



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Differential patterns of gene expression associated with Microsatellite Instability and Chromosomal Instability in Colorectal Cancer

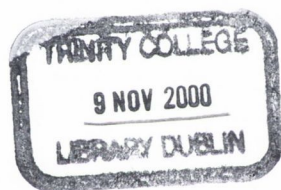
Donncha St. John Dunican

Thesis submitted to the University of Dublin for the
degree of Doctor of Philosophy

Research was carried out in the Department of Biochemistry,
Royal College of Surgeons in Ireland, under the supervision of
Dr. David T. Croke

Department of Biochemistry
University of Dublin
Trinity College

May 2000

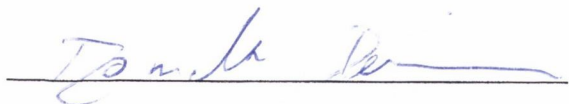


Thesis
5918

DECLARATION

I hereby certify that this thesis, submitted to the University of Dublin for examination for the degree of Doctor of Philosophy, has not been submitted as an exercise for a degree at any other university. I also certify that this thesis is entirely my own work.

This thesis may be made available for consultation within the university library and may be photocopied or loaned to other libraries for the purpose of consultation.

A handwritten signature in blue ink, appearing to read 'Donncha St. John Dunican', is written over a horizontal line.

Donncha St. John Dunican

May, 2000

TO DAD, MUM, DARA & ELEANOR

ACKNOWLEDGEMENTS

I would like to extend thanks to Dr. David Croke for giving me the opportunity to carry out this project. David, you have always been encouraging, patient and willing to listen. Most importantly, you have been a good friend and I will be lucky to work for somebody as understanding.

Thanks to all the members of the DTC group (past and present). Pedro: In **all** fairness. Thanks for keeping me grounded for the last few years. I owe you more than one. Orna: It was tough at the start but without your driving influence I don't know where I'd be now. APMcD: Stuck for words. For some reason I don't play as many 'sounds' anymore. It just doesn't seem right. I will work for you some day (if you'll take me). Macker: You were a bad influence and I enjoyed every minute of it. Where? Everywhere. I should mention the extended family: Karen, Miriam, Kate, Charlie, Jonathon, Eilis (no "h"). It was enjoyable to work with all of you.

To the other members of the Biochemistry Department, RCSI: Paddy, Alan, Aideen, Catherine P., Claire, Vicki, Catherine H., Grace, Brian C., Cathal O Flynn, Aine, Marice, Daphne, David Hardiman. Was I really that noisy? Oh yeah, and Daz: a member of Biochemistry and one of Ireland's leading fishmongers. The newer posse: Sylvia, Olive, Michael, Heidi. I'm sure you'd all agree that it's been real (unusual).

I would also like to thank Tony O' Grady of the Pathology Department who was kind enough to sort out all my 'immuno stuff', Theresa Keane who was instrumental in my tissue culture expertise and Philip Hassell for helping with my data analysis and coding my Perl scripts. Finally, I would like to thank Dr. Pat Vaughan who assisted me in carrying out my phosphorimaging.

A huge shout has to go out to the Stillorgan (ex-Rathmines) posse. Wah, Baz, LoloBall, Ambient and the man KP. Keep it real. I have learnt a considerable amount about debauchery not to mention roguish banter. And to the Galway people – Phil, Cearbhall, Ian (Tipperary?), Sarah (for the early years).

The most important people of all are my family. I dedicate this work to Mam and Dad for their love and attention. Thanks mum for looking after me for the last 26 years and making sure I did my "homework". You always make home such a wonderful place to

go. Dad, this work would have been much more difficult without you in my thoughts everyday and being my constant source of inspiration. I dedicate this work to you. Dara, what can I tell you? You'd think I'd cop on and realise Science was for boffin-anoraks and get a real job (like covering Miles Davis in some dingy pub in East London). Some people never learn. Eleanor, you were always my favourite and I hope you earn mega cash and support your impoverished brothers!

Table of Contents

Summary	i-ii
Abbreviations	iii-iv
Chapter 1 Introduction	1
1.0 Introduction to Colorectal Cancer	2-3
1.1 Colon Anatomy	4
1.2 Staging of Colorectal Cancer	5-6
1.3 Hereditary Non Polyposis Colorectal Cancer (HNPCC)	7-10
1.3.1 Microsatellite Instability	7
1.3.2 Mismatch Repair	7
1.3.3 Mismatch Repair Genes in HNPCC	8-9
1.3.4 Targets of Microsatellite Instability in HNPCC	9-10
1.4 Familial Adenomatous Polyposis	11-13
1.4.1 Familial Adenomatous Polyposis	11
1.4.2 The Adenomatous Polyposis Coli (APC) Gene	11-12
1.4.3 APC Protein Interactions	12-13
1.5 Sporadic Colorectal Cancer	14
1.6 A Model of Colorectal Tumourigenesis	15-16
1.7 Colorectal Cancer and Metastasis	17-18
1.8 Objectives of this Research Project	18-19
1.9 The Sporadic Colorectal Cancer Cell Lines Studied	20
Chapter 2 Materials and Methods	21
2.0 Materials	22-28
2.0.1 Reagents	22-25
2.0.2 Stock Solutions	26-28
2.1 Methods	29-35
2.1.1 Cell Culture	29
2.1.2 General Molecular Biology Techniques	29-31
2.1.3 RNA Isolation	32
2.1.4 RNA Analysis and Quantitation	32
2.1.5 Hybridisation Assay for mRNA Quantitation	33
2.1.6 cDNA Synthesis	33-34
2.1.7 Genomic DNA Screen	34
2.1.8 Semi-quantitative Reverse Transcription PCR	34-35
2.2 Differential Display Reverse Transcription PCR (DDRT-PCR)	36-41
2.2.1 Optimisation of DDRT-PCR	36-37
2.2.2 Cloning of DDRT-PCR Products	38-39
2.2.3 Sequencing of DDRT-PCR Products	39-41
2.2.4 Characterisation of DDRT-PCR Products	41
2.3 High Density cDNA Arrays	42-48
2.3.1 Introduction	42

2.3.2	Prehybridisation	42
2.3.3	cDNA Probe Preparation	42-45
2.3.4	Probe Preparation for DNA Orientation Marker	45
2.3.5	Hybridisation	45
2.3.6	Post- Hybridisation	45-46
2.3.7	Imaging of GDA Filters	46
2.3.8	GDA Library Clones	46
2.3.9	Northern Blotting	47
2.3.10	Screening for Repeat Sequence Instability	48
2.4	Protein Analysis	49-54
2.4.1	Protein Isolation / Subcellular Fractionation	49
2.4.2	Protein Estimation	50-51
2.4.3	Protein Precipitation	51
2.4.4	Protein Electrophoresis	51
2.4.5	Protein Blotting	52
2.4.6	Probing of Western Blots	52-53
2.4.7	Immunohistochemistry	53-54

Chapter 3 Differential Display Reverse Transcription PCR 55

3.0	Introduction	56-57
3.1	Methods	58
3.2	Results	59-62
3.2.1	RNA Integrity & Quantitation	59
3.2.2	Hybridisation Assay to Quantitate mRNA	59
3.2.3	Semi-quantitative Reverse Transcription PCR	59-60
3.2.4	Optimisation of DDRT-PCR Conditions	60-61
3.2.5	Reamplification of DDRT-PCR Products	61
3.2.6	Cloning and Sequencing of DDRT-PCR Products	61-62
3.3	Discussion	63-65

Chapter 4 Gene Discovery Array Analysis 66

4.0	Introduction	67-68
4.1	GDA Filters V1.3	69-71
4.1.1	GDA Clone Library	69
4.1.2	GDA Filter Structure	69-70
4.1.3	Normalisation of Duplicate Filters	70
4.1.4	GDA Hybridisation Protocol	70-71
4.2	Results	72-77
4.2.1	Introduction	72
4.2.2	Implementation of the GDA Protocol	72-73
4.2.3	Scatter Plot	73
4.2.4	Identification of Consistently Differentially Expressed Genes	73-74
4.2.5	Categorisation of the Gene Expression Profiles	74-75

4.2.6	Correlation of GDA Data with the SAGE Database	75
4.2.7	Identification of Simple Sequence Repeats in the Differentially Expressed Genes	76-77
4.3	Discussion	78-84

Chapter 5 Analysis of Differentially Expressed Genes 85

5.0	Introduction	86-88
5.1	Methods	89-91
5.1.1	Semi-quantitative RT-PCR / Northern blot Analysis	89-90
5.1.2	Detection of Instability at Simple Sequence Repeats	90-91
5.2	Results	92-101
5.2.1	Semi-quantitative RT-PCR / Northern blot Analysis of GDA Candidate Genes	92
5.2.2	GDA Data	92-97
5.2.3	Candidate Genes	97-100
5.2.4	Analysis of Instability in TGF β RII / PTPD1 Transcripts	101
5.3	Discussion	102-7
5.3.1	Roles for GDA Genes	102-5
5.3.2	Roles for Candidate Genes	105-7

Chapter 6 Analysis of IPP2 and IPP2 Associated Gene Expression 108

6.0	Introduction	109-12
6.0.1	Differential Display Reverse Transcription PCR	109
6.0.2	Protein Phosphatase 1 Catalytic Subunit Beta	109-10
6.0.3	Inhibitor 2 of Protein Phosphatase 1 (IPP2)	110-11
6.0.4	Identification of IPP2 by DDRT-PCR	112
6.1	Methods	112-14
6.1.1	DDRT-PCR & Semi-quantitative RT-PCR	112
6.1.2	Protein Analysis of PPP1C-Beta	113-14
6.2	Results	115-18
6.2.1	RT-PCR Analysis of IPP2	115-16
6.2.2	RT-PCR of Candidate Genes	116-17
6.2.3	Analysis of PPP1C-Beta Protein Expression	117-18
6.2.3.1	PPP1C-Beta protein expression in archival tissue	117
6.2.3.2	PPP1C-Beta protein expression in Cytospin Preparations	117
6.2.3.3	PPP1C-Beta Western Blot Analysis	118
6.3	Discussion	119-21
6.3.1	PPP1C-Beta	119-20
6.3.2	IPP2	120-21

Chapter 7 General Discussion	122
7.0 Introduction	123
7.1.1 MIN Versus CIN Phenotypes	124
7.1.2 MIN Associated Genes	124-25
7.1.3 CIN Associated Genes	125-26
7.1.4 CIN Genes found by GDA	126-27
7.2 IPP2 and IPP2 Associated Gene Studies	127-28
References	

SUMMARY

Two distinct pathways exist in sporadic colorectal cancer. The microsatellite instability pathway (MIN) is recognised by widespread instability caused by aberrant mismatch repair machinery. This genotype accounts for 15% of sporadic colorectal cancers. The chromosomal instability (CIN) genotype is recognised by gross chromosomal lesions but the underlying mechanism of CIN remains unclear. The differences in MIN and CIN colon cancer are likely to be due to the coordinate activation and repression of a key number of genes. Identification of MIN and CIN targeted genes will increase our understanding of the presence of two distinct genotypes in one type of cancer.

We have addressed differences in gene expression by using an *in vitro* model of MIN and CIN in the form of sporadic colorectal cancer cell lines. We have used the lines in conjunction with DDRT-PCR and high density cDNA array technology. DDRT-PCR analysis yielded four cDNA clones one of which contained an Alu element. The remaining three clones represent novel cDNA sequences of unknown identity. GDA v1.3 filters were used to identify known genes involved in MIN/CIN. This approach generated an initial panel of consistently differentially expressed genes (approximately seventy). A further nine of these genes were selected for confirmatory analysis by independent methods, of which six were confirmed [H2AZ, BTF3, PTPD1, RANBP2 & CCNA2]. The expression of a number of candidate genes was examined based on current literature findings including CYT8, CYT18, COX2 & SPLA2. The differentially expressed genes involve diverse cellular processes such as chromatin architecture, DNA-damage checkpoint, cell cycle regulation and arachidonic acid biosynthesis. Evidence from current literature suggest potential roles for some of these genes in gene ‘silencing’ in MIN and genes possibly implicated in CIN. The expression differences observed in these genes between MIN/CIN were modest which suggests that they may either be under the influence of a key CIN gene or be involved in a distinct CIN pathway. Exposing genes of this nature in the contexts of MIN and CIN may focus current views on the processes and pathways involved in these genotypes.

Supplementary to the MIN/CIN studies was the examination of IPP2 and a number of IPP2 associated genes in two pairs of isogenic cell lines (premetastatic and metastatic). We found elevated expression of IPP2 and PPP1C-Beta at the level of mRNA in both premetastatic cell lines suggesting possible coordinate regulation of the two transcripts. In addition, we observed similar patterns of expression of PPP1C-Beta

protein by Western blot analysis. These transcripts may have a role in colon cancer metastasis.

ABBREVIATIONS

AP3L: anchor primer 3 L
APC: adenomatous polyposis coli
ARD: ankyrin repeat domain
ATCC: american type culture collection
BLAST: basic local alignment search tool
BSA: bovine serum albumin
BTF3: basic transcription factor 3
CAM: cell adhesion molecule
CCNA2: cyclin A2
CIN: chromosomal instability
COX2: cyclooxygenase 2
CYT18: cytokeratin 18
CYT8: cytokeratin 8
DAB: diaminobenzadine
DCC: deleted in colorectal cancer
DDP: clone DDP
DDRT-PCR: differential display reverse transcription PCR
DDT: clone DDT
DEPC: diethylpyrocarbonate
DNTPS: dideoxy nucleoside triphosphates
DSCDNA: double stranded complementary DNA
ECL: electrochemiluminescence
ECM: extracellular matrix
EDTA: ethylenediaminetetraacetic acid
EGTA: ethylene glycol-bis N,N,N',N'-tetraacetic acid
EST: expressed sequence tag
FAP: familial adenomatous polyposis
FTP: file transfer protocol
GAPDH: glyceraldehyde-3-phosphate dehydrogenase
GDA: gene discovery array
GSI: genome systems incorporated
GTP: guanosine triphosphate
H2AZ: histone 2AZ
HDAC2: histone deacetylase A2
HNPCC : hereditary non polyposis colorectal cancer
HPLC: high performance liquid chromatography
HRP: horseradish peroxidase
IGFRII: insulin-like growth factor receptor 2
IMAGE: integrated molecular analysis of genes and their expression
IPP2: inhibitor 2 of protein phosphatase 1
IPTG: isopropylthio- β -D-galactoside
LOH: loss of heterozygosity
MAD2: mitotic arrest deficient yeast homolog like 1
MCR: multiple cluster region
MEM: minimum essential medium
MIN: microsatellite instability

MLC: myosin light chain
MMLV: moloney murine leukaemia virus
MOPS: 3-[N-morpholino]-propane-sulphonic acid
MSI: microsatellite instability
MSS: microsatellite stable
NCBI: national centre for biotechnology information
OMIM: online mendelian inheritance in man
ORF: open reading frame
PBS: phosphate buffered saline
PCR: polymerase chain reaction
PGER2: prostaglandin receptor E2
PLK: polo (*drosophila*) –like kinase
PMSF: phenylmethylsulphonyl fluoride α -toluensulfonyl FLUORIDE
PPP1C: protein phosphatase 1 catalytic subunit
PTPD1: protein tyrosine phosphatase D1
PVDF: polyvinylidene difluoride
RANBP2: ran binding protein 2
RER: replication error
ROCK: rhoA-associated kinase
RPS19: ribosomal protein s19
RTPCR: reverse transcription
SAGE: serial analysis of gene expression
SDS: sodium dodecyl sulphate
SPLA2: secretory phospholipase A2
SSC: sodium saline citrate
SSH: suppression subtractive hybridisation
TAE: tris-acetate-edta
TBE: tris borate-edta
TEMED: NNN'N-tetramethylethylenediamine
TFGBRII: transforming Growth Factor beta receptor II
TIGR: the institute for genomic research
UTR: untranslated region
X-GAL: 5-bromo-4-chloro-3-indolyl- β -D-galactoside

CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION TO COLORECTAL CANCER

Colorectal cancer refers to a cancer of one, or both, of two parts of the digestive tract: the colon and the rectum. Up to 50% of Western populations develop colorectal cancer by the age of 70. Malignancy ensues in approximately 10% of these cases. This makes colorectal cancer the second leading cause of death in the United States (Parker *et al.*, 1996). Colorectal cancer is the second most common type of cancer in Ireland, responsible for 15% of cancer related deaths (National Tumour Registry, Republic of Ireland, 1995). Environmental and heritable factors interact in the pathogenesis of colorectal cancer. Epidemiological evidence has identified environmental factors of lifestyle factors as being important in colorectal cancer progression. People in Western countries are 2-3 times more susceptible to colorectal cancer than Chinese, Japanese and people of most developing countries (Ziegler *et al.*, 1986). Colorectal cancer risk and diets high in fat and low in fibre are correlated strongly (Palmer *et al.*, 1983; Kinzler & Vogelstein, 1996).

Approximately 15% of colorectal tumours are inherited in an autosomal dominant fashion (Houlston *et al.*, 1992). Hereditary predisposition to colorectal cancer assumes two well-defined forms: Hereditary Non-Polyposis Colorectal Cancer (HNPCC) and Familial Adenomatous Polyposis (FAP). These inherited traits account for 5% and 1% of all colorectal cancers respectively (Konishi *et al.*, 1996). FAP is recognisable by the appearance of thousands of adenomas in the colon and rectum which form during the second and third decades of life in FAP patients. Cytogenetic studies identified a deletion on chromosome 5q in a polyposis patient, with further studies demonstrating tight linkage of the disease to markers on chromosome 5q21, which coincides with a tumour suppressor gene – the adenomatous polyposis coli (APC) gene. Mutations in this gene lead to truncated APC protein products, thereby diminishing the tumour suppressor properties of APC (Kinzler & Vogelstein, 1996). FAP tumours are classically left-sided, appearing in the descending colon (Vasen *et al.*, 1990; Lynch *et al.*, 1993). APC mutations may be followed by downstream events including DCC (deleted in colorectal cancer) deletions and loss of expression of p53. HNPCC kindreds are recognisable by widespread genomic instability at the nucleotide level. Errors during DNA replication are normally repaired by DNA mismatch repair genes, which are mutated in HNPCC tumours. Slippage of DNA Polymerases leads to instability (expansion or contraction) at simple mono-, di- and trinucleotide repeat sequences.

Mutations of this nature have been detected in the transforming growth factor- β I receptor II (TGF β IRII) and insulin-like growth factor-II receptor I (IGFIIRI) genes in HNPCC tumours which seem to be the driving force behind abnormal growth in these tumours. Interestingly, HNPCC tumours are right-sided, appearing in the mucosa of the ascending colon (Vasen *et al.*, 1990; Lynch *et al.*, 1993).

The majority of solid colorectal tumours (85%) are sporadic in origin – sporadic denotes affected patients with no familial history of colorectal cancer. However, sporadic colorectal tumours exhibit close similarities to those in the inherited syndromes. Sporadic colorectal cancers exhibit both chromosomal instability (CIN) and MIN microsatellite instability (MIN) phenotypes (Lengauer *et al.*, 1997). The CIN phenotype exhibits gross chromosomal abnormalities such as aneuploidy and loss of heterozygosity (LOH; loss of the paternal or maternal allele). APC or β -catenin mutations are the initial event in the CIN phenotype. The MIN phenotype is not as prevalent in sporadic colorectal cancer accounting for 15% of such malignancies. Similar to HNPCC tumours, MIN sporadic colorectal cancers have widespread genomic instability.

1.1 COLON ANATOMY

The colon and rectum together make up the lower part of the digestive system also known as the large bowel or large intestine. The colon is about five feet long and may be viewed in four sections: ascending, transverse, descending and sigmoid. The ascending colon is usually 15cm in length and ascends to the right lobe of the liver and makes a slight depression in it. The right colic flexure joins the ascending colon and the transverse colon. The high point of the ascending colon turns abruptly forward and to the left. The transverse section is about 50cm long and extends from the right colic flexure turning down, forward and to the left. The transverse colon is concave; its concavity is directed up and back. The posterior surface associates with the descending duodenum and head of the pancreas by means of loose connective tissue. The left colic flexure forms the junction between the transverse colon and the descending colon. The descending colon is about 25cm long and travels downward along the left kidney to the iliac crest. The descending colon then curves downward in front of the iliacus to end in the sigmoid colon at the inlet of the lesser pelvis. The sigmoid colon (pelvic colon) begins at the pelvic inlet and forms a loop about 40cm long that finally descends in contact with the left pelvic cavity between the rectum and bladder in males, and rectum and uterus in females. Finally, it turns back toward the midline level before bending downwards to the rectum which is about 12cm in length and travels downwards to join the anal canal passing through the pelvic diaphragm. Figure 1.1 illustrates the anatomy of the colorectal tract and its tissue composition.

Colorectal cancers (hereditary and sporadic) showing microsatellite instability are known as Replication Error positive (RER+), microsatellite unstable (MSI or MIN). Colorectal cancers with functional mismatch repair machinery are often referred to as Replication Error negative (RER-), microsatellite stable (MSS) or CIN (chromosomal instability). To minimise confusion the MIN and CIN abbreviations are used throughout.

1.2 STAGING OF COLORECTAL CANCER

A clinician encountering a patient with colorectal cancer needs to be able to predict the expected outcome to devise an appropriate therapy. Lockhart-Mummery (1927) documented a relationship between patient prognosis and tumour extent in the colorectal passage. This observation was enhanced by the development of a system for the pathological classification of rectal carcinoma (Dukes *et al.*, 1932). Initially, three pathological stages were described: **A**, growth solely in the rectal wall; **B**, expansion of growth to extrarectal tissues without metastasis; and **C**, metastasis in distant lymph nodes. Later, Stage C was divided into two subclasses: **C₁**, only lymph nodes involved; and **C₂**, nodal spread to extrarectal blood vessels. Eventually, Stage D was adopted to describe metastasis to distant organs. The Dukes' classification was eventually applied to colonic tumours. The Dukes' staging system for colorectal cancers is detailed in Table 1.1.

The Dukes' system was somewhat complicated by two modifications. The Dukes' stages were subdivided to describe the degree of penetration of the tumor through the different tissue types present in the colon and rectum (Kirklin *et al.*, 1949; Astler *et al.*, 1954). This advance made it difficult to compare reported relationships between tumour stage, prognosis and survival rate. There are deficiencies in all versions of the Dukes' classifications. Analysis based on a small number of tissue specimens may miss breakthrough to the muscularis propria. This type of analysis could misclassify a patient into the good prognosis Dukes' A group resulting in patient death sooner than expected. In addition, the number of lymph nodes involved and the histological grade of the tumour are overlooked in Dukes' classification, which are known survival rate indicators (Jass *et al.*, 1987). However, Dukes' classification remains a powerful prognostic tool. Dukes and Bussey (1958) reported 5-year survival rates of 81.2, 64 and 27.4 % for patients with Dukes A, B and C tumours respectively, figures comparable to modern series. Moreover, univariate analysis has emphasised the prognostic importance of tumour staging in colorectal neoplasia .

In summary, Dukes' staging of colorectal cancer progression is not perfect. It fails to account for clinical factors and the number of nodes involved. The modified Dukes' classifications served only to complicate the assessment of tumour stage, prognosis and survival rate. However, due to its prognostic significance and its ease of

use, Dukes' classification remains the best predictor of survival in colorectal cancer after 60 years.

1.3 HEREDITARY NON-POLYPOSIS COLORECTAL CANCER

1.3.1 Microsatellite Instability

Microsatellites are short nucleotide sequences (2-5 base pairs) that appear in repetitive tracts distributed throughout the genome. The initial clue that microsatellite instability has a role in colorectal cancer came from a study of a group of sporadic colorectal tumours which harboured widespread instability at poly(A) tracts in their genomes (Ionov *et al.*, 1993). Further studies revealed instability at poly(CA) tracts leading to the descriptive term 'microsatellite instability' (MSI or MIN) (Thibodeau *et al.*, 1993). This genotype was rarely observed in sporadic cases but was present in the majority of HNPCC tumours. Expansion or contraction of simple repeat sequences is assumed to be due to slippage of DNA Polymerases during DNA replication. Following the exclusion of candidate gene loci, the genes responsible for HNPCC were linkage-mapped to chromosomes 2p16 and 3p21 after evaluation of large kindreds from North America, Europe and New Zealand (Lindblom *et al.*, 1993; Peltomaki *et al.*, 1993).

1.3.2 Mismatch Repair

During normal DNA replication, discrepancies may arise from single base misincorporations by Polymerases or larger mismatches may result from strand slippage. Mutator strains of bacteria and yeast have highlighted mechanisms of genome stability maintenance. One such *E. coli* strain has elucidated the mechanism for correction of replicative errors (Modrich *et al.*, 1991). This repair system recognises the newly synthesised strand containing the misincorporated nucleotide, excises a small number of nucleotides surrounding this base, fills in the correct bases and ligates the corrected sequence (Modrich *et al.*, 1991). Similar systems have been identified in yeast (Reenan *et al.*, 1992). None of the original MIN reports proposed a specific mechanism responsible for the instability. It was noted that microsatellite instability was present in bacteria with defective mismatch repair genes, *mutS* or *mutL*, in addition to *S. cerevisiae* with mutated yeast homologues of either *mutS* or *mutL* (Strand *et al.*, 1993). It was postulated that the MIN genotype might arise from defective mismatch repair. This hypothesis was confirmed by the discovery of mutations in the *mutS* or *mutL* human homologues in HNPCC kindreds (Fishel *et al.*, 1993; Bronner *et al.*, 1994).

1.3.3 Mismatch Repair Genes in HNPCC

To date, more than 70 different germline mutations have been detected in DNA mismatch repair genes and shown to be associated with HNPCC families. The genes identified are *hMSH2* (human *mutS* homologue 2), *hMHS6* (GTBP; human *mutS* homologue 6), *hMLH1* (human *mutL* homologue 1), *hPMS1* (human *mutL* homologue) and *hPMS2* (human *mutL* homologue). Current evidence suggests that two genes, *hMSH2* and *hMLH1*, account for roughly equal proportions of HNPCC kindreds, together being responsible for a majority of these families (Peltomaki *et al.*, 1997). *hMSH2*, the human homologue of *mutS*, was cloned by homology to conserved regions of *mutS* gene homologues in bacteria and yeast. *MutS* gene products function in methyl-directed mismatch repair pathways by binding DNA at the site of base mismatches. The *hMSH2* protein is potentially over 900 amino acids in length and is 41% homologous to the yeast *mutS* homologue (Fischel *et al.*, 1993). The human gene was mapped to chromosome 2q21 in a panel of somatic cell hybrids.

Independent evidence of the role of *hMSH2* arose from the collaborative efforts of the laboratories of Vogelstein, Kinzler and de la Chapelle (Leach *et al.*, 1993). In an HNPCC kindred, all affected individuals carried a *hMSH2* mutant allele resulting in a substitution of leucine for proline at codon 622. DNA from unaffected members of the kindred carried two wildtype *hMSH2* alleles. The proline residue is highly conserved from bacteria to human suggesting its importance in normal gene function. A number of other mutations were found in other kindreds, but were absent in mismatch repair proficient colorectal tumours. In a second kindred, a defect deleting codons 265 to 314 segregated perfectly with HNPCC. MIN colorectal cell lines were shown to replicate endogenous and episomal (CA)_n hgs with 100-fold more errors than mismatch repair proficient colorectal cancer cell lines (Parsons *et al.*, 1993). All of these investigations are convincing evidence of the role of *hMSH2* gene in HNPCC and a small subset of sporadic colorectal cancers.

hMLH1, a *mutL* homologue, on chromosome 3 was found to be mutated in the germline of some 30% of HNPCC kindreds (Bronner *et al.*, 1994). *hMSH6* and the two *hPMS* genes have been found to be mutated in the germline of a few HNPCC patients (Liu *et al.*, 1996). The most important lesson resulting from HNPCC studies concerns its dominant inheritance. In general, patients with HNPCC have one normal allele of the relevant mismatch repair gene. Mismatch repair is only abolished when the wildtype

allele is inactivated resulting in the accumulation of mutations and ultimately clonal expansion. This shows that somatic mutations in repair genes can be selected for during tumourigenesis, even when such mutations do not directly enhance cell growth.

1.3.4 Targets of microsatellite instability in HNPCC

Little is known about the mechanisms underlying MIN and its direct contribution to colorectal tumourigenesis. Defective mismatch repair may promote mutations in protooncogenes or tumor suppressor genes known to be involved in other cancers, i.e. K-Ras, p53, DCC and APC. However, mutations in such genes were shown to be less common in MIN cancers compared with tumours that do not have this genotype (Thibodeau *et al.*, 1993). An accepted alternative is that MIN leads to mutations in novel genes, and perhaps those specifically with simple sequence repeats in their coding regions. This has been shown for the genes encoding transforming growth factor β receptor II (TGF- β -RII) and the insulin-like growth factor II receptor (IGFIIIR).

TGF- β is a multi-functional polypeptide that regulates a number of cellular processes including growth, differentiation, deposition of the extracellular matrix and immunosuppression (Moses *et al.*, 1990; Roberts *et al.*, 1991). TGF- β exerts its effects through binding to specific cell surface proteins. This signalling polypeptide acts as an anti-proliferative signal of colon epithelial cells and is the primary ligand for TGF- β -RII (Wrana *et al.*, 1994). Cells that lose the ability to express or respond to TGF- β are more likely to exhibit uncontrolled cell growth and to become tumourigenic. TGF- β binds TGF- β -RII which interacts transiently with TGF- β -RI, which in turn activates parallel pathways that inhibit cellular proliferation (Alexandrow *et al.*, 1995).

TGF- β -RII has a 10 base-pair poly(A) tract (BAT-RII) at codons 125-128 of its 565-codon open reading frame. BAT-RII was found to be mutated in 90% of MIN tumours. In most cases, both BAT-RII alleles are affected. Tumours heterozygous for BAT-RII mutations exhibited additional mutations in the other TGF- β -RII allele which were expected to inactivate it. These genetic data support that TGF- β -RII behaves like a tumor suppressor and is a critical target of inactivation in mismatch repair deficient tumours.

Due to the existence of microsatellite instability in the coding region of TGF- β -RII, it was reasoned that there may exist further instability in other cancer related genes

in MIN tumours. GenBank and the European Molecular Biology Laboratory (EMBL) databases were searched for genes with exonic microsatellites which resulted in the discovery of the insulin-like growth factor II receptor (IGFRII) gene (Souza *et al.*, 1996). IGFIIR is known to bind and activate the latent TGF- β complex, a process that is essential for the growth suppressive properties of TGF- β (Kojima *et al.*, 1993). In addition, IGFIIR mediates the degradation of IGFII, itself a potent proliferative signal (Lau *et al.*, 1994). A screen of 92 MIN tumours demonstrated IGFIIR mutations in 13% of the specimens (Souza *et al.*, 1996). Interestingly, no IGFIIR mutations were present in 52 non-MIN tumours. All but one of the mutations appeared in an 8bp poly(G) tract spanning nucleotides 4089 and 4096 of IGFIIR. Mutations in either of these genes have been demonstrated to result in neoplasia and are thought to be the driving force of tumourigenesis in MIN tumours.

A number of other targets of the MIN genotype have been identified. Simms *et al.* (1997) screened a range of cancer associated genes and demonstrated a number of polymorphisms in the 5' untranslated region (UTR), 3' UTR or the coding regions of genes. Strikingly, the instability appeared only in mononucleotide repeat sequences. Endometrial cancer is the most common extracolonic neoplasia in HNPCC. Instability at a 10 base pair poly(A) tract has been identified in 55% of endometrial cancers in the Caspase-5 gene (Schwartz *et al.*, 1999). Frameshift mutations were detected in both mono- and dinucleotide repeat sequences of the Bax gene in tumours from 36 HNPCC patients (Sakao *et al.*, 1999).

1.4 FAMILIAL ADENOMATOUS POLYPOSIS

1.4.1 Familial Adenomatous Polyposis

The second genetic predisposition to colorectal cancer is known as Familial Adenomatous Polyposis (FAP) and is inherited in an autosomal dominant fashion. During adolescence affected individuals develop from a few hundred to over a thousand adenomatous polyps in their large bowel. These are sufficiently likely to give rise to adenocarcinomas to make prophylactic removal of the colon usual in diagnosed FAP individuals. Adenomas may occur elsewhere in the gastrointestinal tract and the condition is often associated with other extracolonic lesions, such as epidermoid cysts, jaw osteomata and fibrous desmoid tumours. In contrast to HNPCC syndromes that are genetically heterogeneous, the FAP syndromes appear to be due to defects in a single gene, the Adenomatous Polyposis Coli (APC) gene. However, β -catenin mutations in tumours with normal APC function can lead to cancers closely related to FAP cancers (Morin *et al.*, 1997).

1.4.2 The Adenomatous Polyposis Coli gene

The first evidence of a causative genetic lesion in FAP arose from cytogenetic studies that revealed a chromosomal deletion in a sporadic colorectal cancer patient with polyposis (Herrera *et al.*, 1986). Chromosome analysis of blood cells from a 42-year-old white male with colon carcinoma and possible Gardner Syndrome revealed a deletion at chromosome 5q. This result prompted further molecular studies to map this region of chromosome 5q. Tight linkage was demonstrated of FAP to markers on chromosome 5q21 (Bodmer *et al.*, 1987; Leppert *et al.*, 1987). This region was shown to harbour the APC gene and was subsequently demonstrated to be the causative genetic component of FAP by its cosegregation with mutant APC alleles in affected kindreds (Grodin *et al.*, 1991).

The APC gene transcript is 9,872 bp in length and consists of 15 exons. Over 120 different mutations have been identified in FAP cases. Patients with FAP generally carry one mutant APC allele which is sufficient for polyp formation in the colorectal tract. Loss of the wildtype APC allele in such patients by a process that is known as loss of heterozygosity (LOH) is thought to be the initiating event in the progression of non-

invasive adenoma to carcinoma. Strikingly, most APC mutations yield truncated protein products (Smith *et al.*, 1993). The nature of the APC mutation seems to be a determinant of the severity of the disease, *i.e.*: the extent of polyposis.

1.4.3 APC Protein Interactions

The defined function for the APC protein is currently unclear. The polypeptide chain of the APC gene is illustrated in Figure 1.2. The constant deletion of the carboxy-terminal structure suggests that important growth-suppressing activity resides in this region of the protein. It is difficult to assess how APC mutations may affect its activity as the APC amino acid sequence bears little similarity to any known protein. The majority of germline APC mutations appear to be dispersed throughout the 5' half of the sequence. Somatic mutations have been observed to accumulate between codons 1309 and 1450, a stretch known as the Mutation Cluster Region (MCR) (Miyoshi *et al.*, 1992). This region has been shown to interact with β -catenin (Su *et al.*, 1993). β -catenin was originally viewed as a protein involved in cell adhesion due to its association with various members of the cadherin protein family (Polakis *et al.*, 1995). However, the cell adhesion properties of β -catenin have been cleanly dissected from the discovery of its involvement in the Wntless-Wnt signalling pathway (Gumbiner *et al.*, 1995). In addition, β -catenin binds members of the T cell factor (Tcf) and lymphoid enhancer factor (Lef) families of transcription factors (Behrens *et al.*, 1996; Molenaar *et al.*, 1996). β -catenin is capable of forming a complex with Tcf-4 and this complex is known to activate the transcription of a number of unknown gene targets (Korinek *et al.*, 1997). Wildtype APC regulates gene expression by binding β -catenin and promoting its degradation (Rubinfeld *et al.*, 1993). Mutant APC protein binding to β -catenin is impaired resulting in the inappropriate activation of transcription of these unknown target genes. Dominant mutations in β -catenin that effect its degree of phosphorylation allow constant association of β -catenin with Tcf-4 resulting in inappropriate gene transcription (Morin *et al.*, 1997). The choice of downstream targets for β -catenin is extensive due to its diverse range of functions.

Recent reports have shown APC to interact with EB-1 (a protein of unknown function), the human homologue of *Drosophila* tumor suppressor gene (DLG), various tubulins and γ -catenin. However, the implications of these interactions remain unclear.

c-Myc has been demonstrated as a target of the β -catenin/Tcf-4 complex (He *et al.*, 1998) and is a known nuclear phosphoprotein involved in proliferation. Previously, overexpression of c-Myc has been implicated in colorectal tumourigenesis (Fearon *et al.*, 1990).

Mutations in APC which effect downstream β -catenin mediated signalling or mutations in β -catenin itself are likely initiator components in FAP tumours. Notably, although FAP accounts for only 1% of all colorectal cancers, APC and β -catenin defects are found in the majority of sporadic colorectal neoplasms.

1.5 SPORADIC COLORECTAL CANCER

Sporadic colorectal cancer cases are classified as those that appear in patients with no familial history of colorectal cancer. However, abnormalities associated with the hereditary colorectal syndromes are prevalent in the majority of sporadic neoplasms, particularly APC mutations (Kinzler & Vogelstein, 1996). Specifically, sporadic colorectal cancers can be classified into chromosomal instability (CIN) or microsatellite instability (MIN) genotypes. CIN tumours are recognised by gross chromosomal abnormalities and loss of heterozygosity (LOH) (the loss of a maternal or paternal allele). Karyotypic studies have shown the loss or gain of chromosomes in the majority of cancers. LOH is also widespread in most cancers (Vogelstein *et al.*, 1989). Fluorescence *in situ* hybridisation studies have shown that multiple losses or gains of chromosomes to be 10-100 fold greater in colorectal cancer cell lines than in their normal histological counterparts (Lengauer *et al.*, 1997). The degree of LOH observed in markers surrounding the APRT gene was seen to be 10-fold higher in a mismatch repair proficient cell line than a mismatch repair deficient cell line (Phear *et al.*, 1996). CIN tumours account for 85% of all sporadic colorectal cancers, most of which are initiated by mutations in APC or β -catenin. p53 was initially a candidate for the molecular basis of CIN. However, a number of cancers with mutant p53 have been shown to be diploid with no obvious chromosomal defects and exhibit MIN due to impaired mismatch repair (Eshleman *et al.*, 1998). Studies on yeast have revealed a large array of gene types that can induce chromosomal abnormalities. The mitotic checkpoint system in yeast may prove to be an attractive target for CIN tumours. Disruption of genes of the mitotic checkpoint assembly can lead to CIN because checkpoint defective cells can complete mitosis with inappropriately aligned chromosomes (Paulovich *et al.*, 1997). Mutations of *hsMad2* have been found in some breast cancers (Li *et al.*, 1996) and mutations of *hBub1* in a small fraction of colorectal cancers (Cahill *et al.*, 1998). These genes are components of the mitotic checkpoint system and may contribute to aneuploidy.

1.6 A MODEL OF COLORECTAL TUMOURIGENESIS

Significant evidence exists from experimental investigations defining a common stepwise series of events in colorectal cancer pathogenesis. Initially, an expanded zone of epithelial cell proliferation is seen in colonic crypts (Gerdes *et al.*, 1993). Clonal growth usually results in the formation of an adenomatous polyp, with further progression leading to carcinoma. The probability of invasion through the basement membrane increases with the size of the polyp. A significant body of molecular evidence has enhanced this series of morphological events.

A number of molecular lesions have been found which occur at different stages of tumourigenesis. These observations include the growth promoting genes k-Ras, deleted in colorectal cancer (DCC) and p53. k-Ras protein transmits growth signals to intracellular targets when it binds guanosine triphosphate (GTP), and the signal is terminated by the hydrolysis of GTP. Mutant alleles of k-Ras hydrolyse GTP inefficiently leaving a continuous growth signal. Such mutations have been found in 30-50% of colorectal carcinomas (Vogelstein *et al.*, 1988). The DCC gene was localised to a region on chromosome 18q and identified as being reduced in expression in colorectal cancers in comparison to normal colon (Fearon *et al.*, 1990). The DCC protein encodes a receptor for Netrin-1, a molecule involved in axon guidance and may transmit growth-restraining signals between cells (Narayanan *et al.*, 1992). p53 is a transcriptional regulator that prevents cells from progressing forward through the cell division cycle, thereby providing maximal opportunity for DNA repair before the lesions are retained by DNA replication and mitosis (Kuerbitz *et al.*, 1992). Both p53 and DCC are lost late in tumor progression but are unlikely to be the only genes with loss of expression in sporadic colorectal cancer (See Sections 1.4.3 and 1.5).

The theory that multiple mutations are required for the development of tumours is now widely accepted and is described in a genetic model by Kinzler and Vogelstein (1996) (Figure 1.3). At least seven genetic changes are required in this model, as both alleles of a tumour suppressor gene need to be inactivated. In mismatch repair deficient tumours (HNPCC and sporadic) the mutation rate is 2-3 times higher than in normal cells (Shibata *et al.*, 1994). However, in the 85% of colorectal cancers that are mismatch repair proficient (FAP and sporadic) the mutation rates are comparable to those in normal cells, and are characterised by genetic alterations leading to gross chromosomal abnormalities. A conclusion may be that cancers need to obtain a type of instability, but

that chromosomal instability provides minimal selective growth advantage compared to tumours with mismatch repair deficiency and vice versa. It would seem that there are two ways in which a tumour may develop the genetic lesions required for malignancy: subtle alterations due to defective mismatch repair in a minority of cases while gross chromosomal abnormalities occur in the majority.

1.7 COLORECTAL CANCER AND METASTASIS

Metastasis in colorectal cancer involves the overexpression or underexpression of genes involved in specific biological processes. Angiogenesis, proteolysis, motility and cell adhesion are the principal targeted processes. Angiogenesis is the process by which cancer cells secrete substances that promote the formation of blood vessels into the body of the tumour. Proteolytic degradation of the extracellular matrix (ECM) is a limiting step in metastatic spread of tumour cells. The interplay and relative concentrations of proteolytic enzymes and proteolytic enzyme inhibitors are key events in metastasis. The motility and adhesion of tumours cells are inversely related. Metastatic tumours cells have greater motility and are capable of invading tissues to reach circulating blood. Increased motility is often caused by down-regulation of cell adhesion molecules (CAM) in metastatic tumour cells.

The acquisition of the metastatic phenotype is likely to involve the activation and repression of a set of genes. However, some specific candidate genes have been discovered to be involved in the metastatic process. A number of these involved in colorectal metastasis include the *car* (cell adhesion regulator) gene (Yamamoto *et al.*, 1996), *cox2* (cyclooxygenase 2) (Tsuji *et al.*, 1997), E-cadherin and α -catenin (Gofuku *et al.*, 1999). In particular, these genes were characterised as either overexpressed or underexpressed between primary and metastatic colorectal tumours. Despite these advances, the complexity of metastasis and the variability of tumours from independent sources would suggest that changes in metastatic progression involve a range of genes. The roles for the aforementioned metastasis associated genes are unclear.

Although the formation of primary tumours is well characterised, the progression from primary tumour to metastatic tumour is poorly understood at the molecular level. The isogenic paired cell lines SW480 (Dukes' B, Leibovitz *et al.*, 1976) and SW620 (Dukes' C, Leibovitz *et al.*, 1976) are a convenient model of colorectal metastasis. SW480 was derived from a primary colorectal tumour and SW620 was derived from the lymph nodes of the same patient. Initiation of SW620 coincided with metastasis in this patient (Leibovitz *et al.*, 1976). A number of studies have utilised SW480 and SW620. One such study showed that SW620 injection into the caecum of nude mice induced hepatic metastases while SW480 injection did not (Witty *et al.*, 1994). Moreover,

Sampson-Johannes (1996) demonstrated loss of integrin $\alpha 3$ and NCAM (cell adhesion molecules) in SW620 with respect to SW480.

Studies on SW480 demonstrate a truncated APC (adenomatous polyposis coli) gene product (Smith *et al.*, 1993), p53 mutations (Hermeking *et al.*, 1997) and chromosomal abnormalities (CIN) (Cahill *et al.*, 1998). Hence, SW480 exhibits typical primary tumour properties and the SW480/SW620 *in vitro* cell system maintain their parent pre-metastatic and metastatic characteristics. This makes SW480 and SW620 an ideal *in vitro* model of the metastatic process.

1.8 OBJECTIVES OF THIS RESEARCH PROJECT

It is apparent that two clear pathways exist in sporadic colorectal cancer progression. The MIN (microsatellite instability) pathway is recognised by inactivating mutations in components of DNA mismatch repair, which leads to widespread nucleotide instability. The CIN (chromosomal instability) pathway is associated with inactivating mutations in APC which is followed by mutations in K-Ras, deletions in DCC (deleted in colorectal cancer) and inactivation of p53. These genetic events lead to gross chromosomal abnormalities.

The differences in these pathways are likely to be controlled by the coordinate repression and activation of expression of a number of genes. This project has selected a panel of sporadic colorectal cancer cell lines (Table 1.2) as a convenient *in vitro* model of tumorigenesis. A range of gene expression profiling techniques were used to identify gene expression differences within the panel of mismatch repair deficient and mismatch repair proficient cell lines. The specific goals of this research project were as follows:

- Employ Differential Display Reverse Transcription PCR (DDRT-PCR) to identify putative differentially expressed genes between mismatch repair deficient and mismatch repair proficient sporadic colorectal cancer cell lines
- Implement Gene Discovery Array Filters v1.3 to obtain gene expression profiles from pairs of mismatch repair deficient and mismatch repair proficient sporadic colorectal cancer cell lines
- To confirm differential expression of genes using Semi-Quantitative Reverse Transcription PCR (RT-PCR) and Northern Blotting

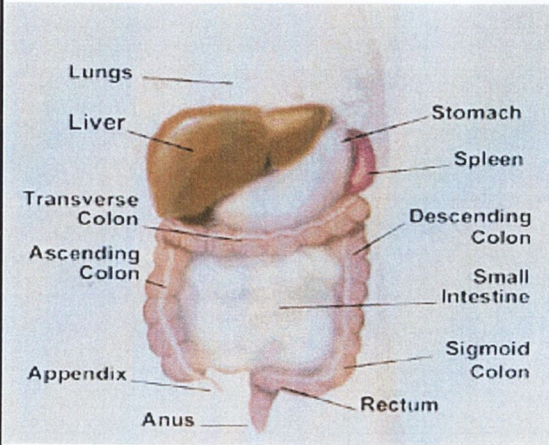
- To investigate differential expression of candidate genes associated with pre-metastatic and metastatic colorectal cancer by semi-quantitative RT-PCR and western blotting

1.9 THE SPORADIC COLORECTAL CANCER CELL LINES STUDIED

The overall goal of this project was to address differential gene expression in a panel of sporadic colorectal cancer cell lines. Within this panel, a number of cell lines are known to exhibit mismatch repair deficiency (LS174T, SW48 & HCT116) and the remaining cell lines are mismatch repair proficient (Caco2, SW480 & HT29) (See Table 1.2). A number of studies which have included these cell lines. Heinen *et al.* (1995) studied a panel of colorectal cancer cell lines, with or without full-length APC protein, for the presence of microsatellite instability. These cell lines were assessed for instability by Polymerase Chain Reaction (PCR) screening of 10 loci containing dinucleotide repeats. The three cell lines (LS174T, SW48 & HCT116) that maintained normal APC protein exhibited the highest degree of instability. These findings supported the idea of a second pathway in colorectal cancer development. It has become apparent that a small fraction of colorectal cancers exhibit widespread genomic nucleotide instability due to defective mismatch repair (microsatellite instability; MIN). The majority of colorectal carcinomas display chromosomal instability (CIN) leading to gross chromosomal abnormalities. Cahill *et al.* (1998) investigated molecular aspects of CIN by studying a panel of MIN and CIN colorectal cancer cell lines. All CIN cell lines tested were defective for the human homologue to yeast *BUB1* (*hBUB1*) gene. This gene is known to be a vital mitotic checkpoint component in yeast and important for chromosome segregation. The normal mitotic checkpoint of cells displaying the MIN genotype becomes defective upon transfer of mutant *hBUB1* alleles from CIN cancers. This study verifies the distinct genetic components of MIN and CIN colorectal cancers.

The previous experiments described above confirm these cell lines as being representative of MIN or CIN cancers and thus, are suitable for use in the study of mismatch repair deficiency in sporadic colorectal cancer.

(A) Lower Intestinal Anatomy



(B) Histology of colorectal passage lining

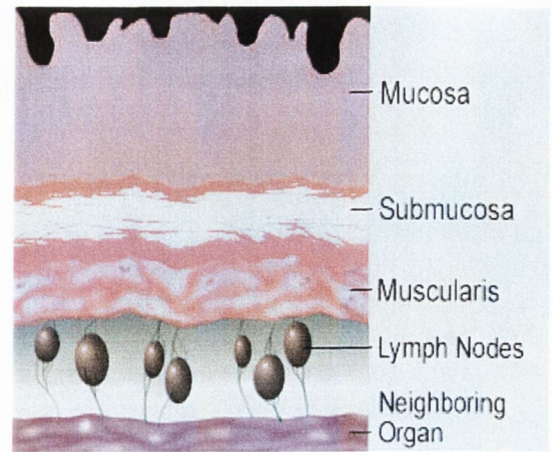


Figure 1.1 Lower intestinal anatomy and tissue composition of the colorectal passage. **A.** Illustration of the anatomy of the lower intestine detailing the architecture of the ascending colon, transverse colon, descending colon and the sigmoid colon. **B.** Illustration of the tissue types lining the colorectal passage where colorectal adenomas are initiated.

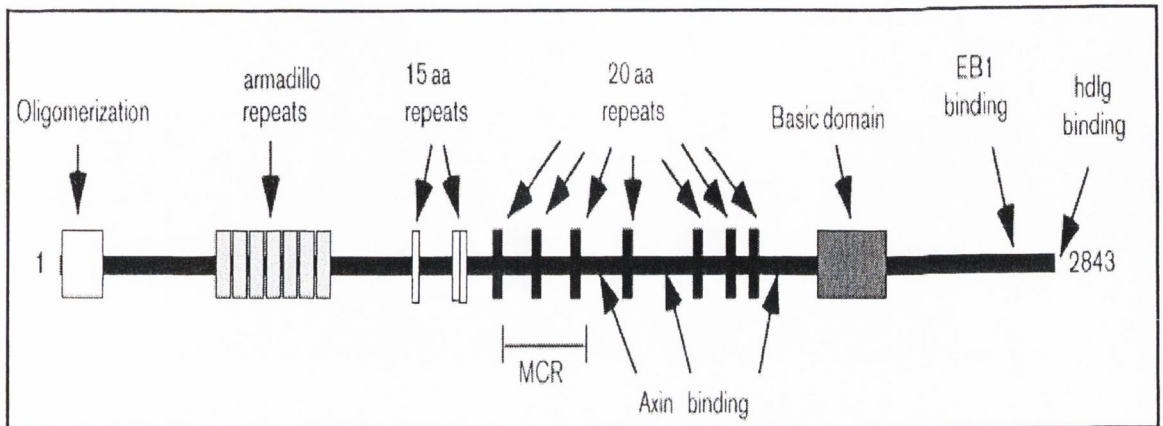


Figure 1.2 Structure of the Adenomatous Polyposis Coli (APC) gene. Linear representation of the APC polypeptide chain detailing the seven armadillo repeat (20 amino acids in length) sequences first identified in the *Drosophila melanogaster* β -catenin gene. The 20 amino acid repeat sequences bind β -catenin in a phosphorylation-dependent manner. The Multiple Cluster Region (MCR) is the target region of the majority of APC-truncating mutations in sporadic colorectal tumours (Diagram reproduced from Polakis, 1999).

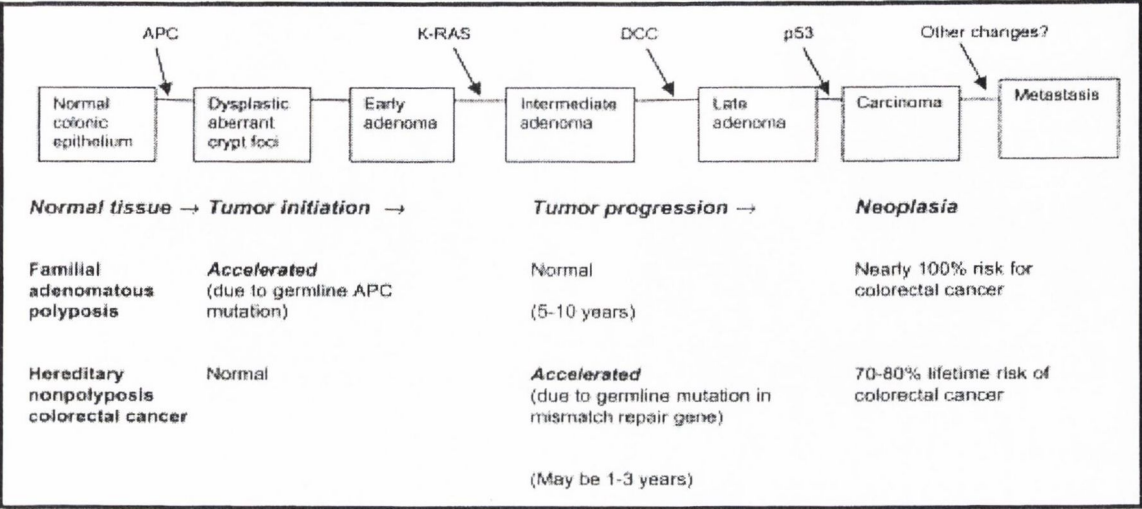


Figure 1.3 A genetic model for colorectal tumourigenesis. APC mutations initiate the neoplastic process and tumour progression results from mutations in other target genes as indicated. FAP patients inherit APC mutations which when coupled with mutations in K-Ras, deletions in DCC (deleted in colorectal cancer) promote tumour progression. HNPCC patients go through a similar but not identical series of inactivating genetic defects; mismatch repair deficiency accelerates the process. Sporadic colorectal cancer follows either of these pathways of tumourigenesis with APC mutations being the more prominent (Schematic diagram reproduced from Vogelstein & Kinzler, 1996)

Table 1.1 Dukes' classification of colorectal tumours including modifications

Dukes' Stage	Pathological description
A	Growth restricted to rectal wall
B	Expansion to extrarectal tissues
C	Metastasis to distant lymph nodes
C ₁	Metastasis restricted to lymph nodes
C ₂	Nodal spread to extrarectal blood vessels

The stages denoted in **bold** represent the original Dukes' staging (Dukes *et al.*, 1932). The substages C₁ and C₂ (Gabriel *et al.*, 1935) represent subsequent modifications.

Table 1.2 Description of the sporadic colorectal cell lines used in this study

A.) MSI sporadic colorectal cancer cell lines	
Cell Line	Description
LS174T	Isolated from a Duke's B adenocarcinoma in a 58-year-old patient. LS174T was produced by trypsinising LS180 cultures isolated from the same patient (Tom <i>et al.</i> , 1976)
SW48	Isolated from a Duke's D adenocarcinoma in a 82 year old Caucasian female (Leibovitz <i>et al.</i> , 1976)
HCT116	Isolated from a male patient with colonic carcinoma (Brattain <i>et al.</i> , 1979)
B.) MSS sporadic colorectal cancer cell lines	
Cell Line	Description
Caco2	Isolated from a primary colonic tumor using the explant culture technique (Fogh <i>et al.</i> , 1977)
SW480	Isolated from a Duke's C adenocarcinoma in a 51 year old Caucasian male (Leibovitz <i>et al.</i> , 1976)
HT29	Isolated from a primary colonic tumor using the explant culture technique (Fogh <i>et al.</i> , 1964)
SW620*	Isolated from a lymph node metastatic lesion in a 51 year old Caucasian male (Leibovitz <i>et al.</i> , 1976)
C.) Additional sporadic colorectal cancer cell lines used in protein analyses	
Cell Line	Description
SW1116	Isolated from a Grade II adenocarcinoma in a 73 year old Caucasian male (Leibovitz <i>et al.</i> , 1976)
498Li	

* denotes a lymph node metastasis sporadic colorectal cancer cell line included in Differential Display RT-PCR experiments

CHAPTER 2

MATERIALS AND METHODS

2.0 MATERIALS

2.0.1 Reagents

General Chemicals and Solvents

Amresco®, USA: 6% Gene-PAGE sequencing mix, Phenol saturated solution.

B.D.H. Ltd., UK: Boric Acid, Chloroform, Ethanol, Ethylenediaminetetra-acetic acid disodium salt (EDTA), Glycine, Hydrochloric acid, Iso-Amyl alcohol, Isopropanol, NNN'-tetramethylethylenediamine (TEMED), Sodium chloride, Sodium Dodecyl Sulphate, Sodium Sulphate, Sodium Tetraborate, Sucrose, Tris-(hydroxymethyl)-methyllamine, Urea.

Boehringer GmbH., Germany: Isopropylthio- β -D-galactoside (IPTG), 5-bromo-4-chloro-3-indoyl- β -D-galactoside (X-gal).

Flowgen, UK: Longer Ranger™ Gel 50% stock solution of acrylamide:bis-acrylamide (19:1).

Merck, Germany: Ethylene glycol-bis (β -amino ethyl ether) N,N,N',N'-tetraacetic acid (EGTA), Potassium dihydrogen phosphate, di-Potassium hydrogen phosphate, di-Sodium hydrogen phosphate.

National Diagnostics, USA: Protogel, ultra pure 30% acrylamide solution.

Scotlab, UK: Easigel™ 6% denaturing acrylamide solution.

Sigma Chemical Company Ltd., UK: Acrylamide/bis-acrylamide 30% solution electrophoresis grade (ratio 37.5:1), Agarose (Type 1A), Ammonium acetate, Ammonium persulphate, Ampicillin, Bovine Serum Albumin Fraction V, Bromophenol Blue, Diethyl Pyrocarbonate (DEPC), Ethidium Bromide, Formaldehyde 37% solution (Formalin), Formamide, Lysozyme, Magnesium chloride, Magnesium sulphate, Mineral oil, 3-(N-Morpholino) propanesulphonic acid (MOPS), Polyvinylpyrrolidone, Sephadex

G-25, Sheared Salmon Sperm DNA, Sodium Dihydrogen Phosphate, Sodium Hydroxide, Xylene cyanol.

Tissue Culture Reagents and Cell Lines

B.D.H. Ltd., UK: Dimethyl Sulphoxide.

Corning Costar, USA: 75cm³-7170 tissue culture flasks, sterile Pasteur pipettes, cell scrapers.

Gibco-BRL, UK: Eagle's MEM (minimum essential medium), McCoy's 5A medium, Leibovitz L-15, Sodium Pyruvate, L-Glutamine, Foetal Bovine Serum, Penicillin-Streptomycin, Trypsin-EDTA.

Oxoid Ltd., UK: Agar, Tryptone, Yeast extract.

American Type Culture Collection (ATCC), USA: Panel of Colorectal Adenocarcinoma cell lines (Table 1.2).

Radiochemicals and Photographic Film

Amersham Pharmacia Biotech, UK: [α -³²P] dCTP (6000Ci/mmol), [γ -³²P] dCTP (6000Ci/mmol).

NENTM Life Science Products, USA: [α -³³P] dCTP (6000Ci/mmol).

Fuji Photo film Co., Japan: Fuji RX film.

Polaroid Corporation, USA: Polaroid 665 film.

Molecular Biology Kits and Reagents

Ambion, USA: NorthernMax Hybridisation Buffer.

Amersham Pharmacia Biotech, UK: Cycle sequencing kit-RPN2436, Ficoll, Hybond-N+ Nylon Transfer Membrane, Nylon Mesh, Oligonucleotide primers, ProbeQuant™ G-50 Micro Columns, Rediprime DNA Labelling System.

Biotechx Labs Inc., USA: Ultraspec™ II RNA Isolation System.

Boehringer GmbH., Germany: *Eco* RI (10U/μL), Glycogen 20μg/μL, RNase A

Clontech, USA: SMART™ PCR cDNA Synthesis Kit.

Genome Systems, Inc., USA: Gene Discovery Array Analyses, Gene Discovery Array Library clones, Gene Discovery Array Human I Version 1.3 Filters.

Gibco-BRL, UK: DNase I amplification grade (1U/μL), Message Maker Poly-A (+) Isolation Kit, Oligo dT(12-18) primer, Poly-A (+) Standard, Random hexamers (3μg/mL), Superscript™ II Moloney Murine Leukaemia Virus Reverse Transcriptase kit, T4 DNA Ligase (1U/μL).

Invitrogen, USA: TOP10 One Shot™ competent *E.coli* cells.

Pharmacia, UK: Cy5 labelled M13 Forward and Reverse primers.

Promega Corporation, USA: dATP, dCTP, dGTP, dTTP, *Hae* III(10U/μL), Polynucleotide Kinase Labelling Kit, , RNasIN(40U/μL) , *Taq* DNA Polymerase(5U/μL), Wizard® PCR preps DNA purification system – A71 70, Wizard® SV plasmid purification kit.

Sigma Ltd., UK: Sheared Salmon Testis DNA (10mg/mL).

Stratagene, USA: pBluescript II SK(+) DNA vector, pCR-Script™ Amp SK(+) cloning kit.

Protein Analysis

B.D.H. Ltd., UK: Copper sulphate, DPX mountant, Isobutanol, Tween-20.

Beckman Instruments Inc., USA: Polyallomer bell-top Quick Seal® centrifuge tubes.

Cadburys, UK: Marvel Non-fat dried milk powder.

Gelman Sciences, UK: PALL Polyvinylidene difluoride (PVDF) transfer membrane.

Merck, Germany: Folin-Ciocalteu's phenol reagent.

Munktell, Sweden: Filter paper, Grade 1F.

New England Biolabs (NEB), UK: Phototope®-HRP Western Blot Detection Kit.

Santa Cruz Biotechnology, Inc.: Polyclonal goat anti-human Protein Phosphatase 1 beta (PP1β) antibody, Horseradish Peroxidase linked anti-goat IgG antibody.

Sigma Ltd., UK: Acrylamide: bis-acrylamide, (205kDa to 29kDa) molecular weight standard marker – SDS-6H, Leupeptin, Mercaptoethanol, Methanol, Phenylmethylsulfonyl fluoride α-toluenesulfonyl fluoride (PMSF), Ponceau S solution, Potassium sodium tartrate, Sodium bicarbonate, Triton X-100.

Shandon, USA: Haematoxylin solution.

2.0.2 Stock solutions

Universal

Loading dye: 0.25g bromophenol blue, 0.25g xylene blue, 30mL glycerol per 100mL.

SSC (20X): 3M NaCl, 0.3M Sodium citrate pH7.0.

TAE (50X): 242g Tris base, 57.1mL glacial acetic acid, 100mL 0.5M EDTA (pH8.0) per litre.

TBE (10X): 0.89M Tris base, 0.89M Boric Acid, 0.02M EDTA disodium, pH7.0

TE: 10mM Tris-HCl, pH7.5, 1mM EDTA.

Cloning media / Plasmid Preparation

Luria Bertani (LB) Broth: 2.5g Yeast Extract, 5g NaCl, 5g Tryptone in 500mL sterile water.

LB agar: 7.5g Agar, 2.5g Yeast Extract, 5g NaCl, 5g Tryptone in 500mL sterile water.

TES: 50mM Tris HCl (pH8.0), 0.25% Sucrose, 2mM EDTA (pH8.0).

M-STET: 5% Triton X-100, 50mM EDTA (pH8.0), 50mM Tris-HCl (pH8.0), 5% Sucrose in 200mL sterile water.

SOB: 2% Tryptone, 0.5% Yeast Extract, 0.05% NaCl, 2.5mM KCl, 10mM MgCl₂.

SOC: 1% (v/v) 2M Glucose in SOB.

RNA Analyses

DEPC-treated H₂O: 500μL DEPC were made up to 500mL with sterile water, mixed thoroughly and incubated at 37°C overnight. The solution was autoclaved for 1 hour.

Hybridisation/prehybridisation Assay buffer: 6.25mL 20X SSPE (3M NaCl, 0.2M NaH₂PO₄, 20mM EDTA), 2.5mL 50X Denhardt's solution (5g Ficoll, 5g polyvinylpyrrolidone, 5g bovine serum albumin and sterile H₂O to 500mL), 1.25mL 10% SDS and 25μL of 100ng/mL sheared salmon sperm DNA in a total volume of 25mL.

MOPS (10X): 0.2M 3-(N-Morpholino)-propanesulphonic acid, 0.05M Sodium acetate (pH7.0), 0.01M EDTA.

RNA denaturing buffer: 30% (v/v) Formamide, 56% 10X MOPS (v/v), 32% Formaldehyde (37%) (v/v).

Sodium Phosphate buffer: 13.14g Na_2HPO_4 , 14.82g NaH_2PO_4 in 200mL DEPC-treated H_2O .

Wash A: 0.5% (w/v) BSA Fraction V, 5% SDS, 40mM Sodium phosphate buffer (pH6.8), 1mM EDTA.

Wash B: 1% SDS, 40mM Sodium phosphate buffer (pH6.8), 1mM EDTA.

Stripping buffer: 0.1XSSC, 0.1% SDS.

Protein Analyses

Electrophoresis running buffer (5X): 15g Tris (pH8.3), 72g Glycine, 5g SDS in 1L H_2O .

Sample buffer (5X): 0.0625M Tris base (pH6.75), 2% SDS, 5% β - mercaptoethanol, 10% Glycerol, 0.001% Bromophenol blue.

Coomassie blue: 0.25% Coomassie⁺ brilliant blue G250, 45% Acetic acid, 45% Methanol, 10% Glacial acetic acid, 10% sterile H_2O . Solution was filtered through Whatman filter paper No.1.

Protein gel destain solution: as for Coomassie stain above with the omission of Coomassie.

Buffer A: 20mM Tris-HCl (pH7.5), 0.25M Sucrose, 2mM EGTA, 2mM EDTA, 1mM PMSF, Leupeptin (10mg/mL).

Buffer B: 20mM Tris-HCl (pH7.5), 1% Nonidet-P40, 150mM NaCl, 1mM EGTA, 1mM EDTA, 1mM PMSF, Leupeptin (10 μ g/mL).

Buffer C: 20mM Tris-HCl (pH7.5), 1% SDS, 150mM NaCl, 1mM EGTA, 1mM EDTA.

Lowry solution: 2% sodium bicarbonate, 0.2% Potassium sodium tartrate, 0.1M Sodium hydroxide.

Bradford solution: 0.01% Coomassie brilliant blue G250, 50mL 95% ethanol, 100mL 85% orthophosphate (H_3PO_4) in 1L distilled H_2O .

Western blot transfer buffer: 0.384g SDS, 5.82g Tris, 2.94g Glycine, 200mL Methanol in 1L distilled water.

PBS (10X): 14.5g Na_2HPO_4 , 2g KH_2PO_4 , 2g KCl, 80g NaCl, pH7.4 in 1L.

Blocking buffer: 1X PBS, 0.1% Tween-20, 5% (w/v) in non-fat dried milk.

Wash buffer (10X): 100mM Tris-HCl (pH9.5), 100mM NaCl, 10mM MgCl_2 .

TBS: 100mL 0.05M Tris (pH7.6), 7.65g NaCl, 25 μL Tween-20 in 1L.

Instrumentation Used

ATTO, Japan: HorizBlot semi-dry transfer blotter - AE-6675, Dual Slab Chamber electrophoresis rigs - AE-6220.

Hybaid, USA: Hybridisation Oven.

Jouan S.A., France: RC 10.10. SpeedVac

LEP, UK: LEPTM Thermal Cycler.

MJ Research, UK: DNA Engine.

Molecular Dynamics, USA: MD-Storm 860 Phosphorimager.

Shandon, USA: Cytospin II.

Unicam, UK: 8625 UV/VIS spectrophotometer.

UVP, USA: Transilluminator.

Wallac, USA: 1214 Rackbeta Liquid Scintillation Counter.

2.1 METHODS

2.1.1 Cell Culture

A panel of human adenocarcinoma cell lines (Table 1.2) was obtained from the American Type Culture Collection (ATTC, USA). The cell line 498Li was received as a kind gift from Dr. Meenhard Herlyn, The Wistar Institute, Philadelphia, USA. These cells were grown in media (Section 2.0.2) as described by ATCC in 75mL tissue culture flasks (specific growth conditions; Table 2.1). Passage was carried out when cells reached 70-80% confluence and typically split in a ratio of 1 : 5. Trypsin-EDTA (1X) was used to lift the cells prior to passage. This was carried out by discarding the exhausted medium and rinsing the flask with PBS Dulbecco's medium (Gibco BRL, UK). An aliquot (1-2mL) of 1X Trypsin-EDTA was added to the flask and was subsequently swirled and finger tapped. Cells were seen to detach from the flask after 10-15 minutes. The Trypsin-EDTA was neutralised by adding 10ml of medium and the cells were centrifuged at 1000rpm for 3 minutes. The supernatant was discarded and the cells resuspended in fresh medium and aliquoted into new tissue culture flasks. Cells were grown in an aseptic incubator at 37°C with 5% CO₂ (except for SW480, SW620, SW1116 & 498Li which grow in the absence CO₂).

2.1.2 General molecular biology techniques

Agarose Gels

Agarose gels ranging from 0.8-2.0% were prepared as follows. A stock (200mL) of 1% agarose was made by adding 2g of Type 1A agarose (Sigma Ltd., UK) to 200mL of 1X TBE or 1X TAE as appropriate. This mixture was boiled gently until the agarose was completely in solution. Gels were cast in gel trays and allowed to stand at room temperature for 20 minutes. Agarose gels were stained in an ethidium bromide bath (prepared by the addition of 75µL of 10µg/mL ethidium bromide to 500mL of distilled water). Agarose gels were placed in distilled water to destain excess ethidium bromide for 30 minutes where appropriate. Loading dye was added to each sample (0.1 to 0.2

volumes) to gauge progress of electrophoresis. Gels were subjected to electrophoresis for 30-60 minutes at 40mA.

Agarose gels for Northern blotting were prepared in formaldehyde as follows: 1.5g agarose was dissolved in a solution containing 15mL 10X MOPS buffer and 109.5mL DEPC-treated water by gentle boiling. This solution was allowed to cool for 15 minutes followed by the addition of 26.5mL formaldehyde (37%) to a final volume of 150mL. Northern blot gels were cast in large gel trays for a minimum of 45 minutes. Loading dye for northern blot gels was prepared in 50% (v/v) glycerol containing 0.25% bromophenol blue (w/v) in DEPC-treated water. Gels for northern blotting were run for approximately 4 hours at 100mA. Denaturing agarose gels for non-Northern analyses were prepared by scaling down the volumes described above accordingly and run at 60mA for 1.5 hours.

Polyacrylamide gels

Polyacrylamide gels were prepared as follows: 24mL of 30% ultrapure acrylamide, 29mL distilled water and 6mL of 10X TBE were added to a graduated cylinder and mixed thoroughly. Polyacrylamide gel plates were rinsed in distilled water and washed with 70% ethanol. Ammonium persulphate (378 μ L of 10% (v/v)) and TEMED (75%) were added to the acrylamide mix to induce polymerisation. Gels were poured rapidly into pre-prepared taped polyacrylamide gel plates and allowed to set for 1 hour and were run for 2 hours at 30mA. Gels for Differential Display RT-PCR were made in a similar manner with the volumes of the components scaled up.

Gels for the ALFexpress™ automated sequencer were made up as described below. ‘Short’ ALF gels were made by mixing 40mL 6% Gene-PAGE or EasiGel™ solution with 400 μ L 10% APS and 40 μ L TEMED and were allowed to polymerise for 1 hour. Sequencing (‘long’) gels were formulated by mixing 25g urea, 7.2mL Longer Ranger™ gel solution, 7.2mL 10X TBE to a volume of 60mL with distilled water. This mixture was dissolved by gentle heat and polymerised by the addition of 375 μ L 10% APS and 38 μ L TEMED.

Molecular Weight Markers

Agarose and polyacrylamide gels were all electrophoresed with the inclusion of molecular weight markers that were prepared by restriction enzyme endonuclease digestion of plasmid DNA. Specifically, 20µg of pBlueScript SK(+) plasmid was digested for 2 hours with 50U of *Hae* III restriction enzyme at 37°C. In other cases (for different sized PCR products) 20µg of pBlueScript SK(+) plasmid was digested for 2 hours with 50U of *Hha* I restriction enzyme. An aliquot of 3µL of marker was used for agarose and acrylamide gels.

Nucleic Acid Precipitations

Samples of DNA and RNA were routinely concentrated by standard nucleic acid precipitation methods. DNA was precipitated by adding 1/10 volume of 3M sodium acetate, 2.5 volumes of 100% ethanol and 50µg glycogen followed by brief vortexing and storage at -70°C for 1 hour or overnight. Precipitated DNA was pelleted by centrifugation at 10,000 rpm for 5 minutes. The supernatant was removed and the DNA pellet washed once in 70% ethanol. The DNA was finally resuspended in a small volume of sterile water.

RNA was precipitated in a similar fashion with a number of modifications. Ethanol precipitation of RNA was carried out by adding 50µg of glycogen to the RNA sample, 0.1 volume of 7.5M ammonium acetate and two volumes of -20°C ethanol. This was mixed by inversion and stored at -20°C for 48 hours. The precipitation mixture was centrifuged at 2,600g for 30 minutes at 4°C and the RNA pellet was washed with 75% ethanol followed by centrifugation for 10 minutes at 2,600g at 4°C, air dried and resuspended in a small volume of RNase-free water. RNA samples were also concentrated using a SpeedVac.

Gel Purification of Nucleic Acids

Impure nucleic acids (contaminated PCR products with non-specific products or excess primers) were routinely gel purified using the Wizard® PCR Preps DNA purification system (Promega, USA). This was carried out largely according to the manufacturer's directions except that the solubilisation of the agarose was carried out at room temperature and the elution volume was reduced to 25µL in most cases.

2.1.3 RNA Isolation

Total RNA was isolated from cultured cells using the Ultraspec II RNA isolation system (Biotech Labs. Inc., USA) after reaching 70% confluence. Exhausted medium was discarded and the cells rinsed briefly in 5mL of PBS Dulbecco's medium (Gibco BRL, UK). The cells were harvested in 1mL of Ultraspec and gently homogenised before transfer to sterile microcentrifuge tubes on ice. Nucleoproteins were allowed to dissociate for 5 minutes and 0.2mL of chloroform : isoamyl alcohol (24 : 1) were added and tubes shaken vigorously before centrifugation at 12,000 rpm at 4°C for 15 minutes. The aqueous phase (0.8 volume) containing the RNA was carefully removed and transferred to a sterile microcentrifuge tube to which a 0.5 volume of isopropanol and a 0.05 volume of RNATackTM Resin were added. This mixture was vortexed briefly followed by a 1 minute spin at 12,000 rpm. The supernatant was removed and the pellet was washed twice in 1mL of 75% ethanol. The pellet was suspended in 30-50 µL of RNase free water. RNA samples were DNase treated using DNase I Amplification Grade. Additions were made as follows to a sterile microcentrifuge tube: 50µg of total RNA, 5µL 10X DNase I buffer, 1 µL RnasIN (10U/µL) (Promega Corporation, USA), 1 µL DNase I (1U/µL) in a 50 µL reaction volume at 25°C for 15 minutes. DNase I treatment was terminated by the addition of 2.5mM EDTA with a 10 minute incubation at 70°C. RNA was aliquoted and stored at -70°C under 70% ethanol.

2.1.4 RNA Analysis and Quantitation

The integrity of the RNA samples was determined by denaturing formamide agarose gel electrophoresis. Specifically, 1µg of total RNA was incubated in RNA denaturing buffer at 65°C for 5 minutes. Each sample was chilled on ice followed by the addition of 1µL RNA loading dye. Samples were electrophoresed for 1-2 hours using 1X MOPS buffer as running buffer. The gels were viewed on a UVP Transilluminator. The 28S and 18S ribosomal RNA bands were clearly evident with a faint smear migrating between them indicating the messenger RNA.

Aliquots of RNA were quantitated in a quartz cuvette using the Unicam 8625 UV/VIS spectrophotometer. All RNA samples were quantitated at wavelengths of 260nm and 280nm and the A260/A280 ratios were calculated, which should fall in the range of 1.8-2.1 (Sambrook *et al.*, 1989).

2.1.5 Hybridisation Assay for mRNA quantitation

The method of oligo-dT hybridisation was employed to accurately quantitate poly-A (+) levels in all total RNA samples (Harley *et al.*, 1987). An oligo-dT primer was radiolabeled by making the following additions to a sterile microcentrifuge tube: 1μL of oligo-dT primer (240pmol), 3μL T4 Polynucleotide Kinase buffer, 1μL [γ - 32 P] dATP, 21μL sterile H₂O, 1μL Polynucleotide Kinase (5U/μL). This labeling reaction was incubated for 1 hour at 37°C. The labeled probe was purified by passing it through a G-25 Sephadex column. The hybridisation/prehybridisation assay buffer solution was incubated for 1 hour at 20°C. RNA samples were denatured in a buffer containing 1X SSC (3M NaCl, 0.3M sodium citrate) and 7.4% formaldehyde. A standard poly A(+) RNA (Amersham Pharmacia Biotech, UK) was dotted in triplicate in the concentration range of 0 – 45 ng/μL onto Hybond-N filters which had been divided into 1cm² sections using a pencil. The unknown RNA samples were also dotted onto the filters in triplicate. The RNA was fixed to the filter by soaking in 0.05M NaOH. The membrane was presoaked in 2X SSPE. Hybridisation was carried out for 2 hours at room temperature and washed in 2X SSC for 5 minute intervals until a Geiger Counter registered counts of 250-500cpm. The blot was exposed to Fuji x-ray film for 12-36 hours. The filter was cut up into squares representing each RNA dot. The emission of radioactivity in each cm² was determined by Cerenkov counting using the Wallac 1214 Rackbeta Liquid Scintillation Counter. Triplicate values for counts were averaged. A graph was plotted between the known poly A(+) concentrations and their equivalent counts per minute. The levels of poly A(+) for each unknown sample were determined by extrapolation from the standard curve.

2.1.6 cDNA Synthesis

cDNA was prepared for DDRT-PCR using the Superscript II Reverse Transcription kit (Gibco BRL, UK). AP3L (anchor primer 3L) was used to prime first strand cDNA synthesis (See Table 2.2 for primer sequence). Typically, 1μg of total RNA was incubated in the presence of 5μM AP3L in a volume of 25μL at 70°C for 10 minutes to denature the RNA. The following additions were made to each cDNA synthesis reaction: 5μL 200μM dNTPS (200μM of each of the four deoxynucleotide triphosphates: dATP, dCTP, dGTP, dTTP), 5μL 1mM DTT, 10μL 5X cDNA synthesis

buffer (20mM Tris-HCl (pH7.5), 100mM NaCl, 0.1mM EDTA, 0.01% (v/v) NP-40, 50% (v/v) glycerol) and 10U of RNasIN (Promega Corporation, USA). Each cDNA synthesis reaction was briefly centrifuged and incubated for a further 2 minutes at 37°C. Finally, 200U of Moloney Murine Leukemia Virus (MMLV) Reverse Transcriptase (Gibco BRL, UK) was added to each reaction to a final volume of 50µL and cDNA was synthesised at 37°C for 90 minutes and the reaction terminated at 95°C for 5 minutes. cDNA was stored at -20°C for further use.

cDNA was prepared for Semi-Quantitative RT-PCR as for DDRT-PCR with the exception that random hexamers (0.2µg/mL) (Gibco BRL, UK) were used to prime first strand cDNA synthesis. cDNA was stored at -20°C for further use.

2.1.7 Genomic DNA Screen

cDNA samples were controlled for the presence of contaminating genomic DNA. Specific primers for the human Polycystic Kidney Disease 1 (PKD1) gene (L43619) were designed to amplify across exon 43, intron 43 and exon 44 of the gene. The following PCR components were added to a microfuge tube: 1U Promega *Taq* DNA Polymerase (5U/µL), 5µM PKD1 U primer, 5µM PKD1 L primer, 2µL 10X PCR buffer, 1µL cDNA or 1µL genomic DNA (50ng/µL), 2µL 200µM dNTPs, 1.2µL Mg²⁺ (25mM) to a final volume of 20µL. PCR reactions were overlayed with mineral oil and cycled as follows: 94°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute for 30 cycles followed by extension at 72°C for 5 minutes. PCR products were electrophoresed on 1% agarose gels. PCR products derived from genomic DNA yield a product of 256bp and the PCR product from cDNA template yields a 183bp fragment.

2.1.8 Semi-Quantitative Reverse Transcription PCR

RNA samples from the panel of six cell lines (quantitated spectrophotometrically) were used as template for RT-PCR reactions. cDNA was synthesized as detailed in Section 2.4. Specific primers were designed for the human β -actin transcript (See Table 2.2 for sequences). The amplification conditions for the β -actin PCR were optimised on control human genomic DNA prior to use in RT-PCR experiments. In a total volume of 50µL the RT-PCR reactions contained 5µM of β -actin U and β -actin L primers, 200µM dNTPs, 5µL *Taq* DNA Polymerase 10X buffer, 1.5mM MgCl₂, 1U of *Taq* DNA

Polymerase (Promega Corporation, USA) and 2 μ L cDNA. A mineral oil overlay was added to each RT-PCR reaction. The cycling conditions used were: 95°C for 1 minute, 60°C for 1 minute, 72°C for 1 minute for 30 cycles, followed by extension at 72°C for 5 minutes. Amplification was paused at 21, 24 and 27 cycles and 12 μ L aliquots of RT-PCR product were removed from the reactions. Aliquots (4 μ L) from each RT-PCR reaction at each cycle number were resolved on 1% agarose gels or 8-12% polyacrylamide gels.

2.2 DIFFERENTIAL DISPLAY REVERSE TRANSCRIPTASE PCR

2.2.1 Optimisation of DDRT-PCR Conditions

Previously, DDRT-PCR was carried out in the laboratory using the protocol documented by Bauer *et al.* (1993) with minor modifications. The following additions were made to a PCR tube: 1.2 μ L MgCl₂ (25mM), 1U Stoffel fragment (5U/ μ L) (Perkin Elmer, USA), 5 μ M anchor primer, 5 μ M arbitrary primer, 2 μ L 10X Stoffel fragment buffer, 2 μ L 200 μ M dNTPs, 0.2 μ L [α -³²P] dCTP, 2 μ L cDNA and DEPC-treated water to a final volume of 20 μ L. The PCR mixture was overlaid with mineral oil and cycled as follows: 95°C for 1 minute, 46°C for 1 minute, 72°C for 1 minute for 30 cycles followed by extension at 72°C for 5 minutes. The resulting display patterns yielded insufficient numbers of bands (15-30) and were not reproducible over multiple experiments (data not shown).

In the present study, the DDRT-PCR conditions documented by Ayala *et al.* (1995) were adopted. cDNA templates for DDRT-PCR reactions were prepared as described in Section 2.1.6. PCR reactions were prepared in microfuge tubes as follows: 1.2 μ L MgCl₂ (25mM), 1U *Taq* DNA Polymerase (5U/ μ L) (Promega Corporation, USA), 5 μ M arbitrary primer, 2 μ L 10X Promega buffer, 2 μ L 200 μ M dNTPs, 0.2 μ L [α -³²P] dCTP, 2 μ L cDNA and DEPC-treated water to a final volume of 20 μ L. Reactions were incubated as follows: 95°C for 1 minute, 38°C for 5 minutes, 72°C for 5 minutes. The anchor primer AP3L (2 μ L) was added to each reaction mixture through the mineral oil overlay and reactions were incubated under the following conditions: 95°C for 1 minute, 60°C for 1 minute, 72°C for 1 minute for 35 cycles followed by extension at 72°C for 5 minutes. Primers sequences for AP3L and the arbitrary upstream primers (DD1A-DD6A) are detailed in Table 2.2.

Taq DNA Polymerase, Stoffel Fragment and *Pfu* DNA Polymerase were all tested to determine which enzyme yielded the most useful DDRT-PCR patterns. All enzymes were assayed on one cDNA sample to identify the optimal enzyme concentration. The optimal concentration was then assayed with multiple cDNA samples to determine the stability of the band pattern. Stoffel Fragment yielded very few bands (5-15) with low reproducibility over repeat experiments. *Pfu* DNA

Polymerase yielded more bands than Stoffel Fragment but also resulted in low reproducibility. *Taq* DNA Polymerase gave the best band patterns (40-60 bands per lane) and the greatest reproducibility of the three enzymes assayed (data not shown).

To determine the optimal concentration of magnesium, a series of DDRT-PCR experiments were carried out varying the magnesium concentration from 1.2mM to 2.2mM. This was carried out by adding 0.9µL MgCl₂(1.2mM), 1.2 µL MgCl₂(1.5mM), 1.5 µL MgCl₂(1.9mM) and 1.8 µL MgCl₂(2.2mM) to separate DDRT-PCR reactions. Initially, one cDNA sample was used in these experiments. Magnesium at a final concentration of 1.9mM was seen to give the desired number of DDRT-PCR bands with high reproducibility. This magnesium concentration was assayed over multiple cDNA samples and yielded the desired band pattern stability (data not shown).

Anchor primer AP3L concentrations of 1µM, 5 µM and 10 µM were used to determine the optimal concentration required. This was carried out by adding 0.5 µL, 2.5 µL and 5 µL to separate DDRT-PCR reactions using one cDNA sample. All three AP3L concentrations yielded useful and reproducible band patterns. The variation in concentration did not have a marked effect on the band pattern. The pattern resulting from using multiple cDNA samples was stable. The 5 µM AP3L concentration produced the most stable pattern with multiple cDNA samples and was chosen for further experiments. A number of arbitrary primers (DD1A-DD6A; Table 2.2) were designed for DDRT-PCR analysis. The concentration of each arbitrary upstream primer was optimised separately. The concentration of each arbitrary primer was varied against the 5µM AP3L concentration and the optimal concentration for each arbitrary primer was chosen.

Aliquots of each DDRT-PCR reaction (3µL) were mixed with 2 µL of loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 40% (w/v) sucrose in water) and electrophoresed in parallel on 6% non-denaturing polyacrylamide sequencing gels at 60W for 2-3 hours. Gels were vacuum dried. A pipette tip was used to dot excess radioactive PCR product containing loading dye on the four corners of the dried gels. This allowed alignment of the autoradiographs with the dried gels. All gels were carefully placed in autoradiography cassettes with Fuji X-ray film for 4-32 hours at -70°C.

2.2.2 Cloning of DDRT-PCR Products

All autoradiographs were carefully examined on a light box. Bands representing potentially differentially expressed mRNA species were selected for excision. Bands that were reproducibly differentially expressed were deemed suitable for further analysis. Gels were carefully aligned with their corresponding autoradiographs and taped together in the correct orientation. These were carefully placed on a light box with the gel above the autoradiograph. The position of the bands of interest on the gels was determined with the aid of the autoradiograph band pattern illuminated by the light box. A sterile blade was used to excise the bands of interest which were placed in sterile microcentrifuge tubes.

Bands were suspended in 50µL sterile water and boiled for 5 minutes. The cDNA was allowed to elute from the gel slice at room temperature for 2 hours. Aliquots (2µL) of the cDNA eluates were reamplified with *Taq* DNA Polymerase (Promega Corporation, USA) under the following amplification conditions: 1U *Taq* DNA Polymerase, 2µL of 10X *Taq* DNA Polymerase buffer (50mM Tris-HCl (pH 8.3), 75mM KCl), 5µM upstream primer, 5µM AP3L with 2µL of gel purified DDRT-PCR product as template. Amplification was carried out under the following cycling conditions: 95°C for 1 minute, 60°C for 1 minute, 72°C for 1 minute for 10 cycles with extension at 72°C for 5 minutes. Aliquots of 2 µL were taken from these reactions and used as template under the same conditions for 30 cycles. PCR reactions were pooled and gel purified using the Wizard® PCR Preps DNA Purification System (Promega Corporation, USA). The total volume of PCR products was reduced to 20 µL using a Jouan Speed Vacuum system. Purified cDNAs were stored for use as template for cloning.

The pCR-Script™ Amp SK(+) Cloning Kit (Stratagene, USA) was used to clone the purified DDRT-PCR products. PCR products generated with *Taq* DNA Polymerase may have single nucleotide overhangs. Treatment of PCR products with *Pfu* DNA Polymerase renders sticky ended oligonucleotides blunt-ended by the addition of complementary nucleotides. Purified DDRT-PCR products were *Pfu* DNA Polymerase ‘polished’ before the ligation step. Each polishing reaction was set up as follows: 10µL of the purified PCR product, 1 µL of 10mM dNTP mix (2.5mM of each of the four deoxynucleotide triphosphates: dATP, dCTP, dGTP, dTTP), 1.3 µL of 10X polishing

buffer, and 1 μL of *Pfu* DNA Polymerase (0.5U/ μL) in a sterile microcentrifuge tube. A 20 μL mineral oil overlay was added to each polishing reaction and incubated at 72°C for 30 minutes. The polished PCR product was used directly in the ligation step. Ligations were set up by adding the following in order to a sterile microcentrifuge tube: 1 μL of the pCR-Script Amp SK(+) cloning vector (10ng/ μL), 1 μL of pCR-Script 10X reaction buffer, 0.5 μL of 10mM rATP, 4 μL of polished PCR product, 1 μL *Srf* I restriction endonuclease enzyme (5U/ μL), 1 μL T4 DNA Ligase in a final volume of 10 μL sterile water. The ligation reactions were carried out at room temperature for 1 hour and terminated by incubation at 70°C for 10 minutes.

Ligations were stored on ice prior to transformation. Plasmids were cloned into TOP10 One ShotTM *E. coli* competent cells (Invitrogen, USA). Cells were thawed carefully on ice and supplemented with 0.5M β -mercaptoethanol. Aliquots of ligation reactions (3 μL) were added to the competent cells and stored on ice for 30 minutes. The transformation reactions were heat-pulsed for 30 seconds at 42°C in a water bath followed by 2 minutes on ice. SOC medium (250 μL ; Section 2.0.2) was added to each of the transformation reactions which were then incubated in a 37°C shaker for 1 hour at 225rpm.

Transformation reactions (50 μL and 150 μL) were plated onto Luria Bertani (LB) Ampicillin agar (Section 2.0.2) plates containing 20 μL 5-Bromo-4-chloro-3-indoyl- β -D-galactopyranoside (x-gal) (40mg/mL). Plates were incubated for 12-16 hours at 37°C, and white colonies were picked and replated onto fresh master plates.

2.2.3 Sequencing of DDRT-PCR Products

Colonies were screened for insert by PCR using the Forward and Reverse sequencing primers which flank the insert region of the pCR-Script vector. A flamed loop was used to take colonies from the master plates and dipped into 50 μL sterile water, which was boiled for 5 minutes and centrifuged at 12,000rpm for 2 minutes. Aliquots (2 μL) of these cell lysates were used as template for the insert screen under the following conditions: 1U *Taq* DNA Polymerase (Promega Corporation, USA), 2 μL 10X reaction buffer, 2 μL 200 μM dNTPS, 3 μL of 25mM MgCl_2 , 2 μM of each primer in a final

volume of 20 μ L. The amplification conditions used were as follows: one cycle at 95°C for 1 minute; 94°C for 45 seconds, 60°C for 45 seconds, 72°C for 1 minute for 32 cycles followed by a two minute extension at 72°C. PCR products were separated on 1% agarose gels thereby identifying clones containing the desired insert.

Clones containing inserts were picked from master plates and placed in 10mL of LB Ampicillin broth overnight at 37°C shaking at 225rpm. These cultures were centrifuged at 1000rpm for 10 minutes and the exhausted LB broth was drained off. Cells were resuspended in 60 μ L TES (Section 2.0.2) and transferred to sterile microcentrifuge tubes. Lysozyme (20 μ L) (40mg/mL in 250mM Tris-HCl, pH 8.0) was then added to each plasmid preparation and incubated for 10 minutes on ice. M-STET (550 μ L; Section 2.0.2) was then added at room temperature for a further 10 minutes. The lids were opened and the tubes boiled for 90 seconds followed by centrifugation at 12,000 rpm for 30 minutes. Pellets were pulled out using toothpicks. RNase A (10 μ L) (10mg/mL) was added and incubated at 37°C for 15 minutes. Samples were phenol extracted, and the aqueous phases transferred to 0.6 volume of 100% isopropanol which was incubated at room temperature for a further 15 minutes. Samples were centrifuged for 10 minutes and the supernatants were discarded. Plasmid DNA was washed with 70% ethanol and suspended in 50 μ L sterile water. Plasmid DNA was stored at -20°C.

Plasmids were sequenced using the Thermo Sequenase cycle sequencing kit (Amersham Pharmacia Biotech, UK) on the ALFexpressTM Automated DNA Sequencer (Amersham Pharmacia Biotech, UK). Sequencing was carried out according to manufacturers' guidelines. Aliquots (1 μ L) of each reagent (A, C, G or T) were added to separate microfuge tubes. A 'master mix' was made by making the following additions to a microfuge tube: 2.5 μ L Forward sequencing primer (Cy5 labelled; 10 pmole/ μ L), 0.5-5.0 μ g plasmid DNA in a final volume of 12.5 μ L with sterile water. Master mix aliquots (3 μ L) were added to each reagent (A, C, G or T) followed by the addition of a mineral oil overlay. PCR reactions were cycled as follows: 95°C for 45 seconds, 55°C for 45 seconds and 72°C for 1 minute for 32 cycles and an extension at 72°C for 5 minutes. PCR products were denatured before electrophoresis at 95°C for 5 minutes and then incubated on ice. Sequencing gels were prepared as described in Section 2.1.2. Cycle sequencing PCR products were electrophoresed on 'long' 15% polyacrylamide gels at 60mA overnight at 55°C. Sequence data was acquired using the ALFwin

Evaluation software (Amersham Pharmacia Biotech, UK) and extracted as PC ‘.txt’ files.

2.2.4 Characterisation of DDRT-PCR Products

Sequences obtained from DDRT-PCR experiments were converted to FASTA format and analysed using the BLASTN software package located at the National Centre for Biotechnology Information (NCBI, USA) (Altschul *et al.*, 1990). The Blast server is located at <http://www.ncbi.nlm.nih.gov/BLAST>.

2.3 HIGH DENSITY cDNA ARRAYS

2.3.1 Introduction

Genome Systems Incorporated (GSI) recommends the use of approximately 2.5µg mRNA to synthesise a cDNA probe for the Gene Discovery Array (GDA) filters. cDNA synthesis is primed by a two base anchored oligo-dT primer (dT₁₈VN) in the presence of [α -³³P] dCTP. Control RNA samples from *Arabidopsis thaliana* and *Drosophila melanogaster* are reverse transcribed in the same manner and are combined with the cDNA probe in the hybridisation mixture. A DNA orientation marker is also random-prime radiolabeled with [α -³³P] dCTP to aid in filter orientation, and is added to the hybridisation mixture prior to hybridisation.

2.3.2 Prehybridisation

Prehybridisation was carried out by dispensing 15mL NorthernMax Hybridisation buffer (Ambion, USA) into a suitable RNase-free container. Filters were carefully laid onto the buffer and soaked for 30 seconds. The filters were carefully rolled in nylon mesh and placed in hybridisation bottles followed by the addition of 15mL pre-warmed (42°C) hybridisation buffer. Pre-hybridisation was carried out overnight at 42°C using a hybridisation oven (Hybaid, USA).

2.3.3 cDNA Probe preparation

Two methods were used to generate cDNA probes for GDA analysis. Initially, total RNA samples from each cell line were pooled amounting to 1mg of total RNA. The MessageMaker Kit (Gibco BRL, UK) was used to isolate high quality poly-A(+) mRNA. The total RNA concentration was adjusted to 0.55mg/ml with RNase-free water. The RNA was heated at 65°C for 5 minutes and chilled on ice. The salt concentration of the RNA sample was adjusted to 0.5M NaCl by adding 0.1 volume of 5M NaCl. A volume (2mL) of oligo(dT) cellulose was added to the salt adjusted RNA, mixed several times by inversion. Hybridisation of the RNA with the oligo(dT) cellulose was carried out by incubation at 37°C for 10 minutes. This sample was passed through a Filter syringe and the liquid containing unbound RNA was retained. The oligo(dT) cellulose/RNA mixture was washed with 5mL of Wash Buffer 1. This step

was carried out by drawing 5mL of Wash Buffer 1 into the Filter Syringe, mixing the contents and expelling the buffer. This step was repeated using 5mL of Wash Buffer 2. The mRNA was eluted from the cellulose by drawing 2mL of distilled water (previously heated to 65°C) into the Filter syringe, mixing the contents thoroughly and expelling the mRNA into an RNase-free 1.5mL tube. A double selection was deemed necessary to obtain high quality mRNA. The single selected RNA was reheated to 65°C for 5 minutes.. The salt concentration was adjusted by adding 0.1 volume of 0.5M NaCl as before. The second mRNA selection was carried out as before with hybridisation of mRNA and oligo(dT) cellulose carried out at room temperature for 10 minutes. The oligo(dT) cellulose sediment fines were pelleted by centrifugation at 3,000 g for 5 minutes. The supernatant was transferred to an RNase-free tube. Ethanol precipitation of the mRNA was carried out by adding 50µg of glycogen to the mRNA sample, 0.1 volume of 7.5M ammonium acetate and two volumes of -20°C ethanol. The mixture was inverted once and stored at -20°C for 48 hours. The precipitation mixture was centrifuged at 2,600g for 30 minutes and the mRNA pellet was washed with 75% ethanol followed by centrifugation for 10 minutes at 2,600g at 4°C, air dried and resuspended in 5µL of RNase-free water. mRNA integrity was assessed by denaturing formamide agarose gel electrophoresis and mRNA concentration was determined by spectrophotometry at a wavelength of 260nm. mRNA was aliquoted and stored at -80°C.

cDNA probes were prepared for each of the mRNA samples separately. The following components were combined in an RNase-free tube: 2.5µg sample mRNA, 2.0µL of 50µM 5'-oligo dT₁₈VN. This mixture was incubated at 70°C for 10 minutes and placed on ice for 5 minutes. The sample was lyophilised on low heat until dry and dissolved in the following: 1.4µL 5X MMLV Buffer, 0.4µL 10mM dA/dG/dT mix, 4.0µL [α -³³P] dCTP (10uCi/µL) and 1.2µL MMLV (200U/µL). The cDNA synthesis reaction was mixed by brief vortexing and spinning in a centrifuge. cDNA synthesis was carried out at 42°C for 2 hours.

Labeled cDNA was diluted to a volume of 50 µL with sterile water. A small volume (1 µL) was removed for cpm quantitation. The excess RNA was degraded by the addition of 12.5 µL of 1N NaOH and incubation at 37°C for 10 minutes. This was followed by the addition of 12.5 µL Tris pH 6.8 and 10 µL 1M HCl. Unincorporated

nucleotides were removed from the radiolabeled probes using ProbeQuantTM G-50 Micro Columns (Amersham Pharmacia Biotech, UK) according to the manufacturers protocol. Labeled probes were incubated on ice prior to hybridisation. The lyophilised control RNA was resuspended in 20mL of DEPC-treated water. Control cDNA probes were prepared using the control RNA in the same manner.

The SMARTTM PCR cDNA Synthesis Kit (Clontech, USA) was used as an alternative method for generating cDNA probes. The first strand synthesis was prepared by making the following additions to a sterile microcentrifuge tube: 1µg total RNA, 1µL CDS primer (10µM), 1µL SMART II oligonucleotide (10µM) and 3µL DEPC-treated water. This mixture was incubated at 70°C for 2 minutes prior to making the following additions: 2µL 5X First-strand buffer, 1µL DTT(20mM), 1µL dNTPs(10mM) and 1µL MMLV Reverse Transcriptase(200U/µL). cDNA synthesis was carried out at 42°C for one hour. The first strand reaction was diluted by adding 40µL TE buffer (10mM Tris[pH 7.6], 1mM EDTA). First strand cDNA synthesis was completed by incubation at 72°C for 7 minutes. For each cDNA sample, PCR-Select reactions were prepared in triplicate as described by the manufacturer: 1µL single stranded cDNA (sscDNA), 74µL DEPC-treated water, 10µL 10X cDNA PCR buffer, 2µL dNTPs(10mM), 2µL PCR primer (10µM), 2µL 50X Advantage cDNA Polymerase Mix. The PCR conditions used were: 95°C for 1 minute; 95°C for 15 seconds, 65°C for 30 seconds, 68°C for 6 minutes for fifteen cycles. The optimal number of PCR cycles was determined for each reaction by subjecting one of each of the triplicate reactions to additional cycles (16, 17, 18, 19, 20, 21 cycles). Aliquots of each PCR cycle(5µL) were electrophoresed in parallel to compare relative PCR amplification between different templates. The optimal cycle number was chosen to yield balanced PCR products and the remaining two triplicate reactions were amplified for the additional number of cycles.

SMARTTM PCR products were ethanol precipitated overnight and resuspended in 6.5µL sterile water. Aliquots (0.5µL) of the products were electrophoresed on 1.2% agarose gels. The SMARTTM PCR products were labeled with [α -³³P] dCTP using the Rediprime II labeling kit (Amersham Pharmacia Biotech, UK) according to the manufacturers' guidelines. Each SMARTTM PCR product was diluted to 45µL with TE buffer [pH8.0]. The samples were denatured at 100°C for 5 minutes, briefly chilled on

ice and then added to the Rediprime reaction tubes. [α - ^{33}P] dCTP (5 μL) was added to each labeling reaction and incubated at 37°C for one hour. Labeling reactions were stopped by the addition of 5 μL of 0.2M EDTA. Unincorporated nucleotides were removed using ProbeQuantTM G-50 Micro Columns (Amersham Pharmacia Biotech, UK). Probes were incubated on ice prior to hybridisation.

2.3.4 Probe Preparation for the DNA Orientation Marker

The DNA Orientation Marker was labeled using the *Rediprime* DNA labeling system (Amersham Pharmacia Biotech, UK). For each filter 45 μL (25ng) of the DNA Orientation Marker were placed in an RNase-free tube and denatured by heating at 100°C for 5 minutes. The denatured DNA was added to the pre-aliquoted *Rediprime* mix and reconstituted by gentle flicking of the tube. The contents were brought to the bottom of the tube by brief centrifugation. [α - ^{33}P] dCTP was added (5 μL) and mixed by gently pipetting up and down five times. This mixture was incubated at 37°C for 10 minutes and the reaction was terminated by the addition of 5 μL of 0.25M EDTA and by heating to 95°C for 5 minutes. The contents were brought to the bottom of the tube by brief centrifugation. The unincorporated nucleotides and [α - ^{33}P] dCTP were removed using G-50 Micro columns as described previously. The labeled probe was incubated on ice prior to hybridisation.

2.3.5 Hybridisation

The pre-hybridisation solution was drained from the hybridisation bottles. Denatured Salmon sperm DNA (75 μL of 10mg/mL) was added to a new aliquot of 20mL of NothernMax hybridisation buffer. The probe/control mixture and labeled orientation marker were added to this buffer and mixed by inversion. This buffer/probe combination was added to the hybridisation bottle. Hybridisation was carried out at 42°C overnight.

2.3.6 Post-hybridisation

The hybridisation buffer was removed from the containers and the filters were rinsed with 200mL of 2XSSC at room temperature for 5 minutes. The filters were washed for 30 minutes at 68°C with 200mL of pre-warmed Wash 1 (2XSSC, 1%SDS). This step

was repeated with a new aliquot of Wash 1. The filters were washed for 30 minutes at 68°C with 200mL of pre-warmed Wash 2 (0.6XSSC,1%SDS).). This step was repeated with a new aliquot of Wash 2. The GDA filters were removed from the bottles and wrapped very carefully in plastic wrap and mounted on exposed sections of film. Filters were stored in autoradiography cassette boxes at -20°C.

2.3.7 Imaging of GDA Filters

The filters were imaged using the STORM 860 phosphorimager. The filters were placed carefully in the phosphorimager cassette (colony side facing up) and the phosphor-sensitive screen placed carefully on top of the filter. Exposure was carried out for three hours at a resolution of 100microns. Image files were generated in PC '.gel' format and uploaded to the Genome Systems GDA server using File Transfer Protocol (FTP). Genome Systems carried out the image analysis using the Gene Discovery Array (GDA) software and the results were uploaded onto a secure GDA server. Data were downloaded from the GDA server and used to extract differential gene expression patterns. GDA clones were selected for further analysis using the GeneMatch v1.0 perl script program (Section 4.2.4).

2.3.8 GDA Library Clones

Clones were obtained from the Genome Systems Inc. Gene Discovery Array Clone Library. PCR probes were prepared from the GDA clones for Northern blot analyses (Section 2.3.9) largely as described for DDRT-PCR analyses with slight modifications. In brief, clones were received as bacterial stabs in agar. Colonies were plated on LB agar and grown overnight at 37°C. Individual colonies were selected and regrown overnight at 37°C. Suitable colonies were selected and grown in LB broth overnight at 225rpm at 37°C. Plasmid DNA was purified from bacterial growth using the protocol outlined in Section 2.2.4. Inserts were excised from plasmid DNA using suitable restriction enzymes and were gel purified (Section 2.1.2). DNA probes were labelled using the RediprimeTM labeling kit (Amersham Pharmacia Biotech, UK) with [α -³²P] dCTP and were purified using ProbeQuantTM G-50 Micro Columns (Amersham Pharmacia Biotech, UK) as described in Section 2.3.9. The clones selected from the GDA Library are detailed in Table 2.3.

2.3.9 Northern Blotting

Formaldehyde agarose gels were prepared as follows: 1.5g agarose was dissolved in a solution containing 15mL 10X MOPS buffer, 109.5mL DEPC-treated water and 26.5mL formaldehyde (37%). Total RNA samples were prepared for electrophoresis as follows: 20µg (6µL) of each RNA sample was incubated in a solution containing 12.5µL deionised formamide, 2.5µL 10X MOPS buffer and 4µL formaldehyde (37%) and incubated at 65°C for 5 minutes. Each sample was chilled on ice followed by the addition of 2.5µL 50% (v/v) glycerol, containing 0.25% bromophenol blue (w/v). Samples were resolved for 5-6 hours using 1X MOPS buffer as running buffer. RNA was transferred to HybondTM- N+ nylon (Amersham Pharmacia Biotech, UK) by capillary blotting (Sambrook *et al.*, 1989) using 20XSSC as the transfer buffer. RNA was fixed to the nylon by carefully placing the nylon blot on Whatman paper soaked in 0.05M NaOH. Blots were stored in plastic wrap at 4°C.

Northern blot probes were generated by RT-PCR using primers specific for the transcript of interest or by the isolation of insert DNA from GDA clones(Section 2.3.8). The primers designed for RT-PCR analyses are detailed in Table 2.2. RT-PCR products were gel-purified and quantitated by spectrophotometry. The RediprimeTM labeling kit was used to label 25ng of RT-PCR products with [α -³²P] dCTP. Probes were purified using ProbeQuantTM G-50 Micro Columns. Prehybridisation was carried out by placing nylon blots in hybridisation bottles in the presence of hybridisation buffer (1M Sodium phosphate pH 6.8) at 65°C for 2-3 hours. Hybridisation was initiated by adding radiolabeled RT-PCR product directly to the hybridisation bottles at 65°C overnight. Nylon blots were washed twice using prewarmed Wash buffer A (2.5g BSA, 25g SDS, 20mL 1M sodium phosphate, 1mL 0.5 M EDTA in 500mL DEPC-treated water) at 65°C for 10 minutes. This was followed by multiple (2-4) washes at 65°C for 10 minutes with Wash buffer B (10g SDS, 40mL 1M sodium phosphate buffer, 2mL 0.5M EDTA in 1000mL DEPC-treated water). Nylon blots were mounted on exposed x-ray film and covered in plastic wrap. Nylon blots were placed in autoradiography cassettes with x-ray film for 4-48 hours.

2.3.10 Screening for repeat sequence instability

Two candidate genes were selected for simple sequence repeat instability in the panel of sporadic colorectal cancer cell lines by two different approaches. Primers specific for Heat Shock Protein 86 (Hsp86) were received as a kind gift from Dr. Peter McWilliam. RT-PCR was carried out using these primers on the panel of MIN and CIN cell lines with the incorporation of [α - ^{32}P] dCTP in the PCR step. See Table 2.3 for the amplification conditions. PCR products were separated on a non-denaturing 6% sequencing polyacrylamide gel.

Primers specific for the human protein tyrosine phosphatase 2A (PTP2A) transcript and the human transforming growth factor β receptor II (RII) transcript were designed using the Oligo 4.0 software package (National Biosciences, Inc., USA). A Cy5 label was including at the 5' end of the sense primers. RT-PCR was carried out on the panel of MIN and CIN cell lines. Products were separated using the ALFexpress™ automated sequencer on 15% denaturing polyacrylamide gels for 500 minutes.

2.4 Protein Analysis

2.4.1 Protein Isolation / Subcellular Fractionation

Cell lines were grown to 70-75% confluence and treated with Trypsin-EDTA to lift the cells. The cells were centrifuged at 1,000 rpm for 3 minutes. The supernatant was discarded. The cells were resuspended in 1ml of ice-cold buffer A and sonicated (Branson Sonifier 250) in an ice water bath for 1 second pulses ten times at 100W. Sonication was deemed successful when the solution appeared partially opaque with a blue shade. The sonicated mixture was centrifuged in a benchtop centrifuge (Sorvall® RT6000) at 1,500 rpm for 5 minutes at 4°C to pellet unlysed cells, nuclear and cellular debris. The supernatant was transferred to 6mL Quick Seal® tubes, overlaid with paraffin oil, and centrifuged at 28,000 rpm in an ultracentrifuge (Beckmann L7-55) for 10 minutes at 4°C. The paraffin oil was removed and the supernatant designated the cytosolic cellular fraction. The protein pellet was washed in Buffer A and resuspended in 500µL of ice-cold buffer B and incubated at 4°C in ice water for 30 minutes with occasional inversions. The suspension was centrifuged at 15,000 rpm for 30 minutes at 4°C. The supernatant was termed the membrane protein fraction. The final pellet was designated the cytoskeletal fraction and was resuspended in 250µL Buffer C by a combination of boiling and homogenisation.

A simplified protocol was also used to obtain cytoskeletal fractions and the remaining total protein fractions less the cytoskeletal component. This was carried out to maximise the yield of the cytoskeletal component. Cells were resuspended in 1mL ice-cold buffer B and lysed by incubated in ice water for 30 minutes. This sample was centrifuged, as described above in an ultracentrifuge, at 28,000 rpm for 30 minutes at 4°C. The supernatant was designated the total protein fraction less the cytoskeletal component. The pellet was washed in Buffer C and resuspended in Buffer C by a combination of boiling and homogenisation. This was labeled the cytoskeletal fraction.

2.4.2 Protein Estimation

Protein estimation was accomplished by two methods. The Lowry assay was carried out largely as described (Lowry *et al.*, 1951). Briefly, a standard dilution curve of BSA was made in the range 0mg/mL to 10mg/mL BSA. The Lowry reagent and 4% copper sulphate were mixed in a ratio of 100:1. Aliquots (10 μ L) of the standards and 5-10 μ L of unknown protein (in a final volume of 10 μ L with distilled water) were added to 1mL volumes of the copper sulphate : Lowry reagent in 1.5mL plastic cuvettes. All samples were thoroughly vortexed and incubated for 10 minutes at 37°C. Folin's reagent (100 μ L) was added to each cuvette, mixed vigorously followed by incubation for 15 minutes at 37°C. All samples were assayed in triplicate and absorbances determined spectrophotometrically at 660nm. The absorbance value for 0mg/mL BSA was subtracted from the absorbances of the standards and from the unknown sample absorbances. Standard curves were plotted between absorbance at 660nm versus concentration of BSA standard. Absorbance values for unknown protein were used to approximate the concentration of these samples using the standard curve. Fractions containing membrane protein (membrane and total minus cytoskeletal fractions) were observed to precipitate after addition of Folin's reagent. The detergent Triton X-100 (10 μ L) was added to these samples to solubilise the precipitate and the standard curve repeated including an equivalent amount of Triton X-100. Protein concentrations for these samples were determined as described above.

The Bradford assay (Bradford *et al.*, 1976) was used to determine unknown protein concentrations. Bradford reagent was prepared by making the following additions: 100mg Coomassie brilliant blue G250, 50mL 95% ethanol, 100mL 85% orthophosphate (H₃PO₄) to a final volume of 1 litre with distilled water. The Bradford reagent as filtered through a No.1 Whatman filter. A standard curve of BSA was prepared in the range 0mg/mL to 1.0mg/mL BSA. Aliquots of standards (100 μ L) and 10-20 μ L of unknown protein (in a final volume of 100 μ L with 1X PBS buffer) were added to 5mL of Bradford reagent in plastic tubes. These samples were mixed thoroughly and 1mL aliquots were analysed spectrophotometrically at 595nm. All samples were analysed in triplicate and absorbance values of the standards corrected for the blank value corresponding to 0mg/mL BSA. A standard curve was plotted between absorbance versus BSA concentration. This plot was used to determine the

concentration of the unknown protein samples. The two protocols were used to minimise inaccuracies in either assay. Both assays yielded reproducible and comparable results.

2.4.3 Protein Precipitations

Dilute protein samples were concentrated for Western blot analysis by acetone precipitation. In brief, 4 volumes of 100% acetone were added to 1 volume of protein fraction. This was vortexed gently and briefly centrifuged at 4°C. The precipitation mixtures were incubated at -20°C for 1 hour with intermittent inversion to prevent the protein from freezing. The mixtures were centrifuged at 15,000 rpm for 5 minutes at 4°C. The supernatants were removed and the protein gently resuspended in 20µL 1X protein loading dye.

2.4.4 Protein Electrophoresis

Polyacrylamide gels were prepared for protein electrophoresis as described below. The resolving section of the gel was prepared by making the following additions in the order shown: 6.66mL acrylamide:bis-acrylamide (37.5:1) (30%(v/v)), 8.04mL distilled water, 5mL Tris-HCl pH8.8 buffer, 200µL 10% SDS, 100 µL 10% APS, 10µL TEMED. This mixture was poured into the 70% ethanol washed gel plates. The top level of the gel was overlayed with iso-butanol to prevent evaporation and to eliminate air bubbles. The resolving gel was allowed to set for 45 minutes. The stacking gel was formulated by mixing the following components: 0.65mL acrylamide:bis-acrylamide (37.5:1) (30%(v/v)), 3.05mL distilled water, 1.25mL Tris-HCl pH6.8 buffer, 50 µL 10% SDS, 25 µL 10% APS, 5µL TEMED. The isobutanol was washed out of the plates using distilled water. The stacking gel was poured on top of the protein gel, the comb placed between the plates and allowed to set for 1 hour. Protein samples were prepared in 1X protein loading dye and boiled in a water bath for 5 minutes. Samples were stored in ice water prior to loading. Proteins were electrophoresed at 25mA per gel (200V) until the dye front had migrated to the end of the gel. Duplicate gels were stained in Coomassie stain overnight on a shaking platform and destained repeatedly in protein gel destain solution.

2.4.5 Protein Blotting

Gels were retrieved from the electrophoresis plates and the stacking gel discarded. Polyvinylidene difluoride (PVDF) membrane and filter paper (Grade 1F, Munkell, Sweden) were cut to the same size as the gel dimensions. Sheets of filter paper were soaked in transfer buffer. The membrane was soaked in methanol for 30 seconds followed by brief immersion in transfer buffer. The transfer sandwich was prepared as follows (from bottom to top): 10 sheets of filter paper, the PVDF membrane, the polyacrylamide protein gel and 10 sheets of filter paper. A small volume of transfer buffer was poured on top of the sandwich and the area surrounding the gel sandwich was dried carefully. Protein was transferred to the membrane using the HorizoBlot (ATTO, Japan) semi-dry apparatus. Transfer was carried out at 100mA for 1 hour at room temperature. The transferred gel was retrieved from the transfer sandwich and stained in Coomassie to assess the success of protein transfer. Duplicate sections of transferred membrane were stained in Ponceau S and destained in distilled H₂O to visualise protein transfer onto the membrane.

2.4.6 Probing of Western blots

The polyvinylidene difluoride (PVDF) membrane was washed briefly in 1X PBS to remove any traces of transfer buffer and then blocked overnight at 4°C in blocking buffer. The blot was washed for 5 minutes in 1X PBS (0.1% (v/v) Tween-20) 3 times. The primary antibody PP1 β was diluted 1:400 with blocking buffer and incubated with the membrane for 1 hour at room temperature with gentle agitation. The antibody mixture was retained and the blot washed 3 times as before. The Phototope®-HRP Western blot detection kit was used to probe western blots. The membrane was incubated with the HRP (horseradish peroxidase)-linked anti-goat IgG secondary antibody at a dilution of 1:1,500 and the anti-biotin HRP-linked antibody (1:1,000) for the molecular weight marker for 1 hour at room temperature. The blot was washed 3 times as before. Protein bands were detected using electrochemiluminescence (ECL) reaction using the LumiGlo® 20X ECL kit (New England Biolabs, UK). The substrate reaction was prepared by mixing by 0.5mL of 20X LumiGlo reagent A and 0.5mL 20X peroxide reagent B to a final volume of 10mL in distilled water. This mixture was poured onto the protein side of the PVDF membrane lying on a bed of plastic wrap. The

reaction was allowed to proceed for 1 minute. The excess reaction solution was drained off and the membrane was wrapped in plastic wrap and mounted in an autoradiography cassette. The blot was exposed to Kodak x-ray film for 15-60 seconds and the film was developed as described for DDRT-PCR.

2.4.7 Immunohistochemistry

Sectioning of paraffin-embedded human colon tumours was carried out by Tony O'Grady, Department of Pathology, RCSI. Preliminary experiments to optimise the polyclonal antibody to human Protein Phosphatase 1 β (PP1 β) were also carried out by Tony O'Grady, Department of Pathology, RCSI.

The Dako Envision™+ system HRP (DAB) anti-mouse antibody was used following the manufacturer's instructions. An antigen retrieval step was carried out on the sections for the PP1 β antibody. Slides were placed in 0.01M tri-Sodium citrate (pH6.0) and microwaved at 850W for 20 minutes. Slides were incubated in goat serum diluted (1:10) with 0.05M Tris buffer (pH7.6) for approximately 10 minutes. PP1 β was diluted in a concentration gradient from 1:10 to 1:500 with 0.05M Tris buffer (pH7.6) and placed on slides for 30 minutes at room temperature. Slides were washed 2X5 minutes in TBS. The secondary antibody was incubated on the slides for a further 30 minutes followed by 2X5 minute washes in TBS. Diaminobenzadine (DAB) was diluted 1:10 with 0.05M Tris buffer (pH7.6) adding 5 μ L 30% hydrogen peroxide per 10mL. DAB was added to the slides for 5 minutes followed by rinsing with distilled H₂O. Haemotxylin was used to counterstain the slides for 15 seconds followed by washing in tap water for 3 minutes. The slides were dehydrated by 2X10 dips of alcohol and 3X10 dips of xylene. The slides were finally coverslipped using DPX mountant.

Cytospins were also analysed by immunohistochemistry. Cell lines (SW480, SW620, SW1116 & 498Li, Table 1.2) were grown to 50-70% confluence and isolated by treatment with Trypsin-EDTA and centrifugation at 1,000rpm for 3 minutes. The supernatant was decanted with a sufficient amount left in the container to produce a monolayer of cells. The pellets were resuspended and 1-3 drops of fluid were added to each cytofunnel and cytospin. Cytospin assemblies were centrifuged at 1000rpm for 5 minutes. Thick cytospin preparations were gently 'spider' cell dotted with an applicator stick. Cytospins were fixed immediately with cyto-fix and allowed to dry for 5 minutes.

Preparations were post-fixed in IMS 99 (spirit) for 20 minutes in a fumehood and washed in tap water to remove carbowax for 5 minutes. Cytospins were analysed immunohistochemically as for the colon tumour paraffin-embedded material.

Table 2.1 Cell lines, media formulations and growth conditions

Cell Line	Medium Formulation
LS 174T	Eagle's MEM with non-essential amino acids and Earle's BSS, 90%; fetal bovine serum, 9%, penicillin streptomycin, 1%. Growth at 37°C, 5% CO ₂
SW48	Leibovitz's L-15 medium, 90%; fetal bovine serum, 9%; penicillin streptomycin, 1%. Growth at 37°C.
HCT116	McCoy's 5a medium, 90%; fetal bovine serum, 10%; penicillin streptomycin, 1%. Growth at 37°C, 5% CO ₂ .
Caco2	Eagle's MEM with non-essential amino acids and Earle's BSS, 85%; fetal bovine serum, 15%, penicillin streptomycin, 1%. Growth at 37°C, 5% CO ₂
SW480	Leibovitz's L-15 medium, 90%; fetal bovine serum, 10%; penicillin streptomycin, 1%. Growth at 37°C.
HT29	McCoy's 5a medium, 90%; fetal bovine serum, 10%; penicillin streptomycin, 1%. Growth at 37°C, 5% CO ₂ .
SW620	Leibovitz's L-15 medium, 90%; fetal bovine serum, 10%; penicillin streptomycin, 1%. Growth at 37°C.
SW1116	Leibovitz's L-15 medium, 90%; fetal bovine serum, 10%; penicillin streptomycin, 1%. Growth at 37°C.
498Li	Leibovitz's L-15 medium, 90%; fetal bovine serum, 10%; penicillin streptomycin, 1%. Growth at 37°C.

Table 2.2 Oligonucleotide primer sequences

Primer Name	Sequence
β -Actin U *	5' -Cy5-ACAGGGGAGGTGATAGCATTG-3'
β -Actin L	5' -CGAAGGCTCATCATTCAAAAT-3'
AP3L	5' -GGGGATCCTTTTTTTTTTTTTT{A/G/C}T-3'
DD1A	5' -CGACGGATCCTATGATTTAA-3'
DD2A	5' -AGGCGGATCCAATACTATTA-3'
DD3A	5' -CCAGGGATCCTAACAAATAA-3'
DD4A	5' -CGGAGGATCCATTAAGAATA-3'
DD5A	5' -CAGCGGATCCTAATGATTAA-3'
DD6A	5' -AGCCGGATCCATATTCTATA-3'
SPLA 2 U *	5' -Cy5-CCGCACTCAGTTATGGCTTTCT-3'
SPLA 2 L	5' -ACATCCACGTTTCTCCAGACG-3'
COX 2 U *	5' -Cy5-AGCCATTTTGCTAAGAAGACA-3'
COX 2 L	5' -TGCAAAAGTTCAGGTAAGAA-3'
Cytokeratin 8 U	5' -CGGGCTGAGGCTGAGAGCAT-3'
Cytokeratin 8 L	5' -GGCCAGCTTGACGTTTCATCA-3'
Cytokeratin 18 U	5' -GAGGCTGGAGAGCAAAATCC-3'
Cytokeratin 18 L	5' -CGGGCATTGTCCACAGTATT-3'
Forward sequencing primer	5' -GTAAAACGACGGCCAGT-3'
Reverse sequencing primer	5' -GGAAACAGCTATGACCATG-3'
Forward sequencing primer *	5' -Cy5-GTAAAACGACGGCCAGT-3'
PKD1 U	5' -CGCCGCTTCACTAGCTTCGAC-3'
PKD1 L	5' -CCAAGGTGACCCCCAGGAG-3'
RANBP2 U	5' -AATGGCAAACATAAGAATG-3'
RANBP L	5' -ACATCCATGCTCTATCTG-3'
PGER2 U	5' -TCCCGGGGTCAATTTCGTC-3'
PGER2 L	5' -GCGCGAGTTGCACAGCAC-3'

Table 2.2 Oligonucleotide primer sequences (continued)

CCNA2 U	5' -ACGGCGCTCCAAGAGGAC-3'
CCNA2 L	5' -CTTCGGCCTCTGCTGCTG-3'
PTPD1 U *	5' -Cy5-TCTGGGAAGAAGAACATC-3'
PTPD1 L	5' -AGGGAGAGTCCATTTAGG-3'
PLK U	5' -CCCGGAGGTCTAGTGGA-3'
PLK L	5' -CGGCTTGAGCAGCAGAGA-3'
TGFBRII U *	5' -Cy5-CTTTATTCTGGAAGATGCTGC-3'
TGFBRII L	5' -GAAGAAAGTCTCACCAGGC-3'
DDP U	5' -AGACATCTTAGCCAGGAAA-3'
DDP L	5' -TTCTTCAGGTGAGAGGTCA-3'
DDT U	5' -GTAATTGATAGGGTATGAACT-3'
DDT L	5' -AGACAAGGTAACACCAAATAG-3'
PP1 α U	5' -TAAGGCCACCTGTCACG-3'
PP1 α L	5' -CCCCAGGATCCTGCTCAC-3'
PP1 β U	5' -AAAGATGTGCTACACTGT-3'
PP1 β L	5' -TAAAACCACATTACACAG-3'
PP1 γ U	5' -CCAAAAAGAAAAGTGTTT-3'
PP1 γ L	5' -TTATGGGTGTACTGAAT-3'
GAP U	5' -ACTTTACCACAAATACAGTGT-3'
GAP L	5' -AATTAGCAATCTGTAAATCAG-3'

The primers marked * were Cy5-labelled at the 5' end including:

β -actin U

SPLA2 U

COX2 U

Forward sequencing primer

PTPD1 U

TGFBRII U

Table 2.3 PCR conditions

Gene	Method	Study	T _A °C	Size	Mg ²⁺ mM
β-actin	RTPCR	Control	50	109	1.5
GAPDH	RTPCR	Control	55	598	1.5
M1	RTPCR/ N. blot	DDRT- PCR	50	160	2.5
PTPD1	RTPCR	GDA/MIN	49	116	2.5
PLK	RTPCR	GDA	58	157	1.5
RANBP	RTPCR	GDA	47	125	1.5
PGER2	RTPCR	GDA	58	120	1.5
CCNA2	RTPCR	GDA	58	120	1.5
Cytokeratin 8	RTPCR	CG	55	326	1.5
Cytokeratin 18	RTPCR	CG	60	129	1.5
Cox2	RTPCR	CG			
sPLA2	RTPCR	CG			
GAP	N. Blot	CG	50	874	1.5
TGFβRII	RTPCR	MIN	55	73	1.5
HSP86	RTPCR	MIN	50	134	1.25
DDP	RTPCR	MET	55	109	1.5
DDT	RTPCR	MET	55	168	1.5
PP1α	RTPCR	MET	55	137	1.5
PP1β	RTPCR	MET	45	195	1.5
PP1γ	RTPCR	MET	45	149	1.5

Abbreviations:

CG : Candidate gene

DDRT-PCR : confirmation of Differential Display Reverse Transcription PCR data

GDA : Confirmation of Gene Discovery Array data

MET : Investigation of gene in metastatic process

MIN : Investigation of nucleotide instability at simple sequence repeat

Mg²⁺ mM : Optimal magnesium concentration in PCR amplification

N. blot : Northern blot

RTPCR : Reverse Transcription PCR

Size : PCR product size (bp)

T_A °C : Annealing temperature used in PCR amplification

CHAPTER 3

DIFFERENTIAL DISPLAY REVERSE

TRANSCRIPTASE PCR

3.0 INTRODUCTION

Differential Display Reverse Transcriptase PCR (DDRT-PCR) was developed in 1992 (Liang *et al.*, 1992) and subsequently modified by other investigators (Bauer *et al.*, 1993). The DDRT method involves multiple steps: isolation of undegraded RNA; reverse transcription of the RNA to cDNA primed by an anchored oligo-dT primer; amplification of the cDNA via PCR and display of the PCR products by electrophoresis (see Figure 3.1).

Efficient Differential Display Reverse Transcription PCR (DDRT-PCR) experiments rely on optimising each component of the reaction and maintaining a useful number of DDRT-PCR bands. The concentration of enzyme, magnesium and primers has the greatest effect on display band patterns (unpublished observations). The enzyme concentration was optimised first, followed by the magnesium concentration and finally the primer concentrations. When assaying a number of cDNA samples using DDRT-PCR a balance was required between obtaining a display pattern with a useful number of bands (40-60) and sufficient reproducibility so that small variations in conditions did not introduce spurious variation. In the optimisation of enzyme, magnesium and primer concentrations one cDNA sample was used. The concentration of each reaction component that gave a suitable band pattern was chosen and assayed using multiple cDNA samples to examine the stability of the experimental design. Varying the components of the DDRT-PCR yields appreciably different display patterns (unpublished observations).

DDRT-PCR experiments from the present study yielded five sequences. The most interesting of these clones proved to be clone M6 which was overexpressed in the MIN sporadic colorectal cancer cell lines. This clone shows a weak similarity to the human regulator of mitotic spindle assembly 1 protein (hRMSA1) – a gene which has been implicated in cell cycle checkpoint control. The remaining four sequences may represent novel sequences. The enormous amount of data obtained from GDA analysis (Chapters 4 & 5) allowed us to discontinue further analysis of the DDRT-PCR sequences.

To illustrate the scope of downstream investigations possible post DDRT-PCR a number of experiments have been carried out on genes associated with Inhibitor 2 of protein phosphatase 1 (IPP2). This gene is overexpressed in the pre-metastatic colorectal cancer cell line SW480 with respect to its isogenic metastatic counterpart

SW620 (McWilliam *et al.*, in preparation). This observation was originally made using DDRT-PCR. Semi-quantitative RT-PCR experiments were carried out on the above pair of cell lines and an additional pair of cell lines (SW1116 and 498Li) directed towards genes associated with IPP2. In addition, western blot analysis was carried out on the Protein Phosphatase I Catalytic subunit Beta isoform (PPP1C-Beta). The results of these investigations are outlined in Chapter 6.

3.1 METHODS

A panel of sporadic colorectal cancer cell lines (Table 2.1) was cultured as recommended by the American Type Culture Collection (ATCC). Total RNA was extracted from the cell lines using the UltraspecTM II RNA Isolation System (Biotech Labs Inc., USA). RNA samples were quantitated by spectrophotometry, the hybridisation assay, semi-quantitative RT-PCR using the 'housekeeping' gene β -actin and visualised on denaturing agarose gels. cDNA was synthesised from total RNA using the SuperscriptTM II Moloney Murine Leukaemia Virus Reverse Transcriptase kit (Gibco-BRL, UK). Contaminating genomic DNA was screened for using an intron –spanning PCR assay with specific primers for the Polycystic Kidney Disease 1 (PKD1) gene. The primers were used to amplify control human genomic DNA and cDNA synthesised from cell line total RNA in parallel. The genomic DNA product is 256bp and the RT-PCR product is 183bp.

Extensive optimisation of the Differential Display Reverse Transcriptase PCR (DDRT-PCR) was carried out (Chapter 2, Section 2.2.1). DDRT-PCR was carried out on normalised total RNA from each of the sporadic colorectal cancer cell lines (Table 2.1) with the exception of HT-29. cDNA for DDRT-PCR was synthesised from total RNA and was primed with the anchor primer AP3L. cDNA was amplified by PCR under the modified conditions described in Chapter 2, Section 2.2.1. DDRT-PCR products were electrophoresed on 6% non-denaturing polyacrylamide sequencing gels which were dried and autoradiography was carried out. Differentially expressed products were excised, reamplified by PCR, cloned and sequenced (Chapter 2, Sections 2.2.3 & 2.2.4). DDRT-PCR sequences were characterised using the BLAST server [<http://www.ncbi.nlm.nih.gov/BLAST>] where possible.

3.2 RESULTS

3.2.1 RNA Integrity and Quantitation

Total RNA was visualised by electrophoresis on 1% denaturing formaldehyde agarose gels (Section 2.1.4). The majority of RNA samples were found to be undegraded with strong 28S and 18S ribosomal RNA bands (Figure 3.2A). A faint smear was seen migrate between the ribosomal RNA bands indicating the presence of mRNA. A number of RNA samples were seen to contain trace levels of high molecular weight genomic DNA adjacent to the origin of electrophoresis. Such samples were DNase I treated and separated by electrophoresis (Figure 3.2B). This treatment resulted in the degradation of contaminant DNA leaving intact RNA.

RNA samples were quantitated by spectrophotometry. Absorbances were determined at wavelengths of 260nm and 280nm. The A₂₆₀/A₂₈₀ ratios were calculated. The ratio values obtained were seen to fall in the accepted range of 1.8-2.1 (Sambrook *et al.*, 1989). RNA samples with ratios outside the recommended range may have contained contaminant phenol or protein and were deemed unsuitable for further analysis.

3.2.2 Hybridisation Assay to quantitate mRNA content

To obtain useful gene expression data, the absolute levels of mRNA were determined in each RNA sample using a hybridisation assay (Harley *et al.*, 1987; Section 2.1.5). A standard curve was plotted of the counts per min. (cpm) for each standard mRNA versus mRNA quantity of each corresponding mRNA standard. A linear plot was obtained and was used to determine the absolute mRNA levels of the MIN and CIN RNA samples (Figure 3.3). An image of a hybridisation assay autoradiograph is shown in Figure 3.4.

3.2.3 Semi-Quantitative Reverse Transcription PCR

The technique of semi-quantitative RT-PCR was employed to normalise the MIN and CIN RNA samples. β -actin was selected as a typical constitutively expressed gene between cell types. cDNA was synthesised from RNA samples as described in Section 2.1.6. Input RNA quantity in cDNA syntheses was determined by spectrophotometry and hybridisation assay. PCR was performed on the cDNA samples using primers

specific for the β -actin gene. Initially, experiments revealed slight variations in β -actin RT-PCR product between the MIN and CIN samples. Input RNA was varied accordingly and RT-PCR repeated. An example of a typical semi-quantitative RT-PCR experiment is illustrated in Figure 3.5.

3.2.4 OPTIMISATION OF DIFFERENTIAL DISPLAY REVERSE TRANSCRIPTION-PCR CONDITIONS

Different cycling conditions were tested to yield the most reproducible DDRT-PCR patterns. A low annealing temperature of 40°C was used in the initial DDRT-PCR experiments. The display patterns obtained were unreproducible with a high degree of background 'noise'. A higher annealing temperature of 60°C was subsequently adopted. This was seen to yield less background and greater reproducibility in display patterns (data not shown).

Three types of DNA Polymerase were used in DDRT-PCR reactions. Stoffel fragment yielded very few bands with no discernable degree of reproducibility. *Pfu* DNA Polymerase was also used. This high fidelity DNA Polymerase enzyme has a 5' exonuclease activity and was expected to improve the priming from cDNA species in DDRT-PCR. Display patterns obtained with *Pfu* DNA Polymerase yielded greater numbers of bands without a useful level of reproducibility. *Taq* DNA Polymerase was found to give the greatest reproducibility over repeat experiments with adequate number of bands in display gels (Figure 3.6).

The optimal magnesium concentration of 1.9mM was determined as described in Section 2.2.1. Lower magnesium concentrations yielded faint display patterns. Higher magnesium concentrations yielded complex display patterns with high background noise levels. Magnesium at a final concentration of 1.9mM was used for all DDRT-PCR reactions, yielding optimal display patterns. Consistent display patterns are clearly illustrated in Figure 3.6 at the optimised magnesium concentration.

High percentages of false positives have been associated with DDRT-PCR. An experimental design was adopted to minimise the selection of false positive DDRT-PCR products. Reactions were carried out over a range of arbitrary primer concentrations against a fixed anchor primer concentration. cDNA species seen to be consistently differentially expressed over the range of arbitrary primary concentrations were selected

for reamplification. cDNA species of inconsistent differential expression were not selected for further analysis.

3.2.5 Reamplification of DDRT-PCR products

DDRT-PCR products were excised from display gels and reamplified as described in Section 2.2.2. Fragments of 700bp and less were readily reamplified. The majority of reamplified products contained contaminant non-specific PCR products. Such fragments were gel purified as described in Section 2.2.3. The reamplification of a number of DDRT-PCR clones is illustrated in Figure 3.7. The DDRT-PCR products excised and reamplified, their expression patterns and sizes are outlined in Table 3.1.

3.2.6 Cloning and Sequencing of DDRT-PCR products

Blunt-ended cloning was used to clone DDRT-PCR products as described in Materials and Methods, Section 2.2.2. The limiting factor in the cloning efficiency was determined to be the concentration of PCR-product insert used in the ligation step. Very few white colonies were obtained using 100ng of PCR-product insert. Using 0.5-1 μ g of PCR insert increased the proportion of white colonies. Plasmid DNA was isolated from white colonies and screened for the presence of insert as described in Section 2.2.2. insert sizes obtained matched the reamplified cDNA sizes. The percentage of insert-positive plasmids varied between the DDRT-PCR products analysed. This may have been due to partial deletion of the β -galactosidase gene in the plasmid by the *E. coli* cells used for transformation, or to unevenly distributed X-Gal on the LB culture plates. A PCR screen of DDRT-PCR clones is demonstrated in Figure 3.8. A number of clones were sequenced from each of five DDRT-PCR products. A product yielding at least four identical sequences was deemed to be representative of the original DDRT-PCR band. Additional clones were sequenced (four) which did not yield a significant consensus sequence which could not be analysed further. The DDRT-PCR clone sequences are given in Figures 3.9-3.13. BLAST homology search results for each DDRT-PCR cDNA are shown in Figures 3.9B – 3.13B. Clones M1, M2, M5 showed no significant homologies to known sequences in GenBank. The BLASTX algorithm was used to determine homologies between DDRT-PCR sequences and known protein sequences. The BLASTX algorithm compares a nucleotide query sequence translated in

all reading frames against a protein sequence database (Gish *et al.*, 1993). A weak similarity between M6 and the human regulator of mitotic spindle assembly 1 protein (RMSA1) was observed which is likely to be due to an Alu element in the clone. In addition, clone AV showed significant homology to a novel protein containing ankyrin repeat motifs.

3.3 DISCUSSION

Various optimisations and modifications of the DDRT protocol were carried out. Initially, a number of DNA Polymerase enzymes were assayed for their performance in DDRT-PCR. Stoffel fragment of *Taq* DNA Polymerase is reported to maintain its activity for PCR amplification over a considerable magnesium concentration range [0.8-2.5mM] (Larrick *et al.*, 1995). This property makes troublesome PCR optimisation more convenient with a larger magnesium range to operate with. *Pfu* DNA Polymerase is a high fidelity polymerase with a 5' exonuclease activity that repairs DNA mismatches (Larrick *et al.*, 1995). *Taq* DNA Polymerase supplied by Promega Corporation (USA) yielded the most useful patterns in DDRT-PCR experiments. The concentrations of the components of the PCR reaction were optimised to yield a useful number (approximately 50) of bands reproducibly. The most sensitive component of the PCR with respect to band pattern proved to be the concentration of the upstream arbitrary primer. The sequences of the arbitrary primers used were randomly designed and of different base composition. Primers of this design have different annealing properties on cDNA templates. This may explain the unique performance of each arbitrary primer assayed in DDRT-PCR experiments. DDRT-PCR experiments were performed over a range of three arbitrary primer concentrations to control for potential PCR artefacts. Display bands were selected for further analysis if consistent differential expression was seen over the primer concentration gradient used.

Overall examination of the display patterns obtained between the different cell line RNA samples showed a reproducible fingerprint pattern of bands appearing in all RNA samples assayed. The fingerprint pattern typically took the form of 10-15 display bands appearing in parallel across all the cell lines assayed. The remaining display pattern showed a high degree of variability, i.e. bands were distributed variably across the cell lines assayed. This is likely to be due to expression differences between the MIN and CIN genotypes. It may also be due to tissue specific expression of mRNA transcripts between the sporadic colorectal cancer cell lines selected. Each cell line was established from a particular tumour from different patients (Table 2.1). A certain number of PCR artefacts may have also contributed to the pattern heterogeneity. A number of bands (seventeen; Table 3.1) were excised from gels and reamplified. Some cDNA products were successfully reamplified under the initial conditions, others

yielded high background and non-specific products. A proportion (30%) yielded no product after reamplification, despite further optimisation of the PCR conditions.

Products were gel purified where necessary, cloned and sequenced resulting in five DDRT-PCR product sequences. The identity of the clones was determined using the Basic Local Alignment of Sequences Tool (BLAST; <http://www.ncbi.nlm.nih.gov/BLAST>) algorithm which identifies sequences in GenBank homologous to the query, in this case a DDRT-PCR sequence. Of the clones sequenced clone M6 proved to be the most interesting. Initially, clone M6 yielded no appreciable homologies to any known nucleotide sequences. The BlastX algorithm was used to identify any homology between clone M6 and known protein sequences. This approach led to the identification of a weak similarity between clone M6 and the human regulator of mitotic spindle assembly 1 protein (RMSA1). However, BLAST analysis also revealed weak similarity to approximately 1,500 additional sequences diagnostic of the presence of an Alu repeat.

The cell line SW620 was included in DDRT-PCR experiments. SW620 is the isogenic metastatic counterpart of the primary colorectal cancer cell line SW480. Clone AV was isolated from a DDRT-PCR experiment and was overexpressed in SW480 with respect to SW620. This clone shows a 96% homology to clone AK000100 (Figure 3.13B). This clone corresponds to a protein of unknown function found in a colon cDNA library (Kawabata *et al.*, unpublished data). PSI-Blast analysis of this protein sequence yielded multiple alignments to proteins containing ankyrin motifs. This observation was verified by P-Fam analysis of the protein sequence which identified three ankyrin repeat domains (Figure 3.14). Visual analysis of the protein sequence identified a fourth ankyrin domain which is presumably below the significance threshold of the P-Fam program. A recent study suggests that the Ankyrin Repeat Domain (ARD) associates with CD44 and promotes the cytoskeletal activation in hyaluronic acid-mediated functions such as ovarian tumour cell migration (Zhu *et al.*, 2000). Further analysis of clone AV may clarify a role for proteins harbouring ankyrin repeat elements in the progression from primary to metastatic colorectal cancer.

Sequencing of the five DDRT-PCR products yielded three sequences with no homologies to known sequences or ESTs. This may be due to a number of factors. Some of the sequences were seen to contain Alu repeat elements. This may be due, in part, to a PCR or cloning artefact. Inherent in the DDRT-PCR experimental design is a

cDNA synthesis primed by an anchored oligo-dT primer, followed by PCR using the same oligo-dT primer with an upstream primer of arbitrary sequence. A design of this nature may generate cDNA molecules biased to the 3'- untranslated region (3'-UTR) of candidate genes. Sequences biased to the 3'-UTR may not be uniformly represented in the public nucleotide databases. Conceivably, the experimental design depends on the length of the 3'-UTR of a particular mRNA transcript and whether cDNA synthesis is sufficiently efficient to reach the open reading frame (ORF) of the mRNA transcript. In addition, the size of DDRT-PCR products must be considered. It is possible that the size range of PCR products separated on 6% polyacrylamide gels may not contain a high number of PCR products encompassing the ORF of these genes. It may be possible to circumvent this problem by introducing the SMART cDNA synthesis system (Clontech, UK). This protocol preferentially generates full length cDNA molecules which are amplified using a 'Long PCR'. cDNA generated in this manner may alleviate the problems associated with the DDRT-PCR protocol.

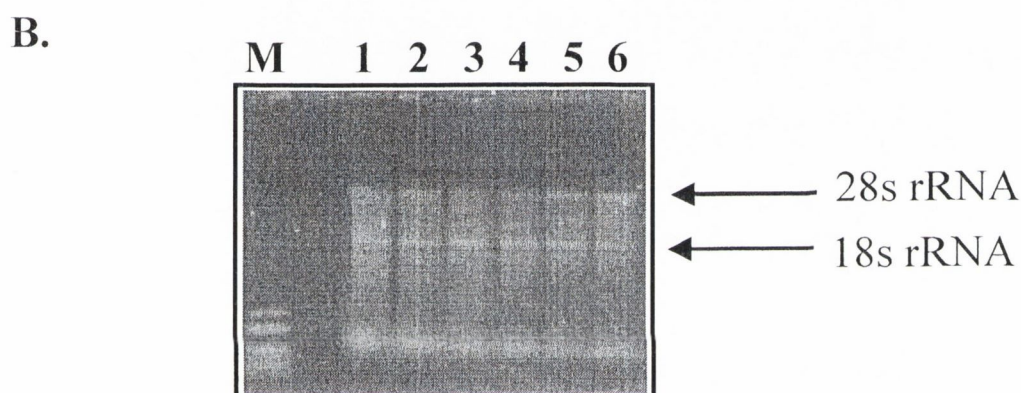
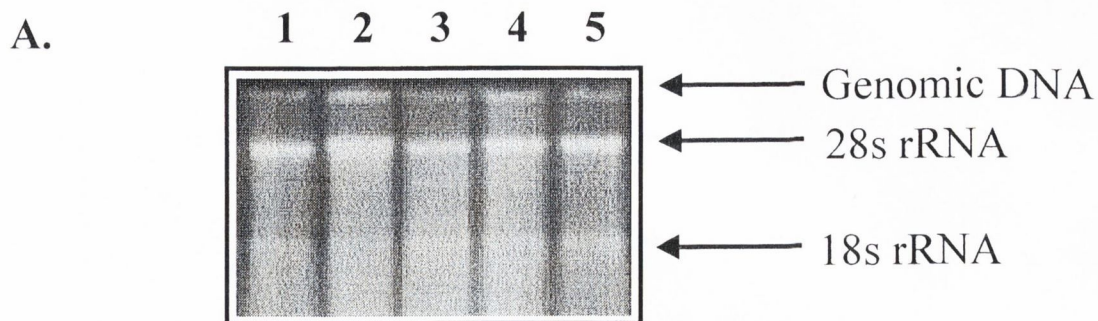


Figure 3.2 Analysis of RNA integrity and DNase I treatment of RNA. **A.** An illustration of total RNA electrophoresis. The integrity of total RNA for Differential Display Reverse Transcriptase PCR (DDRT-PCR) was assessed. RNA was electrophoresed on a denaturing formaldehyde 1% agarose gel stained with ethidium bromide. Intact ribosomal RNA bands (28S & 18S) are indicated by the arrows. Faint genomic DNA contamination is evident in the wells of the gel as indicated by the arrow. 1-4= total RNA samples isolated from the cell line SW480. **B.** An illustration of DNase I treated total RNA. Total RNA was DNase I treated and electrophoresed on a 1% denaturing formaldehyde agarose gel stained with ethidium bromide. High molecular weight genomic DNA is degraded. M= pBS-*Hae* III molecular weight marker; 1-6= DNase I treated RNA isolated from the cell line SW480 with no detectable genomic DNA.

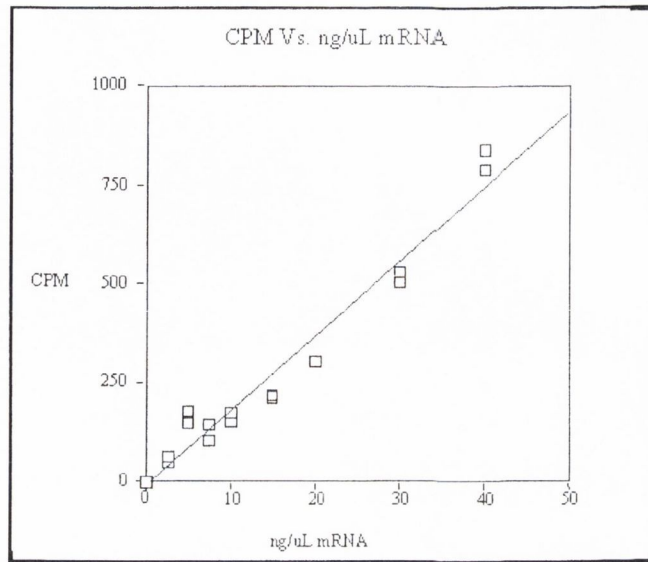


Figure 3.3 Hybridisation assay standard curve. A graph of scintillation counts per minute (cpm) with respect to ng/μL poly-A(+) mRNA standard. The hybridisation assay was performed and counts per minute (CPM) for each mRNA standard were plotted against the standard concentrations of mRNA. A linear plot was obtained. CPM for the unknown RNA samples were calculated and the standard was used to determine the concentrations of the unknown RNA samples.

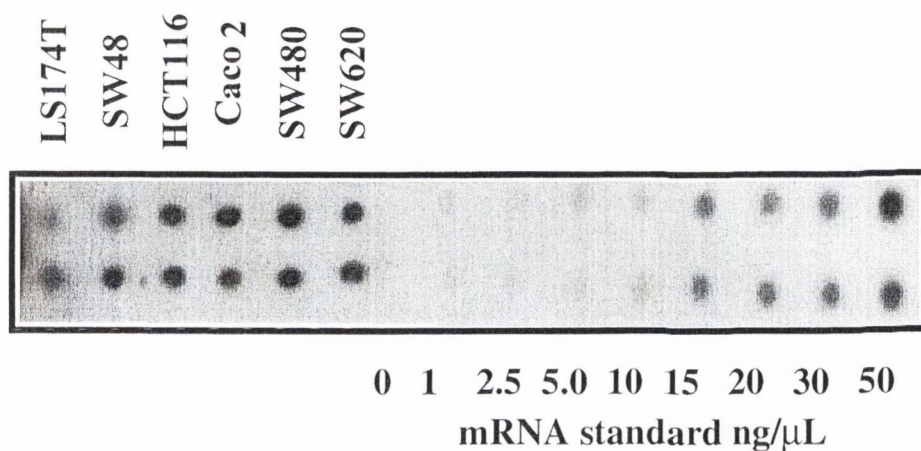


Figure 3.4 Autoradiograph of a typical hybridisation assay. An illustration of an autoradiograph obtained from a typical hybridisation assay. A dilution curve of standard mRNA was prepared from 0 – 50 ng/μL. Total RNA samples from the panel of sporadic colorectal cancer cell lines (Table 2.1) were assayed by the hybridisation assay in parallel with the mRNA standard. The MIN sporadic colorectal cancer cell lines used were LS174T, SW48 & HCT116 and the CIN sporadic colorectal cancer cell lines used were Caco2, SW480. The sporadic colorectal cancer cell line SW620 was included. Cell lines are listed adjacent to the respective hybridisation signals detected. The concentration of each standard is detailed adjacent to the respective hybridisation signals. All samples were assayed in duplicate.

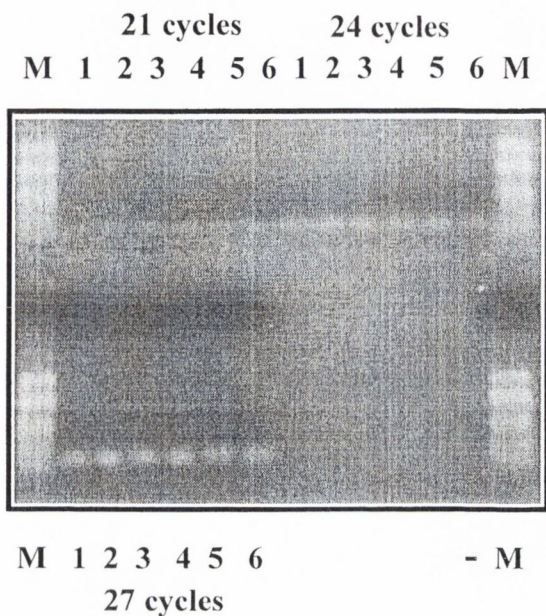


Figure 3.5 Expression of β -actin in a panel of sporadic colorectal cancer cell lines.

An illustration of a typical semi-quantitative RT-PCR experiment using specific primers for the human β -actin transcript. RT-PCR products were electrophoresed on 1% agarose gels stained with ethidium bromide. PCR products were sampled at the cycle numbers indicated above each panel of RT-PCR products. M= pBS-*Hae* III molecular weight marker; - = negative control; 1= LS174T,; 2= SW48; 3= HCT116; 4= Caco-2; 5= SW480; 6= SW620.

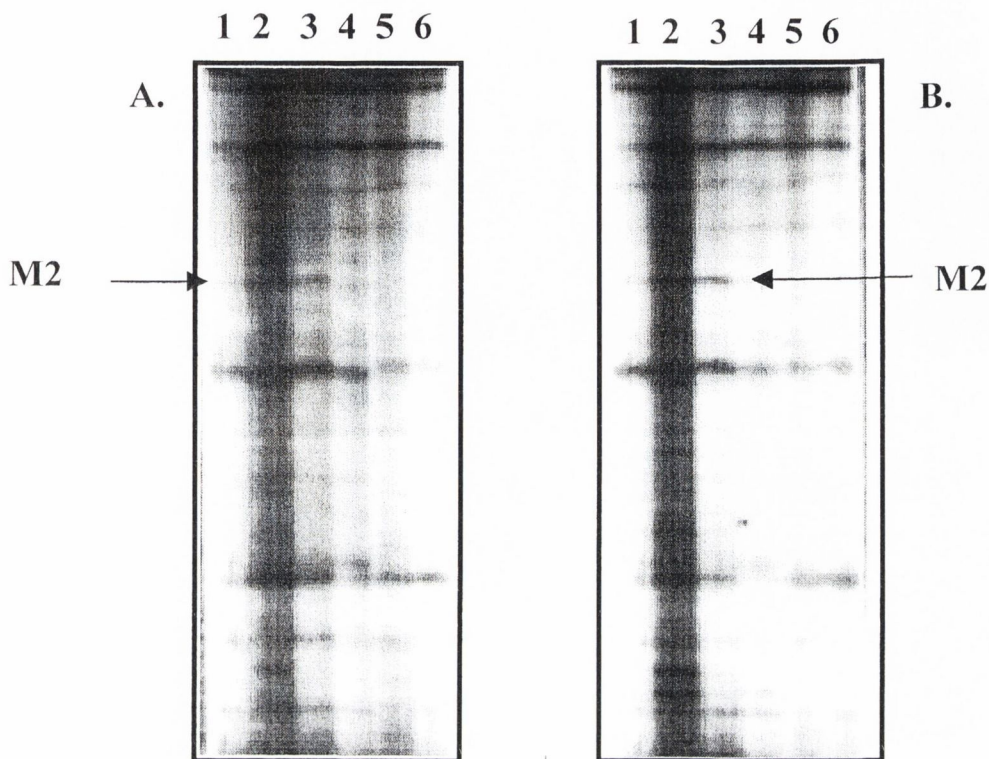


Figure 3.6 Typical Differential Display Reverse Transcriptase PCR pattern. An illustration of a typical Differential Display Reverse Transcription PCR (DDRT-PCR) pattern. In the complete DDRT-PCR analysis an anchor primer (AP3L; Table 2.2) was used in conjunction with six arbitrary upstream primers (DD1A-DD6A; Table 2.2). DDRT-PCR experiments were carried out on normalised total RNA from the panel of sporadic colorectal cancer cell lines (Table 2.1). DDRT-PCR products were electrophoresed on non-denaturing 6% polyacrylamide sequencing gels. Band patterns were visualised by autoradiography. DDRT-PCR reactions were carried out at various concentrations of arbitrary upstream primer. **A.** DDRT-PCR analysis using the anchor primer AP3L (5 μ M) and arbitrary upstream primer DD5A (2.5 μ M). **B.** DDRT-PCR analysis using the anchor primer AP3L (5 μ M) and the arbitrary upstream primer DD5A (5.0 μ M). M2 denotes a DDRT-PCR product overexpressed consistently over two DD5A concentrations in cell lines designated 1, 2 & 3 compared to the cell lines designated 4, 5 & 6. MIN cell lines: 1=LS174T; 2=SW48 & 3=HCT116. CIN cell lines: 4=Caco2 & 5=SW480. Additional cell line: 6=SW620.

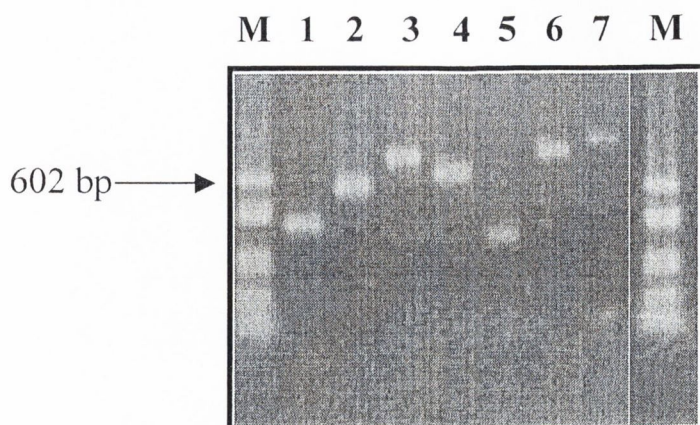


Figure 3.7 DDRT-PCR fragment reamplification. An illustration of the reamplification of DDRT-PCR products. Putative differentially expressed products were excised from dried sequencing gels and reamplified by PCR using the original DDRT-PCR amplification conditions (Section 2.2.1). Reamplified products were electrophoresed on 1% agarose gels and stained in ethidium bromide. The absence of contaminant non-specific products demonstrated successful reamplification for each DDRT-PCR product. These products were cloned directly into the p-BlueScript cloning vector for sequence analysis. M= pBS-*Hae*-III molecular weight marker; 1= M1; 2= M2, 3= M5; 4= M6; 5= M7; 6= M9 & 7= M11.

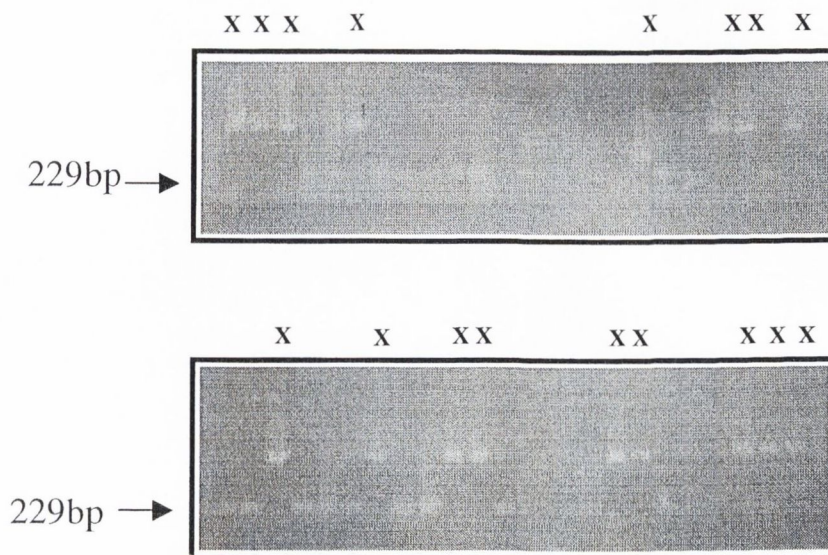


Figure 3.8 PCR screen of DDRT-PCR clones. DDRT-PCR clones were screened for inserts by PCR using the M13 Universal and M13 Reverse sequencing primers from pPCR ScriptTM Amp SK(+) cloning vector. Products were resolved on 1% agarose gels stained in ethidium bromide. **229bp** denotes insert negative PCR products. **X** denotes insert positive PCR products of different sizes verifying successful cloning of DDRT-PCR products.

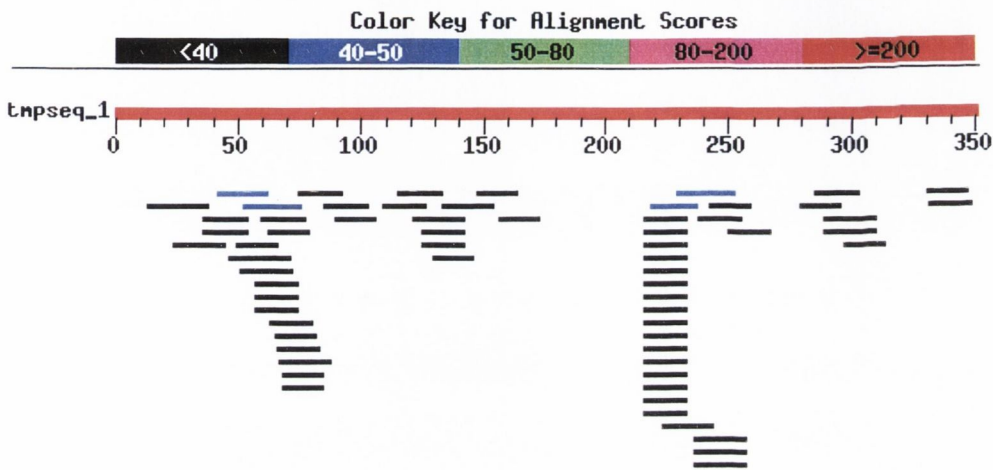

```

1 AGACACAAAG ATTTTCATTG CTCTTACATG AAGAGCATTG TAAAGTGTTA
51 GAATAAAGTA TACTTCTAAA ATGTAAAAAA GTTAAAGGAG ATTCTCACC
101 TCAATTCTTA GGAGATGCTG TATCTGGAAA GGGTAAACAG AAGCACACAG
151 TTTCTGAAA TTGTGCCAAG GCTACATGTC AACTGCCTGG TAGGAAATGC
201 TTTGGAGAAC CTAGATAAAC AGACTATCAT CTTTCTCTGG CTTTTTATAA
251 AGGAAATGTA CTGAGTTAAA CTAGCATCAT AGACACAACA AGCAGAAAAG
301 GGCACCGAGA AATTTTAGAT GATAGTACCT GTAGGATTTA TGACAATTAA
351 A

```

Figure 3.9 Clone M1 sequence. Clone M1 was generated with the primers DD1A and AP3L. DDRT-PCR showed clone M1 as being overexpressed in the cell lines LS174T, SW48, HCT116 with respect to the cell lines SW480 and SW620. Partial expression was detected in the cell line Caco-2.

Figure 3.9B BLAST analysis of DDRT-PCR clone M1. An illustration of the output obtained using the BLAST algorithm [<http://www.ncbi.nlm.nih.gov/BLAST>] with clone M1 as the query sequence. The illustration below shows the query sequence (in red) and the alignments found by the BLAST algorithm. The top ten alignments are shown textually. No significant matches were found.



Ten most significant hits.

gb AC007275.3 AC007275	Homo sapiens clone NH0109F19, comple...	42	0.13
gb AC007052.4 AC007052	Homo sapiens chromosome 18, clone hR...	42	0.13
gb AE001163.1 AE001163	Borrelia burgdorferi (section 49 of ...	42	0.13
gb AE003620.1 AE003620	Drosophila melanogaster genomic scaff...	40	0.50
gb AC007920.18 AC007920	Homo sapiens 3q27 BAC RP11-208N14 (...	38	2.0
gb AE001818.1 AE001818	Thermotoga maritima section 130 of 1...	38	2.0
gb U67460.1 U67460	Methanococcus jannaschii section 2 of 15...	38	2.0
gb U22376.1 HSU22376	Human (c-myb) gene, complete primary c...	38	2.0
emb AL109839.11 HSJ1069P2	Human DNA sequence from clone RP5...	38	2.0
emb AL009178.4 HS431P23	Human DNA sequence from clone 431P2...	38	2.0

gb|AC007275.3|AC007275 Homo sapiens clone NH0109F19, complete sequence
Length = 169904

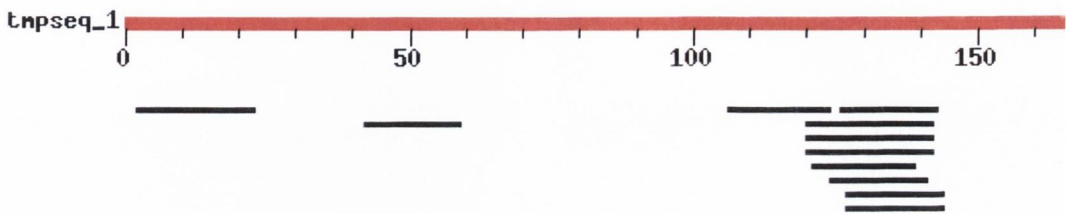
Score = 42.1 bits (21), Expect = 0.13
Identities = 24/25 (96%)
Strand = Plus / Minus

Query: 53 ataaagtatacttctaaaaagttaa 77
 |||||
Sbjct: 120580 ataaagtatacttctaaaaagttaa 120556

1 TCATCTGGGA ATCACCATCA CATACTCAGG CAAAGAACGC TGCCAGGCAC
51 CAGCTGGTGG AACCCCAACA CTCAGGAAGC GGAAGGTCAG GGCAAGGCAG
101 AGACTATGGT GCCCAAGCCT GACACACCAG CTGCCTTGCA GAGTTATCAG
151 ATGATCCTAC AAGCT

Figure 3.10 Clone M2 sequence. Clone M2 was generated with the primers DD1A and AP3L. DDRT-PCR identified clone M2 as being overexpressed in the cell lines LS174T, SW48, HCT116 with respect to the cell lines SW480 and SW620.

Figure 3.10B BLAST analysis of DDRT-PCR clone M2. An illustration of the output obtained using the BLAST algorithm [<http://www.ncbi.nlm.nih.gov/BLAST>] with clone M2 as the query sequence. The illustration below shows the query sequence (in red) and the alignments found by the BLAST algorithm. The top ten alignments are shown textually. No significant matches were found.



Ten most significant hits.

gb AC004149.1 AC004149	Homo sapiens chromosome 17, clone hC...	38	0.88
dbj AP001423.1 AP001423	Homo sapiens genomic DNA, chromosom...	38	0.88
dbj AP000163.1 AP000163	Homo sapiens genomic DNA, chromosom...	38	0.88
dbj AB020863.1 AB020863	Homo sapiens genomic DNA of 8p21.3-...	38	0.88
dbj AP000021.2 AP000021	Homo sapiens genomic DNA of 21q22.2...	38	0.88
gb AF095725.1 AF095725	Homo sapiens PAC LLNLP704E02527Q3 ol...	36	3.5
gb AF095145.1 HSPOMT1G09	Homo sapiens protein O-mannosyl-tr...	36	3.5
gb AC007106.6 AC007106	Homo sapiens chromosome 4 clone C038...	36	3.5
gb AF052041.1 AF052041	Homo sapiens olfactory receptor gene...	36	3.5
gb AC005900.1 AC005900	Homo sapiens chromosome 17, clone hR...	36	3.5

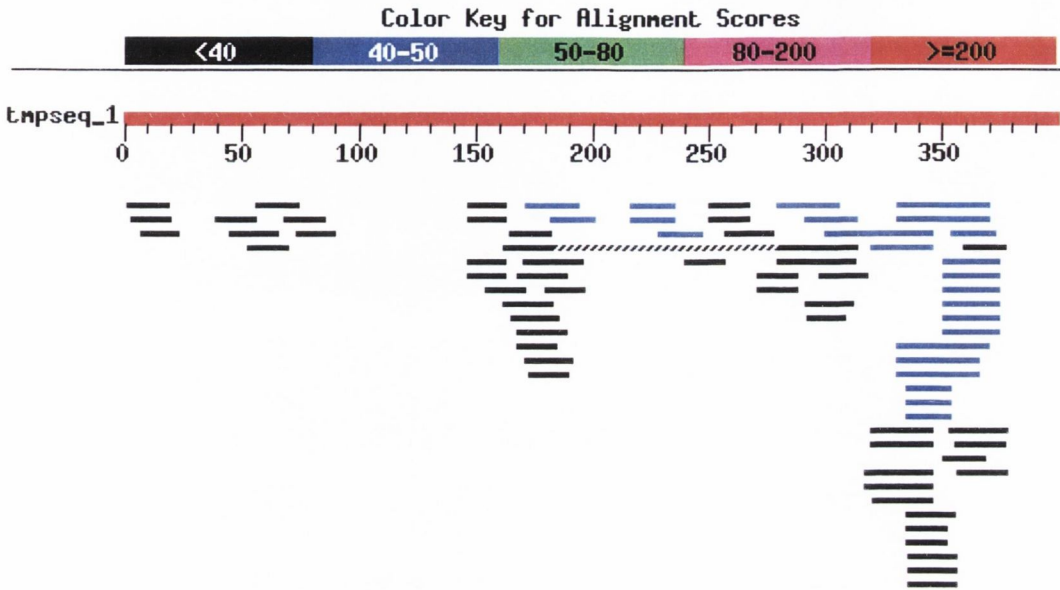
gb AC004149.1 AC004149	Homo sapiens chromosome 17, clone hCIT.501_O_10, complete sequence	
	Length = 84767	
	Score = 38.2 bits (19), Expect = 0.88	
	Identities = 19/19 (100%)	
	Strand = Plus / Minus	

Query: 122	acacaccagctgccttgca	140
Sbjct: 68171	acacaccagctgccttgca	68153

1 TGATAGTCCT GATACTGAGA AGGGTTAGCT AGCTAGTCCG TGTTTAGTGA
51 AGTAAACAAA GCATTAACCT CCACACAACA CAGATGCCCT CCATAAAAAA
101 ATGTGGTGTG AACAGTCTGT TCTTACAATA GACTCCCAAT TTTTTTAAAA
151 AAGGGGAGAA TGGAAATAGA AATTTATCAT TTAGCAAAAA TTACAGTGAT
201 AATCACTGCG AGACAGAATC GTCAATGGAA TGCTAAGATC AGTGGGTGAA
251 AGTGAGGAAC AGAATGTTGC AGTATCTATG TATCTTCCCA GAAGATATTT
301 CTTAATTATA AAGGGGGAAC AGTAACCTTG TAGTGGAGAA ACCTGGCTTC
351 CATTACCTTA ACCAAATGAT CCAAGTTAGA GCCACCTGTA AATTAAGTTG

Figure 3.11 Clone M5 sequence. Clone M5 was generated with the primers DD4A and AP3L. DDRT-PCR showed clone M5 as being overexpressed in the cell lines LS174T and Caco-2 compared to the remaining cell lines of the panel: SW48, HCT116, SW480 and SW620.

Figure 3.11B BLAST analysis of DDRT-PCR clone M5. An illustration of the output obtained using the BLAST algorithm [<http://www.ncbi.nlm.nih.gov/BLAST>] with clone M5 as the query sequence. The illustration below shows the query sequence (in red) and the alignments found by the BLAST algorithm. The top ten alignments are shown textually. No significant matches were found.



Ten most significant hits.

gb AC007372.4 AC007372	Homo sapiens chromosome 14 BAC conta...	48	0.002
emb AL049837.4 CNS0000F	Human chromosome 14 DNA sequence **...	48	0.002
gb AC007551.1 AC007551	Homo sapiens clone UWGC:djs104 from ...	46	0.009
emb AL049797.8 HSJ661G12	Human DNA sequence from clone 661G...	46	0.009
ref NC_001145.1	Saccharomyces cerevisiae chromosome XIII,...	42	0.15
emb X58693.1 SCYAP1	Yeast YAP1 gene for a transcriptional a...	42	0.15
emb X63268.1 SCSNQ3	S.cerevisiae SNQ3 gene confering mutage...	42	0.15
emb X53830.1 SCPDR4	Yeast PDR4 gene for pdr4 protein	42	0.15
emb X60780.1 SCPAR1G	S.cerevisiae PAR1 gene	42	0.15
emb Z49810.1 SC9571X	S.cerevisiae chromosome XIII cosmid 9571	42	0.15

gb AC007372.4 AC007372	Homo sapiens chromosome 14 BAC containing gene for type 2 iodothyronine deiodinase (DIO2) gene, complete CDS, complete sequence Length = 154796
------------------------	--

Score = 48.1 bits (24), Expect = 0.002
Identities = 36/40 (90%)
Strand = Plus / Minus

Query: 332 agtgagaaacctggcttcattaccttaaccaaagatc 371
 ||||| ||| ||||| |||
Sbjct: 7257 agtgagaaacctggcagacatcaccttaaccaaagatc 7218

```

1  GGTATATAA  GTCGGTTTAC  TGGAACCTGG  AGACTGTTTT  CTAAAAAAG
51 TTATACTCTG AATATTAAGA AAGAGAATAA AACTAGTACA CACACACACA
101 CACGCACACA CACACACCCT GTCCCAATTT TCCCTGTAGT AGAGCATATA
151 CAAAGGGGAA AAAAAAGACA GAGAAGAAAA TGTCTCTGTG TACCTTAAAT
201 CCTGGCACTG AGAAGTACTG GTATATTTAG AAGACGCCAT CAGTTTTATG
251 AGGAATGGGA GGGAAATTTT AAATTCCTTG AAGGGCTAAA ATATACTGTA
301 TGGTTTTTCA TTTTTTTTTT TCTTGATGCA TCTTGTTAGT AATACTAATA
351 TTGTGTTTTA CCTTTAATCA GCTATGCCTT TTTTTTTTTT TTTTTTTTTT
401 TTTTGTAGAC GCAGTTCGCT CTGTCGCCCA GGCTGGAGAG CAATGGCATG

```

Figure 3.12 Clone M6 sequence. Clone M6 was generated with the primers DD4A and AP3L. DDRT-PCR showed clone M6 as being overexpressed in the cell lines LS174T, HCT116 and Caco-2 with respect to the cell lines SW480 and SW620. Partial expression was detected in the cell line SW48.

Figure 3.12B BLAST analysis of DDRT-PCR clone M6. An illustration of the output obtained using the BLAST algorithm [<http://www.ncbi.nlm.nih.gov/BLAST>] with clone M6 as the query sequence. The illustration below shows the query sequence (in red) and the alignments found by the BLAST algorithm. The top ten alignments are shown textually. A total of 1,748 alignments were found. The ten most significant matches are likely to represent Alu alignments. This may be accounted for by the match to AC003029.1. Clone M6 matches to 29 distinct regions of the PAC clone AC003029.1 which suggests that M6 contains an Alu element from approximately base 416 to the end of the clone. The first match to the AC003029.1 clone is shown textually.

Ten most significant hits.

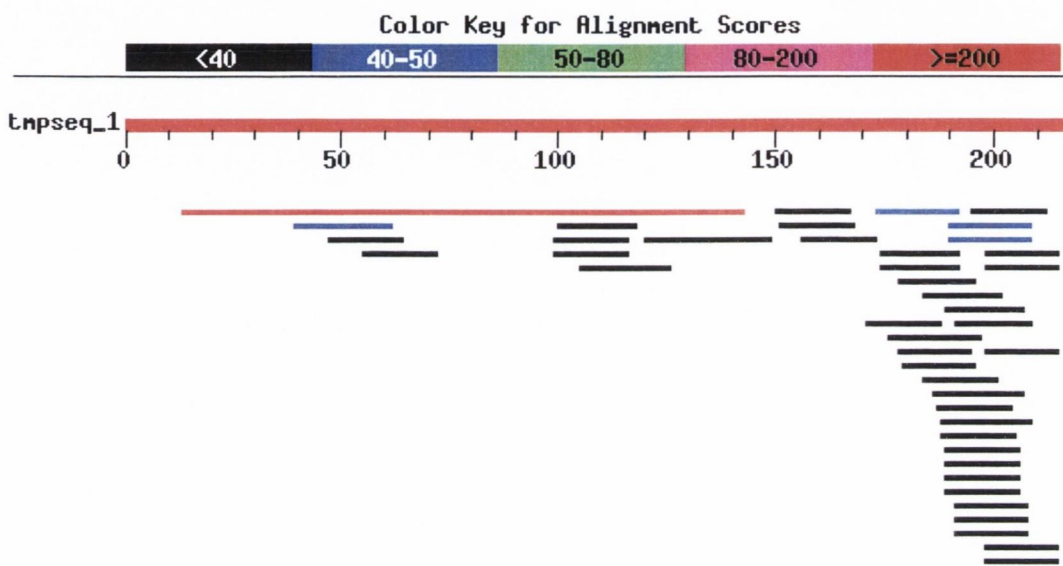
gb|AC003029.1|AC003029 Homo sapiens Chromosome 12q24 PAC RPCI3-462E2 (Roswell Park
Cancer
Institute Human PAC library) complete sequence
Length = 139166

Query: 416 tcgctctgtcgcccaggctggagagcaatggc 447
|||||
Sbjct: 136365 tcactctgtcgcacgaagctgagagcaatgac 136334

1 AGCCGGATCC ATATTCTATA ACTATGGCAT CTTAAATGGG GGTCATCAGA
51 ATGTATTTCC TAATCATAAT AGTGTGGGAA CAAATCGAAA GAGATGCTAT
101 GGAAGACTCA GAGGACTTTG GAGTAAAGAA AGCTAGAACT GAAGGTGAGA
151 CCGCTTTGGG TGGGAAGAGC ACACTTATTT TTCCTTTCTG TAATATGTTT
201 TCTTTTATA GCTGAG

Figure 3.13 Clone AV sequence. Clone AV was generated with the primers DD5A and AP3L. To complement previous metastasis studies carried out on the isogenic pair of cell lines SW480 and SW620 DDRT-PCR showed clone AV as being overexpressed in the cell line SW480 compared to the cell line SW620.

Figure 3.13B BLAST analysis of DDRT-PCR clone AV. An illustration of the output obtained using the BLAST algorithm [<http://www.ncbi.nlm.nih.gov/BLAST>] with clone A6 as the query sequence. The illustration below shows the query sequence (in red) and the alignments found by the BLAST algorithm. The top ten alignments are shown textually. A significant match (4e-53) was found between the query sequence (clone AV) and the clone AK000100. This is highlighted textually in bold text below the graphical representation.



Four most significant hits.

dbj AK000100.1 AK000100	Homo sapiens cDNA FLJ20093 fis, clo...	212	4e-53
gb AC007071.6 AC007071	Arabidopsis thaliana chromosome II s...	40	0.30
gb AC007317.3 AC007317	Homo sapiens BAC clone RP11-414K7 fr...	40	0.30
gb AC006336.3 AC006336	Homo sapiens BAC clone RP11-508K5 fr...	40	0.30

dbj|AK000100.1|AK000100 Homo sapiens cDNA FLJ20093 fis, clone COL04263
Length = 2509

Score = 212 bits (107), Expect = 4e-53
Identities = 126/131 (96%), Gaps = 1/131 (0%)
Strand = Plus / Plus

Query: 14 ttctataactatggcatcttaaatgggggtcatcagaatgtatttcctaatacataatagt 73
 ||||| |||||
Sbjct: 697 ttctataacaatggcatcttaaatgggggtcatcagaatgtatttcctaatacatattagt 756

Query: 74 gtgggaacaaatcgaaagagatgctatggaagactcagaggactttggagtaaagaaagc 133
 ||||| |||||
Sbjct: 757 gtgggaacaaatcgaaagagatgct-tggaggactcagaagactttggagtaaagaaagc 815

Query: 134 tagaactgaag 144
 |||||
Sbjct: 816 tagaactgaag 826

Model	Seq-from	Seq-to	HMM-from	HMM-to	Score	E-value	Description
!! ank	57	89	1	33	26.4	2.4e-05	Ank repeat
!! ank	91	123	1	33	19.8	0.0023	Ank repeat
!! ank	124	156	1	33	36.3	2.4e-08	Ank repeat



ank 57-89
ank 91-123
ank 124-156

Figure 3.14 Ankyrin repeat motifs in unknown protein. An illustration of ankyrin repeats motifs in the unknown protein derived from the clone AK000100. The protein sequence from clone AK000100 was entered into the P-Fam database [<http://pfam.wustl.edu/hmmsearch.shtml>]. This reveals three ankyrin domains at the amino terminal of the protein. Visual analysis of the protein sequence reveals a possible fourth ankyrin domain upstream of the first domain which is presumably below the significance threshold value for the P-Fam program.

DDRT-PCR Band Name	Approximate Size (bp)	Arbitrary Primer	Expression Pattern
1A	800	DD1A	1,2,3+
1B	800	DD1A	3,4,5+
M1	250	DD1A	1,2,3,,4+
M2	350	DD1A	1,2,3+
M3	900	DD2A	3,4,5+
M4	400	DD2A	5+
M5	450	DD4A	1,3,4+
M6	400	DD4A	1,2,3+
M7	250	DD4A	1,4+
M8	850	DD5A	3,4,5+
M9	600	DD5A	1,2,3,4,,5+
M10	600	DD6A	1,3,4+
M11	500	DD6A	4,5,6+
M12	550	DD6A	4,5,6+
M13	1000	DD6A	1,2,3+
M14	1000	DD6A	1,2+
PC	600	DD6A	1,2,3+

Table 3.1 Summary of the DDRT-PCR products isolated. DDRT-PCR was carried out as described and products were excised from gels. Each product was named and the approximate fragment sizes are detailed. The arbitrary primer used in the generation of each particular product is listed and the expression patterns for each DDRT-PCR product are summarised. Expression profiles are read numerically, i.e. 1+ = a product detected in cell line 1. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480 & 6=SW620.

CHAPTER 4

GENE DISCOVERY ARRAY

ANALYSIS

4.0 INTRODUCTION

The technology available to examine global gene expression patterns in biological specimens has advanced enormously over the last three years. The techniques typically employed include Differential Display Reverse Transcriptase PCR (DDRT-PCR) (Liang *et al.*, 1992), Serial Analysis of Gene Expression (SAGE) (Velculescu *et al.*, 1995), Suppression Subtractive Hybridisation (SSH) (Diatchenko *et al.*, 1996), cDNA microarray (DeResi *et al.*, 1996; Lockhart *et al.*, 1996) and commercially available high density cDNA filters supplied by companies such as Genome Systems Incorporated, Research Genetics and Clontech.

The original subtractive hybridisation techniques were labour intensive with a low degree of sensitivity towards low abundance transcripts (St. John *et al.*, 1979; Sargent *et al.*, 1983). The sensitivity of this technology has been improved by the addition of a PCR step in the protocol (Hubank *et al.*, 1994). The advantages of DDRT-PCR include the requirement for small amounts of starting material (as little as 200ng of total RNA), which makes it suitable for small tissue samples, and its applicability to multiple RNA samples. However, a high degree of false positives has been widely associated with DDRT-PCR analyses (Francia *et al.*, 1996) and the technique yields specific differences in gene expression rather than a complete gene expression profile. DDRT-PCR is discussed in more detail in Chapter 2.

Serial analysis of gene expression (SAGE) (Velculescu *et al.*, 1995) generates thousands of tags (which represent cDNA species) from each of the two RNA populations under investigation. Tag abundances correspond to the level of gene expression of a particular transcript. The advantage of this technique is that a complete gene expression profile is obtained. Large amounts of starting material (up to 5µg poly-A(+) mRNA) are required for SAGE which limits the technique to cultured cells or large tissue samples, although a recent report details a new protocol termed microSAGE (Datson *et al.*, 1999) which may be useful for small tissue samples. Moreover, the technique is limited to the analysis of two samples simultaneously and places heavy demands on sequencing time.

cDNA microarrays were first described by Schena *et al.* (1995). These are high capacity systems used to monitor the expression of multiple genes in parallel. High speed robotics are used to print thousands of cDNA molecules on glass slides, which

may be probed with fluorescent cDNA probes. This technique has been applied to the analysis of human cancer cells (DeRisi *et al.*, 1996). The technology is expensive and specialised and is not amenable to general molecular biology laboratories.

A number of companies (Genome Systems Inc., Research Genetics, Clontech) have developed nylon membranes harbouring thousands of cDNA species. cDNA probes labeled with [α - 33 P]dCTP may be hybridised using regular hybridisation protocols to duplicate membranes and gene expression profiles between different RNA populations may be obtained using software which quantitates hybridisation signals on the phosphorimage files of the probed membranes. Small amounts of poly-A(+) mRNA (1 μ g) are required, making this technology suitable for both scarce tissue samples and cultured material. The technical aspects of cDNA array filters are based on standard molecular biology techniques making the methodology applicable to most non-specialised laboratories. Rapid results (two weeks) may be expected from multiple RNA samples at a relatively low cost. cDNA arrays have been used to examine the gene expression responses between untreated keratinocytes and keratinocytes treated with epidermal growth factor (Trenkle *et al.*, 1998).

The main objective of this work was to implement the Gene Discovery Array (GDA) v1.3 filters to determine gene expression differences between MIN sporadic colorectal cancer cell lines (LS174T, SW48, HCT116) and CIN sporadic colorectal cancer cell lines (Caco-2, SW480, HT-29). The experimental design involved sequential probing of duplicate GDA filters with three pairs of MIN and CIN total cDNA probes to analyse the entire panel of cell lines.

4.1 GENE DISCOVERY ARRAY (GDA) FILTERS v1.3

4.1.1 GDA clone library

The GDA v1.3 filters consist of two duplicate 22x22-cm nylon filters and two randomly cut 6x6-cm test GDA filters. Genome Systems Inc. (GSI) generated the clones on the GDA v1.3 filters by retrieving human EST sequences (203,911 sequences) from the I.M.A.G.E. consortium clone libraries (Miller *et al.*, 1997). Sequences matching vector, bacteria, mitochondria or ribosomal RNA were removed leaving approximately 400,000 high quality EST sequences. The ESTs from this database were grouped into unique clusters to eliminate redundancy in the database. ESTs that matched a known full-length gene were assigned to a Unigene cluster corresponding to that particular gene. The remaining ESTs were compared to each other using sequence alignment software. All ESTs with at least 95% sequence identity over a 40 base pair minimal overlap with another EST were placed into the same cluster. Of the 203,911 clones, 3,230 matched full-length cDNA sequences exactly (1,291 of these were discarded because they were grown in SOLR cells which are incompatible with DH10B cells, in which the majority of the clones were grown). GDA v1.3 filters contain 1,939 representative clones from the Human Gene Index clusters, 14,898 representative clones from other clusters and 1,489 singleton clones (clones which did not match any other clones in the original 203,911 clone library). Figure 4.1 details the bioinformatics strategy adopted in choosing the cDNA species which are spotted onto GDA v1.3 filters.

4.1.2 GDA Filter Structure

The clones chosen by GSI were re-arrayed into 384 well culture dishes. These cultures were gridded robotically onto the nylon membranes in a double spotted pattern. Each 22x22-cm nylon filter is divided into six sections, which are termed fields. Each field on the grid is sub-divided into 384 small squares. These are assigned a position by means of a numbered column and an alphabetical row. A positive hybridisation signal in the "1A" small square corresponds to the cDNA clone in position "1A" in one of the 48 microtitre dishes that house the GDA library. Each square is further divided into 16 units which have been assigned a microtitre dish number. Each microtitre dish number is repeated twice in each small square, *i.e.* within one small square in a particular field

each individual clone has been spotted in duplicate. The details of the structure of the GDA v1.3 filters are illustrated in Figure 4.2.

4.1.3 Normalisation of duplicate filters

GSI provide run-off RNA corresponding to the *Arabidopsis thaliana* and *Drosophila melanogaster* internal control clones spotted on the GDA v1.3 filters. Hybridisation intensity data from these spots may be plotted against the amounts of RNA controls on the filters. The slope of this line can be used to normalise the hybridisation intensities of all the spots on one filter against their counterpart spots on the second filter. In a case where the controls do not yield a linear plot, the filters may be normalised by global means. For each point, X_B , on filter B a new value is calculated to compare to the X_A value. X_B is multiplied by a constant value, which is derived by dividing the mean of filter A over the mean of filter B. The total summation of intensity values for both filters is identical. The ratios X_A/X_B for each point are calculated and are sorted by the absolute difference between the average background and the X_A/X_B value for each point. The Gene Discovery Array software is based on an algorithm which carries out the normalisation process rapidly [Gene Discovery Array v1.3 documentation, GSI, USA].

4.1.4 GDA hybridisation protocol

Poly-A(+) mRNA (2.5µg) or total RNA (1 µg) is isolated from the samples of interest and reverse transcribed in the presence of [α - 33 P]dCTP using an HPLC purified cDNA synthesis primer of the structure 5'-dT₁₈VN-3' (Chapter 2, Section 2.3). This is a 20-mer oligonucleotide composed of 18 T residues followed by an A, G or C in the 19th position and a mixture of A, G, C or T at the 3' position. This structural feature anchors the oligonucleotide to the junction of the poly-A tail with the 3'-untranslated region of mRNA molecules immediately prior to cDNA synthesis. RNA corresponding to the *Arabidopsis thaliana* and *Drosophila melanogaster* internal control clones on the filters is reverse transcribed using a similar method in parallel. The cDNA products from the experimental samples and the control samples are combined and purified using G-50 sephadex columns.

DNA orientation markers (derived from the *E. coli* vector pBR325) are radiolabeled with [α -³³P]dCTP by random-priming and are purified using G-50 sephadex columns. The radiolabeled cDNA samples and DNA orientation markers are added to a GDA v1.3 filter in hybridisation mix overnight at 42°C. Filters are washed using standard washing solutions at 68°C and analysed overnight using a phosphorimager set at a resolution of 100 microns. The image files may be analysed using the Gene Discovery Array software or may be sent to GSI who perform GDA analyses. A summary of the GDA v1.3 protocol is detailed in Figure 4.3.

Results from GDA analyses can be obtained from a secure GDA server [<http://www.genomesystems.com/secure>] which is protected by private User Identity and Password. There are three ways of viewing GDA v1.3 expression data. The raw data output is a complete gene expression profile of all the cDNA species arrayed on the filters. The investigator may exclude ESTs from the data outputs to yield an expression profile of full length transcripts. Zero intensity signals [clones which shows no detectable difference in hybridisation signal with respect to the background noise signal] may also be eliminated from the expression profile. In addition, the gene expression profile may be sorted in three different ways: i) on the basis of score (which accounts for expression level and intensity difference); ii) intensity difference or iii) ratio of expression levels. The investigator may choose the most suitable parameters.

4.2 RESULTS

4.2.1 Introduction

Initially, the protocol was tested using the GDA v1.3 test filters that represent randomly sectioned 6x6-cm squares of actual GDA v1.3 filters. Probes were synthesised as described in Chapter 2 from poly-A(+) mRNA isolated from the cell lines LS174T (MIN) and SW480 (CIN). Both probes were hybridised separately to each test filter overnight. The filters were subjected to autoradiography at -70°C for 72 hours. This revealed a clear pattern of hybridisation signals for both filters (data not shown), thereby verifying the quality of the mRNA samples isolated and the success of the GDA v1.3 hybridisation protocol.

Duplicate GDA v1.3 filters were probed in parallel with radiolabeled cDNA samples derived from paired MIN and CIN sporadic colorectal cancer cell lines. Following phosphorimager analysis the images obtained were saved in PC “gel” format. The images were transferred to GSI by means of file transfer protocol (ftp) to the Genome Systems ftp server [ftp://ftp.incyte.com]. GSI carried out the analysis of the filters using the Gene Discovery Array software. Results were obtained from the password-secured GSI web-site.

4.2.2 Implementation of the GDA V1.3 Protocol

Total RNA and poly-A (+) mRNA were isolated as described in Section 2.1.3 and 2.3.3. RNA integrity was assessed by denaturing formaldehyde gel electrophoresis and was of comparable quality to the RNA utilised in DDRT-PCR analyses (Figure 3.2). Figure 4.4 illustrates the isolation of high quality poly-A (+) mRNA from the cell lines LS174T and SW480. RNA of this quality was deemed suitable for reverse transcription in the presence of $[\alpha\text{-}^{33}\text{P}]\text{dCTP}$. Due to the large amounts of total RNA required for probe synthesis, the SMARTTM cDNA synthesis kit (Clontech, UK) was also used to generate cDNA for probe synthesis. SMART cDNA was amplified by PCR according to the Clontech protocol (Section 2.3.3) over a range of PCR cycle timepoints. Figure 4.5 illustrates the amplification of SMART cDNA by PCR from the cell lines SW48 (MIN) and Caco2 (CIN). Equal amounts of PCR product were observed after 18 cycles for SW48 and 19 cycles for Caco2. SMART cDNA PCR products were purified from

each cell line by ammonium acetate precipitation (Figure 4.6). The SMART cDNA PCR protocol was chosen to prepare cDNA probes from the third pair of cell lines HCT116 (MIN) and HT29 (CIN).

A measure of the success of the GDA protocol is illustrated in Figure 4.7 in an example of typical expression differences seen on the GDA v1.3 filters. The Solute Carrier Family 24 gene is clearly shown to be overexpressed in the Caco-2 cell line with respect to the cell line SW48. It is worth noting that the clone is clearly dotted in duplicate.

4.2.3 Scatter Plot

High-throughput mRNA analysis platforms such as high density cDNA array technologies are capable of providing large scale expression profiles which have the potential to increase our understanding of cancer biology at the molecular level. It is useful to gain an overall impression of how the entire complement of genes is being expressed in these analyses. This gives the investigator clues as to how successful an experiment has been and the degree of reproducibility of the experimental design between repeat experiments (Cole *et al.*, 1999). A simple way of achieving this is by plotting a graph between the expression levels of genes on the filters for both experimental samples. The majority of genes (between similar cell types) are likely to be expressed at comparable levels. This should be clearly seen on a scatter plot with the highest density of data points appearing coincident with the line $y=x$.

To achieve this, we have downloaded [<http://www.genomesystems.com/GDA/>] the complete expression profiles obtained from the analysis of the sporadic colorectal cancer cell lines described in Table 1.2. The expression values for each clone from each experiment were extracted using a simple Perl script. Values from MIN clones were plotted against values for CIN clones producing a scatter plots for each MIN/CIN pairwise comparison. The scatter plots obtained are shown in Figure 4.8.

4.2.4 Identification of consistently differentially expressed genes

One problem associated with high density cDNA array methodology is the volume of data generated from a single experiment. A related problem is the difficulty of determining the consistently differentially expressed genes between the three pairwise comparisons of MIN and CIN colorectal cancer cell lines. A simple Perl Script program

(GeneMatch v1.0) was coded (Mr. Philip Hassell) to facilitate the identification of this subset of key genes from the GDA analyses. The complete raw expression profiles were downloaded from <http://www.genomesystems.com/secure> in tab-delimited format and saved as Unix compatible .txt files. In each instance, the 1,000 most highly differentially expressed genes from the three pairwise comparisons were entered into GeneMatch. This Perl script generated an output .txt file containing the subset of genes (based on their Unigene Identifiers) which appeared in all three lists of top 1,000 differentially expressed genes. These genes are detailed in Tables 4.1 and 4.2. The genes selected for further analysis are highlighted in bold font. The number of differentially expressed cDNA species between LS174T and SW480 was greater than 1,000. A similar number of differentially expressed genes was observed in the second and third comparisons.

4.2.5 Categorisation of the gene expression profiles

The nature of the genes found by high density cDNA array screening is a determining factor of the downstream experiments possible to further analyse these genes. Biological information of this nature leads the investigator into the 'next-step' experiments to be carried out which often involve functional analyses. The candidate genes obtained from the GDA expression profiles were thoroughly scrutinised for potential function by the utilisation of bioinformatics. The process used is summarised in Figure 4.9. The Gene Discovery Array (GDA) library [<http://www.genomesystems.com/GDA>] details the clones and matches each one to a unique Unigene code number. The Unigene numbers were entered into the Unigene database [<http://www.ncbi.nlm.nih.gov/UniGene>] at the National Centre for Biotechnology Information (NCBI). Information is available for each gene and Internet links to the protein sequences in human and the homologous protein sequences in *M. musculus*, *D. melanogaster* and *S. cerevisiae* where available. The functional properties of proteins in these model organisms give useful clues to functions in human.

The Unigene entry also gives an Internet link to Online Mendelian Inheritance in Man (OMIM; <http://www.ncbi.nlm.nih.gov/omim>). This is a biological server dedicated to the accumulation of the biological functions, properties and the most recent research findings of all the characterised genes in human disease. In the first instance, a short summary of the gene of interest is given, including all of the published alternative

names for the gene. A highly detailed catalogue follows including genetic information such as chromosomal location, alternative splice variants, characterised mutations and deletions. In addition, tissue expression profiles are given where known and Internet links to the most recent scientific abstracts are incorporated. This procedure was followed for all 70 differentially expressed genes. The functions were divided into seven broad categories as described at The Institute for Genomic Research (TIGR) and by Craig Venter (Adams *et al.*, 1995) and are detailed in Tables 4.1 and 4.2. The broad categories are cell division, cell signaling/cell communication, cell structure/motility, cell/organism defense, gene/protein expression, metabolism and unclassified. These categories have also been sub-divided into more specific functions.

4.2.6 Correlation of GDA data with the SAGE database

It was deemed necessary to measure the significance of the expression differences found between the MIN and CIN cell lines using the GDA v1.3 filters. A useful control of these differences was found at the public access SAGE database (Lal *et al.*, 1999). The method of SAGE (Velculescu *et al.*, 1996) has been carried out on various tissue types and cell lines including the sporadic colorectal cancer cell lines HCT116 (MIN) and Caco-2 (CIN). The expression profiles from these tissues and cell lines have been uploaded for public access at <http://www.ncbi.nlm.nih.gov/SAGE>. The process adopted to determine the expression differences between the genes tabulated in Tables 4.1 and 4.2 using the public SAGE data is detailed in Figure 4.10. The SAGE database may be accessed using the Internet location above. The Unigene code number for the gene of interest is entered and the abundance of SAGE tags is displayed for each cell line. The SAGE database may be accessed indirectly from the Unigene database [<http://www.ncbi.nlm.nih.gov/UniGene>] for specific genes also. The Unigene database is integrated with the SAGE database and provides a convenient direct Internet link between Unigene and SAGE for the genes under investigation. Moreover, the Unigene database details all expressed sequence tags (EST) which show significant alignment with the gene of interest. The results of the SAGE database searches and EST abundances from both the MIN and CIN expression profiles are detailed in Tables 4.3 and 4.4.

4.2.7 Identification of simple repeat sequences in the differentially expressed genes

The MIN genotype is clearly recognised by nucleotide level instability at simple repeat sequences. Specifically, expansions and contractions occur at simple repeat sequences such as mono-, di-, tri- and tetranucleotide repeats altering the number of these repeats. This phenomenon has been observed in the coding region of the TGF β RII and IGFR1 in sporadic colorectal cancers leading to mutant proteins (Section 1.3.4). All genes found to be differentially expressed between MIN and CIN sporadic colorectal cancer cell lines were analysed for the presence of simple repeat sequences. The route chosen to find these sequence motifs is outlined in Figure 4.11.

The initial GDA analysis expression profile yields accession numbers and Unigene identifier numbers for all of the clones arrayed on the GDA v1.3 filters. These Unigene unique identifiers may simply be entered in the Unigene database [<http://www.ncbi.nlm.nih.gov/UniGene>] and the mRNA sequences can be obtained directly. However, the Unigene database is dynamic and the Unigene identifiers are subject to change (often in the case where a sequence is found to be a constituent of a different gene). To overcome this, the GDA clone sequences (corresponding to the potentially differentially expressed transcripts) were entered into the GenBank database and the BLAST algorithm was used to determine which known sequence the particular clone exhibited similarity to. The gene names were then entered into the Unigene database and full length mRNA sequences were obtained.

Mononucleotide repeat sequences in the open reading frames of transcripts were deemed the most likely to be significant in the MIN phenotype (See Section 4.3). The presence of such sequence motifs was determined by visually scanning the coding regions of genes or by use of a simple text editor (Microsoft WordPad v1.0, USA). The presence of mononucleotide repeat sequences are outlined in Tables 4.5 and 4.6.

An alternative method for unveiling genes containing simple sequence repeats involved using the Expasy server [<http://www.expasy.ch>]. The FindPattern algorithm on this server searches through protein sequences for amino acid motifs of interest. The protein motif Lys-Lys-Lys was entered into the database. Lysine is coded for by the codons {aaa} or {aag}. This approach led to the identification of a number of genes containing coding region poly-A tracts. A similar approach with the Phe-Phe-Phe protein motif led to the identification of a modest number of genes harbouring poly-T

tracts. A similar approach was used to identify poly-G and poly-C tracts by searching the Expasy server for the protein sequences Gly-Gly-Gly and Pro-Pro-Pro respectively. The distribution of the different mononucleotide repeats in the SwissProt database as identified using FindPattern is illustrated in Figure 4.12.

4.3 DISCUSSION

Gene expression profiling techniques are constantly improving. Techniques such as Serial Analysis of Gene Expression (Velculescu *et al.*, 1995) and Suppression Subtractive Hybridisation (Diatchenko *et al.*, 1996) have been used to identify gene expression profiles of multiple genes in a number of cell systems (de Waard *et al.*, 1999; Kal *et al.*, 1999; Nupponen *et al.*, 1999). We have utilised the GDA v1.3 filters supplied by Genome Systems Inc. (GSI) to explore expression differences between the MIN and CIN sporadic colorectal cancer cell lines previously described in Chapter 2.

Initially, the GDA protocol was followed as described by Genome Systems Inc. which involves the isolation of poly-A(+) mRNA for probe synthesis. Total RNA was harvested and pooled followed by poly-A(+) mRNA purification on oligo-dT cellulose columns. The concentrations of experimental mRNA samples were balanced spectrophotometrically and visualised on 1.2% agarose denaturing formaldehyde gels. This protocol was found to be inefficient with poor yields of mRNA retrieved (as low as 20% of the expected yield). This approach was used for the GDA v1.3 test filters followed by the pairwise comparison between the cell lines LS174T (MIN) and SW480 (CIN) with satisfactory array data being obtained.

Due to the high amounts of starting material required (~1.5mg total RNA) and the inefficiency of the poly-A(+) purification, the GDA protocol was modified as follows. The SMART cDNA synthesis kit (Clontech, USA) was employed using 1µg of total RNA. The protocol takes advantage of the terminal transferase activity of reverse transcriptase, which adds a few nucleotides (primarily deoxycytidine) to the 3'-end of cDNA molecules. The SMARTTM oligonucleotide is designed with an oligo-G sequence at its 3'-end which anneals to the deoxycytidine stretch creating an extended template. The reverse transcriptase enzyme switches template and replicates to the end of the oligonucleotide. The first-strand synthesis full length cDNA is then amplified using a long PCR protocol yielding full length double stranded cDNA (dscDNA). The dscDNA samples were balanced spectrophotometrically, visualised on 1.0% agarose gels and labeled with [α -³³P]dCTP. This approach has two outstanding advantages over the original GDA protocol. Firstly, the protocol can be carried out on as little as 1µg of total RNA making this approach suitable for small tissue samples. In addition, the normalisation of dscDNA samples proved to be more convenient than the normalisation

of poly-A(+) mRNA which seemed to be operator dependent. This modified method for generating cDNA probes was used in the second and third pairwise comparisons of the cell lines SW48 (MIN) with respect to Caco-2 (CIN), and HCT116 (MIN) with respect to HT-29 (CIN) yielding satisfactory expression data.

Viewing the first gene expression profile between LS174T (MIN sporadic colorectal cancer cell line) and SW480 (CIN sporadic colorectal cancer cell line) it became clear that there were a significant number of expression differences [over a thousand cDNA clones showing significant differences; GSI suggest five-fold differences in gene expression as being significant]. It is likely that many of these differences represent tissue specific gene expression differences, which may not be associated with mismatch repair or sporadic colorectal cancer in general. The colorectal cancer cell lines selected for this study were not isolated from the same patients and may have been removed from different areas of the large intestine which would contribute to differences in gene expression unlikely to be related to mismatch repair or colorectal cancer. The experimental rationale involved three pairwise comparisons between total cDNA probes prepared from MIN and CIN colorectal cancer cell lines to identify a subset of consistently differentially expressed cDNA species to reduce the number of tissue specific differences. When the second and third pairwise comparison gene expression profiles were obtained, the expression profiles were searched for consistently differentially expressed genes in all three. This proved to be an exhaustive process with the consideration that the filters harbour over 18,000 cDNA species. A simple Perl script program (GeneMatch v1.0) was coded which allowed the investigator to compare the gene expression profiles in tab-delimited text format and obtain an output text file of the most significant differences in expression present in the profiles. With this approach, the thousand most differentially expressed cDNA species from each profile were considered. This significantly reduced the number of differentially expressed cDNA species: approximately 30-40 cDNA clones were found to be overexpressed in MIN cell lines with respect to CIN cell lines. Similar numbers of clones were found to be overexpressed in CIN cell lines with respect to MIN cell lines. Tables 4.1 and 4.2 summarise the differentially expressed genes. This subset of cDNA species is more likely to play a role in mismatch repair, chromosomal abnormalities and sporadic colorectal cancer.

There are a number of formats in which the investigator may view the expression data. A high percentage (up to 50%) of expressed sequence tags (ESTs) appear in the output using raw GDA data. Despite the progress of the Human Genome Project many genes remain uncharacterised. This was clearly illustrated in the raw data output. It may be possible to identify a portion of these ESTs in GenBank or by the construction of tentative human EST contigs using software like the ESTAssembler [this program aligns multiple ESTs and can be found at <http://hercules.tigem.it/cgi-bin/uniestass.pl>]. ESTs may be omitted from the gene expression profiles to allow identification of known genes. The profiles were examined with the omission of either ESTs or zero intensity signals or of both. Zero intensity signals may be considered as clones that yield no significant difference in hybridisation in comparison with the background signal. Currently, the significance of zero intensity signals is unclear. These may represent undetectable expression of a given transcript (the sensitivity limit of the filters is approximately five copies of a transcript per cell) or may represent a hybridisation artifact in some cases. Northern blot analysis of zero intensity clones will help to clarify these possibilities.

While initially exciting, data sets from high density cDNA arrays comparing two distinct colorectal tumour phenotypes become difficult to manage and immediately raise a number of questions. Is the validity and accuracy of the observed changes authentic? Which genes are to be selected for further analysis and which genes are a consequence rather than a cause of colorectal cancer? The legitimacy of the changes observed in a high density cDNA screening may be gauged by an independent gene expression technique (RT-PCR or Northern blotting; See Chapter 5). A convenient method of focusing on the genes likely to be associated with mismatch repair or chromosomal instability was to carry out three pairwise comparisons of MIN and CIN sporadic colorectal cancer cell lines. However, it is also useful to gain a general scrutiny of gene expression within the entire complement of clones on the high density cDNA array. Hybridisation intensity values for the entire set of cDNA clones (~18,000) were extracted from the raw GDA data output using a Perl script. The intensities for each MIN/CIN comparison were plotted.

Knowledge of a panel of differentially expressed genes between the MIN and CIN phenotypes in colorectal cancer will only be useful if put into specific biological

contexts. We expected to observe differential expression of genes associated with mismatch repair and chromosomal instability and a considerable number of genes previously unassociated with colorectal cancer. A clear idea of the biological functions of these genes will allow the investigator to design directed downstream experiments. We used a bioinformatic approach to identify the genes from our GDA experiments. The GenBank, Unigene and OMIM databases are interconnected at the National Centre for Biotechnology Information server. Using these resources, we were able to classify (in general terms) most of the panel of differentially expressed genes. The broad categories were based on the categories established at The Institute for Genomic Research (TIGR). Initially, there seemed to be a high proportion of receptors and transcription factors in the gene set. An obvious link between this phenomenon and colorectal tumor progression is not easily apparent. However, on inspection there are a number of genes involved with cell cycle progression, cell cycle checkpoint and genes involved in chromatin structure. There is no clear pattern of expression of these transcripts; they are overexpressed in both patterns (MIN with respect to CIN and CIN with respect to MIN). The MIN phenotype has been associated with hypermethylation of the mismatch repair genes and their promoters. It is conceivable that decreased or increased expression of acetyltransferases and histone deacetylases, which are integral components of chromatin maintenance, might influence the methylation state of genes in MIN tumours. In contrast, CIN tumours exhibit gross chromosomal abnormalities including loss of heterozygosity (LOH), translocations and aneuploidy. Aberrant cell cycle machinery has been associated with these chromosomal abnormalities. Differential activation or repression of cell cycle related genes may clarify the genotypic differences seen between MIN and CIN tumours.

High density cDNA array based methodologies are subject to error due to poorly quantitated mRNA, preferential labelling of different samples or uneven hybridisation of probes to filters. As a consequence, it is important to assess how significant the observed differences in expression are. Semi-quantitative RT-PCR and Northern blotting are the techniques of choice to confirm differential gene expression. Prior to using these methodologies, we took advantage of the public access "Gene to Tag Mapping" SAGE database. A complete SAGE analysis (approximately 60,000 tags were sequenced from each cell line) has been performed previously on the sporadic colorectal cancer cell lines HCT116 (MIN) and Caco-2 (CIN) (Lal *et al.*, 1999). The

SAGE database was screened for tag abundances from each gene found from the GDA analyses and the results are detailed in Tables 4.3 and 4.4. Most of the SAGE expression levels correspond with the differences observed in the GDA analyses. A modest number of genes disagree with the GDA analyses (overexpressed instead of underexpressed or *vice versa*). Genes expressed at low levels may be difficult to quantitate by either of the two techniques. However, it is likely that PCR-based techniques, such as SAGE, are more suitable for detecting low-level expression of genes. SAGE data were also useful for assessing the likely level of expression of genes. This was used as a guide to decide whether to use northern blotting (suitable for the detection of highly expressed transcripts) or semi-quantitative RT-PCR (suitable for detection of lowly expressed genes) as techniques for confirming the differential expression of selected genes from the GDA analyses.

Also relevant to this study was the possibility of detecting differentially expressed genes with instability at a mononucleotide repeat sequence. As discussed in Section 1.3.4, many mismatch repair deficient colorectal cancers harbour frameshift mutations in the TGF β RII and IGFRII genes diminishing the expression of these transcripts. This phenomenon is seen as one of the driving forces behind tumourigenesis in MIN colorectal cancers. Two methods were used to identify the presence of simple sequence repeats in the panels of differentially expressed genes (Tables 4.1 & 4.2). Multiple full mRNA sequences were extracted from the Unigene database consistent with the panel of genes obtained from the GDA analyses. The mRNA sequences were thoroughly scrutinised to identify 5' untranslated regions (UTR), 3'UTR, open reading frames (ORF), exons and introns. Published reports on instability at simple repeat sequences suggest that 8-10bp polyA tracts are the most likely target of expansion or contraction due to a compromised mismatch repair machinery in MIN tumours (Section 1.3.4). The size of repeat sequences found was usually 6-7 nucleotides in length. Repeats of this short length have not been reported to be sites of instability. A number of 5'UTR and 3'UTR repeat sequences were found but are unlikely to have as significant a biological affect as coding repeat sequences. A 9bp polyA tract was found in the coding region of the protein tyrosine phosphatase type D1 (PTPD1) gene. This was analysed for instability by PCR (see Chapter 5) in conjunction with a PCR spanning the well characterised 9bp polyA tract in the transforming growth

factor beta receptor II (TGF β RII) gene. One problem noticed was the quality of particular mRNA sequences submitted to Unigene. Coding region polynucleotide repeat sequences were often abandoned as the initial sequence was inconsistent with mRNA sequences of the same transcript from other sources. This problem can be overcome by analysis of multiple ESTs derived from a given mRNA transcript to analyse sequencing discrepancies more closely.

The second method to find coding region mononucleotide repeat sequences was a 'shotgun' approach based on the Expasy FindPattern algorithm. Simple protein motifs such as Lys-Lys-Lys were entered into the FindPattern server. This yielded hundreds of proteins whose mRNA counterparts were analysed for the presence of a 9bp polyA tract. This approach led to many proteins with aag codons instead of the desired aaa codons. However, a significant number of genes with polyA tracts were found in this manner. The same approach was taken with the protein motifs potentially resulting from mononucleotide repeats composed of polyT, polyC or polyG tracts. The distribution of the presence of these simple repeat sequences is detailed in Figure 4.11. The presence of these genes on the GDA v1.3 filters was determined using the clone search facility at the GDA server [<http://www.genomesystems.com/clone.html>]. None of the genes present on the filters exhibited consistent differential expression between MIN cells and CIN cells.

In addition to considering the greatest expression differences we have also adopted a more directed approach. The raw data profiles have been routinely searched for candidate genes with a possible role in mismatch repair and sporadic colorectal cancer. One example of this method led to the identification of the Mad-2 gene. This gene was seen to be undetectable in the CIN cell line SW480 and underexpressed in the CIN cell line Caco-2 with respect to the MIN cell lines analysed. However, Mad-2 is expressed at equivalent levels in HCT116 and HT29 (third MIN/CIN GDA comparison). Therefore, this transcript did not appear in the panels of genes shown in Tables 4.1 & 4.2. Mad-2 is a known mitotic spindle assembly checkpoint protein (Cahill *et al.*, 1999). Mad-2 stabilises the anaphase promoting complex (APC) which can check the cell cycle at mitosis, allowing for the correct alignment of chromosomes at the kinetochores. Mouse models deficient in Mad-2 expression are subject to gross chromosomal instabilities such as polyploidy and loss of heterozygosity. It is possible

that underexpression of Mad-2 in some CIN cell lines is a contributory factor to the type of instabilities seen in these lines. A number of additional genes associated with the mitotic checkpoint have also been seen to be underexpressed in the CIN cell lines analysed and are detailed in Tables 4.1 and 4.2.

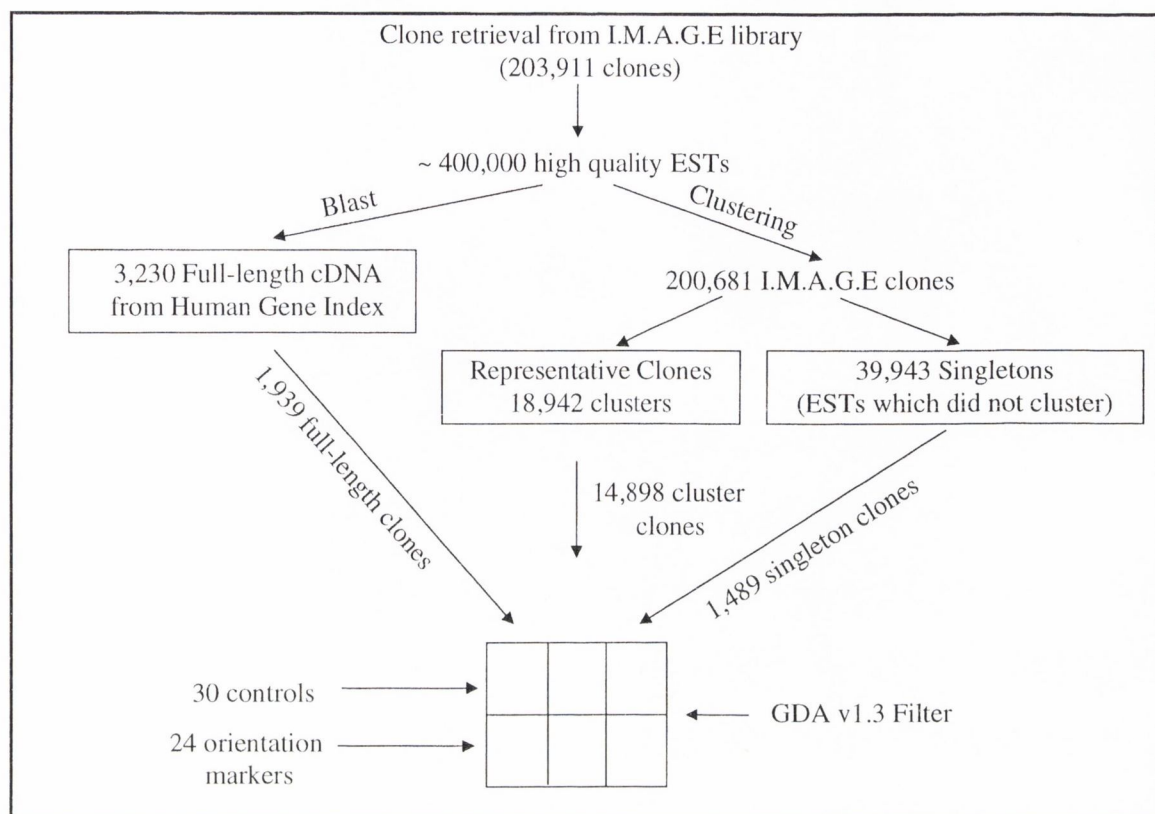


Figure 4.1 Gene Discovery Array (GDA) clone library. Over 200,000 clones (corresponding to 400,000 high quality ESTs (Expressed Sequence Tag) were obtained from the IMAGE consortium database (Integrated Molecular Analysis of Genomes and their Expression). BLAST (Basic Local Alignment Sequence Tool) was used to categorise 3,200 of these clones as previously characterised full length genes. The remainder of the clones were categorised as overlapping EST clusters or as singleton ESTs with no overlap counterparts. Clones were microdotted onto nylon GDA filters in duplicate including 30 controls and 24 orientation markers.

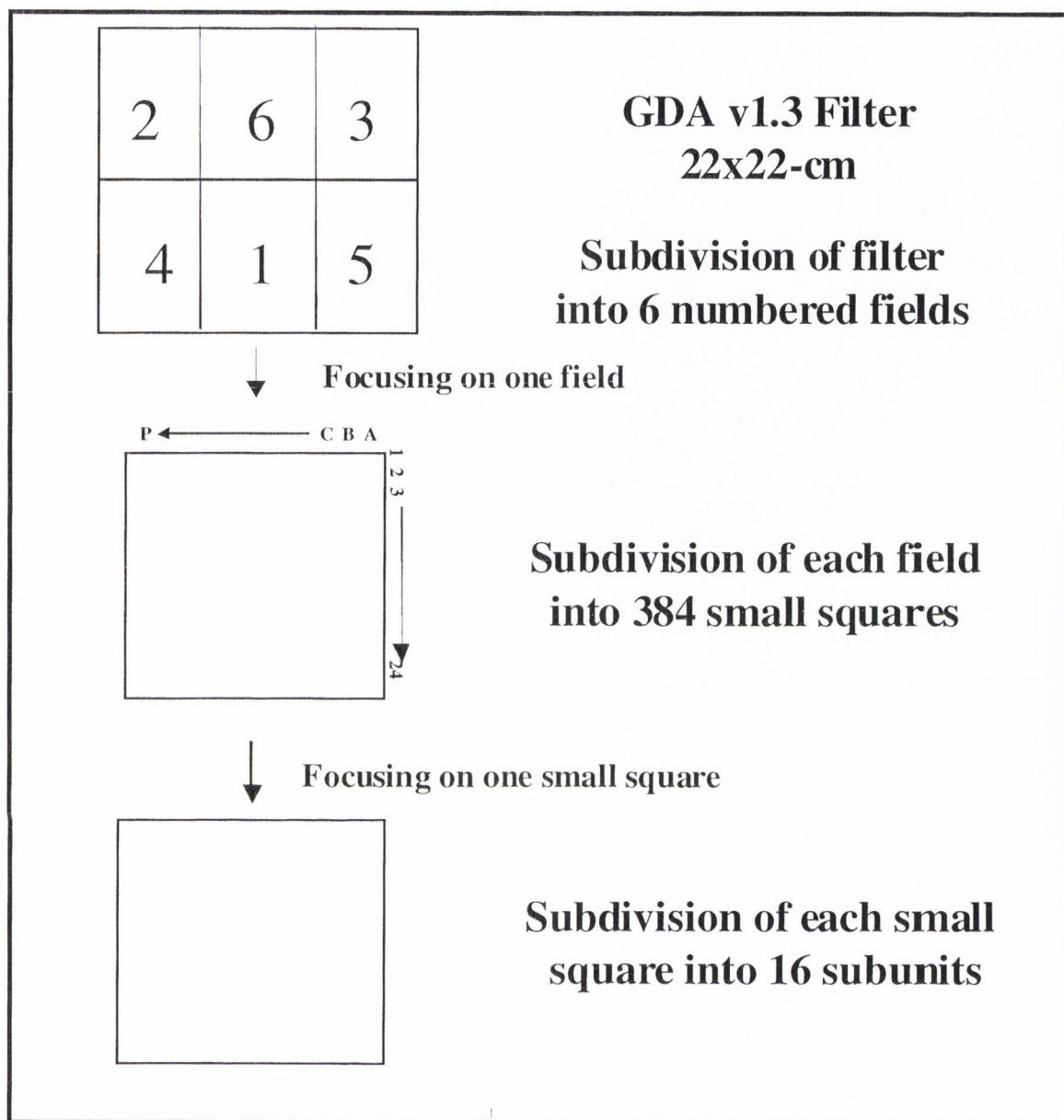


Figure 4.2 Structure of GDA V1.3 filters. An illustration of the hierarchical organisation of the GDA V1.3 filter. Each filter is divided into six fields, designated 1-6. Each field is divided into 384 small squares, divided A to P horizontally and 1 to 24 vertically. Each small square is subdivided into sixteen subunits each harbouring a GDA clone. A detailed description of the GDA V1.3 filter structure may be found at <http://www.genomesystems.com/GDA/geneID.html>.

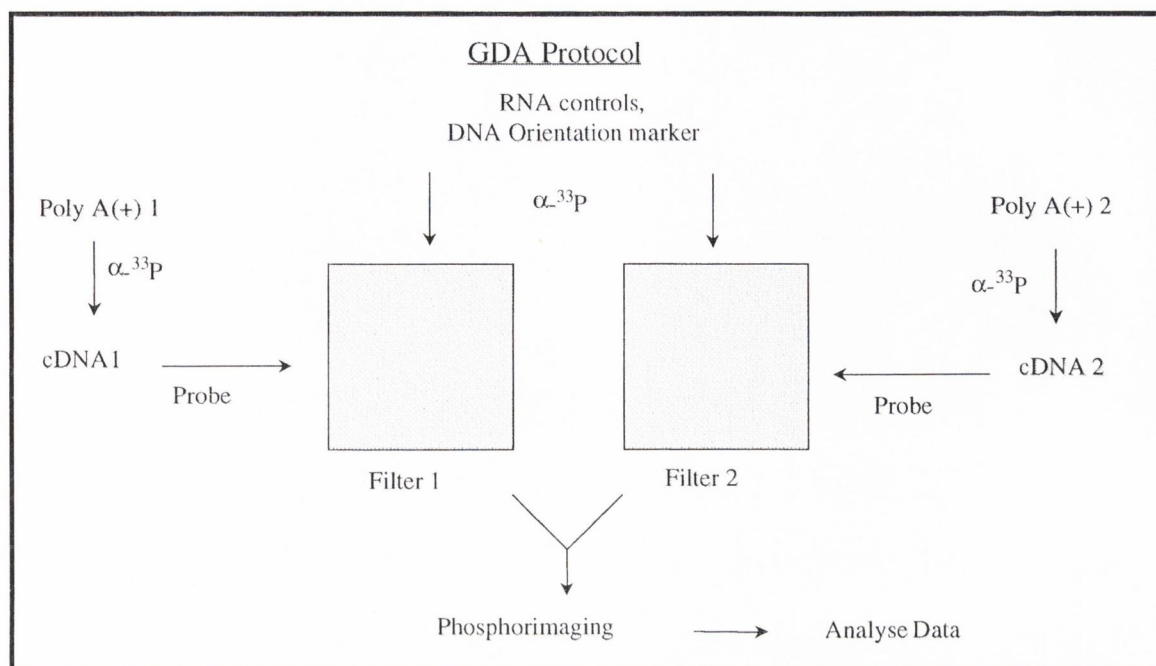


Figure 4.3 A schematic representation of the GDA protocol. Sample poly-A(+) mRNA and the RNA controls are reverse transcribed in the presence of $[\alpha\text{-}^{33}\text{P}]\text{dCTP}$. DNA orientation markers are labelled by random-priming with $[\alpha\text{-}^{33}\text{P}]\text{dCTP}$. Duplicate filters are probed with the combination of cDNA 1, control cDNA and orientation marker or the combination of cDNA 2, control cDNA and orientation marker. Images of the filters are obtained by phosphorimaging and the image files are analysed using the Gene Discovery Array software.

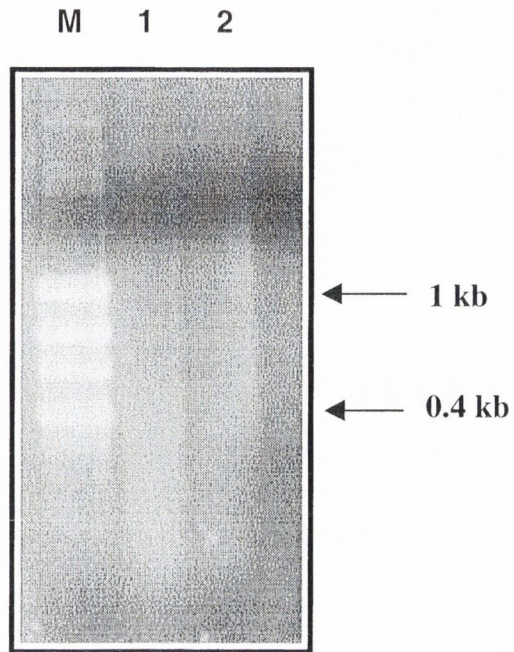


Figure 4.4 Analysis of poly-(A)+ mRNA integrity. An illustration of electrophoresis of poly-(A)+ mRNA on a denaturing formaldehyde 1.0% agarose gel stained with ethidium bromide. The faint smear ranging from below 0.4kb to greater than 1kb is consistent with the expected molecular weight of poly-A(+) mRNA. M=pBS-*Bgl* II/*A*/*w*44 I molecular weight marker; 1= double selected poly-(A)+ mRNA from LS174T sporadic colorectal cancer cells; 2= double selected poly-(A)+ mRNA from SW480 sporadic colorectal cancer cells.

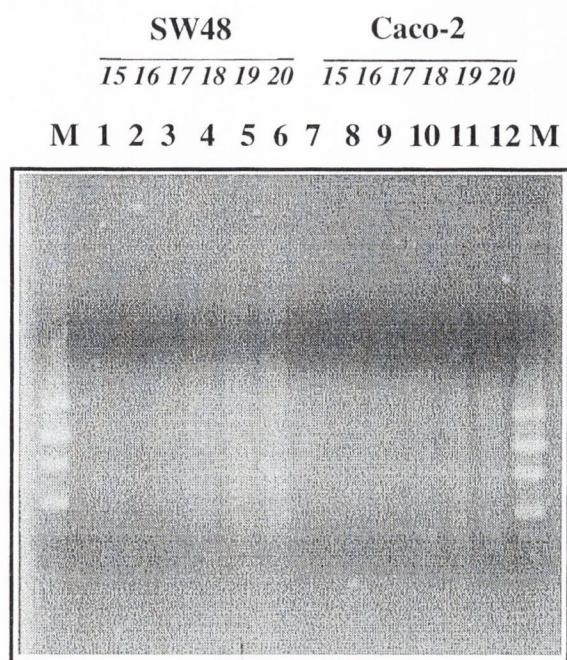


Figure 4.5 Analysis of SMARTTM cDNA PCR amplification. SMARTTM cDNA PCR products were separated on 1.5% agarose gels and stained with ethidium bromide. The PCR cycle numbers carried out are indicated above each well in italics. M=pBS-*Bgl* II/*Alw*44 II molecular weight marker; 1, 2, 3, 4, 5 & 6= SW48; 7, 8, 9, 10, 11, 12= Caco-2.

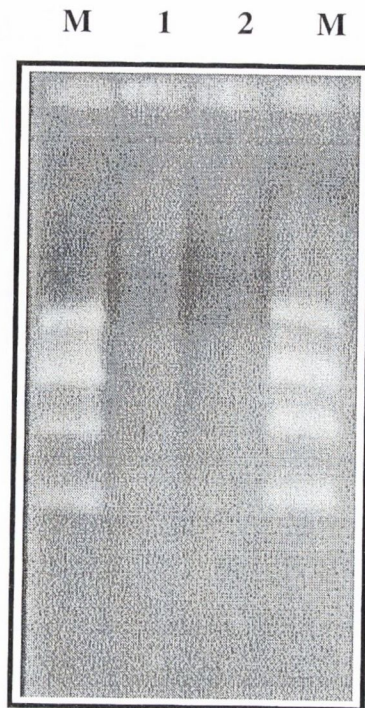


Figure 4.6 Analysis of ammonium acetate precipitated SMARTTM cDNA. Aliquots (0.5 μ L) of SMARTTM cDNA PCR products were separated on 1.5% agarose gels and stained with ethidium bromide. M= pBS-*Bgl* II/ *Alw*44 I molecular weight marker; 1= pooled SW48 SMARTTM cDNA; 2= pooled Caco-2 SMARTTM cDNA.

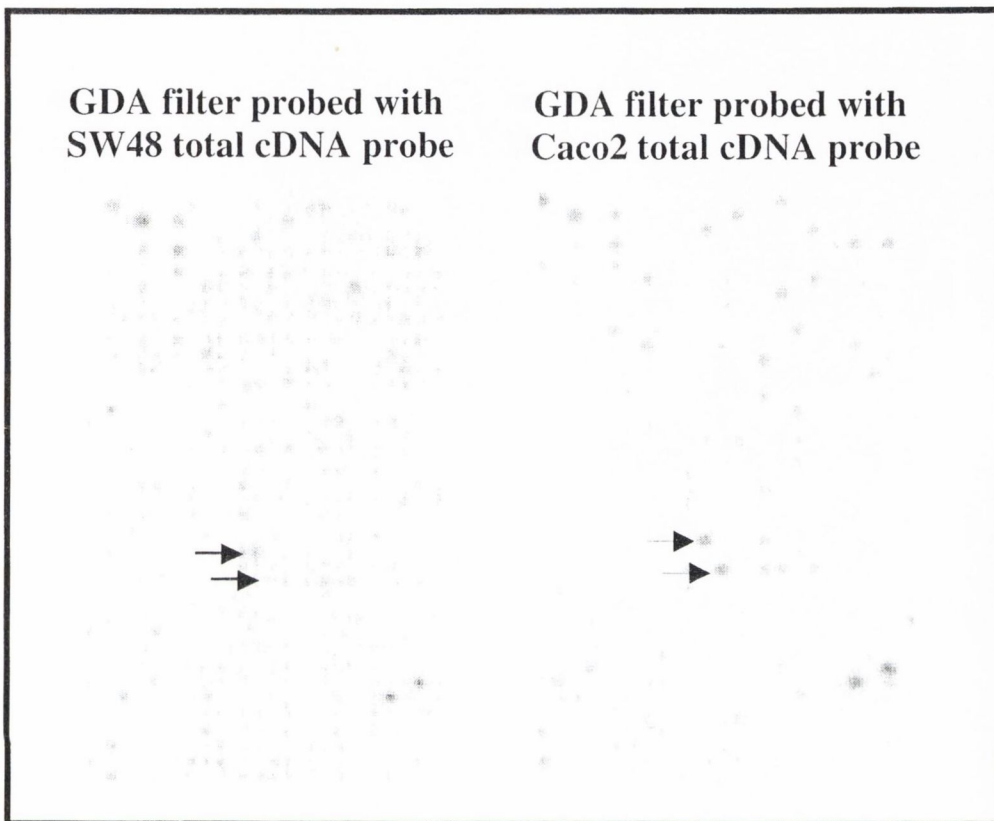


Figure 4.7 Illustration of a differentially expressed clone on GDA v1.3 filters. The arrows denote the location of the clone corresponding to the human solute carrier family 24 gene. The clone is clearly more highly expressed in the Caco-2 cDNA probe in relation to the SW48 cDNA probe.

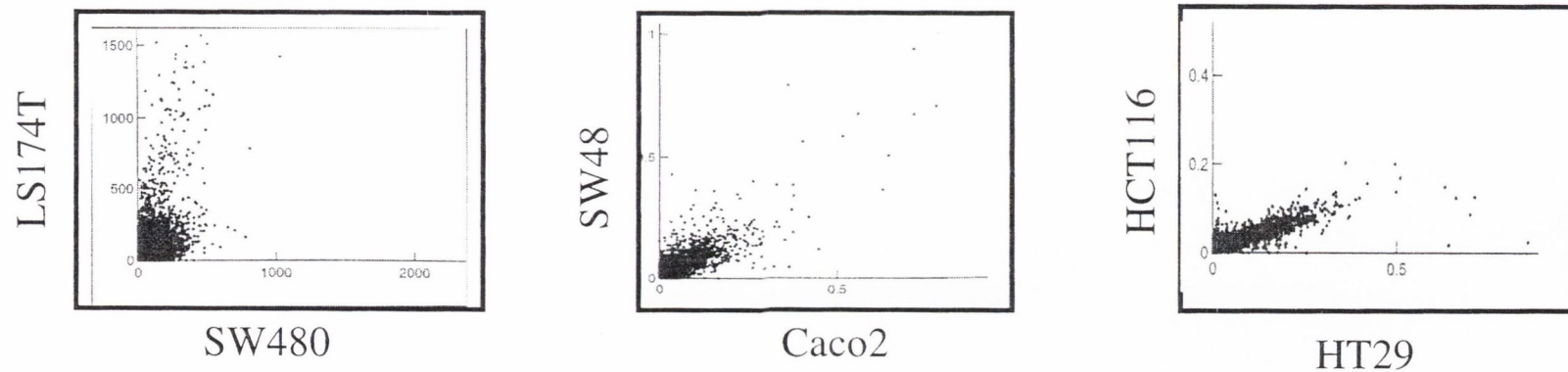


Figure 4.8 Scatter plots for the three MIN/CIN GDA analyses. An illustration of the scatter plots obtained from GDA analysis of three MIN/CIN pairwise comparisons. These were plotted using MatLab v1.0 on a PC workstation. The expression level of each cDNA clone was plotted between the cell lines, i.e. the expression values determined by GDA analysis for each of the 18,000 cDNA clones were plotted between LS174T/SW480, SW48/Caco2 and HCT116/HT29. These profiles give an overall view of the expression of cDNA species between the cell lines examined.

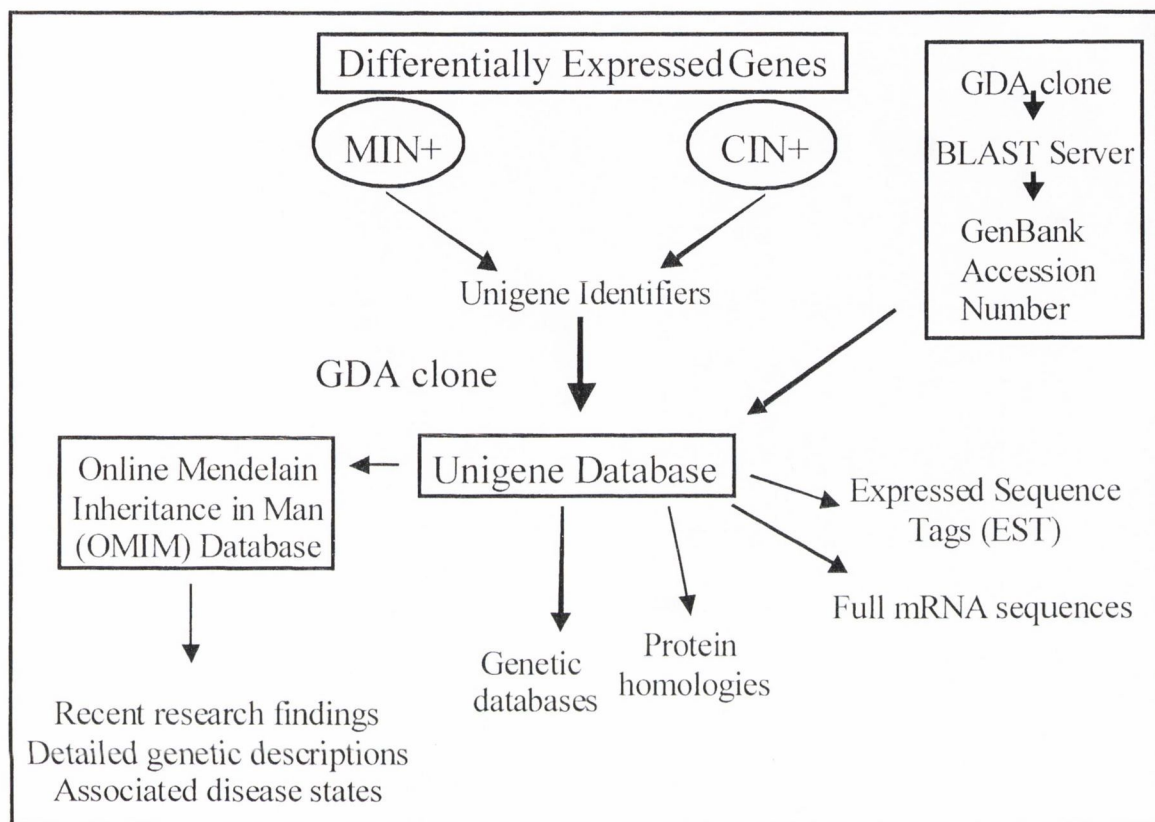


Figure 4.9 Categorisation of the differentially expressed mini-database of genes. The Gene Discovery Array (GDA) database details the relevant unigene identifier with each GDA clone. This identifier is entered in the Unigene database and a descriptive catalogue of information on the gene of interest is obtained. A link to Online Mendelian Inheritance in Man (OMIM) can be followed which yields current research findings, genetic descriptions, tissue specific expression data and associated disease states if appropriate. Differentially expressed GDA clones were entered in the BLAST server [<http://www.ncbi.nlm.nih.gov/BLAST>] to confirm the identity of clones from the GDA library [See inset section].

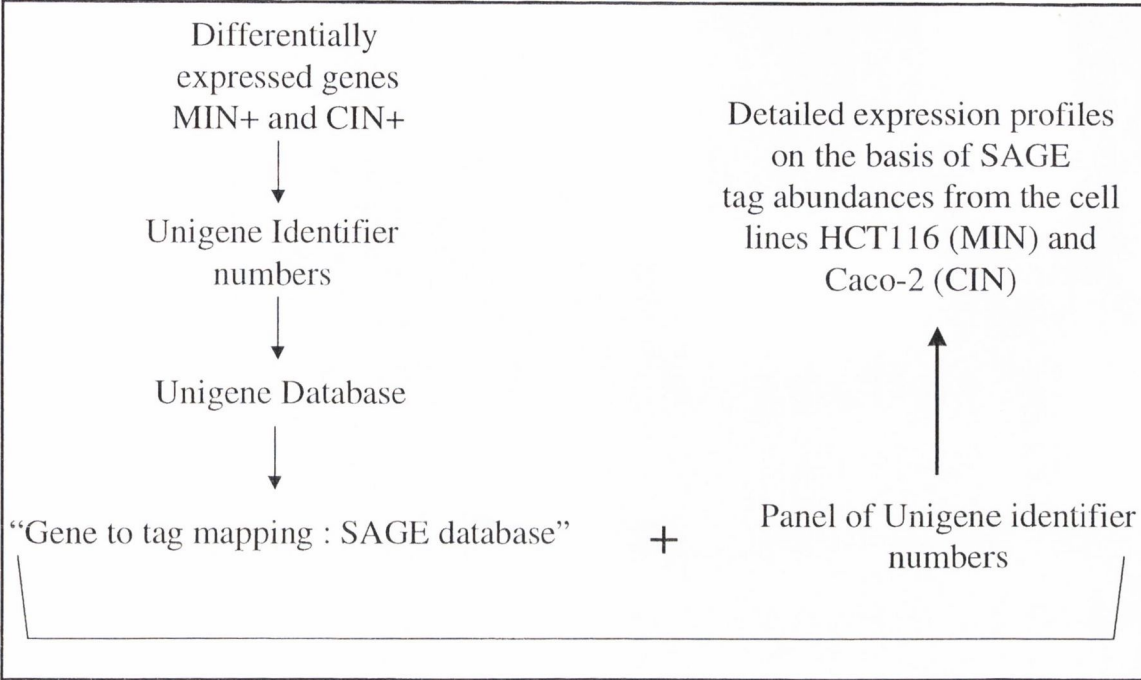


Figure 4.10 Correlation of GDA expression data with SAGE data from the cell lines HCT116 (MIN) and Caco-2 (CIN). An illustration of the procedure used to correlate the expression differences seen using GDA v1.3 filters and the public SAGE database [<http://www.ncbi.nlm.nih.gov/SAGE>]. Each gene from the GDA analyses (MIN+ and CIN+) was entered in the Unigene database. A link to the SAGE database is provided which yields detailed expression information of the genes for the sporadic colorectal cancer cell lines HCT116 (MIN) and Caco-2 (CIN).

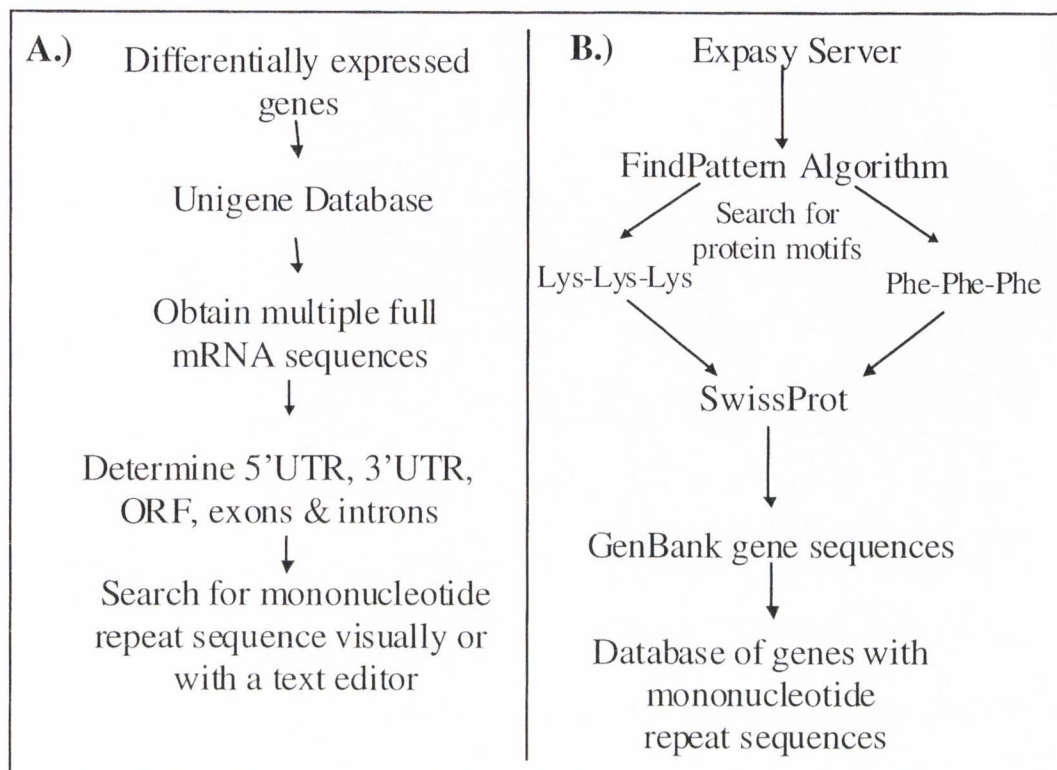


Figure 4.11 Schematic representation of procedures employed to search for coding region simple repeat sequences. **A.)** Schematic of search for mononucleotide repeat sequences in the database of genes differentially expressed between MIN and CIN sporadic colorectal cancer cell lines. Full length mRNA sequences for each gene were obtained from the Unigene database. These were analysed for the presence of simple repeat sequences visually or by using a text editor (MicroSoft WordPad v1.0, USA). **B.)** Schematic of search for mononucleotide repeat sequences in public database proteins. The FindPattern algorithm on the Expasy server [<http://www.expasy.ch>] was used to search for the simple protein motifs Lys-Lys-Lys and Phe-Phe-Phe. The codons aaa and aag code for the amino acid lysine, and the codons ttg and ttc code for phenylalanine. Searching through proteins with these motifs yielded a modest number of genes with either coding polyA tracts or coding polyT tracts.

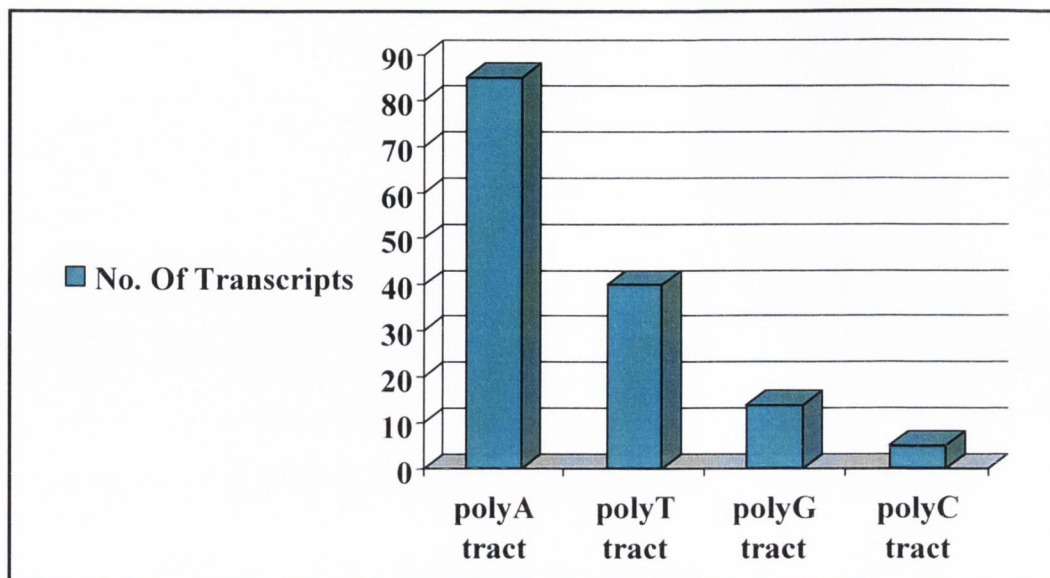


Figure 4.12 Bar chart of the distribution of coding region mononucleotide repeat sequences in man. Protein sequences containing simple protein motifs comparable to potential mononucleotide repeats were identified using the FindPattern algorithm at the Expasy server [<http://www.expasy.ch>]. The protein sequences were extracted from the SwissProt database and the parent mRNA transcripts were extracted from the Unigene server. The mRNA sequences were analysed visually or by using a text editor (Microsoft WordPad v1.0, USA) for the presence of mononucleotide repeat sequences. The vertical axis represents the number of sequences containing a mononucleotide repeat and the horizontal axis describes the nature of the repeat sequence. These numbers are based on the analysis of 500, 500, 300 and 300 potential polyA tract, polyT tract, polyC tract and polyG tract harbouring proteins respectively [For example, 500 proteins with the KKK amino acid motif were searched yielding 90 polyA tracts. The remaining proteins harbour polyA tracts interspersed with T nucleotides. Similar results were found with polyC, PolyG and polyT tract searches.]

Table 4.1 MIN genotype gene expression profile. A summary of the genes showing consistently elevated expression in the microsatellite instability (MIN) sporadic colorectal cancer cell lines LS174T, SW48 & HCT116 with respect to the microsatellite stable (CIN) cell lines Caco2, SW480 & HT29. Each gene is named with the inclusion of a GenBank accession number. Tentative gene functions were assigned to each gene where possible using the approach described in Figure 4.9. Genes highlighted in bold text were selected for further expression analysis by semi-quantitative RT-PCR analysis or Northern blot analysis.

Gene Transcript	Accession Number	Tentative Gene Function
Hemoglobin, alpha 1	W04832	Oxygen transport
Heat shock 27kD protein 1	W69499	Activates apoptosis / caspases
Cell division cycle 2-like 1	N98592	Cell division
H2AZ histone	W45695	Nucleosome assembly
Spectrin, beta, non-erythrocytic 1	N70585	Cytoskeletal component of red blood cells
Metallothionein-Ie	H52525	Heavy metal transport
Transcription factor AP-2	N63770	Transcription factor / gene expression
Isovaleryl Coenzyme A	H37754	Metabolism
ETS-related transcription factor ELF-1	N74959	Transcription factor / gene expression
Cathepsin S precursor	N53585	Cysteine protease; protein degradation / processing
Collagen, type V, alpha 1	W52829	Structural protein; bone architecture
Basic transcription factor 3	AA203284	Transcription factor / gene expression
CTP synthetase	H11731	Pyrimidine metabolism
Retinol-binding protein 3	H64488	Involved in Vit A regulation in eye cells
Carbonic anhydrase I	R46266	Hydration of CO ₂
40S ribosomal protein S19	AA046433	Monocyte chemotactic factor
DNA primase	R43686	DNA polymerase / DNA replication
Kinase insert domain receptor	R35713	VEGF-A tyrosine kinase receptor
Dual specificity phosphatase 5	R43931	Hydrolyses substrates phosphorylated on both tyrosine and serine/threonine residues
Integrin, alpha L	T83460	Expressed on lymphocytes and phagocytic cells; involved in T-cell adhesion to target cells
HLA class I histocompatibility	R74335	Cell signalling / cell communication
ADP-ribosylation factor 4	N52015	Purine metabolism
Microfibrillar protein	R01211	Cytoskeletal organisation
Alu RNA transcript	W31678	Unclassified
Trans-1,2-dihydrobenzene-1,2-diol dehydrogenase	W63727	General aldehyde and ketone oxidation
Ribonuclease, RNase A family, 1 pancreatic	R62231	RNA metabolism; encourages thyroid tumor cells to apoptose
Elastin	T79116	Cytoskeletal

Table 4.1 MIN genotype gene expression profile (continued)

Gene Transcript	Accession Number	Tentative Gene Function
Prepro form of corticotropin	R33777	Cell signalling / communication

Table 4.2 CIN genotype gene expression profile. A summary of the genes showing consistently elevated expression in the microsatellite stable (CIN) sporadic colorectal cancer cell lines Caco2, SW480 & HT29 with respect to the microsatellite stable (CIN) cell lines LS174T, SW48 & HCT116. Each gene is named with the inclusion of a GenBank accession number. Tentative gene functions were assigned to each gene where possible using the approach described in Figure 4.9. Genes highlighted in bold text were selected for further expression analysis by semi-quantitative RT-PCR analysis or Northern blot analysis.

Gene Transcript	Accession Number	Tentative Gene Function
Guanosine 5'-monophosphate synthase	R94457	Purine Metabolism
Creatine Kinase B	AA069373	Cell signalling/ cell communication
Phosphatidic Acid Phosphatase Type 2A	R59009	Cell signalling involving lipid mediators and androgens
Pregnancy-specific beta 1-glycoprotein 7	R80286	Foetal development; embryonic growth and development
TBP-associated factor TAFII80	R44112	Gene expression; general transcription factor
Receptor protein-tyrosine kinase (EDDR1)	R56392	Cell signalling / cell communication ; promotor has p53 binding site and is upregulated by p53 in osteosarcoma cells
Ribosomal protein S6 kinase	AA039599	Growth factor regulated serine / threonine kinase
Protein tyrosine phosphatase D1	R23593	Gene / protein expression
Coagulation factor XI (plasma thromboplastin antecedent)	N52835	Cell / organism defense
Guanine nucleotide binding protein	H61281	Cell signalling / G protein intracellular messenger
Complement component 4-binding protein	H53489	Cell / organism defense
Cholinesterase-related cell division controller	R83723	Cell division / regulates passage through mitosis
Granulocyte-macrophage colony-stimulating factor receptor	N99581	Cytokine receptor
Calcium channel L-type alpha 1 subunit	AA136049	Cell signalling / cell communication
Iduronate 2-sulfatase	N26477	Hunter Syndrome; prevents deposition of mucopolysaccharides in lysosomes
Fibronectin 1	N90769	Cell adhesion / morphology
Chronic granulomatous disease (b-245 beta polypeptide)	AA149281	Cellular respiration in killing bacteria in phagocytic vacuoles
HMGCo-A reductase	N93874	Cholesterol biosynthesis; controls primordial germ cell migration

Table 4.2 CIN genotype gene expression profile (continued).

Gene Transcript	Accession	Tentative Gene Function
RYK receptor-like tyrosine kinase	W32536	Unlike classical RTKs; activates MAPK pathway
Ran binding protein 2	R82691	Nuclear pore complex activator; involved in mitosis
Collagen, type XVIII, alpha 1	AA156812	Structural peptide; contains motif (endostatin) capable of inhibiting proliferation, angiogenesis and tumour growth
Electron-transfer-flavoprotein, alpha polypeptide	W85863	Mitochondrial respiration
Monokine induced by α -interferon	AA1131406	Chemotactic cytokine
Secretogranin II precursor	R41603	Neuropeptide precursor of secretogranin
Amylo-1,6-glucosidase, 4-alpha-glucanotransferase	AA017181	Glycogen debranching enzyme; glycogen metabolism
Small inducible cytokine A2	AA024754	Involved in monocyte recruitment in inflammatory response
Glucose-6-phosphate dehydrogenase	AA017135	Glucose metabolism
Phospholipase C, gamma 2	H48246	Phospholipid metabolism; controlled by receptor tyrosine kinases
HE4	AA025750	Epididymal specific protein with possible role in sperm maturation
Glucosidase II	W47502	Glycoprotein metabolism
Angiodema inhibitor	AA056092	Possible role in reproduction
Alpha crystallin A chain	H82655	Chaperone protein in lens
Polo(Drosophila)like kinase	R99810	Cell cycle regulator; localises to mitotic spindle
Glycogenin	N35332	Glycogen biosynthesis
Gem GTPase	AA129719	Ras-related GTPase; receptor-mediated signal transduction
ATPase, Cu ⁺⁺ transporting, beta	R41429	Copper metabolism
Estrogen sulfotransferase	H79311	Involved in the sulfate conjugation of monoamine neurotransmitters
Rearranged immunoglobulin lambda	T71328	Cell / organism defense ; immunoresponsive
Prostaglandin E receptor 2	N28920	Prostaglandins biosynthesis pathway Mediates signalling of prostaglandins to hepatocyte growth factor
Cyclin A2	AA001916	Cell cycle regulator; DNA replication
Phosphoribosyl pyrophosphate synthetase subunit I	H14586	Purine and pyrimidine biosynthesis
DNAX activation protein	W92376	Protein tyrosine kinase binding protein
Insulin-like growth factor 2	R32132	Autocrine growth factors associated with colorectal cancer; stimulates proliferation

Table 4.2 CIN genotype gene expression profile (continued).

Gene Transcript	Acession Number	Tentative Gene Function
Solute carrier family 2	N62266	Glut5 transport fructose from lumen of small intestine

Table 4.3 SAGE data and EST abundances for genes showing elevated expression in MIN sporadic colorectal cancer cell lines. Tabulated below are the genes differentially expressed between the MIN and CIN sporadic colorectal cancer cell lines. Data were extracted from both the SAGE and Unigene databases to give supplementary expression data for these candidate genes using the process outlined in Figure 4.10. Columns 1 and 2 list each transcript name and the Unigene identifier number. Columns 3 and 4 illustrate the differences in tag abundance for each given gene in the SAGE database. A measure of the general expression of these genes is indicated by the number of ESTs derived from each gene reported in the Unigene database and is listed in Column 5. The transcripts highlighted in bold were selected for further expression analysis.

Gene Transcript	Unigene Identifier	SAGE Tags per million in HCT116	SAGE Tags per million in Caco2	EST abundance
Hemoglobin, alpha 1	Hs.75792	-	-	3789
Heat shock protein 1, 27kD	Hs.76067	-	-	1609
Cell division cycle-2 like 1	Hs.183418	645	48	534
H2AZ histone	Hs.119192	530	407	280
Spectrin, beta, non-erythrocytic 1	Hs.107164	-	-	348
Metallothionein-Ie	Hs.74170	82	0	137
Transcription factor AP-2	Hs.18387	-	-	138
Isovaleryl Coenzyme A	Hs.77510	363	264	85
ETS-related transcription factor ELF-1	Hs.154365	-	-	116
Cathepsin S precursor	Hs.181301	64	72	126
Collagen, type V, alpha 1	Hs.146428	0	32	350
BTF3	Hs.101025	629	505	1190
CTP synthetase	Hs.253653	49	32	160
Retinol-binding protein 3	Hs.857	-	-	14
Carbonic anhydrase I	Hs.23118	-	-	25
40S ribosomal protein S19	Hs.126701	-	-	2000
DNA primase	Hs.74519	0	48	32
Kinase insert domain receptor	Hs.12337	-	-	76
Dual specificity phosphatase 5	Hs.1218	49	16	62
Integrin, alpha L	Hs.174103	-	-	17
HLA class I histocompatibility	Hs.639	447	440	639
ADP-ribosylation factor 4	Hs.75290	66	97	312
Microfibrillar protein	Hs.61418	0	16	40
Alu RNA transcript	-	-	-	-

Table 4.3 SAGE data and EST abundances for genes showing elevated expression in MIN sporadic colorectal cancer cell lines (continued).

Gene Transcript	Unigene Identifier	SAGE Tags per million in HCT116	SAGE Tags per million in Caco2	EST abundance
Trans-1,2- dihydrobenzene- 1,2-diol dehydrogenase		0	32	148
Ribonuclease, RNase A family, 1 (pancreatic)	Hs.78224	-	-	143
Elastin	Hs.203197	-	-	10
Prepro form of corticotropin releasing factor	Hs.75294	-	-	27

Table 4.4 SAGE data and EST abundances for genes overexpressed in CIN cell lines. Tabulated above are the genes differentially expressed between the CIN and MIN sporadic colorectal cancer cell lines. Data were extracted from both the SAGE and Unigene databases to give supplementary expression data for these candidate genes using the process outlined in Figure 4.10. Columns 1 and 2 list each transcript name and the Unigene identifier number. Columns 3 and 4 illustrate the differences in tag abundance for each given gene in the SAGE database maintained by Bert Vogelstein. A measure of the general expression of these genes is indicated by the number of ESTs derived from each gene reported in the Unigene database and is listed in Column 5. The transcripts highlighted in bold were selected for further expression analysis.

Gene Transcript	Unigene Identifier	SAGE Tags per million in HCT116	SAGE Tags per million in Caco2	EST abundance
Guanosine 5'- monophosphate synthase	Hs.5398	49	16	37
Creatine Kinase B	Hs.173724	182	505	285
Phosphatidic Acid Phosphatase Type 2A	Hs.41569	0	16	170
Pregnancy-specific beta 1-glycoprotein 7	Hs.173609	0	16	196
TBP-associated factor TAFII80	Hs.78865	33	16	43
Receptor protein- tyrosine kinase (EDDR1)	Hs.75562	265	114	120
Ribosomal protein S6 kinase	Hs.149957	0	32	57
Protein tyrosine phosphatase D1	Hs.155693	0	32	45
Coagulation factor XI (plasma thromboplastin antecedent)	Hs.1430	0	16	10
Guanine nucleotide binding protein	Hs.233771	82	32	56
Complement component 4- binding protein	Hs.1012	0	16	39
Cholinesterase- related cell division controller	Hs.155266	16	0	34
Granulocyte- macrophage colony-stimulating factor	Hs.182378	7900	7300	9
Calcium channel L- type alpha 1 subunit	Hs.89925	-	-	13
Iduronate 2- sulfatase	Hs.172458	0	16	279
Fibronectin 1	Hs.118162	0	717	1433

Table 4.4 SAGE data and EST abundances for genes showing elevated expression in CIN sporadic colorectal cancer cell lines (continued).

Gene Transcript	Unigene Identifier	SAGE Tags HCT116	SAGE Tags Caco2	EST Abundance
Chronic granulomatous disease (b-245 beta polypeptide)	Hs.88974	16	8	40
HMGC _o -A reductase	Hs.11899	0	65	156
RYK receptor-like tyrosine kinase	Hs.79350	0	32	94
Ran binding protein 2	Hs.179825	16	32	194
Collagen, type XVIII, alpha 1	Hs.78409	49	163	202
Electron-transfer-flavoprotein, alpha polypeptide	Hs.169919	49	65	101
Monokine induced by α -interferon	Hs.77367	-	-	54
Secretogranin II precursor	Hs.75426	16	0	46
Amylo-1,6-glucosidase, 4-alpha-glucanotransferase	Hs.904	49	16	52
Small inducible cytokine A2	Hs.340	-	-	90
Glucose-6-phosphate dehydrogenase	Hs.80206	49	114	208
Phospholipase C, gamma 2	Hs.75648	0	32	37
HE4	Hs.2719	0	32	99
Glucosidase II	Hs.76847	645	603	233
Angiodema inhibitor	Hs.151242	-	-	414
Alpha crystallin A chain	Hs.184085	-	-	29
Polo(Drosophila)like kinase	Hs.77597	33	163	68
Glycogenin	Hs.174071	33	16	174
Gem GTPase	Hs.79022	33	16	61
ATPase, Cu ⁺⁺ transporting, beta	Hs.84999	0	16	42
Estrogen sulfotransferase	Hs.181327	0	16	74
Rearranged immunoglobulin lambda	Hs.181125	-	-	1221
Prostaglandin E receptor 2	Hs.199248	-	-	38
Cyclin A2	Hs.85137	0	32	27
Phosphoribosyl pyrophosphate synthetase subunit I	Hs.251024	16	0	43
DNA X activation protein	Hs.9963	0	16	50

Table 4.4 SAGE data and EST abundances for genes showing elevated expression in CIN sporadic colorectal cancer cell lines (continued).

Gene Transcript	Unigene Identifier	SAGE Tags per million in HCT116	SAGE Tags per million in Caco2	EST abundance
Insulin-like growth factor 2	Hs.251024	110	587	932
Solute carrier family 2	Hs.33084	0	16	74
Complement component 4-binding protein	Hs.1012	0	16	39
Cholinesterase-related cell division controller	Hs.155266	16	0	34
Granulocyte-macrophage colony-stimulating factor	Hs.182378	7900	7300	9
Calcium channel L-type alpha 1 subunit	Hs.89925	-	-	13
Iduronate 2-sulfatase	Hs.172458	0	16	279
Fibronectin 1	Hs.118162	0	717	1433
Chronic granulomatous disease (b-245 beta polypeptide)	Hs.88974	16	8	40
HMGCo-A reductase	Hs.11899	0	65	156
RYK receptor-like tyrosine kinase	Hs.79350	0	32	94
Ran binding protein 2	Hs.179825	16	32	194
Collagen, type XVIII, alpha 1	Hs.78409	49	163	202
Electron-transfer-flavoprotein, alpha polypeptide	Hs.169919	49	65	101
Monokine induced by α -interferon	Hs.77367	-	-	54

Table 4.5 Genes overexpressed in MIN cell lines with simple sequence repeat motifs. Each gene showing elevated expression in MIN sporadic colorectal cancer cell lines was screened for simple sequence repeats in the form of short mononucleotide repeats. The type of repeat is indicated for each gene and the number of nucleotides is indicated by a subscript number *i.e.* a repeat of eight A residues is denoted A₈. The region where simple sequence repeats occur are listed for each repeat. The genes selected for further analysis are highlighted in bold text.

Gene	Region	Type of repeat
HBA1	-	none
HSP27	ORF	C ₆ , C ₆
CDCL1	ORF	A ₆ , C ₆ , G ₆
H2AZ	-	none
SPTBN1	ORF	A ₆ , A ₆
hMT-Ie	-	none
AP-2	ORF	A ₆
IVD	ORF	G ₆
ELF-1	-	none
CTSS	-	none
Pro-alpha-1 (V) collagen	ORF	9(C ₆), 2(C ₇), G ₆
BTF3	ORF	A ₆
CTP synthetase	ORF	T ₆
RBP3	-	none
Carbonic anhydrase I	-	none
RS19	-	none
DNA primase	-	none
VEGFR2	ORF	A ₆
DUSP5	-	none
Integrin alpha L	ORF	A ₆ , 3(G ₆)
HLA-E	-	none
ARF4	-	none
MFAP1	ORF	A ₆
AKR1C3	ORF	A ₆
RNASE1	-	none
Elastin	-	none
CRH	-	none

Abbreviations:

ORF: Open Reading Frame of gene

Table 4.6 Genes overexpressed in CIN cell lines with simple sequence repeat motifs. Each gene showing elevated expression in CIN sporadic colorectal cancer cell lines was screened for simple sequence repeats in the form of short mononucleotide repeats. The type of repeat is indicated for each gene and the number of nucleotides is indicated by a subscript number *i.e.* a repeat of eight A residues is denoted A₈. The region where simple sequence repeats occur are listed for each repeat. The genes selected for further analysis are highlighted in bold text.

Gene	Region	Type of repeat
Guanosine 5'-monophosphate synthase	-	None
Creatine kinase B	ORF	C ₆ , G ₆
Phosphatidic Acid Phosphatase Type 2A	-	None
Pregnancy-specific beta 1-glycoprotein 7	-	None
TBP-associated factor TAFII80	ORF	C ₆ , G ₆
EDDR1	ORF	C ₆ , C ₆
Ribosomal protein S6 kinase	ORF	C ₆ , C ₇
PTP D1	ORF	A ₈ , C ₆ , T ₆
Coagulation factor XI	ORF	T ₆ , T ₆
C4b-binding protein	-	None
CHED	ORF	A ₆ , A ₇
GM-CSF	ORF	C ₆ , T ₆
HLCC105	ORF	C ₆
Iduronate 2-sulfatase	-	none
Fibronectin I	-	none
X-CGD	ORF	A ₆
HMGCo-A reductase	ORF	A ₆
H-RYK	ORF	A ₉
RanBP2	ORF	A ₆ , A ₆
Col18A1	ORF	C ₇ , G ₆
Electron-transfer-flavoprotein, alpha	ORF	A ₆
Humig	ORF	A ₆
Secretogranin II precursor	-	None
Glycogen debranching enzyme	ORF	A ₈ , A ₈
Small inducible cytokine A2	-	none
G6PD	-	none
Phospholipase C	ORF	A ₆ , A ₆ , G ₆
HE4	-	none
KIAA0088 (Glucosidase II)	3'UTR, ORF, 3'UTR	G ₆ , T ₆ , T ₈

Table 4.6 Genes overexpressed in CIN cell lines with simple sequence repeat motifs.
(continued)

Gene	Region	Type of repeat
C1-inhibitor	-	none
Alpha crystallin A chain	-	none
PLK	-	none
Glycogenin	ORF	T ₆
Gem GTPase	-	none
ATPase, Cu++ transporting, beta	ORF	C ₆ , T ₆
Estrogen sulfotransferase	-	none
IgG lambda	-	none
Prostaglandin E2R	5'UTR, 5'UTR	A ₈ , G ₆
Cyclin A2	ORF	A ₆ , C ₆
Phosphoribosyl pyrophosphate synthetase	ORF	T ₆
DAP12	ORF	G ₆
IGFII	5'UTR, ORF, 5'UTR	A ₁₀ , C ₇ , C ₁₂
Glut5	ORF	G ₆

Abbreviations:
 ORF: Open Reading Frame of gene
 UTR: Untranslated Region of gene

CHAPTER 5

**ANALYSIS OF DIFFERENTIALLY
EXPRESSED GENES**

5.0 INTRODUCTION

A panel of sporadic colorectal cancer cell lines was analysed for global gene expression patterns between the MIN (microsatellite unstable) and CIN (microsatellite stable) genotypes. The techniques employed were Differential Display Reverse Transcription PCR (DDRT-PCR) and high density cDNA array technology. Examination of the cell lines by DDRT-PCR yielded a data set of anonymous differentially expressed cDNA species. Gene Discovery Array (GDA) v1.3 filters yielded a considerable amount of data including a panel of differentially expressed characterised genes (approximately 70 genes). The panel included genes involved in maintenance of chromatin/DNA integrity and genes implicated in cell cycle regulation and cell cycle checkpoints.

Despite the uncharacterised status of the cDNA species obtained by DDRT-PCR further confirmatory investigations needed to be carried out on these cDNAs. DDRT-PCR has emerged as a useful technique with the drawback of a high rate of false positive results. Confirmation of the DDRT-PCR data will allow the efficiency of the technique to be assessed and will determine which cDNA species are suitable for further study. High density cDNA array technology is currently the technique of choice for high throughput gene expression profiling. However, the significance of the differences seen in such analyses remains unclear. Verification of differences seen on GDA v1.3 filters is essential and should be carried out by more established independent techniques. In addition, substantiated expression data will allow for planning of downstream experiments. The established techniques for confirmatory are Northern blot and related methods (Sambrook *et al.*, 1989) and techniques based on Reverse Transcription PCR (He *et al.*, 1995; Larrick *et al.*, 1995). DDRT-PCR yields clones from which radiolabeled oligonucleotide probes can be synthesised and probed against Northern blots. The technique does not lend itself to RT-PCR based techniques as oligonucleotide primers are required in this procedure. GDA v1.3 analyses yields extensive expression data without the benefit of clones or PCR products – one of the drawbacks of the technique. Taking these facts into account, a combination of the Northern blotting and RT-PCR techniques were used to confirm DDRT-PCR and GDA data.

There are three main forms of RT-PCR: competitive PCR (Gilliland *et al.*, 1990), semi-quantitative RT-PCR (He *et al.*, 1995; Larrick *et al.*, 1995) and semi-quantitative RT-PCR with an internal control template (Noonan *et al.*, 1990). Competitive PCR involves the use of a competitive template that is co-amplified with the same primers as the transcript of interest. This approach allows the investigator to gauge the efficiency of the RT and PCR steps. However, this technique is not amenable to the study of an extensive panel of genes, as a separate competitive template would be needed for each transcript of interest. Semi-quantitative RT-PCR involves the use of a 'housekeeping' gene as a control. cDNA generated from the RNA under scrutiny is amplified using primers specific for the housekeeping gene. Genes such as β -actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) are largely expressed at equivalent levels between tissue samples and cell lines. Hence, they are seen as useful controls of the amount of RNA present in different experimental samples. Once the 'housekeeping' gene RT-PCR products for each experimental sample are equal, the technique is repeated in the same manner with primers specific for the gene transcripts under examination. It is important to carry out RT-PCR experiments in the exponential phase of amplification of the PCR. Samples that exceed the exponential phase of PCR tend to saturate. Therefore, semi-quantitative RT-PCR analyses are carried over a range of cycle numbers. Semi-quantitative RT-PCR with the inclusion of an internal control involves the co-amplification of the reference control in the same PCR tube as the target transcript. There are a number of drawbacks associated with this technique. Of most importance is the necessity for both sets of PCR conditions to be the same. In addition, the exponential phase of the two PCRs need to be in the same window. Standard semi-quantitative RT-PCR was considered the technique of choice for this study.

The semi-quantitative RT-PCR protocol used was based on previously described protocols (He *et al.*, 1995; Larrick *et al.*, 1995). Some minor modifications were made with the inclusion of a Cy5 label on a number of the primers. This renders the RT-PCR products fluorescent so that they can be analysed using the ALFexpress™ automated DNA sequencer. Expression values for each PCR (usually a peak area output from the ALFexpress™ product) are compared to the corresponding β -actin or GAPDH expression levels. In the majority of RT-PCR experiments, samples were resolved on 12% acrylamide gels and quantitated by densitometry (Vilber Lourmat Densitometry

Instrument and Vilber Lourmat Bio1D v6.23 Densitometry software) and compared to the expression levels of a housekeeping gene.

cDNA clones were obtained directly from the GDA Clone Library. These cDNA species were routinely isolated (Section 2.3.8) and used directly to probe Northern blots. One drawback of this approach is related to the expression level of the transcript under investigation. Northern blotting is sensitive to five copies of a transcript per cell or greater (Sambrook *et al.*, 1989). A number of the transcripts did not register when probed onto Northern blots. These species were analysed using semi-quantitative RT-PCR, which is more sensitive to gene transcripts expressed at low levels. RT-PCR and Northern blot analyses were used to confirm the differential expression of GDA analysis genes. A subset of the GDA genes were selected for these experiments (Table 5.1A).

In addition to the confirmatory analysis of the GDA analysis genes, we selected a number of candidate genes based upon indications in the literature which may have a role in mismatch repair mechanisms (MIN genotype) or chromosomal lesions (CIN genotype). These genes were analysed using semi-quantitative RT-PCR as above and are listed in Table 5.1B.

The MIN genotype is recognised by instability at microsatellites and simple sequence repeats (Section 1.3.4.). We screened the panel of GDA genes for the presence of coding region simple sequence repeats (Section 4.2.7) and the results are listed in Tables 4.5 and 4.6. The Protein tyrosine phosphatase D1 (PTPD1) gene was found to contain a 9bp poly-A tract in its coding region. We screened this region for instability using PCR primers which flank the repeat sequence and analysed the products on 15% denaturing polyacrylamide gels using the ALFexpress™ automated DNA sequencer.

5.1 METHODS

The techniques of northern blotting and/or semi-quantitative RT-PCR were used to confirm the differential expression of DDRT-PCR and GDA clones. Clones (five in total) were obtained from DDRT-PCR investigations and were analysed where possible. An extensive panel of differentially expressed genes was generated from GDA analyses. A number of these (ten) were selected based on their expression differences and particularly on their possible roles in mismatch repair and cell cycle equilibrium in sporadic colorectal cancer. Sequences were retrieved (where available) using the BLAST algorithm [<http://www.ncbi.nlm.nih.gov/BLAST>] and specific primers were designed using the Oligo 4.0 package. All PCR reactions were optimised prior to amplification on genomic DNA or cDNA where suitable. The major components altered during optimisations were primer concentration, Mg^{2+} concentration and annealing temperature (Section 2.1.6). The optimal conditions for each PCR are shown in Table 2.4.

5.1.1 SEMI QUANTITATIVE RT-PCR/NORTHERN BLOT ANALYSIS OF CONTROLS

Northern blot and semi-quantitative RT-PCR experiments were controlled with the β -actin and GAPDH genes. A housekeeping gene is used as a control since the expression of genes of this nature remains largely constant between different tissues and cell lines. The choice of housekeeping control gene is crucial as differential expression of some housekeeping genes has been reported previously (Goldsworthy *et al.*, 1993). Both of these transcripts have been widely used as controls in gene expression studies on sporadic colorectal tumours (Backert *et al.*, 1999). In addition, all RNA samples were stringently quantitated spectrophotometrically, visually on denaturing gels or by the hybridisation assay. Samples analysed in this way showed similar levels of expression of both β -actin and GAPDH using Northern blotting and RT-PCR.

Semi Quantitative RT-PCR

Primers were designed specific for the human β -actin and GAPDH transcripts using the Oligo 4.0 program. The primers for β -actin and GAPDH produced RT-PCR products of 109bp (bases 1551 to 1659 from GenBank entry NM_001101) and 598bp (bases 212 to 809 from GenBank entry NM_002046) respectively. The PCR for β -actin was initially

optimised on genomic DNA. Suitable amplification conditions were extended to RT-PCR analysis of MIN and CIN total RNA. β -actin experiments were repeated a number of times yielding equivalent amounts of product across the panel of experimental samples. β -actin was observed to be highly expressed in sporadic colorectal cancer cells and products were sampled after 17, 19 and 21 PCR cycles. Products were resolved on 12% polyacrylamide gels. The same reverse transcription reaction was used for both the β -actin analysis and the target genes. An example of a β -actin RT-PCR is illustrated in Figure 5.1.

The GAPDH PCR was also optimised on genomic DNA. The optimal PCR conditions were extended to RT-PCR on the MIN and CIN total RNA samples. GAPDH products were separated on 1% agarose gels or 12% polyacrylamide gels and are illustrated in Figure 5.2.

Northern blot analysis

MIN and CIN total RNA was quantitated spectrophotometrically and equivalent amounts (10-20 μ g) were separated on denaturing 1% agarose gels. Northern blot probes for the β -actin and GAPDH transcripts were prepared by RT-PCR (with the incorporation of [α - 32 P]dCTP) for each transcript. Northern blots were probed with β -actin and GAPDH probes. Balanced Northern blots for β -actin and GAPDH are shown in Figures 5.3a and 5.4a respectively with densitometry values listed in Figures 5.3b and 5.4b for each blot.

5.1.2 DETECTION OF INSTABILITY AT SIMPLE SEQUENCE REPEAT MOTIFS

Specific primers for the Transforming Growth Factor beta Receptor II (TGF β RII) were obtained which span the 10bp poly-A tract in the open reading frame of the transcript. Specific primers were designed for the Protein Tyrosine Phosphatase D1 (PTPD1) gene using the Oligo 4.0 primer design software which flanks a 9bp poly-A tract in the open reading frame of the transcript. RT-PCR was carried out on both transcripts as described in Section 2.1.8 with minor modifications. The sense TGF β RII and PTPD1 primers were fluorescently labelled with a Cy5 label. RT-PCR products were electrophoresed on 15% denaturing polyacrylamide sequencing gels using the ALFexpressTM automated sequencer. MIN and CIN RT-PCR products were analysed in

parallel to identify electrophoretic shifts due to instability in the poly-A tracts in both transcripts.

5.2 RESULTS

5.2.1 SEMI QUANTITATIVE RT-PCR/NORTHERN BLOT ANALYSIS OF TARGET GENES

Primer sets were designed for the genes of interest using the Oligo 4.0 package. Semi-quantitative RT-PCR experiments were carried out on previously normalised RNA samples (RNA samples were balanced with respect to 'housekeeping genes'). All experiments were carried out in triplicate and utilised the same reverse transcription reactions used in the 'housekeeping' gene analysis. Reaction products were resolved on non-denaturing 12% polyacrylamide gels and the amount of each target product compared to the relative amount of 'housekeeping' product. The primers for each target gene and the conditions for each PCR are detailed in Tables 2.2 and 2.4

Total RNA quantitated spectrophotometrically, visualised on denaturing 1% agarose gels and normalised by semi-quantitative RT-PCR using a 'housekeeping' gene were used in Northern blot analyses. Total RNA (10-20µg) was resolved on 1% denaturing agarose gels and transferred to nylon membranes (Section 2.3.9). Probes for target genes were prepared in two ways. Primers were designed for the target under investigation and the target amplified from a cDNA template followed by gel purification (See Chapter 2). Probes were also generated from cDNA clones by excising the target sequence from a vector with restriction enzymes. The excised products were gel purified (Chapter 2). PCR primers used to cDNA generate probes and the PCR conditions are detailed in Tables 2.2 and 2.4. The source of clones and the conditions used to generate cDNA probes are described in Table 2.5.

5.2.2 Gene Discovery Array Data

The consistently differentially expressed genes found by GDA analysis selected for further analysis are summarised in Table 5.1.

Polo(*Drosophila*)like Kinase (PLK)

A GDA clone (Clone no. 23 m 21) was obtained from the GDA Clone library. Plasmid DNA was isolated from the clone and was digested with the restriction enzymes *Bgl*I and *Eco*RI. The digest was resolved on a 2% TAE agarose gel. It was clear from the undigested and digested plasmid DNA that more than one clone was present in the

original bacterial stab received from the GDA clone library (data not shown). The bacterial clone was plated into distinct bacterial colonies on LB agar plates. Plasmid DNA minipreps from separate colonies were digested with *Bgl*I and *Eco*RI. In addition, the insert sequence was amplified directly from plasmid DNA by PCR using the Forward and Reverse sequencing primers. This approach yielded digestion products and PCR products of various sizes confirming a heterogeneous mixture of cDNA clones. The study of this clone was discontinued as a result.

Primers were designed for the PLK transcript yielding a 158bp PCR product between the bases 183 and 339 from the sequence U01038 (See Table 2.4 for PCR conditions). The PLK RT-PCR was sampled after 24, 26 and 28 cycles. Levels of the PLK transcript were compared to the levels of both β -actin and GAPDH. A modest decrease in expression of PLK in the MIN cell lines (LS174T, SW48 & HCT116) with respect to the CIN cell lines (Caco2, SW480 & HT29) was observed after electrophoresis. A representative PLK RT-PCR is illustrated in Figure 5.5. This corresponds to the levels of PLK expression observed in GDA analyses. In addition, the public SAGE database yields a ratio of 33/163 of tags (per million) between the cell lines HCT116 (MIN) and Caco2 (CIN) (Table 4.4). These data confirm reduced expression of the PLK transcript in MIN sporadic colorectal cancer cell lines.

Ran binding protein 2 (RANBP2)

Primers were designed specific for the human RANBP2 transcript amplifying a PCR product of 118bp between nucleotides 245 and 361 of AF012086 (See Table 2.4 for PCR conditions). Expression of the RANBP2 transcript was compared to the expression of the two control genes. Examination of the RANBP2 electrophoresis data shows a decrease in expression of RANBP2 in MIN cell lines compared to the panel of CIN cell lines. A similar observation was seen in the original analysis of these cell lines using GDA v.13 filters. A RANBP2 semi-quantitative RT-PCR is illustrated in Figure 5.6. These data correlate with the public SAGE data which yields a 16/32 tag (per million) ratio between HCT116 and Caco2 for the RANBP2 transcript. These data taken together confirm the differential expression of RANBP2 between MIN and CIN colorectal cancer cells.

Prostaglandin E Receptor 2 (PGER2)

Oligonucleotide primers were designed to specifically amplify the human PGER2 transcript. The PCR product generated is 131bp in size from bases 594 to 723 of the GenBank entry L25124 (See Table 2.4 for PCR conditions). Semi-quantitative RT-PCR experiments were carried out on this transcript and levels were compared to the expression of GAPDH and β -actin. High expression of PGER2 was observed in the MIN cell lines LS174T and SW48 with low levels in the remaining cell lines. This contradicts the GDA data which suggested overexpression of PGER2 in CIN cell lines. The reason for this discrepancy remains unclear. The public SAGE database yielded no reliable tags from the PGER2 transcript preventing comparison of our PGER2 data with public information. A representative RT-PCR experiment for the PGER2 transcript is shown in Figure 5.7. One possibility is that the clone placed in the positions denoted for PGER2 on the GDA v1.3 filters does not represent PGER2. This hypothesis may explain the difference in expression pattern seen between the filters and RT-PCR analysis.

Cyclin A2 (CCNA2)

PCR primers were designed to specifically amplify the human CCNA2 transcript (PCR product is 136bp) between the bases 177 and 311 from the GenBank entry NM_001237. Semi-quantitative RT-PCR experiments were carried out using the CCNA2 primers and were compared to the expression of β -actin and GAPDH. Expression of CCNA2 was not consistently differentially expressed between the MIN and CIN cell lines. However, the differences in CCNA2 expression in each MIN/CIN comparison from GDA analysis correlate directly with the data obtained from CCNA2 RT-PCR, i.e. CCNA2 expression between LS174T and SW480 was consistent for both GDA and RT-PCR analyses and a similar phenomenon was observed for the two other comparisons. Moreover, the comparison of CCNA2 expression by examination of SAGE tag abundances revealed a 0/32 ratio between HCT116/Caco2. An example of a CCNA2 RT-PCR experiment is illustrated in Figure 5.8. In conclusion, these data show elevated CCNA2 expression in CIN cells with detectable CCNA2 present in MIN cells.

Basic transcription factor 3 (BTF3)

A BTF3 clone (Clone no. 21 f 24) was obtained from the GDA clone library. Plasmid DNA was isolated and the identity of the clone was confirmed by cycle sequencing (data not shown). The plasmid was digested with the restriction enzymes *Bgl*II and *Eco*RI and the digest was resolved on a 2% TAE agarose gel. The expected BTF3 insert (in reference to the clone data received from Genome Systems) was gel purified and radiolabeled with [α -³²P]dCTP. Northern blots were prepared as described (Section 2.3.9) and were probed with the BTF3 clone. The same Northern blot was also probed with a human β -actin specific probe. The β -actin probe was prepared by radiolabelling the β -actin 109bp PCR product. Levels of BTF3 and β -actin expression were quantitated by densitometry. The expression level of BTF3 was clearly seen to be higher in MIN cells compared to CIN cells although the transcript was present at quite high levels in Caco2 (CIN). Furthermore, use of the public SAGE database showed an expression ratio of 605/250 tags (per million) for the cell lines HCT116(MIN) and Caco2(CIN) for the BTF3 transcript. These data together confirm lower expression of BTF3 in CIN cells with respect to MIN cells. A northern blot of BTF3 is illustrated in Figure 5.9a with the densitometry data listed in Figure 5.9b.

Histone deacetylase 2 (HDAC2)

A HDAC2 clone (Clone no. 45 p 3) was obtained from the GDA clone library. Plasmid DNA was isolated and the identity of the clone was confirmed by cycle sequencing (data not shown). Plasmid DNA was digested with the restriction enzymes *Not*I and *Eco*RI and the digest was resolved on a 2% TAE agarose gel. The expected HDAC2 insert fragment was gel purified and radiolabelled with [α -³²P]dCTP. A northern blot was probed with HDAC2 and GAPDH probes. Densitometry showed the HDAC2 transcript to be expressed at lower levels in LS174T and HCT116 (MIN) with respect to SW480 and HT29 (CIN) with similar levels of expression in the cell lines SW48 and Caco2. This result correlates directly with GDA data for the HDAC2 transcript. In addition, the 'gene-to-tag' public SAGE database suggests a 16:48 tags (per million) ratio between HCT116 and Caco2 for this transcript. The northern blots and densitometry data are shown in Figures 5.10a and 5.10b. These findings confirm diminished expression of the HDAC2 transcript in MIN cells with the exception of SW48 (MIN).

40S Ribosomal protein S19 (RPS19)

A RPS19 clone was received as a kind gift from Dr. Peter McWilliam. The clone had previously been sequence verified and a PCR probe was generated by PCR using the Forward and Reverse sequencing primers. The expected size of the PCR product was verified by resolving the PCR product on a 2% TAE agarose gel. The RPS19 product was gel-purified and labeled with [α - 32 P]dCTP by random priming (Chapter 2). A northern blot was probed with RPS19 and GAPDH probes. GDA analysis had shown RPS19 to be overexpressed in MIN cell lines. However, densitometry data based on RPS19 and GAPDH show RPS19 to be expressed at higher levels in CIN cell lines. The northern blot data and densitometry are shown in Figure 5.11a and Figure 5.11b. Closer examination of GDA data for the RPS19 transcript showed it to be present on the GDA filters in two different locations each associated with the same Unigene identifiers but derived from different GenBank entries (GenBank R68033 and GenBank AA046433). In contrast to the clone initially considered (R68033), the GDA expression differences observed in the AA046433 clone show elevated RPS19 expression in CIN cells. Further analysis of the origins of these clones and information on possible RPS19 variants will be required to clarify these expression discrepancies. Unfortunately, no reliable tags for the RPS19 transcript were available in the public SAGE database. To summarise, differential expression of RPS19 by GDA analysis was not confirmed by the independent method of RT-PCR.

Histone 2A Z (H2AZ)

A H2AZ clone was received from Dr. Peter McWilliam, which had been previously sequence verified. The clone insert was amplified by PCR using the Forward and Reverse sequencing primers which flank the insert region. The PCR product was gel purified and radiolabeled with [α - 32 P]dCTP by random priming. A northern blot was probed with H2AZ and GAPDH specific probes. An increase was observed in expression of the H2AZ transcript in MIN cell lines with respect to CIN cells. Densitometry on both H2AZ and GAPDH verified the above observation. In addition, the 'gene-to-tag' public SAGE database yields a slight increase in H2AZ expression in the MIN cell line HCT116 with respect to the CIN cell line Caco2. The H2AZ and

GAPDH Northern blots are shown in Figure 5.12a and 5.12b. These data indicate an overall higher expression of H2AZ in the panel of MIN cells

5.2.3 Analysis of Candidate gene expression data

A number of genes were selected for expression analysis on the basis of their potential roles in mismatch repair mechanisms or chromosomal lesions from indications in the literature. The panels of genes selected are listed in Tables 5.1A and 5.1b.

Cytokeratin 8 (Cyt8)

Primers were designed to amplify the human Cyt8 transcript between the bases 882 and 1207 from the GenBank entry NM_002273, yielding a PCR product of 326bp. The Cyt8 product was sampled at 17, 20 & 23 cycles of semi-quantitative RT-PCR. β -actin and GAPDH RT-PCR experiments were carried out on the same reverse transcription reaction for control purposes. Expression of Cyt8 was found to be lower in MIN cells with respect to CIN cells. Two clones corresponding to Cyt8 (GenBank entries N73980 and H60463) are harboured on the GDA v1.3 filters. The expression of these clones concurred with the expression differences seen by RT-PCR. The ‘gene-to-tag’ mapping showed a slight increase in Cyt8 expression in HCT116 (MIN) compared to Caco2 (CIN). The expression differences are illustrated in Figure 5.13. These data show elevated expression of Cyt8 in CIN cells.

Cytokeratin 18 (Cyt18)

Primers were designed for the human Cyt18 gene between the bases 369 and 497 of the GenBank entry NM_000224 yielding a PCR product of 129bp. Semi-quantitative RT-PCR was performed using the primers on MIN and CIN cell line total RNA and products were sampled at 17, 19 & 21 cycles of amplification. RT-PCR was carried out using β -actin primers and GAPDH primers on the same total RNA. These experiments revealed higher expression of Cyt18 in CIN cells. This contradicts the data for a Cyt18 clone (AA044720) on the GDA v1.3 filters, which suggests equivalent expression of Cyt18 between both MIN and CIN cell lines. The ‘gene-to-tag’ mapping database yielded a 500:700 ratio for the Cyt18 transcript between HCT116 (MIN) and Caco2 (CIN). The different results may be clarified by examining in detail the source of

AA044720 and its listed identity as a Cyt18 clone. However, this result may simply indicate false positive hybridisation on the filters. The Cyt18 semi-quantitative RT-PCR is illustrated in Figure 5.14. In summary, the expression of Cyt18 in MIN and CIN cells by GDA and RT-PCR analyses is inconclusive. A northern blot of Cyt18 may prove useful in elucidating the above findings.

Inhibitor 2 of Protein Phosphatase 1 (IPP2)

Primers were designed to amplify a 109bp human IPP2 transcript (See Chapter 6 for a detailed description of the IPP2 transcript) PCR product (Table 2.2 DDP U and DDP L) between nucleotides 589 and 697 of the GenBank entry NM_006241. The DDP PCR product is from the IPP2 open reading frame upstream of the IPP2 (GAP) probe used for Northern blot analysis (See below). Semi-quantitative RT-PCR was carried out on IPP2 (products were sampled at 21, 23 & 25 cycles), β -actin and GAPDH on the same reverse transcription reaction. This revealed higher expression of IPP2 in CIN cells. The IPP2 RT-PCR is shown in Figure 5.15. The IPP2 expression results obtained by Northern and RT-PCR analyses confirm overexpression of IPP2 in CIN cells (with particularly high expression in HT29).

A set of IPP2 primers (Table 2.2 GAPU and GAPL) were accepted as a gift from Dr. Peter McWilliam. The primers generate an 872bp product from 821 to 1694 of GenBank entry AJ133812. An IPP2 cDNA probe was generated by PCR. The PCR product size was verified by 2% TAE agarose gel electrophoresis and the IPP2 product was gel purified and radiolabelled as before. A northern blot was probed with IPP2 and GAPDH. The IPP2 northern blot and densitometry are illustrated in Figures 5.16a and 5.16b. This clearly showed elevated expression of both IPP2 transcripts (the probe detects two alternative polyadenylation IPP2 transcripts) in CIN cells with respect to MIN cells.

A GDA clone present on the GDA v1.3 filters identified as IPP2 (H85608) exhibits a complicated pattern of expression across the panel of sporadic colorectal cancer cell lines. In addition, no IPP2 expression data was available in the SAGE database. The reason for this discrepancy between the GDA analysis and the northern blot/RT-PCR data is unclear. Non-specific hybridisation to the IPP2 clone on the GDA filters is a likely cause of the conflicting data sets. The identity of the IPP2 clone may

not be verified in the particular plate of the GDA clone library which 'houses' the clone. Errors may exist in the GDA clone library (personal observation) and the concurring data sets between northern blotting and RT-PCR suggests that the IPP2 transcript is expressed more highly in CIN cell lines than MIN cell lines. A convenient internal control for these observations is the established knowledge that IPP2 is overexpressed in SW480 with respect to SW620 (McWilliam *et al.*, MS in preparation) which is observed in the IPP2 RT-PCR.

Cyclooxygenase 2 (COX2)

Oligonucleotide primers were designed to specifically amplify the human COX2 transcript. The COX2 PCR product is 114bp from 3214 to 3336 of the GenBank entry NM_000963. COX2 semi-quantitative RT-PCR experiments were carried out and products were sampled at 24, 27 & 30 cycles. RT-PCR was also carried out on the β -actin transcript as a control. The RT-PCR polyacrylamide gel for COX2 is illustrated in Figure 5.17. No COX2 data were found in the public SAGE database possibly due to low expression of the transcript. Expression of COX2 was observed in the cell lines LS174T (MIN), HCT116 (MIN) and Caco2 (CIN) with undetectable levels in the remaining cell lines.

Secretory Phospholipase A2 (sPLA2)

Oligonucleotide primers were designed to specifically amplify the human sPLA2 transcript. The PCR product is 129bp between bases 245 and 372 of the GenBank entry NM_000300. sPLA2 semi-quantitative RT-PCR experiments were carried out and products were sampled at 21, 24 & 27 cycles. RT-PCR was also carried out on the β -actin transcript as a control. Unfortunately, no SAGE data were available for the sPLA2 transcript. The RT-PCR polyacrylamide gel for sPLA2 is illustrated in Figure 5.18. Expression of sPLA2 was detectable in the cell lines LS174T (MIN) and Caco2 (CIN) with undetectable levels in the remaining cell lines.

Mitotic arrest deficient yeast homolog like 1 (MAD2)

Preliminary GDA data (after two pairwise comparisons of MIN and CIN cells) yielded higher expression of the MAD2 transcript in MIN cells compared to CIN cells. The

third comparison showed MAD2 expressed at similar levels in HCT116 (MIN) and HT29 (CIN). A clone for MAD2 was received as a gift from Dr. Peter McWilliam. A northern blot was probed with MAD2 and GAPDH. MAD2 was expressed at very low levels on the Northern blot and SAGE database (data not shown) and the study of this transcript was not extended.

5.2.4 Analysis of Instability in the TGF β RII and PTPD1 transcripts

Transforming growth factor beta receptor II (TGFBRII)

Previously published primers (Parsons *et al.*, 1995) specific for the human TGFBRII transcript were obtained, yielding a 73bp PCR product between bases 130 and 202 of GenBank entry U69148. A Cy5 label was included on the TGFBRII sense primer. The primers were designed to flank a 10bp polyA tract. Semi-quantitative RT-PCR was carried out as before on MIN and CIN cell line total RNA including the control gene β -actin. The TGFBRII product was examined for instability at the polyA tract by resolving TGFBRII products on 15% denaturing polyacrylamide gels on the ALFexpress™ automated DNA sequencer. TGFBRII instability in MIN cell lines is illustrated in Figure 5.19. The expected shifts indicating frameshift mutations were observed in MIN cells with no discernable shift in CIN cells. In conclusion, this result confirms the reported instability in the TGF β RII gene (Section 1.3.4) in MIN colorectal cells.

Protein Tyrosine Phosphatase D1 (PTPD1)

Primers specific for the human PTPD1 transcript were designed yielding a 116bp PCR product between bases 2802 and 2917 of GenBank entry X79510. The primers were designed to flank an 8bp polyA tract with the inclusion of a Cy5 label on the sense PTPD1 primer. Semi-quantitative RT-PCR was carried out as before on MIN and CIN cell line total RNA including the control gene β -actin. The expression differences in PTPD1 are illustrated in Figure 5.20. This data combined with the GDA data confirm overexpression of PTPD1 in MIN colorectal cells.

The PTPD1 product was examined for instability at the polyA tract by resolving RT-PCR PTPD1 products on 15% denaturing polyacrylamide gels on the ALFexpress™ automated DNA sequencer. No discernable shift in PTPD1 PCR products was observed between MIN and CIN cell lines (data not shown). This suggests that the PTPD1 is not a target in the MIN genotype of colorectal cancer (Section 1.3.4).

5.3 DISCUSSION

5.3.1 ROLES FOR GDA GENES IN COLORECTAL CANCER

PLK

A potential cause of aneuploidy involves centrosomes. Multipolar spindles have been observed in human cancers *in situ* and an abnormal number of centrosomes have been observed in colorectal cancer (Doxsey *et al.*, 1998). Polo (*Drosophila*)-like Kinase (PLK) is a protein which is associated with proliferating cells. Wolf *et al.* (1997) found elevated expression of PLK in non-small cell lung cancer (NSCLC) tumours compared to low expression of PLK in normal lung tissue. The extent of elevated expression correlated directly with poor patient prognosis. Another PLK study found that downregulation of PLK is necessary for withdrawal of cardiomyocytes from the cell cycle (Georgescu *et al.*, 1997). PLK is a centrosome-associated kinase and, similar to aurora 2/STK15, when exogenously expressed in mammalian cells can affect centrosome number and chromosome segregation (Wolf *et al.*, 1997). Overexpression of PLK in CIN cells (observed both by GDA analysis and RT-PCR) may suggest a role for overexpression of PLK in colorectal cancer chromosomal abnormalities.

RANBP2

Components of the cell that regulate nuclear organization may have a role in proper execution of chromosome replication and segregation. One of the important components for the control of nuclear structure is the Ran/RCC1 molecular switch system (Ouspenski *et al.*, 1995). Ran is an abundant nuclear GTPase with a defined role in nuclear transport. However, the role of Ran in mitotic regulation is less clearly understood. RanBP2 binds Ran GTPase and is thought to mediate the Ran pathway from the cytoplasm into the nucleus. In addition, overproduction of RanBP2 in yeast induces mitotic chromosome disjunction (Ouspenski *et al.*, 1995). Due to these findings we investigated the expression of RanBP2 in MIN and CIN colorectal cancer cell lines by RT-PCR which showed elevated RANBP2 expression in CIN cells as predicted by GDA analysis. This finding is consistent with the possibility that the overproduction of RANBP2 affects mitosis in CIN colorectal cancer cells lines thereby contributing to chromosomal instability.

methyated gene, which is associated with a deacetylated nucleosome structure, could be relieved by the deacetylase inhibitor, trichostatin A (Razin A., 1998). In addition, HDAC2 may have a role in DNA damage during the cell cycle by its association with ataxia telangiectasia mutated (ATM)-and Rad-3-related protein (ATR) (Schmidt *et al.*, 1999). HDAC2 inhibits the BAX promoter, a natural p53-responsive system, in a p53-dependent manner and acetylates p53 silencing its transcriptional activity (Juan *et al.*, 2000). Underexpression of HDAC2 in MIN cells was observed by GDA and Northern blot analysis suggesting a potential role involving gene silencing in the MIN genotype.

H2AZ

Histone 2AZ (H2AZ) is a replication-independent member of the histone gene family. The gene has both regions of low and high similarity to other H2A family members. However, H2AZ contains introns and encodes a polyadenylated mRNA in contrast to other histones. The gene was localised to chromosome 4q24 by FISH (Popescu *et al.*, 1994). Histones are transported to the nucleus by Importin β 1 and this process is dependent on the acetylation state of the histones (Johnson-Salibu *et al.*, 2000). Interestingly, centromeric heterochromatin does not contain H2AZ (Rusanova *et al.*, 1989) and Dnmt1 (DNA Methyltransferase I) seems to methylate this region of chromosomes specifically (Rhee *et al.*, 2000). We have found overexpression of H2AZ in MIN cells. This result and the HDAC2 data taken together may suggest reciprocal roles for H2AZ and HDAC2 in gene silencing in MIN cells.

PTPD1

Inhibition of protein tyrosine phosphatases (PTP) induces cell proliferation with concomitant increase in the phosphotyrosine protein product. Induction of PTPs has the opposite effect contributing to PTPs being considered as tumour suppressor genes. PTPs are downregulated in colorectal cancer and upregulated in breast cancer (Partanen *et al.*, 1996). PTPs have been shown to be sites of LOH in renal cell carcinomas (Druck *et al.*, 1995). The PTPD1 protein contains an ezrin domain which joins microfilaments together. We found consistently raised expression of PTPD1 in MIN cells by both GDA analysis and RT-PCR. CIN tumours are recognised by gross chromosomal lesions including LOH. Expression of PTPD1 was undetectable in CIN cells by RT-PCR. This

suggests either extremely low expression of the transcript in CIN cells or potential LOH at the PTPD1 locus in CIN cells. In addition, the PTPD1 gene has a 9bp poly-A tract in its coding sequence. We screened the gene for instability at this region by PCR flanking the repeat sequence. We were unable to detect a discernable shift in the MIN cell PTPD1 RT-PCR products suggesting that the poly-A sequence is not a target for the MIN genotype.

5.3.2 ROLES FOR CANDIDATE GENES IN COLORECTAL CANCER

CYT8

A large-scale study of gene expression in colorectal cancer was carried out by the isolation and sequencing of 2,166 clones from a human cDNA library (Frigerio *et al.*, 1995). The expression of Cyt8 (Cytokeratin 8) was seen to be decreased 5-fold in carcinoma compared to normal colon mucosa. Cytokeratins are key components of cytoskeletal maintenance and are generally found in the epithelia that line internal body cavities (Lodish *et al.*, 1995).

We investigated the expression of Cyt8 in MIN and CIN cell lines by RT-PCR. Repeat experiments showed consistent reductions in expression of Cyt8 in MIN cell lines (albeit modest differences). Western blot analysis of Cyt8 would clarify whether the expression differences are maintained post-translationally and the extent of the differences.

CYT18

The comprehensive analysis of colorectal gene expression carried out by Frigerio *et al.* (1995) also revealed differential expression of Cyt18 (cytokeratin 18) between normal colon mucosa and colon carcinoma. Intriguingly, the ratio of Cyt18 between normal and carcinoma was 1:5 (the opposite ratio seen in Cyt8). In light of this, we also investigated expression of Cyt18 in MIN and CIN cells.

RT-PCR analysis of Cyt18 revealed higher expression in the CIN cell lines compared to the MIN cell lines. A similar phenomenon was observed in Cyt8 RT-PCR analysis.

IPP2

Two small heat-stable proteins were first described in 1976 (Huang *et al.*, 1976) in rabbit skeletal muscle (IPP1 and IPP2) as potent inhibitors of Protein Phosphatase 1. Inhibitor 2 of Protein Phosphatase I (IPP2) was cloned and characterised in 1994 (Sanseau *et al.*, 1994). The IPP2 transcript maps to chromosome 6 which yields a 203 amino acid residue protein of 22.8kDa. DDRT-PCR analysis generated clone PM208 which was identified as human IPP2 (McWilliam *et al.*, MS in preparation). Expression of IPP2 was investigated by RT-PCR analysis and Northern analysis in MIN and CIN colorectal cancer cell lines as a potential transcript involved in the distinct genotypes. Elevated expression of IPP2 was observed in CIN cells by both techniques (particularly high expression of IPP2 was observed in HT29 cells). IPP2 is currently seen as a chaperone protein for Protein Phosphatase I (Alessi *et al.*, 1993) which has been associated with diverse cellular processes including completion of cell division (Shenolikar *et al.*, 1991). Elevated IPP2 presumably accentuates PP1 activity thereby stimulating cell division. Abnormally high cell division may be a contributory factor to the chromosomal lesions seen in CIN colorectal cancer cells. Investigation of the expression of the PP1 isozymes (PPP1Calpha, PPP1Cbeta & PPP1Cgamma) in the MIN and CIN cell lines would help to clarify this hypothesis.

COX2

Cyclooxygenase 2 (COX2) was selected as a candidate gene potentially associated with the different nature of molecular instabilities seen in MIN and CIN colorectal cancer cell lines. Epidemiological studies have shown a link between long-term intake of nonsteroidal anti-inflammatory drugs (NSAID) and a reduced mortality from colorectal cancer (Thun *et al.*, 1993). The chemopreventative properties of NSAIDs may be attributable to their inhibitory activity against COX2 which catalyses the conversion of arachidonic acid to prostaglandins (Donnelly *et al.*, 1997). COX2 and prostaglandin levels are increased in sporadic colorectal adenomas (Eberhart *et al.*, 1994), carcinomas (Rigas *et al.*, 1993) and in FAP (Giardiello *et al.*, 1997) in human.

We investigated the expression of COX2 in MIN and CIN colorectal cancer cell lines by RT-PCR. Expression of COX2 was observed in LS174T (MIN) and Caco2 (CIN) after a high number of PCR cycles (30). This may be due to low expression of COX2 in the panel of cell lines selected.

SPLA2

SPLA2 (Secretory phospholipase A2) was selected as a candidate gene involved with the MIN and CIN genotypes. SPLA2 has been shown to release arachidonic acid for prostaglandin production (Hara *et al.*, 1991). SPLA2 has been shown to act as a modifier of the *Apc* gene in the *Min* mouse (a murine model for Familial Adenomatous Polyposis). Functional SPLA2 is required in a protective role against polyp formation in the *Min* mouse (MacPhee *et al.*, 1995). Therefore, it appeared to be worth investigating the expression of the transcript in MIN and CIN cell lines due to their specific molecular lesions (Chapter 1).

RT-PCR was used to analyse SPLA2 expression and the results were similar to the COX2 expression findings. SPLA2 expression was detected in LS174T (MIN), HCT116 (MIN) and in the CIN cell line Caco2. This is comparable to the expression of COX2 with the exception of SPLA2 expression in HCT116. The COX2 and SPLA2 results taken together do not suggest a distinguishing role for either of the genes between the MIN and CIN genotypes. However, both transcripts operate in the prostaglandin synthesis pathway which may be compromised in the remaining cell lines (SW48, SW480 & SW620). In addition, both SW480 and SW620 harbour truncating APC mutations and the absence of COX2 and SPLA2 may have the same affect on these cells as seen in the *Min* mouse.

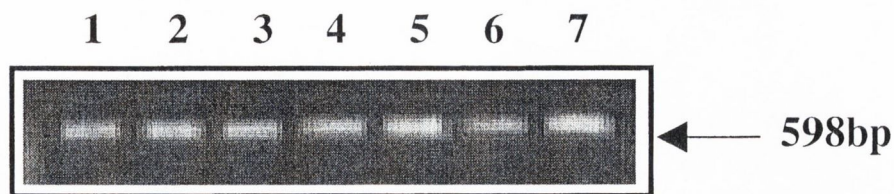
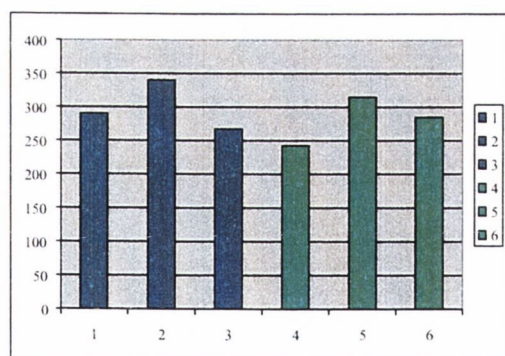


Figure 5.2 Semi-quantitative RT-PCR analysis of the ‘housekeeping’ gene GAPDH in MIN and CIN cell lines. RT-PCR analysis was carried out using specific human GAPDH primers on the MIN and CIN sporadic colorectal cancer cell lines. PCR products were electrophoresed on 12% non-denaturing polyacrylamide gels and stained in ethidium bromide. This gel shows GAPDH PCR products sampled after 17 PCR cycles. The expected product size of 598bp was obtained when compared to the molecular weight marker pBluescript-*Hae*III. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480; 6=HT28 & 7=SW620.



Figure 5.3A Northern blot analysis of the ‘housekeeping’ gene β -actin in MIN and CIN sporadic colorectal cancer cell lines. An illustration of β -actin Northern blot analysis of the panel of sporadic colorectal cancer cell lines. Northern blot procedures were carried out as described (Section 2.3.9). The expected approximate signal of 1.761 kb was obtained as judged by the migration of the 28S and 18S ribosomal RNA bands. Expression signal intensities of β -actin were estimated using densitometry. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480 & 6=SW620.



No.	Cell Line	Expression Value
1	LS174T	289
2	S48	341
3	HCT116	267
4	Caco2	242
5	SW480	315
6	HT29	285

Figure 5.3B. Densitometry analysis of the β -actin transcript. Expression of the β -actin transcript was measured densitometrically. The bar chart shows the expression of β -actin in each of the MIN and CIN colorectal cancer cell lines. The MIN cells are highlighted in **navy text** and the CIN cells are highlighted in **green text**. The densitometry values are listed in the Table above.

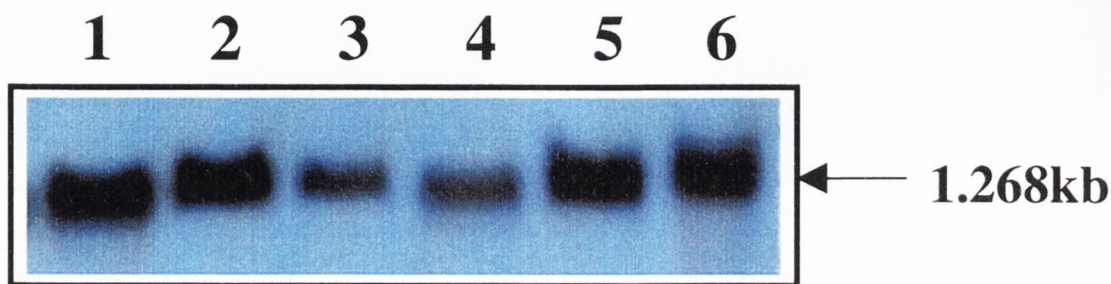
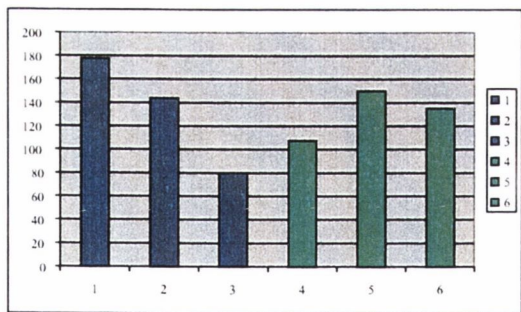


Figure 5.4A Northern blot analysis of the ‘housekeeping’ gene GAPDH in MIN and CIN sporadic colorectal cancer cell lines. An illustration of GAPDH Northern blot analysis of the panel of sporadic colorectal cancer cell lines. Northern blot procedures were carried out as described (Section 2.3.9). The expected approximate signal of 1.268 kb was obtained as judged by the migration of the 28S and 18S ribosomal RNA bands. Expression signal intensities of β -actin were estimated using densitometry. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480 & 6=SW620.



No.	Cell Line	Expression Value
1	LS174T	177
2	S48	144
3	HCT116	80
4	Caco2	108
5	SW480	150
6	HT29	135

Figure 5.4B. Densitometry analysis of the GAPDH transcript. Expression of the GAPDH transcript was measured densitometrically. The bar chart shows the expression of GAPDH in each of the MIN and CIN colorectal cancer cell lines. The MIN cells are highlighted in navy text and the CIN cells are highlighted in green text. The densitometry values are listed in the Table above.

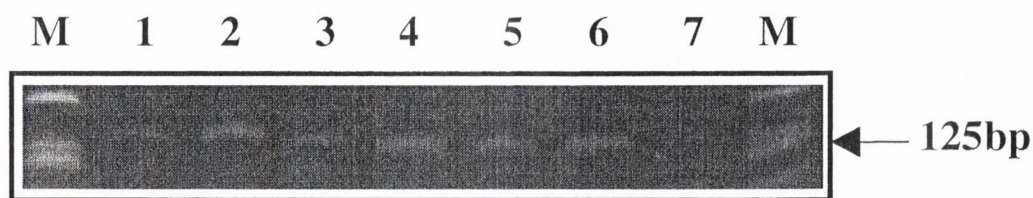


Figure 5.6 Semi-quantitative RT-PCR analysis of the candidate gene Ran binding protein 2 (RANBP2) in MIN and CIN cell lines. RT-PCR analysis was carried out using specific human RANBP2 primers on the MIN and CIN sporadic colorectal cancer cell lines. PCR products were electrophoresed on 12% non-denaturing polyacrylamide gels and stained with ethidium bromide. This gel shows RANBP2 PCR products sampled after 28 cycles. The expected product size of 125bp was obtained compared to the molecular weight marker pBluescript-*Hha* I. M= pBluescript-*Hha* I molecular weight marker; 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480; 6=HT29 & 7=SW620.

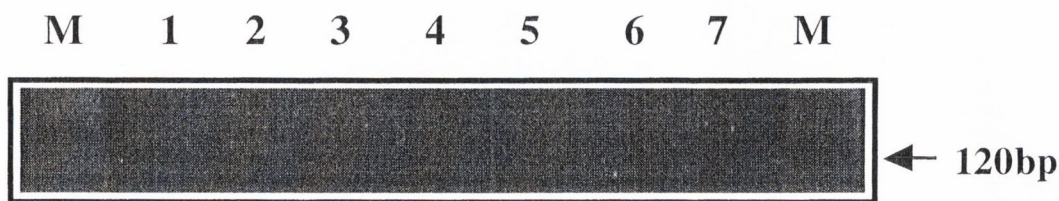


Figure 5.7 Semi-quantitative RT-PCR analysis of the candidate gene Prostaglandin E2 receptor (PGER2) in MIN and CIN cell lines. RT-PCR analysis was carried out using specific human PGER2 primers on the MIN and CIN sporadic colorectal cancer cell lines. PCR products were electrophoresed on 12% non-denaturing polyacrylamide gels and stained in ethidium bromide. This gel shows PGER2 PCR products sampled after 30 PCR cycles. The expected product size of 120bp was obtained when compared to the molecular weight marker pBluescript-*Hae*III. M= pBluescript-*Hae*III molecular weight marker; 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480; 6=HT28 & 7=SW620.

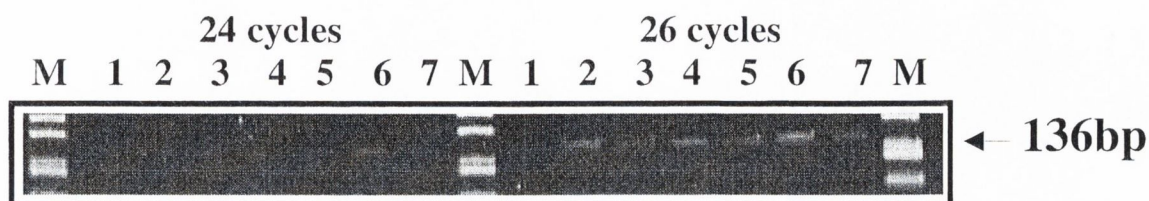


Figure 5.8 Semi-quantitative RT-PCR analysis of the candidate gene Cyclin A2 (CCNA2) in MIN and CIN cell lines. RT-PCR analysis was carried out using specific human CCNA2 primers on the MIN and CIN sporadic colorectal cancer cell lines. PCR products were electrophoresed on 12% non-denaturing polyacrylamide gels and stained in ethidium bromide. This gel shows CCNA2 PCR products sampled after 24 and 26 PCR cycles. The expected product size of 136bp was obtained when compared to the molecular weight marker pBluescript-*Hae*III. M=pBluescript-*Hae*III; 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480; 6=HT29 & 7=SW620.

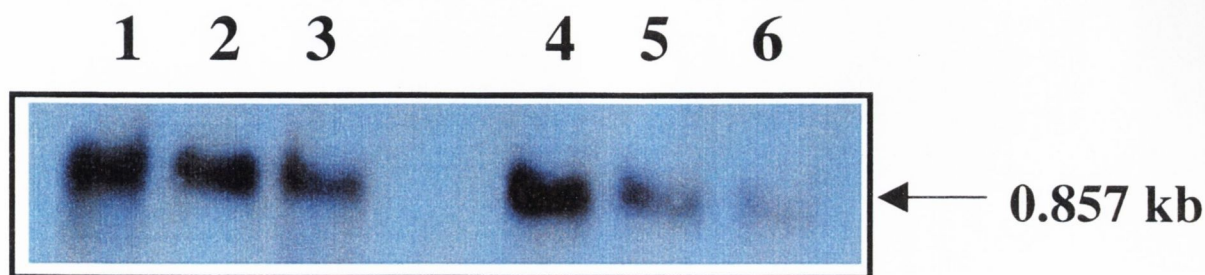
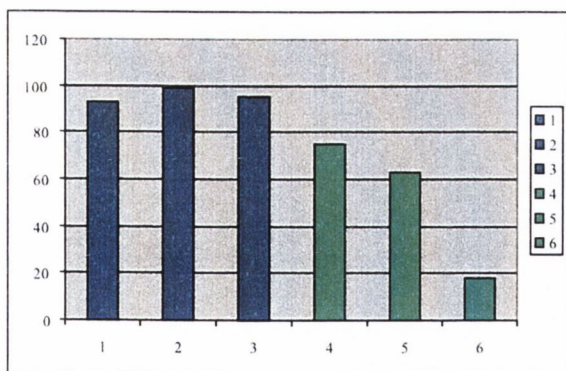


Figure 5.9A Northern blot analysis of the ‘housekeeping’ gene Basic Transcription factor 3 (BTF3) in MIN and CIN sporadic colorectal cancer cell lines. An illustration of BTF3 Northern blot analysis of the panel of sporadic colorectal cancer cell lines. Northern blot procedures were carried out as described (Section 2.3.9). The expected approximate signal of 0.857 kb was obtained as judged by the migration of the 28S and 18S ribosomal RNA bands. Expression signal intensities of BTF3 were estimated using densitometry and normalised to the expression levels of the ‘housekeeping’ gene β -actin (Figure 5.3). It is clear that BTF3 exhibits elevated expression in the MIN cell lines LS174T, SW48 & HCT 116. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480 & 6=SW620.



No.	Cell Line	Normalised Expression Value
1	LS174T	93
2	S48	99
3	HCT116	95
4	Caco2	75
5	SW480	63
6	HT29	18

Figure 5.9B. Densitometry analysis of the BTF3 transcript. Expression of the BTF3 transcript was normalised to the expression of the ‘housekeeping’ gene β -actin. The bar chart shows the expression of BTF3 in each of the MIN and CIN colorectal cancer cell lines. The MIN cells are highlighted in navy text and the CIN cells are highlighted in green text. The normalised expression values of BTF3 in each cell line are detailed in the Table above.

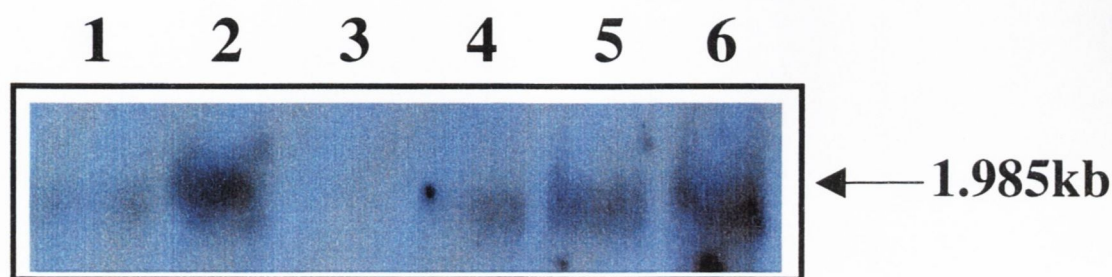
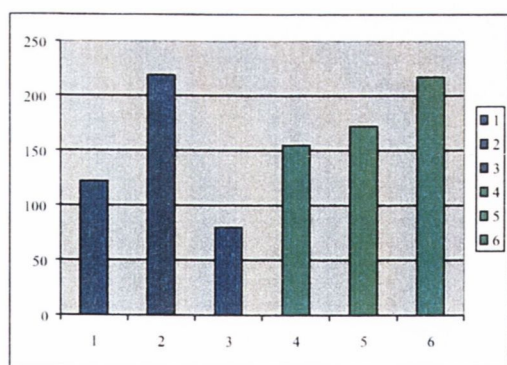


Figure 5.10A Northern blot analysis of Histone Deacetylase 2 (HDAC2) in MIN and CIN sporadic colorectal cancer cell lines. An illustration of diminished expression of HDAC2 in two out of three MIN cell lines. Northern blot procedures were carried out as described (Section 2.3.9). The expected approximate HDAC2 signal of 1.985 kb was obtained as judged by the migration of the 28S and 18S ribosomal RNA bands. Expression signal intensities of HDAC2 were estimated using densitometry and normalised to the expression levels of the ‘housekeeping’ gene β -actin (Figure 5.3). Decreased levels of HDAC2 are evident in the MIN Cell lines LS174T and HCT116. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480 & 6=SW620.



No.	Cell Line	Normalised Expression Value
1	LS174T	122
2	S48	219
3	HCT116	80
4	Caco2	155
5	SW480	172
6	HT29	217

Figure 5.10B. Densitometry analysis of the HDAC2 transcript. Expression of the HDAC2 transcript was normalised to the expression of the ‘housekeeping’ gene GAPDH. The bar chart shows the expression of HDAC2 in each of the MIN and CIN colorectal cancer cell lines. The MIN cells are highlighted in navy text, and the CIN cells are highlighted in green text. The normalised expression values of HDAC2 in each cell line are detailed in the table above.

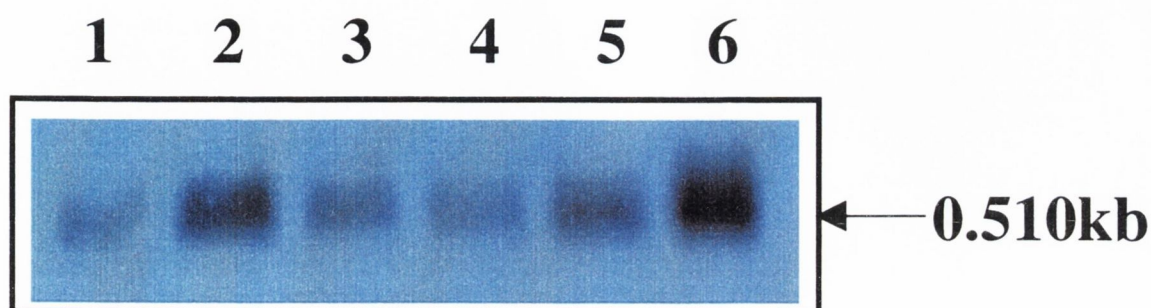


Figure 5.11A Northern blot analysis of ribosomal protein S19 (RPS19) in MIN and CIN sporadic colorectal cancer cell lines. An illustration of the expression of RPS19 in MIN and CIN sporadic colorectal cancer cell lines by Northern blot analysis. Northern blot procedures were carried out as described (Section 2.3.9). The expected approximate signal of 0.510 kb was obtained compared to the migration of the 28S and 18S ribosomal RNA bands. Expression levels of RPS19 were measured by densitometry and normalised to the levels of the ‘housekeeping’ gene GAPDH. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480 & 6=SW620

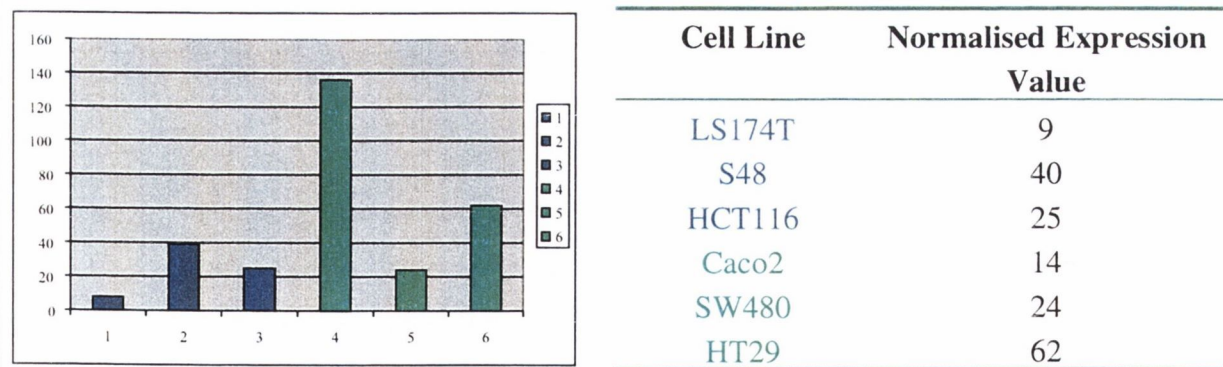


Figure 5.11B. Densitometry analysis of the RPS19 transcript. Expression of the RPS19 transcript was normalised to the expression of the ‘housekeeping’ gene GAPDH. The bar chart shows the expression of RPS19 in each of the MIN and CIN colorectal cancer cell lines. The MIN cells are highlighted in navy text, and the CIN cells are highlighted in green text. The normalised expression values of RPS19 in each cell line are detailed in the Table above.

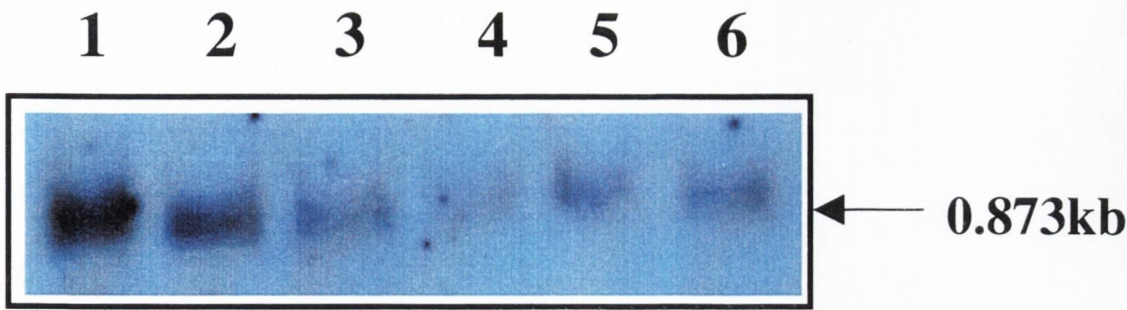
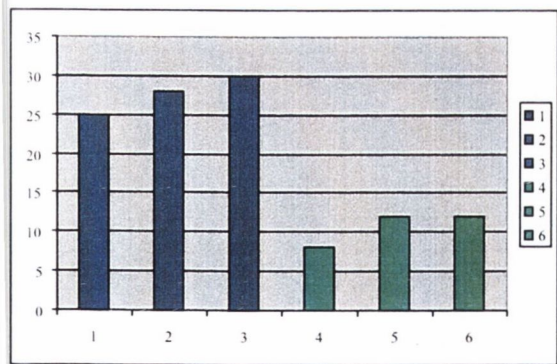


Figure 5.12A Northern blot analysis of the Histone 2AZ (H2AZ) in MIN and CIN sporadic colorectal cancer cell lines. An illustration of H2AZ Northern blot analysis of the panel of sporadic colorectal cancer cell lines. Northern blot procedures were carried out as described (Section 2.3.9). The expected approximate signal of 0.873 kb was obtained compared to the migration of the 28S and 18S ribosomal RNA bands. Expression signal intensities of H2AZ were estimated using densitometry and normalised to the expression levels of GAPDH. The expression of H2AZ is elevated in the MIN cell lines LS174T, SW48 & HCT116. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480 & 6=SW620.



No.	Cell Line	Normalised Expression Value
1	LS174T	25
2	S48	28
3	HCT116	30
4	Caco2	8
5	SW480	12
6	HT29	12

Figure 5.12B. Densitometry analysis of the H2AZ transcript. Expression of the H2AZ transcript was normalised to the expression of the 'housekeeping' gene GAPDH. The bar chart shows the expression of H2AZ in each of the MIN and CIN colorectal cancer cell lines. The MIN cells are highlighted in navy text, and the CIN cells are highlighted in green text. The normalised expression values of H2AZ in each cell line are detailed in the table above.

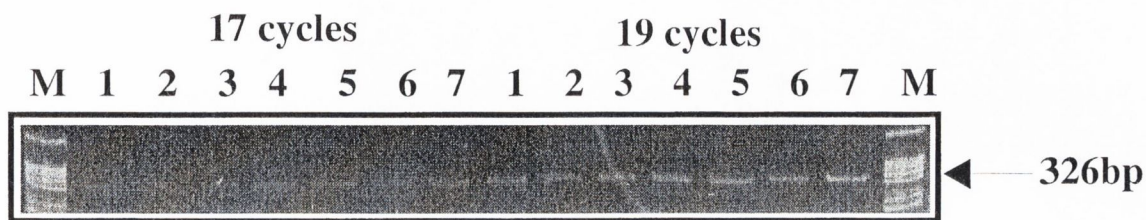


Figure 5.13 Semi-quantitative RT-PCR analysis of the candidate gene Cytokeratin 8 (Cyt8) in MIN and CIN cell lines. RT-PCR analysis was carried out using specific human Cyt8 primers on the MIN and CIN sporadic colorectal cancer cell lines. PCR products were electrophoresed on 12% non-denaturing polyacrylamide gels and stained in ethidium bromide. This gel shows Cyt8 PCR products sampled after 17 and 19 PCR cycles. The expected product size of 157bp was obtained when compared to the molecular weight marker pBluescript-*HaeIII*. M= pBluescript-*HhaI* molecular weight marker; 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480; 6=HT28 & 7=SW620.

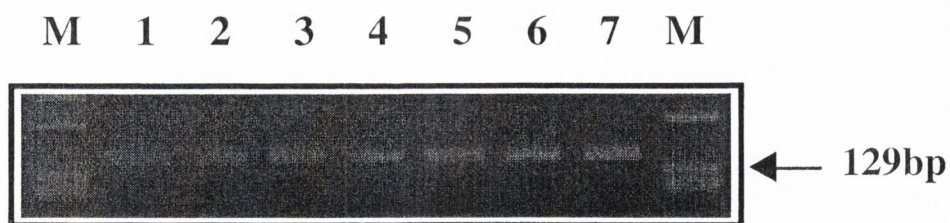


Figure 5.14 Semi-quantitative RT-PCR analysis of the candidate gene Cytokeratin 18 (Cyt18) in MIN and CIN cell lines. RT-PCR analysis was carried out using specific human Cyt18 primers on the MIN and CIN sporadic colorectal cancer cell lines. PCR products were electrophoresed on 12% non-denaturing polyacrylamide gels and stained in ethidium bromide. This gel shows Cyt18 PCR products sampled after 28 PCR cycles. The expected product size of 129bp was obtained when compared to the molecular weight marker pBluescript-*Hae*III. M= pBluescript-*Hae*III molecular weight marker; 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480; 6=HT28 & 7=SW620.

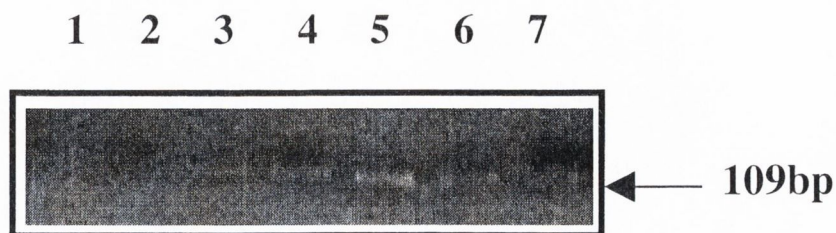


Figure 5.15 Semi-quantitative RT-PCR analysis of the candidate gene **Inhibitor 2 of Protein Phosphatase I (IPP2)** in MIN and CIN cell lines. RT-PCR analysis was carried out using specific human IPP2 primers on the MIN and CIN sporadic colorectal cancer cell lines. PCR products were electrophoresed on 12% non-denaturing polyacrylamide gels and stained in ethidium bromide. This gel shows IPP2 PCR products sampled after 28 PCR cycles. The expected product size of 109bp was obtained when compared to the molecular weight marker pBluescript-*Hae*III. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480; 6=HT28 & 7=SW620.

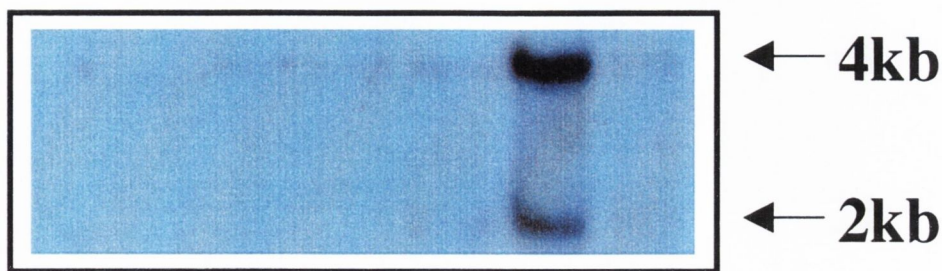
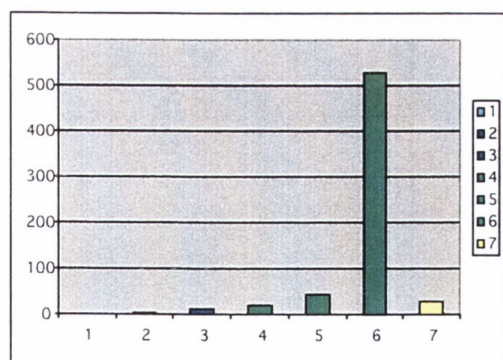


Figure 5.16A Northern blot analysis of Inhibitor 2 of Protein Phosphatase I (IPP2) in MIN and CIN sporadic colorectal cancer cell lines. An illustration of Northern blot analysis of the panel of sporadic colorectal cancer cell lines. Northern blot analysis was carried out as described (Section 2.3.9). The IPP2 probe hybridised to two IPP2 transcripts of approximately 2kb and 4kb. Highest expression of IPP2 was detected in HT29. The overall expression of IPP2 is higher in the CIN cell lines (Caco2, SW480, HT29 & SW620) than in the MIN cell lines (LS174T, SW48 & HCT116). 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480; 6=SW620 & 7=HT29.



Cell Line	Normalised Expression Value
LS174 T	0
S48	6
HCT116	12
Caco2	20
SW480	44
HT29	529
SW620	36

Figure 5.16B. Densitometry analysis of the IPP2 transcript. Expression of the IPP2 transcript was normalised to the expression of the ‘housekeeping’ gene GAPDH. The bar chart shows the expression of IPP2 in each of the MIN and CIN colorectal cancer cell lines. The MIN cells are highlighted in navy text, and the CIN cells are highlighted in green text. SW620 was Included in this experiment and is highlighted in yellow text. The normalised expression values of IPP2 in each cell line are detailed in the table above.

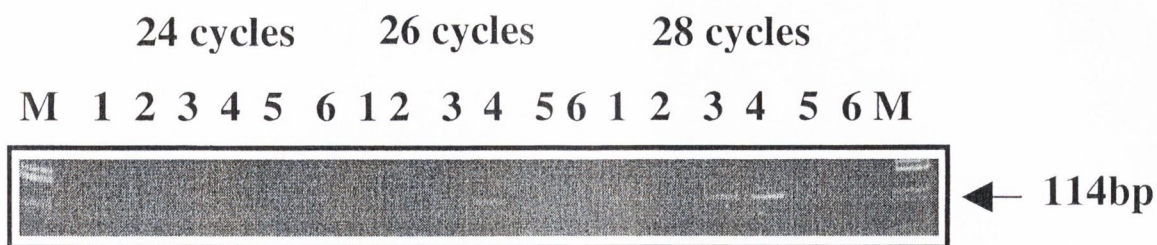


Figure 5.17 Expression of cyclooxygenase 2 (COX2) in MIN and CIN sporadic colorectal cancer cell lines. An illustration of a COX2 semi-quantitative RT-PCR on MIN and CIN sporadic colorectal cancer cell lines. COX2 RT-PCR products (114bp) were electrophoresed in parallel on 12% polyacrylamide gels and stained in ethidium bromide. RT-PCR products were sampled at 24, 26 & 28 cycles. Expression of the COX2 transcript was detected in LS174T, HCT116 & Caco-2. M= pBS-*Hae* III molecular weight marker; 1= LS174T; 2= SW48; 3= HCT116; 4= Caco-2; 5= SW480 & 6= SW620.

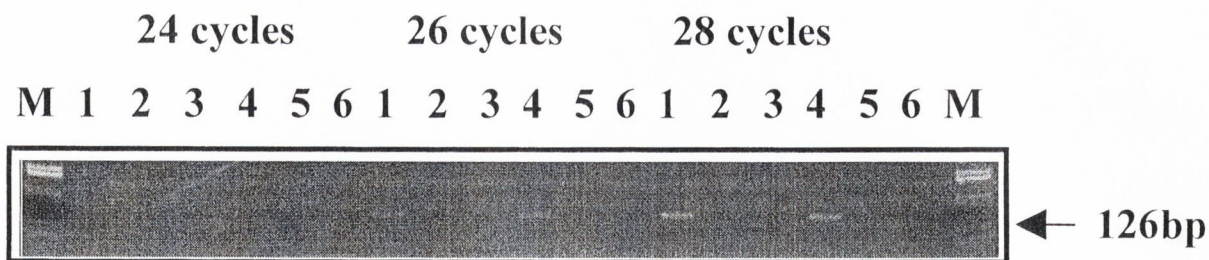


Figure 5.18 Expression of secretory phospholipase A2 (SPLA2) in MIN and CIN sporadic colorectal cancer cell lines An illustration of a sPLA2 semi-quantitative RT-PCR on MIN and CIN sporadic colorectal cancer cell lines. SPLA2 RT-PCR products (126bp) were electrophoresed in parallel on 12% polyacrylamide gels and stained in ethidium bromide. RT-PCR products were sampled at 24, 26 & 28 cycles. Expression of the SPLA-2 transcript was detected in LS174T and Caco-2. M= pBS-*Hae* III molecular weight marker; 1= LS174T; 2= SW48; 3= HCT116; 4= Caco-2; 5= SW480 & 6= SW620.

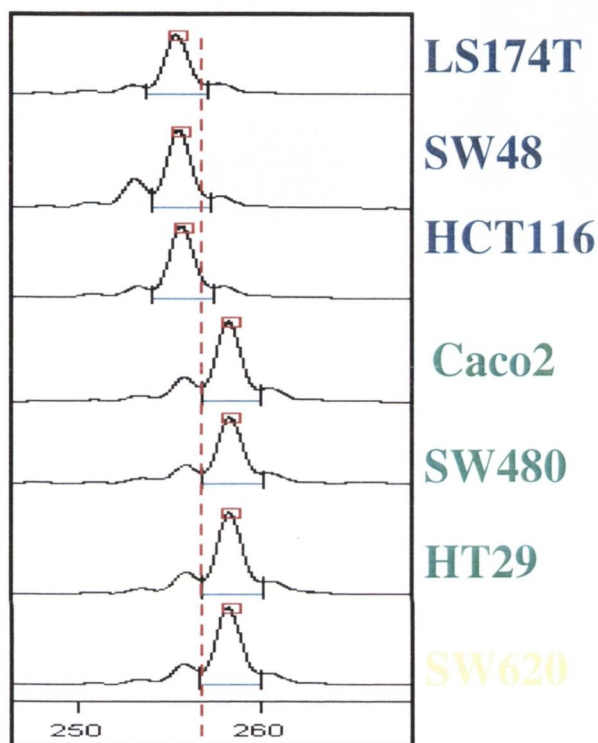


Figure 5.19 Analysis of instability in the TGF β RII transcript in MIN and CIN colorectal cancer cells. An illustration of a fluorogram generated by instability analysis of the TGF β RII transcript. RT-PCR products flanking the 8bp poly-A tract in TGF β RII were electrophoresed on 15% denaturing polyacrylamide sequencing gels and fluorograms were generated using Allele Links Version 1.00. MIN cell lines are highlighted in navy text and CIN cell lines in green text. SW620 was included in this experiment and is highlighted in yellow text. The dashed red line shows the difference in mobility of the TFG β RII PCR product between MIN and CIN cells.

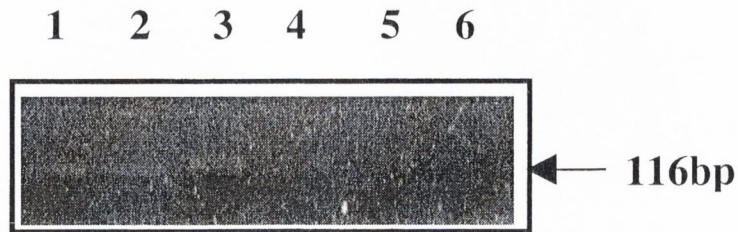


Figure 5.20 Semi-quantitative RT-PCR analysis of the candidate gene **Protein Tyrosine Phosphatase D1 (PTPD1)** in MIN and CIN cell lines. RT-PCR analysis was carried out using specific human PTPD1 primers on the MIN and CIN sporadic colorectal cancer cell lines. PCR products were electrophoresed on 12% non-denaturing polyacrylamide gels and stained in ethidium bromide. This gel shows PTPD1 RT-PCR products sampled after 28 cycles. The expected product size of 116bp was obtained when compared to the molecular weight marker pBluescript-*HhaI*. 1=LS174T; 2=SW48; 3=HCT116; 4=Caco2; 5=SW480 & 6=HT29.

Table 5.1A GDA analysis genes selected for further analysis. A summary of the genes consistently differentially expressed between MIN and CIN cells selected for further analysis. The expression profile predicted by GDA analysis for each transcript is listed in Column 2. The techniques (RT-PCR analysis or Northern blot analysis) used to confirm expression difference is given in Column 3. The results of the confirmatory analyses are given in Column 4.

GDA Gene	GDA analysis	Confirmatory technique	Confirmation
PLK	CIN +	RT-PCR	Yes
RANBP2	CIN +	RT-PCR	Yes
PGER2	CIN +	RT-PCR	No
CCNA2	CIN +	RT-PCR	Yes
BTF3	MIN +	Northern blot	Yes
HDAC2	CIN+	Northern blot	Yes*
RPS19	MIN +	Northern blot	No
H2AZ	MIN +	Northern blot	Yes
MAD2	MIN +	Northern blot	No
PTPD1	MIN +	RT-PCR	Yes

*Northern analysis of this gene resulted in diminished HDAC2 expression in MIN cells with the exception of SW48. This suggests inactivation of HDAC2 in a subset of MIN colorectal cancers.

Table 5.1B Expression analysis of candidate genes. A summary of the candidate genes selected for gene expression analysis. The technique used to confirm expression differences is listed in Column 2 and the expression profiles are given in Column 3.

Gene	Technique	Expression Profile
Cyt8	RT-PCR	CIN +
Cyt18	RT-PCR	No difference
IPP2	RT-PCR, Northern blot	CIN +
COX2	RT-PCR	Mixed*
SPLA2	RT-PCR	Mixed*

* Expression of COX2 and SPLA2 did not yield exclusive MIN+ or CIN+ expression (Section 5.1.2.2)

CHAPTER 6

ANALYSIS OF IPP2 AND IPP2

ASSOCIATED GENES

6.0 INTRODUCTION

6.0.1 Differential Display Reverse Transcription PCR

Differential Display Reverse Transcription PCR (DDRT-PCR) (See Section 1.7) (Liang *et al.*, 1992) was previously used (McWilliam *et al.*, MS in preparation) to identify gene expression differences in the isogenic pair of colorectal cancer cell lines SW480 (pre-metastatic) and SW620 (metastatic)[Table 1.2]. DDRT-PCR is a technique that yields expression differences between mRNA populations. In brief DDRT-PCR involves the isolation of high quality total RNA (or mRNA) from experimental tissue, reverse transcription primed with an anchored oligo-dT primer, subsequent PCR amplification with the anchor primer and an arbitrary upstream primer. PCR products are electrophoresed on polyacrylamide sequencing gels and putative difference products are excised. Differences products are reamplified by PCR, cloned and sequenced. A schematic of the protocol is illustrated in Figure 3.1. The principal advantage of DDRT-PCR is the facility to analyse multiple sets of difference products in parallel (up to forty samples on a sequencing gel). Furthermore, the technique generates clones for further analysis and is relatively rapid. However, the technique is technically demanding, operator-sensitive and is associated with a high false positive rate (Liang *et al.*, 1992; Bauer *et al.*, 1993; Dunican *et al.*, 1997; McWilliam *et al.*, MS in preparation). A previous study in our laboratory identified overexpression of Inhibitor 2 of Protein Phosphatase 1 (IPP2) in SW480 cells (premetastatic) compared to SW620 cells (metastatic) by DDRT-PCR (McWilliam *et al.*, MS in preparation).

6.0.2 Protein Phosphatase 1 (PPP1C-Beta)

PPP1C-Beta regulates many biological processes including synaptic plasticity, cell cycle, gene transcription and carbohydrate/lipid metabolism. Four mammalian isoforms of the PPP1C subunit are generated from three genes. With the exception of PPP1C-Gamma2, which is expressed predominantly in testes, the other isoforms, PPP1C-Alpha, Beta and Gamma1, are widely expressed in mammalian tissues. As the levels of the PPP1 catalytic subunit (PPP1C) do not change in response to physiological stimuli, regulation is thought to occur through endogenous inhibitors (Brautigan *et al.*, 1990) (Section 6.0.3). One of the best examples of PPP1C regulation operates during

operates during the control of the cAMP-response binding protein (CREB). Protein kinase A (PKA) activates CREB by phosphorylating CREB-serine₁₃₃ to activate it as a transcription factor. Active CREB is dephosphorylated by CREB phosphatase (identified as nuclear PP1) to control the transcription of a variety of genes (Hagiwara *et al.*, 1992). Treatment of cells with okadaic acid, a potent phosphatase inhibitor, leads to prolonged phosphorylation of CREB. Similarly, the expression of IPP1 was found to promote CREB phosphorylation and gene transcription in the absence of elevated second messenger levels (Alberts *et al.*, 1994). These findings raise the possibility that inhibitors of PPP1C activity may be important mediators of signal transduction in the nucleus.

It has been established that the Retinoblastoma (RB) gene product must be dephosphorylated as cells exit mitosis. This dephosphorylation process activates the function of RB as a growth suppressor that blocks cells in the G1 phase of the cell cycle (Sherr, 1996). This was precisely shown by expressing mutant PPP1C in cancer cells which blocked their entry into S phase (Berndt *et al.*, 1997). As PPP1C levels do not change during the cell cycle, it is presumed that cell cycle-dependent changes in PPP1C activity must be mediated by either covalent modification of PPP1C or by endogenous PPP1C regulators. In this regard, a nuclear PPP1C-binding protein has been identified that may modulate the activity of PPP1C (Allen *et al.*, 1998).

6.0.3 Inhibitor 2 of Protein Phosphatase 1 (IPP2)

The heat-stable proteins IPP1 (Inhibitor 1 of Protein Phosphatase 1) and IPP2 were first isolated from rabbit skeletal muscle extracts (Huang *et al.*, 1976) and IPP2 was found to be a potent inhibitor of PPP1C (Protein Phosphatase 1) (Foulkes *et al.*, 1983). Unlike IPP1, IPP2 does not need to be phosphorylated to inhibit PPP1C but forms a stable and inactive complex with PPP1C. This inactive complex was isolated as an Mg^{2+} -ATP-dependent phosphatase which was inactive until incubated with Mg^{2+} -ATP (Park *et al.*, 1994). Activation of the latent complex is accompanied by the phosphorylation of IPP2 on threonine₇₂ by GSK-3 (glycogen synthase kinase-3) and the phosphorylation of three serine residues by casein kinase II (CK2). Phosphorylation by CK2 does not alter IPP2 activity but greatly facilitates the subsequent phosphorylation of IPP2 by GSK3 (Depaoli-Roach, 1984; Holmes *et al.*, 1986). The fundamental aspect of Mg^{2+} -ATP-

dependent phosphatase activation is that IPP2 phosphorylation by GSK-3 promotes a conformational change in the PPP1C/IPP2 complex, which by itself does not activate the enzyme. The slow autodephosphorylation of IPP2 correlates best with the increase in PPP1C activity (Park *et al.*, 1994). A study showed that the addition of exogenous IPP2 can inhibit the activity of the active PPP1C/IPP2 complex bringing the role of IPP2 as a PPP1C inhibitor into question (Bollen *et al.*, 1994).

Recent *in vitro* studies suggest that IPP2 interacts with denatured PPP1C to promote the refolding of this protein. In addition, recombinant PPP1C behaves like native PPP1C after incubation with IPP2 and reactivation by GSK-3 (Alessi *et al.*, 1993; Mackintosh *et al.*, 1996). The expressed isoforms of PPP1C were 100-600-fold less sensitive to inhibitor-1, 3-5-fold less sensitive to okadaic acid, 5-100-fold less sensitive to microcystin-LR and approximately 20-fold more active in dephosphorylating histone H1 than native PPP1C. This has led to the proposal that IPP2 is a chaperone for PPP1C. Studies with Glc8, the yeast homologue of IPP2, suggest that under different circumstances, Glc8 was either an inhibitor or an activator of the yeast PPP1C (Tung *et al.*, 1995).

Perhaps the most intriguing property of IPP2 is that its mRNA and protein levels peak twice during the cell cycle (Brautigan *et al.*, 1990). Studies have shown that IPP2 is cytosolic during G1 and translocates to the nucleus during S phase (Kakinoki *et al.*, 1997). Further investigation of IPP2 has identified four phosphorylation sites in IPP2 that are required for nuclear translocation.

6.0.4 Identification of IPP2 by DDRT-PCR

DDRT-PCR was carried out as described in Chapters 2 & 3 by Dr. Peter McWilliam on the isogenic colorectal cancer cell lines SW480 (premetastatic) and SW620 (metastatic). The cells were cultured as described in Table 2.1. A number of putative differentially expressed DDRT-PCR cDNAs were isolated, cloned and sequenced between the colorectal cancer cell lines SW480 and SW620 (McWilliam *et al.*, MS in preparation). In particular, clone PM208 showed increased expression in the primary tumour cell line (SW480) compared to the metastatic cell line (SW620). Clone PM208 was sequenced in both directions resulting in a 560bp product and the sequence was deposited to GenBank under the accession number Y13810 (McWilliam *et al.*, MS in preparation). Through a combination of *in silico* bioinformatic analysis and sequencing, PM208 was identified as Inhibitor 2 of Protein Phosphatase 1 (IPP2) (McWilliam *et al.*, MS in preparation) (Figure 6.1).

6.1 METHODS

6.1.1 DDRT-PCR and Semi-Quantitative RT-PCR Analysis

RT-PCR analysis was used to confirm the differential expression of clone PM208 (IPP2 3'-untranslated region) and DDP (IPP2 coding region) in SW480 and SW620. The study was extended to the transcripts for Protein Phosphatase 1 Catalytic Subunit Isoform Alpha (PPP1C-Alpha), Protein Phosphatase 1 Catalytic Subunit Isoform Beta (PPP1C-Beta) and Protein Phosphatase 1 Catalytic Subunit Isoform Gamma (PPP1C-Gamma) and the additional pair of isogenic cell lines SW1116 (premetastatic) and 498Li (metastatic). RT-PCR analysis was carried out as described in Chapter 5. Specific primers for each transcript were designed using the Oligo 4.0 primer design software (Table 2.2). PCR reactions were optimised prior to RT-PCR analysis on genomic DNA or cDNA where suitable. The optimal conditions for each PCR are shown in Table 2.3. RT-PCR experiments were controlled using the 'housekeeping' genes β -actin and GAPDH.

Semi Quantitative RT-PCR

Primers were designed specific for the human β -actin and GAPDH transcripts using the Oligo 4.0 program. The primers for β -actin and GAPDH produced RT-PCR products of 109bp and 602bp respectively. The PCR for β -actin was initially optimised on genomic DNA. Suitable amplification conditions were extended to RT-PCR analysis of the premetastatic (SW480 & SW1116) and metastatic (SW620 & 498Li) total RNA samples. β -actin experiments were repeated a number of times yielding equivalent amounts of product across the panel of experimental samples. β -actin was observed to be highly expressed in sporadic colorectal cancer cells and products were sampled after 18, 21 and 24 PCR cycles. Products were resolved on 8% polyacrylamide gels and quantitated using densitometry. The same reverse transcription reaction was used for the β -actin analysis and the target genes. An example of a β -actin RT-PCR is illustrated in Figure 6.2. This shows equivalent expression of β -actin across the four cell lines.

The GAPDH PCR was optimised on genomic DNA. The optimal PCR conditions were extended to RT-PCR on the premetastatic (SW480 & SW1116) and metastatic (SW620 & 498Li) total RNA samples. GAPDH products were separated on 8% polyacrylamide gels and are illustrated in Figure 6.3. The same reverse transcription reaction was used for the GAPDH analysis and the candidate genes. Similar to β -actin RT-PCR analysis, GAPDH was highly expressed (sampled at 21, 24 and 27 cycles) and at comparable levels between the cell lines.

6.1.2 Protein expression Analysis of PPP1C-Beta

A number of techniques were used to assess the expression of PPP1C-Beta including immunohistochemistry on paraffin embedded tissue, immunohistochemistry on cytospin preparations from colorectal cancer cell lines and Western blot analysis of cell line subcellular protein fractions. Paraffin-embedded colorectal cancer tissue was analysed for PPP1C-Beta expression using a polyclonal PPP1C-Beta antibody. Positive staining for PPP1C-Beta was not detected by this technique. The antibody was thus deemed unsuitable for analysis using paraffin-embedded tumour material. Cytospins were prepared from the cell lines SW480, SW620, SW1116 and 498Li. Low positive

staining was obtained for PPP1C-Beta in the metastatic cell lines SW1116 and 498Li (data not shown). However, positive staining also appeared in the negative control which may be due to endogenous biotin in the cells (Tony O' Grady, pers. communication). The cytospin studies were discontinued as a result.

Western blot analysis was used to measure PPP1C-Beta protein expression in the cell lines SW480, SW620, SW1116 and 498Li. The procedure was carried out as described in Chapter 2, Section 2.4. In short, cells were cultured under standard conditions (Table 2.1). Protein was isolated and subcellular fractionation was carried out resulting in the isolation of membrane-rich, cytosol-rich and cytoskeletal-rich subcellular fractions. Protein was quantitated using the Bradford or Lowry Assays and electrophoresed on 10% denaturing SDS-polyacrylamide gels. Protein was transferred to membranes using semi-dry transfer. Protein was detected using a Horseradish Peroxidase coupled polyclonal PPP1C-Beta antibody at a dilution of 1/400 by chemiluminescence.

6.2 RESULTS

6.2.1 SEMI-QUANTITATIVE RT-PCR ANALYSIS OF IPP2

DDT

Oligonucleotide primers were designed to specifically amplify the human DDT transcript [A full contig for the IPP2 transcript is shown in Figure 6.1]. The DDT primers were designed from the original DDRT-PCR clone PM208 and are from the 3'-untranslated region of the IPP2 gene. The PCR product generated is 168bp in size from bases 2736 to 2903 of the GenBank entry AJ133812 (See Table 2.4 for PCR conditions). Semi-quantitative RT-PCR experiments were carried out on this transcript. RT-PCR products were electrophoresed on 8% nondenaturing polyacrylamide gels and stained in ethidium bromide. An illustration of DDT RT-PCR analysis is shown in Figure 6.4. Semi-quantitative RT-PCR showed elevated expression of DDT in both pre-metastatic cell lines SW480 and SW1116 with respect to their metastatic counterparts SW620 and 498Li. The original DDRT-PCR analysis detected raised expression of PM208 (DDT) in SW480 compared to SW620. This independent analysis has thus confirmed the observation with the extension of the phenomenon to a second pair of isogenic colorectal cancer cell lines.

DDP

Specific primers for the DDP (IPP2 open reading frame) transcript were designed using the Oligo 4.0 package. PCR amplification using the DDP primers yields a PCR product of 109bp between bases 589 and 697 of the GenBank entry NM_006241. RT-PCR was carried out using optimal DDP PCR conditions (Table 2.4) and DDP products were resolved on 8% nondenaturing polyacrylamide gels stained in ethidium bromide. As with DDT, elevated expression of DDP was observed in both pre-metastatic cell lines (SW480 & SW1116) compared to the metastatic cell lines (SW620 & 498Li). A typical DDP RT-PCR experiment is illustrated in Figure 6.5. This result confirms the data observed for clone PM208 (DDT) following DDRT-PCR analysis. In addition, the data confirm the differential expression of the functionally important coding region of

the DDP (IPP2) transcript. Moreover, the IPP2 expression change is extended to a second isogenic pair of colorectal cancer cell lines (SW1116 & 498Li).

6.2.2 SEMI-QUANTITATIVE RT-PCR ANALYSIS OF CANDIDATE GENES

IPP2 seems to act as both an inhibitor of and chaperone protein for the Protein Phosphatase Catalytic subunit (PPP1C). We explored the role of IPP2 in colorectal cancer metastasis by extending the study to the three isoforms of PPP1C: PPP1C-Alpha, PPP1C-Beta and PPP1C-Gamma. We examined the expression of these transcripts by RT-PCR in the cell lines SW480, SW620, SW1116 and 498Li.

PPP1C-Alpha

The Oligo 4.0 primer design software was used to design specific primers for the human PPP1C-Alpha transcript. These primers generate a 137bp PCR product between bases 1107 and 1243 of the GenBank entry NM_002708. RT-PCR analysis was carried out on PPP1C-Alpha, β -actin and GAPDH. A PPP1C-Alpha RT-PCR experiment is illustrated in Figure 6.6. Elevated expression of PPP1C-Alpha is clear in SW480 with respect to SW620. However, differential expression of PPP1C-Alpha was not clearly demonstrated between SW1116 and 498Li.

PPP1C-Beta

Specific primers were designed for the human PPP1C-Beta transcript. These primers generate a 195bp PCR product between bases 1402 and 1596 of the GenBank entry NM_002709. RT-PCR analysis was carried out on PPP1C-Beta, β -actin and GAPDH. A PPP1C-Beta RT-PCR experiment is illustrated in Figure 6.7. RT-PCR analysis revealed elevated expression of the PPP1C-Beta transcript in both pre-metastatic cell lines (SW480 and SW1116) compared to the metastatic cell lines SW620 and 498Li. The analysis appeared to reveal a larger differential in PPP1C-Beta expression between SW1116 and 498Li than in SW480 and SW620. This results may indicate a generalised overexpression of the PPP1C-Beta transcript in premetastatic colorectal cancer.

PPP1C-Gamma

The Oligo 4.0 primer design software was used to design specific primers for the human PPP1C-Gamma transcript. These primers generate a 149bp PCR product between bases 1311 and 1459 of the GenBank entry NM_002710. RT-PCR analysis was carried out on PPP1C-Gamma, β -actin and GAPDH. A PPP1C-Gamma RT-PCR experiment is illustrated in Figure 6.8. Elevated expression of PPP1C-Gamma is clear in SW1116 with respect to 498Li. However, differential expression of PPP1C-Gamma was not clearly demonstrated between SW480 and SW620.

6.2.3 Analysis of PPP1C-Beta protein expression

6.2.3.1 PPP1C-Beta protein expression in paraffin-embedded tissue

A polyclonal PPP1C-Beta antibody was obtained (Santa Cruz Biotech., USA) and used in immunohistochemical experiments. Immunohistochemistry was carried out by Tony O' Grady as described in Section 2.4.7. Analysis was carried out on paraffin-embedded tissue sections from Dukes' B colorectal cancers. Positive staining for the antibody was not immediately evident. Possible staining was observed in extracellular regions. As PPP1C-Beta is not recognised as a secretory protein this result was deemed a false positive and the experiments were discontinued.

6.2.3.2 PPP1C-Beta protein expression in cytospin preparations

Paraffin-embedded tissue was deemed unsuitable for analysis with the PPP1C-Beta antibody (Section 6.2.2.1). The next step was to examine protein expression in the pairs of isogenic cell lines. Cytospins were prepared from the cell lines SW620 and 498Li as described (Section 2.4.7) and analysed immunohistochemically with anti-PPP1C-Beta. A low frequency of positive staining was observed in cytospins from both cell lines. However, positive staining was also observed in the negative control [cytospin analysed without primary antibody]. This may have been due to endogenous biotin in the cell lines analysed (Tony O' Grady, pers. communication). Cytospin protein expression analysis was discontinued as a result.

6.2.3.3 PPP1C-Beta Western Blot Analysis

Subcellular fractionation was carried out as described in Section 2.4.1 and protein lysates were electrophoresed on 10% denaturing polyacrylamide gels. This facilitated protein analysis in three cellular environments: the cytosol, the membrane and the cytoskeleton. An example of protein electrophoresis is shown in Figure 6.9. Protein concentrations were estimated using the Lowry assay and the Bradford assay. A typical standard curve using Bovine Serum Albumin is illustrated in Figure 6.10. Western blot analysis was carried out as described (Sections 2.4.5 & 2.4.6). Several Western blots were rinsed in Ponceau S solution showing successful protein transfer (data not shown). The PPP1C-Beta antibody was initially used at a dilution of 1:400. A Western blot optimisation experiment is shown in Figure 6.11. PPP1C-Beta is a 38 kDa protein and the expected protein band was observed to co-migrate with the 37.5 kDa protein standard.

Western blot analysis was designed to investigate the three subcellular fractions mentioned above for the cell lines: SW480, SW620, SW1116 & 498Li. Elevated expression of PPP1C-Beta was observed in the pre-metastatic cell lines SW480 and SW1116 with respect to their metastatic counterparts SW1116 and 498Li in all subcellular fractions (Figure 6.12). The most marked differential was observed in the cytoskeletal fractions. Interestingly, the cytosolic PPP1C-Beta luminescent signals migrated marginally further than in the other fractions. This may be due to specific dephosphorylation of PPP1C-Beta in the cytosol or a post-translational modification of PPP1C-Beta. A PPP1C study revealed two protein products (37kDa and 39kDa) which react with PPP1C-Beta antibody (Takizawa *et al.*, 1994). The authors suggest that the 37kDa fragment is a proteolysis product of PPP1C-Beta which has lost a short peptide fragment but retains the antigenic site. This finding is consistent with our observation of two strong PPP1C-Beta fragments by Western blot analysis.

6.3 DISCUSSION: ROLES OF IPP2 AND PPP1C-BETA IN COLORECTAL CANCER METASTASIS

6.3.1 Role of PPP1C-Beta in Colorectal Cancer Metastasis

Protein phosphatases regulate many diverse processes in cells. Four isoforms of the catalytic subunit have been isolated: PPP1C-Alpha, PPP1C-Beta (also termed PPP1C-Delta), PPP1C-Gamma1 and PPP1C-Gamma2. Regulation of the PPP1C family is thought to be controlled by endogenous inhibitors as the levels of PPP1C are largely unchanged under the influence of physiological stimuli (Brautigan *et al.*, 1990). We have shown underexpression of PPP1C-Beta by RT-PCR in the metastatic cell lines SW620 and 498Li compared to the premetastatic cell lines SW480 and SW1116. In addition, we have demonstrated a similar expression pattern for the PPP1C-Beta protein in the cytosol-rich and cytoskeletal-rich subcellular fractions in the cell lines examined.

Recent studies have demonstrated a RhoA-mediated pathway which determines the status and organisation of actin-myosin bundles (Maier *et al.*, 1995; Kimura *et al.*, 1996; Yoshioka *et al.*, 1999). This pathway begins with the activation of RhoA-GDP to RhoA-GTP. Active RhoA-GTP interacts with and activates ROCK (RhoA-Associated Kinase) which regulates the myosin phosphatase complex by dephosphorylating the MBS (Myosin Binding Subunit of myosin phosphatase) (Kimura *et al.*, 1996). The active form of myosin phosphatase dephosphorylates MLC (Myosin Light Chain) thereby blocking polarisation of actin-myosin bundles (Figure 6.13).

Non-metastatic Lewis lung carcinomas cells (LLC) become more motile when PPP1C and PP-2A are inhibited by okadaic acid (Maier *et al.*, 1995). Non-metastatic LLC are adherent and exhibit spread morphology in culture. Addition of okadaic acid leads to contraction of extended processes, formation of spikes and membrane blebs. PP1 and PP2A maintain cytoskeletal organisation and disruption of this regulation reduces cytoskeletal integrity and increases tumour cell motility. Overexpression of small GTP-binding protein RhoA promotes invasion of mesothelial cells *in vitro*, and overexpression of wild-type RhoA conferred invasive ability *in vivo* in syngeneic rats (Yoshioka *et al.*, 1999). Furthermore, an inhibitor of ROCK effectively inhibits the invasion of MM1 cells *in vitro* and *in vivo* (Itoh *et al.*, 1999).

The RhoA-mediated regulation of cytoskeletal architecture is beginning to be unravelled. The decreases in expression of PPP1C-Beta at the mRNA and protein levels shown here would suggest a role for this gene in colon cancer metastasis. The RhoA pathway may be compromised by the downregulation of PPP1C-Beta. We argue that a decrease in expression of PPP1C-Beta may reduce activity of the myosin phosphatase complex, of which PPP1C-Beta is a component. A decrease in myosin phosphatase activity would lead to a decrease in the regulation of MLC (myosin light chain) by its dephosphorylation thereby increasing Actin-Myosin bundle polarisation and the formation of stress fibres. Further support for this hypothesis is given by Villa-Moruzzi *et al.* (1998) who have demonstrated PPP1C-Beta enrichment at focal adhesions in HeLa cells, fibroblasts, endothelial cells and keratinocytes. Highest expression of PPP1C-Beta seems to occur in the cytoskeletal-rich fraction in the premetastatic and metastatic colon cancer cell lines examined (Figure 6.12) and a large differential is apparent. Current findings in the literature and our PPP1C-Beta results taken together suggest a conceivable role for PPP1C-Beta in colon cancer metastasis involving a possible role in RhoA-mediated cytoskeletal maintenance.

6.3.2 Role for IPP2 in Colorectal Cancer Metastasis

As discussed in Section 6.0.3, IPP2 was originally viewed as a highly specific inhibitor of the Protein Phosphatase Catalytic Subunit (PPP1C) (Huang *et al.*, 1976). Recent findings suggest that IPP2 can act as a chaperone protein for PPP1C (Alessi *et al.*, 1993; Tung *et al.*, 1995). The chaperone property of IPP2 is consistent with our expression data for IPP2 and PPP1C-Beta. The coding and 3'-untranslated regions of IPP2 were shown to be overexpressed in premetastatic cells lines (SW480 & SW1116) compared to metastatic cell lines (SW620 & 498Li). Examination of the expression of the PPP1C isoforms showed consistent overexpression of PPP1C-Beta in premetastatic cell lines. This may suggest that IPP2 is acting as a chaperone of PPP1C-Beta. It is possible that IPP2 is a more efficient chaperone for PPP1C-Beta than for PPP1C-Alpha and PPP1C-Gamma. Decreased expression of IPP2 is consistent with decreased expression of PPP1C-Beta in the metastatic cell lines studied. This may suggest coordinate regulation of IPP2 and PPP1C-Beta. Whether IPP2 is necessary for appropriate functioning of PPP1C-Beta in MLC dephosphorylation remains to be

clarified. The observation that PPP1C-Beta expression is decreased is consistent with this possibility.

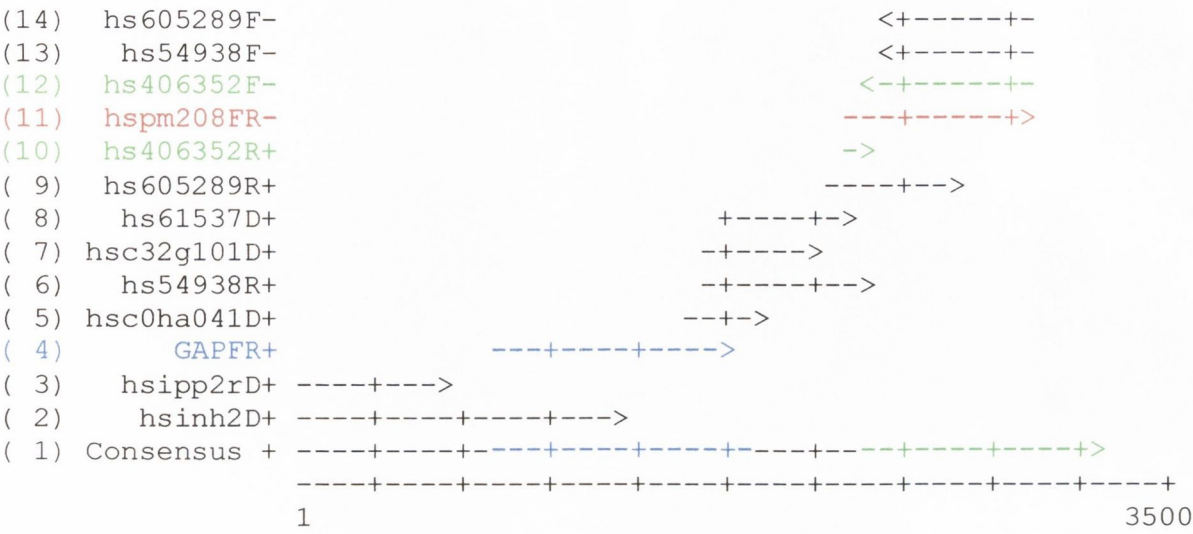


Figure 6.1 Contig construct of IPP2 transcript. The above sequence contig was assembled using the Wisconsin GCG package, Version 9.1 by McWilliam *et al.*, 2000 (MS in preparation). The complete sequence has been submitted to GenBank and assigned the accession number AJ133812. The contig consists of sequences from ESTs (Expressed Sequence Tags), IPP2 submitted entries and clones PM208 and GAP. Clones were sequenced with M13 Forward (F) and Reverse (R) primers or the sequences were obtained directly from GenBank. The sequences are designated as F, R or D as appropriate. The contig clearly shows that DDRT-PCR clone **PM208 (hspm208FR)** (shown in red) originated from a longer IPP2 sequence that extends beyond the published sequences for IPP2 (hsipp2rD and hsinh2D). The IPP2 specific probe used in the Northern blot (Figure 5.16) is highlighted in blue (**GAP**).

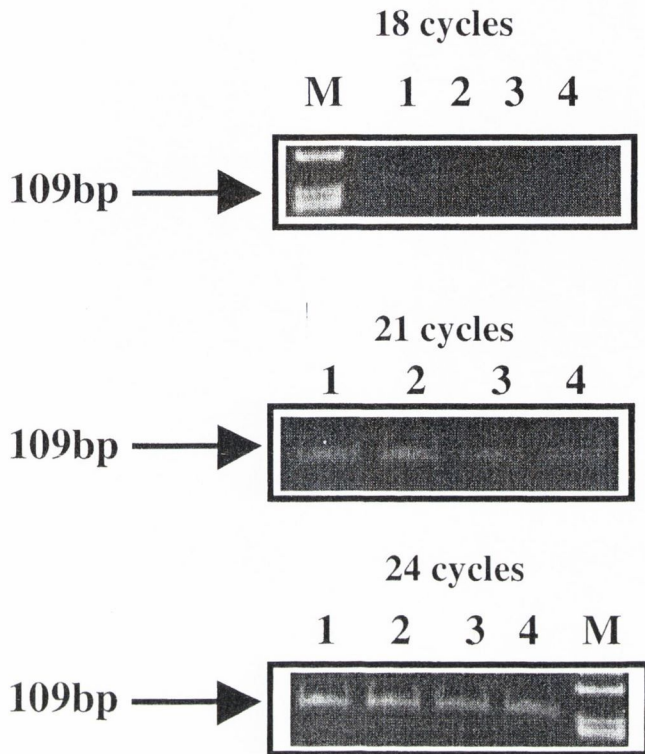


Figure 6.2 Semi-quantitative RT-PCR analysis of the ‘housekeeping’ gene β -actin in premetastatic and metastatic colorectal cancer cell lines. An illustration of a semi-quantitative RT-PCR experiment directed towards the β -actin transcript. RT-PCR reactions were carried out as described and electrophoresed on non-denaturing 8% polyacrylamide gels stained in ethidium bromide. PCR products were sampled at 18, 21 & 24 cycles as indicated. Equivalent expression of β -actin is evident in all four sporadic colorectal cancer cell lines across all samplings. M=pBluescript – *Hha* I molecular weight marker; 1=SW480; 2=SW620; 3=SW1116 & 4=498Li.

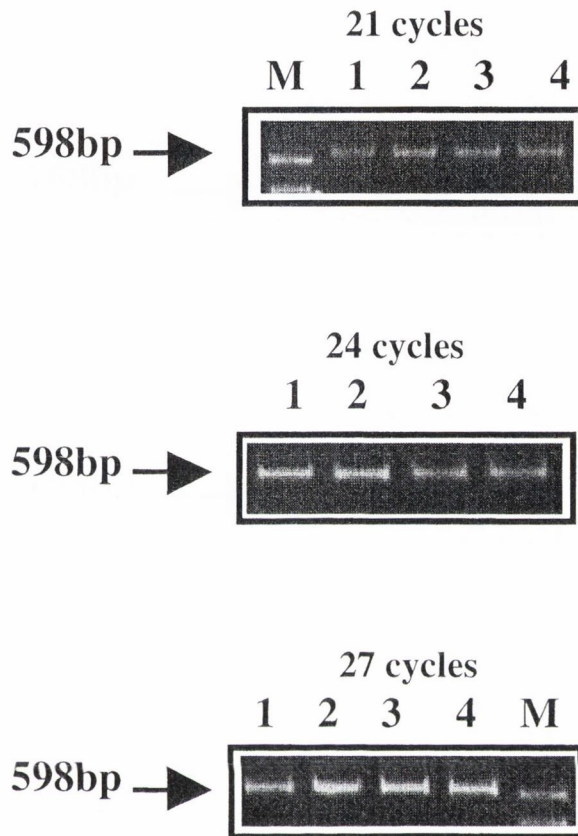


Figure 6.3 Semi-quantitative RT-PCR analysis of the ‘housekeeping’ gene **Glyceraldehyde Phosphate Dehydrogenase (GAPDH)** in **premetastatic and metastatic colorectal cancer cell lines**. An illustration of a semi-quantitative RT-PCR experiment directed towards the ‘housekeeping’ GAPDH transcript. RT-PCR reactions were carried out as described and electrophoresed on non-denaturing 8% polyacrylamide gels stained in ethidium bromide. PCR products were sampled at 21, 24 & 27 cycles as indicated. Equivalent expression of GAPDH is evident in all four sporadic colorectal cancer cell lines across all samplings. M=pBluescript – *Hha* I molecular weight marker; 1=SW480; 2=SW620; 3=SW1116 & 4=498Li.

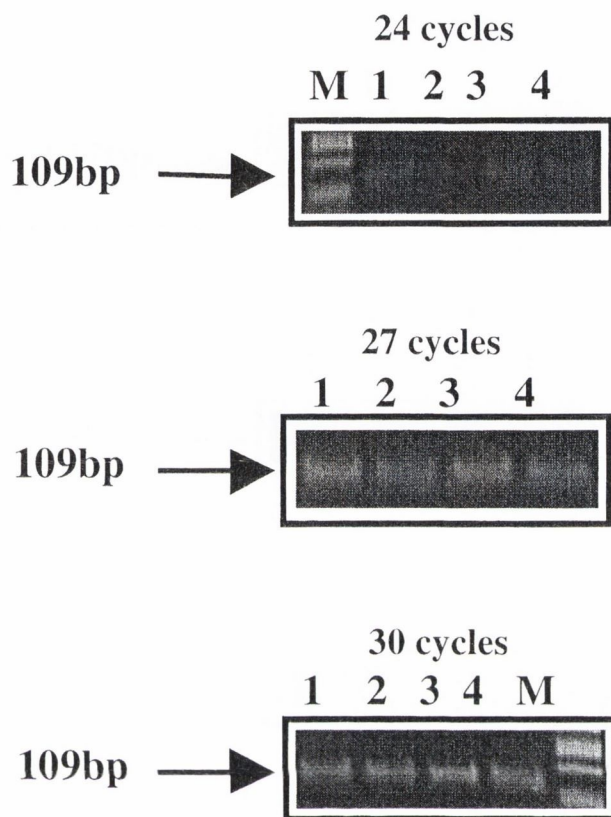


Figure 6.5 Semi-quantitative RT-PCR analysis of DDP. An illustration of a semi-quantitative RT-PCR experiment directed towards the DDP transcript (IPP2 open reading frame). RT-PCR reactions were carried out as described and electrophoresed on non-denaturing 8% polyacrylamide gels stained in ethidium bromide. PCR products were sampled at 24, 27 & 30 cycles as indicated. Elevated expression of DDP is evident in both pre-metastatic cell lines (SW480 & SW1116) with compared to the metastatic cell lines (SW620 & 498Li) across the three cycle numbers sampled. M=pBluescript – *Hha* I molecular weight marker; 1=SW480; 2=SW620; 3=SW1116 & 4=498Li.

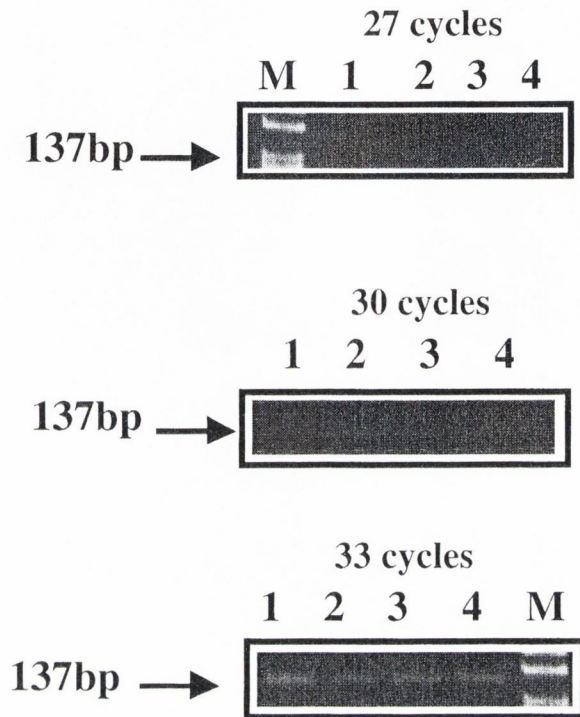


Figure 6.6 Semi-quantitative RT-PCR analysis of PPP1C-Alpha. An illustration of a semi-quantitative RT-PCR experiment directed towards the PPP1C-Alpha transcript. RT-PCR reactions were carried out as described and electrophoresed on non-denaturing 8% polyacrylamide gels stained in ethidium bromide. PPP1C-Alpha PCR products were sampled at 27, 30 & 33 cycles. Raised expression of PPP1C-Alpha in SW480 with respect to SW620 is apparent for the final cycle number sampled. PPP1C-Alpha seems to be expressed at comparable levels in the pair of cell lines SW1116 and 498Li. M=pBluescript – *Hha* I molecular weight marker; 1=SW480; 2=SW620; 3=SW1116 & 4=498Li.

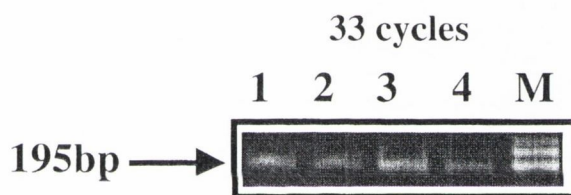
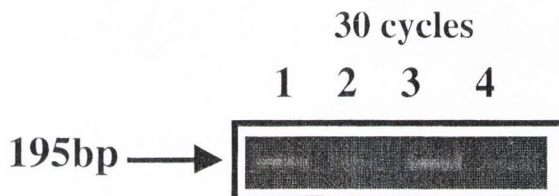
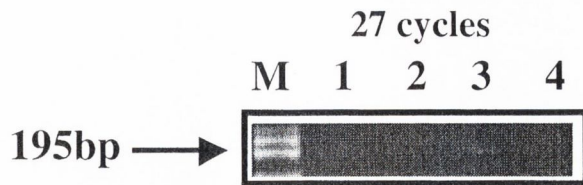


Figure 6.7 Semi-quantitative RT-PCR analysis of PPP1C-Beta. An illustration of a semi-quantitative RT-PCR experiment directed towards the PPP1C-Beta transcript. RT-PCR reactions were carried out as described and electrophoresed on non-denaturing 8% polyacrylamide gels stained in ethidium bromide. PPP1C-Beta PCR products were sampled at 27, 30 & 33 cycles. Raised expression of PPP1C-Beta in both pre-metastatic cell lines (SW480 & SW1116) is evident with respect to the metastatic cell lines SW620 and 498Li. M=pBluescript – *Hha* I molecular weight marker; 1=SW480; 2=SW620; 3=SW1116 & 4=498Li.

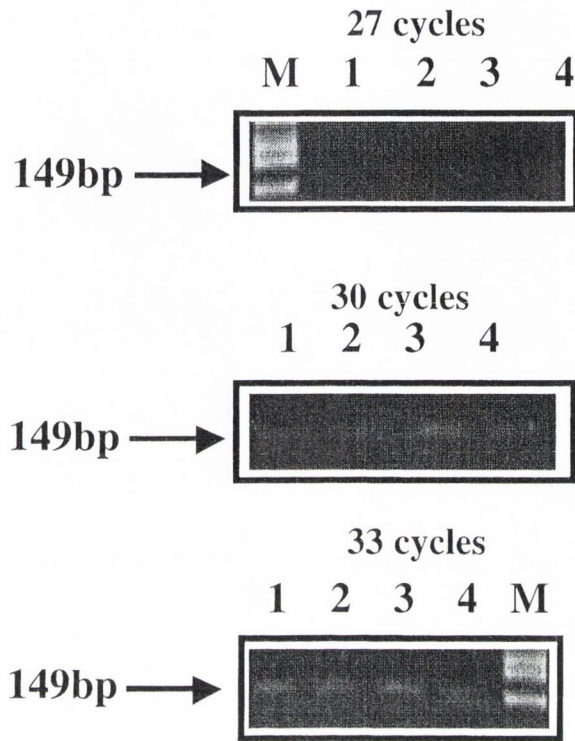
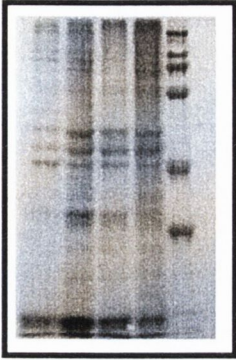


Figure 6.8 Semi-quantitative RT-PCR analysis of PPP1C-Gamma. An illustration of a semi-quantitative RT-PCR experiment directed towards the PPP1C-Gamma transcript. RT-PCR reactions were carried out as described and electrophoresed on non-denaturing 8% polyacrylamide gels stained in ethidium bromide. PPP1C-gamma PCR products were sampled at 27, 30 & 33 cycles. Raised expression of PPP1C-Gamma in SW1116 with respect to 498Li is clear for the final cycle number sampled with comparable levels in the pair of cell lines SW480 and SW620. M=pBluescript – *Hha* I molecular weight marker; 1=SW480; 2=SW620; 3=SW1116 & 4=498Li.

A. 1 2 3 4 M



B. M 1 2 3 4

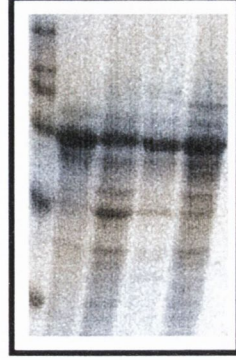


Figure 6.9 SDS polyacrylamide gel electrophoresis. An illustration of denaturing SDS polyacrylamide gel electrophoresis of cytoskeletal-rich and cytosol-rich protein fractions. **A.** Cytoskeletal-rich protein fractions were isolated as described in Section 2.4.1. Approximately 10 μ g protein were electrophoresed on 10% SDS polyacrylamide gels for 1 hour and stained in Coomassie reagent. **B.** Cytosol-rich fractions were isolated as described in Section 2.4.1. Approximately 15mg protein were electrophoresed on 10% SDS polyacrylamide gels for 1 hour and stained in Coomassie reagent. Similar results were obtained with membrane-rich protein fractions (data not shown). M= molecular weight standard marker (205kDa to 29kDa); 1=SW480; 2=SW620; 3=SW1116 & 4=498Li.

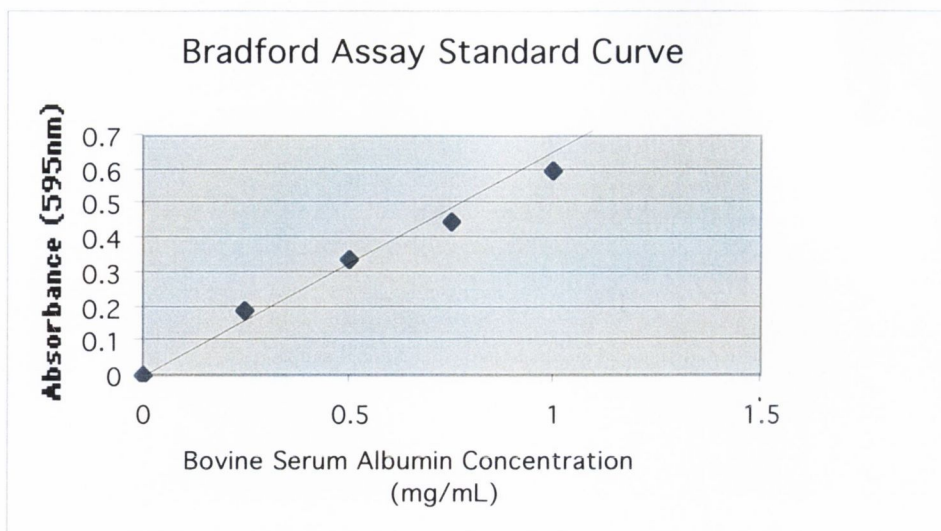


Figure 6.10 Bradford Assay Standard Curve. The Bradford assay was carried out as described in Section 2.4.2. A typical Bradford assay standard curve is shown above. Serial dilutions of a bovine serum albumin solution were prepared and the absorbances of each standard were determined at 595nm. Bovine serum albumin concentration was plotted against absorbance and the standard curve was used to estimate unknown protein concentrations.

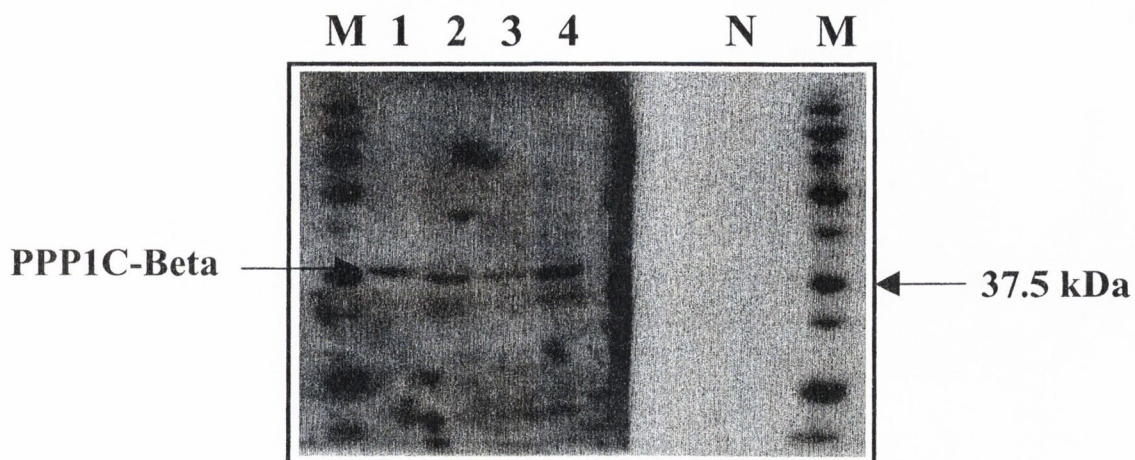


Figure 6.11 Protein phosphatase 1 catalytic subunit beta isoform (PPP1CB) Western blot analysis optimisation. An illustration of a typical PPP1CB Western blot optimisation experiment. Cytoskeletal-rich protein fractions were isolated from cell lysates as described (Section 2.4.1) and electrophoresed (approximately 25µg protein in each lane) on 10% denaturing polyacrylamide-SDS gels. Western blot analysis was carried out with a 1/400 dilution of PPP1CB antibody. This yielded a protein signal which co-migrated with the 37.5kD size marker. The PPP1CB protein is approximately 38kD in size. A duplicate Western blot was prepared with the omission of the primary antibody (PPP1CB) as a negative control which yielded no protein signals. M=205kDa to 29kDa high range protein molecular weight marker; 1=SW480; 2=SW620; 3=SW1116; 4=498Li & N=negative control.

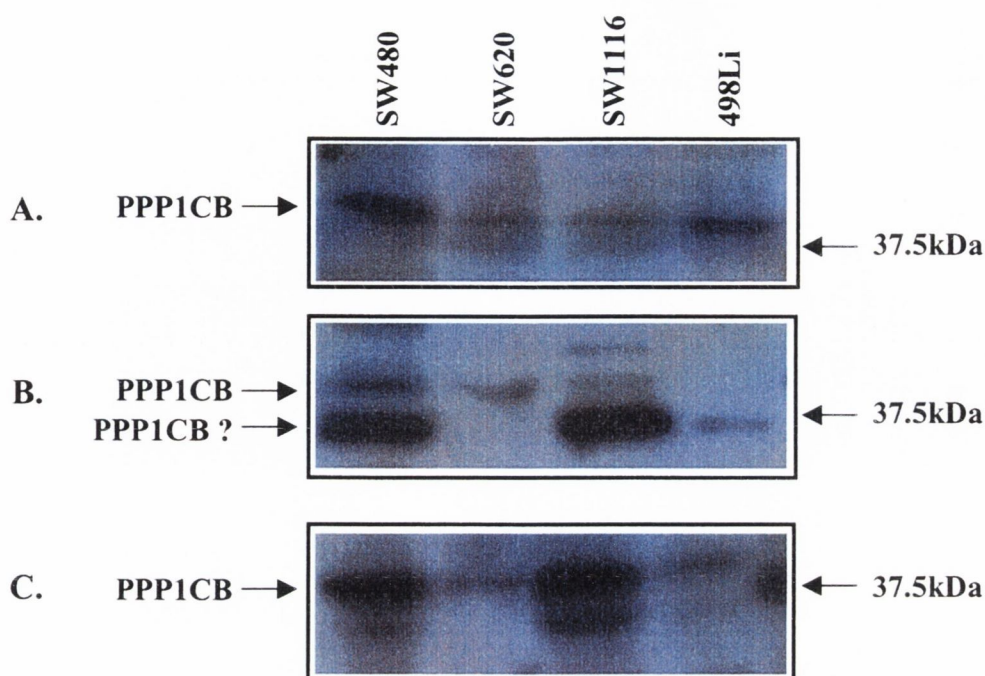


Figure 6.12 Protein phosphatase I, catalytic subunit, isozyme B (PPP1CB) Western blot analysis in subcellular fractions from the cell lines SW480, SW620, SW1116 & 498Li. An illustration of the expression of PPP1CB protein in subcellular fractions from pre-metastatic (SW480 & SW1116) and metastatic (SW620 & 498Li) sporadic colorectal cancer cell lines. **A.** Expression of PPP1CB in the membrane protein rich subcellular fractions. Elevated expression of PPP1CB is evident in SW480 with respect to SW620. A consistent overexpression of PPP1CB in premetastatic cells is not clear due to the elevated expression of PPP1CB in 498Li compared to SW1116. **B.** Expression of PPP1CB in the cytosolic protein rich subcellular fractions. Elevated expression of PPP1CB is clear in both pre-metastatic cell lines (SW480 & SW1116). A very strong protein signal (PPP1CB?) exists immediately below PPP1CB which may represent a dephosphorylated form of PPP1CB. **C.** Expression of PPP1CB in the cytoskeletal rich subcellular fractions. Elevated expression of PPP1CB is evident in both pre-metastatic cell lines SW480 & SW1116 compared to the metastatic cell lines SW620 & 498Li.

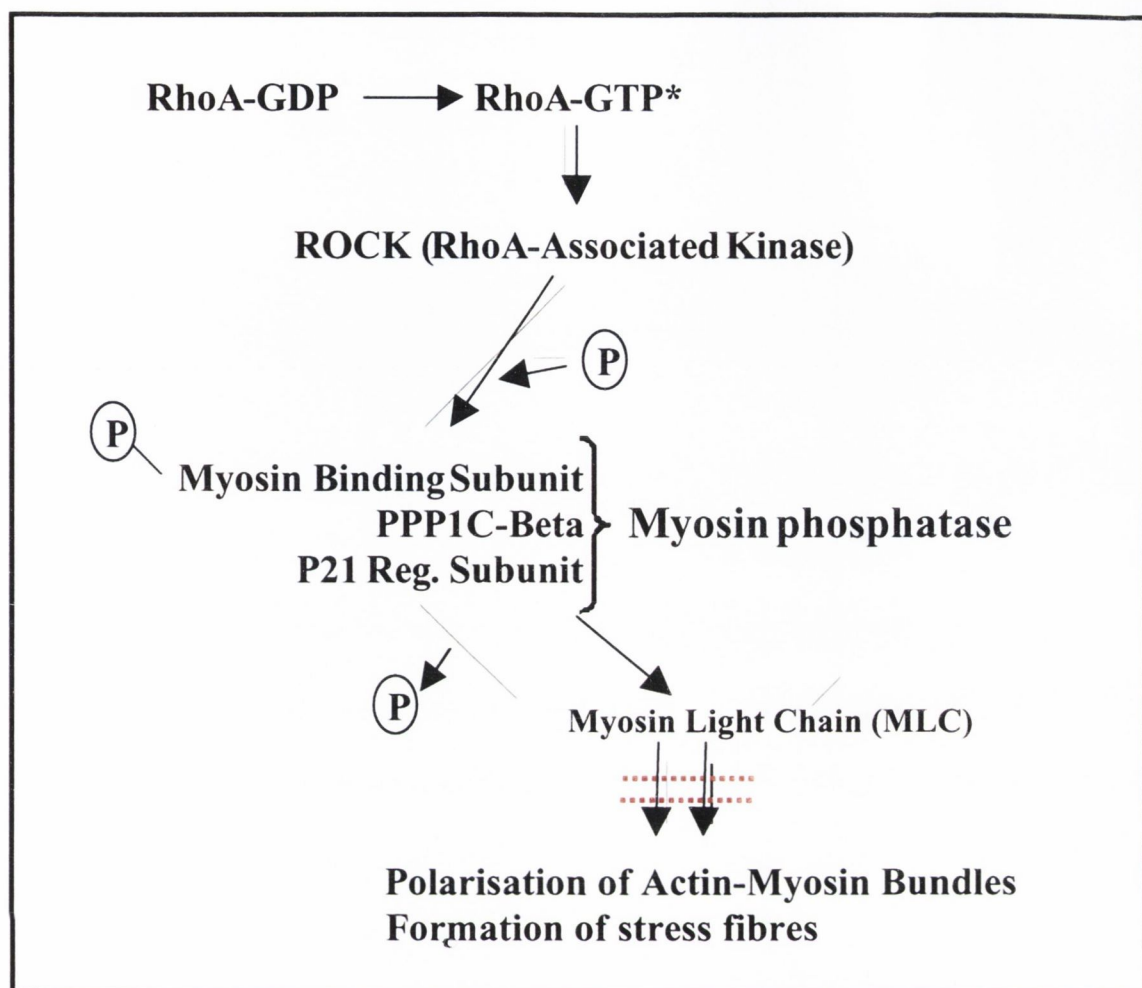


Figure 6.13 RhoA-mediated maintenance of cell motility. A schematic diagram of RhoA-mediated regulation of the cytoskeleton. RhoA-GDP is activated to RhoA-GTP by CD44 (not shown) which interacts with and activates ROCK (RhoA-Associated Kinase). ROCK phosphorylates the Myosin Binding Subunit of the trimeric complex myosin phosphatase. Activated myosin phosphatase dephosphorylates Myosin Light Chain (MLC) blocking the polarisation of Actin-Myosin bundles and formation of stress fibres. RhoA indirectly regulates cytoskeletal structure which determines the extent of cell invasion and migration. Downregulation of PPP1C-Beta in the cytoskeletal-rich fractions of the metastatic cell lines SW620 and 498Li may reduce the activity of myosin phosphatase. The downstream events are likely to involve increased polarisation of Actin-Myosin bundles and formation of stress fibres. This possibility is consistent with the decreased expression of PPP1C-Beta in the metastatic cell lines which have increased motility.

CHAPTER 7

GENERAL DISCUSSION

7.0 INTRODUCTION

The major goal of this study was to identify differences in gene expression between the MIN (microsatellite instability) and CIN (chromosomal instability) phenotypes in sporadic colorectal cancer. This was achieved by screening a panel of MIN and CIN cell lines using Differential Display Reverse Transcriptase PCR and GDA high density cDNA filters. These cell lines provide a convenient model for this type of analysis as the MIN lines are mismatch repair deficient and the CIN lines exhibit gross chromosomal abnormalities (Section 1.8). These techniques allowed the identification of a panel of differentially expressed candidate genes that may be associated with the two distinct phenotypes. This panel also formed the basis of further analysis. In addition to these studies, a secondary goal was to analyse the expression of IPP2 and IPP2 associated genes in colon cancer metastasis. The cell lines SW480 and SW620 are a convenient *in vitro* model for metastasis and had allowed the identification of IPP2 prior to this project (McWilliam *et al.*, MS in preparation). This study was extended to a number IPP2 associated genes.

To identify additional genes involved in MIN and CIN, we used Gene Discovery Array (GDA) v1.3 high density cDNA filters to determine expression differences between the MIN and CIN phenotypes. This involved three pairwise (one MIN probe and one CIN probe in each instance) comparisons of MIN and CIN cell line expression patterns on duplicate GDA filters. The panel of consistently differentially expressed genes (approximately seventy in number) between the three comparisons was obtained by bioinformatic approaches. A number of genes were selected for confirmatory analysis of which a number showed confirmed expression patterns using independent methods (including PTPD1, PLK, RANBP2, CCNA2, H2AZ). A total of six out of nine selected genes showed comparable expression patterns. This would suggest that a considerable fraction of the remaining genes from the original panel represent true expression differences. With this consideration, GDA analysis provides a convenient, cost-effective and straightforward method for obtaining global gene expression patterns.

7.1 MIN VERSUS CIN PHENOTYPES

Two distinct phenotypes exist in sporadic colorectal cancer. The MIN phenotype accounts for approximately 15% of sporadic colon tumours. MIN sporadic colorectal cancers are recognised by the prevalence of widespread genomic instability at both microsatellite and simple sequence repeats. The underlying mechanism for this phenotype is well understood. Inactivating germline mutations and epigenetic events such as hMLH1 promotor methylation in mismatch repair genes lead to errors in DNA replication. Mutations of this nature in coding regions of genes (TGF β RII, IGFR II & BAX) are seen as the driving force behind tumourigenesis in MIN tumours.

The molecular basis of CIN in tumours is just beginning to be explored. CIN tumours are characterised by gross chromosomal lesions such as aneuploidy and loss of heterozygosity and account for approximately 80% of sporadic colorectal cancers. Inactivating mutations in APC (tumour suppressor gene) is an early event in the CIN phenotype. The underlying mechanisms of the CIN pathway remain unclear. Loss of function in APC (a tumour suppressor gene) is an early event in the model of colorectal cancer proposed by Kinzler & Vogelstein (1996) (Figure 1.3). However, the underlying mechanisms initiating the molecular lesions in CIN tumours remain unclear. Two comparisons between MIN and CIN are informative. Firstly, both types of instability occur early during tumour evolution, but the genetic variations resulting from these instabilities increase with tumour progression. Secondly, when MIN and CIN cells are fused the resulting hybrids exhibit the CIN but not MIN phenotype. This shows that the CIN phenotype has a dominant quality, indicating that it may result from gain of function of an expressed protein rather than from gene inactivation. This, in turn, suggests that only a single mutational ‘hit’ may be acquired to produce the CIN phenotype.

7.1.1 MIN ASSOCIATED GENES

The panel of genes obtained by GDA analysis did not include genes previously associated with mismatch repair. However, the genes H2AZ and HDAC2 may have a role in the MIN phenotype. A considerable body of evidence has indicated that gene ‘silencing’ in MIN tumours can be carried out by a combination of CpG island methylation and the acetylation of histones in chromatin. Genes can be ‘silenced’ due to

steric hindrance of transcription factors at promotor regions by CpG island methylation. Hyperacetylation (or low deacetylation) of histones has a ‘tightening’ effect on DNA which decreases the efficiency of transcription from these regions. The combinatory effect of elevated expression of H2AZ in MIN cells and low expression of HDAC2 in CIN cells (with the exception of SW48) may represent a gene ‘silencing’ pathway in MIN tumours. It is worth noting that centromeric heterochromatin does not contain H2AZ and the DNMT1 (DNA Methyltransferase 1) has been reported to preferentially methylate heterochromatin. With the consideration of the current understanding of the MIN phenotype, further exploration of gene ‘silencing’ processes in MIN & CIN. However, examination of the EST (~14,000 EST are harboured on the GDA filters) expression data may also reveal novel genes directly associated with MIN and mismatch repair mechanisms.

7.1.2 CIN ASSOCIATED GENES

p53 was an obvious candidate for CIN as cells in culture become grossly aneuploid coincident with p53 inactivation. However, several cancers with mutant p53 are diploid and chromosomally stable. Studies in yeast have given clues to the candidate genes behind MIN, but many gene alterations can give rise to CIN in yeast. One such class of genes are those involved in cell cycle checkpoints. Alterations in the expression or sequence of human mitotic-checkpoint genes might be pivotal in CIN. Decreased expression of hMAD2 has been seen in breast cancers and might be important in aneuploidy. A number of other spindle checkpoint genes, including MAD1L1, MAD2, BUB3, TTK and CDC20 have been screened for mutations in nineteen aneuploid colorectal tumours. No new mutations were found suggesting that genes yet to be discovered are responsible for most of the checkpoint defects observed in CIN tumours. A small fraction of colorectal cancers harbour mutations in either hBUB1 or hBUBR1. The DNA-damage checkpoint has also been linked to CIN via genes such as ATM, ATR, BRCA1 and BRCA2. Yeast studies have revealed that chromosome instability arising from failed DNA-damage checkpoints is often associated with enhanced mitotic recombination and aberrant chromosome checkpoint. Another potential source of CIN involves centrosomal abnormalities. Genes that regulate centrosome number including

STK15 and PLK1 may also be linked to chromosome segregation. In summary many candidate genes for CIN have been studied resulting in occasional expression differences between CIN and MIN or occasional mutations in CIN related genes. Therefore, non-discriminative targeting of numerous genes involved in diverse cellular processes seems to be the trend in the CIN phenotype. It is also likely that many candidates are yet to be discovered.

7.1.3 CIN GENES FOUND BY GDA

The genes selected for confirmatory analysis were selected on the basis of current findings in the literature. In line with the genes currently being investigated in CIN colorectal cancer, we selected genes potentially involved in the different checkpoints described above. Generally, the expression differences seen by GDA analysis were considerably larger than the differences seen by RT-PCR or Northern blot analyses. This is likely to be due to the variability in hybridisation of complex probes (total cDNA probes) to targets and hybridisation of specific probes (Northern blot analysis) to targets or the sensitivity of RT-PCR. The low expression differences obtained would suggest that genuine 'hits' have not been selected in the confirmatory analyses. Such differences are unlikely to be due to mutations of the candidate genes. These genes are more likely to be under the influence of an upstream target gene at the focal point of CIN or represent genes involved in a particular pathway influenced by an upstream gene. Further confirmatory analysis of the original panel of genes may identify larger expression differences which are more likely to represent genuine targets in the CIN phenotype. Furthermore, similar confirmatory analysis of a subset of the 14,000 Expressed Sequence Tags (EST) harboured on GDA filters is likely to reveal genes involved in CIN.

GDA analysis has generated expression profiles of over 18,000 cDNA species between the MIN and CIN phenotypes. As suggested previously, a large proportion of the genes obtained (PTPD1, PLK, RANBP2, CCNA2, H2AZ) may be under the influence of an upstream gene or be involved in a particular pathway. Use may be made of the Yeast 2-hybrid database [<http://portal.curagen.com>] which allows the *in silico* identification of protein-protein interactions (in yeast) between a query yeast protein (protein of interest) and 6,000 yeast proteins. Some of the genes may be involved in

domains (ARD) have been shown to interact with CD44 inducing tumour migration in ovarian cancer cells. Further analysis of the AK000100 derived protein may clarify a role for it in colon cancer metastasis.

A previous study identified overexpression of IPP2 (Inhibitor 2 of Protein Phosphatase 1) in SW480 with respect to SW620. To illustrate the scope of downstream experiments post-DDRT-PCR we examined the expression of IPP2 in two isogenic pairs of cells lines. As IPP2 is reported to act as a chaperone of PPP1C (Protein Phosphatase 1 Catalytic Subunit) we examined the expression of three isoforms of PPP1C and the expression of PPP1C-Beta protein in the isogenic pairs of cell lines. We confirmed the elevated expression of IPP2 in SW480 and SW1116 (premetastatic) compared to SW620 and 498Li (metastatic). We found a similar expression pattern for PPP1C-Beta by RT-PCR analysis. As a consequence, we examined PPP1C-Beta protein expression and found a similar phenomenon, i.e. raised expression of PPP1C-Beta in both premetastatic cell lines with respect to metastatic cell lines. This would suggest that IPP2 and PPP1C-Beta may be under coordinate regulation. An observation of this nature leads to many possibilities for further investigation of IPP2/PPP1C-Beta, perhaps involving the RhoA pathway described in Chapter 6.

REFERENCES

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne J.D., White O. *et al.* Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*. 1995; 377(6547 Suppl): 3-174

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J.D. "Molecular biology of the cell." Garland Publishing Inc., NY (1994)

Alessi, D.R., Street, A.J., Cohen, P., Cohen, P.T. Inhibitor-2 functions like a chaperone to fold three expressed isoforms of mammalian protein phosphatase-1 into a conformation with the specificity and regulatory properties of the native enzyme. *Eur J Biochem*. 1993; 213(3): 1055-1066

Allen, P.B., Kwon, Y.G., Nairn, A.C. & Greengard, P. Isolation and characterization of PNUTS, a putative protein phosphatase 1 nuclear targeting subunit. *J Biol Chem*. 1998; 273(7): 4089-4095

Alexandrow, M.G. & Moses, H.L. Transforming growth factor beta and cell cycle regulation. *Cancer Res*. 1995; 55(7): 1452-1457

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol*. 1990; 215: 403-410

Astler, V.B. & Collier F.A. The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg*. 1954; 139: 846-852

Ayala, M., Balint, R.F., Fernandez-de-Cossio, M.E., Canaan-Haden, L., Larrick, J.W. & Gavilondo J.V. New primer strategy improves precision of differential display. *Biotechniques*. 1995; 18(5): 842-850

Backert, S., Gelos, M., Kobalz, U., Hanski, M.L., Bohm, C., Mann, B., Lovin, N., Gratchev, A., Mansmann, U., Moyer, M.P., Riecken, E.O. & Hanski C. Differential gene expression in colon carcinoma cells and tissues detected with a cDNA array. *Int. J. Cancer*. 1999; 82(6): 868-874

Badie, C., Itzhaki, J.E., Sullivan, M.J., Carpenter, A.J. & Porter, A.C. Repression of CDK1 and other genes with CDE and CHR promoter elements during DNA damage-induced G(2)/M arrest in human cells. *Mol. Cell. Biology*. 2000; 20(7): 2358-2366

Badie, C., Bourhis, J., Sobczak-Thepot, J., Haddada, H., Chiron, M., Janicot, M., Janot, F., Tursz, T. & Vassal, G. p53-dependent G2 arrest associated with a decrease in cyclins A2 and B1 levels in a human carcinoma cell line. *Br. J. Cancer*. 2000; 82(3): 642-650

Bauer, D., Muller, H., Reich, J., Riedel, H., Ahrenkiel, V., Warthoe, P. & Strauss M. Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Res*. 1993; 21(18): 4272-4280

Beavon, I.R. Regulation of E-cadherin: does hypoxia initiate the metastatic cascade? *Mol. Pathology*. 1999; 52(4): 179-188

Behrens, J., von Kries, J.P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R. & Birchmeier, W. Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature*. 1996; 382(6592): 638-642

Behrens, J., Jerchow, B., 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*. 1998; 280(5363): 596-599

Berndt, N., Dohadwala, M. & Liu, C.W. Constitutively active protein phosphatase 1alpha causes Rb-dependent G1 arrest in human cancer cells. *Curr Biol*. 1997; 7(6): 375-386

Bodmer, W.F., Bailey, C.J., Bodmer, J., Bussey, H.J., Ellis, A., Gorman, P., Lucibello, F.C., Murday, V.A., Rider, S.H., Scambler, P. *et al*. Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature*. 1987; 328(6131): 614-616

Bollen, M., DePaoli-Roach, A.A. & Stalmans, W. Native cytosolic protein phosphatase-1 (PP-1S) containing modulator (inhibitor-2) is an active enzyme. *FEBS Letters* 1994; 344(2-3): 196-200

Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Anal. Biochem*. 1972; 72: 248-254

Datson, N.A., van der Perk-de Jong, J., van der Berg, M.P., de Kloet, E.R. & Vreugdenhil, E. MicroSAGE: a modified procedure for serial analysis of gene expression in limiting amounts of tissue. *Nucleic Acids Res.* 1999; 27: 1300-1307

DePaoli-Roach, A.A. Synergistic phosphorylation and activation of ATP-Mg-dependent phosphoprotein phosphatase by F A/GSK-3 and casein kinase II (PC0.7). *J. Biol. Chem.* 1984; 259(19): 12144-12152

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. & Trent, J.M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet.* 1996; 14: 457-460

de Waard, V., van den Berg, B.M., Veken, J., Schultz-Heienbrok, R., Pannekoek, H. & van Zonneveld, A.J. Serial analysis of gene expression to assess the endothelial cell response to an atherogenic stimulus. *Gene.* 1999; 226(1):1-8

Diatchenko, L., Lau, C.Y., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. & Siebert, P. Suppression subtractive hybridisation: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A.* 1996; 93: 6025-6030

Donnelly, M.T. & Hawkey., C.J. COX-II inhibitors--a new generation of safer NSAIDs? *Aliment Pharmacol Ther.* 1997; 11(2): 227-236

Doxsey, S. The centrosome--a tiny organelle with big potential. *Nat Genet.* 1998; 20(2): 104-106

Druck, T., Kastury, K., Hadaczek, P., Podolski, J., Toloczko, A., Sikorski, A., Ohta, M., LaForgia, S., Lasota, J., McCue P., *et al.* Loss of heterozygosity at the familial RCC t(3;8) locus in most clear cell renal carcinomas. *Cancer Res.* 1995; 55(22): 5348-5353

Dukes, C.E. The classification of cancer of the rectum. *J Pathol Bacteriol.* 1932; 35: 323-332

Dukes, C.E., Bussey, H.J.R. The spread of rectal cancer and its effect on prognosis. *Br J Cancer.* 1958; 12: 309-320

Dunican, D., ParleMcDermott, A. & Croke, D.T. Gene expression profiling of cells and tissues. *Curr. Diag. Pathol.* 1997; 4: 170-175

Eberhart, C.E., Coffey, R.J., Radhika, A., Giardiello, F.M., Ferrenbach, S. & DuBois, R.N. Up-regulation of cyclooxygenase 2 gene expression in human colorectal adenomas and adenocarcinomas. *Gastroenterology.* 1994; 107(4): 1183-1188

Eshleman, J.R., Casey, G., Kochera, M.E., Sedwick, W.D., Swinler, S.E., Veigl, M.L., Willson, J.K., Schwartz, S. & Markowitz, S.D. Chromosome number and structure both

are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53. *Oncogene*. 1998; 17(6):719-725

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

Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M. & Kolodner, R. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell*. 1993; 75: 1027-1038

Foulkes, J.G., Strada, S.J., Henderson, P.J. & Cohen, P. A kinetic analysis of the effects of inhibitor-1 and inhibitor-2 on the activity of protein phosphatase-1. *Eur J Bioch*. 1983; 132(2): 309-313

Francia, G., Mitchell, S.D., Moss, S.E., Hanby, A.M., Marshall, J.F. & Hart I.R. Identification by differential display of annexin-VI, a gene differentially expressed during melanoma progression. *Cancer Res*. 1996; 56(17): 3855-3858

Frigerio, J.M., Berthezene, P., Garrido, P., Ortiz, E., Barthelémy, S., Vasseur, S., Sastre, B., Seleznieff, I., Dagorn, J.C. & Iovanna JL. Analysis of 2166 clones from a human colorectal cancer cDNA library by partial sequencing. *Hum Mol Genet*. 1995; 4(1): 37-

Gabriel, W.B, Dukes, C. & Bussey, H.J.R. Lymphatic spread in cancer of the rectum. Br J Surg. 1935; 23: 395-413

Georgescu, S.P., Komuro, I., Hiroi, Y., Mizuno, T., Kudoh, S., Yamazaki, T. & Yazaki, Y. Downregulation of polo-like kinase correlates with loss of proliferative ability of cardiac myocytes. J Mol Cell Cardiol. 1997; 29(3): 929-937

Gerdes, H., Gillin, J.S., Zimbalist, E., Urmacher, C., Lipkin, M. & Winawer, S.J. Expansion of the epithelial cell proliferative compartment and frequency of adenomatous polyps in the colon correlate with the strength of family history of colorectal cancer. Cancer Res. 1993; 53(2): 279-282

Giardiello, F.M., Petersen, G.M., Piantadosi, S., Gruber, S.B., Traboulsi, E.I., Offerhaus, G.J., Muro, K., Krush, A.J., Booker, S.V., Luce, M.C., Laken, S.J., Kinzler, K.W., Vogelstein, B. & Hamilton, S.R. APC gene mutations and extraintestinal phenotype of familial adenomatous polyposis. Gut. 1997; 40(4): 521-525

Gilliland, G., Perrin, S., Blanchard, K. & Bunn, H.F. Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. Proc of Natl Acad Sci USA. 1990; 87(7): 2725-2729

Gofuku, J., Shiozaki, H., Tsujinaka, T., Inoue, M., Tamura, S., Doki, Y., Matsui, S., Tsukita, S., Kikkawa, N. & Monden, M. Expression of E-cadherin and alpha-catenin in

He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogelstein, B. & Kinzler, K.W. Identification of c-MYC as a target of the APC pathway. *Science*. 1998; 281(5382):1509-1512

Heinen, C.D., Richardson, D., White, R. & Groden J. Microsatellite instability in colorectal adenocarcinoma cell lines that have full-length adenomatous polyposis coli protein. *Cancer Res*. 1995; 55:4797-4799

Hermeking, H., Lengauer, C., Polyak, K., He, T.C. Zhang, L., Thiagalingam, S., Kinzler, K.W. & Vogelstein, B. 14-3-3 sigma is a p53-regulated inhibitor of G2/M progression. *Molecular Cell*. 1997; 1(1): 3-11

Herrera, L., Kakati, S., Gibas, L., Pietrzak, E. & Sandberg, A.A. Gardner syndrome in a man with an interstitial deletion of 5q. *Am J Med Genet*. 1986; 25(3):473-476

Holmes, C.F., Kuret, J., Chisholm, A.A. & Cohen, P. Identification of the sites on rabbit skeletal muscle protein phosphatase inhibitor-2 phosphorylated by casein kinase-II. *Biochim Biophys Acta*. 1986; 870(3): 408-416

Houlston, R.S., Collins, A., Slack, J. & Morton, N.E. Dominant genes for colorectal cancer are not rare. *Ann Hum Genet*. 1992;56 (Pt 2): 99-103

Huang, F.L. & Glinsmann, W.H. Separation and characterization of two phosphorylase phosphatase inhibitors from rabbit skeletal muscle. *Eur J Biochem.* 1976; 70(2): 419-426

Hubank, M. & Schatz, D.G. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res.* 1994; 22: 5640-5648

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature.* 1993; 363(6429): 558-561

Jass, J., Love, S. & Northover, J. A new prognostic classification for rectal cancer. *Lancet.* 1987; **i**: 1333-1335

Johnson-Saliba, M., Siddon, N.A., Clarkson, M.J., Tremethick, D.J & Jans, D.A. Distinct importin recognition properties of histones and chromatin assembly factors. *FEBS Lett.* 2000; 467: 169-174

Juan, L.J., Shia, W.J., Chen, M.H., Yang, W.M., Seto, E., Lin, Y.S., Wu, C.W. & Kakinoki, Y.J. Histone Deacetylases Specifically Down-regulate p53-dependent Gene Activation. *J. Biol. Chem.* 2000; (MS in press)

Kal, A.J., van Zonneveld, A.J., Benes, V., van den Berg, M., Koerkamp, M.G., Albermann, K., Strack, N., Ruijter, J.M., Richter, A., Dujon, B., Ansorge, W. & Tabak, H.F. Dynamics of gene expression revealed by comparison of serial analysis of gene

expression transcript profiles from yeast grown on two different carbon sources. *Mol Biol Cell*. 1999; 10(6): 1859-1872

Kakinoki, Y.J., Somers, J. & Brautigan, D.L. Multisite phosphorylation and the nuclear localisation of phosphatase inhibitor 2-green fluorescent protein fusion protein during S phase of the cell growth cycle. *J. Biol. Chem*. 1997; 272: 32308-32314

Kennedy, B.P., Soravia, C., Moffat, J., Xia, L., Hiruki, T., Collins, S., Gallinger, S. & Bapat B. Overexpression of the non pancreatic secretory group II PLA2 messenger RNA and protein in colorectal adenomas from familial adenomatous polyposis patients. *Cancer Res*. 1998; 58(3): 500-503

Kimura, K., Ito, M., Amano, M., Chihara, K., Fukata, Y., Nakafuku, M., Yamamori, B., Feng, J., Nakano, T., Okawa, K., Iwamatsu, A. & Kaibuchi, K. Regulation of Myosin Phosphatase by Rho and Rho-Associated Kinase (Rho-Kinase). *Science*. 1996; 273: 245-248

Kinzler, K.W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell*. 1996; 87(2): 159-70

Kirklin, J.W., Dockerty, M.B. & Waugh. J.M. The role of the peritoneal reflection in the prognosis of carcinoma of the rectum and sigmoid colon. *Surg Gynecol Obstet*. 1949; 88: 326

Kojima, S., Nara, K., Rifkin, D.B. Requirement for transglutaminase in the activation of latent transforming growth factor-beta in bovine endothelial cells. *J Cell Biol.* 1993; 121(2): 439-448

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.* 1996; 111:307-317

Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B. & Clevers, H. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC^{-/-} colon carcinoma. *Science.* 1997; 275(5307):1784-1787

Kuerbitz, S.J., Plunkett, B.S., Walsh, W.V. & Kastan, M.B. Wild-type p53 is a cell cycle checkpoint determinant following irradiation. *Proc Natl Acad Sci U S A.* 1992; 89(16):7491-7495

Larrick, J.W. & Siebert, P.D. Reverse transcriptase PCR. Published by Ellis Horwood. 1995

Lau, M.M., Stewart, C.E., Liu, Z., Bhatt, H., Rotwein, P. & Stewart, C.L. Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev.* 1994; 8(24): 2953-2963

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L.A., Nystrom-Lahti, M. *et al.* Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell.* 1993; 75(6): 1215-1225

Leibovitz, A., Stinson, J.C., McCombs, W.B, 3rd., McCoy, C.E., Mazur, K.C. & Mabry, N.D. Classification of human colorectal adenocarcinoma cell lines. *Cancer Res.* 1976; 36(12): 4562-4569

Lengauer, C., Kinzler, K.W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature.* 1997; 386(6625): 623-627

Leppert, M., Dobbs, M., Scambler, P., O'Connell, P., Nakamura, Y., Stauffer, D., Woodward, S., Burt, R., Hughes, J., Gardner, E. *et al.* The gene for familial polyposis coli maps to the long arm of chromosome 5. *Science.* 1987; 238(4832): 1411-1413

Li, Y. & Benezra, R. Identification of a human mitotic checkpoint gene: hsMAD2. *Science.* 1996; 274(5285): 246-248

Liang, P., Averboukh, L., Zhu, W. & Pardee, A.B. Ras activation of genes: Mob-1 as a model. *Proc Natl Acad Sci U S A*. 1994; 91(26): 12515-12519

Liang, P. & Pardee, A.B. Differential display of eukaryotic messenger RNA by means of a polymerase chain reaction. *Science*. 1992; 257: 967-971

Lindblom, A., Tannergard, P., Werelius, B. & Nordenskjold, M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet*. 1993; 5(3): 279-282

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

Lockhart, D.J., Dong, H., Byrne, M.C., Folletie, M.T., Gallo, G.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. & Brown, E.L. Expression monitoring by hybridisation to high-density oligonucleotide arrays. *Nat Biotech*. 1996; 14(13): 1675-1680

Lockhart-Mummery, J.P. Two hundred cases of cancer of the rectum treated by perineal excision. *Br J Surg*. 1926-1927; 14: 110

Lodish, H., Baltimore, D., Berk, A., Zipursky, S.L., Matsudaira, P. & Darnell, J.

“Molecular Cell Biology.”. Scientific American Books, Inc., USA, 1995

Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. Protein measurement with the Folin phenol reagent. J. Biol. Chem. 1951; 193: 265-275

Lynch, H.T., Smyrk, T.C., Watson, P., Lanspa, S.J., Lynch, J.F., Lynch, P.M., Cavalieri, R.J. & Boland, C.R. Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. Gastroenterology. 1993; 104: 1535-1549

MacKintosh, C., Garton, A.J., McDonnell, A., Barford, D., Cohen, P.T., Tonks, N.K. & Cohen, P. Further evidence that inhibitor-2 acts like a chaperone to fold PP1 into its native conformation. FEBS Letters. 1996; 397(2-3): 235-238

MacPhee, M., Chepenik, K.P., Liddell, R.A., Nelson, K.K., Siracusa, L.D. & Buchberg A.M. The secretory phospholipase A2 gene is a candidate for the Mom1 locus, a major modifier of ApcMin-induced intestinal neoplasia. Cell. 1995; 81(6):957-966

Maier, G.D., Wright, M.A., Lozano, Y., Djordjevic, A., Matthews, J.P. & Young, M.R. Regulation of cytoskeletal organization in tumor cells by protein phosphatases-1 and -2A. Intl J Cancer. 1995; 61(1): 54-61

McWilliam, P., Dunican, D., Parle-McDermott, A., Tighe, O. & Croke D.T. MS in preparation. 2000

Miyoshi, Y., Ando, H., Nagase, H., Nishisho, I., Horii, A., Miki, Y., Mori, T., Utsunomiya, J., Baba, S., Petersen, G. *et al.* Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. *Proc Natl Acad Sci U S A.* 1992; 89: 4452-4456

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

Molenaar, M., van de Wetering, M., Oosterwegel, M. Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H.X. Tcf-3 transcription factor mediates beta-catenin-induced axis formation in *Xenopus* embryos. *Cell.* 1996; 86(3): 391-399

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.* 1997; 275(5307): 1787-1790

Moses, H.L., Yang, E.Y. & Pietsenpol, J.A. TGF-beta stimulation and inhibition of cell proliferation: new mechanistic insights. *Cell.* 1990; 63(2): 245-247

Narayanan, R., Lawlor, K.G., Schaapveld, R.Q., Cho, K.R., Vogelstein, B., Bui-Vinh, Tran, P., Osborne, M.P. & Telang, N.T. Antisense RNA to the putative tumor-suppressor gene DCC transforms Rat-1 fibroblasts. *Oncogene*. 1992; 7(3):553-561

Noonan, K.E., Beck, C., Holzmayer, T.A., Chin, J.E., Wunder, J.S., Andrulis, I.L., Gazdar, A.F., Willman, C.L., Griffith, B., Von Hoff, D.D. *et al.* Quantitative analysis of MDR1 (multidrug resistance) gene expression in human tumors by polymerase chain reaction. *Proc Nat Acad Sci USA*. 1990; 87(18): 7160-7164

Nupponen, N.N., Porkka, K., Kakkola, L., Tanner, M., Persson, K., Borg, A., Isola, J. & Visakorpi T. Amplification and overexpression of p40 subunit of eukaryotic translation initiation factor 3 in breast and prostate cancer. *Am J Pathol*. 1999; 154(6): 1777-1783

Oshima, M., Dinchuk, J.E., Kargman, S.L., Oshima H., Hancock, B., Kwong, E., Trzaskos, J.M., Evans JF. & Taketo MM. Suppression of intestinal polyposis in Apc delta716 knockout mice by inhibition of cyclooxygenase 2 (COX-2). *Cell*. 1996; 87(5):803-809

Ouspenski, I.I., Mueller, U.W., Matynia, A., Sazer, S., Elledge, S.J. & Brinkley, B.R. Ran-binding protein-1 is an essential component of the Ran/RCC1 molecular switch system in budding yeast. *J Biol Chem*. 1995; 270(5): 1975-1978

Peltomaki, P., Aaltonen, L.A., Sistonen, P., Pylkkanen, L., Mecklin, J.P., Jarvinen, H., Green, J.S., Jass, J.R., Weber J.L., Leach, F.S. *et al.* Genetic mapping of a locus predisposing to human colorectal cancer. *Science*. 1993; 260: 810-812

Peltomaki, P. & de la Chapelle, A. Mutations predisposing to hereditary nonpolyposis colorectal cancer. *Adv Cancer Res*. 1997; 71: 93-119

Phear, G., Bhattacharyya, N.P. & Meuth, M. Loss of heterozygosity and base substitution at the APRT locus in mismatch-repair-proficient and -deficient colorectal carcinoma cell lines. *Mol Cell Biol*. 1996; 16(11): 6516-6523

Polakis, P. Mutations in the APC gene and their implications for protein structure and function. *Curr Opin Genet Dev*. 1995; 5(1): 66-71

Polakis, P. The oncogenic activation of beta-catenin. *Curr Opin Genet Dev*. 1999; 9(1): 15-21

Popescu, N., Zimonjic, D., Hatch, C. & Bonner, W. Chromosomal mapping of the human histone gene H2AZ to 4q24 by fluorescence in situ hybridization. *Genomics*. 1994; 20(2): 333-335

Razin A. CpG methylation, chromatin structure and gene silencing-a three-way connection. *EMBO Journal*. 1998; 17(17): 4905-4908

Reenan, R.A. & Kolodner, R.D. Isolation and characterization of two *Saccharomyces cerevisiae* genes encoding homologs of the bacterial HexA and MutS mismatch repair proteins. *Genetics*. 1992; 132(4): 963-973

Rhee, I., Jair, K., Yen, R.C., Lengauer, C., Herman, J.G., Kinzler, K.W., Vogelstein, B., Baylin, S.B. & Schuebel, K.E. CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature*. 2000; 27; 404(6781): 1003-1007

Rigas, B., Goldman, I.S. & Levine, L. Altered eicosanoid levels in human colon cancer. *Journal of Laboratory & Clinical Medicine*. 1993; 122(5): 518-523

Roberts, A.B. & Sporn, M.B. Peptide growth factors and their receptors. Springer-Verlag, Heidelberg. 1991. 419-472

Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S.H., Masiarz, F.R., Munemitsu, S. & Polakis, P. Association of the APC gene product with beta-catenin. *Science*. 1993; 262(5140): 1731-1734

Russanova, V., Stephanova, E., Pashev, I. & Tsanev, R. Histone variants in mouse centromeric chromatin. *Mol Cell Biochem*. 1989; 90(1): 1-7

Sakao, Y., Noro, M., Sekine, S., Nozue, M., Hirohashi, S., Itoh, T. & Noguchi, M. Microsatellite instability and frameshift mutations in the Bax gene in hereditary nonpolyposis colorectal carcinoma. *Jpn J Cancer Res.* 1998; 89(10):1020-1027

Sambrook, J., Fritsch, E.L. & Maniatis, T. "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory Press, Plainview, NY, 2nd Edition, 1989

Sampson-Johannes, A., Wang, W. & Shtivelman, E. Colonization of human lung grafts in SCID-hu mice by human colon carcinoma cells. *Intl J Cancer.* 1996; 65(6): 864-869

Sanseau, P., Jackson, A., Alderton, R.P., Beck, S., Senger, G., Sheer, D., Kelly, A. & Trowsdale J. Cloning and characterization of human phosphatase inhibitor-2 (IPP-2) sequences. *Mamm Genome.* 1994; 5(8): 490-496

Sargent, T.D. & Dawid, I.B. Differential gene expression in the gastrula of *Xenopus laevis*. *Science.* 1983; 222: 135-139

Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science.* 1995; 270(5235): 467-470

Schmidt, D.R. & Schreiber, S.L. Molecular association between ATR and two components of the nucleosome remodeling and deacetylating complex, HDAC2 and CHD4. *Biochemistry.* 1999; 38(44): 14711-14717

Schwartz, S., Jr., 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.* 1999; 59(12): 2995-3002

Shenolikar, S. & Nairn, A.C. Protein phosphatases: recent progress. *Advances in Second Messenger & Phosphoprotein Research.* 1991; 23: 1-121

Sherr, C.J. Cancer cell cycles. *Science.* 1996; 274(5293): 1672-1677

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.* 1994; 6(3): 273-281

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. *New Engl J Med.* 1996; 335: 1727-1732

Simms, L.A., Zou, T.T., Young, J., Shi Y.Q., Lei, J., Appel, R., Rhyu, M.G., Sugimura, H., Chenevix-Trench, G., Souza, R.F., Meltzer, S.J. & Leggett, B.A. Apparent protection from instability of repeat sequences in cancer-related genes in replication error positive gastrointestinal cancers. *Oncogene.* 1997; 14(21): 2613-2618

Su, L.K., Vogelstein, B. & Kinzler, K.W. Association of the APC tumor suppressor protein with catenins. *Science*. 1993; 262(5140): 1734-1737

Thibodeau, S.N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science*. 1993; 260(5109): 816-819

Thun, M.J., Namboodiri, M.M., Calle, E.E., Flanders, W.D. & Heath, C.W. Jr. Aspirin use and risk of fatal cancer . *Cancer Res*. 1993; 53(6): 1322-1327

Tom, B.H., Rutzky, L.P., Jakstys, M.M., Oyasu, R., Kaye, C.I. & Kahan, B.D. Human colonic adenocarcinoma cells. I. Establishment and description of a new line. *In Vitro*. 1976; 12(3): 180-191

Trenkle, T., Welsh, J., Jung, B., Mathieu-Daude, F. & McClelland, M. Non-stoichiometric reduced complexity probes for cDNA arrays. *Nucleic Acids Res*. 1998; 26(17): 3883-3891

Tsujii, M., Kawano, S. & DuBois, R.N. Cyclooxygenase-2 expression in human colon cancer cells increases metastatic potential. *Proc Natl Acad Sci USA*. 1997; 94(7): 3336-3340

Tung, H.Y., Wang, W. & Chan C.S. Regulation of chromosome segregation by Glc8p, a structural homolog of mammalian inhibitor 2 that functions as both an activator and an inhibitor of yeast protein phosphatase 1. *Mol Cell Biol*. 1995; 15(11): 6064-6074

Vasen., H.F., Offerhaus, G.J., den Hartog Jager, F.C., Menko, F.H., Nagengast, F.M. & Griffioen, G. *et al.* The tumour spectrum in hereditary non-polyposis colorectal cancer: a study of 24 kindreds in the Netherlands. *Int J Cancer*. 1990; 46: 31-34

Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler K.W. Serial analysis of gene expression. *Science*. 1995; 270: 484-487

Villa-Moruzzi, E., Tognarini, M., Cecchini, G. & Marchisio, P.C. Protein phosphatase 1 delta is associated with focal adhesions. *Cell Adhes Commun*. 1998; 5(4): 297-305

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-tumour development. *N. Engl. J. Med*. 1988; 319: 525-532

Vogelstein, B., Fearon, E.R., Kern, S.E., Hamilton, S.R., Preisinger, A.C., Nakamura, Y. & White, R. Allelotype of colorectal carcinomas. *Science*. 1989; 244(4901): 207-211

Witty, J.P., McDonnell, S., Newell, K.J., Cannon, P., Navre, M., Tressler, R.J., Matrisian, L.M. Modulation of matrilysin levels in colon carcinoma cell lines affects tumorigenicity in vivo. *Cancer Res*. 1994; 54(17): 4805-4812

Wolf, G., Elez, R., Doermer, A., Holtrich, U., Ackermann, H., Stutte, H.J., Altmannsberger, H.M., Rubsamen-Waigmann, H. & Strebhardt, K. Prognostic

significance of polo-like kinase (PLK) expression in non-small cell lung cancer. *Oncogene*. 1997; 14(5): 543-549

Wrana, J.L., Attisano, L., Wieser, R., Ventura, F. & Massague, J. Mechanism of activation of the TGF-beta receptor. *Nature*. 1994; 370(6488): 341-347

Yamamoto, H., Itoh, F., Hinoda, Y. & Imai, K. Inverse association of cell adhesion regulator messenger RNA expression with metastasis in human colorectal cancer. *Cancer Res*. 1996; 56(15): 3605-3609

Yeo, J.P., Alderuccio, F. & Toh, B.H. A new chromosomal protein essential for mitotic spindle assembly. *Nature*. 1994; 367(6460): 288-291

Yoshioka, K., Nakamori, S. & Itoh, K. Overexpression of small GTP-binding protein RhoA promotes invasion of tumor cells. *Cancer Res*. 1999; 59(8): 2004-2010

Zhu, D. & Bourguignon, L.Y. Interaction between CD44 and the repeat domain of ankyrin promotes hyaluronic acid-mediated ovarian tumor cell migration. *J Cell Physiol*. 2000 May; 183(2): 182-195

Zheng, X.M., Black, D., Chambon, P. & Egly, J.M. Sequencing and expression of complementary DNA for the general transcription factor BTF3. *Nature*. 1990; 344(6266): 556-559

Ziegler, R.G., Devesa, S.S., & Fraumeni, J.F.Jr. Epidemiologic patterns of colorectal cancer. Important Adv Oncol. 1986: 209-232