an atypical hereditary nonpolyposis colorectal cancer kindred. *Cancer Res.* **57**, 3920-3923 (1997).

148. Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M., & Mori, T. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat.Genet.* **17**, 271-272 (1997).

149. Verma, L., Kane, M. F., Brassett, C., Schmeits, J., Evans, D. G., Kolodner, R. D., & Maher, E. R. Mononucleotide microsatellite instability and germline MSH6 mutation analysis in early onset colorectal cancer. *J.Med.Genet.* **36**, 678-682 (1999).

150. Maehama, T. & Dixon, J. E. PTEN: a tumour suppressor that functions as a phospholipid phosphatase. *Trends Cell Biol.* **9**, 125-128 (1999).

151. Stambolic, V., Suzuki, A., de la Pompa, J. L., Brothers, G. M., Mirtsos, C., Sasaki, T., Ruland, J., Penninger, J. M., Siderovski, D. P., & Mak, T. W. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell* **95**, 29-39 (1998).

152. Sakurada, K., Furukawa, T., Kato, Y., Kayama, T., Huang, S., & Horii, A. RIZ, the retinoblastoma protein interacting zinc finger gene, is mutated in genetically unstable cancers of the pancreas, stomach, and colorectum. *Genes Chromosomes Cancer* **30**, 207-211 (2001).

153. Piao, Z., Fang, W., Malkhosyan, S., Kim, H., Horii, A., Perucho, M., & Huang, S. Frequent frameshift mutations of RIZ in sporadic gastrointestinal and endometrial carcinomas with microsatellite instability. *Cancer Res.2000.Sep.1.;60.(17.):4701.-4.* **60**, 4701-4704 .

154. Chadwick, R. B., Jiang, G. L., Bennington, G. A., Yuan, B., Johnson, C. K., Stevens, M. W., Niemann, T. H., Peltomaki, P., Huang, S., & de la Chapelle, A. Candidate tumor suppressor RIZ is frequently involved in colorectal carcinogenesis. *Proc.Natl.Acad.Sci.U.S.A.2000.Mar.14.;97.(6.):2662.-7.* **97**, 2662-2667 .

155. Kim, N. G., Choi, Y. R., Baek, M. J., Kim, Y. H., Kang, H., Kim, N. K., Min, J. S., & Kim, H. Frameshift mutations at coding mononucleotide repeats of the hRAD50 gene in gastrointestinal carcinomas with microsatellite instability. *Cancer Res.* **61**, 36-38 (2001).

156. Haber, J. E. The many interfaces of Mre11. *Cell* **95**, 583-586 (1998).

157. Petrini, J. H. The mammalian Mre11-Rad50-nbs1 protein complex: integration of functions in the cellular DNA-damage response. *Am.J.Hum.Genet.* **64**, 1264-1269 (1999).

158. Wyllie, A. H. Apoptosis and carcinogenesis. *Eur.J.Cell Biol.* **73**, 189-197 (1997).

159. Wyllie, A. H. Apoptosis: an overview. *Br.Med.Bull.* **53**, 451-465 (1997).

160. McCormick, F. Signalling networks that cause cancer. *Trends Cell Biol.* **9**, M53-M56 (1999).

161. Shimizu, S., Narita, M., & Tsujimoto, Y. Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC. *Nature* **399**, 483-487 (1999).

162. Miyashita, T. & Reed, J. C. Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. *Cell* **80**, 293-299 (1995).

163. Oltvai, Z. N., Milliman, C. L., & Korsmeyer, S. J. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. *Cell* **74**, 609-619 (1993).

164. Calin, G. A., Gafa, R., Tibiletti, M. G., Herlea, V., Becheanu, G., Cavazzini, L., Barbanti-Brodano, G., Nenci, I., Negrini, M., & Lanza, G. Genetic progression in microsatellite instability high (MSI-H) colon cancers correlates with clinico-pathological parameters: A study of the TGRbetaRII, BAX, hMSH3, hMSH6, IGFIIR and BLM genes. *Int.J.Cancer 2000.May.20.;89.(3.):230.-5.* **89**, 230-235 .

165. Ionov, Y., Yamamoto, H., Krajewski, S., Reed, J. C., & Perucho, M. Mutational inactivation of the proapoptotic gene BAX confers selective advantage during tumor clonal evolution. *Proc.Natl.Acad.Sci.U.S.A.2000.Sep.26.;97.(20.):10872.-7.[Record.as.supplied.by.publisher.]* **97**, 10872-10877 .

166. Schwartz, S. J., Yamamoto, H., Navarro, M., Maestro, M., Reventos, J., & Perucho, M. Frameshift mutations at mononucleotide repeats in caspase-5 and other target genes in endometrial and gastrointestinal cancer of the microsatellite mutator phenotype. *Cancer Res.* **59**, 2995-3002 (1999).

167. Yamamoto, H., Gil, J., Schwartz, S. J., & Perucho, M. Frameshift mutations in Fas, Apaf-1, and Bcl-10 in gastro-intestinal cancer of the microsatellite mutator phenotype [letter; comment]. *Cell Death.Differ.2000.Feb.;7.(2.):238.-9.* **7**, 238-239 .

168. Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S. M., Ahmad, M., Alnemri, E. S., & Wang, X. Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479-489 (1997).

169. Thornberry N and Lazebnik Y. Caspases: Enemies Within. *Science* **281**, 1312-1316 (1998).

170. Song, Q., Lees-Miller, S. P., Kumar, S., Zhang, Z., Chan, D. W., Smith, G. C., Jackson, S. P., Alnemri, E. S., Litwack, G., Khanna, K. K., & Lavin, M. F. DNA-dependent protein kinase catalytic subunit: a target for an ICE-like protease in apoptosis. *EMBO J.* **15**, 3238-3246 (1996).

171. Bischof, O., Kim, S. H., Irving, J., Beresten, S., Ellis, N. A., & Campisi, J. Regulation and localization of the bloom syndrome protein in response to dna damage. *J.Cell Biol.* **153**, 367-380 (2001).

172. Hoekstra, M. F. Responses to DNA damage and regulation of cell cycle checkpoints by the ATM protein kinase family. *Curr.Opin.Genet.Dev.* **7**, 170-175 (1997).

173. Zhang, Q., Siebert, R., Yan, M., Hinzmann, B., Cui, X., Xue, L., Rakestraw, K. M., Naeve, C. W., Beckmann, G., Weisenburger, D. D., Sanger, W. G., Nowotny, H., Vesely, M., Callet-Bauchu, E., Salles, G., Dixit, V. M., Rosenthal, A., Schlegelberger, B., & Morris, S. W. Inactivating mutations and overexpression of BCL10, a caspase recruitment domain-containing gene, in MALT lymphoma with t(1;14)(p22;q32). *Nat.Genet.* **22**, 63-68 (1999).

174. Yan, M., Lee, J., Schilbach, S., Goddard, A., & Dixit, V. mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J.Biol.Chem.* **274**, 10287-10292 (1999).

175. Thome, M., Martinon, F., Hofmann, K., Rubio, V., Steiner, V., Schneider, P., Mattmann, C., & Tschopp, J. Equine herpesvirus-2 E10 gene product, but not its cellular homologue, activates NF-kappaB transcription factor and c-Jun N-terminal kinase. *J.Biol.Chem.* **274**, 9962-9968 (1999).

176. Koseki, T., Inohara, N., Chen, S., Carrio, R., Merino, J., Hottiger, M. O., Nabel, G. J., & Nunez, G. CIPER, a novel NF kappaB-activating protein containing a caspase recruitment domain with homology to Herpesvirus-2 protein E10. *J.Biol.Chem.* **274**, 9955-9961 (1999).

177. Willis, T. G., Jadayel, D. M., Du, M. Q., Peng, H., Perry, A. R., Abdul-Rauf, M., Price, H., Karran, L., Majekodunmi, O., Wlodarska, I., Pan, L., Crook, T., Hamoudi, R., Isaacson, P. G., & Dyer, M. J. Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types [see comments]. *Cell* **96**, 35-45 (1999).

178. Ruland, J., Duncan, G. S., Elia, A., del, B. B., I, Nguyen, L., Plyte, S., Millar, D. G., Bouchard, D., Wakeham, A., Ohashi, P. S., & Mak, T. W. Bcl10 is a positive regulator of antigen receptor-induced activation of NF-kappaB and neural tube closure. *Cell* **104**, 33-42 (2001).

179. Lee, S. H., Shin, M. S., Kim, H. S., Park, W. S., Kim, S. Y., Lee, H. K., Park, J. Y., Oh, R. R., Jang, J. J., Park, K. M., Han, J. Y., Kang, C. S., Lee, J. Y., & Yoo, N. J. Point mutations and deletions of the Bcl10 gene in solid tumors and malignant lymphomas. *Cancer Res.* **59**, 5674-5677 (1999).

180. Simms, L. A., Young, J., Wicking, C., Meltzer, S. J., Jass, J. R., & Leggett, B. A. The apoptotic regulatory gene, BCL10, is mutated in sporadic mismatch repair deficient colorectal cancers. *Cell Death.Differ.* **7**, 236-237 (2000).

181. Gryfe, R., Kim, H., Hsieh, E. T., Aronson, M. D., Holowaty, E. J., Bull, S. B., Redston, M., & Gallinger, S. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer [see comments]. *N.Engl.J.Med.* **342**, 69-77 (2000).

182. Mirabelli-Primdahl, L., Gryfe, R., Kim, H., Millar, A., Luceri, C., Dale, D., Holowaty, E., Bapat, B., Gallinger, S., & Redston, M. Beta-catenin mutations are specific for colorectal carcinomas with microsatellite instability but occur in endometrial carcinomas irrespective of mutator pathway. *Cancer Res.* **59**, 3346-3351 (1999).

183. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N., & Srivastava, S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248-5257 (1998).

184. Cottu, P. H., Muzeau, F., Estreicher, A., Flejou, J. F., Iggo, R., Thomas, G., & Hamelin, R. Inverse correlation between RER+ status and p53 mutation in colorectal cancer cell lines. *Oncogene* **13**, 2727-2730 (1996).

185. Vermeulen, S. J., Debruyne, P. R., Marra, G., Speleman, F. P., Boukamp, P., Jiricny, J., Cuthbert, A. P., Newbold, R. F., Nollet, F. H., van Roy, F. M., & Mareel, M. M. hMSH6 deficiency and inactivation of the alphaE-catenin invasion-suppressor gene in HCT-8 colon cancer cells. *Clin.Exp.Metastasis* **17**, 663-668 (1999).

186. Stone, J. G., Rowan, A. J., Tomlinson, I. P., & Houlston, R. S. Mutations in Bcl10 are very rare in colorectal cancer. *Br.J.Cancer* **80**, 1569-1570 (1999).

187. Lee, S. H., Shin, M. S., Kim, H. S., Park, W. S., Kim, S. Y., Lee, H. K., Park, J. Y., Oh, R. R., Jang, J. J., Park, K. M., Han, J. Y., Kang, C. S., Lee, J. Y., & Yoo, N. J. Point mutations and deletions of the Bcl10 gene in solid tumors and malignant lymphomas. *Cancer Res.* **59**, 5674-5677 (1999).

188. Lakin, N. D. & Jackson, S. P. Regulation of p53 in response to DNA damage. *Oncogene* **18**, 7644-7655 (1999).

189. Taylor, W. R. & Stark, G. R. Regulation of the G2/M transition by p53. *Oncogene* **20**, 1803-1815 (2001).

190. Zou, H., Henzel, W. J., Liu, X., Lutschg, A., & Wang, X. Apaf-1, a human protein homologous to C. elegans CED-4, participates in cytochrome c-dependent activation of caspase-3. *Cell* **90**, 405-413 (1997).

191. Denecker, G., Vercammen, D., Declercq, W., & Vandenabeele, P. Apoptotic and necrotic cell death induced by death domain receptors. *Cell Mol.Life Sci.* **58**, 356-370 (2001).

192. Lees-Miller, S. P. & Anderson, C. W. The DNA-activated protein kinase, DNA-PK: a potential coordinator of nuclear events. *Cancer Cells* **3**, 341-346 (1991).

193. Jhappan, C., Morse, H. C., III, Fleischmann, R. D., Gottesman, M. M., & Merlino, G. DNA-PKcs: a T-cell tumour suppressor encoded at the mouse scid locus. *Nat.Genet.* **17**, 483-486 (1997).

194. Taccioli, G. E., Amatucci, A. G., Beamish, H. J., Gell, D., Xiang, X. H., Torres Arzayus, M. I., Priestley, A., Jackson, S. P., Marshak, R. A., Jeggo, P. A., & Herrera, V. L. Targeted disruption of the catalytic subunit of the DNA-PK gene in mice confers severe combined immunodeficiency and radiosensitivity. *Immunity.* **9**, 355-366 (1998).

195. Shieh, S. Y., Ikeda, M., Taya, Y., & Prives, C. DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. *Cell* **91**, 325-334 (1997).

196. Jimenez, G. S., Bryntesson, F., Torres-Arzayus, M. I., Priestley, A., Beeche, M., Saito, S., Sakaguchi, K., Appella, E., Jeggo, P. A., Taccioli, G. E., Wahl, G. M., & Hubank, M. DNA-dependent protein kinase is not required for the p53-dependent response to DNA damage. *Nature* **400**, 81-83 (1999).

197. Kurimasa, A., Ouyang, H., Dong, L. J., Wang, S., Li, X., Cordon-Cardo, C., Chen, D. J., & Li, G. C. Catalytic subunit of DNA-dependent protein kinase: impact on lymphocyte development and tumorigenesis. *Proc.Natl.Acad.Sci.U.S.A* **96**, 1403-1408 (1999).

198. Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433-439 (2000).

199. Shiloh, Y. ATM and ATR: networking cellular responses to DNA damage. *Curr.Opin.Genet.Dev.* **11**, 71-77 (2001).

200. Durocher, D. & Jackson, S. P. DNA-PK, ATM and ATR as sensors of DNA damage: variations on a theme? *Curr.Opin.Cell Biol.* **13**, 225-231 (2001).

201. Tibbetts, R. S., Brumbaugh, K. M., Williams, J. M., Sarkaria, J. N., Cliby, W. A., Shieh, S. Y., Taya, Y., Prives, C., & Abraham, R. T. A role for ATR in the DNA damage-induced phosphorylation of p53. *Genes Dev.* **13**, 152-157 (1999).

202. Lakin, N. D., Hann, B. C., & Jackson, S. P. The ataxia-telangiectasia related protein ATR mediates DNA-dependent phosphorylation of p53. *Oncogene* **18**, 3989-3995 (1999).

203. Cliby, W. A., Roberts, C. J., Cimprich, K. A., Stringer, C. M., Lamb, J. R., Schreiber, S. L., & Friend, S. H. Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. *EMBO J.* **17**, 159-169 (1998).

204. Brown, E. J. & Baltimore, D. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.2000.Feb.15.;14.(4.):397.-402.* **14**, 397-402 .

205. Liu, Q., Guntuku, S., Cui, X. S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A., Donehower, L. A., & Elledge, S. J. Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev.2000.Jun.15.;14.(12.):1448.-59.* **14**, 1448-1459 .

206. Flaggs, G., Plug, A. W., Dunks, K. M., Mundt, K. E., Ford, J. C., Quiggle, M. R., Taylor, E. M., Westphal, C. H., Ashley, T., Hoekstra, M. F., & Carr, A. M. Atm-dependent interactions of a mammalian chk1 homolog with meiotic chromosomes. *Curr.Biol.* **7**, 977-986 (1997).

207. Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J., & Qin, J. BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* **14**, 927-939 (2000).

208. Frei, C. & Gasser, S. M. RecQ-like helicases: the DNA replication checkpoint connection. *J.Cell Sci.* **113 (Pt 15)**, 2641-2646 (2000).

209. Kawabe, T., Tsuyama, N., Kitao, S., Nishikawa, K., Shimamoto, A., Shiratori, M., Matsumoto, T., Anno, K., Sato, T., Mitsui, Y., Seki, M., Enomoto, T., Goto, M., Ellis, N. A., Ide, T., Furuichi, Y., & Sugimoto, M. Differential regulation of human RecQ family helicases in cell transformation and cell cycle. *Oncogene* **19**, 4764-4772 (2000).

210. Langland, G., Kordich, J., Creaney, J., Heppner, G. K., Lillard-Wetherell, K., Bebenek, K., Kunkel, T. A., & Groden, J. The BLM helicase interacts with hMLH1 but is not required for DNA mismatch repair. *J.Biol.Chem.* (2001).

211. Kirkpatrick, D. T. & Petes, T. D. Repair of DNA loops involves DNA-mismatch and nucleotide-excision repair proteins. *Nature* **387**, 929-931 (1997).

212. van Brabant, A. J., Ye, T., Sanz, M., German III, J. L., Ellis, N. A., & Holloman, W. K. Binding and melting of D-loops by the Bloom syndrome helicase. *Biochemistry* **39**, 14617-14625 (2000).

213. Calin, G., Herlea, V., Barbanti-Brodano, G., & Negrini, M. The coding region of the Bloom syndrome BLM gene and of the CBL proto-oncogene is mutated in genetically unstable sporadic gastrointestinal tumors. *Cancer Res.* **58**, 3777-3781 (1998).

214. Nicholson, D. W. Caspase structure, proteolytic substrates, and function during apoptotic cell death. *Cell Death.Differ.* **6**, 1028-1042 (1999).

215. Langlois, R. G., Bigbee, W. L., Jensen, R. H., & German, J. Evidence for increased in vivo mutation and somatic recombination in Bloom's syndrome. *Proc.Natl.Acad.Sci.U.S.A* **86**, 670-674 (1989).

216. Groden, J. & German, J. Bloom's syndrome. XVIII. Hypermutability at a tandem-repeat locus. *Hum.Genet.* **90**, 360-367 (1992).

217. Abdel-Rahman, W. M., Georgiades, I. B., Curtis, L. J., Arends, M. J., & Wyllie, A. H. Role of BAX mutations in mismatch repair-deficient colorectal carcinogenesis. *Oncogene* **18**, 2139-2142 (1999).

218. Fujiwara, T., Stolker, J. M., Watanabe, T., Rashid, A., Longo, P., Eshleman, J. R., Booker, S., Lynch, H. T., Jass, J. R., Green, J. S., Kim, H., Jen, J., Vogelstein, B., & Hamilton, S. R. Accumulated clonal genetic alterations in familial and sporadic colorectal carcinomas with widespread instability in microsatellite sequences. *Am.J.Pathol.* **153**, 1063-1078 (1998).

219. Willis, T. G., Jadayel, D. M., Du, M. Q., Peng, H., Perry, A. R., Abdul-Rauf, M., Price, H., Karran, L., Majekodunmi, O., Wlodarska, I., Pan, L., Crook, T., Hamoudi, R., Isaacson, P. G., & Dyer, M. J. Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types [see comments]. *Cell* **96**, 35-45 (1999).

220. Zhang, Q., Siebert, R., Yan, M., Hinzmann, B., Cui, X., Xue, L., Rakestraw, K. M., Naeve, C. W., Beckmann, G., Weisenburger, D. D., Sanger, W. G., Nowotny, H., Vesely, M., Callet-Bauchu, E., Salles, G., Dixit, V. M., Rosenthal, A., Schlegelberger, B., & Morris, S. W. Inactivating mutations and overexpression of BCL10, a caspase recruitment domain-containing gene, in MALT lymphoma with t(1;14)(p22;q32). *Nat.Genet.* **22**, 63-68 (1999).

221. Gryfe, R., Di Nicola, N., Gallinger, S., & Redston, M. Somatic instability of the APC I1307K allele in colorectal neoplasia [see comments]. *Cancer Res.* **58**, 4040-4043 (1998).

222. Yan, M., Lee, J., Schilbach, S., Goddard, A., & Dixit, V. mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J.Biol.Chem.* **274**, 10287-10292 (1999).

223. Koseki, T., Inohara, N., Chen, S., Carrio, R., Merino, J., Hottiger, M. O., Nabel, G. J., & Nunez, G. CIPER, a novel NF kappaB-activating protein containing a caspase recruitment domain with homology to Herpesvirus-2 protein E10. *J.Biol.Chem.* **274**, 9955-9961 (1999).

224. Srinivasula, S. M., Ahmad, M., Lin, J. H., Poyet, J. L., Fernandes-Alnemri, T., Tsichlis, P. N., & Alnemri, E. S. CLAP, a novel caspase recruitment domain-containing protein in the tumor necrosis factor receptor pathway, regulates NF-kappaB activation and apoptosis. *J.Biol.Chem.* **274**, 17946-17954 (1999).

225. Srinivasula, S. M., Ahmad, M., Lin, J. H., Poyet, J. L., Fernandes-Alnemri, T., Tsichlis, P. N., & Alnemri, E. S. CLAP, a novel caspase recruitment domain-containing protein in the tumor necrosis factor receptor pathway, regulates NF-kappaB activation and apoptosis. *J.Biol.Chem.* **274**, 17946-17954 (1999).

226. Apostolou, S., De Rienzo, A., Murthy, S. S., Jhanwar, S. C., & Testa, J. R. Absence of BCL10 mutations in human malignant mesothelioma. *Cell* **97**, 684-686 (1999).

227. Apostolou, S., Murthy, S. S., Kolachana, P., Jhanwar, S. C., & Testa, J. R. Absence of post-transcriptional RNA modifications of BCL10 in human malignant mesothelioma and colorectal cancer. *Genes Chromosomes Cancer* **30**, 96-98 (2001).

228. Dyer, M. J. Bcl10 mutations in malignancy [editorial]. *Br.J.Cancer* **80**, 1491 (1999).

229. Takahashi, H., Maeda, Y., Seto, M., & Hosokawa, Y. Nucleotide insertions and deletions within the homopolymeric runs of adenines and thymidines of BCL10 cDNAs in normal peripheral blood leukocytes. *Blood* **95**, 2728-2729 (2000).

230. Lambers, A. R., Gumbs, C., Ali, S., Marks, J. R., Iglehart, J. D., Berchuck, A., & Futreal, P. A. Bcl10 is not a target for frequent mutation in human carcinomas. *Br.J.Cancer* **80**, 1575-1576 (1999).

231. Stone, J. G., Rowan, A. J., Tomlinson, I. P., & Houlston, R. S. Mutations in Bcl10 are very rare in colorectal cancer. *Br.J.Cancer* **80**, 1569-1570 (1999).

232. Yoshitaka, T., Matsubara, N., Ikeda, M., Tanino, M., Hanafusa, H., Tanaka, N., & Shimizu, K. Mutations of E2F-4 trinucleotide repeats in colorectal cancer with microsatellite instability. *Biochem.Biophys.Res.Commun.* **227**, 553-557 (1996).

233. Komatsu, N., Takeuchi, S., Ikezoe, T., Tasaka, T., Hatta, Y., Machida, H., Williamson, I. K., Bartram, C. R., Koeffler, H. P., & Taguchi, H. Mutations of the E2F4 gene in hematological malignancies having microsatellite instability. *Blood 2000.Feb.15.;95.(4.):1509.-10.* **95**, 1509-1510 .

234. Chang, Y. C., Nakajima, H., Illenye, S., Lee, Y. S., Honjo, N., Makiyama, T., Fujiwara, I., Mizuta, N., Sawai, K., Saida, K., Mitsui, Y., Heintz, N. H., & Magae, J. Caspase-dependent apoptosis by ectopic expression of E2F-4. *Oncogene* **19**, 4713-4720 (2000).

235. Yamamoto, H., Sawai, H., Weber, T. K., Rodriguez-Bigas, M. A., & Perucho, M. Somatic frameshift mutations in DNA mismatch repair and proapoptosis genes in hereditary nonpolyposis colorectal cancer. *Cancer Res.* **58**, 997-1003 (1998).

236. Yin, C., Knudson, C. M., Korsmeyer, S. J., & Van Dyke, T. Bax suppresses tumorigenesis and stimulates apoptosis in vivo. *Nature* **385**, 637-640 (1997).

237. Chen, W. D., Eshleman, J. R., Aminoshariae, M. R., Ma, A. H., Veloso, N., Markowitz, S. D., Sedwick, W. D., & Veigl, M. L. Cytotoxicity and mutagenicity of frameshift-inducing agent ICR191 in mismatch repair-deficient colon cancer cells. *J.Natl.Cancer Inst.* **92**, 480-485 (2000).

238. Hancock, J. M. Simple sequences and the expanding genome. *Bioessays* **18**, 421-425 (1996).

239. Tran, H. T., Keen, J. D., Kricker, M., Resnick, M. A., & Gordenin, D. A. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol.Cell Biol.* **17**, 2859-2865 (1997).

240. Greene, C. N. & Jinks-Robertson, S. Frameshift intermediates in homopolymer runs are removed efficiently by yeast mismatch repair proteins. *Mol.Cell Biol.* **17**, 2844-2850 (1997).

241. Sia, E. A., Kokoska, R. J., Dominska, M., Greenwell, P., & Petes, T. D. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol.Cell Biol.* **17**, 2851-2858 (1997).

242. Michalickova, K., Bader, G. D., Isserlin, R., & Hogue, C. W. V. SeqHound biologoical sequence database system as a platform for bioinformatics research. *manuscript in preparation* (2001).

243. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. J., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. The protein data bank: a computer-based archival file for macromolecular structures. *Arch.Biochem.Biophys.* **185**, 584-591 (1978).

244. Nakamura, Y., Gojobori, T., & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic.Acids.Res.2000.Jan.1.;28.(1.):292.* **28**, 292 (0 AD/1/1).

245. Karlin, S. & Mrazek, J. What drives codon choices in human genes? *J.Mol.Biol.* **262**, 459-472 (1996).

246. Hanai, R. & Wada, A. Novel third-letter bias in Escherichia coli codons revealed by rigorous treatment of coding constraints. *J.Mol.Biol.* **207**, 655-660 (1989).

247. Knight, R. D., Freeland, S. J., & Landweber, L. F. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol.* **2**, RESEARCH0010 (2001).

248. Kreitman, M. The neutral theory is dead. Long live the neutral theory. *Bioessays* **18**, 678-683 (1996).

249. Xia, X. Maximizing transcription efficiency causes codon usage bias. *Genetics* **144**, 1309-1320 (1996).

250. Sharp, P. M., Stenico, M., Peden, J. F., & Lloyd, A. T. Codon usage: mutational bias, translational selection, or both? *Biochem.Soc.Trans.* **21**, 835-841 (1993).

251. Morton, B. R. & So, B. G. Codon usage in plastid genes is correlated with context, position within the gene, and amino acid content. *J.Mol.Evol.* **50**, 184-193 (2000).

252. Berg, D. & Chamberlin, M. Physical studies on ribonucleic acid polymerase from Escherichia coli B. *Biochemistry* **9**, 5055-5064 (1970).

253. Wagner, L. A., Weiss, R. B., Driscoll, R., Dunn, D. S., & Gesteland, R. F. Transcriptional slippage occurs during elongation at runs of adenine or thymine in Escherichia coli. *Nucleic Acids Res.* **18**, 3529-3535 (1990).

254. Knight, R. D. & Landweber, L. F. The early evolution of the genetic code. *Cell 2000.Jun.9.;101.(6.):569.-72.* **101**, 569-572 (0 AD/6/9).

255. Sagher, D., Hsu, A., & Strauss, B. Stabilization of the intermediate in frameshift mutation. *Mutat.Res.* **423**, 73-77 (1999).

256. Strauss, B. S. Frameshift mutation, microsatellites and mismatch repair. *Mutat.Res.* **437**, 195-203 (1999).

257. Dasika, G. K., Lin, S. C., Zhao, S., Sung, P., Tomkinson, A., & Lee, E. Y. DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis. *Oncogene* **18**, 7883-7899 (1999).

258. Molinari, M. Cell cycle checkpoints and their inactivation in human cancer. *Cell Prolif.* **33**, 261-274 (2000).

259. McDonald, E. R., III & El Deiry, W. S. Checkpoint genes in cancer. *Ann.Med.* **33**, 113-122 (2001).

260. Nurse, P. Cell cycle. Checkpoints and spindles. *Nature* **354**, 356-358 (1991).

261. DeWeese, T. L., Shipman, J. M., Larrier, N. A., Buckley, N. M., Kidd, L. R., Groopman, J. D., Cutler, R. G., te, R. H., & Nelson, W. G. Mouse embryonic stem cells carrying one or two defective Msh2 alleles respond abnormally to oxidative stress inflicted by low-level radiation. *Proc.Natl.Acad.Sci.U.S.A* **95**, 11915-11920 (1998).

262. Zhang, H. & Sun, X. F. Loss of p27 expression predicts poor prognosis in patients with Dukes' B stage or proximal colorectal cancer. *Int.J.Oncol.* **19**, 49-52 (2001).

263. Kong, S., Wei, Q., Amos, C. I., Lynch, P. M., Levin, B., Zong, J., & Frazier, M. L. Cyclin d1 polymorphism and increased risk of colorectal cancer at young age. *J.Natl.Cancer Inst.* **93**, 1106-1108 (2001).

264. Holler, N., Zaru, R., Micheau, O., Thome, M., Attinger, A., Valitutti, S., Bodmer, J. L., Schneider, P., Seed, B., & Tschopp, J. Fas triggers an alternative, caspase-8-independent cell death pathway using the kinase RIP as effector molecule. *Nat.Immunol.* **1**, 489-495 (2000).

265. Strasser, A., O'Connor, L., & Dixit, V. M. Apoptosis signaling. *Annu.Rev.Biochem.* **69**, 217-245 (2000).

266. Harrington, J. R. SODD-silencer of death domains. *Stem Cells* **18**, 388-389 (2000).

267. Yang, X., Khosravi-Far, R., Chang, H. Y., & Baltimore, D. Daxx, a novel Fas-binding protein that activates JNK and apoptosis. *Cell* **89**, 1067-1076 (1997).

268. Chang, H. Y., Nishitoh, H., Yang, X., Ichijo, H., & Baltimore, D. Activation of apoptosis signal-regulating kinase 1 (ASK1) by the adapter protein Daxx. *Science* **281**, 1860-1863 (1998).

269. Hollenbach, A. D., Sublett, J. E., McPherson, C. J., & Grosveld, G. The Pax3-FKHR oncoprotein is unresponsive to the Pax3-associated repressor hDaxx. *EMBO J.* **18**, 3702-3711 (1999).

270. Ozawa, F., Friess, H., Zimmermann, A., Kleeff, J., & Buchler, M. W. Enhanced expression of Silencer of death domains (SODD/BAG-4) in pancreatic cancer. *Biochem.Biophys.Res.Commun.* **271**, 409-413 (2000).

271. Dixit, V. M. RIPs: an emerging family of kinases involved in pro-inflammatory and apoptotic signaling. *Cold Spring Harb.Symp.Quant.Biol.* **64**, 359-362 (1999).

272. Kunkel, T. A. & Bebenek, K. DNA replication fidelity. *Annu.Rev.Biochem.* **69**, 497-529 (2000).

273. Luo, G., Yao, M. S., Bender, C. F., Mills, M., Bladl, A. R., Bradley, A., & Petrini, J. H. Disruption of mRad50 causes embryonic stem cell lethality, abnormal embryonic development, and sensitivity to ionizing radiation. *Proc.Natl.Acad.Sci.U.S.A* **96**, 7376-7381 (1999).

274. Bogre, L., Meskiene, I., Heberle-Bors, E., & Hirt, H. Stressing the role of MAP kinases in mitogenic stimulation. *Plant Mol.Biol.* **43**, 705-718 (2000).

275. Chang, L. & Karin, M. Mammalian MAP kinase signalling cascades. *Nature* **410**, 37-40 (2001).

276. Pelicci, G., Lanfrancone, L., Grignani, F., McGlade, J., Cavallo, F., Forni, G., Nicoletti, I., Grignani, F., Pawson, T., & Pelicci, P. G. A novel transforming protein (SHC) with an SH2 domain is implicated in mitogenic signal transduction. *Cell* **70**, 93-104 (1992).

277. Henthorn, P. S., Stewart, C. C., Kadesch, T., & Puck, J. M. The gene encoding human TFE3, a transcription factor that binds the immunoglobulin heavy-chain enhancer, maps to Xp11.22. *Genomics* **11**, 374-378 (1991).

278. de Leeuw, B., Suijkerbuijk, R. F., Balemans, M., Sinke, R. J., de Jong, B., Molenaar, W. M., Meloni, A. M., Sandberg, A. A., Geraghty, M., Hofker, M., & . Sublocalization of the synovial sarcoma-associated t(X;18) chromosomal breakpoint in Xp11.2 using cosmid cloning and fluorescence in situ hybridization. *Oncogene* **8**, 1457-1463 (1993).

279. Suijkerbuijk, R. F., Meloni, A. M., Sinke, R. J., de Leeuw, B., Wilbrink, M., Janssen, H. A., Geraghty, M. T., Monaco, A. P., Sandberg, A. A., & Geurts, v. K. Identification of a yeast artificial

chromosome that spans the human papillary renal cell carcinoma-associated t(X;1) breakpoint in Xp11.2. *Cancer Genet.Cytogenet.* **71**, 164-169 (1993).

280. Murphy, G. A., Solski, P. A., Jillian, S. A., Perez, d. l. O., D'Eustachio, P., Der, C. J., & Rush, M. G. Cellular functions of TC10, a Rho family GTPase: regulation of morphology, signal transduction and cell growth. *Oncogene* **18**, 3831-3845 (1999).

281. Murphy, G. A., Jillian, S. A., Michaelson, D., Philips, M. R., D'Eustachio, P., & Rush, M. G. Signaling Mediated by the Closely Related Mammalian Rho Family GTPases TC10 and Cdc42 Suggests Distinct Functional Pathways. *Cell Growth Differ.* **12**, 157-167 (2001).

282. Davis, A. P., Witte, D. P., Hsieh-Li, H. M., Potter, S. S., & Capecchi, M. R. Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11. *Nature* **375**, 791-795 (1995).

283. Lamm, G. M., Nicol, S. M., Fuller-Pace, F. V., & Lamond, A. I. p72: a human nuclear DEAD box protein highly related to p68. *Nucleic Acids Res.* **24**, 3739-3747 (1996).

284. Huang, J., Papadopoulos, N., McKinley, A. J., Farrington, S. M., Curtis, L. J., Wyllie, A. H., Zheng, S., Willson, J. K., Markowitz, S. D., Morin, P., Kinzler, K. W., Vogelstein, B., & Dunlop, M. G. APC mutations in colorectal tumors with mismatch repair deficiency. *Proc.Natl.Acad.Sci.U.S.A* **93**, 9049-9054 (1996).

285. Mai, M., Qian, C., Yokomizo, A., Smith, D. I., & Liu, W. Cloning of the human homolog of conductin (AXIN2), a gene mapping to chromosome 17q23-q24. *Genomics* **55**, 341-344 (1999).

286. Webster, M. T., Rozycka, M., Sara, E., Davis, E., Smalley, M., Young, N., Dale, T. C., & Wooster, R. Sequence variants of the axin gene in breast, colon, and other cancers: an analysis of mutations that interfere with GSK3 binding. *Genes Chromosomes Cancer* **28**, 443-453 (2000).

287. Gould, K. L., Moreno, S., Tonks, N. K., & Nurse, P. Complementation of the mitotic activator, p80cdc25, by a human protein-tyrosine phosphatase. *Science* **250**, 1573-1576 (1990).

288. Moreno, S. & Nurse, P. Clues to action of cdc25 protein. *Nature* **351**, 194 (1991).

289. Krause, K., Haugwitz, U., Wasner, M., Wiedmann, M., Mossner, J., & Engeland, K. Expression of the Cell Cycle Phosphatase cdc25C Is Down-Regulated by the Tumor Suppressor Protein p53 but Not by p73. *Biochem.Biophys.Res.Commun.* **284**, 743-750 (2001).

290. Sanchez, Y., Wong, C., Thoma, R. S., Richman, R., Wu, Z., Piwnica-Worms, H., & Elledge, S. J. Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science* **277**, 1497-1501 (1997).

291. Peng, C. Y., Graves, P. R., Thoma, R. S., Wu, Z., Shaw, A. S., & Piwnica-Worms, H. Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science* **277**, 1501-1505 (1997).

292. Hernandez, S., Bessa, X., Bea, S., Hernandez, L., Nadal, A., Mallofre, C., Muntane, J., Castells, A., Fernandez, P. L., Cardesa, A., & Campo, E. Differential expression of cdc25 cell-cycle-activating phosphatases in human colorectal carcinoma. *Lab Invest* **81**, 465-473 (2001).

293. Canman, C. E. Replication checkpoint: Preventing mitotic catastrophe. *Curr.Biol.* **11**, R121-R124 (2001).

294. Klein, H. L. A radical solution to death. *Nat.Genet.* **25** , 132-134 (2000).

295. Takai, H., Tominaga, K., Motoyama, N., Minamishima, Y. A., Nagahama, H., Tsukiyama, T., Ikeda, K., Nakayama, K., Nakanishi, M., & Nakayama, K. Aberrant cell cycle checkpoint function and early embryonic death in Chk1(-/-) mice. *Genes Dev.* **14**, 1439-1447 (2000).

296. Chan, T. A., Hermeking, H., Lengauer, C., Kinzler, K. W., & Vogelstein, B. 14-3-3Sigma is required to prevent mitotic catastrophe after DNA damage. *Nature* **401**, 616-620 (1999).

297. Shieh, S. Y., Ahn, J., Tamai, K., Taya, Y., & Prives, C. The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev.* **14**, 289-300 (2000).

298. Liao, H., Winkfein, R. J., Mack, G., Rattner, J. B., & Yen, T. J. CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis. *J.Cell Biol.* **130**, 507-518 (1995).

299. Zhu, X., Mancini, M. A., Chang, K. H., Liu, C. Y., Chen, C. F., Shan, B., Jones, D., Yang-Feng, T. L., & Lee, W. H. Characterization of a novel 350-kilodalton nuclear phosphoprotein that is specifically involved in mitotic-phase progression. *Mol.Cell Biol.* **15**, 5017-5029 (1995).

300. Weinberg, R. A. The molecular basis of carcinogenesis: understanding the cell cycle clock. *Cytokines Mol.Ther.* **2**, 105-110 (1996).

301. Ohtani, K., DeGregori, J., & Nevins, J. R. Regulation of the cyclin E gene by transcription factor E2F1. *Proc.Natl.Acad.Sci.U.S.A* **92**, 12146-12150 (1995).

302. Jares, P., Donaldson, A., & Blow, J. J. The Cdc7/Dbf4 protein kinase: target of the S phase checkpoint? *EMBO Rep.* **1**, 319-322 (2000).

303. Sclafani, R. A. Cdc7p-Dbf4p becomes famous in the cell cycle. *J.Cell Sci.* **113 (Pt 12)**, 2111-2117 (2000).

304. Puranam, K. L. & Blackshear, P. J. Cloning and characterization of RECQL, a potential human homologue of the Escherichia coli DNA helicase RecQ. *J.Biol.Chem.* **269**, 29838-29845 (1994).

305. Fossey, S. C., Kuroda, S., Price, J. A., Pendleton, J. K., Freedman, B. I., & Bowden, D. W. Identification and characterization of PRKCBP1, a candidate RACK-like protein. *Mamm.Genome* **11**, 919-925 (2000).

306. O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C., & Wood, R. D. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* **371**, 432-435 (1994).

307. Samec, S., Jones, T. A., Corlet, J., Scherly, D., Sheer, D., Wood, R. D., & Clarkson, S. G. The human gene for xeroderma pigmentosum complementation group G (XPG) maps to 13q33 by fluorescence in situ hybridization. *Genomics* **21**, 283-285 (1994).

308. Hyytinen, E. R., Frierson, H. F., Jr., Sipe, T. W., Li, C. L., Degeorges, A., Sikes, R. A., Chung, L. W., & Dong, J. T. Loss of heterozygosity and lack of mutations of the XPG/ERCC5 DNA repair gene at 13q33 in prostate cancer. *Prostate* **41**, 190-195 (1999).

309. Diamond, A., Cooper, G. M., Ritz, J., & Lane, M. A. Identification and molecular cloning of the human Blym transforming gene activated in Burkitt's lymphomas. *Nature* **305**, 112-116 (1983).

310. Billings, P. C., Weichselbaum, R. R., & Kennedy, A. R. Expression and methylation of the Blym gene in human tumor cell lines. *Eur.J.Cancer Clin.Oncol.* **23**, 331-337 (1987).

311. Cooper, G. M. Structure and function of ras and Blym oncogenes. *Hamatol.Bluttransfus.* **29**, 273-276 (1985).

312. Neiman, P. The Blym oncogenes. *Adv.Cancer Res.* **45**, 107-123 (1985).

313. Diamond, A., Devine, J. M., & Cooper, G. M. Nucleotide sequence of a human Blym transforming gene activated in a Burkitt's lymphoma. *Science* **225**, 516-519 (1984).

314. Inagaki, H., Matsushima, Y., Nakamura, K., Ohshima, M., Kadowaki, T., & Kitagawa, Y. A large DNA-binding nuclear protein with RNA recognition motif and serine/arginine-rich domain. *J.Biol.Chem.* **271**, 12525-12531 (1996).

315. Wei, Y. F., Robins, P., Carter, K., Caldecott, K., Pappin, D. J., Yu, G. L., Wang, R. P., Shell, B. K., Nash, R. A., Schar, P., & . Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol.Cell Biol.* **15**, 3206-3216 (1995).

316. Tillman, D. M., Harwood, F. G., Gibson, A. A., & Houghton, J. A. Expression of genes that regulate Fas signalling and Fas-mediated apoptosis in colon carcinoma cells. *Cell Death.Differ.* **5**, 450-457 (1998).

317. Li, Y., Kanki, H., Hachiya, T., Ohyama, T., Irie, S., Tang, G., Mukai, J., & Sato, T. Negative regulation of Fas-mediated apoptosis by FAP-1 in human cancer cells. *Int.J.Cancer* **87**, 473-479 (2000).

318. Ohmiya, N., Matsumoto, S., Yamamoto, H., Baranovskaya, S., Malkhosyan, S. R., & Perucho, M. Germline and somatic mutations in hMSH6 and hMSH3 in gastrointestinal cancers of the microsatellite mutator phenotype. *Gene* **272**, 301-313 (2001).

319. Ma, A. H., Xia, L., Littman, S. J., Swinler, S., Lader, G., Polinkovsky, A., Olechnowicz, J., Kasturi, L., Lutterbaugh, J., Modrich, P., Veigl, M. L., Markowitz, S. D., & Sedwick, W. D. Somatic mutation of hPMS2 as a possible cause of sporadic human colon cancer with microsatellite instability. *Oncogene 2000.Apr.27.;19.(18.):2249.-56.* **19**, 2249-2256 .

320. Liu, B., Nicolaides, N. C., Markowitz, S., Willson, J. K., Parsons, R. E., Jen, J., Papadopolous, N., Peltomaki, P., de la, C. A., Hamilton, S. R., & . Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat.Genet.* **9**, 48-55 (1995).

321. Nicolaides, N. C., Carter, K. C., Shell, B. K., Papadopoulos, N., Vogelstein, B., & Kinzler, K. W. Genomic organization of the human PMS2 gene family. *Genomics* **30**, 195-206 (1995).

322. Gordenin, D. A. & Resnick, M. A. Yeast ARMs (DNA at-risk motifs) can reveal sources of genome instability. *Mutat.Res.* **400**, 45-58 (1998).

323. Kunkel, T. A., Resnick, M. A., & Gordenin, D. A. Mutator specificity and disease: looking over the FENce. *Cell* **88**, 155-158 (1997).

324. Kroutil, L. C. & Kunkel, T. A. Deletion errors generated during replication of CAG repeats. *Nucleic Acids Res.* **27**, 3481-3486 (1999).

325. Kroutil, L. C., Register, K., Bebenek, K., & Kunkel, T. A. Exonucleolytic proofreading during replication of repetitive DNA. *Biochemistry* **35**, 1046-1053 (1996).

326. Galgoczy, D. J. & Toczyski, D. P. Checkpoint adaptation precedes spontaneous and damage-induced genomic instability in yeast. *Mol.Cell Biol.* **21**, 1710-1718 (2001).

327. Woerner, S. M., Gebert, J., Yuan, Y. P., Sutter, C., Ridder, R., Bork, P., & von Knebel, D. M. Systematic identification of genes with coding microsatellites mutated in DNA mismatch repair-deficient cancer cells. *Int.J.Cancer* **93**, 12-19 (2001).

328. Bader, S., Walker, M., Hendrich, B., Bird, A., Bird, C., Hooper, M., & Wyllie, A. Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* **18**, 8044-8047 (1999).

329. Bertoni, F., Codegoni, A. M., Furlan, D., Tibiletti, M. G., Capella, C., & Broggini, M. CHK1 frameshift mutations in genetically unstable colorectal and endometrial cancers. *Genes Chromosomes.Cancer* **26**, 176-180 (1999).

330. Codegoni, A. M., Bertoni, F., Colella, G., Caspani, G., Grassi, L., D'Incalci, M., & Broggini, M. Microsatellite instability and frameshift mutations in genes involved in cell cycle progression or apoptosis in ovarian cancer. *Oncol.Res.* **11**, 297-301 (1999).

331. Duval, A., Gayet, J., Zhou, X. P., Iacopetta, B., Thomas, G., & Hamelin, R. Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability. *Cancer Res.* **59**, 4213-4215 (1999).

332. Guanti, G., Resta, N., Simone, C., Cariola, F., Demma, I., Fiorente, P., & Gentile, M. Involvement of PTEN mutations in the genetic pathways of colorectal cancerogenesis. *Hum.Mol.Genet.2000.Jan.22.;9.(2.):283.-7.* **9**, 283-287 .

333. Singer, G. A. & Hickey, D. A. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol.Biol.Evol.* **17**, 1581-1588 (2000).

334. Kreitman, M. & Comeron, J. M. Coding sequence evolution. *Curr.Opin.Genet.Dev.* **9**, 637-641 (1999).

335. Eyre-Walker, A. DNA mismatch repair and synonymous codon evolution in mammals. *Mol.Biol.Evol.* **11**, 88-98 (1994).

336. Thomas, D. C., Umar, A., & Kunkel, T. A. Microsatellite instability and mismatch repair defects in cancer. *Mutat.Res.* **350**, 201-205 (1996).

337. Brown, T. C. & Jiricny, J. Repair of base-base mismatches in simian and human cells. *Genome* **31**, 578-583 (1989).

338. Osawa, S. & Jukes, T. H. Codon reassignment (codon capture) in evolution. *J.Mol.Evol.* **28**, 271-278 (1989).

339. Islam, M. Q. & Islam, K. A new functional classification of tumor-suppressing genes and its therapeutic implications. *Bioessays* **22**, 274-285 (2000).

340. Ilyas, M. & Tomlinson, I. P. Genetic pathways in colorectal cancer. *Histopathology* **28**, 389-399 (1996).

341. Elsaleh, H., Joseph, D., Grieu, F., Zeps, N., Spry, N., & Iacopetta, B. Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet 2000.May.20.;355.(9217.):1745.-50.* **355**, 1745-1750 .

342. Mirabelli-Primdahl, L., Gryfe, R., Kim, H., Millar, A., Luceri, C., Dale, D., Holowaty, E., Bapat, B., Gallinger, S., & Redston, M. Beta-catenin mutations are specific for colorectal carcinomas with microsatellite instability but occur in endometrial carcinomas irrespective of mutator pathway. *Cancer Res.* **59**, 3346-3351 (1999).