



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 metastasis in colon cancer

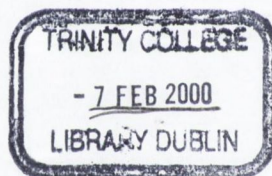
Anne Parle-McDermott

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

October 1999

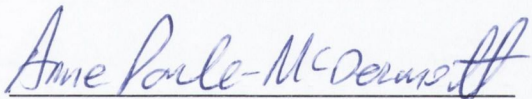


Thesis
5323.

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.



Anne Parle-McDermott

October, 1999.

For Mam, Dad and Paul

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. David Croke, for giving me the opportunity to carry out this research, for his constant encouragement and support throughout the project and finally for the much appreciated help in editing this thesis. David, you were a great supervisor.

Thanks to all in the DTC lab. past and present: Orna, Peter, Donncha, Kate, Charlie, Miriam, Brian McH. and Karen, it was a pleasure to work with you all. Orna, thanks for all your help, common sense and taste in men (Robson Greene springs to mind)! Peter, thanks for all your advice and experience and for always seeing the bright side (not to mention 'Snood' advice). Donncha, what can I say, the life and soul of the lab. Thanks for all your help with my PC dilemmas and for furthering my music education.

To everyone in the Biochemistry department, RCSI past and present: Paddy, Alan, Aideen, Belinda, Catherine P., Claire, Vicky, Darren, Silvia, Olive, Marice, Mary, Catherine H., Jim, Mary, Grace, Cathal and not forgetting Mr. 'Flame Face' himself Brian C. A special thanks to Dr. Aideen Long for all her help with the protein work.

Thanks to all in the Pathology department past and present: David B., Gillian, William, Bernie, Martina and Christine. A special thanks to Tony for his patience and help with the immunohistochemistry and to Elaine for the blocks.

I would also like to thank the Health Research Board for their financial support. Thanks to Drs. Jill Powell and Victor Velculescu for helpful SAGE advice and to Dr. Noel Williams for his continued interest and support of the project. I would also like to thank Teresa Keane, Dept. of Pharmacology, for all her help with tissue culture.

A special thanks to everyone in Ravens basketball club especially my friends and team-mates. Thanks for unknowingly providing all the 'light relief' when I needed it the most. Long may it continue!

Thanks to my good friend Clodagh for all the long distance phone calls and for keeping me up-to-date on the gossip!

Finally, I wish to thank my family for all their love and support: Ray, Breda, Mary, Jim, Nicky, Chris, Jed, Pat, Liz, Ron, Betty and all my nieces and nephews. A special

thanks to my sister Mary for always showing an interest. Not forgetting the 'in-laws'!
Thanks to Pat, Brendan, Miriam, Paul, Margaret and Patsy. A special thanks to Margaret for her friendship and the much appreciated loan of her printer and to Patsy for the computer.

This thesis is dedicated to my parents, Mary and Jack and to my husband, Paul. Mam and Dad, thanks for your continued love and encouragement and for always being supportive. Paul, thanks for your constant love and support, for listening to my continuous thesis monologues and for always making me smile.

TABLE OF CONTENTS

	Page Number
SUMMARY	i-ii
ABBREVIATIONS	iii-iv
 CHAPTER 1: INTRODUCTION	 1
1.0 INTRODUCTION	2
1.1 COLORECTAL TUMOURS: STAGING, CLASSIFICATION AND GENETICS	2
1.1.1 Staging of colorectal tumours	2
1.1.2 Classification of colorectal tumours	3
1.1.2.1 Familial Adenomatous Polyposis (FAP)	3
1.1.2.2 Hereditary Nonpolyposis Colorectal Cancer (HNPCC)	6
1.1.2.3 Sporadic colorectal cancer	8
1.1.3 A genetic model for colorectal tumourigenesis	10
1.2 THE METASTATIC PROCESS	13
1.2.1 Angiogenesis	13
1.2.2 Cell adhesion and motility	15
1.2.3 Proteolysis	16
1.2.4 Summary of metastasis and associated genes	18
1.3 AN EXAMINATION OF COLON TUMOUR METASTASIS	20
1.3.1 Aim	20
1.3.2 Differential gene expression profiling methodologies	20
1.3.3 The cell lines SW480 and SW620	21
1.3.4 Objectives	22
 CHAPTER 2: MATERIALS AND METHODS	 24
2.1 MATERIALS	25
2.1.1 Reagents	25
2.1.2 Stock solutions	28
2.1.3 Cell lines	30
2.2 TISSUE CULTURE METHODS	31
2.2.1 Cell Culture	31
2.3 MOLECULAR BIOLOGY METHODS	31
2.3.1 Electrophoresis methods	31
2.3.2 RNA extraction and purification	33
2.3.3 Estimation of nucleic acid concentration	33

2.3.4 Ethanol precipitations	34
2.3.5 Serial analysis of gene expression (SAGE)	34
2.3.6 Computer software	35
2.3.7 Semi-quantitative RT-PCR	35
2.3.8 Gel purification of PCR products/DNA fragments	37
2.3.9 Plasmid DNA preparation and DNA sequencing	38
2.3.10 Cloning	39
2.3.11 Northern blotting	40
2.4 PROTEIN ANALYSIS METHODS	42
2.4.1 Electrophoresis methods	42
2.4.2 Subcellular fractionation	42
2.4.3 Protein concentration estimation	43
2.4.4 Western blotting	44
2.4.5 Immunohistochemistry	45

CHAPTER 3: SERIAL ANALYSIS OF GENE EXPRESSION 47

3.0 INTRODUCTION	48
3.1 MATERIALS	53
3.1.1 Reagents	53
3.1.2 Stock solutions	54
3.2 METHOD	55
3.2.1 RNA preparation	55
3.2.2 cDNA synthesis	55
3.2.3 Cleavage of biotinylated cDNA with anchoring enzyme	56
3.2.4 Binding to streptavidin-coated magnetic beads	57
3.2.5 Ligating linkers to bound cDNA	57
3.2.6 Release of cDNA tags by digestion with tagging enzyme	58
3.2.7 Blunt ending of released cDNA tags	59
3.2.8 Ligating tags to form ditags	59
3.2.9 PCR amplification of ditags	59
3.2.10a Purification of the 102bp product	60
3.2.10b Release and purification of ditags	61
3.2.11 Ligation of ditags to form concatamers	63
3.2.12 Cloning of concatamers and sequence analysis	63
3.3 RESULTS	65
3.3.1 SW480 SAGE analysis	66
3.3.2 SW620 SAGE analysis	66

3.3.3 Gene expression differences between SW480 and SW620	67
3.4 DISCUSSION	69
3.4.1 SAGE data overview	69
3.4.2 Methodological considerations	71
3.4.2.1 Implementation of the SAGE protocol	71
3.4.2.2 Improvements to the SAGE protocol	72
3.4.2.3 Current gene expression methodologies	74

CHAPTER 4: ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES

4.0 INTRODUCTION	78
4.1 RESULTS	81
4.1.1 Controls for semi-quantitative RT-PCR and northern analysis	81
4.1.2 Confirmed differentially expressed SAGE tags	83
4.1.2.1 Keratin K5	83
4.1.2.2 Cystatin S	84
4.1.2.3 Serum Amyloid A (SAA)	84
4.1.2.4 Human homologue of yeast ribosomal S28	85
4.1.2.5 p32 subunit of Splicing Factor 2 (SFp32)	86
4.1.3 Unconfirmed differentially expressed SAGE tags	86
4.1.3.1 Myosin light chain	86
4.1.3.2 hnRNP L	87
4.1.3.3 High mobility group box protein (SSRP1)	87
4.1.3.4 β -tubulin	87
4.1.3.5 hnRNP C	88
4.1.3.6 PACAP type-3/VIP type-2 receptor	89
4.1.4 Analysis of expression in other metastatic cell lines	90
4.1.5 Summary of results	91
4.2 DISCUSSION	92
4.2.1 Re-assessment of the SAGE data	92
4.2.2 Comparison of SAGE data to other studies in SW480 and SW620	93
4.2.3 Expression of confirmed differences in other cell lines	95
4.2.4 Identification of ESTs	97
4.2.5 Roles of differentially expressed genes in metastasis	99
4.2.5.1 Cytoskeletal and motility associated genes	99

4.2.5.2 Cystatin S	104
4.2.5.3 Serum Amyloid A (SAA)	105
4.2.5.4 p32 subunit of Splicing Factor 2 (SFp32)	107
4.2.5.5 The ribosomal proteins	108
4.2.5.6 Other notable differentially expressed transcripts	110
4.2.5.7 Conclusions	111
CHAPTER 5: ANALYSIS OF DDRT CLONE PM208	113
5.0 INTRODUCTION	114
5.1 RESULTS	116
5.1.1 Identification of DDRT clone PM208	116
5.1.2 Expression analysis of Hox11, an associated gene product	118
5.1.3 Expression of IPP-2 in SW1116 and 498LI	119
5.2 DISCUSSION	121
5.2.1 Alternative polyadenylation signals	121
5.2.2 Roles of IPP-2 and Hox11 in metastasis	122
5.2.3 Conclusions	124
CHAPTER 6: ANALYSIS OF PROTEIN EXPRESSION IN CELL LINES AND TUMOURS	125
6.0 INTRODUCTION	126
6.1 RESULTS	128
6.1.1 Subcellular fractionation and western blot analysis	128
6.1.2 Immunohistochemical analysis	130
6.2 DISCUSSION	133
CHAPTER 7: GENERAL DISCUSSION	135
7.1 SUMMARY OF RESULTS	136
7.2 CONTRIBUTION TO COLON TUMOUR METASTASIS RESEARCH	140
7.3 FUTURE WORK	144
7.3.1 Selection of candidate genes for further analysis	144
7.3.2 Further analysis of candidate genes	145
7.4 CONCLUSIONS	147

REFERENCES

APPENDICES

SUMMARY

Metastatic spread is the main cause of colon cancer related death and therefore a greater understanding of this process is a prerequisite for improved treatment. The acquisition of metastatic properties confers tumour cells with the ability to leave the primary tumour, invade the circulatory system, survive transport and establish at distant sites. This process must involve the coordinate induction and repression of a range of genes involved in processes such as angiogenesis, proteolysis and growth-independence. Efforts to reach a greater understanding of tumour metastasis has involved analysis of gene expression patterns in tumours with different invasive potentials. However, the majority of studies have focused on the comparison of normal and tumour tissue and an analysis of late stage tumourigenesis, i.e. comparison of primary and metastatic tumours, has not been carried out extensively.

The work described here applied the method of Serial Analysis of Gene Expression (SAGE) to a pair of isogenic cell lines SW480 (primary colon tumour) and SW620 (lymph node metastasis of the same patient). The cell lines provided a model system to examine gene expression changes that may occur during the process of colon tumour metastasis. The SAGE analysis involved sequencing of >10,000 tags (tag= transcript identifier) and resulted in identification of a panel of differentially expressed genes that may have a role in colon tumour metastasis. The SAGE data also identified genes that did not show altered expression and provided information on the most abundant gene transcripts in the cell lines.

A subset of differentially expressed SAGE gene transcripts, and a previously isolated Differential Display Reverse Transcription-PCR clone (PM208), were analysed further. Expression of these genes was analysed by semi-quantitative RT-PCR and/or northern blot analyses in SW480 and SW620 and in several other colon tumour cell lines. A number of 'false positives' were identified from the comparative SAGE analysis and led to a re-assessment of the SAGE data. Clone PM208 was successfully identified by a combination of bioinformatics and limited cloning and sequencing. Protein levels were analysed for two confirmed differences, i.e. Keratin

K5 and Serum Amyloid A, by subcellular fractionation and western blot and expression in clinical material was examined by immunohistochemistry.

The results described in this work provide a reference database of the most abundant transcripts in the primary and metastatic cell lines SW480 and SW620 and identified a panel (51) of differentially expressed gene transcripts, some of which have been confirmed and showed altered expression in additional metastatic cell lines.

ABBREVIATIONS

AA:	amyloid A component
AMF:	autocrine motility factor
APC:	adenomatous polyposis coli
APS:	ammonium persulphate
ATCC:	American Type Culture Collection
BLAST:	basic local alignment search tool
CIN:	chromosomal instability phenotype
DCC:	Deleted in Colon Cancer
ddNTPs:	dideoxy nucleoside-triphosphates
DDRT:	Differential Display Reverse Transcription-PCR
DEPC:	diethylpyrocarbonate
DMEM:	Dulbecco's modified Eagle medium
DMSO:	dimethyl sulphoxide
dNTPs:	deoxy nucleoside-triphosphates
DPC4:	Deleted in Pancreatic Cancer 4
ECM:	extracellular matrix
EDTA:	ethylene diaminetetra acetic acid disodium salt
EGTA:	ethylene glycol-bis (β -amino ethyl ether) N, N, N', N'-tetraacetic acid
EST:	expressed sequence tag
FAP:	familial adenomatous polyposis
GAPDH:	glyceraldehyde 3-phosphate dehydrogenase
GSK-3 β :	glycogen synthase kinase-3 β
HA:	hyaluronic acid
HFL:	human fetal lung
HGF:	hepatocyte growth factor
HIP:	heparin/heparan sulphate interacting protein
HNPCC:	hereditary nonpolyposis colorectal cancer
ICAM:	intercellular adhesion molecule
IPP-2:	inhibitor-2 of Protein Phosphatase-1
IPTG:	isopropylthio- β -galactoside
LBP:	laminin-binding protein
LCM:	laser capture microdissection
Lef:	lymphoid enhancing factor
LOH:	loss of heterozygosity
MIN:	microsatellite instability phenotype
MOPS:	3-[N-Morpholino]-propane-sulphonic acid

MMP: matrix metalloproteinase
 MRP-1: motility related protein-1
 MSI: microsatellite instability
 MSS: microsatellite stable
 NCAM: neural cell adhesion molecule
 NEB: New England Biolabs
 NER: nucleotide-excision repair
 NIN: NER-associated instability
 ORF: open reading frame
 pBS: pBluescript II SK(+) DNA vector
 PBS: phosphate buffered saline
 PCR: polymerase chain reaction
 PMSF: phenylmethylsulfonyl fluoride a-toluenesulfonyl fluoride
 PP-1: Protein Phosphatase-1
 PP2A: Protein Phosphatase 2A
 RER: replication error phenotype
 RT-PCR: reverse transcription-polymerase chain reaction
 SAA: acute phase serum amyloid A
 SAGE: serial analysis of gene expression
 SCID: severe combined immunodeficient phenotype
 SDS-PAGE: sodium dodecyl sulphate-polyacrylamide gel electrophoresis
 SFp32: p32 subunit of splicing factor 2
 SSH: suppression subtractive hybridization
 SSRP1: high mobility group box protein
 S28: human homologue of yeast ribosomal S28
 Tcf: T cell factor
 TIGR: The Institute of Genomic Research
 TIMP: tissue inhibitor of metalloproteinase
 TNF α : tumour necrosis factor α
 TNM: tumour, nodes and metastasis
 uPA: urokinase plasminogen activator
 uPAR: urokinase type plasminogen activator receptor
 UTR: untranslated region
 VEGF: vascular endothelial growth factor
 X-gal: 5-bromo-4-chloro-3 indoyl- β -D-galactoside
 ZO-1: zonula occludens-1

CHAPTER 1

Introduction

1.0 INTRODUCTION

Colorectal cancer is the second most common form of cancer in Ireland, accounting for approximately 13% of cancer related deaths (National Tumour Registry, Republic of Ireland, 1995). The risk of colorectal cancer is two to three times higher in the Western world than in developing countries and this is believed to be largely due to a diet high in fat and low in fibre (Kinzler & Vogelstein, 1996). Dietary factors are thought to act as irritants rather than mutagens which may lead to tissue regeneration. Thus, dietary fibre is thought to absorb these irritants explaining their protective effect (Giovannucci & Willett, 1994; Smith, 1994). Despite advanced surgical and chemotherapeutic techniques, the overall mortality rate for most forms of epithelial cancer have not declined in the past twenty five years (Hong *et al.*, 1997).

1.1 COLORECTAL TUMOURS: STAGING, CLASSIFICATION AND GENETICS

1.1.1 Staging of colorectal tumours

The clinicopathological classification of tumours examines the extent of tumour invasion (metastasis) to assess patient prognosis. Tumours of an advanced stage have a greater chance of metastasis and, thus, the survival rates for these patients is poor. The staging of colorectal tumours may be described using the Dukes' classification system (Dukes, 1932). Dukes developed a system for the pathological classification of rectal carcinoma and it was first introduced at St. Mark's Hospital (London) in 1928. The classification was soon applied to colonic as well as rectal tumours (Simpson, 1939). The original Dukes' classification divided tumours into stages A to C and the Dukes' stage C was later subdivided into C₁ and C₂ (Gabriel *et al.*, 1935). The original Dukes' classification system is described in Table 1.1.

Modifications and 'improvements' on the Dukes' method have been introduced, such as the TNM (Tumour, Nodes and Metastasis) classification (Sobin *et al.*, 1988) and a reclassification of the Dukes' stages (Kirklin *et al.*, 1949). These changes have resulted in more confusion than clarity (Deans *et al.*, 1992). However,

the Dukes' classification is not perfect. Two important prognostic indicators, i.e. the histological grade of the tumour and the number of lymph nodes involved, are not considered in the Dukes' method (Deans *et al.*, 1992). In addition, the Dukes' B stage exhibits a variation in survival rates and new criteria are required to further subdivide this class (Bennett *et al.*, 1995). One improvement on the Dukes' method is the use of Dukes' C subdivisions as described by Astler & Coller (1954). The subdivisions are also named as C₁ and C₂, which is rather confusing, but C₁ is described as a lesion limited to the bowel wall and C₂ is described as a lesion penetrating through all the layers of the bowel with positive nodes (Deans *et al.*, 1992). This subdivision has been found to be superior in predicting prognosis to a single Dukes' C class by multivariate analysis (Fisher *et al.*, 1989).

The Dukes' classification of tumour staging still remains the best prognostic indicator for assessing patient survival. Patients with Dukes' A have an expected five year survival rate of approximately 80%, with Dukes' B and C having rates of 64% and 27% respectively (Deans *et al.*, 1992). Once the lymph nodes are involved however, prognosis is poor and disimproves with increasing number of nodes involved (Morson & Dawson, 1990).

1.1.2 Classification of colorectal tumours

Colorectal cancer may be divided into hereditary and sporadic (nonhereditary) types. Inherited colorectal cancers are inherited as autosomal dominant traits and account for at least 15% of all cases (Kinzler & Vogelstein, 1996). The hereditary forms consist of two major types, i.e. Familial Adenomatous Polyposis (FAP) and Hereditary Nonpolyposis Colorectal Cancer (HNPCC).

1.1.2.1 Familial Adenomatous Polyposis (FAP)

FAP accounts for approximately 1% of all colorectal cancer cases (Smith, 1994; Sheng *et al.*, 1998). FAP patients usually develop numerous benign adenomatous polyps (polyposis) during adolescence. These can range from hundreds

to thousands in number and at least one of these adenomas is likely to progress to a carcinoma, if untreated, by the age of forty. Therefore, FAP patients have accelerated tumour initiation (Kinzler & Vogelstein, 1996). Among the variant forms of FAP, Gardners' Syndrome refers to cases that have associated extracolonic tumours such as fibrous desmoid tumours, epidermoid cysts or jaw osteomata. Linkage of genetic markers on chromosome 5q21 to FAP (Bodmer *et al.*, 1987; Leppert *et al.*, 1987; Solomon *et al.*, 1987) resulted in identification of the APC (Adenomatous Polyposis Coli) gene (Grodin *et al.*, 1991). The APC gene has 15 exons and approximately 120 different mutations have been documented in FAP cases. Mutation of just one APC allele is required for polyp formation with mutation or loss of the second allele, referred to as LOH (Loss of Heterozygosity), being the initiating event for progression of the adenoma to carcinoma. The majority of FAP patients usually possess a germline mutation in just one allele of APC. The mutations predominantly yield truncated APC protein (Smith *et al.*, 1993) with the type of truncation mutation affecting the degree of polyposis (Kinzler & Vogelstein, 1996; Spirio *et al.*, 1998). In addition, specific mutant alleles of APC have also been shown to affect the spectrum of mutational events in developing colonocytes that lead to tumourigenesis (Spirio *et al.*, 1998). Moreover, a private mutation in Ashkenazi jews (i.e. a mutation unique to that particular population) identified a hypermutable tract in APC that results in incorrect processing by polymerases and therefore indirectly caused a cancer predisposition (Laken *et al.*, 1997).

Recent advances (Inomata *et al.*, 1996; Korinek *et al.*, 1997; Morin *et al.*, 1997; Behrens *et al.*, 1998; He *et al.*, 1998; Mann *et al.*, 1999; Roose *et al.*, 1999; Seeling *et al.*, 1999) in APC protein interactions have elucidated, in part, the role which this essential protein has in colorectal cancer progression (see Figure 1.1). A number of proteins are known to interact with APC and these include EB-1 (unknown function), DLG (human homologue of *Drosophila* tumour suppressor gene), γ -catenin, tubulin, conductin (Behrens *et al.*, 1998), B56 subunit of PP2A (Seeling *et al.*, 1999) and β -catenin (Morin *et al.*, 1997; Sheng *et al.*, 1998). Moreover, FAP

tumours show an accumulation of and an alteration in the subcellular localisation of β -catenin compared to normal cells (Inomata *et al.*, 1996; Sheng *et al.*, 1998). β -catenin was originally identified based on its interaction with α -catenin and E-cadherin in cell-cell interactions (Sherbet & Lakshmi, 1997) and has more recently been identified as an essential element in the Wntless-Wnt signalling cascade (important for growth and development) through its ability to bind to members of the Tcf (T cell factor) and Lef (lymphoid enhancer factor) families of transcription factors (Behrens *et al.*, 1996; Molenaar *et al.*, 1996). β -catenin forms a complex with Tcf-4 and this complex is capable of activating the transcription of a number of unknown target genes (Korinek *et al.*, 1997; Morin *et al.*, 1997). In normal colon, the wildtype APC protein controls gene transcription by interacting with β -catenin and promoting its degradation (Rubinfeld *et al.*, 1993; Su *et al.*, 1993). Phosphorylation of β -catenin by GSK-3 β (Glycogen synthase kinase-3 β) at multiple sites is necessary for interaction with APC and occurs in the absence of Wnt signalling (Morin *et al.*, 1997; Nathke, 1999). Colorectal tumours with recessive inactivating mutations of both APC alleles, or dominant mutations in β -catenin that effect its phosphorylation, allow uncontrolled association of β -catenin with Tcf-4 and therefore a likely overexpression of yet to be identified Tcf- target genes (Korinek *et al.*, 1997; Morin *et al.*, 1997). Mutations in conductin, another recently identified protein that can also regulate β -catenin degradation independently of APC, may also be involved (Behrens *et al.*, 1998). Moreover, expression of the B56 subunit of Protein Phosphatase 2A (PP2A) was also found to decrease the amount of β -catenin and this interaction required an intact APC protein (Seeling *et al.*, 1999). A number of targets of the β -catenin/Tcf transcription complex have recently been identified and these include *c-MYC* (He *et al.*, 1998), *c-jun* and *fra-1* (Mann *et al.*, 1999), and *Tcf1* (Roose *et al.*, 1999). uPAR (urokinase-type plasminogen activator receptor) and ZO-1 (zonula occludens-1) also showed altered expression levels in response to overexpression of β -catenin but a direct interaction with the complex was not demonstrated (Mann *et al.*, 1999). *c-MYC* is an oncogene which encodes a nuclear phosphoprotein and is involved in cell

proliferation (Scott & Quirke, 1993) The overexpression of *c-MYC* has been extensively documented during colorectal tumourigenesis (Fearon & Vogelstein, 1990; He *et al.*, 1998). *c-jun* and *fra-1* are components of the AP-1 complex which has been shown to increase expression of uPAR in colorectal cancer (Mann *et al.*, 1999). *Tcf1* may act as a feedback transcriptional repressor of β -catenin-Tcf4 target genes through its interaction with APC (Roose *et al.*, 1999). Thus, uncontrolled interaction of β -catenin with Tcf4 through mutations in APC, β -catenin itself or possibly conductin is the initiating step in FAP and results in uncontrolled activation of gene transcription, the targets of which are just beginning to be elucidated.

Although FAP only accounts for approximately 1% of colorectal cancers, mutations in APC or β -catenin are found in the vast majority of sporadic cases and are indistinguishable from those in FAP tumours.

1.1.2.2 Hereditary Nonpolyposis Colorectal Cancer (HNPCC)

Hereditary Nonpolyposis Colorectal Cancer (HNPCC) accounts for approximately 3-6% of all colorectal cancer cases and is an autosomal dominant disorder (Liu *et al.*, 1995; Konishi *et al.*, 1996). These patients do not present with polyposis and adenomatous polyps are observed to form at approximately the same frequency in patients with HNPCC as they do in people with sporadic cancer (Smith, 1994). However, HNPCC tumours exhibit accelerated tumour progression once the tumour is initiated (Kinzler & Vogelstein, 1996). HNPCC families are defined using the Amsterdam Criteria as described by The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC) (Vasen *et al.*, 1990):

- At least three relatives are affected, one should be a first degree relative to the other two and polyposis should be excluded.
- At least two successive generations should be affected.
- One of the relatives should be diagnosed under 50 years of age.

HNPCC cases are heterogeneous and patients may present with colon cancer only (Lynch syndrome I), colon cancer and/or endometrial, ovarian, gastric, duodenal or

proximal ureter/renal pelvic carcinomas (Lynch syndrome II) and/or cutaneous tumour manifestations (Muir-Torre syndrome) or as randomly scattered colorectal tumours. In addition, HNPCC tumours tend to be right-sided and diploid (Liu *et al.*, 1995). Linkage analysis initially associated HNPCC with chromosome 2p15-p16 (Smith, 1994). LOH analysis at this locus led to the discovery of genome-wide microsatellite instability (MSI) in most HNPCC tumours (Aaltonen *et al.*, 1993; Ionov *et al.*, 1993; Thibodeau *et al.*, 1993; Liu *et al.*, 1996; Boland *et al.*, 1998). Microsatellites are repetitive nucleotide sequences of varying lengths which occur in the human genome, between and within genes (Sherbet & Lakshmi, 1997). MSI is observed by the lengthening or shortening of the repetitive units within microsatellites. This tumour phenotype of MSI was named RER, for *replication errors* (Smith, 1994) and is also referred to as MIN (microsatellite instability phenotype) (Lengauer *et al.*, 1998). An association between MSI and mutant mismatch repair genes was first identified in yeast by Strand *et al.* (1993). A comparison was drawn between mutations in the mismatch repair genes *mutS* and *mutL* in *E. coli* and increased MSI and led to the discovery of mutant yeast homologues, i.e. *PMS1*, *MLH1* and *MSH2* that also caused MSI in these yeast strains (Strand *et al.*, 1993). HNPCC is now associated with at least five DNA mismatch repair genes: *hMSH2*, *hMLH1*, *hPMS2*, *hPMS1* and *hMSH6* (Lu *et al.*, 1998) with resulting genetic instability. However, mutations in *hMSH2*, *hMLH1* and *hPMS2* account for the majority of HNPCC cases (Kinzler & Vogelstein, 1996). Approximately 92% of HNPCC families displaying MSI have germline mutations in at least one of the mismatch repair (MMR) genes (Liu *et al.*, 1996). Thus, MSI reflects a malfunction in the replication or repair of DNA and results in instability in genes with repeated nucleotides within their sequence (Lothe *et al.*, 1997). Genes affected by MSI identified to date include TGF- β receptor type II, BAX and APC (Prolla *et al.*, 1998). The mutational spectrum for APC mutations in MSI tumours is different to that in tumours lacking an MSI phenotype (known as microsatellite stable (MSS) tumours (Boland *et al.*, 1998) and usually exhibit a chromosomal instability (CIN) phenotype (Lengauer *et al.*, 1998). APC mutations in

MSI tumours show a high frequency of frame-shift mutations in close proximity to one of the many (A)_n tracts within the APC sequence (Huang *et al.*, 1996). Thus, APC mutations occur in the majority of MSI and MSS colon tumours, suggesting a shared pathway for colorectal carcinogenesis (Huang *et al.*, 1996) but in the case of MSI tumours, defects in mismatch repair precede APC mutations and thus are the initiating events in tumourigenesis in HNPCC patients.

Approximately 12-14% of sporadic tumours are found to exhibit MSI (Smith, 1994) and are likely to follow the same pathway as HNPCC tumours, i.e. MSI leads to accelerated mutation in genes involved in colorectal tumourigenesis.

1.1.2.3 Sporadic colorectal cancer

Sporadic colorectal cancer describes those patients who have no family history of colorectal cancer and accounts for the majority of cases. However, mutations in genes associated with inherited colorectal cancer are also found in most sporadic cases, particularly APC mutations (Kinzler & Vogelstein, 1996). Similarly to the hereditary cases, sporadic colorectal cancer cases can be associated with a chromosomal instability (CIN) or microsatellite instability (MIN) phenotype (Lengauer *et al.*, 1998). CIN is characterised by the gains and losses of whole chromosomes resulting in aneuploidy and is likely to occur in most human malignancies (Lengauer *et al.*, 1998). CIN accounts for approximately 85% of sporadic colorectal cancers, the majority of which are initiated through mutations in either APC or β -catenin as described above for FAP. The molecular basis of CIN in colorectal cancer is just beginning to be unravelled and *hBUB1*, which is a mitotic checkpoint gene, has recently been associated with CIN (Cahill *et al.*, 1998). MIN accounts for a relatively small percentage of sporadic colorectal cancers (approximately 15%) but occurs in most cancers in patients with HNPCC as described above. Somatic mutations in the same mismatch repair genes associated with HNPCC are associated with some sporadic colorectal cancers exhibiting MIN but mutations in genes other than those known to be associated with mismatch repair

(Liu *et al.*, 1995) may also be involved. Thus, somatic mutations in repair genes can be selected for during tumourigenesis, even when such mutations do not directly enhance cell growth. Defects in a different type of repair mechanism, i.e. nucleotide-excision repair (NER) has also been associated with tumourigenesis and has been described as NER-associated instability or NIN (Lengauer *et al.*, 1998). The NER is responsible for repairing damage caused by exogenous mutagens and therefore, in addition to defective NER genes, exposure to an environmental mutagen (such as ultraviolet light) is a necessary step for development of NIN with resulting tumourigenesis. Moreover, patients with NER defects, as exemplified in xeroderma pigmentosum, usually develop numerous skin tumours which indicates that ultraviolet light is one of the major environmental mutagens (Lengauer *et al.*, 1998). Thus, a NIN phenotype has not been associated with colorectal tumours which are characterised instead by either a MIN or CIN phenotype as described above. The CIN and MIN phenotypes are mutually exclusive, i.e. cancers exhibiting mismatch repair deficiencies usually have a diploid chromosome number and chromosomal changes occur at the same frequency as in normal cells. CIN cancers, on the other hand, are usually aneuploid and generally do not display microsatellite instability. In addition, fusion of colorectal cancer cells with MIN and CIN phenotypes results in complementation of mismatch repair but not in chromosomal instability which suggests that a single gain of function mutation is necessary for initiation of the CIN phenotype (Lengauer *et al.*, 1998).

Regardless of the type of instability, both types occur early in tumourigenesis and the instabilities increase with tumour progression. The majority of colorectal cancers, either inherited or sporadic, involve mutations in either APC or β -catenin which results in uncontrolled activation of gene transcription (see Figure 1.1). However, in the case of MIN tumours, the microsatellite instability precedes the mutations in APC. β -catenin has also been implicated in human skin tumours (Chan *et al.*, 1999) and melanomas (Rubinfeld *et al.*, 1997) and, similarly to colorectal cancer, β -catenin-Tcf/Lef misregulation results in tumourigenesis in these tissues.

1.1.3 A genetic model for colorectal tumourigenesis

Colorectal carcinomas usually develop from benign neoplastic polyp(s) known as adenomas (Morson & Dawson, 1990). Adenomas are described as villous or tubular with intermediate forms known as tubulovillous. The progression from the adenoma to the carcinoma stage is facilitated by the accumulation of a number of genetic changes and these are outlined in the genetic model described by Kinzler & Vogelstein (1996) and is shown in Figure 1.2. These include, in addition to APC, mutation or deletion of the oncogene *k-ras*, a number of genes on chromosome 18q21 including *DCC* (Deleted in Colon Cancer) and *DPC4* (Deleted in Pancreatic Cancer locus 4) (Thiagalingam *et al.*, 1996), and the tumour suppressor gene *p53*. However, allelic loss of *DPC4* is only found in a subset of colorectal cancers and therefore is not always a target for allelic loss (Thiagalingam *et al.*, 1996). *DCC* is deleted in approximately 75% of colorectal cancers and is probably mutant in the remainder (Scott & Quirke, 1993; Cho *et al.*, 1994). *DCC* encodes a transmembrane glycoprotein with similarities to N-CAM (neural cell adhesion molecule) and is a receptor for netrin-1, a molecule involved in axon guidance (Mehlen *et al.*, 1998). *DCC* is highly expressed in normal colonic mucosa and many other tissues (Cho *et al.*, 1994; Sherbet & Lakshmi, 1997) and has recently been shown to induce apoptosis (Mehlen *et al.*, 1998). The tumour suppressor gene *p53* is located on chromosome 17p13.1 and is mutated in a wide variety of cancers including colorectal cancer (Sherbet & Lakshmi, 1997). *p53* acts as a transcription factor and is an important regulator of the cell cycle (Levine, 1997). *p53* transcriptional activation is negatively regulated by the adenovirus E1B-55kd protein and human MDM2 (Levine, 1997). Expression of *p53* in response to DNA-damaging agents induces either a stable growth arrest or apoptosis (Polyak *et al.*, 1997; Sherbet & Lakshmi, 1997) and thus is known as the "guardian of the genome" (Wynford-Thomas, 1997). *p53* has been shown to affect the expression of a number of genes either directly or indirectly and some of these include *p21^{WAF1/CIP1}*, cyclin G, GADD45, BAX, IGF-BP3, TGF- α , *c-fos* and *c-jun* (Madden *et al.*, 1997). *p53*-dependent growth arrest can occur at

either the G1/S or G2/M phases of the cell cycle (Agarwal *et al.*, 1995) and allows repair of damaged cells prior to cell replication (G1/S) or premitotic chromosome condensation (G2/M) and chromosome partition (Schwartz & Rotter, 1998). Growth arrest *via* p53 is dependent on the transcriptional induction of the protein p21^{WAF1/CIP1} (Agarwal *et al.*, 1995; Waldman *et al.*, 1995), which is an inhibitor of most of the cyclin-dependent kinases which help regulate the cell cycle (Agarwal *et al.*, 1995). p53-mediated apoptosis occurs in response to DNA damage or in response to expression of either a viral or cellular oncogene or the absence of a critical tumour suppressor gene product such as Rb (Levine, 1997). Two targets of p53 regulation, i.e. BAX and IGF-BP3 are thought to have an important role in p53-mediated apoptosis. In addition, overexpression of BCL2 blocks apoptosis and is antagonised by BAX (Levine, 1997). The majority of p53 mutations, which are found in 50-55% of all human cancers, result in a mutant p53 protein that fails to bind to DNA in a sequence-specific fashion (Levine, 1997). Somatic p53 mutations can occur as early or late events in tumourigenesis depending on the tumour type (Levine, 1997). A p53 mutation occurs as an early event in skin cancers, i.e. in pre-malignant lesions. In contrast, mutations both in p53 and DCC occur as late events in colorectal tumourigenesis (Levine, 1997).

As described in the previous sections, colorectal tumours with MSS exhibit a relatively high degree of chromosomal losses (CIN), while tumours displaying MSI can often show no allelic loss but have a higher mutation rate (MIN) (Kinzler & Vogelstein, 1996; Lengauer *et al.*, 1998). However, both types of tumours seem to go through a similar series of mutations during colorectal cancer progression and it is not simply the accumulation of mutations, but rather it is also the order, that determines the propensity to neoplasia (Kinzler & Vogelstein, 1996). In addition to the genetic alterations mentioned above, a recent study has identified somatic mutations in the mitochondrial genome in 70% of colorectal cancer cell lines examined (Polyak *et al.*, 1998). Moreover, the mutations were also confirmed in the tumours from which the cell lines were derived. The mutated mitochondria displayed a selective growth

advantage during tumour evolution and could have significant effects on cellular processes controlled by mitochondria (Polyak *et al.*, 1998).

Once the primary tumour is established a subpopulation of cells within the tumour may acquire additional invasive capabilities and become metastatic. The characteristics and changes that primary tumour cells undergo in order to become metastatic are discussed in the next section.

1.2 THE METASTATIC PROCESS

Metastasis is the process by which malignant cells leave their primary site and spread to distant locations throughout the body (reviewed in Ahmad & Hart, 1996; Duffy, 1996; Ahmad & Hart, 1997; Cockett *et al.*, 1998) and was first recognised by Jean-Claude Recamier in 1829 (Beavon, 1999). In order to do this, invading cells have to overcome a number of obstacles. The metastatic process may be divided into several sequential steps: detachment of cancer cells from the primary tumour, penetration into the circulatory system, dissemination throughout the body, arrest in capillary bed of distant organs, extravasation, induction of angiogenesis and finally, growth at the secondary site. The metastatic process is therefore quite complex and involves a number of gene products associated with angiogenesis, adhesion and motility, and basement membrane breakdown. An outline of the metastatic process is shown in Figure 1.3. Some of the molecules known to be involved in metastasis are described in the following sections.

1.2.1 Angiogenesis

There is a strong inverse association of tumour vascularisation with overall survival, as well as mode of metastasis in colorectal cancer (Tanigawa *et al.*, 1997). In order for tumours to grow beyond 2mm in size, the formation of new blood vessels is required in a process known as angiogenesis (Ahmad & Hart, 1997). New blood vessels are thin walled and leaky and are therefore more permeable to tumour cells. Similarly to the metastatic process itself, angiogenesis may also be divided into a number of steps as follows (Ahmad & Hart, 1996; Duffy, 1996; Ahmad & Hart, 1997):

- Degradation of the basement membrane
- Endothelial cell migration
- Endothelial cell proliferation
- Vascular loop formation
- Development of a new basement membrane

Moreover, the mechanisms used by endothelial cells in angiogenesis are comparable to those used by the invading tumour cells (Ahmad & Hart, 1997). Matrix-degrading proteases such as MMPs (Matrix Metalloproteinases) are necessary for ECM (extracellular matrix) degradation to allow new vessel formation (Cockett *et al.*, 1998). Protease inhibitors have been shown to inhibit angiogenesis and thus inhibit metastasis (Ahmad & Hart, 1996 & 1997). The stimulus for angiogenesis may be provided by the tumour cells or by host inflammatory cells within the tumour (Duffy, 1996). The process of angiogenesis involves both pro-angiogenic and anti-angiogenic factors. Pro-angiogenic factors include vascular endothelial growth factor (VEGF), hepatocyte growth factor, IL-8 and tumour necrosis factor- α (TNF α) and anti-angiogenic factors include angiostatin, thrombospondin-1 and platelet factor-4. Thus, induction of angiogenesis requires a balance in favour of pro-angiogenic factors. Subsequent research has found that tumour cells produce proteases which cleave plasminogen to generate the angiogenesis inhibitor angiostatin (O'Reilly *et al.*, 1994). Serum angiostatin was no longer detected five days after removal of the primary tumour and this presumably led to increased angiogenesis and growth of metastases (O'Reilly *et al.*, 1994). Thus, tumours can be self-inhibitory by halting angiogenesis *via* production of angiostatin. In addition, expression of angiostatin cDNA induced long term dormancy, i.e. a balance between proliferation and apoptosis, in lung metastases in experimental mice (Cao *et al.*, 1998). Angiogenesis is an essential step in the metastatic process by allowing continued growth of the primary tumour and providing an entry route for invasive cells into the circulatory system. Moreover, the established metastatic tumour will also require a blood supply at the distant site. Thus, angiogenesis is important both at the initial and ultimate stages of metastasis and is proving to be an important target in the prevention and treatment of cancer.

1.2.2 Cell adhesion and motility

The spread of cells from the primary tumour to distant sites involves alteration of cell adhesion, signalling and motility. It is likely that the ability of tumour cells to invade requires an intermediate expression of adhesive capability (Aznavorian *et al.*, 1993). Down-regulation of intercellular adhesion is initially required in order for tumour cells to leave the primary tumour. However, increased cellular adhesive capacity is required at a later stage for reattachment at the metastatic site (Ahmad & Hart, 1997). The major classes of cell adhesion molecules include the cadherins, integrins, intercellular adhesion molecules (ICAMs), immunoglobulin receptors and selectins. Alteration in expression of a number of these cell adhesion molecules has been implicated in metastatic spread. Expression of cadherins has received much attention in tumour metastasis. The extracellular domains of cadherins contain several calcium ion binding domains and these self-associate and cause intercellular adhesion (Aznavorian *et al.*, 1993; Sherbet & Lakshmi, 1997). Cadherins are expressed in a wide variety of cell types and their adhesion function plays a role in cell fate, segregation and differentiation (Takeichi, 1990; Braga, 1999). Loss or downregulation of E-cadherin in particular has been inversely correlated with invasive potential in several epithelial cancers including colorectal cancer (Ahmad & Hart, 1997; Sherbet & Lakshmi, 1997; Beavon, 1999; Perry *et al.*, 1999).

Altered expression of integrins is also associated with increased invasiveness. Integrins are receptors that recognise signals from the extracellular matrix and these are transduced *via* the cytoskeletal structures (Sherbet & Lakshmi, 1997). A large number of classical signaling pathways are now known to be activated by the interactions of cells with matrix proteins *via* integrins (Kumar, 1998). These receptors are composed of α and β subunits which combine to form a variety of heterodimers that are capable of binding a number of ligands in the extracellular matrix (ECM) including fibronectin, fibrin, vitronectin, laminin, type I collagen and thrombospondin (Kumar, 1998). The integrins, therefore, mediate cell-matrix interactions during cell migration through the ECM. Variable expression and

distribution of integrins has been associated with metastasis in epithelial cancers and expression of the fibronectin receptor was associated with enhanced colon tumour proliferation (Ahmad & Hart, 1996). In epithelial cancers, fewer integrins tend to be expressed when compared to their normal epithelial cell counterparts (Ahmad & Hart, 1997). Thus, an association of integrins with metastasis is related to their role in signal transduction which is central to cytoskeletal organisation, cell survival and cell proliferation pathways (Kumar, 1998).

Altered motility has been associated with an invasive tumour cell phenotype. Tumour cell lines known to be more highly invasive show a higher degree of motility than their low metastatic counterparts (Aznavorian *et al.*, 1993). Motility factors associated with increased movement in tumour cells include autocrine motility factor (AMF) or hepatocyte growth factor (HGF). In addition, altered expression of Motility Related Protein-1 (MRP-1/CD9), which has been shown to effect motility, has also been associated with metastasis in a number of cancers including colorectal (Ikeyama *et al.*, 1993; Higashiyama *et al.*, 1995; Miyake *et al.*, 1995 & 1996; Cajot *et al.*, 1997; Uchida *et al.*, 1999).

1.2.3 Proteolysis

Proteolytic breakdown of the extracellular matrix (ECM) is also essential for metastatic spread. The ECM is comprised of the interstitial stroma and the basement membrane. The major components of the basement membrane include type IV collagen, laminin and heparan sulphate proteoglycan. Traversal of the basement membrane is required at least three times during invasion: escape from the primary site, and both intra-and extra-vasation from the vasculature and therefore breakdown of the ECM is an essential step. The major classes of proteases include the serine proteinases (includes urokinase plasminogen activator uPA and plasmin), cysteine proteinases (includes cathepsins B and L), aspartyl proteinases (cathepsin D), integral membrane proteinases and the metalloproteinases (MMP) (Duffy, 1996; Ahmad & Hart, 1996; Sherbet & Lakshmi, 1997). Some of the proteases implicated in

metastasis include urokinase plasminogen activator, cathepsins B and D, and the MMPs. Plasminogen activator (uPA), as its name suggests, catalyses the conversion of plasminogen to active plasmin and this reaction requires the binding of uPA to its receptor. Plasmin is a broad spectrum protease and also activates latent proteases (Duffy, 1996). Interestingly, angiostatin is also derived from plasminogen (see Section 1.2.1) which emphasises the importance of plasminogen in metastasis. The cathepsins are lysosomal enzymes and the expression of cathepsins B and D has been extensively studied in relation to metastasis (Sherbet & Lakshmi, 1997). Cathepsin B over-expression has been correlated with basement membrane degradation and lymph node invasion in adenocarcinomas (Sherbet & Lakshmi, 1997). In addition, cathepsin B can activate certain MMPs and receptor-bound uPA. Enhanced expression of cathepsin D has been found in a variety of human neoplasms including colon carcinomas but its correlation with metastasis was variable (Sherbet & Lakshmi, 1997). The MMP family of proteases consists of three main groups: the type IV collagenases, the interstitial collagenases and the stromelysins. Altered expression of MMPs and their inhibitors, known as TIMPs (tissue inhibitor of metalloproteinases), have been associated with a variety of cancer metastases (Cockett *et al.*, 1998). MMPs are produced not only by neoplastic cells but also by the surrounding stromal fibroblasts (Ahmad & Hart, 1996). However, MMP-7 was found to be expressed in a majority of gastric and colon tumours, but this expression was restricted to the tumour cells and was not found in stromal cells or lymphocytes (McDonnell *et al.*, 1991). In addition, inhibitors of cysteine proteases, metalloproteinases and serine proteases are capable of blocking tumour cell invasion (Aznavorian *et al.*, 1993). Thus, it is the balance between the local concentration of activated enzymes and their endogenous inhibitors that is the determinant of proteolytic activity during metastasis (Aznavorian *et al.*, 1993) and further studies are required to decipher the interplay between proteases and their inhibitors and their role in metastasis.

1.2.4 Summary of metastasis and associated genes

In summary, metastasis is a complex process that requires altered expression of a range of molecules associated with angiogenesis, proteolysis, motility and cell adhesion. The metastatic phenotype is independent of the tumourigenic phenotype (Sobel, 1990) and therefore not all tumours will have the same metastatic capabilities. Moreover, individual tumours consist of subpopulations of cells with widely differing invasive and metastatic capabilities (Sherbet & Lakshmi, 1997). Acquisition of the metastatic phenotype is likely to involve the altered expression of a range of gene products rather than the induction of unique and specific genes. However, a number of metastasis-associated genes have been identified to date and these include *ras* (Sobel, 1990), *nm23* (Steeg *et al.*, 1988), *mta1* (Toh *et al.*, 1997), *18A2/mts1* (Sherbet & Lakshmi, 1997) and more specifically for colorectal metastasis: *car* (cell adhesion regulator) (Yamamoto *et al.*, 1996), *cox-2* (cyclooxygenase-2) (Tsujji *et al.*, 1997), acylphosphatase (Riley *et al.*, 1997), MRP-1/CD9 (Mori *et al.*, 1998), E-cadherin and α -catenin (Gofuku *et al.*, 1999), heparin/heparan sulphate (Wang *et al.*, 1999) and CD44 variants (Sherbet & Lakshmi, 1997; Ghigna *et al.*, 1998; Weg-Remers *et al.*, 1998; Yamane *et al.*, 1999). The genes listed above were all identified based on their differential expression being positively or negatively correlated with metastatic spread. Increased expression of metastasis-associated genes either enhances or suppresses the metastatic ability of tumours and, therefore, must exert control over the metastatic cascade. However, due to the complexity of the metastatic cascade and differences between tumours and individuals, a definitive role in metastasis still has to be established for the majority of metastasis-associated genes. Moreover, development of metastasis is unlikely to involve just one specific gene, but rather an accumulation of changes in a number of genes which effect angiogenesis, adhesion, motility, proteolysis and cell survival.

Early events in colon tumourigenesis which lead to the establishment of the primary tumour have been relatively well characterised as described above. However, the changes involved in primary tumour cells becoming metastatic is

'lagging behind'. Metastatic spread is the main cause of cancer-related death and therefore a greater understanding of this process is a pre-requisite for improved treatment. The behaviour and specificity of cell types is determined, in part, by differences in gene expression. Thus, tumour cells with differences in metastatic ability are likely to have different gene expression profiles. Examination of the gene expression differences between pre-metastatic and metastatic cell lines is likely to identify differentially expressed gene products may be important contributors to the metastatic process.

1.3 AN EXAMINATION OF COLON TUMOUR METASTASIS

1.3.1 Aim

The development of colon cancer is now understood to be a multistage process involving genetic alteration of specific oncogenes and tumour suppressor genes (Kinzler & Vogelstein, 1996). However, the genetic changes leading to the development of colorectal metastasis are not well known (Sampson-Johannes *et al.*, 1996). Some of the changes that primary tumour cells undergo before acquiring a metastatic phenotype will be reflected in their gene expression patterns. The overall aim of this study was to arrive at a better understanding of colon tumour metastasis by identifying mRNA species whose induction or repression is involved in the acquisition of metastatic competence. Cultured cell lines provide a convenient *in vitro* model for analysis and avoids the complications arising from mixed cell types in clinical material by providing a relatively pure population of tumour cells. In addition, the use of isogenic cell lines, i.e. those isolated from multiple tumours in the same patient, eliminates the possibility of isolating gene expression differences that are likely to exist between different individuals. Moreover, a relatively unlimited supply of RNA means that an extensive examination of gene expression patterns can be carried out.

1.3.2 Differential gene expression profiling methodologies

A number of techniques have been developed to identify differences in gene expression in cells of different types. Some of the more recent methods include Differential Display Reverse Transcription-PCR (DDRT) (Liang & Pardee, 1992), Serial Analysis of Gene Expression (SAGE) (Velculescu *et al.*, 1995), Suppression Subtractive Hybridisation (SSH) (Diatchenko *et al.*, 1996), cDNA microarray (DeRisi *et al.*, 1996; Lockhart *et al.*, 1996) and commercially available high-density cDNA blots such as Gene Discovery Array (Genome Systems Incorporated) or Application-Targeted Atlas cDNA expression Array (Clontech).

SAGE allows the quantitative analysis of thousands of transcripts at once, therefore generating important information on the relative abundances of mRNA transcripts between cell types. The final step of the SAGE protocol produces a library of clones for each cell type of interest. The SAGE clones contain short sequence tags that uniquely identify specific cDNAs (analogous to a 'barcode') and are subsequently quantified. Tag abundance from separate cell types of interest are compared and quantitative information regarding the differentially expressed species is produced (Velculescu *et al.*, 1995). SAGE was used for the gene expression analysis in this study and will be discussed further in Chapter 3.

1.3.3 The cell lines SW480 and SW620

The isogenic cell lines SW480 and SW620 were derived respectively from a primary colorectal adenocarcinoma and a lymph node metastasis in the same patient, respectively (Leibovitz *et al.*, 1976). SW480 was classified as Dukes' stage B and SW620 was classified as Dukes' stage C (Leibovitz *et al.*, 1976). Moreover, initiation of SW620 coincided with widespread metastasis in the patient and therefore, SW620 cells are likely to be representative of cells that metastasised to various visceral organs of the patient (Sampson-Johannes *et al.*, 1996). These cell lines provide an *in vitro* model of pre-metastatic and metastatic stages of colon tumourigenesis. The cell lines have been examined previously by others (Witty *et al.*, 1994; Sampson-Johannes *et al.*, 1996) and were shown to maintain the invasive characteristics of the tumours from which they were derived. Witty *et al.* (1994) showed that only SW620 cells, and not SW480 cells, were capable of producing hepatic metastases when injected into the caecum of nude mice. A separate study employed a colonisation assay using SCID (severe combined immunodeficient phenotype) mice engrafted with human fetal lung (HFL) tissues (Sampson-Johannes *et al.*, 1996). HFL grafts can maintain endothelium of human origin for up to 20 weeks after implantation. SW480 cells did not form tumours within the HFL grafts whereas SW620 cells did. The SCID-hu mouse model allows the basis of interactions between human cancer cells and the

human tissue microenvironment to be investigated. Various human tumour cell lines, in addition to SW480 and SW620, were analysed and the tumourigenic activity in the SCID-hu mice faithfully reflected the clinical features of the original cancer (Namikawa & Shtivelman, 1999). Analysis of adhesion molecules, previously implicated in metastasis, in SW480 and SW620 showed a loss of integrin $\alpha 3$ and NCAM in SW620 relative to SW480 (Sampson-Johannes *et al.*, 1996). Two separate studies have also found that the SW480 primary tumour cells exhibited a higher migration potential *in vitro* relative to the SW620 metastatic cells and this was also found in similar paired cell lines (Cajot *et al.*, 1997; Kubens & Zanker, 1998). This may be due to expression of Motility Related Protein-1 (MRP-1) and is related to poor prognosis (Ikeyama *et al.*, 1993; Higashiyama *et al.*, 1995; Miyake *et al.*, 1995 & 1996; Cajot *et al.*, 1997; Uchida *et al.*, 1999).

Additional studies on SW480 showed that it harbours a typical APC truncation mutation (Smith *et al.*, 1993; Mann *et al.*, 1999), a mutant p53 (Hermeking *et al.*, 1997; Polyak *et al.*, 1997), a predominant localisation of β -catenin in the nuclei (Sheng *et al.*, 1998) and exhibits the chromosome instability (CIN) phenotype (Cahill *et al.*, 1998). Therefore, SW480 represents the most common form of colorectal cancer (see Sections 1.1.2 and 1.1.3) and both SW480 and SW620 maintain their tumour and metastatic characteristics in culture and, thus, could be used to study gene expression in colon tumour metastasis.

1.3.4 Objectives

The overall objective of this study was to examine the gene expression patterns in metastatic colon tumour cell lines, i.e. SW480 and SW620, and identify any differences between them. Genes that demonstrated differential expression would be analysed further in additional metastatic cell lines and, if possible, in clinical material in order to identify metastasis-associated genes. The specific objectives were as follows:

- Implementation of the method of Serial Analysis of Gene Expression (SAGE) under licence from Vogelstein & Kinzler (Johns Hopkins University, MD, USA) in the cell lines SW480 and SW620.
- Analysis of the SAGE data using SAGE software version 1.0 (Vogelstein & Kinzler) and comparison to Genbank flatfile release 101.0. Selection of a subset of differentially expressed genes with possible association to the metastatic phenotype and confirmation of differential expression by an independent method.
- Examination of expression of candidate metastasis-associated genes in other cell lines and in fresh-frozen/paraffin-embedded tumour specimens by semi-quantitative RT-PCR (He *et al.*, 1995)/immunohistochemistry if possible.
- Further analysis of DDRT clone PM208. A previous study applied DDRT-PCR to SW480 and SW620, and a number of differentially expressed products were identified, one being PM208.

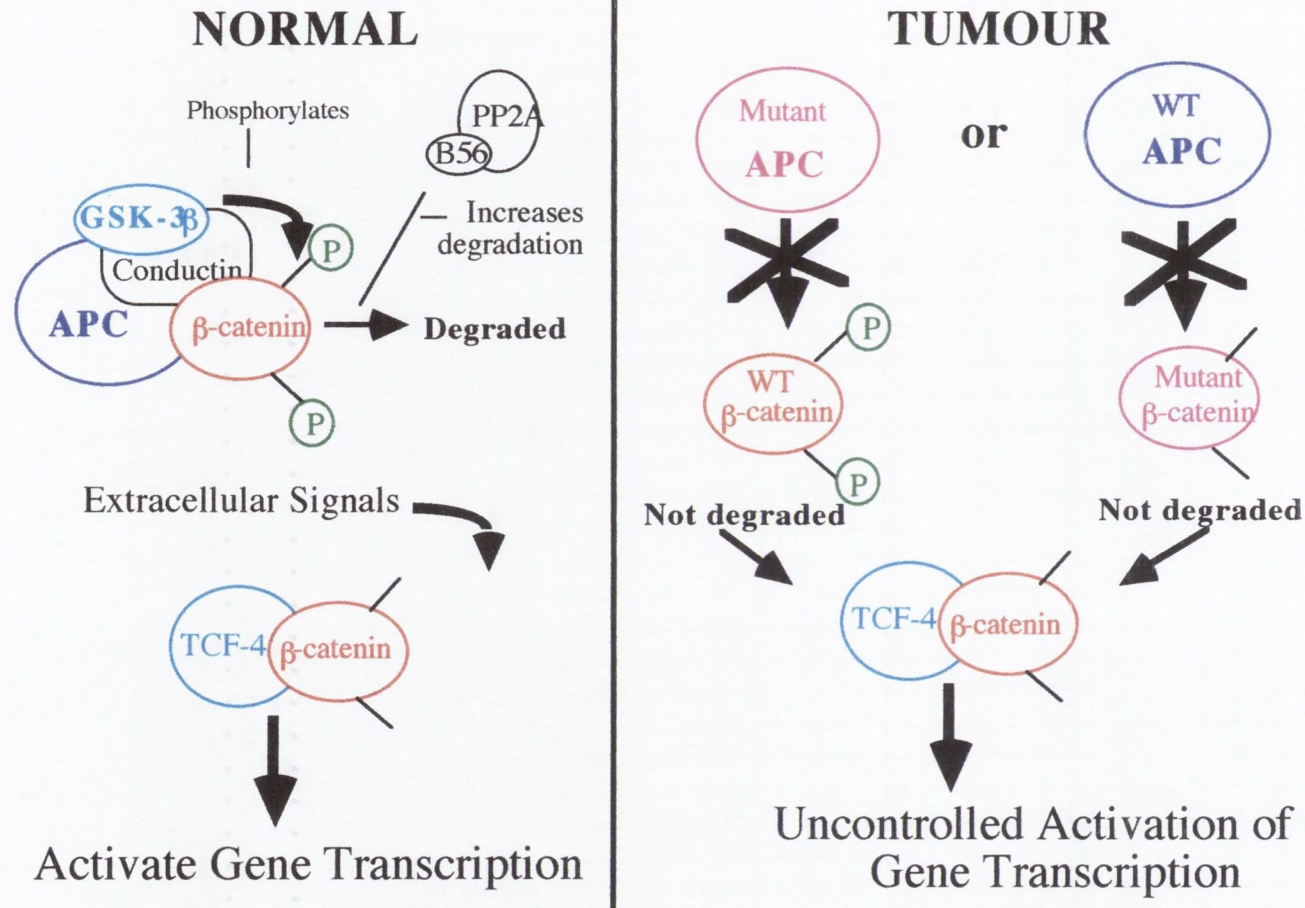


Figure 1.1 APC and β -catenin interactions in normal and tumour cells. The interactions illustrated above are based on a number of recent publications (see text for references). In normal cells in response to extracellular signals, β -catenin interacts with the transcription factor Tcf-4 and this complex activates transcription of a number of target genes. APC indirectly controls transcription by interacting with phosphorylated β -catenin in a complex which consists of GSK-3 β , conductin and possibly B56 bound PP2A, and promotes its degradation. Thus, mutations in APC or β -catenin that occur in colon tumours can disrupt binding and lead to uncontrolled gene transcription.

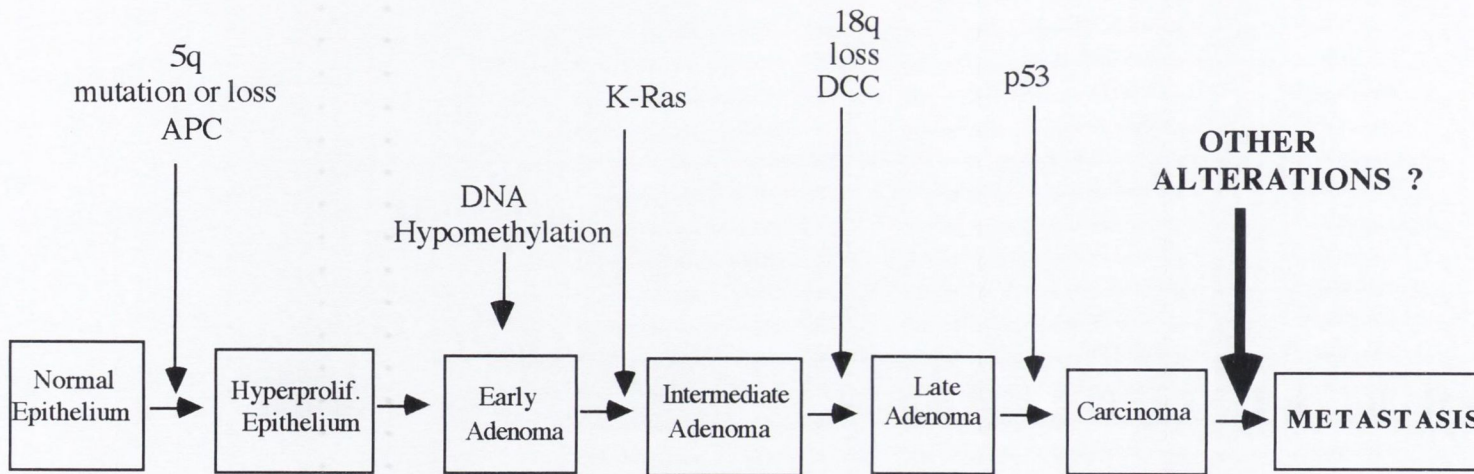


Figure 1.2. The genetic model for colorectal cancer progression. (Adapted from Fearon & Vogelstein, 1990; Kinzler & Vogelstein, 1996). Mutation or loss of the APC gene (or β -catenin) is the initiating event and results in hyperproliferation leading to an adenoma. An accumulation of genetic alterations including K-Ras, DCC, p53 allows the adenoma to progress to the carcinoma stage. Other alterations in the established tumour give certain cells the characteristics required to spread to other parts of the body, i.e. to become metastatic.

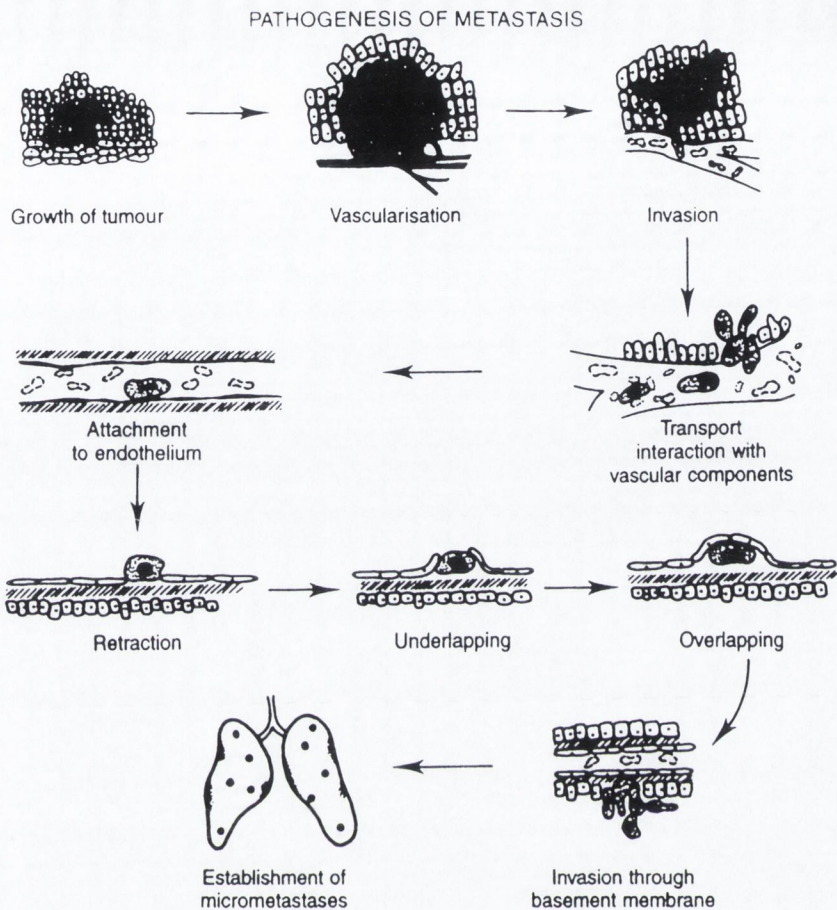


Figure 1.3 The metastatic process (taken from Sherbet & Lakshmi, 1997).

Invasive cells are capable of leaving the primary tumour and spreading to distant locations throughout the body. This process, known as metastasis, involves gene products associated with basement membrane breakdown, angiogenesis, adhesion and motility.

Table 1.1 Dukes' original classification of colorectal tumours

Dukes' Stage	Description
A	Growth of tumour limited to the rectal wall.
B	Tumour growth extended to extrarectal tissues but no metastases in regional lymph nodes.
C	Metastases in regional lymph nodes.
C ₁	Only regional lymph nodes involved.
C ₂	Nodal spread to the level of the point of ligature of the blood vessels.

The stages written in **bold** were described originally by Dukes (1932) and Dukes' C was later subdivided C₁ and C₂ (Gabriel *et al.*, 1935).

CHAPTER 2

Materials and Methods

2.1 MATERIALS

2.1.1 Reagents

General chemicals

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

B.D.H. Ltd., UK: Chloroform, Copper sulphate, Ethanol, Ethylenediaminetetraacetic acid disodium salt (EDTA), Glycine, Iso-amyl alcohol, Iso-propanol, Nonidet P40, NNN'N'-Tetramethylethylenediamine (TEMED), Paraffin liquid, Potassium sodium tartrate, Sodium bicarbonate, Sodium chloride, tri-Sodium citrate, Sodium dodecyl sulphate, Sodium hydroxide, tetra-Sodium pyrophosphate, Sucrose, Tris (hydroxymethyl) methylamine, Urea.

Boehringer GmbH., Germany: Isopropylthio- β -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).

Oxoid, UK: Agar, Tryptone, Yeast extract.

Sigma-aldrich, UK: Acrylamide/bis-acrylamide 30% solution electrophoresis grade (ratio 37.5:1), Agarose, Ammonium acetate, Ammonium persulphate, Ampicillin, Bovine serum albumin Fraction V, Bromophenol blue, Dextran blue, Diethylpyrocarbonate (DEPC), Dimethyl Sulphoxide (DMSO), Ethidium bromide, Formamide, Formaldehyde 37% solution (formalin), Leupeptin, Magnesium chloride, Mineral oil, N,N-Dimethylformamide, 3-[N-Morpholino]-propane-sulphonic acid (MOPS), Phenol red, Phenylmethylsulfonyl fluoride α -toluenesulfonyl fluoride (PMSF), Sephadex G-50, Sodium sulphate, Triton X-100.

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

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

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

Photographic and X-ray film

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

Polaroid 665 Film: Polaroid Corp., USA.

Tissue Culture

GibcoBRL, Life Technologies, UK: DMEM (Dulbecco's modified Eagle's medium), DMEM (Dulbecco's modified Eagle's medium): F12 nutrient mixture (Ham's) (1:1) medium with L-glutamine, 15mM hepes buffer and pyridoxine, F-12K nutrient mixture (Kaighn's modification) medium, Foetal bovine serum, Leibovitz L-15 medium, Dulbecco's PBS (Phosphate buffered saline), Penicillin-streptomycin (10,000u/ml; 10,000 μ g/ml), RPMI 1640 medium with 25mM hepes buffer and L-glutamine, Trypsin-EDTA (2.5%), Versene.

Corning Costar, NL: 75cm³ flasks-7170, cell scrapers.

Molecular biology

Amersham International PLC, England: [α -³²P] dCTP (6000Ci/mmol), Cycle sequencing kit-RPN2436, Hybond N⁺nylon membrane, Megaprime™ labelling kit, Rediprime® labelling kit II.

Biotech laboratories Inc., USA: Ultraspec™ II RNA Isolation System.

Boehringer GmbH., Germany: Glycogen 20µg/µl, *Hin* dIII (10u/µl), *Eco* RI (10u/µl), RNase A.

Corning Costar, NL: SpinX centrifuge tubes-8160.

GibcoBRL, Life technologies, UK: DNaseI (1u/µl) amplification grade, T4 DNA Ligase high concentration (5u/µl), T4 DNA Ligase (1u/µl), λ DNA (0.45µg/µl), Random primers (3µg/µl), Superscript™ II Moloney Murine Leukaemia Virus Reverse Transcriptase.

I.M.A.G.E (Integrated Molecular Analysis of Genomes and their Expression) Consortium: EST clones W58406, N52605, T65549 obtained *via* HGMP (Human Genome Mapping Project), UK.

NEB (New England Biolabs), UK: *Not* I (10u/µl), *Ksp* I (10u/µl).

Perkin-Elmer, USA: AmpliTaq™ (5u/µl).

Pharmacia, UK: Cy5 labelled M13 Forward and Reverse primers (Table 2.1).

Promega, USA: dNTPs, *Hae* III (10u/µl), RNasin (40u/µl), *Taq* DNA Polymerase (5u/µl), Wizard® PCR preps DNA purification system- A7170, Wizard® SV plasmid purification kit.

Sigma-aldrich, UK: Lysozyme.

Stratagene, USA: pBluescript II SK(+) DNA vector, pCR-Script Amp SK(+) cloning kit, Epicurian® Coli XL1-Blue MRF Kan supercompetent cells.

Protein analysis

B.D.H. Ltd., UK: DPX mountant.

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

Dako, DK: Monoclonal mouse anti-human amyloid A component, Envision™+ system HRP (DAB) anti-mouse antibody kit.

GelmanSciences, UK: BioTrace™ transfer membrane.

NEB (New England Biolabs), UK: Phototope® -Star western blot detection kit.

NeoMarkers, UK: Monoclonal mouse anti-keratin 5/8 Ab-1 (clone C-50).

Munktell, SW: Filter paper.

Sigma-aldrich, UK: High (205kDa to 29kDa) and Low (66kDa to 14.2 kDa) range molecular weight markers, Ponceau S solution.

Shandon, USA: Haematoxylin solution.

2.1.2 Stock solutions

General

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

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

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

Loading dye 0.25% bromophenol blue (w/v), 0.25% xylene blue (w/v), 30% glycerol.

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

Plasmid DNA preparation

LB: 10g NaCl, 10g tryptone, 5g yeast extract per litre of H₂O, pH7.0.

LB agar: LB plus 1.5% (w/v) agar.

SOB: 2% tryptone, 0.5% yeast extract, 0.05% NaCl, 2.5mM KCl, 10mM MgCl₂.

TES: 50mM Tris HCl, pH 8.0, 25% Sucrose, 2mM EDTA, pH 8.0.

M-STET: 5% Triton X-100, 50mM EDTA, pH 8.0, 50mM Tris HCl, pH8.0, 5%.

RNA

DEPC-H₂O: 0.01% (v/v) DEPC was added to dH₂O, mixed well and allowed to sit overnight at room temperature. The treated water was then autoclaved for 30 minutes to remove any traces of DEPC.

MOPS (10x): 0.2M 3-[N-Morpholino]-propane-sulphonic 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, 1M, pH6.8: 13.14g Na₂HPO₄, 14.82g NaH₂PO₄ up to 200ml with H₂O.

Wash A: 0.5% (w/v) BSA Fraction V, 5% SDS, 40mM Sodium phosphate buffer pH 6.8, 1mM EDTA.

Wash B: 1% SDS, 40mM Na phosphate buffer pH6.8, 1mM EDTA.

Strip buffer: 0.1xSSC, 0.1% SDS .

Protein

Electrophoresis buffer (5x): 15g Tris, pH8.3, 72g Glycine, 5g SDS per litre.

Sample buffer (5x): 4ml 1% Bromophenol blue, 20ml 1.5M Tris-HCl, pH6.8, 17.5ml 20% SDS, 53.5ml glycerol, 2.5ml β -mercaptoethanol per 100ml.

Coomassie blue: 0.1% Coomassie blue G250, 50% Acetic acid, 20% Methanol.

Buffer A: 20mM Tris-HCl, pH7.5, 0.25M sucrose, 2mM EGTA, 2mM EDTA, 1mM PMSF, leupeptin (10 μ g/ml).

Buffer B: 20mM Tris-HCl, pH7.5, 1% (w/v) 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.

Transfer buffer: 0.384g SDS, 5.82g Tris, 2.94g Glycine, 200ml Methanol per litre.

PBS (10X): 14.5g Na₂HPO₄, 2g KH₂PO₄, 2g KCl and 80g NaCl, pH7.4.

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

Wash buffer (10x): 100mM Tris-HCl, pH9.5, 100mM NaCl, 10mM MgCl₂.

TBS (Tris buffered saline): 100ml 0.05M Tris pH7.6, 7.65g NaCl, 25 μ l Tween 20 per litre.

2.1.3 Cell lines

The following cell lines were obtained from the American Type Culture Collection (ATCC), MD, USA: CCL-228 (SW480), CCL-227 (SW620), CCL-233 (SW1116), CCL-229 (LoVo), CCL-224 (Colo 201), CCL-248 (T84).

The cell line 498LI was the kind gift of Dr. Meenhard Herlyn, The Wistar Institute, Philadelphia, USA.

2.2 TISSUE CULTURE METHODS

2.2.1 Cell Culture

The cell lines SW480, SW620, SW1116 and 498LI were cultured in Leibovitz L15 medium supplemented with 10% (v/v) foetal bovine serum at 37°C in the absence of CO₂. The cell lines LoVo, Colo201, T84 and mouse 3T3 fibroblasts were all maintained at 37°C in the presence of 5% CO₂ in their respective growth media. LoVo was cultured in 90% F-12K medium with 10% foetal bovine serum, Colo201 was cultured in RMPI 1640 supplemented with 20% foetal bovine serum and T84 was cultured in 95% DMEM: Ham's F12 (1:1) and 5% foetal bovine serum. All cell lines were grown in the presence of 100units/ml penicillin and 100µg/ml streptomycin. Cells were passaged by incubating with 0.25% Trypsin-EDTA (GibcoBRL) at 37°C for 10 to 15 minutes. The subcultivation ratio was usually between 1:2 to 1:10. The approximate passage number for each cell line was as follows: SW480: 99; SW620: 86; SW1116: 41; 498li: unknown; LoVo: 30; Colo201: unknown; T84: 52.

2.3 MOLECULAR BIOLOGY METHODS

2.3.1 Electrophoresis methods

Agarose gels

Agarose gels were made by boiling 1-2.5% agarose (w/v) in 1xTAE or 1xTBE (Section 2.1.2). The electrophoresis buffer was 1xTAE or 1xTBE as appropriate. Samples were usually mixed with at least 0.2 volumes of loading dye (Section 2.1.2). After electrophoresis, gels were stained in an ethidium bromide bath (1.5µg/ml) and visualised on a UV transilluminator.

Formaldehyde agarose gels were composed of 1-1.5% (w/v) agarose dissolved in 1x MOPS buffer, cooled to 50°C and formaldehyde solution added to a final concentration of 6.3%. The electrophoresis buffer was 1x MOPS.

Polyacrylamide gels

Non-denaturing polyacrylamide gels were made using 30% Protogel (National Diagnostics) acrylamide solution. Per 60ml gel mix- 16-24ml (8-12%) 30% Protogel, 6ml 10x TBE, 21-29 ml dH₂O, 378µl 10% APS and 35µl TEMED. Gels were allowed to polymerise for at least 40 minutes.

Denaturing gels for the ALFexpress™ automated sequencer consisted of two types: sequencing 'long' gels and 'short' gels. Sequencing gels were made by mixing 25g urea, 7.2ml Longer Ranger™ gel solution, 7.2ml 10x TBE, the volume was made up to 60ml with dH₂O and polymerised by addition of 375µl 10% APS and 38µl TEMED. Sequencing gels were allowed to set for at least 2 hours. 'Short' gels were made by mixing 40ml 6% Gene-PAGE or Easigel™ solution with 400µl 10% APS and 40µl TEMED. 'Short' gels were allowed to polymerise for at least 1 hour.

Size markers

The λ-*Hin* dIII marker was made by incubating 4.5µg of λ DNA (GibcoBRL) with 40 units of *Hin* dIII (Boehringer) restriction enzyme in a final volume of 100µl containing 1x reaction buffer (Boehringer) at 37°C for 2-3 hours. This yielded fragment sizes of 23,130bp, 9,416bp, 6,557bp, 4,361bp, 2,322bp, 2,027bp, 564bp and 125bp. An aliquot of 5µl of this digest was used for agarose gels.

The pBS-*Hae* III marker was made by incubating 15µg pBluescript II SK(+) DNA vector with 50 units of *Hae* III (Promega) in a final volume of 300µl containing 1x reaction buffer (Promega) overnight at 37°C. This yielded fragment sizes of 602bp, 458bp, 434bp, 290bp, 267bp, 254bp, 174bp, 142bp, 102bp, 80bp, 79bp, 50bp, 18bp, 11bp. An aliquot of 3µl was used for agarose gels and a 5µl aliquot was used for nondenaturing polyacrylamide gels.

2.3.2 RNA extraction and purification

The cell lines were harvested for total RNA extraction at approximately 70% confluence. The cells were lysed directly in the culture flask using the Ultraspec™ II Isolation System (2ml/75cm³ flask) according to the manufacturer's instructions. Yield and quality of RNA were determined by A₂₆₀/A₂₈₀ spectrophotometer readings (Sambrook *et al.*, 1989). A ratio of 1.8 to 2.0 was indicative of intact, good quality total RNA. Integrity of RNA was assessed by mixing a 1:1 ratio of approximately 1µg of an RNA sample and denaturing buffer, heating to 70°C for 10 minutes, chilling on ice and electrophoresing on a 1% agarose gel. Good quality RNA has a 28s:18s ratio of 2:1 as shown in Figure 2.1.

2.3.3 Estimation of nucleic acid concentration

Spectrophotometry

DNA or RNA samples were diluted to a volume of 1ml with dH₂O and placed in a quartz cuvette. Concentrations were determined from readings at A_{260nm} as described previously (Sambrook *et al.*, 1989).

Ethidium Bromide Dot Quantitation

DNA standards were prepared using pure pUC19 plasmid DNA and diluted to the following concentrations: 0ng/µl, 1ng/µl, 2.5ng/µl, 5ng/µl, 7.5ng/µl, 10ng/µl and 20ng/µl. RNA standards were prepared using either single stranded DNA or an RNA sample of known concentration. An equal volume of each standard (usually 4µl) or each diluted unknown DNA/RNA sample was added to the same volume of ethidium bromide solution (1µg/ml). Samples were mixed well and spotted on plastic wrap ('Clingfilm') overlaying a UV transilluminator. DNA/RNA concentration was estimated by visual comparison of the intensities of the standards to those of the samples as shown in Figure 2.2.

DNA DipStick™ kit

The DNA DipStick™ kit (Invitrogen) can be used for estimating the concentration of single- or double-stranded DNA, RNA, or oligonucleotides. It has a similar basis to the ethidium bromide dot quantitation, i.e. comparison of dot intensities to a known standard. The kit was used following the manufacturer's instructions.

2.3.4 Ethanol precipitations

DNA

DNA was precipitated by raising the sample volume to 200µl, adding 3µl glycogen (20µg/µl), 100µl 10M ammonium acetate and 700µl of absolute ethanol, vortexing well and storing at -20°C for at least 1 hour. The DNA pellet was collected by centrifuging in a microfuge for 15 minutes at full speed at 4°C. The pellet was washed 1-2 times with 70% ethanol, air dried and resuspended in sterile H₂O.

RNA

RNA was ethanol precipitated by addition of 50µg glycogen/ml of RNA sample, 0.1 volume of 7.5M ammonium acetate, two volumes of ethanol at -20°C and mixed. After incubation overnight at -20°C the RNA was collected by centrifuging at 2,600 x g at 4°C for 30 minutes. The RNA pellet was washed once with 75% ethanol and centrifuged for a further 10 minutes. The pellet was finally air dried and resuspended in DEPC-treated H₂O.

2.3.5 Serial analysis of gene expression (SAGE)

Implementation of the SAGE protocol is detailed in Chapter 3.

2.3.6 Computer software

Oligo 4 program: National Biosciences Inc., USA.

Evaluation™, Allele Links™: Pharmacia, UK.

GraphPad: Prism™, USA.

Bio1D v6.32: Vilber Lourmat Biotechnology, France.

SAGE software, version 1.0: Johns Hopkins University, USA.

GCG Wisconsin Package, ACNUC retrieval system: Irish National Centre for
Bioinformatics.

BLAST, Entrez, Unigene: National Centre for Bioinformatics, USA
(<http://www.ncbi.nlm.nih.gov/>)

2.3.7 Semi-quantitative RT-PCR¹

DNase I treatment

All RNA samples used for reverse transcriptions were first treated with DNase I to remove any potentially contaminating genomic DNA. Reactions were assembled in a final volume of 40µl consisting of 1 unit of DNase I amplification grade, 20µg of total RNA, 40 units of RNasin and 1x reaction buffer (20mM Tris-HCl, pH8.4, 2mM MgCl₂, 50mM KCl). Reactions were allowed to proceed for 15 minutes at 25°C and then terminated by addition of 2.5mM EDTA and heated at 70°C for 10 minutes. DNase treated RNA samples were stored at -70°C.

Reverse transcription

Reverse transcription reactions were assembled in a volume of either 20µl or 40µl in a reaction mix containing 1-2 µg of total RNA, 100-200pmol random primers, 1x first strand buffer (25mM Tris-HCl pH8.3, 37.5mM KCl, 1.5mM MgCl₂) (GibcoBRL), 0.01M DTT, 0.625mM dNTPs, 40 units RNasin and 200 units Superscript™ II Moloney Murine Leukaemia Virus Reverse Transcriptase. The RNA and random primers were denatured at 70°C for 10 minutes, then placed on ice for 2 minutes

¹ Abbreviation : Reverse Transcription-Polymerase Chain Reaction (RT-PCR).

before addition of the remaining components except the Superscript™ II. The reaction mix was incubated at room temperature for 10 minutes, then preheated at 37°C for 2 minutes before adding the Superscript™ II. Reactions were incubated for at least 1 hour at 37°C. Completed reverse transcription reactions were stored at -20°C.

Genomic DNA contamination test

All treated RNA samples and cDNA products were tested for genomic contamination by a PCR test with an intron-spanning primer set for PKD1 (nucleotides 50619 and 50875 of the genomic sequence L39891) (The European Polycystic Kidney Disease Consortium, 1994). PCR reactions were carried out in a 20µl volume containing 1x Thermophilic DNA Polymerase buffer (Promega), 1.25mM MgCl₂, 0.25mM dNTPs, 10pmol of primers 3A3C1 and 3A3C3 (The European Polycystic Kidney Disease Consortium, 1994), 0.2 units *Taq* DNA Polymerase and 2µl of template (RNA or cDNA). Reactions were amplified for 36 cycles of 94°C 1 minute, 61°C 1 minute, 72°C 1 minute. Genomic DNA is detected by the presence of a PCR product of 256bp in size. Reverse transcribed cDNA amplifies as a PCR product of 181bp in size. The difference in size is due to the presence of an intron in the genomic sequence and its absence in the spliced reverse transcribed mRNA.

Semi-quantitative PCR amplification

Semi-quantitative RT-PCR amplifications were based on the method of He *et al.* (1995) with modifications including a Cy5 labelled primer in all PCR amplifications. Gene specific primer sets were designed using the Oligo 4 program and are listed in Table 2.1. PCR amplifications were usually carried out in a 50µl reaction mix containing 1x PCR buffer II (10mM Tris-HCl, pH8.3, 50mM KCl, 0.01% (w/v) gelatin) (Perkin-Elmer), 0.2mM dNTPs, 0.4µM primers and 1.25 units AmpliTaq™. Other conditions such as the amount of MgCl₂ and the annealing temperature were

optimised for each specific primer pair as described in Table 2.2. All PCR amplifications were 'hotstarted', i.e. the AmpliTaq™ was only added after the reaction had reached a temperature of greater than 70°C and was then heated for 2 minutes at 95°C. This reduces the number of 'mispriming' events that could potentially occur during the initial stages of PCR amplification. PCR reactions were amplified at 94°C for 1 minute, the optimised annealing temperature for 1 minute and 72°C for 1 minute. A range of different cycle numbers (optimised for each PCR) were sampled to ensure that the exponential phase of the PCR reaction was being observed. Amplification of β -actin was carried out in parallel to ensure equivalent amounts of input RNA and to control for efficiencies of reverse transcription and PCR amplification. PCR reactions were performed on a Hybaid™ thermal cycler or a DNA Engine (MJ Research, USA).

Electrophoretic analysis

The fluorescently labelled PCR products were initially analysed on 12% nondenaturing polyacrylamide gels stained with ethidium bromide and visualised on a UV transilluminator. The amount of each specific PCR product was quantified by mixing with β -actin PCR products and electrophoresing on 6% denaturing polyacrylamide gels on an ALFexpress™ automated sequencer. The peak areas generated by the Allele Links™ software represent the amount of fluorescently labelled PCR product. All peak area values were expressed as a ratio relative to the amount of β -actin PCR product.

2.3.8 Gel purification of PCR products/DNA fragments

Gel separated PCR products or DNA restriction fragments were purified from high melting temperature agarose using either the Wizard® PCR Preps DNA purification system or SpinX centrifuge tubes.

Fragments purified by the Wizard® kit were electrophoresed in 1xTAE agarose gels and purification was carried out according to the manufacturer's protocol

except that elution was carried out at 37°C rather than 65°C and the elution volume was reduced to 30µl.

Fragments purified using Spin-X centrifuge tubes were electrophoresed in 1xTBE agarose gels and purification was carried out as follows. The DNA band of interest was excised from an agarose gel and sheared by placing the gel slice in a 0.5ml eppendorf tube with a single hole pierced in the bottom. The gel was sheared and collected by placing the 0.5ml tube into a 2ml tube and centrifuging in a microfuge at full speed for approximately 5 minutes. TE or H₂O was added to the gel (usually 300µl) and incubated for 15 minutes at 37°C. The gel mix was transferred into the filter cup of the Spin-X tube and centrifuged for 5 minutes at full speed in a microfuge. The purified DNA was then ethanol precipitated from the eluate and resuspended in a suitable volume of TE or H₂O.

2.3.9 Plasmid DNA preparation and DNA sequencing

Plasmid preparation

Plasmid DNA was prepared for sequencing using either the Wizard® SV plasmid purification kit according to the manufacturer's protocol or a modified boiling lysis plasmid preparation as detailed below.

A 10ml overnight culture was grown in LB or SOB plus appropriate antibiotic at 37°C shaking at approximately 220rpm. Cells were harvested by centrifugation at between 2-3 krpm. LB was drained off and the pellet was resuspended in 60µl of TES and transferred to an eppendorf tube. A 20µl volume of lysozyme (Sigma) (40mg/ml in 250mM Tris HCl, pH 8.0) was added to the resuspended pellet and incubated on ice for 10 minutes. M-STET (550µl) was then added to the tube and incubated at room temperature for 10 minutes. The lid of the tube was pierced and the contents were boiled for 1-1.5 minutes and centrifuged for 15-30 minutes in a microfuge. The resulting pellet was removed with a toothpick. The supernatant was RNAase treated by adding 5-10µl of heat treated RNAase (10mg/ml) and incubated at 37°C for 15 minutes. The supernatant was then phenol-chloroform extracted and the

upper phase was transferred to a new eppendorf tube. The plasmid DNA was precipitated by the addition of 0.6 volumes of isopropanol. The tube was vortexed well and incubated at room temperature for 15 minutes. The DNA pellet was collected by centrifuging for 10 minutes at room temperature. The pellet was washed (often not visible) with 70-80% ethanol and air dried. The pellet was resuspended in 30 µl of sterile water.

Cycle sequencing

Sequencing was performed using a Cycle Sequencing kit (Amersham) with Cy5 labelled M13 Forward or Reverse primers (Table 2.1) following the manufacturer's instructions with minor modifications. Briefly, 2.5µl of Cy5 Forward or Reverse primer (20pmol/µl) was added to approximately 200ng of template DNA in 10µl of sterile water and vortexed. Four separate tubes were labelled A, C, G and T for each reaction. An equal volume (1µl) of the A, C, G or T reagent from the sequencing kit was added to the appropriate tube. Each reagent is supplied premixed containing the appropriate concentrations of buffer, dNTPs and ddNTPs and Thermo Sequenase™. To each labelled tube, 3µl of the primer and template mix was added and mixed gently with the pipette tip. Mineral oil was overlaid and the reactions were cycle sequenced using the following conditions: 95°C for 1min., 1 cycle; 94°C for 45sec., 60°C for 45sec., 72°C for 1 min., 32 cycles; 72°C for 2 mins. The supplied loading dye (4µl) was added and each reaction was loaded on a 6% denaturing polyacrylamide gel. Sequence reactions were analysed by the ALFexpress™ sequencer using the Evaluation™ software (Pharmacia, UK).

2.3.10 Cloning

Hox11 Cloning

The 309bp Hox11 PCR product was 'blunt-end' cloned into the *Srf* I site of pCR-Script AmpSK(+) (Stratagene). The PCR product was blunt ended with *Pfu* DNA Polymerase, which has the lowest error rate of any thermostable DNA Polymerase,

prior to ligating into the pre-digested pCR-Script according to the manufacturer's protocol (Stratagene). The ligation was transformed into Epicurian® Coli XL1-Blue MRF Kan supercompetent cells (Stratagene) and positive clones were selected on LB agar plates containing ampicillin (50µg/ml), X-gal (35µl of 50mg/ml stock solution per plate) and IPTG (20µl of 100mM stock solution per plate). Positive clones were cycle sequenced as outlined above and their identity confirmed as Hox11 by a BLAST homology search (Altschul *et al.*, 1990). One of the clones was renamed pAPD.1 and used to generate a probe for Northern blot analysis (Section 2.3.11).

2.3.11 Northern Blotting

Radiolabelling

DNA fragments were radiolabelled with [α -³²P] dCTP by the random prime method using the Megaprime™ Labelling Kit (Amersham) or the Rediprime® labelling kit II (Amersham) under conditions recommended by the manufacturer. Radiolabelled probes were purified by separating through a sephadex G-50 column.

Probes

β -actin-1: an *Eco* RI purified fragment from a cloned β -actin cDNA product including bases 245-356 (X00351 cDNA sequence).

β -actin-2: a gel purified cDNA product (254bp) generated by PCR using primers β -actin-L.U and β -actin-L.L (Table 2.1) under the conditions described in Table 2.2.

Ipp2-1 : the upstream IPP-2 probe was amplified from a cloned IPP-2 cDNA ("GAP clone") using the primers GAPU and GAPL (Table 2.1). The GAP clone contains an insert size of 896bp and overlaps the 2kb IPP-2 sequence (X78873) and the cDNA clone HscOha41 (F08719) (see Chapter 5).

Ipp2-2: the downstream IPP-2 probe, which overlaps clone PM208, was a gel purified *Not I/Eco RI* fragment from I.M.A.G.E clone 780-B15 (W58406) (see Chapter 5).

Hox11: the Hox11 probe was a gel purified *EcoR I/Ksp I* digested fragment from cDNA clone pAPD.1 (Section 2.2.10).

Cystatin S, Keratin 5, hnRNP C, hnRNP L, SAA, SFp32, homologue of yeast S28: these probes were all gel purified cDNA products generated by PCR using the primers listed in Table 2.1 under the conditions described in Table 2.2.

Capillary blotting

Total RNA was isolated as described in Section 2.2.2. RNA samples with an estimated total amount of 20µg were resolved on 1% denaturing formaldehyde gels. The RNA was then transferred to Hybond N⁺ nylon membrane by capillary blotting with 20xSSC and fixed in 0.05M NaOH according to the manufacturer's instructions (Amersham). RNA samples were electrophoresed in parallel and stained in ethidium bromide to ensure that gel conditions were optimal. RNA samples that were blotted were not stained with ethidium bromide as this may have interfered with transfer. RNA size was determined by migration relative to 18s and 28s rRNA and λ *Hin* dIII marker.

Hybridisation

Radiolabelled DNA probes were hybridised to the blot using the following procedure. The blot was placed in a 'rolling bottle' (Hybaid™) with 10ml of hybridisation solution containing 0.1g BSA Fraction V, 4ml H₂O, 0.7g SDS, 5ml 1M Na Phosphate buffer pH 6.8 and 20µl 0.5M EDTA, pH 8.0 and allowed to prehybridise for 30 minutes at 65°C. The radiolabelled probe was denatured at 95°C for 5 minutes,

chilled on ice and added to the hybridisation solution. Hybridisations were carried out overnight at 65°C.

Washing and Stripping

The blot was usually washed for two intervals of 10 minutes with pre-heated Wash A at 65°C. Further wash steps were determined based on the background signal. Further washing was for 2-4 intervals of 10 minutes with pre-heated Wash B solution at 65°C. After washing, the blot was placed in a cassette with intensifying screens and X-ray film (Fuji) and exposed for an appropriate number of hours at -70°C. Probes were removed by preheating strip buffer to 95°C and adding to the nylon filter and washing 2-4 times at room temperature.

2.4 PROTEIN ANALYSIS METHODS

2.4.1 Electrophoresis methods

SDS-PAGE gels were made by mixing 3.33ml (10%) or 5.0ml (15%) of 30% acrylamide solution (Sigma) with 3.8ml 1M Tris pH9.1, 100µl 10% SDS, 75µl 10% APS, 5µl TEMED and raising the total volume to 10ml with dH₂O. This gel mix formed the bottom two thirds of the gel with the stacking gel forming the top one third. The stacking gel was made of 0.65ml 30% acrylamide solution, 1.25ml dH₂O, 3.1ml 0.5M Tris pH6.8, 50µl 10% SDS, 25µl 10% APS, 5µl TEMED. Protein samples were mixed with 1x sample buffer and electrophoresed in 1x electrophoresis buffer. Gels that were not blotted were stained with Coomassie blue for at least 15 minutes at room temperature. Gels were destained with 10% acetic acid.

2.4.2 Subcellular fractionation

Cells were harvested and the pellet resuspended in 1ml ice cold buffer A and incubated on ice for 15 minutes. The resuspended cells were then sonicated (Branson Sonifier 250) by placing the probe directly into the cell mixture and pulsed for 5 seconds. Unlysed cells and nuclei were collected by centrifugation in a bench top

centrifuge (Sorvall® RT6000) for 5 minutes at 1,500rpm. The supernatant was transferred into 6ml Quick Seal® tubes, topped up with liquid paraffin and spun at 28,000 rpm in a 50.4 VTi rotor in an ultracentrifuge (Beckmann L7-55) at 4°C for 10 minutes. The supernatant was removed and designated the cytosolic fraction. The pellet was washed once with buffer A and then resuspended in 500µl of buffer B and transferred to a 1.5ml eppendorf tube. The resuspended pellet was incubated on ice for 30 minutes with intermittent mixing. The tube was then centrifuged at 15,000g for 30 minutes at 4°C. The supernatant was removed and designated the detergent soluble membrane fraction. The pellet was dissolved by boiling in buffer C and represents the detergent-resistant cytoskeletal fraction.

A simpler protein extraction method was also performed which yielded isolation of a whole cell preparation and the cytoskeletal fraction. The procedure was carried out as follows. Cells were harvested and resuspended in 1ml of ice cold buffer B and allowed to lyse on ice for 30 minutes. The cells were then spun at 28,000rpm in an ultracentrifuge, as described above, for 30 minutes at 4°C. The supernatant was designated the detergent-soluble fraction. The pellet was boiled in 250µl buffer C and was designated the detergent resistant cytoskeletal fraction.

2.4.3 Protein concentration estimation

Protein concentration was assessed by using the Markwell-Lowry assay (Markwell *et al.*, 1978). Protein standards were diluted from a stock of BSA (50mg/ml) with buffer C to yield the following concentrations: 10mg/ml, 5mg/ml, 2.5mg/ml, 1.25mg/ml and 0.625mg/ml. Lowry solution and 4% copper sulphate were mixed together at a ratio of 100:1. To 1ml of this solution 10µl of each protein standard/sample was added and incubated at 37°C for 10 minutes. Then, 100µl of Folin reagent was added, consisting of a 1:1 ratio of Folin & Ciocalteu's phenol: H₂O. The samples were mixed immediately and incubated at 37°C for 15 minutes. All samples were analysed in a spectrophotometer at a wavelength of 660nm. The concentration of the samples was

determined by first calculating the relationship between the concentration of the BSA standards and their absorbance values. This was achieved by dividing the concentration of each standard by their absorbance and then the mean of these values was determined. This mean value was then used to estimate the concentration of each sample by multiplying their absorbance values by the mean.

2.4.4 Western blotting

Protein samples in 1x sample buffer and a biotin-labelled molecular weight marker were denatured by boiling for 5 minutes, cooled on ice and separated on either 10 or 15% SDS-PAGE gels. Proteins were transferred onto polyvinylidene difluoride membrane (Biotrace®) using transfer buffer in a semi-dry transfer apparatus according to the manufacturer's instructions (Atto). Sufficient transfer was assessed by staining the membrane with Ponceau S solution and then destaining with dH₂O. The membrane was then washed in 1x PBS for 5 minutes at room temperature. The Phototope® -Star western blot detection kit was used for antibody incubations with minor modifications. In brief, the blot was incubated in blocking buffer for 1 hour at room temperature or overnight at 4°C. The primary antibody was diluted in 10ml blocking buffer (keratin 5/8 1:200, keratin 8 1:400, amyloid A 1:200) and incubated on the blot for 1 hour at room temperature with gentle agitation. The blot was washed 3x 5 minutes with blocking buffer and then incubated with the secondary antibody (1:1000 dilution) and the biotin antibody (1:1000 dilution) for the molecular weight marker for 45 minutes at room temperature. The membrane was washed 3x 5 minutes with blocking buffer and then twice with 1x wash buffer. The blot was then incubated with CDP-Star™ (1:100-1:500 dilution) in 3ml assay buffer with gentle agitation for 5 minutes at room temperature. The membrane was drained of excess developing solution and wrapped in 'Clingfilm'. The blot was then exposed to x-ray film, initially for 1 minute, to assess the appropriate exposure time.

2.4.5 Immunohistochemistry

Preliminary experiments to optimise the monoclonal antibodies to human keratin 8 (kind gift of Dr. David Easty, UCD), keratin 5/8 (Neomarkers) and amyloid A (Dako) for immunohistochemistry were performed by Tony O'Grady, Department of Pathology, RCSI. The keratin 8 antibody proved to be unsuitable for immunohistochemistry on paraffin embedded material, even when used neat and was not used further.

Sectioning of paraffin-embedded human colon tumours including 20 Duke's B, 20 Duke's C and 20 metastatic lymph nodes from the Duke's C patients, was carried out by Tony O'Grady, Department of Pathology, RCSI.

Immunohistochemistry was performed on the 60 sections mentioned above using the antibodies to keratin 5/8 and amyloid A including negative controls. The Dako Envision™+ system HRP (DAB) anti-mouse antibody kit was used following the manufacturer's instructions with minor modifications. In brief, the sections were dewaxed by immersion in xylene 2x 15 minutes, cleared in alcohol 2x 10 minutes and then washed in running tap water. Endogenous peroxidase activity in the sections was blocked by incubation with 0.9% hydrogen peroxide in dH₂O (freshly prepared) for 10 minutes and then washed in running tap water for 5 minutes.

An antigen retrieval step was carried out on sections for the keratin 5/8 antibody only, as conditions for amyloid A antibody were optimal without any form of antigen retrieval. Slides for keratin 5/8 were immersed in 0.01M tri-Sodium citrate pH6.0 and microwaved at 850w for 20 minutes. The slides were cooled immediately by immersion in cold water.

All slides were washed with TBS which was subsequently removed and replaced with goat serum diluted 1:10 with 0.05M Tris buffer pH7.6. The goat serum was allowed to incubate for at least 10 minutes. The primary antibodies were prepared by diluting keratin 5/8 1:30 and amyloid A 1:500 with 0.05M Tris buffer pH7.6. The primary antibodies were added to the appropriate slides, excluding the negative controls, and incubated for 30 minutes at room temperature. The slides were

then washed in two changes of TBS for 5 minutes each before addition of the secondary antibody. The secondary antibody was incubated on all slides for 30 minutes at room temperature and then washed with two changes of TBS for 5 minutes each. DAB (Diaminobenzidine) was prepared by diluting 1:10 with 0.05M Tris buffer pH7.6 plus 5 μ l 30% hydrogen peroxide per 10ml. DAB was added to the slides and incubated for 5 minutes before rinsing with dH₂O. The slides were counter stained with haemotoxylin for 15 seconds and then washed in tap water for 3 minutes. The slides were then dehydrated by 2x 10 dips of alcohol and 3x 10 dips of xylene. The slides were finally coverslipped using DPX mountant.

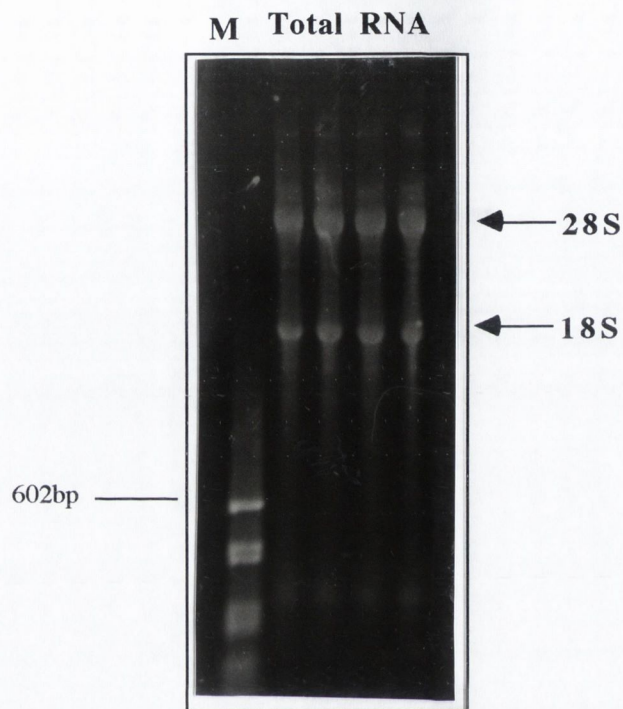


Figure 2.1 Denatured total RNA resolved on a 1 % agarose gel. The 28S:18S ratio is approximately 2:1 and the RNA is clearly undegraded and of good quality. The largest fragment of the pBS-*Hae* III (M) marker is indicated as 602bp.

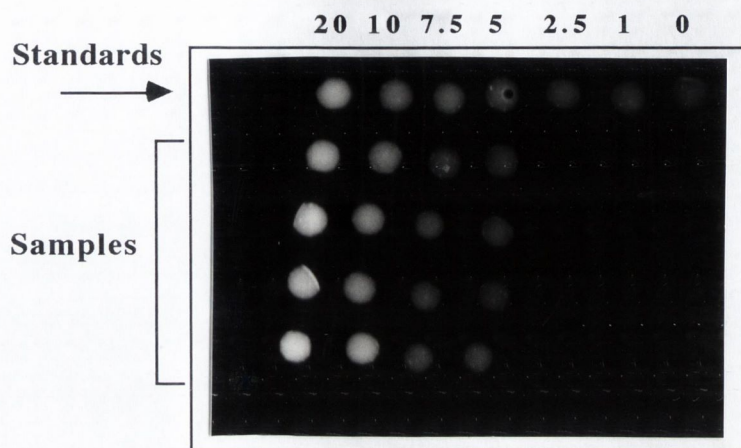


Figure 2.2 Ethidium bromide dot quantitation. DNA standards of known concentration ranging from 20 to 0 ng/ μ l were mixed with an equal volume of ethidium bromide (1 μ g/ml) and spotted onto plastic wrap overlaying a UV transilluminator (top row). Samples of unknown concentration were similarly mixed with ethidium bromide and the concentrations were estimated by comparison of the intensities of the standards to those of the samples.

Table 2.1 Oligonucleotide primer sequences

Primer name	Sequence
M13 Forward ¹	5' GTAAACGACGGCCAGT 3'
M13 Reverse ¹	5' GGAAACAGCTATGACCATG 3'
β-actin-L.U	5' *CTGTCTGGCGGCACCACCAT 3'
β-actin-L.L	5' GCAACTAAGTCATAGTCCGC 3'
β-actin-S.U	5' *ACAGGGGAGGTGATAGC 3'
β-actin-S.L	5' CGAAGGCTCATCATTCAA 3'
β-tubulin.U	5' *CGCCGAGCTGGTTGATTCTG 3'
β-tubulin.L	5' GCCCATTCCAGAGCCTGTGC 3'
Cystatin S.U	5' *ATGGCCCGGCCTCTGTGTA 3'
Cystatin S.L	5' GGCACGCTGTACCCACTCA 3'
Hox11.U	5' *CCTGCCCATTCTGCCCTGTG 3'
Hox11.L	5' CCTCCCTTGCCCTTCTCTGA 3'
DDP.U (IPP-2)	5' *AGACATCTTAGCCAGGAAA 3'
DDP.L (IPP-2)	5' TTCTTCAGGTGAGAGGTCA 3'
DDT.U (IPP-2)	5' *GTAATTGATAGGGTATGAACT 3'
DDT.L (IPP-2)	5' AGACAAGGTAACACCAAATAG 3'
hnRNP.C.U	5' *GGGGCAGTTCAACACAAATG 3'
hnRNP.C.L	5' CCAAACATCAGGCTGTTTAC 3'
hnRNP.L.U	5' *AACCCCAACAAACGCCAGAG 3'
hnRNP.L.L	5' ACTGGGGGCCGTAGCGACTT 3'
Keratin5.U	5' *GCAGCCGGAGCCTCTACAA 3'
Keratin5.L	5' ACCGCCGAAACCAAATCCA 3'
Myosin.U	5' *TCCCAGTCTCCCCAGAGTCC 3'
Myosin.L	5' ACGGCAAACATCATCCAAGA 3'
SAA.U	5' *AAGGGGACCTGGGGGTGTCT 3'
SAA.L	5' TGGGGTCTTTGCCACTCCTG 3'
SF2p32.U	5' *CGAGGGGTGGACAACACTTT 3'
SF2p32.L	5' TGGCTTTCAGCATCTGTCTG 3'
SSRP1.U	5' *TCCCCACCCATATTTTGGTA 3'
SSRP1.L	5' CTGCCTCCCACCCTATCTCT 3'
S28.U	5' *CACCGACGAGACCAGAAGTG 3'
S28.L	5' TGGCAGAATTTGGCTGTTTG 3'
GAPU	5' *ACTTTACCACAAATACAGTGT 3'
GAPL	5' AATTAGCAATCTGTAAATCAG 3'

¹ M13 Forward and Reverse primers were obtained Cy5 labelled and unlabelled.

* denotes Cy5 label.

Note: Positions of all primer sets are indicated in the relevant sections in the text.

Table 2.2 Optimised conditions for semi-quantitative RT-PCR

Gene transcript	Product Size	MgCl ₂	Annealing temperature
β-actin-L ¹	254bp	0.50mM	60°C
β-actin-S ²	109bp	1.50mM	50°C
β-tubulin	118bp	1.00mM	58°C
Cystatin S	141bp	1.50mM	57°C
Hox11	309bp	1.25mM	60°C
IPP-2 (DDP)	109bp	0.75mM	50°C
IPP-2 (DDT)	168bp	0.75mM	50°C
hnRNP C	151bp	1.50mM	50°C
hnRNP L	193bp	1.50mM	60°C
Keratin 5	146bp	1.50mM	57°C
Myosin Light Chain	122bp	1.50mM	53°C
SAA	140bp	1.50mM	57°C
SF2p32	136bp	1.50mM	53°C
SSRP1	179bp	1.50mM	54°C
Homologue of yeast S28	151bp	1.50mM	53°C

Two different primer pairs were used to amplify β-actin. ¹ β-actin-L refers to the set that amplifies a PCR product of 254bp (positions 936 and 1189 of X00351). ² β-actin-S refers to the set that amplifies a PCR product of 109bp (positions 1519 and 1627 of X00351).

CHAPTER 3

Serial Analysis of Gene Expression

3.0 INTRODUCTION

Serial Analysis of Gene Expression or SAGE is a gene expression profiling method that was first described by Velculescu *et al.* (1995). The SAGE technique was published simultaneously with another gene expression profiling method, the cDNA microarray (Schena *et al.*, 1995). The publication of both SAGE and the cDNA microarray techniques introduced, for the first time, the possibility of assessing the activity patterns of thousands of genes at once in a relatively short period of time (Nowak, 1995). Most methods prior to this were based on subtractive hybridization techniques (St. John & Davis, 1979; Timberlake, 1980; Zimmerman *et al.*, 1980; Sargent & Dawid, 1983; Hedrick *et al.*, 1984; Duguid *et al.*, 1989; Wieland *et al.*, 1990; Hara *et al.*, 1991; Hubank & Schatz, 1994) or Differential Display Reverse Transcription-PCR (DDRT) (Liang & Pardee, 1992). These techniques allowed the identification of differences in expression between two samples of interest, but, unlike SAGE or cDNA microarray, did not allow a detailed analysis of thousands of transcripts at once. In addition, the original subtractive hybridization techniques were rather labour intensive, required a lot of starting material and rarely detected low abundance transcripts (St. John & Davis, 1979; Timberlake, 1980; Zimmerman *et al.*, 1980; Sargent & Dawid, 1983; Hedrick *et al.*, 1984). However, the subsequent incorporation of a PCR step in subtraction-based techniques greatly improved these methods (Duguid *et al.*, 1989; Wieland *et al.*, 1990; Hara *et al.*, 1991; Hubank & Schatz, 1994). The DDRT-PCR methodology requires as little as 1µg of total RNA and can detect low abundance transcripts but is prone to a high level of false positives if not carefully controlled (Liang & Pardee, 1995). DDRT-PCR is described in more detail in Chapter 5.

The SAGE method is based on the isolation of short sequence 'tags', usually 9 to 11 bp in length, from a defined position within a given transcript, which contain enough sequence information to uniquely identify it (Velculescu *et al.*, 1995). SAGE yields both qualitative and quantitative information on the gene expression pattern of a given cell type. The various steps in the SAGE protocol are shown in Figure 3.1.

The initial step of the protocol involves isolation of intact total RNA from the tissues or cells of interest. The mRNA, which represents expressed gene transcripts, is usually polyadenylated at the 3' end [referred to as Poly(A)+] during processing of the primary transcript (Wahle & Keller, 1992) and therefore can be purified from ribosomal RNA and transfer RNA by means of an oligo d(T) cellulose column. Purified mRNA is then converted into double stranded cDNA using a biotinylated oligo d(T) primer. The synthesised cDNA is therefore biotinylated which facilitates further purification steps. In order for the majority of transcripts to be 'sampled', a restriction enzyme is required that will cleave most transcripts at least once. Restriction enzymes that have a 4-bp recognition site usually cleave every 256bp on average, and most transcripts are much larger than this (Velculescu *et al.*, 1995). The most popular 4-bp 'cutter' used in SAGE to date is *Nla* III and is referred to as the 'anchoring' enzyme (Velculescu *et al.*, 1995; Hermeking *et al.*, 1997; Madden *et al.*, 1997; Polyak *et al.*, 1997; Velculescu *et al.*, 1997c; Zhang *et al.*, 1997; Hibi *et al.*, 1998; de Waard *et al.*, 1999). The anchoring enzyme digests the biotinylated cDNA and fragments that have been cut closest to the 3' end of the cDNA are purified from the rest of the cleaved fragments by incubation with streptavidin coated magnetic beads. Streptavidin has a very high affinity for biotin (Stryer, 1988) and therefore the biotinylated cDNA fragments readily bind to the magnetic beads, which can then be easily removed from the remaining cDNA mix by capturing the beads with a magnet. The purified bound fragments are then split into two pools and each ligated with two types of linkers (referred to in Figure 3.1 as A and B). The linkers are designed to include a number of important features:

- An 'overhanging' anchoring enzyme site, usually *Nla* III, for ligation with the *Nla* III digested cDNAs. This reconstitutes the *Nla* III site and is necessary for a later step.
- A Type IIS restriction enzyme site (tagging enzyme). The most common Type IIS restriction enzyme site used in SAGE is *BsmF* I. Type IIS restriction enzymes are characterised by cleaving up to 20 bases away from their recognition site. *BsmF* I

ditags may be amplified preferentially during the initial cycles of PCR and artificially increase their abundance in the final analysis. This problem is circumvented in the SAGE analysis by counting any one ditag combination only once. This is because the probability of any two tags being coupled in the same ditag more than once is considered small, even for abundant transcripts (Velculescu *et al.*, 1995). Thus, if a ditag appears more than once it may potentially have been produced by PCR bias and is excluded from the final analysis.

In summary, SAGE isolates short sequence tags, which represent specific mRNA transcripts and are analogous to a transcript 'barcode'. The tags are produced in a way that allows large numbers of transcripts to be analysed in a relatively small number of sequencing runs. In addition, the number of tags identified for a given gene product represents the actual abundance of that particular transcript in the tissue or cell type of interest.

The main aim of the project was to identify differentially expressed genes that may be involved in colon cancer metastasis. The isogenic cell lines SW480 and SW620 provide a good model system for the pre-metastatic and metastatic stages of colon cancer. Briefly, SW480 is derived from a primary colon tumour and SW620 is derived from a lymph node metastasis from the same patient (Leibovitz *et al.*, 1976). A more detailed description of these cell lines is given in Chapter 1.

At the initial stages of the project the methods available to identify differentially expressed genes included subtractive hybridization based methods (Diguid *et al.*, 1989; Wieland *et al.*, 1990; Hara *et al.*, 1991) such as cDNA Representational Difference Analysis (RDA) (Hubank & Schatz, 1994), DDRT-PCR (Liang & Pardee, 1992), cDNA microarrays (Schena *et al.*, 1995) or SAGE (Veculescu *et al.*, 1995). SAGE was the method of choice for a number of reasons. Both the subtractive hybridization and DDRT-PCR methods only identify differences between two cell types of interest and the transcripts that are isolated are essentially random. Moreover, DDRT-PCR had previously been applied to the same two cell lines in our laboratory by Dr. Peter McWilliam (see Chapter 5). Initially, the use of

cDNA microarrays was restricted to laboratories having specialised equipment such as a custom-built arraying instrument and a laser fluorescent scanner (Schena *et al.*, 1995). In addition, cDNA microarrays lack the possibility of identifying unknown transcripts as the arrays are based on cloned cDNAs already in the database. SAGE, on the other hand, utilises basic molecular biology techniques and could be applied in most laboratories. SAGE also has the potential to identify unknown transcripts and, depending on the number of tags sequenced, can generate a partial or complete gene expression profile for a cell type of interest. Comparison of two or more gene expression profiles from cells of different types allows one to identify differentially expressed transcripts between them. In addition, SAGE has the added advantage of allowing genes expressed at the same level to be identified, which may prove to be as interesting as identifying gene expression differences.

The SAGE protocol and its associated software were received under licence from Johns Hopkins University (Velculescu *et al.*, 1997a & 1997b). The disclosure of details is not permitted under the terms of the materials transfer agreement concluded between the Royal College of Surgeons in Ireland and Johns Hopkins University. Thus, the details of the SAGE protocol described here are strictly confidential.

3.1 MATERIALS

General materials are described in Chapter 2, the materials described here are specific to the SAGE protocol.

3.1.1 Reagents

Competent Cells

Epicurian® Coli XL1-Blue MRF Kan supercompetent cells :Stratagene, USA.

Top10 One Shot™ competent cells: Invitrogen, NL.

Enzymes and reaction buffers

Nla III (10u/μl), *BsmF* I(2u/μl), *Sph* I(10u/μl), T4 Polynucleotide Kinase (10u/μl), T4 DNA Polymerase (3u/μl): NEB (New England Biolabs), UK.

T4 DNA Ligase high concentration (5u/μl); T4 DNA Ligase (1u/μl): GibcoBRL, UK.

AmpliTaq (5u/μl): Perkin-Elmer, UK.

Molecular Biology Kits

MessageMaker™ Reagent assembly kit, cDNA synthesis system : GibcoBRL, UK.

PolyATtract® System 1000, pGEM5 : Promega, USA.

Zero Background™ PLUS Cloning kit, Zeocin, DNA Dipstick™ Kit : Invitrogen, NL.

Oligonucleotides

Modified 3' C7 or dideoxy linkers, HPLC purified biotinylated oligo-(dT)₁₈, Primers 1 and 2: Genosys, UK.

M13 Forward and Reverse primers: GibcoBRL, UK.

(Table 3.1)

Miscellaneous

Dynabeads[®] M-280 Streptavidin : Dynal, UK.

Hotstart PCR reaction tubes : Molecular *Bio*Products[™] Inc., USA.

Glass microfibre filters GF/C 25mm: Whatman[®], UK.

3.1.2 Stock Solutions

LoTE buffer: 3mM Tris-HCl (pH7.5), 0.2mM EDTA (pH7.5).

Binding & Washing buffer, 10x: 10mM Tris-HCl (pH7.5), 1mM EDTA, 2M NaCl.

SAGE PCR buffer, 10x: 166mM (NH₄)SO₄, 670mM Tris (pH8.8), 67mM MgCl₂, 100mM β-mercaptoethanol.

SAGE loading dye: 0.25% bromophenol blue, 40% sucrose.

SOB media: 2% tryptone, 0.5% yeast extract, 0.05% NaCl, 2.5mM KCl, 10mM MgCl₂.

3.2 METHOD

The SAGE method (Velculescu *et al.*, 1997a & 1997b) may be divided into a number of steps:

3.2.1 RNA preparation

The starting amount of Poly(A)+ mRNA recommended for SAGE was between 5 and 10 μ g (V. Velculescu & J. Powell, pers. commun.). Yields from the PolyATtract® System 1000 kit (Promega), which purifies Poly(A)+ mRNA directly from the cells, were well below 5 μ g. However, sufficient amounts of Poly(A)+ mRNA were obtained by isolating total RNA first and then subsequently selecting Poly(A)+ mRNA. Total RNA was isolated from the cell lines SW480 and SW620 as described in Chapter 2 Section 2.3.2. Poly(A)+ mRNA was purified from total RNA using the MessageMaker™ Reagent assembly kit according to the manufacturer's protocol. Double selection through the oligo d(T) cellulose column was necessary to remove visible amounts of ribosomal RNA (data not shown). The integrity of the Poly(A)+ mRNA was checked by electrophoresis on 1% denaturing agarose gels (Figure 3.2a) and quantified using a DNA Dipstick™ Kit according to the manufacturer's instructions. A total RNA quantity of 1-2mg yielded 5 to 10 μ g of Poly(A)+ mRNA. Assuming that 5% of total RNA consists of mRNA this suggests a yield of approximately 10%.

3.2.2 cDNA synthesis

Reverse transcription and double strand cDNA synthesis of purified Poly(A)+ mRNA (5-10 μ g) was carried out using the cDNA synthesis system (GibcoBRL). Reactions were assembled according to the manufacturer's instructions with substitution of the supplied oligo-(dT)₁₈ primer with an HPLC-purified biotinylated oligo-(dT)₁₈ primer (Genosys). The use of a purified biotinylated oligo-(dT)₁₈ primer ensures that all of the primer is biotinylated and eliminates any loss of cDNA product in a later step due to the presence of unbiotinylated cDNA. A radioactive tracer

reaction was carried out with the first strand synthesis reactions and consisted of 1 μ Ci of [α -³²P] dCTP as recommended by the manufacturer's protocol. A radioisotope tracer (10 μ Ci) was also included in the second strand reaction. The radioisotope allowed the yield of cDNA to be assessed. Yield was calculated by measuring the amount of incorporated radioisotope according to the manufacturer's instructions. In brief, aliquots of the first strand tracer and the second strand reactions were spotted onto two glass fibre filters. One filter was dried and the second filter was precipitated with trichloroacetic acid and washed with 95% ethanol. The radioactive counts per minute (cpm) were determined in a liquid scintillation counter (Wallac-1214 rackbeta). The specific activity of the radioisotope was calculated from the cpm of the unprecipitated filter and used to calculate the yield of first or second strand synthesis determined from the acid-precipitable counts of the washed filter. Yields of at least 2 μ g of double stranded cDNA (in a volume of 20 μ l) were used for further steps of SAGE. Aliquots (12,000cpm per lane) of the tracer first strand and the second strand synthesis reactions were electrophoresed on 1.4 % denaturing alkaline agarose gels to assess the quality of the cDNA. Gels were then dried and exposed overnight (room temperature) to x-ray film (Figure 3.2b).

3.2.3 Cleavage of biotinylated cDNA with anchoring enzyme

The double stranded cDNAs were then cleaved with the anchoring enzyme *Nla* III. The digestion reactions were assembled in a final volume of 200 μ l containing 1x Buffer 4, 1x BSA and 50 units of *Nla* III (NEB). Digestions were incubated for 1 hour at 37°C. After digestion, the cDNA was phenol extracted, ethanol precipitated (as outlined in Chapter 2), washed twice with 70% ethanol, air dried and resuspended in 20 μ l LoTE. Note: *Nla* III is an unstable enzyme and is normally stored at -70°C. Therefore, fresh stocks of *Nla* III were used when possible and stocks stored for longer than 3 to 4 months were discarded.

3.2.4 Binding to streptavidin-coated magnetic beads

Two aliquots of Dynabeads (Dynal) were prepared by placing 100µl of the Dynabead slurry into each of two 1.5ml microcentrifuge tubes and removing the supernatant by capturing the beads with a magnet. The beads were washed once with 200µl 1x Binding & Washing buffer. The beads were then captured with the magnet and 100µl 1x Binding & Washing buffer, 90µl dH₂O and 10µl cDNA restriction products (half of the total in each tube) were added to the beads and allowed to incubate at room temperature for 15 minutes. The tubes were mixed intermittently. After incubation, the beads were washed, as above, with 200µl 1x Binding & Washing buffer and once with 200µl LoTE.

3.2.5 Ligating linkers to bound cDNA

The cDNA fragments bound to the beads had an *Nla* III overhang at the 5' end. This allowed ligation of the linkers to the bound fragments as they contained compatible *Nla* III ends. They also possess a *BsmF* I site and a priming site for PCR amplification. Some linkers are better PCR templates than others (Table 3.1) when amplified in a later step (Section 3.2.9). The linkers were phosphorylated and annealed to their complementary linker before ligation.

The phosphorylation reaction was carried out on linkers A2 and B2 (Table 3.1) as follows: 9µl of each linker (200ng/µl) was added to a 20µl reaction mix containing 1x kinase buffer, 2mM ATP and 10 units of T4 Polynucleotide Kinase (NEB) and incubated at 37°C for 30 minutes. The reaction was terminated by heating at 65°C for 10 minutes. The kinased linkers were then mixed with 9µl of their appropriate complementary linker (linkers A1 or B1), heated to 95°C for 2 minutes and allowed to reanneal by cooling to room temperature for 20 minutes. The reannealed linkers were tested for kinasing efficiency by performing a self-ligation test. The ligation reaction consisted of 1µl of linker A or B, 1x ligase buffer, 1.25 units T4 DNA Ligase and incubated at 16°C for 4 hours. Only kinased linkers are capable of self-ligating, forming linker-linker dimers of between 80-100bp in size.

The self-ligated linkers were compared to unligated linkers following electrophoresis on a 2.5% agarose gel stained with ethidium bromide and the majority of linkers clearly were ligated (Figure 3.3). The final concentration of linkers after phosphorylating and annealing was 125ng/μl estimated by ethidium bromide dot quantitation as outlined in Chapter 2 Section 2.3.3. This was less than the recommended concentration of 200ng/μl but helped to reduce the problem of linker contamination described later in Section 3.2.10.

Each linker (A or B) was ligated to one of the two cDNA bound beads by mixing the beads with 10μl of linker A or B (125ng/μl), 8μl 5x ligase buffer and 20μl LoTE. The tubes were gently mixed and heated at 50°C for 2 minutes and then allowed to sit at room temperature for 15 minutes. The ligation reaction was initiated by adding 2μl T4 DNA Ligase (5u/μl) (GibcoBRL) and allowed to proceed for 2 hours at 16°C. Unligated linkers were removed by washing the beads 4 times with 200μl 1x Binding & Washing buffer and then twice with 1x Buffer 4 (NEB).

3.2.6 Release of cDNA tags by digestion with tagging enzyme

The tags and linkers were released from the magnetic beads by digestion with the tagging enzyme, *BsmF* I. The beads were digested in a reaction mix containing 1x buffer 4, 2x BSA and 4 units *BsmF* I (NEB). Incubation was for 1 hour at 65°C with intermittent mixing. Since *BsmF* I cleaves 14bp away from its recognition site (leaving a 4 base 5' overhang), and the *Nla* III site overlaps the *BsmF* I site by 1bp, a 15bp SAGE tag was released after digestion with *BsmF* I. Thus, the released construct consisted of the 40bp linker (A or B) and a 15bp tag.

After digestion, the beads were immobilised with the magnet and the supernatant collected. The supernatant was phenol extracted, ethanol precipitated, washed twice with 70% ethanol, air dried and resuspended in 10μl LoTE.

3.2.7 Blunt ending of the released cDNA tags

SAGE tag overhangs were blunt ended by incubating with 3 units of T4 DNA Polymerase¹ (NEB) in a reaction mix containing 1x T4 Polymerase buffer, 2x BSA (NEB), 100μM dNTPs (Promega) at 11°C for 15 minutes. The sample volumes were then increased to 200μl with LoTE, phenol extracted and precipitated as described above. The cDNA pellets were resuspended in 6μl LoTE.

3.2.8 Ligating tags to form ditags

The linker-tags from each pool (A and B) were combined and ligated to each other in a reaction volume of 6μl containing 2μl of each linker-tag pool, 1x ligase buffer and 4 units of T4 DNA Ligase (high concentration) (GibcoBRL) and incubated overnight at 16°C. The volume was increased to 200μl with LoTE, phenol extracted and precipitated as described above. The pellet was resuspended in 10μl of LoTE buffer.

3.2.9 PCR amplification of ditags

The ditag ligation was diluted between 1/10 and 1/200 to determine the optimum dilution. PCR reactions were performed in 1x SAGE PCR buffer, with 6%DMSO, 1.5mM dNTPs, 700ng of each primer (1 and 2) per 100μl (Table 3.1) and 10 units of AmpliTaq™ (Perkin-Elmer). All PCR reactions were 'hotstarted' using Hot Start tubes (Molecular BioProducts) and amplified for 25 cycles of 30 seconds at 95°C, 1 minute at 55°C, 1 minute at 70°C and 1 cycle of 70°C for 5 minutes. These conditions were optimal for a Hybaid™ PCR machine. The ligated ditags with linker amplify as a band of 102bp. Figure 3.4 shows the amplified 102bp product plus a number of expected background bands. The background bands include: linker ligated to linker, i.e. no ditag in between : 80bp, linker-tag only (unligated) : approximately 60bp and linker only : 40bp. Once the ligation dilution was optimised,

¹ Although T4 DNA Polymerase was successfully used here, a later version of the SAGE protocol (Velculescu *et al.*, 1997b) recommends use of the Klenow fragment (large fragment of DNA Polymerase I), which is a more processive polymerase than T4 DNA Polymerase.

PCR amplifications were performed in bulk (usually 50 x 100µl) using the optimised ligation dilution as template. Bulk PCR products were pooled into a suitable number of tubes and then phenol extracted, ethanol precipitated, washed with 70% ethanol and resuspended in a total volume of approximately 200µl of LoTE.

3.2.10a Purification of the 102bp product

The presence of background bands, which were essentially various forms of linker-linker contamination, necessitated gel purification of the 102bp product. The PCR products were mixed with one fifth volume of SAGE loading dye and separated on 3 x 10 well, 12% polyacrylamide gels electrophoresed at approximately 130 volts and stained in an ethidium bromide bath (1.5µg/ml). The 102bp bands were gel purified using Spin-X centrifuge tubes as outlined in Chapter 2 Section 2.3.8, except that three excised bands were pooled together and 300µl of LoTE was added. The tubes were heated at 65°C for 20 minutes and then rotated at 4°C overnight². The gel was removed using one Spin-X tube per 300µl. The original procedure (Velculescu *et al.*, 1995, 1997a) omits gel purification at this stage, as released ditags were gel purified at a later step (see below). However, in a later publication (Velculescu *et al.*, 1997b, 1997c) this extra step was included. This was to reduce the problem of linker contamination during subsequent concatamerisation and cloning. The ditags were released from the linkers by digestion with *Nla* III. This meant that both the ditags and the linkers had *Nla* III compatible overhangs. This was necessary for ditags to ligate to each other to form concatamers in a later step (see Section 3.2.11). However, the presence of contaminating linkers would have been problematic by ligating to the ditags and subsequently interfering with concatamerisation. Once a linker binds to one or both ends of a concatamer, this inhibits its extension by preventing any further ligation of ditags. Moreover, this also renders the concatamer (whatever its length) unclonable (J. Powell and V. Velculescu, personal communications).

² Rotation overnight at 4°C did not significantly increase the yield of purified PCR product compared to processing straight away.

After gel purification of the 102bp product the yield was quantified by ethidium bromide dot quantitation as described in Chapter 2 Section 2.3.3. A yield of approximately 15µg was required for subsequent steps, therefore if the yield was below this amount, PCR reamplification was carried out. This was a modification (J. Powell, personal communication) to both versions of the detailed protocol (Velculescu *et al.*, 1997a & 1997b). This was preceded by a cycle curve to assess the optimum number of PCR cycles required. This was necessary as the purified 102bp product was susceptible to 'over-amplification'. Figure 3.5 shows how increasing the cycle number beyond the optimum results in loss of the 102bp band. This was possibly caused by its accumulation as a high molecular weight product due to aggregation. Once the optimum number of cycles was determined, bulk PCR reactions were carried out as above. The PCR products were again pooled, a small aliquot checked by electrophoresis on a 4% Nusieve agarose gel, and then phenol extracted, ethanol precipitated as before and resuspended in a total volume of 240µl of LoTE.

3.2.10b Release and purification of ditags

Ditags were released from the linkers by digestion with *Nla* III. The digestion reactions were usually split into three tubes each containing 80µl of PCR products, 1x buffer 4, 1x BSA and 50 units of *Nla* III (NEB). The digestions were incubated at 37°C for between 1 and 3 hours. Complete digestion of the 102bp product was never achieved with *Nla* III, even after extended incubation (up to 3 hours) and increasing the concentration of enzyme in the reaction. This resulted in reduced yields. The digests were phenol extracted, pooled into two tubes, the volume brought up to 200µl with LoTE and ethanol precipitated by adding 100µl 10M ammonium acetate, 3µl glycogen (20µg/µl) and 1.1ml 100% ethanol (stored at -20°C). The tubes were vortexed and placed in a dry ice/ethanol bath for 10 minutes. The cDNA was collected by spinning in a microfuge at top speed for 15 minutes at 4°C. The pellets were washed once with 70% ethanol and resuspended in 15µl LoTE. The digested

fragments were resolved by electrophoresis on a 12% polyacrylamide gel (2 wells) and stained with ethidium bromide as described above. Figure 3.6 shows a typical pattern achieved after digestion of the 102bp product with *Nla* III. The ditags migrate as a 26bp band and were excised from the gel and purified using Spin-X centrifuge tubes as described above, except that 300µl of LoTE was added to each excised gel band and incubated at 37°C for 20 minutes. Once the gel has been removed the purified ditags were ethanol precipitated exactly as described for precipitation of the *Nla* III digested 102bp PCR product (Section 3.2.10a). Purified ditags were resuspended in 7.5µl LoTE and the yield determined by ethidium bromide dot quantitation. Yields usually ranged from 300 to 900ng.

An alternative method to ensure a purer ditag sample is to use biotinylated primers for PCR amplification of the 102bp product. This results in PCR products that are biotinylated at the linker ends. The 102bp product is not purified from the background bands at this stage but is *Nla* III digested as described in the original protocol (Velculescu *et al.*, 1997a). The 26bp ditags are gel purified and then bound to streptavidin beads and incubated as described in Section 3.2.4 and any remaining contaminating linkers (which are biotinylated) will bind to the beads. This results in a purer ditag sample in the supernatant (Powell, 1998).

An attempt was made to adapt this step in combination with gel purification of the 102bp product.³ Therefore, biotinylated primer 1 and 2 would have been used in step 3.2.9. The 102bp product was purified and then digested with *Nla* III. The ditags were gel purified and then bound to the beads as described above. However, the extra purification step resulted in a substantial loss of 26bp product as it also included an additional precipitation step. Therefore, gel purification only (as outlined above) was used for the generation of the SAGE libraries in this study.

³ The low efficiencies of *Nla* III digestion meant that attempting a digestion without purifying the 102bp band first (as described in the previous paragraph) was not desirable as this was likely to reduce the efficiency of digestion even more. This is due to the presence of 'background' bands that would also be cleaved by *Nla* III.

3.2.11 Ligation of ditags to form concatamers

Purified ditags were ligated together to form concatamers. The ligation reaction consisted of 7µl of purified ditags, 2µl 5x ligation buffer and 1µl T4 DNA Ligase (5u/µl, high concentration) (GibcoBRL). The ligation time was normally between 3 and 4 hours at 16°C. The entire ligation reaction was electrophoresed on an 8% non-denaturing polyacrylamide gel and stained with ethidium bromide. Figure 3.7 shows an example of a concatamer smear. Concatamers sized between 400 and 2000bp were excised and purified using Spin-X centrifuge tubes as described in Chapter 2 Section 2.3.8. The concatamers were ethanol precipitated by adding 3µl glycogen (20µg/µl), 100µl 10M ammonium acetate, 1ml 100% ethanol and incubated at -20°C for several hours. The concatamers were collected by centrifugation in a microcentrifuge at full speed for 15 minutes at 4°C. The DNA pellet was washed twice with 70% ethanol, air dried and resuspended in 5µl LoTE.

3.2.12 Cloning of concatamers and sequence analysis.

Any *Sph* I cleaved vector can be used for cloning of the *Nla* III produced concatamers. Initially, pGem5 (Promega) was used in combination with Epicurian® Coli XL1-Blue MRF Kan supercompetent cells (Stratagene). This was a lacZ based blue/white selection system (Sambrook *et al.*, 1989) and cloning and transformation efficiencies were not satisfactory. Therefore, an *Sph* I cleaved pZero cloning vector (Figure 3.8) was used instead. The cut vector was gel purified using the Wizard PCR preps DNA purification system (Promega) as described in Chapter 2 Section 2.3.8. Concatamers were ligated into the prepared pZero vector (25ng) by incubating with 1x ligase buffer and 1 unit of T4 DNA Ligase overnight at 16°C. Ligations were transformed into Top10 One Shot™ competent cells (Invitrogen) and selected on low salt, Zeocin (50µg/ml) LB plates according to the manufacturer's protocol.

Transformed colonies were picked into 96 well plates containing SOB media with 5% glycerol and 50µg/ml of Zeocin antibiotic. These plates were grown overnight in a shaking incubator at 37°C and then frozen at -70°C until screened.

Clones were screened for insert size by taking 1 μ l of culture and mixing with 1x PCR buffer, 250 μ M dNTPs, 2.5mM MgCl₂ (Promega), 40pmol of M13 Forward and Reverse primers (GibcoBRL) (Table 3.1) and 0.125 units of *Taq* DNA Polymerase (Promega). The PCR reactions were heated at 95°C for 3 minutes and then amplified for 30 cycles of 1 minute at 94°C, 1 minute at 55°C, 1 minute at 72°C. The PCR products were resolved on 2% agarose gels and stained with ethidium bromide (Figure 3.9). Clones containing at least 10-15 tags (i.e. a PCR product of approximately 450bp) were plasmid prepped and cycle sequenced as outlined in Chapter 2. Figure 3.10 illustrates the sequence of a typical SAGE clone with the punctuated 'CATG' denoting the beginning of one ditag and the end of another.

3.3 RESULTS

The SAGE protocol was applied to the cell lines SW480 and SW620. A SAGE library was generated for each cell line and clones were screened and sequenced until a target of approximately 5,000 tags was reached for each library. All clones were sequenced, as described above, on an ALFexpress™ automated DNA sequencer. Sequences were saved as 'text' files once the vector and ambiguous sequences had been removed. All sequence files were then analysed using the SAGE software version 1.0 (Johns Hopkins University, MD, USA).

The SAGE software identifies the anchoring enzyme site (*Nla* III), extracts the intervening tags, and records them in a database (Velculescu *et al.*, 1997c). Cleavage with *BsmF* I should yield an 11bp tag plus the 4bp CATG (*Nla* III enzyme site). However, for the initial analysis a 9bp tag (plus CATG) was used to identify the expressed transcripts. This was due to reports of *BsmF* I occasionally cleaving closer to its recognition sequence (Velculescu *et al.*, 1995). Any ambiguity in identification of a particular transcript could be clarified by using the extra two bases (11bp in total). Ditags that occurred repeatedly were counted only once. This eliminates any PCR bias, as the probability of any two tags being coupled in the same ditag more than once is small, even for abundant transcripts (Velculescu *et al.*, 1995). Tag number was quantified for each library and tags were identified using the SAGE software. The software identifies SAGE tags by comparison to a modified Genbank database (similar to the BLAST algorithm; Altschul *et al.*, 1989) which was derived from Genbank Flatfile release 101.0 (June, 1997). The program classified the Genbank entries into class A: identified mRNA transcripts, class B: EST sequences and class C: gene sequences. SAGE tags were compared to all the Genbank entries and all tag matches were noted. However, only tags that originate from the extreme 3' end of the sequence are collected and displayed. A separate tag profile was generated for each library, providing information on tag abundance and comparisons to Genbank. Thus, gene expression profiles were produced for each library, which were then compared to identify differences between the two cell lines.

3.3.1 SW480 SAGE analysis

The SAGE library for the premetastatic cell line SW480 consisted of over 1,300 colonies. A total of 453 clones were sequenced which yielded 341 duplicate ditags. The duplicate ditags were only counted once as they may have occurred repeatedly due to PCR bias as discussed in Section 3.0. This left a total of 5,079 tags. The average insert size for this library was approximately 12 tags per clone. The tags were analysed by the SAGE software as described above. The gene expression profile generated for SW480 is detailed in Table 3.2 and includes the first 183 most abundant transcripts that were expressed at a level of 3 tags or greater. Among these transcripts approximately 21% were cDNA clones/ESTs, 19% were ribosome associated transcripts and the remaining transcripts consisted of a mixture of known gene transcripts associated with a variety of functions including the cytoskeleton, transcription, motility, signalling and protein synthesis and metabolism. A further 16% were present in Genbank but were not uniquely identified based on the 9bp tag and are referred to as 'equivocal'. Attempts were made to identify equivocal tags only if they showed a difference in expression when compared to SW620. A total of 1.2% of the tags originated from the linker sequences and, since these do not represent 'real' transcripts, they were disregarded.

3.3.2 SW620 SAGE analysis

The SAGE library for the metastatic cell line SW620 consisted of approximately 700 colonies. A total of 304 clones were sequenced, which yielded 414 duplicate ditags. The duplicate ditags were again only counted once as they may have occurred repeatedly due to PCR bias. This left a total of 5,044 tags. The average insert size for this library was approximately 18 tags per clone. The tags were analysed by the SAGE software as described above. The gene expression profile generated for SW620 is detailed in Table 3.3 and includes the first 170 most abundant transcripts that are expressed at a level of 3 tags or greater. The gene expression profile for SW620 had a similar pattern to SW480 as expected. Again, the

cDNA/EST clones and the ribosomal associated transcripts were the most abundant transcripts comprising 19 and 20% of the total respectively. A further 20% were described as equivocal as described in the last section. A total of 0.58% of the tags were derived from linker sequences and thus were disregarded.

3.3.3. Gene expression differences between SW480 and SW620

The gene expression profiles of SW480 and SW620 were compared using the SAGE software as described above. A total of 81 gene transcripts showed a difference of at least three fold between the two cell lines. These differentially expressed gene products consisted of known genes, ESTs (Expressed Sequence Tags) and unidentified transcripts (see Table 3.4). The known gene products fall into a number of different functional categories including ribosomal associated proteins, cytoskeletal proteins, motility related proteins and a range of proteins associated with transcription and transcript splicing. A full list of the differentially expressed transcripts observed is given in Table 3.5. Two housekeeping genes β -actin and Glyceraldehyde 3-phosphate dehydrogenase displayed an approximate equal tag abundance. The housekeeping genes are normally constitutively expressed (Cleveland *et al.*, 1980) and show similar patterns of expression between different cell and tissue types. Similar expression levels of such genes were expected between these isogenic cell lines.

A significant number of the differentially expressed transcripts identified have been designated as ESTs (Expressed Sequence Tags). ESTs are partial, single pass sequences from either the 5' or 3' end of a randomly selected cDNA clone. The transcript from which the EST originated may be identified by taking advantage of a number of database tools (Borsani *et al.*, 1998). These include the Human Gene Index (HGI; The Institute of Genomic Research), the EST Assembly Machine (Tigem, Italy) or ESTblast (Human Genome Mapping Project Resource Center, Cambridge, UK). The accession number of the human EST of interest is entered and a sequence contig can be constructed *in silico* from the submitted EST with overlapping

sequences from the available database. If the contig extends into coding region, the EST will then be identified. The use of these databases can allow the identification of ESTs without the usual experimental work required (Borsani *et al.*, 1998). Identification of some of the differentially expressed ESTs was attempted by this method and this is described in the next chapter (Chapter 4).

An analysis of a moderate number of SAGE tags has allowed identification of the most highly abundant gene transcripts in human colon tumour cell lines isolated from different stages of progression to metastasis. Comparison of their gene expression profiles has resulted in identification of a panel of putative differentially expressed genes that may be involved in colon tumour progression. Confirmation of some of the differentially expressed genes is described in the next chapter (Chapter 4).

3.4 DISCUSSION

3.4.1 SAGE data overview

A SAGE analysis of over 10,000 tags has allowed the expression of highly abundant transcripts to be analysed in colon tumour cell lines of different stages of metastasis. SAGE had previously been used to analyse gene expression in colon cancer (Zhang *et al.*, 1997) but that study focused on gene expression differences between normal and tumour tissues/cell lines. The SAGE analysis described here focused on a later stage of colon tumourigenesis, i.e. the changes that tumour cells undergo in order to become metastatic.

Examination of the gene expression profiles from each of the cell lines showed a similar expression level for the majority of transcripts. The most abundant transcripts for the SW480 cell line consisted of 21% cDNA clones or ESTs, a further 21% of transcripts involved in protein synthesis/turnover and between 1.5 to 5% of transcripts which were categorised as being involved in cell division, cell signalling/communication, cell structure/motility, homeostasis/defense, metabolism and RNA synthesis. A total of 19% of transcripts did not fall into any of the above categories and were therefore unclassified. Tags which could not be uniquely identified from 9 bases consisted of 16% of the total.

The gene expression profile for SW620 was very similar to SW480 with minor fluctuations. These included slight decreases such as 0.6% of transcripts from the cell division category compared to 3% for SW480 and 3% from the homeostasis/defense category as compared to 5% in SW480. The slight increases included a 2-3% increase in the number of transcripts associated with cell signalling/communication, cell structure/motility, RNA synthesis and metabolism in the SW620 expression profile compared to SW480. A total of 19% of transcripts were cDNA/ESTs, 22% were classified as being involved protein synthesis/turnover and 14% were unclassified. Tags which could not be uniquely identified from 9 bases consisted of 20% of the total.

A significant portion of the protein synthesis/turnover category consisted of ribosomal proteins. Ribosomal proteins form a large fraction of the cellular protein content. Seventy nine ribosomal proteins have been described to date, but their functions, apart from being constituents of the ribosome, are largely unknown. A number of recent studies have found that certain ribosomal proteins have functions apart from the ribosome and protein synthesis (Wool, 1996). This is especially significant when altered expression of ribosomal proteins is observed. This is discussed in Chapter 4, Section 4.2.5.5. Other major categories of highly abundant gene products included proteins involved with transcription, mRNA processing, translation, some of the major metabolic pathways such as oxidative-phosphorylation and glycolysis, signalling pathways and the cytoskeleton. Comparison of the SAGE analysis with an expression analysis based on 37 human tissues (Adams *et al.*, 1995) showed that the average proportion of each of the categories described above were relatively similar. Adams *et al.* (1995) examined human gene expression patterns by analysing over 83 million nucleotides of cDNA sequence derived by sampling 300 cDNA libraries. The analysis was based on high to moderate abundance transcripts. The average proportion of transcripts calculated for the 37 tissues were: cell signalling/cell communication, 12%; RNA synthesis and processing, 6%; protein synthesis and processing, 15%; metabolism, 16%; cell division and DNA synthesis, 4%; cell structure/motility, 8%; cell/organism defence and homeostasis, 12%; unclassified, 24%. Thus, the categories with the largest proportion of transcripts were protein synthesis and metabolism (Adams *et al.*, 1995). This SAGE analysis also identified protein synthesis as the dominant category and metabolism and cell structure/motility were the second most abundant categories in both cell lines. Housekeeping genes have been defined as genes that are essential for cell viability and therefore are expressed in all cells. In depth analysis of 30 human tissues identified a range of genes that were widely expressed (Adams *et al.*, 1995). These included a range of ribosomal proteins and gene transcripts such as Elongation Factor 1 α , β -actin, Glyceraldehyde 3-phosphate dehydrogenase and ferritin light and heavy

chains, all of which were expressed at a similar level in both tumour cell lines as determined from the SAGE analysis. The gene expression profiles for both cell lines demonstrate that a large proportion of transcripts produced are destined for roles involved in protein synthesis. Thus, a large amount of a cell's energy is devoted to this process both in tumours (Tables 3.2 and 3.3) and in a range of human tissues (Adams *et al.*, 1995). This would be expected in cells that are actively growing and dividing as large numbers of proteins are produced to carry out a multitude of functions within the cell.

Comparison of the two gene expression profiles has identified 81 transcripts showing a difference in expression level of at least three fold. This included all of the transcripts detected by SAGE which included transcripts with tag abundances of <3 and which therefore may not be listed in Table 3.2 or 3.3. The majority of these transcripts were identified as known genes or EST/cDNA clones using the SAGE software as described. A moderate SAGE analysis requires that any gene expression differences identified need verification by an independent method. Confirmation of a subset of differentially expressed genes and further analysis and discussion of the SAGE data is described in the next chapter (Chapter 4).

3.4.2 Methodological considerations

3.4.2.1 Implementation of the SAGE protocol

Successful implementation of the SAGE protocol required basic molecular biology methods, but attention to a number of details was necessary and these are discussed below. The extraction and purification of good quality, undegraded, Poly(A)+ mRNA was essential. In addition, a yield of 5 to 10 μ g was necessary to ensure that a sufficient amount of double stranded cDNA was synthesised. Moreover, use of high quality reagents, purified linkers and biotinylated oligo-(dT) primer was necessary to reduce loss of material throughout the procedure. Introduction of an extra purification step, either gel purification (Velculescu *et al.*, 1997a, 1997b, 1997c) or the biotin-streptavidin method (Powell, 1998), was important to remove potentially

contaminating linkers from the concatamerisation and cloning steps. In addition, transformation of the final ligation products required highly efficient competent cells. Batches of competent cells (Invitrogen, Stratagene) were found to vary in their transformation efficiencies of ligations even though a control plasmid (pUC19) showed comparable efficiencies (data not shown).

The biggest problem that remained unresolved was incomplete digestion of the 102bp product with the anchoring enzyme *Nla* III. This may have been due to the inherent instability of this enzyme and unreliable batches of *Nla* III have been reported (V. Velculescu and J. Powell, pers. comms.). Incomplete digestion of the 102bp product substantially reduced the yield of ditags which meant less concatamers for the final cloning step. This resulted in reduced cloning efficiency and therefore less clones per SAGE library. For an extensive SAGE analysis many thousands of clones are required, however, the libraries produced in this study were sufficient for a moderate SAGE analysis.

Sequencing of the SAGE clones was performed using the plasmid DNA directly rather than PCR product generated using M13 Forward and Reverse primers as previously described (Velculescu *et al.*, 1997c; Zhang *et al.*, 1997). This reduced the sequencing error rate due to misincorporation by *Taq* DNA Polymerase (error rate: 8.0×10^{-6} , Stratagene) during PCR amplification. The SAGE tag is only 15bp in length, thus alteration of just one base may prevent its identification.

In summary, the SAGE protocol was applied soon after its initial introduction (Velculescu *et al.*, 1995) and required a certain amount of optimisation. The most essential technical aspects of using SAGE have been highlighted and discussed. The SAGE method has since been improved and further optimised by a number of different laboratories and this is discussed in the next section.

3.4.2.2 Improvements to the SAGE protocol

SAGE has proved to be a very elegant technique which has allowed gene expression profiling to be carried out in a number of different systems (Velculescu *et*

al., 1995; Hermeking *et al.*, 1997; Madden *et al.*, 1997; Polyak *et al.*, 1997; Velculescu *et al.*, 1997c; Zhang *et al.*, 1997; Hibi *et al.*, 1998; de Ward *et al.*, 1999). However, SAGE has a number of disadvantages.

One of the first obvious disadvantages to SAGE is the requirement for a large amount of Poly(A)+ RNA (approximately 5µg). This restricts SAGE to cultured material or large tissue samples. However, a modified version of SAGE, known as microSAGE, has recently been described (Datson *et al.*, 1999). This method allowed SAGE to be performed on a single punch of a 300µm tissue slice, estimated to contain 1-5ng Poly(A)+ RNA. The initial steps of microSAGE are carried out on total RNA in a single tube coated with streptavidin. This substantially reduces any loss of material due to phenol extractions and precipitations. PCR amplification of the ditags is carried out in a manner similar to the original procedure (Velculescu *et al.*, 1997c) except that a reamplification step is carried out after gel purification of the 102bp product. (This is similar to the modification described in Section 3.2.10a.) The remaining steps are carried out in much the same way as the original procedure. The introduction of microSAGE is likely to broaden the applications of SAGE enormously (Datson *et al.*, 1999).

Another major disadvantage of SAGE is the large amount of sequencing involved. The greater the cloning efficiency, the larger the insert size, which means sequencing targets are obtained more rapidly and cheaply. The original procedure (Velculescu *et al.*, 1997b) produced clones with an average of 20 tags per clone (Powell, 1998; Kenzelmann & Muhlemann, 1999). Introduction of the biotin-streptavidin method for purifying ditags (as described above, Section 3.2.10b) increased the average number of tags per clone to approximately 36 (Powell, 1998). This is due to a 'purer' ditag sample, i.e. less contaminating linkers which effectively poison the concatamerisation and cloning steps. A further improvement to the cloning efficiency of SAGE has recently been described by simply including an extra heating step (Kenzelmann & Muhlemann, 1999). In the original procedure, concatamers are resolved on an 8% polyacrylamide gel for size selection (usually

>400bp) and purification. Despite this step, inserts of a smaller than selected size are usually found in a significant number of clones. This may be due to incomplete ligation of concatamers which leaves a large amount of small sized fragments. These small fragments may migrate with large concatamer molecules by aggregating (not ligating) together during concatamerisation. The modified procedure includes extending the concatamer ligation to an overnight incubation and heating the concatamer ligation to 65°C for 15 minutes. The concatamers are chilled on ice for 10 minutes and then separated by electrophoresis on an 8% polyacrylamide gel. This extra heating step produced a visible shift in banding pattern towards molecules of smaller size. The average insert size from SAGE libraries using this step was 67 tags (Kenzelmann & Muhlemann, 1999). This modification allows a high throughput of SAGE tags with minimal sequencing costs.

The SAGE protocol has been further improved since its introduction (Velculescu *et al.*, 1995) and could be applied in any laboratory. SAGE can be used at any level, but for an extensive analysis access to a high-throughput sequencing facility with technical support is desirable.

3.4.2.3 Current gene expression methodologies

Two of the major differential gene expression methods developed since SAGE include Suppression Subtractive Hybridization and further advances in cDNA microarrays.

Suppression Subtractive Hybridization (SSH) is an improved PCR based subtractive hybridization method for identifying differentially expressed gene transcripts (Diatchenko *et al.*, 1996). It shares several similarities to cDNA Representational Difference Analysis (Hubank & Schatz, 1994) but does not require multiple rounds of subtraction. SSH enriches for low abundance transcripts by including a 'normalization' step which results in approximately equal concentrations of high and low abundance cDNAs. The adaptors used in the technique are designed so that only subtracted, differentially expressed cDNAs are exponentially amplified in

the PCR step. This is the suppression effect, hence the name Suppression Subtractive Hybridization. The SSH protocol and reagents are available as a kit from Clontech. SSH has been implemented in the laboratory (by Dr. Peter McWilliam) to complement the SAGE analysis described here, which focused on high abundance transcripts. However, the method is limited similarly to other methods, such as DDRT-PCR (Liang & Pardee, 1992), in only identifying differences in gene expression and does not, like SAGE and cDNA microarrays, provide a complete gene expression profile.

The use of cDNA microarrays (Schena *et al.*, 1995) has led to the generation of gene expression profiles for processes such as tumour suppression (DeRisi *et al.*, 1996), inflammatory disease (Heller *et al.*, 1997), metabolic regulation in *Saccharomyces cerevisiae* (DeRisi *et al.*, 1997) and alveolar rhabdomyosarcomas (Khan *et al.*, 1998). These microarrays use a specialised robotic system to bind EST/cDNA clones or gene specific oligonucleotides onto glass slides ('chips') to generate an array. Fluorescently labelled cDNAs (each cDNA population is labelled with a different fluorophore) from each cell type of interest are hybridized to the array. Fluorescence patterns are measured with a computerised scanner which translates this information into actual levels of expression of a particular gene transcript from each cell type of interest. Most of the arrays to date have analysed the expression of over 1,000 cDNAs but arrays can be scaled up to contain over 20,000 cDNA targets (Schena *et al.*, 1995). This form of microarray technology requires specialised equipment and is not amenable to most laboratories. However, the introduction of cDNA arrays on nylon membranes by companies such as Research Genetics, Genome systems and Clontech has opened array technology to the majority of laboratories. The arrays consist of EST/cDNA clones or PCR products dotted (in duplicate) onto nylon membranes and contain a variable number of cDNAs depending on the company source (588 known genes : Clontech; 5,000 cDNAs: Research Genetics; 18,000 cDNAs : Genome Systems). Duplicate filters are compared after hybridizing with ³³P-labelled cDNAs synthesised from total RNA, Poly(A)+ RNA or

an RAP-PCR fingerprint (Trenkle *et al.*, 1998). Use of an RAP-PCR fingerprint reduces the complexity of the probe and allows rarer transcripts to be identified. Subtracted cDNAs could also be used in this manner. Several software programs are now available from these companies to allow the probed filters to be analysed. The main disadvantage with microarrays is the inability to identify novel transcripts as the arrays are based on cDNA clones already deposited in the database. SAGE has the advantage in this regard. However, this is likely to become less of a problem with the continued expansion of the EST/cDNA databases.

In the context of the current gene expression profiling methods, SAGE remains a valuable tool and can provide a complete gene expression profile for a cell type of interest, including the possibility of identifying unknown transcripts. For a high throughput approach to SAGE, access to a large sequencing facility is essential. The glass 'chip' technology is not currently available to most laboratories but is the most amenable to automation and allows the simultaneous analysis of multiple samples (Strachan *et al.*, 1997). Both SAGE and glass 'chips' are likely to be the tools for most large scale gene expression operations. However, for the average laboratory, commercial cDNA microarray filters currently provide the least labour intensive method to analyse thousands of transcripts at once and are likely to become the most popular choice for gene expression profiling within the present environment.

B. **AE**= anchoring enzyme, *Nla* III. **TE**= tagging enzyme, *Bsm* FI. The final step of the SAGE protocol yields a library of clones consisting of concatenated ditags as shown above. The '**CATG**' sequence of the anchoring enzyme indicates the beginning of one ditag and the end of another.

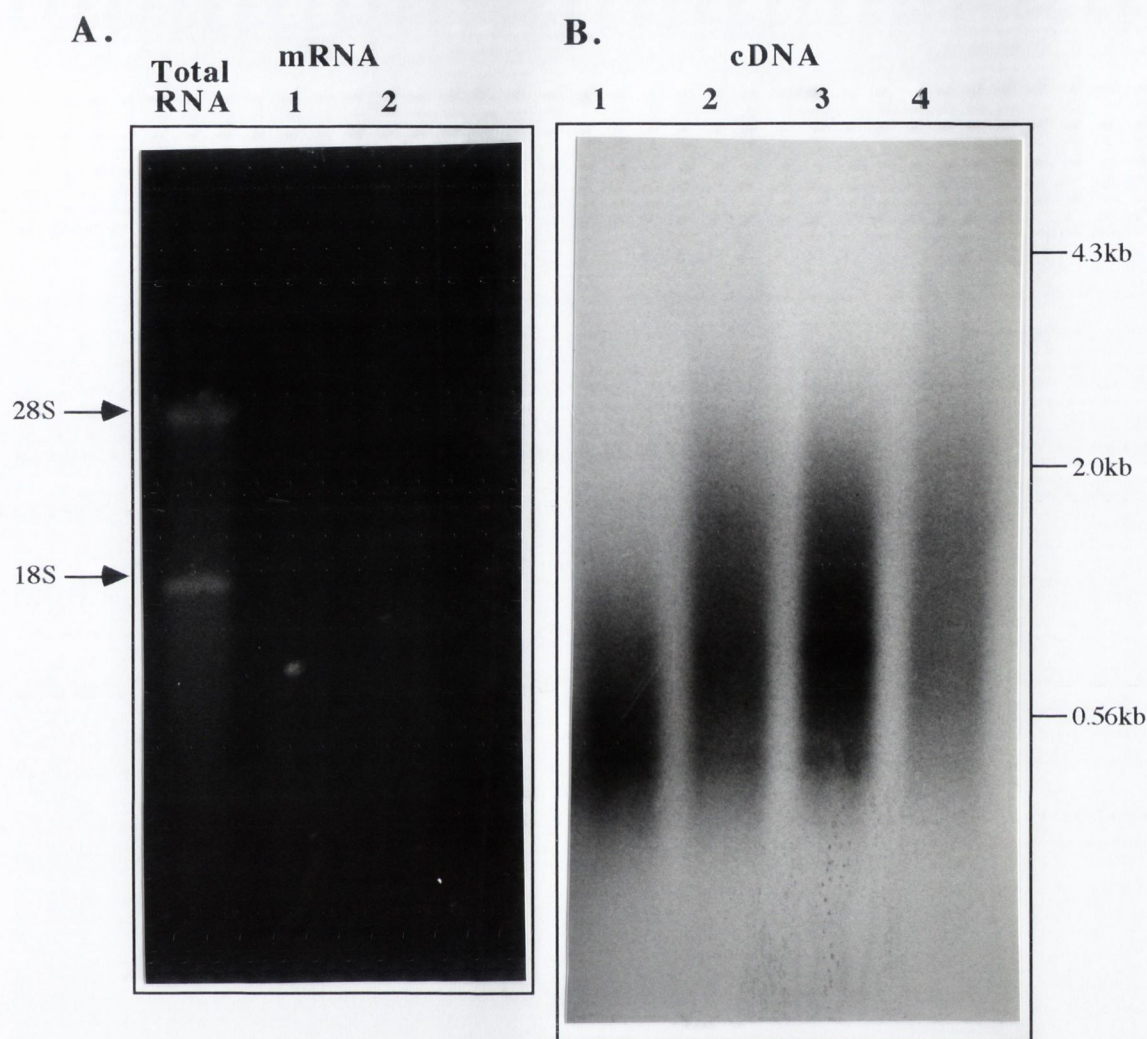


Figure 3.2 Purified mRNA and cDNA synthesis for SAGE. (A) Poly(A)+ mRNA was double selected from total RNA using the MessageMaker™ Reagent assembly kit. The integrity of purified mRNA was assessed by electrophoresis on a 1% denaturing agarose gel and stained with ethidium bromide. Total RNA was electrophoresed in parallel in order to estimate size. The RNA was visualised on a UV transilluminator and the mRNA (lanes 1 and 2) was detected as a faint high molecular weight 'smear' as expected for intact, undegraded mRNA. (B) Radiolabelled first and second strand cDNA synthesis reactions were resolved by electrophoresis on 1.4% denaturing alkaline agarose gels. The gels were dried and then exposed overnight to x-ray film. Lanes 1 & 3 show the first strand reactions from two separate samples and lanes 2 & 4 show the second strand reactions from the same two samples. The high molecular smears detected for each reaction demonstrated successful synthesis of double stranded cDNA for each sample.

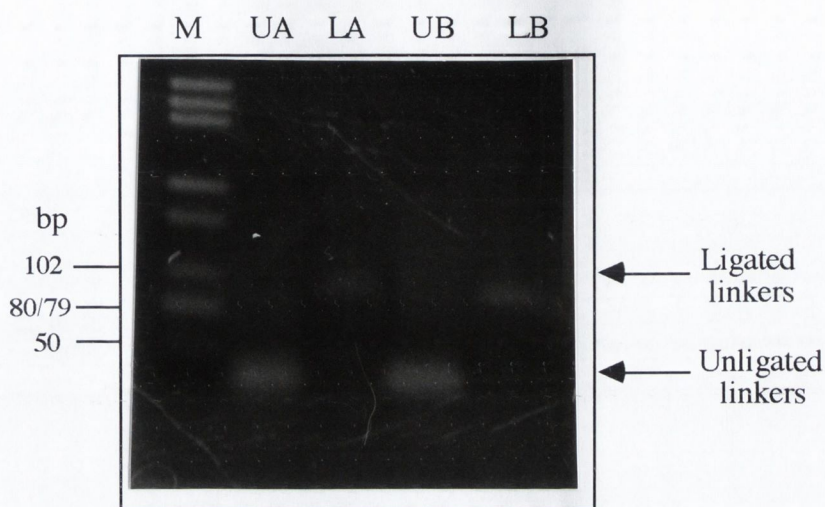


Figure 3.3 SAGE linker ligation test. The reannealed and kinased SAGE linkers (A and B) (Table 3.1) were tested for kinasing efficiency by a self-ligation test. Unligated and ligated linkers were resolved by electrophoresis on a 2.5% agarose gel and stained with ethidium bromide. M= pBS- *Hae* III molecular weight marker and sizes are indicated in base pairs (bp). UA= unligated linker A. LA= ligated linker A. UB= unligated linker B. LB= ligated linker B. The ligated linkers show a clear increase in size (approximately 80-100bp) relative to the unligated linkers and this indicated that the majority of linkers were kinased efficiently as they were capable of forming linker-linker dimers.

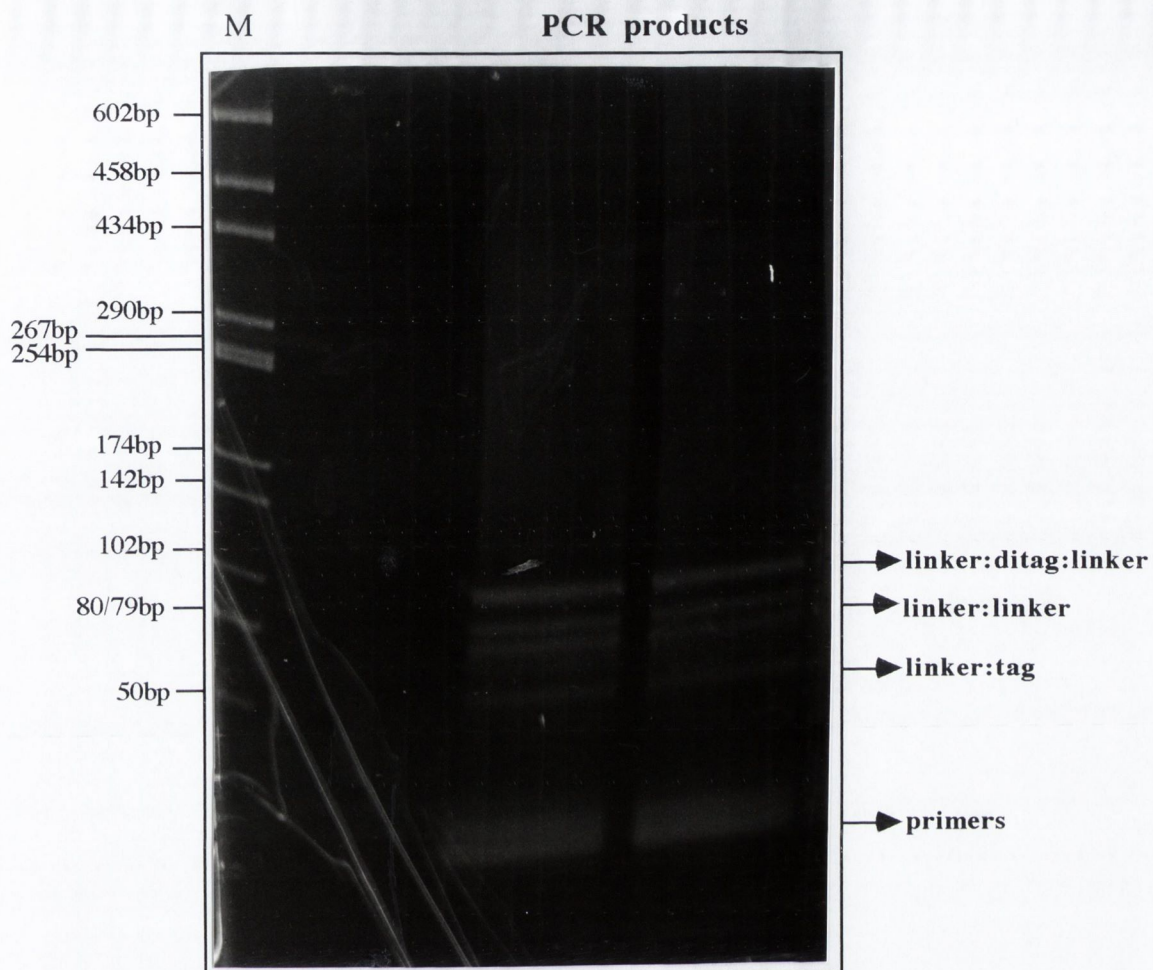


Figure 3.4 PCR amplification of ditags. PCR amplification of the linker:tag ligation yielded a number of PCR products which were resolved on a 12% non-denaturing polyacrylamide gel and stained with ethidium bromide. M= pBS-*Hae* III molecular weight marker and the sizes are described in base pairs (bp). Ligated ditags with linker amplified as 102bp PCR product (linker:ditag:linker). A number of 'background' products were also detected and these included linker ligated to linker, i.e. no ditag in between (linker:linker), and was resolved as two bands of approximately 80bp in size due to the slight size differences between linker A and B (see Table 3.1). An additional background band of approximately 50 to 60bp in size was also detected and was due to linker-tag only (unligated). The smallest band detected (primers) probably consisted of excess primers and 'free' linkers.

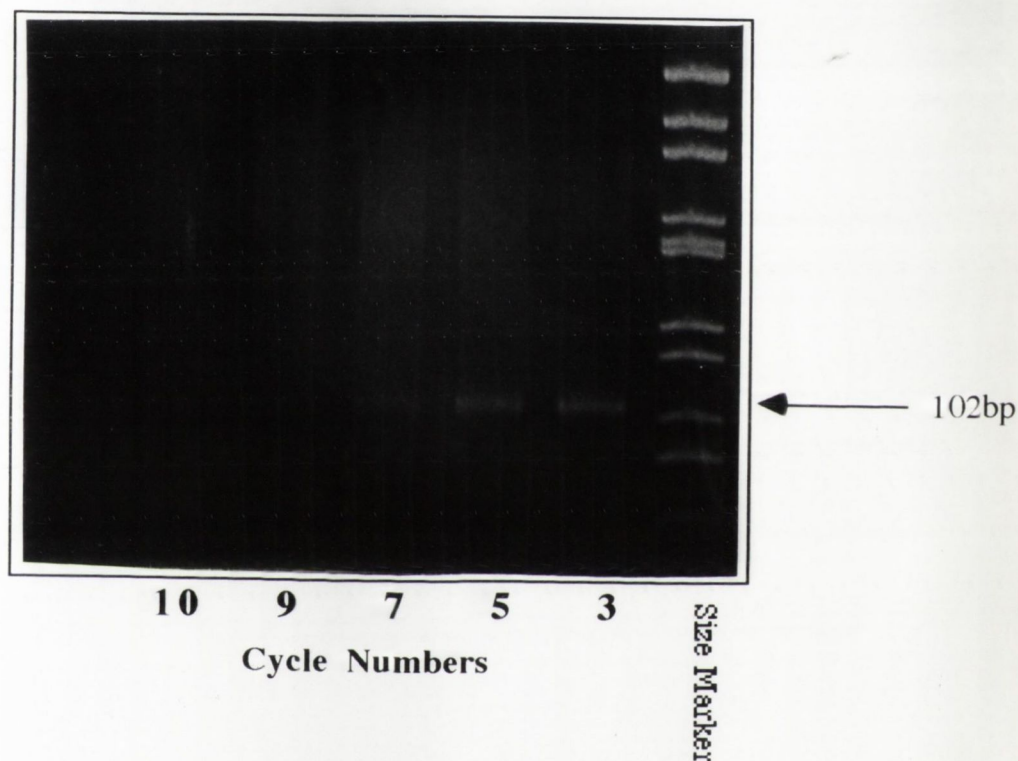


Figure 3.5 A cycle curve to assess the optimum number of cycles for reamplification of the 102bp PCR product. The optimum cycle number was determined by amplifying an aliquot of purified 102bp PCR product and sampling over a range of cycle numbers, i.e. 3 to 10 cycles in the experiment shown above. Samples were resolved on a 12% non-denaturing polyacrylamide gel and stained with ethidium bromide. Size Marker= pBS-*Hae* III. The optimum cycle number in the experiment shown above was 3 and excessive cycles beyond the optimum, i.e. >5, resulted in an apparent 'disappearance' of the 102bp PCR product which may be due to its accumulation as high molecular weight aggregates.

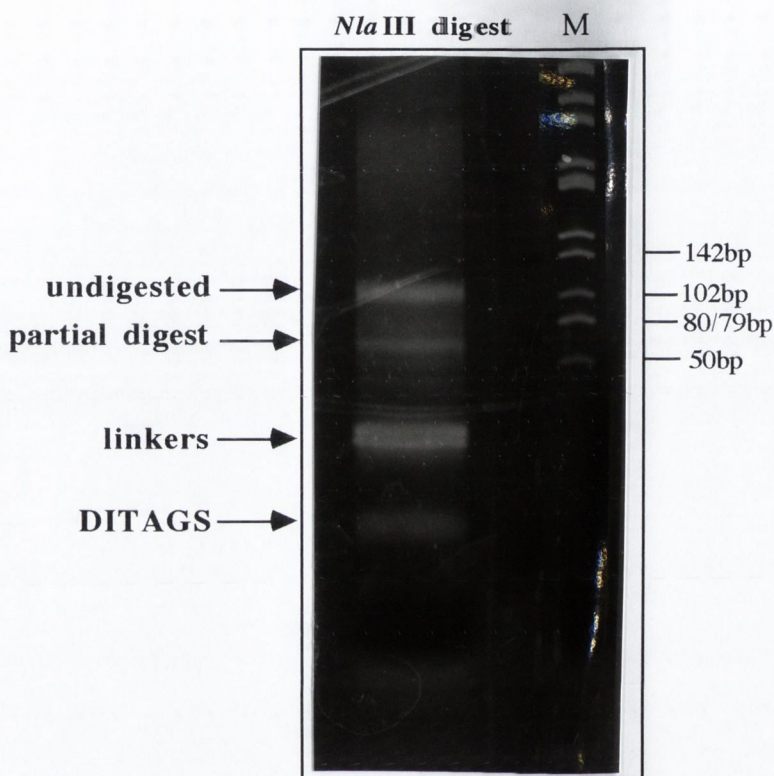


Figure 3.6 Release of ditags from the linkers. The ditags were released from the linkers by digestion with *Nla* III. Ethanol precipitated *Nla* III digested 102bp PCR products were resolved on 12% non-denaturing polyacrylamide gels and stained with ethidium bromide. A partial digestion pattern was typically observed as shown above. M= pBS-*Hae* III molecular weight marker and fragment sizes are described in base pairs (bp). The partially digested product (partial digest) of approximately 60bp in size resulted from cleavage of only one of two *Nla* III sites within the 102bp PCR product. The linkers resolved as two bands of approximately 40bp in size due to slight size differences between the linkers. The released ditags (approximately 26bp in size) were excised from the gel and purified for subsequent ligation.

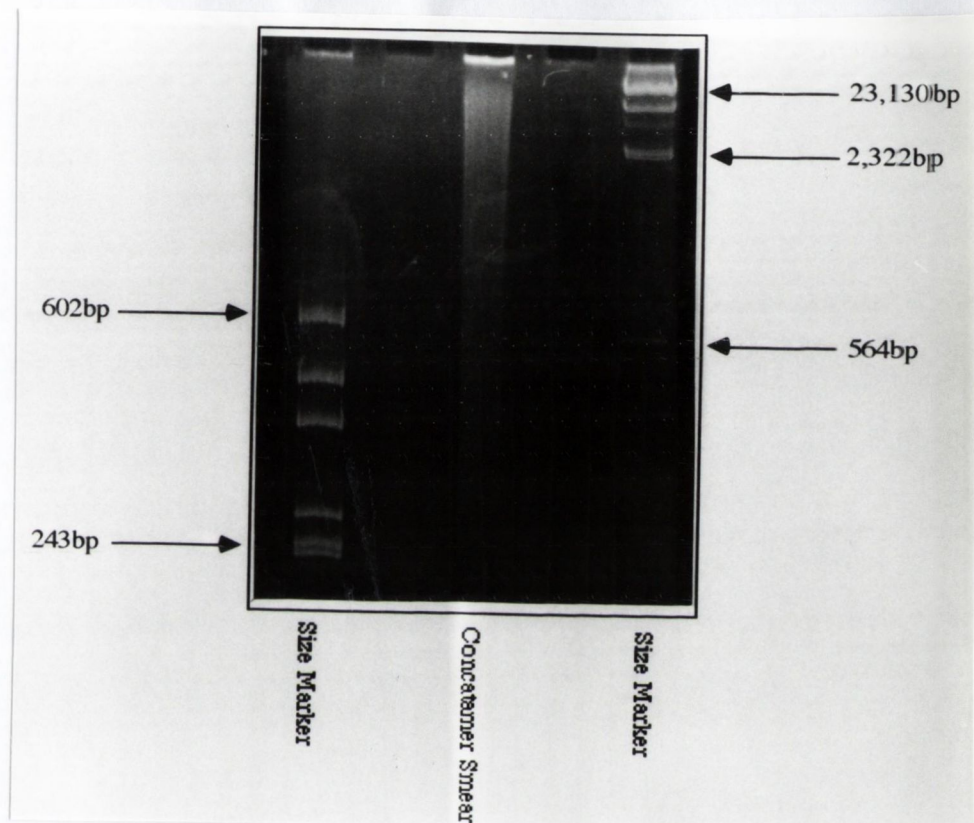
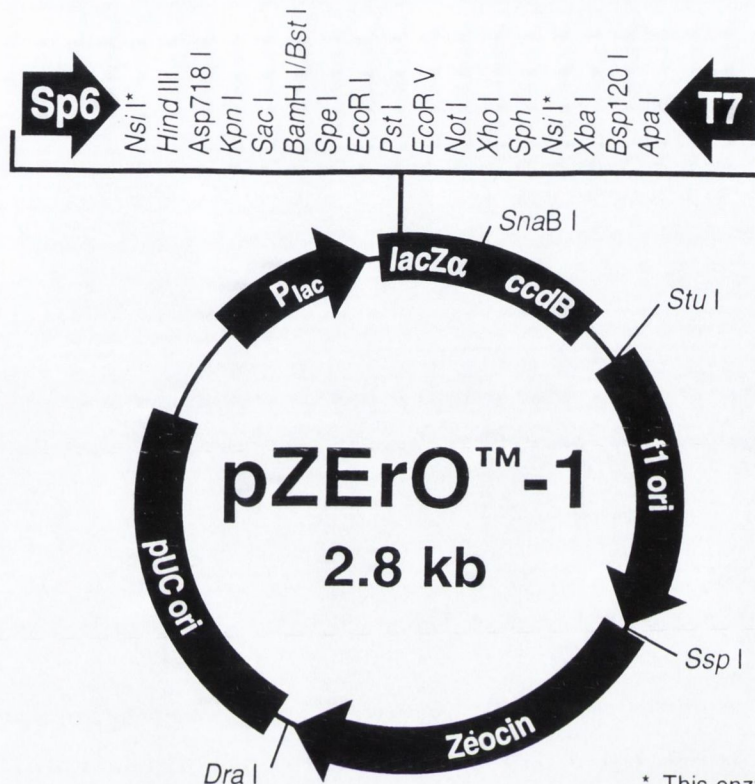


Figure 3.7 Ligation of ditags to form concatamers. The ditag ligation reactions were resolved by electrophoresis on 8% non-denaturing polyacrylamide gels and stained with ethidium bromide. Size Markers= pBS-*Hae* III (left) and λ -*Hin* dIII (right) and fragment sizes are described in base pairs (bp). The ditag ligations resulted in concatamers which resolved as a high molecular weight 'smear' as shown above. Concatamers sized between 400 and 2000bp were gel purified for cloning.



Comments for pZErO™-1
2808 nucleotides

* This enzyme has two recognition sites both of which are found in the multiple cloning site.

Figure 3.8 Map of pZero cloning vector (Invitrogen). The pZero vector was digested with *Sph* I, gel purified and ligated with purified concatamers in the final step of the SAGE protocol. The multiple cloning site is within the lethal gene, *ccdB*, and this allowed growth of recombinants only due to disruption of this gene. Transformed colonies were selected by resistance to the antibiotic Zeocin.

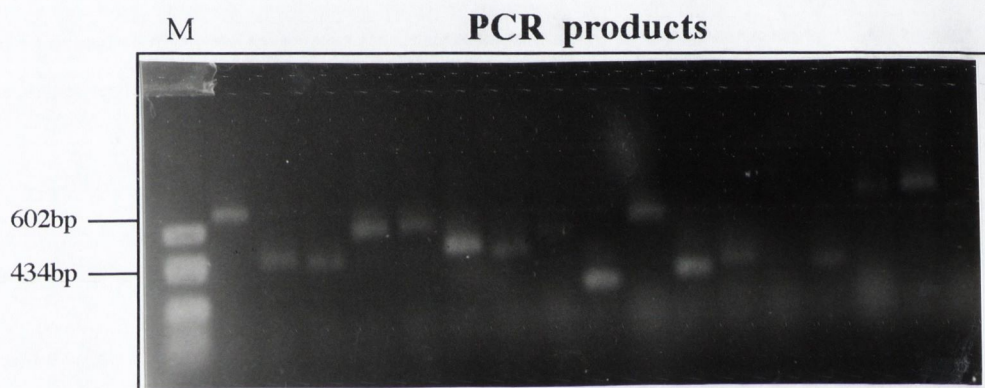


Figure 3.9 PCR screen of SAGE clones. SAGE clones were screened for size by PCR with M13 Forward and Reverse primers. PCR products were visualised on ethidium bromide stained 2% agarose gels as exemplified above. M= pBS-*Hae* III molecular weight marker with fragment sizes shown in base pairs (bp). Clones that produced a PCR product of approximately 450bp, i.e. 10-15 SAGE tags, were selected for sequencing.

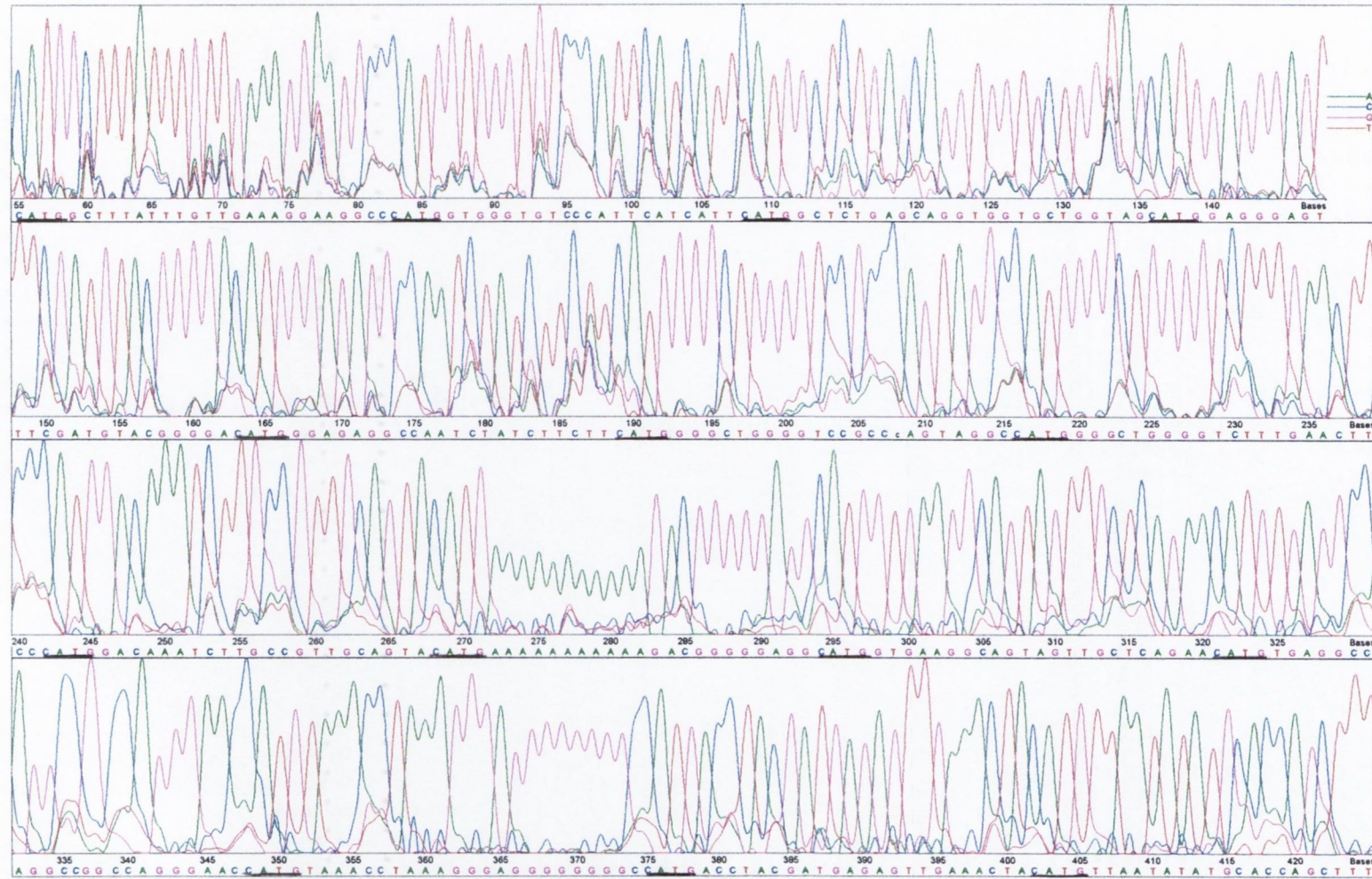


Figure 3.10 The sequence of a typical SAGE clone. The SAGE clones contained the predicated 'CATG' punctuated sequence (underlined) with the ditags in between. The average ditag was between 22 to 26 bases in length and the mixture of sizes was due to the slight variation in cleavage distance by the *Bsm* FI tagging enzyme. All sequence reactions were analysed by the ALFexpress™ automated sequencer (Pharmacia).

Table 3.1 Oligonucleotides for the SAGE protocol

Linkers in **blue** were described in the original SAGE paper (Velculescu *et al.*, 1995) and proved to be suboptimal for PCR amplification (Velculescu *et al.*, 1997a). The anchoring enzyme site (*Nla* III) is underlined and the tagging enzyme site (*BsmF* I) is written in **red**.

Primer/linker Name	Sequence
Linker 1A	5' TTTTACCAGCTTATTCAATTCGGTCCTCTCGCACAGGGACATG 3'
Linker 1B	5' TCCCTGTGCGAGAGGACCGAATTGAATAAGCTGGTA dideoxy 3'
Linker 2A	5' TTTTGTAGACATTCTAGTATCTCGTCAAGTCGGAAGGGACATG 3'
Linker 2B	5' TCCCTTCCGACTTGACGAGATACTAGAATGTCTACAA dideoxy 3'
Linker A1	5' TTTGGATTTGCTGGTGCAGTACAACCTAGGCTTAATAGGGACATG 3'
Linker A2	5' TCCCTATTAAGCCTAGTTGTACTGCACCAGCAAATCC * 3'
Linker B1	5' TTTCTGCTCGAATTCAAGCTTCTAACGATGTACGGGGACATG 3'
Linker B2	5' TCCCCGTACATCGTTAGAAGCTTGAATTCGAGCAG * 3'
Primer 1	5' GGATTTGCTGGTGCAGTACA 3'
Primer 2	5' CTGCTCGAATTCAAGCTTCT 3'
M13 Forward	5' GTAAAACGACGGCCAGT 3'
M13 Reverse	5' GGAAACAGCTATGACCATG 3'
Biotinylated oligo dT	5' [biotin]T ₁₈

* Denotes an amino-modified C₇ at the 3' end of the linker sequence. This modification prevented 'blunt-ended' ligation of linkers at their 5' ends.

Table 3.2 SW480 SAGE gene expression profile

Sequences were analysed using SAGE software version 1.0 (Johns Hopkins University) and tags were identified by comparison to Genbank Flatfile release 101.0 (June, 1997). 'SAGE tag' refers to the 9 to 11 bp sequence adjacent to the 4bp anchoring *Nla* III site. The tag identifies specific mRNA transcripts as it is taken from a defined position within the sequence (Velculescu *et al.*, 1995). 'Tag abundance' is the number of tags counted for any one particular mRNA transcript. Tag abundances of 3 and over are listed only. 'Gene transcript' indicates the match identified for each SAGE tag by comparison to Genbank. 'Acc. No.' is the designated Genbank accession or EST number for each identified gene transcript.

SAGE tag 9/11 *	Tag abundance	Gene transcript	Acc. No.
TGTGTTGAG	42	Elongation Factor 1 α	X16869
CCCATCGTC	37	Cox II	X15759
TTGGTCCTC	36	L41 ribosomal homologue	S64030
GAGGGAGTT	35	L27a ribosomal protein	U14968
TACCATCAA	32	Glyceraldehyde 3-phosphate dehydrogenase	J02642
CACCTAATT	29	cDNA clone	D57986
CTGGGTAA	25	S19 ribosomal protein	M81757
GGATTTGGC	25	EST	11365
AGGGCTTCC	23	equivocal ¹	
AGGCTACGG	22	SMCX mRNA	Z29650

* The majority of transcripts were identified using 9 bases. In some cases 9 bases were not sufficient to identify a specific transcript. In these cases 11 bases were used to identify a transcript and any changes in tag abundance noted.

¹ Tags that were not uniquely identified using 9bp are described as 'equivocal'. The majority of these could be resolved using the extra 2bp as described above.

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
GGGCTGGGG	22	equivocal	
CAAACCATC	21	Cytokeratin 18	X12883
AAGGTGGAG	19	ORF	X80822
CCTCCAGCT	19	Cytokeratin 8	X12882
TTATGGGAT	19	mRNA for MHC protein homologous to chicken B complex	M24194
TAGGTTGTC	18	Human homologue to mouse p21	X64899
ATGGCTGGT	17	LLRep 3 mRNA	X17206
GCCCCTGCT	17	Keratin type II K5	M21389
GCAGCCATC	16	L28 ribosomal protein	U14969
TTCATACAC	15	cDNA clone	D29344
AAAAAAAAA	14	equivocal	
AGCCCTACA	14	cDNA clone	C03782
CCCGTCCGG	14	EST	17304
GTACACACA	14	Cystatin S	X54667
TGGTGTTGA	14	S18 ribosomal protein	X69150
TTCAATAAA	14	equivocal	
CCTTCGAGA	13	S5 ribosomal protein	U14970
CTCAACATC	13	Ribosomal phosphoprotein Po	M17885
GAAAAATGG	13	mRNA for laminin binding protein	X61156
GTTCCCTGG	13	Fau mRNA	X65923
AAGATTGGT	12	Motility related protein-1	X60111

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
AAGGAGATG	12	L31 ribosomal protein	X69181
ACATCATCG	12	L12 ribosomal protein	L06505
TGTACCTGT	12	equivocal	
TTGGGGTTT	12	Ferritin H chain mRNA	M11146
ATGTGTAAC	11	mts 1/CAPL protein	Z33457
CCACTGCAC/TC	11/9	SMA3/SMA5	X83299
CCCTGGGTT	11	Ferritin light chain subunit mRNA	M11147
GAACACATC	11	L19 ribosomal protein	X63527
AAGACAGTG	10	L37a ribosomal protein	X66699
CCGTCCAAG	10	S16 ribosomal protein	M60854
CCTCGGAAA	10	L38 ribosomal protein	Z26876
CTAAGACTT	10	cDNA clone	D56531
GAAGCAGGA	10	Cofilin, non-muscle type	X95404
GGACCACTG	10	L3 ribosomal protein	X73460
GTTTCGTGCC	10	EST	11444
TAAGGAGCT	10	RPS26 mRNA	X77770
TGCACGTTT	10	L32 ribosomal protein	X03342
AATCCTGTG	9	L8 ribosomal protein	Z28407
ACTGGGTCT	9	c-myc transcription factor (puf)	L16785
AGGAAAGCT	9	EST	15369
CCCCAGCCA	9	Hums3 mRNA for 40s ribosomal protein s3	X55715

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
GAGACTCCT/GC	9	EST	79832
GCCGTGTCC	9	S6 ribosomal protein	J03537
GCTTTATTT	9	β -actin	X00351
GTGTTGCAC	9	S13 ribosomal protein	X79239
AAAGTCTAG	8	Prad 1	X59798
AGCACCTCC	8	Elongation factor 2	Z11692
CAAGCATCC	8	cDNA clone	N88134
CCTAGCTGG	8	T cell cyclophilin	Y00052
CGCCGCCGG	8	L35 ribosomal protein	U12465
GCATAATAG	8	L21 ribosomal protein	U14967
GGGAAATC	8	Thymosin β -10 mRNA	M92381
GTGAAACCC	8	equivocal	
TCAGATCTT	8	scar protein mRNA	M22146
TGATTCAC	8	Human autonomously replicating sequence mRNA	L08441
TTACCATAT	8	L39 ribosomal protein	D79205
TTGGTGAAG	8	Thymosin β -4 mRNA	M17733
AAGAAGATA	7	L23a ribosomal protein	U02032
AGGTCCTAG	7	Glutathione S-transferase	X15480
ATAATTCTT	7	S29 ribosomal protein	U14973
CTGTACAGA	7	EST	186648
CTTCCAGCT	7	Lipocortin II mRNA	D00017

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
GCCTGCTGG	7	Glutathione peroxidase-4, GPX-4 mRNA	X71973
GGAGTGGAC	7	L18 ribosomal protein	L11566
GGCCGCGTT	7	S17 ribosomal protein	M13932
GGCTTTACC	7	equivocal	
GTGCTGAAT/GG	7	Smooth muscle myosin alkali light chain mRNA	U02629
TGGCAAAGC	7	equivocal	
ACCGCCGTG	6	p22 phox (CYBA) gene	M61107
ATCAAGGGT	6	L9 ribosomal protein	U09953
ATTCTCCAG	6	L17 ribosomal protein	X52839
CAGGCCCCA/CC	6	cDNA clone	W58644
CCCATCCGA	6	L26 ribosomal protein	X69392
CCTGTAATC	6	equivocal	
CGAGGGGCC	6	cDNA clone	W65372
CGCTGGTTC	6	L11 ribosomal protein	X79234
GCGACCGTC	6	Fibroblast mRNA for aldolase A	M21190
GTGAAGGCA	6	S3a ribosomal protein	M77234
GTGCGGAGG	6	Serum amyloid A mRNA	M10906
AATAGGTCC	5	S25 ribosomal protein	M64716
ACTCCAAAA	5	equivocal	
AGTCTGATG	5	EST	91620
CCAGAACAG	5	L30 ribosomal protein	X79238

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
CCCAAGCTA	5	HSP27	X54079
CGCCGGAAC	5	HRPL4 mRNA	X73974
CTCTTCGAG	5	Glutathione peroxidase-1, GPX-1	Y00433
GATGCTGCC	5	Acute myeloid leukaemia associated protein	L21756
GCTTTTAAG	5	Human mRNA for ribosomal protein	D23660
TCCCCGTAA	5	unidentified	
TCGTCTTTA	5	S7 ribosomal protein	M77233
TGAGGCCAG/GC	5	High mobility group box (SSRP1) mRNA	M86737
ACCGCCTGT/GG	4	EST	186727
AGAAAAAAA	4	equivocal	
CAGCTCACT	4	L14 ribosomal protein	D87735
CCAATAAAG	4	equivocal	
CCAGGAGGA	4	EST	26757
CCAGTGGCC	4	S9 ribosomal protein	U14971
CCCCCTGGA/TC	4	EST	100997
CGGTTACTG/TG	4	EST	19196
CTCATAGCA	4	cDNA clone	N75639
GAAATTTAA	4	mRNA for high mobility group-1	X12597
GAAGCTTTG	4	HSP86	X07270
GACGACACG	4	EST	101103
GATCCCAAC	4	equivocal	

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
GCCGAGGAA	4	S12 ribosomal protein	X53505
GCCTGTATG	4	EST	10576
GCGACGAGG	4	EST	12066
GCTGCCCTT	4	α -tubulin	K00557
GTGTTAACC	4	EST	13598
GTTTCTTCC	4	equivocal	
TCAACTTCT/GG	4	cDNA clone	N57419
TCTCCTGCA/TA	4	Hox 7 mRNA	M97676
TCTGTACAC/CT	4	EST	15947
TCTTGTGCA	4	mRNA for lactate dehydrogenase-A	X02152
TGAGGGAAT	4	Triosephosphate isomerase mRNA	M10036
TTCTTGTGG	4	EST	15728
TTTGTTAAT	4	hnRNP H mRNA	L22009
AAAAATAAA	3	equivocal	
AACTGCTTC	3	cDNA clone	W32468
ACAAGAATT/GT	3	h-Sp1 mRNA	X68194
ACCCTTGGC	3	cDNA clone	D52188
ACTTACCTG	3	mRNA for cytochrome c oxidase subunit V1b	X13923
AGCAGATCA	3	p11, cellular ligand of Annexin II	M38591
AGCAGGGCT	3	SOM172 mRNA	X91195
AGGAGCTGC/TG	3	NADH dehydrogenase-ubiquinone 23kDa subunit	U65579

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
AGGTCAGGA/GA	3	Phosphorylase kinase catalytic subunit PH KG2	S73483
ATCACGCCC	3	cDNA clone	W49840
ATCAGTGGC/TT	3	mRNA for proteasome subunit HsN3	D26600
ATTGGCTTA	3	Prohibitin (PHB) mRNA	S85655
ATTGTTTAT	3	equivocal	
ATTTGAGAA	3	equivocal	
CAAGCCCCA	3	equivocal	
CAATAAACT	3	equivocal	
CACAAACGG	3	EST	15030
CAGTCTCTC	3	EST	11713
CATCTAAAC	3	mRNA for KIAA0038 gene	D26068
CATCTTCAC/CA	3	EST	177313
CCAAACGTG	3	EST	12182
CCGGCGCGT/GT	3	unidentified	
CCGGGTGAT	3	Copper transport protein HAH1 mRNA	U70660
CCTTTGAAC/AG	3	EST	181572
CGTGTTAAT	3	equivocal	
GACATCAAG	3	Keratin 19	Y00503
GAGGAGGGT	3	equivocal	
GAGTGGCTA/TT	3	cDNA clone	W93625
GCAAGCCCC	3	equivocal	

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
GCATTTAAA	3	equivocal	
GCCCCCAAT	3	mRNA for beta-galactoside-binding lectin	X14829
GGCTGATGT/GG	3	Glycyl tRNA synthetase mRNA	U09510
GGCTGGGGG	3	equivocal	
GGCTGGTCT	3	equivocal	
GTGACAGAA	3	equivocal	
GTGACCTCC	3	Cytochrome c oxidase subunit VIII (COX8) mRNA	J04823
GTGAGACCC/CA	3/2	Human clone 2C2 Cri-du-chat critical region mRNA	U10510
TACCCTAAA/AC	3	EST	22307
TCACCCACA	3	cDNA clone	N93688
TCAGACGCA	3	Pro-thymosin-alpha mRNA	M14794
TCCCTATAA	3	unidentified	
TCCCTTTAA	3	cDNA clone	AA0584
TCCTATTAA/GC	3	unidentified	
TCCTTCTCC/AC	3	α -actinin mRNA	K15804
TGAAATAAA	3	equivocal	
TGCAGAACG	3	equivocal	
TGTGATCAG	3	EST	17867
TGTGTGTTT/GT	3	Histone Ha(0) gene	X03473
TTCATTATA	3	equivocal	
TTGGAGATC	3	NADH: ubiquinone oxidoreductase MLRQ subunit	U94586

Table 3.2 SW480 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
TTGGGTTTT/CC	3	mRNA for KiAA0023 gene	D14689
TTGTAATCG	3	Ornithine decarboxylase antizyme (Oaz) mRNA	U09202
TTGTTGTTG	3	mRNA for calmodulin	D45887
TTTAATTGT	3	equivocal	

Table 3.3 SW620 SAGE gene expression profile

Sequences were analysed using SAGE software version 1.00 (Johns Hopkins University) and tags were identified by comparison to Genbank Flatfile release 101.0 (June, 1997). 'SAGE tag' refers to the 9 to 11 bp sequence adjacent to the 4bp anchoring *Nla* III site. The tag identifies specific mRNA transcripts as it is taken from a defined position within the sequence (Velculescu *et al.*, 1995). 'Tag abundance' is the number of tags counted for any one particular mRNA transcript. Tag abundances of 3 and over are listed only. 'Gene transcript' indicates the match identified for each SAGE tag by comparison to Genbank. 'Acc. No.' is the designated Genbank accession or EST number for each identified gene transcript.

SAGE tag 9/11*	Tag abundance	Gene transcript	Acc. No.
TGTGTTGAG	67	Elongation Factor 1 α	X16869
TTGGTCCTC	55	L41 ribosomal homologue	S64030
CACCTAATT	43	cDNA clone	D57986
GAGGGAGTT	31	L27a ribosomal protein	U14968
CTCAACATC	29	Ribosomal phosphoprotein Po	M17885
AGGGCTTCC	28	equivocal ¹	
CCCATCGTC	28	Cox II	X15759
TACCATCAA	26	Glyceraldehyde 3-phosphate dehydrogenase	J02642
CCCGTCCGG	25	EST	17304
CTAAGACTT	24	cDNA clone	D56531

* The majority of transcripts were identified using 9 bases. In some cases 9 bases were not sufficient to identify a specific transcript. In these cases 11 bases were used to identify a transcript and any changes in tag abundance noted.

¹ Tags that were not uniquely identified using 9bp are described as 'mixed'. The majority of these could be resolved using the extra 2bp as described above.

Table 3.3 SW620 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
AGCCCTACA	23	cDNA clone	C03782
ATGGCTGGT	22	LLRep 3 mRNA	X17206
GGATTTGGC	22	EST	11365
TTCAATAAA	21	equivocal	
TTCATACAC	21	cDNA clone	D29344
TCAGATCTT	20	scar protein mRNA	M22146
AGGCTACGG	18	SMCX mRNA	Z29650
TGGTGTTGA	18	S18 ribosomal protein	X69150
TTGGGGTTT	18	Ferritin H chain mRNA	M11146
ACTAACACC	17	Human PACAP type-3/VIP type-2 receptor mRNA	U18810
CCGTCCAAG	17	S16 ribosomal protein	M60854
CCTCCAGCT	17	Cytokeratin 8	X12882
CGCCGCCGG	17	L35 ribosomal protein	U12465
GTGAAACCC	18	equivocal	
CTGTTGGTG	16	Human homologue of yeast ribosomal S28	D14530
GAAGCAGGA	16	Cofilin, non-muscle type	X95404
TGATTTAC	16	Human autonomously replicating sequence mRNA	L08441
ACCCTTGGC	15	cDNA clone	D52188
CCTAGCTGG	15	T cell cyclophilin	Y00052
TTGGTGAAG	15	Thymosin β -4 mRNA	M17733
AGGAAAGCT	14	EST	15369

Table 3.3 SW620 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
CCCCAGCCA	14	Hums3 mRNA for 40s ribosomal protein s3	X55715
CTAGCCTCA	14	mRNA for cytoskeletal γ actin	X04098
GAAAAATGG	14	mRNA for laminin binding protein	X61156
GGGCTGGGG	14	equivocal	
GGGGAAATC	14	Thymosin β -10 mRNA	M92381
TGCACGTTT	14	L32 ribosomal protein	X03342
TGGCAAAGC	14	equivocal	
TGTACCTGT	14	equivocal	
AGCACCTCC	13	Elongation factor 2	Z11692
GCCGAGGAA	13	S12 ribosomal protein	X53505
GTGTTAACC	13	L10 ribosomal protein	L25899
AAGGAGATG	12	L31 ribosomal protein	X69181
ATGTGTAAC	12	mts 1/CAPL protein	Z33457
TTACCATAT	12	L39 ribosomal protein	D79205
TTATGGGAT	12	mRNA for MHC protein homologous to chicken B complex	M24194
TGGGCAAAG/CT	21/11	equivocal	
AAGACAGTG	11	L37a ribosomal protein	X66699
CAAACCATC	11	Cytokeratin 18	X12883
CTGGGTAA	11	S19 ribosomal protein	M81757
GCCGTGTCC	11	S6 ribosomal protein	J03537
GGCTGGGGG	11	equivocal	

Table 3.3 SW620 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
AAGGTGGAG	10	ORF	X80822
CCCATCCGA	10	L26 ribosomal protein	X69392
CCTGTAATC	10	equivocal	
CGCCGGAAC	10	HRPL4 mRNA	X73974
AAAAAAAAA	9	equivocal	
ACACAGCAA/GA	9	cDNA clone	D58814
ATAATTCTT	9	S29 ribosomal protein	U14973
CGCTGGTTC	9	L11 ribosomal protein	X79234
GGCAAGCCC/CA	9	Human clone csa-19	U12404
ATCAAGGGT	8	L9 ribosomal protein	U09953
ATTCTCCAG	8	L17 ribosomal protein	X52839
CCAAACGTG	8	equivocal	
CCTCGGAAA	8	L38 ribosomal protein	Z26876
CCTTCGAGA	8	S5 ribosomal protein	U14970
GCAGCCATC	8	L28 ribosomal protein	U14969
GCCTGTATG	8	EST	10576
GCTTTATTT	8	β -actin	X00351
GTTAACGTC/CC	8	EST	10074
GTTCGTGCC	8	EST	11444
ACTCCAAAA	7	equivocal	
CACAAACGG	7	S27 ribosomal protein	U57847

Table 3.3 SW620 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
GGAGTGGAC	7	EST	10058
TAAAAAATA	7	equivocal	
TAGGTTGTC	7	Human homologue to mouse p21	X64899
TGAAGGAGC/CG	7	Mitochondrial ATPase subunit 9	M16439
AAAACATTC	6	cDNA clone	D56512
AACGCGGCC	6	Macrophage migration inhibitory factor	M95775
CAATAAATG	6	EST	11594
CACTACTCA/CC	6	cDNA clone	D52319
CCAGTGGCC	6	S9 ribosomal protein	U14971
CCGTGGTCG/TG	6	Human fibrillarin (Hfib1) mRNA	M59849
CTGCTATAC/GA	6	L5 ribosomal protein	U14966
CTGTTGATT/GC	6	Liver mRNA DNA binding protein	X04347
GAACACATC	6	L19 ribosomal protein	X63527
GCTTTTAAG	6	Human mRNA for ribosomal protein	D23660
GTGTTGCAC	6	S13 ribosomal protein	X79239
GTTCCCTGG	6	Fau mRNA	X65923
TAAGGAGCT	6	RPS26 mRNA	X77770
TACAAGAGG	6	L6 ribosomal protein	X69391
TACTCTTGG/CA	6	hnRNP L mRNA	X16135
TCACAAGCA/AA	6	alpha NAC mRNA	X80909
TCAGACGCA	6	Prothymosin α mRNA	M14794

Table 3.3 SW620 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
TTGTTGTTG	6	mRNA for calmodulin	D45887
AACGACCTC/GT	5	β -tubulin	V00599
AACTAATAC/TA	5	cDNA clone	W67256
AATCCTGTG	5	L8 ribosomal protein	Z28407
AGGCATTGA/AA	5	Putatively partially transcribed sequence	Z20711
ATAGACATA/AA	5	Pre-mRNA splicing factor SF2p32	M69039
CAGATCTTT/GT	5	Uba52, adrenal mRNA for ubiquitin-52	X56998
CAGTCTCTC	5	S10 ribosomal protein	U14972
GATCCCAAC	5	equivocal	
GCAAAAAAA/AA	5	cDNA clone	N22959
GCGACCGTC	5	Fibroblast mRNA for aldolase A	M11560
GGGTTTTTA/TT	5	DNA binding protein	M24070
GTGAAGGCA	5	S3a ribosomal protein	M77234
TCGAAGCCC/CC	5	EST	F18587
TGAGGGAAT	5	Triosephosphate isomerase mRNA	M10036
TGGCCCCAC	5	M2-type pyruvate kinase mRNA	M23725
TTACCTCCT/TC	5	equivocal	
ACTGGGTCT	4	c-myc transcription factor (puf) mRNA	L16785
ATTGTTTAT	4	equivocal	
ATTTGTCCC	4	equivocal	
CAAGCATCC	4	cDNA clone	N88134

Table 3.3 SW620 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
CCAGAACAG	4	L30 ribosomal protein	X79238
CCCCCACCT/AA	4	Differentiation-dependent A4 protein	L09604
CCCTGGGTT	4	Ferritin light chain subunit mRNA	M11147
CTCAACAGC/AA	4	Translation initiation factor 3	U94855
GCATTTAAA	4	equivocal	
GCCCCCAAT	4	mRNA for beta-galactoside-binding lectin	X14829
GCCCCTCCG	4	EST	112496
GCGACGAGG	4	EST	12066
GCTGCCCTT	4	α tubulin	K00557
GGCAAGAAG	4	equivocal	
GTGAGCCCA	4	EST	178726
TAGACTTAT/TG	4	Mitochondrial aspartate aminotransferase mRNA	M22632
TAGATAATG	4	EST	27859
TCAAATGCA/TC	4	hnRNP C	M16342
TCATCATCT/GA	4	EST	15190
TCCGCGAGA/AG	4	EST	76286
TCTTGTGCA	4	mRNA for lactate dehydrogenase-A	X02152
TTCTTGTGG	4	EST	15728
AAAAGTGA/AA	3	mRNA for surface glycoprotein	Z50022
AACTAAAAA	3	equivocal	
AAGATTGGT	3	Motility related protein-1	X60111

Table 3.3 SW620 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
AAGCAGGGT	3	EST	15697
ACTTACCTG	3	mRNA for cytochrome c oxidase subunit V1b	X13923
AGGAATGCT	3	Threonyl tRNA synthetase	M63180
ATGCAGCCA	3	equivocal	
ATTATTTTT	3	equivocal	
CAATAAACT	3	equivocal	
CCAATAAAG	3	equivocal	
CCACTGCAC/TC	3	SMA3/SMA5	X83299
CCGCCGGCT	3	unidentified	
CGCACCATT	3	RT14 mRNA	D64007
CGGATAACC	3	EST	51115
CTCCTCACC	3	Bak mRNA	U16811
CTTCCAGCT	3	Lipocortin II mRNA	D00017
GAAACAAGA/TG	3	Phosphoglycerate kinase	V00572
GAAATAAAA	3	equivocal	
GGCCCTGAG	3	RNA polymerase II subunit (hsRPB10)	U37690
GGCCGCGTT	3	S17 ribosomal protein	M13932
GGCTCCCAC	3	90kDa heat shock protein mRNA	M16660
GGCTTTACC	3	equivocal	
GGGAAATCG	3	cDNA clone	N85634
GGGGCAGGG	3	equivocal	

Table 3.3 SW620 SAGE gene expression profile

SAGE tag 9/11	Tag abundance	Gene transcript	Acc. No.
GGTCCAGTG	3	equivocal	
GTGACAGAA	3	equivocal	
GTTTTTGCT	3	equivocal	
TAATAAAGG	3	equivocal	
TACCTTCAT	3	equivocal	
TAGCTGAGA/CA	3	Nuclear localisation sequence receptor	U28386
TCACCCACA	3	cDNA clone	W35285
TCCCCGTAA	3	unidentified	
TGTAATCAA	3	equivocal	
TGTGATCAG	3	EST	17867
TTCATTATA	3	equivocal	
TTGGTCCTT	3	equivocal	
TTGTAATCG	3	mRNA for ornithine decarboxylase antizyme (Oaz) mRNA	U09202

Table 3.4 Summary of differentially expressed SAGE tags

	Number of tags increased in the primary tumour cell line ^a	Number of tags increased in the metastatic cell line ^a	Total
Identified genes ^b	27	22	49
EST/cDNA clones ^b	12	17	29
Unidentified tags	3	-	3

^a The number of differences were based on at least a three fold difference in tag abundance and the analysis consisted of approximately 5,000 tags from each cell line.

^b Tags were identified by comparison to Genbank flatfile release 101.0.

Table 3.5 Gene expression differences between SW480 and SW620

SAGE tag is the 9 or 11bp sequence used to identify the gene transcript. SW480/SW620 refers to the relative tag abundance ratio in the cell lines. 'Gene transcript' is the identity of the tag by comparison to Genbank. Acc. No. is the designated Genbank accession number for each identified transcript.

SAGE tag	SW480/SW620 Normalised ^a	SW480/SW620 Actual ^b	Gene transcript	Acc. No.
GCCCCTGCT/GA	17/0	18/1	Keratin type II K5	M21389
GTACACACA	14/0	15/1	Cystatin S	X54667
AAGATTGGT	12/3	13/4	Motility related protein-1 (MRP-1)	X60111
ACATCATCG	12/2	13/3	L12 ribosomal protein	L06505
GGACCACTG	10/0	11/1	L3 ribosomal protein	X73460
AAAGTCTAG	8/2	9/3	Prad1 ¹	X59798
GCATAATAG	8/1	9/2	L21 ribosomal protein	U14967
CTGTACAGA	7/1	8/2	EST	186648
GCCTGCTGG	7/1	8/2	Glutathione peroxidase -4 (GPX-4)	X71973
GTGCTGAAT/GG	7/0	8/1	Myosin light chain mRNA	U02629
ACCGCCGTG	6/1	7/2	p22-phox (CYBA) gene	M21186
CAGGCCCCA/CC	6/0	7/1	cDNA clone	W58644
GTGCGGAGG	6/0	7/0	Serum Amyloid A (SAA)	M10906
AGTCTGATG	5/0	6/1	EST	91620

^a Tag abundances based on normalising each library to 5,000 tags.

^b Tag abundances based on actual numbers observed, i.e. non-normalised.

¹ Identical to Bcl 1 (Acc. No. M73554).

Table 3.5 Gene expression differences between SW480 and SW620

SAGE tag	SW480/SW620 Normalised	SW480/SW620 Actual	Gene transcript	Acc. No.
CCCAAGCTA	5/1	6/2	HSP27	X54079
CTCTTCGAG	5/0	6/1	Glutathione peroxidase-1 (GPX-1)	M21304
GATGCTGCC	5/0	6/1	Acute myeloid leukaemia associated protein	L21756
TGAGGCCAG	5/0	6/0	High mobility group box (SSRP1) mRNA	M86737
TCAACTTCT/GG	4/0	5/0	cDNA clone	N57419
TCTCCTGCA/TA	4/0	5/1	Hox 7 homeobox protein, region 7, mRNA	M97676
TCTGTACAC/CT	4/0	5/1	cDNA clone	D58940
TTTGTTAAT	4/0	5/1	hnRNP H mRNA	L22009
AACTGCTTC	3/0	4/0	cDNA clone	W32468
ACAAGAATT/GT	3/0	4/1	h-Sp1	X68194
AGCAGATCA	3/0	4/0	p11, cellular ligand of Annexin II	M38591
AGGAGCTGC/TG	3/0	4/0	NADH dehydrogenase-ubiquinone 23kDa subunit	U65579
AGGTCAGGA/GA	3/0	4/1	Phosphorylase kinase catalytic subunit PH KG2	S73483
ATCACGCCC	3/0	4/1	cDNA clone	W49840
ATCAGTGGC/TT	3/0	4/1	mRNA for proteasome subunit HsN3	D26600
ATTGGCTTA	3/0	3/0	Prohibitin mRNA	S85655
CATCTTCAC/CA	3/0	4/0	EST	177313
CCGGCGCGT/GT	3/0	4/0	unidentified	
CCTTTGAAC/AG	3/0	4/1	EST	60138
GAGTGGCTA/TT	3/0	4/0	cDNA clone	N83778

Table 3.5 Gene expression differences between SW480 and SW620

SAGE tag	SW480/SW620	SW480/SW620	Gene transcript	Acc. No.
	Normalised	Actual		
GGCTGATGT/GG	3/0	4/0	Glycyl-tRNA synthetase mRNA	U09510
TACCCTAAA/AC	3/0	4/1	EST	22307
TCCCTATAA	3/0	4/1	unidentified	
TCCTATTAA/GC	3/0	4/1	unidentified	
TCCTTCTCC	3/0	4/0	α -actinin mRNA	X15804
TGGGGTACC/TT	3/0	4/0	EST	F20308
TGTGTGTTT/GT	3/0	4/0	Histone H1(0) gene	X03473
TTGGGTTTT/CC	3/0	4/0	mRNA for KiAA0023 gene	D14689
TGGGCAAAG/CT	0/10	1/11	cDNA clone	N31199
ACTAACACC	2/17	3/18	PACAP type 3/VIP type 2 receptor	U18810
CTGTTGGTG/AT	0/16	1/17	Human homologue of yeast ribosomal S28	D14530
ACCCTTGGC	3/15	4/16	cDNA clone	D52188
CTAGCCTCA/CG	2/14	3/15	γ -actin	X04098
GCCGAGGAA/GG	4/13	5/14	S12 ribosomal protein	X53505
GTGTTAACC/AG	4/13	5/14	L10 ribosomal protein	L25899
GGCTGGGGG/CC	3/11	4/12	Profilin mRNA	J03191
ACACAGCAA/GA	0/9	1/10	cDNA clone	C03810
GGCAAGCCC/CA	1/9	2/10	csa-19 cDNA clone	U12404
GTTAACGTC/CC	1/8	2/9	EST	10074
TGAAGGAGC/CG	1/7	2/8	Mitochondrial ATPase subunit 9	M16439

Table 3.5 Gene expression differences between SW480 and SW620

SAGE tag	SW480/SW620	SW480/SW620	Gene transcript	Acc. No.
	Normalised	Actual		
AAAACATTC	1/6	2/7	cDNA clone	D56338
AACGCGGCC	1/6	2/7	Macrophage migration inhibitory factor	M95775
CACTACTCA/CC	1/6	2/7	cDNA clone	C02633
CTGCTATAC/GA	0/6	1/7	L5 ribosomal protein	U14966
CTGTTGATT/GC	0/6	1/7	Liver mRNA DNA binding protein	X04347
TACTCTTG/CA	0/6	0/7	hnRNP L mRNA	X16135
AACGACCTC/GT	0/5	1/6	β -tubulin	VOO599
AGGCATTGA/AA	0/5	0/6	Putatively transcribed partial sequence	Z20711
ATAGACATA/AA	0/5	1/6	Pre-mRNA splicing factor SF2p32	M69039
CAGATCTTT/GT	1/5	2/6	Uba52, adrenal mRNA for ubiquitin-52	X56998
GCAAAAAAA/AA	0/5	1/6	cDNA clone	N22959
TCGAAGCCC/CC	0/5	1/6	EST	F18587
CTCAACAGC/AA	0/4	0/5	Translation initiation factor 3	U94855
GCCCCTCCG	0/4	0/5	EST	112496
GTGAGCCCA	0/4	0/5	EST	178726
TAGACTTAT	0/4	0/5	Mitochondrial aspartate aminotransferase mRNA	M63180
TCAAATGCA/TC	0/4	0/5	hnRNP C	M16342
TCCGCGAGA/AG	0/4	0/5	EST	76286
AAAAGTGA/AA	0/3	0/4	mRNA for surface glycoprotein	Z50022
AAGCAGGGT	0/3	0/4	EST	15697

Table 3.5 Gene expression differences between SW480 and SW620

SAGE tag	SW480/SW620	SW480/SW620	Gene transcript	Acc. No.
	Normalised	Actual		
AGGAATGCT	0/3	1/4	Threonyl rRNA synthetase mRNA	M63180
CGGATAACC	0/3	1/4	EST	51115
CTCCTCACC	0/3	1/4	Bak mRNA	U16811
GAAACAAGA/TG	0/3	1/4	Phosphoglycerate kinase	V00572
GCACCCGCC/TG	0/3	0/4	cDNA clone	T40544
GGGAAATCG	0/3	1/4	cDNA clone	N85634
TAGCTGAGA/CA	0/3	1/4	Nuclear localisation sequence receptor	U28386

CHAPTER 4

Analysis of differentially expressed genes

4.0 INTRODUCTION

Comparison of SAGE tag abundances from the cell lines SW480 and SW620 identified a total of 81 gene transcripts that showed a difference in expression level of at least three fold (see Chapter 3). The three main categories of differentially expressed gene transcripts encode proteins whose functions are associated with RNA synthesis/processing, protein synthesis/processing and cell structure/motility. Comparisons of SAGE tag abundances have previously been shown to correspond to actual levels of gene expression (Hermeking *et al.*, 1997; Madden *et al.*, 1997; Polyak *et al.*, 1997; Velculescu *et al.*, 1997; Zhang *et al.*, 1997; Hibi *et al.*, 1998) and usually involve sequencing in the region of 20,000 to >50,000 tags. However, as the tags sequenced in this study numbered approximately 10,000 in total, the significance of the differences identified had to be confirmed by an independent method. Some of the more conventional methods available for this include northern blotting (Sambrook *et al.*, 1989), reverse northern blotting (Mou *et al.*, 1994) and quantitative RT-PCR based methods (Gilliland *et al.*, 1990; Noonan *et al.*, 1990; He *et al.*, 1995). The SAGE method yields 15 bases of sequence information which are used to identify specific Genbank entries as described in Chapter 3. A disadvantage of the SAGE method is the lack of cDNA clones containing enough sequence for further experimental manipulation. The SAGE method yields 15 bp of cDNA sequence ('tag') per transcript which cannot be easily extracted from other tags in the clone and would not contain a sufficient amount of sequence for practical usefulness. Therefore, the most suitable method for confirmation of differential expression was a quantitative RT-PCR based method, which involved designing specific primer sets for gene transcripts of interest.

The main quantitative RT-PCR based methods include semi-quantitative RT-PCR (He *et al.*, 1995), coamplification of an internal reference template (Noonan *et al.*, 1990) or competitive PCR (Gilliland *et al.*, 1990). Semi-quantitative RT-PCR involves reverse transcription of the target mRNA to cDNA, followed by PCR amplification using gene-specific primer pairs and previously optimised conditions.

A housekeeping gene such as β -actin is amplified in parallel to control for the amount of input RNA and efficiencies of the subsequent reactions. It is essential that the exponential phase of the PCR is being observed and, therefore, a range of different cycle numbers are sampled or a series of cDNA template dilutions are amplified in parallel. The second method is similar to semi-quantitative RT-PCR except the reference template (such as a housekeeping gene) is amplified in the same tube as the target template. Therefore, two different primer sets are present in the same reaction mix and require the same amplification conditions. The need for the same reaction conditions for each primer set is often not practical. In addition, this method also requires the exponential phase of each primer set to occur within an overlapping window and therefore, when abundances of mRNAs are significantly different this method would be unsuitable as the exponential phases of each PCR cannot be observed simultaneously. The final quantitative RT-PCR method, the competitive PCR method, is based on inclusion of a modified competitor template which is amplified using the same primers as the target cDNA. The competitive template differs by having a small intron or a mutant internal restriction enzyme site. The competitor template may also be included in the reverse transcription step by transcribing it from an RNA expression vector. The use of a competitor template is an excellent way of controlling for efficiencies of reverse transcription and PCR amplification. However, the major limitation of this method is that it does not control for the amount of input RNA or the level of degradation of different samples. In addition, the need to construct a competitor template does not make it a suitable method for analysing a range of different genes. Therefore, the method of choice for confirming differential expression of some of the SAGE data described here was semi-quantitative RT-PCR.

Semi-quantitative RT-PCR was performed according to the method of He *et al.* (1995) with modifications including a Cy5 labelled primer in all PCR amplifications. This rendered the amplified PCR products fluorescently labelled which allowed accurate quantitation by resolving the PCR products on an

ALFexpress™ automated sequencer as described in Chapter 2. The experiments were carried out in triplicate (at least) and the amount of PCR product was expressed as a ratio to that of β -actin. Semi-quantitative RT-PCR was used initially to confirm differential expression of a subset of putative differentially expressed transcripts identified from the SAGE analysis in the cell lines SW480 and SW620 and which are listed in Table 4.1. Based on the semi-quantitative RT-PCR data the panel of differentially expressed tags identified from the SAGE analysis (Chapter 3, Table 3.5) was re-assessed (see Section 4.2.1). The confirmed differences were further analysed by northern blot using gel purified cDNA probes as described in Chapter 2. Northern blot analysis was primarily performed to ensure that the amplified RT-PCR products hybridized to the expected size of transcript. Moreover, some of the cDNA probes had the potential to detect alternatively spliced or polyadenylated transcripts that would not be detected by semi-quantitative RT-PCR. Differentially expressed splice variants of the transmembrane glycoprotein CD44 are well documented in colon cancer progression (Sherbet & Lakshmi, 1997) and a similar phenomenon may also occur in other candidate metastasis-associated genes.

Confirmed differentially expressed transcripts in SW480 and SW620 were also examined in other colon carcinoma cell lines (Table 4.2) by semi-quantitative RT-PCR. These included a second isogenic pair of primary colon and metastatic cell lines (SW1116 & 498LI) and three other metastatic cell lines all originating from primary colon tumours (LoVo, Colo201 and T84). The potential role for the identified candidate metastasis-associated genes in colon cancer progression is discussed.

4.1 RESULTS

The expression of a subset of gene transcripts (Table 4.1) that were differentially expressed based on SAGE tag abundances were analysed further by semi-quantitative RT-PCR and/or northern blot. This included a total of 11 transcripts that showed either a substantial difference in expression level or encoded proteins with functions that could contribute to tumour metastasis. Genbank entries were retrieved using either the Entrez program (<http://www3.ncbi.nlm.nih.gov/Entrez/>) or the ACNUC Retrieval system (Gouy *et al.*, 1985) provided by the Irish National Centre for Bioinformatics (<http://acer.gen.tcd.ie>). Gene specific primer sets were designed using the Oligo 4 program and are described in Chapter 2. PCR conditions were optimised using either human genomic DNA or cDNA as template and variables such as primer concentration, MgCl₂ concentration and annealing temperature were optimised for each primer pair (data not shown). The optimised conditions are described in Chapter 2, Section 2.3.7.

4.1.1 Controls for semi-quantitative RT-PCR and northern analysis

The semi-quantitative RT-PCR and northern blot analyses were both controlled using the "housekeeping" gene β -actin. Housekeeping genes are usually constitutively expressed and may be used as an internal standard as their expression levels often remain constant within the same tissue type and across experimental samples. However, altered expression of housekeeping genes between different treatments or cell types has been widely reported (Fort *et al.*, 1985; de Leeuw *et al.*, 1989; Goldsworthy *et al.*, 1993) and therefore the validity of using a housekeeping gene is normally assessed prior to any quantitative analysis. The observation of similar SAGE tag abundance for β -actin between the isogenic cell lines SW480 and SW620 (Table 4.1) suggested that β -actin was a suitable control for the semi-quantitative experiments. In addition, total RNA samples that were balanced using spectrophotometer readings, ethidium bromide dot quantitation and by assessment on

denaturing agarose gels (described in Chapter 2) showed similar expression levels of β -actin by both semi-quantitative RT-PCR and northern blot analysis.

Semi-quantitative RT-PCR

Reverse transcription reactions were carried out on total RNA samples that were balanced as mentioned above and under the same conditions as described in Chapter 2. Amplification of β -actin was carried out from the same reverse transcription reaction as the target genes and all the components of the PCR reaction were combined as a 'master mix' and added to each separate cDNA template. The exponential phase of each new set of reverse transcription reactions was assessed by sampling an extended range of cycle numbers. β -actin PCR products (109bp; see Table 2.2 for primer positions and PCR conditions) were initially resolved by electrophoresis on 2% agarose gels or 12% polyacrylamide gels and stained with ethidium bromide. Figure 4.1a shows β -actin levels for SW480 and SW620 for the initial analysis and also for the subsequent analysis of the 5 additional cell lines with SW480 and SW620 included as controls. Aliquots of β -actin PCR products were also resolved by electrophoresis on 6% denaturing polyacrylamide gels on the ALFexpress™ automated sequencer to facilitate quantitation as described in the following sections.

A control was also required to ensure that potentially contaminating genomic DNA had been successfully removed. This was tested using an intron-spanning primer set which amplified a band of 256bp from genomic DNA and a band of 181bp from cDNA as described in Chapter 2 Section 2.3.7. Figure 4.1b clearly shows the absence of genomic DNA in all reverse transcribed RNA samples from each cell line.

Northern blot

Equivalent amounts of total RNA (assessed as described above) were northern blotted as described in Chapter 2 Section 2.3.11. The blot was hybridised with a β -actin cDNA probe and Figure 4.1c shows equivalent amounts of input RNA and

efficiency of transfer. Transcript size was estimated by migration relative to 28s and 18s rRNA as shown in Figure 4.1d.

4.1.2 Confirmed differentially expressed SAGE tags

Semi-quantitative RT-PCR experiments were performed at least three times. Conditions and primers for each PCR are described in Chapter 2 Section 2.3.7. PCR products, sampled from a number of time-points, were resolved by electrophoresis on 2% agarose gels and subsequently quantitated using the ALFexpress™ automated sequencer as described in Chapter 2 Section 2.3.7. Expression levels were compared between samples as a ratio relative to the amount of β -actin PCR product and the data analysed using GraphPad (Prism™, USA).

4.1.2.1 Keratin K5

Keratin K5 specific primers amplified a PCR product of 146bp between bases 573 and 700 of the Genbank sequence (M21389) (Figure 4.2a). Semi-quantitative RT-PCR analysis in the primary tumour (SW480) and metastatic (SW620) cell lines revealed a high level of expression in the primary cell line and barely detectable levels in the metastatic cell line. This agreed with the SAGE tag abundances of 18/1 (tumour/metastatic cell line). Figure 4.2b shows the resolved ethidium bromide stained PCR products from one of the experiments sampled at three time-points. This is represented graphically (Figure 4.2c) using β -actin: target gene ratios which were quantified from the peak areas generated by the Allele Links™ software as described in Chapter 2 Section 2.3.7. Fold differences are shown in Table 4.3. Subsequent northern blot analysis (as described in Chapter 2) revealed the same pattern of expression as the semi-quantitative RT-PCR analysis and the probe hybridized to a transcript which migrated just above the 18s rRNA band. This correlated with the expected size of 2,234 bases (M21389) and is shown in Figure 4.2d.

4.1.2.2 Cystatin S

PCR amplification with Cystatin S specific primers generated a product size of 141bp from bases 57 to 179 of the Genbank sequence (X54667) (Figure 4.3a). Semi-quantitative RT-PCR analysis showed a similar expression pattern to the above mentioned Keratin K5, i.e. high expression in the primary tumour cell line (SW480) and undetected in the metastatic cell line (SW620). This agrees with the SAGE tag abundances of 15/1 (tumour/metastatic cell line). The experiment was repeated >3 times and Figure 4.3b shows the ethidium bromide stained PCR products from one of the experiments which is represented graphically in Figure 4.3c as described above. Fold differences are shown in Table 4.3. Northern analysis revealed a similar expression pattern as the semi-quantitative RT-PCR analysis. The Cystatin S specific probe hybridised to a transcript that migrated below the 18s rRNA and therefore was consistent with an expected transcript size of 736 bases (X54667) (Figure 4.3d).

4.1.2.3 Serum Amyloid A (SAA)

The human SAA gene family includes the acute phase SAA1 and SAA2 which are approximately 90% identical to each other, SAA3 is a nontranscribed pseudogene and SAA4 is referred to as constitutive SAA and shares 55% sequence identity with the acute phase SAA1 and SAA2 (Kluve-Beckermann & Song, 1995). Moreover, SAA1 and SAA2 are polymorphic; SAA1 has three alleles: SAA1 α , SAA1 β and SAA1 γ , while SAA2 has two: SAA2 α and SAA2 β (Sipe *et al.*, 1985; Woo *et al.*, 1987; Kluve-Beckermann *et al.*, 1988; Betts *et al.*, 1991; Beach *et al.*, 1992; Baba *et al.*, 1995; Steel *et al.*, 1993). Alignment of the available nucleotide sequences of SAA1 and SAA2 (Figure 4.4) revealed that the SAGE tag does not distinguish between SAA1 α or SAA2 α . Primers were designed specifically for SAA1 α but each primer shared 19/20 bases with both forms of SAA2. It is likely that the PCR product amplified was specific for SAA1 α but coamplification of SAA2 could not be ruled out. Therefore, further mention of SAA refers to the acute phase SAAs.

PCR amplification with SAA specific primers generated a PCR product of 140bp in size from bases 192 to 312 (M10906) (Figure 4.5a). The semi-quantitative RT-PCR analysis of the cell lines SW480 and SW620 agreed with the SAGE tag abundances of 7/0 (tumour/metastatic cell line), i.e. a high level of expression in the tumour cell line (SW480) and expression in the metastatic cell line (SW620) was not detected. The experiment was repeated >3 times and quantitated as described above. Ethidium bromide stained PCR products sampled for three time-points are shown in Figure 4.5b and relative expression is shown graphically in Figure 4.5c as β -actin ratios and as described above. Fold differences are shown in Table 4.3. The SAA cDNA probe detected a transcript which migrated below 18s and further below Cystatin S (736 bases), which is consistent with the expected transcript size of approximately 369 bases (M10906) (Figure 4.5d).

4.1.2.4 Human homologue of yeast ribosomal S28

Primers specific for human homologue of yeast ribosomal S28 (referred to hereafter as S28) were derived from bases 65 and 196 of the Genbank sequence (D14530) (Figure 4.6a) and amplified a product of 151bp in size. SAGE tag abundance data suggested an increased expression of 17 fold, i.e. 1/17 (tumour: metastatic) in the metastatic cell line (SW620) relative to the primary tumour cell line (SW480). The semi-quantitative RT-PCR analysis did show increased expression of S28 in the metastatic cell line but the difference relative to the primary tumour cell line was approximately 2.2 fold (see Table 4.3). The semi-quantitative RT-PCR results are shown in Figure 4.6b and 4.6c. Northern blot analysis revealed a similar expression difference as the semi-quantitative RT-PCR data. The S28 cDNA hybridized to a transcript that migrated below the 18s rRNA and in between the Cystatin S (736 bases) and SAA (369 bases) transcripts (Figure 4.6d). This correlated with an expected transcript size of approximately 488 bases. A greater difference in expression between the two cell lines was expected based on the SAGE data and the reason for this discrepancy is unclear at this point.

4.1.2.5 p32 subunit of Splicing Factor 2 (SFp32)

SFp32 specific primers amplified a PCR product of 136bp from bases 724 and 840 of the Genbank sequence M69039 (Figure 4.7a). Semi-quantitative RT-PCR analysis showed increased expression of SFp32 in the metastatic cell line (SW620) relative to the primary tumour cell line (SW480) (Figure 4.7b). The relative expression difference was approximately 2 fold (Table 4.3) which was expected from the SAGE tag abundances of 1/6 (tumour/metastatic cell line) (Figure 4.7c and Table 4.3). Northern blot analysis agreed with the semi-quantitative RT-PCR and SAGE analysis. The SFp32 cDNA probe hybridised to a transcript that migrated below the 18s rRNA band and above the SAA transcript (736 bases) which correlated with an expected transcript size of 1201 bases (M69039) (Figure 4.7d).

4.1.3 Unconfirmed differentially expressed SAGE tags

Semi-quantitative RT-PCR experiments were performed at least three times. Conditions and primers for each PCR are described in Chapter 2 Section 2.3.7. PCR products, sampled from a number of timepoints, were resolved by electrophoresis on 2% agarose gels and quantitated using the ALFexpress™ automated sequencer as described in Chapter 2 Section 2.3.7. Expression levels were compared between samples as a ratio relative to the amount of β -actin PCR product and the data analysed using GraphPad (Prism™, USA).

4.1.3.1 Myosin light chain

Myosin light chain specific primers amplified a PCR product of 122bp designed from bases 545 and 647 of the Genbank sequence (U02629) (Figure 4.8a). The SAGE data showed increased expression in the primary tumour cell line (SW480) relative to the metastatic cell line, i.e. 8/1 (tumour/metastatic cell line). However, repeated semi-quantitative RT-PCR experiments did not show a consistent difference in expression (Figure 4.8b, 4.8c).

4.1.3.2 hnRNP L

hnRNP L specific primers amplified a PCR product of 193bp between bases 830 and 1003 from the Genbank sequence X16135 (Figure 4.9a). Repeated semi-quantitative RT-PCR experiments did not consistently show a difference in expression level of hnRNP L between the cell lines SW480 and SW620 (Figure 4.9b). Northern blot analysis confirmed the semi-quantitative RT-PCR analysis (Figure 4.9c). These results are contradictory to the SAGE tag abundances of 0/7 (tumour/metastatic). The hnRNP L cDNA probe hybridized to a transcript that migrated higher than the 18s rRNA band which is consistent with a transcript of approximately 2.3kb in size as described previously (Pinol-Roma *et al.*, 1989).

4.1.3.3 High mobility group box protein (SSRP1)

PCR primers specific for SSRP1 amplified a product size of 179bp designed from bases 2452 and 2611 of the Genbank sequence M86737 (Figure 4.10a). The semi-quantitative RT-PCR analysis did not agree with the SAGE tag abundances of 6/0 (tumour/metastatic cell line). In contrast, repeated experiments showed a consistent, slight increase in the metastatic cell line (SW620) relative to the primary tumour cell line (SW480). The results are shown in Figure 4.10b, 4.10c.

4.1.3.4 β -tubulin

β -tubulin specific primers amplified a PCR product of 118bp from bases 306 and 404 of the Genbank sequence V00599 (Figure 4.11a). Repeated semi-quantitative RT-PCR analysis showed a marginal increase in expression of β -tubulin in the metastatic cell line (SW620) relative to the primary tumour cell line (SW480) as shown in Figure 4.11b, 4.11c. A more definitive difference was expected from the SAGE tag abundances of 1/6 (tumour/metastatic cell line).

4.1.3.5 hnRNP C

hnRNP C has two alternative polyadenylation transcripts referred to as C1 and C2 (Nakagawa *et al.*, 1986; Swanson *et al.*, 1987). The SAGE tag originated from the longer C1 transcript and showed increased expression in the metastatic cell line (SW620) relative to the primary tumour cell line (SW480) (0/5; tumour: metastatic). Primers specific for hnRNP C1 were designed from bases 1465 and 1596 of the sequence (M16342) (Figure 4.12a) and generated a PCR product of 151bp. The sequence of the SAGE tag from the shorter transcript (C2) was compared to the gene expression profiles of each cell line using the SAGE software as described in Chapter 3. This sequence was identified from the 9 bases adjacent to the last *Nla* III site at the 3' end of the C2 transcript. This may have allowed the expression pattern of C2 to be determined. However, no matches were found, which suggested that the C2 transcript is either not expressed at a high level in either cell line or not expressed at all. Semi-quantitative RT-PCR analysis with primers specific for C1 and subsequent northern blot analysis revealed no difference in expression between the two cell lines (Figure 4.12b, 4.12c) which was in contrast to the SAGE data already mentioned. In addition, the cDNA probe used in the northern analysis originated from the longer transcript (see Figure 4.12a) and therefore was expected to detect a single transcript of approximately 1.9kb in size as described previously (Swanson *et al.*, 1987). However, the single transcript that was detected migrated with the 28s rRNA band and therefore is much larger than the expected size of 1.9kb, i.e approximately 5kb. Larger transcripts for hnRNP C have been reported previously but were much lower in abundance than the two primary transcripts already mentioned (Nakagawa *et al.*, 1986). As hnRNP C1 did not appear to be differentially expressed based on the above experiments, the discrepancy in the size of transcript was not further investigated.

4.1.3.6 PACAP type-3/VIP type-2 receptor

The sequence for PACAP type-3/VIP type-2 receptor was deposited in Genbank in July, 1996 and designated as U18810 (Wei & Mojssov, 1996). The Genbank flatfile release 101.0 used in this work to identify SAGE tags was last updated in June 1997. Therefore, a SAGE tag showing a difference of 3/18 (tumour: metastatic cell line) was identified as PACAP type-3/VIP type-2 receptor. However, subsequent BLAST searches (Altschul *et al.*, 1990) against a more recent update of Genbank (October 1998) identified a region in the 3' untranslated region of the PACAP type-3/VIP type-2 receptor sequence (bases 1668 to 1814) which shared 100% identity to the mitochondrial-encoded NADH dehydrogenase subunit 2 (ND2) gene (AF014892, partial) (Figure 4.13). It is likely that the submitted sequence of the PACAP type-3/VIP type-2 receptor is due either to a recombination event or to a cloning artefact. Regardless of which of the latter two scenarios is correct, the identity of the SAGE tag was ambiguous as the shared region contained the tag sequence. The sequence of the ND2 gene was submitted to Genbank on the 6th November 1997, which was after submission of the PACAP type-3/VIP type-2 receptor sequence and therefore the identical region would not have been detected in the latter sequence at the time of submission. A pair of primers were designed to test whether the sequence of PACAP type-3/VIP type-2 receptor is a real transcript (Set 1). A second set of primers were designed from within the open reading frame of the PACAP type-3/VIP type-2 receptor sequence (Set 2) as shown in Figure 4.13. Extensive optimisation experiments were carried out for both sets of primers on both genomic and cDNA templates. Neither set yielded specific products of expected size (data not shown). The absence of a PCR product using Set 1 suggested that the sequence of the PACAP type-3/VIP type-2 receptor (U18810) is the result of a cloning artefact. However, Set 2 should have produced a PCR product from genomic DNA at least and the lack of a PCR product may have been due to technical difficulties or additional artefacts in the sequence for PACAP type-3/VIP type-2 receptor. The inability to uniquely assign the SAGE tag to one specific sequence

meant that the PACAP type-3/VIP type-2 receptor sequence was not investigated further.

4.1.4 Analysis of expression in other metastatic cell lines

The level of expression of the five confirmed differentially expressed genes (Section 4.1.2) was examined in other colon carcinoma cell lines by semi-quantitative RT-PCR as described in the previous section. These included a pair of isogenic cell lines: SW1116, a human colon adenocarcinoma line and 498LI, a metastatic variant of SW1116 selected *in vitro* (Lieberman *et al.*, 1991). Three metastatic cell lines, LoVo, Colo201 and T84, all originating from primary colon tumours, were also examined (the cell lines are described in Table 4.2). The cell lines SW480 and SW620 were included as controls. Fold differences were all quantified relative to β -actin levels in each cell line (Table 4.3) and resolved β -actin PCR products are shown in Figure 4.1a.

Expression of Keratin type II K5 (Figure 4.14) and Cystatin S (Figure 4.15) was barely detected in all five cell lines tested. Increasing the number of cycles to 30-40 yielded a faint PCR product of correct size but also resulted in a number of non-specific products (data not shown). Serum Amyloid A (SAA) showed decreased expression in all the metastatic cell lines tested (Figure 4.16) which was consistent with the observed downregulation of SAA in the metastatic cell line (SW620) relative to the primary tumour cell line (SW480). SAA showed a 16 fold decrease in LoVo, an 8 fold decrease in Colo201 and was undetected in T84. There was a slight decrease in expression in 498LI when compared to SW1116 (Table 4.3). Expression of the human homologue of yeast ribosomal S28 in the metastatic cell lines showed a 6 fold increase in the LoVo cell line but a similar or slightly lower level of expression in Colo201 and T84 relative to SW480. A slight decrease in expression was observed in 498LI relative to SW1116 (Figure 4.17). Finally, the p32 subunit of human pre-mRNA splicing factor SF2 showed a 2 fold increase in 498LI compared to SW1116

and its expression in the three other metastatic cell lines showed a 1.6 fold increase in LoVo but a slightly lower level of expression in Colo201 and T84 (Figure 4.18).

4.1.5 Summary of results

A subset of 11 genes, identified as being differentially expressed following the SAGE analysis, were analysed further by semi-quantitative RT-PCR and/or northern blot. This analysis resulted in confirmation of differential expression for five of the genes which included Keratin K5, Cystatin S, Serum Amyloid A, the human homologue of yeast ribosomal S28 and the p32 subunit of human pre-mRNA splicing factor SF2. Keratin K5, Cystatin S and Serum Amyloid A all showed decreased expression in the metastatic cell line SW620 while the human homologue of yeast ribosomal S28 and the p32 subunit of human pre-mRNA splicing factor SF2 both showed increased expression in the metastatic cell line relative to the primary tumour cell line (SW480). The PCR products from all five of the above mentioned genes hybridized to the expected size of transcript. Analysis of the five confirmed differentially expressed gene transcripts in other metastatic colon tumour cell lines was also performed. Keratin K5 and Cystatin S were not detected in the five additional metastatic cell lines (Table 4.2). Serum Amyloid A showed varying degrees of decreased expression in all metastatic cell lines examined relative to the primary tumour cell lines. The human homologue of yeast ribosomal S28 and the p32 subunit of human pre-mRNA splicing factor SF2 both showed increased expression in at least one other metastatic cell line besides SW620.

Differential expression of the following gene transcripts was not confirmed: Myosin light chain, hnRNP L, SSRP1 and hnRNP C. In addition the hnRNP C specific PCR product hybridized to a transcript size that was larger than expected. Although β -tubulin showed an increase in expression in the metastatic cell line (SW620), this difference was only marginal. Finally, analysis of the sequence of PACAP type-3/VIP type-2 receptor (U18810) revealed peculiarities within the sequence which may have arisen due to a cloning artefact.

4.2 DISCUSSION

4.2.1 Re-assessment of the SAGE data

A SAGE analysis, based on a total of over 10,000 tags, identified a panel of 81 gene transcripts which showed a difference in expression level, of at least three fold, between a primary tumour cell line (SW480) and a metastatic cell line (SW620) (see Chapter 3). A subset of 11 differentially expressed SAGE tags were analysed further by semi-quantitative RT-PCR and/or northern blot. This analysis showed that only 5 of these genes, which included Keratin type II K5, Cystatin S, Serum Amyloid A, the human homologue of yeast ribosomal S28 (S28) and the p32 subunit of human pre-mRNA splicing factor SF2 (SFp32), represented true differences in expression. The remaining 6 which did not correlate with SAGE tag abundances included Myosin light chain, hnRNP L, SSRP1, β -tubulin, hnRNP C and PACAP type-3/VIP type-2 receptor. Myosin light chain, hnRNP L, and hnRNP C did not show a consistent difference in expression after repeated experiments. SSRP1, which was expected to show increased expression in the primary tumour cell line based on SAGE tag abundances, in fact showed a slight decrease in expression relative to the metastatic cell line. β -tubulin only showed a very marginal increase in the metastatic cell line, while identification of PACAP type-3/VIP type-2 receptor from its corresponding SAGE tag could not be established due to a possible cloning artefact.

This confirmatory analysis has shown that a SAGE analysis based on approximately 5,000 tags per cell type results in a number of false positives despite the 'housekeeping' genes GAPDH and β -actin showing a similar tag number (Table 4.1). SAGE tags which were uniquely identified and proved to be false positives were based on tag abundances of <10. This suggests that differences in expression identified from tag abundances of <10 are not reliable. However, this does not mean that all tags below ten in number do not represent true differences as exemplified by Serum Amyloid A and the p32 subunit of pre-mRNA splicing factor SF2. Moreover, tag abundances >10 seemed to represent true differences in gene expression. Thus,

increasing the number of tags sequenced should reduce or eliminate any false positives.

In light of the results described above, the 81 gene transcripts which showed differences in SAGE tag abundances (Table 3.5, Chapter 3) were re-assessed. Table 4.4 lists gene transcripts which have not been confirmed to date, but are likely to represent true differences in expression based on their tag abundances. This list has been extended to include gene transcripts which showed at least a 2 fold difference in expression. A second list (Table 4.5) details gene transcripts with tag abundances of <10 in one cell line but which were not detected at all in the other cell line. The second list of genes is more 'hit and miss' and they may prove to be false positives. However, these genes cannot be dismissed at this point as they may represent actual differences in expression.

Re-assessment of the SAGE data based upon the observations above reduced the number of identified differentially expressed transcripts from 81 to 49. This analysis has shown that with a moderate level of SAGE tags, comparison of the gene expression profiles must be interpreted with caution as tag abundances <10 do not consistently represent true differences in expression. This agrees with another recent SAGE analysis on endothelial cells which consisted of over 12,000 tags (de Waard *et al.*, 1999) and again false positives were all derived from tag abundances of <10. However, a cautious analysis of SAGE data based on approximately 10,000 tags has allowed identification of real gene expression differences amongst highly abundant transcripts.

4.2.2 Comparison of SAGE data to other studies in SW480 and SW620

There are at least two studies which have applied differential gene expression methodologies to the colon tumour cell lines SW480 and SW620 (Kondoh *et al.*, 1992; Cajot *et al.*, 1997). The first study used the method of Differential Display Reverse Transcriptase-PCR (DDRT) (described in Chapter 5) to identify gene expression differences in a matched series of human carcinoma cell lines derived

from a primary cancer of the ascending colon and both liver and peritoneal metastases from the same patient. This study showed high levels of expression of Motility-related Protein (MRP1/CD9) in the primary cell line relative to the two metastatic cell lines (Cajot *et al.*, 1997). They also examined expression in the cell lines SW480 and SW620 and found the same pattern of expression, i.e. increased expression in the primary tumour cell line (SW480) relative to the matched metastatic cell line (SW620). This agreed with this SAGE analysis which showed a tag abundance of 13/4 (SW480/SW620) for the MRP1/CD9 transcript. The fold difference is also similar as northern blot analysis showed a 2.5 fold higher level of expression in SW480 relative to SW620 (Cajot *et al.*, 1997) and the SAGE data indicated an approximate 3 fold difference.

The second study showed increased expression of both the ribosomal protein S19 and laminin-binding protein (LBP) in a tumour-enriched cDNA library in which cDNAs in normal colon tissue are subtracted out (Kondoh *et al.*, 1992). However, the analysis was extended to 12 cell lines including SW480 and SW620 and they found increased expression of both genes in the metastatic cell line SW620 relative to SW480 by northern blotting. This is in contrast to the SAGE data which shows a slight increase in expression of S19 in the primary cell line SW480 relative to SW620 (25/11; SW480/SW620) and an approximately equal level of expression of LBP (13/14; SW480/SW620). Although expression of S19 and LBP have not been analysed in this work by a method other than SAGE, the tag abundances for each occur within the range that is likely to be significant, i.e. >10 (see Section 4.2.1). In addition, S19 has recently been isolated from an SW480 cDNA library subtracted against SW620 (Peter McWilliam, personal communication). This suggests that S19 is increased in expression in SW480 relative to SW620, which agrees with the SAGE tag abundances for S19. The reason for the contrasting results may be due to the growth conditions of the two cell lines. For the SAGE analysis (Chapter 3) and the SW480 cDNA subtracted library (Peter McWilliam, personal communication), RNA was harvested from subconfluent cell cultures. In addition, the study which identified

differential expression of MRP1/CD9 (Cajot *et al.*, 1997) referred to above, also harvested RNA from subconfluent cultures of SW480 and SW620. However, Kondoh *et al.* (1992) maintained the cell lines for 48 hours in a confluent condition before harvesting. Maintaining cells for such a long period of time and in an overgrown state is not desirable as cells usually become stressed (unpublished observations) which is likely to affect their gene expression patterns. Moreover, a control probe was not used in their northern blot analysis to check for equal efficiency of transfer. Therefore, the SAGE tag abundances for S19 and LBP are likely to represent the actual levels of expression of these genes in the cell lines SW480 and SW620.

4.2.3 Expression of confirmed differences in other cell lines

The lack of access to a bank of matched tumour specimens for RNA extraction, which included normal colon, primary tumour and most importantly, metastatic tumour meant that further analysis of transcript levels was restricted to cell lines. The expression of genes which were confirmed as being differentially expressed in the cell lines SW480 and SW620 (see Section 4.1.2) was examined in five other colon tumour cell lines. A similar pattern of expression in another cell line with similar metastatic characteristics would associate that gene with the development of metastasis. The availability of matched colon tumour cell lines similar to SW480 and SW620 is quite limited. However, a second pair of matched isogenic cell lines was obtained and included SW1116 and 498LI (Lieberman *et al.*, 1991) and these are described in Table 4.2. Three other metastatic cell lines (LoVo, Colo201 and T84) which all originated from primary colon tumours were also obtained, but these lacked matched primary tumour cell lines. Therefore, expression levels (relative to β -actin) were compared to the primary tumour cell line SW480 to detect any possible up- or down-regulation of the five genes in the metastatic cell lines. This is feasible when substantial differences in expression of specific genes is observed but may not be suitable to detect small differences such as 2-4 fold as there are likely to be small

fluctuations in β -actin levels between different individuals. Relative expression differences for the 5 genes in all 7 cell lines are summarised in Table 4.3.

Keratin type II K5 and Cystatin S were not expressed at a significant level in the four other metastatic cell lines examined (LoVo, Colo201, T84 and 498LI) which is consistent with the original observation of loss of expression in the metastatic cell line SW620. However, expression was not detected in the primary tumour cell line SW1116 either. If loss of expression of Keratin type II K5 or Cystatin S is a feature of tumour cells becoming metastatic then expression would have been expected in SW1116 similarly to SW480. This may be explained by considering that the absence of expression in another primary tumour cell line may indicate that loss of expression of Keratin type II K5 and Cystatin S may occur as an early or late event in tumour progression. In support of this, a recent SAGE analysis of non-small cell lung cancer has identified loss of expression of another type II Keratin in primary tumours as compared to normal tissue (Hibi *et al.*, 1998). Alternatively, Keratin type II K5 and Cystatin S may not be expressed at all during colon tumourigenesis and detection of expression in SW480 and the absence in SW620 may be due to a cell line artefact. The latter explanation is unlikely at least in the case of Cystatin S, which occurs with a high frequency in a cDNA library constructed from Poly(A)+ isolated from an adenocarcinoma of the left colon with ganglionic and liver metastasis (Dukes D) obtained from a male with no previous treatment (Frigerio *et al.*, 1995).

Serum Amyloid A (SAA) showed varying degrees of loss of expression in all of the metastatic cell lines that were examined with complete loss of expression in T84 similarly to SW620. In the second pair of cell lines a slight, but definite decrease in expression was observed in the metastatic cell line (498LI) relative to the primary tumour cell line (SW1116) (see Table 4.3 and Figure 4.5). Therefore, expression of SAA in all the metastatic cell lines examined is inversely correlated with colon tumour metastatic ability.

The expression pattern of the human homologue of yeast ribosomal S28 (S28) amongst the cell lines was rather varied. The increased expression shown in the

metastatic cell line SW620 was only observed in one of the four other metastatic cell lines, i.e. LoVo. Expression in the remaining metastatic cell lines was either approximately the same or slightly lower than the primary tumour cell lines. Interestingly, an increase in expression of S28 had previously been described in colon tumours as compared to normal tissue (Zhang *et al.*, 1997). The further increased expression observed in two of the metastatic cell lines as compared to the primary tumour cell lines indicates that expression of S28 may be correlated with colon cancer progression.

The expression of p32 subunit of Splicing Factor 2 (SFp32) was increased by approximately 2 fold in both metastatic cell lines (SW620, 498LI) relative to their matched primary cell line (SW480, SW1116) (see Table 4.3). There was also a slight increase in expression observed in one of the other metastatic cell lines, i.e. LoVo. An increase of just 2 fold would be difficult to detect amongst the unmatched metastatic cell lines as comparison should ideally be with the primary tumour from which they were derived and not compared to another primary tumour cell line isolated from a different person. The observed increase in expression of SFp32 in both matched metastatic cell lines is significant and shows the importance of detecting small differences in expression which may prove to be more important than gross changes in gene expression.

4.2.4 Identification of ESTs

The reassessed list of differentially expressed SAGE tags (see Section 4.2.1) contains approximately 36% EST or cDNA clones. As described in Chapter 3, identification of the transcript from which the EST (Expressed Sequence Tag) or cDNA clone is facilitated by a number of publicly available database tools such as Unigene (<http://www.ncbi.nlm.nih.gov/Schuler/Unigene/>), the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>), The EST Assembly Machine (<http://hercules.tigem.it/cgi-bin/uniestass.pl>), or EST blast (<http://www.hgmp.mrc.ac.uk/ESTBlast/>). ESTs are nucleotide sequences generated

by sequencing either end of a randomly selected cDNA clone (Borsani *et al.*, 1998). Attempts were made to identify 3 ESTs using either the Unigene or TIGR Human Gene Index. These included EST 1 (N31199) (Table 4.4), EST 7 (178726) (Table 4.5) and EST 8 (76286) (Table 4.5). EST 1 was identified as a HL23 ribosomal protein homologue (X55954) using the Unigene website and its amino acid sequence was found to be 100% identical to Elongation Factor 1- γ (EF1- γ). Moreover, the 9 base SAGE tag was not sufficient to uniquely assign EST 1 as it could not be distinguished from other ESTs or EF1- γ and therefore the extra two bases were needed to uniquely identify the tag as EST 1 (see Chapter 3). Thus, the sequences for the HL23 ribosomal protein homologue and EF1- γ are identical at the amino acid level with slight variations of the nucleotide sequence in the vicinity of the SAGE tag. This suggests that these cDNA sequences are either highly homologous or are the same gene sequence with minor differences due to polymorphisms or sequencing errors. EST 7 was identified as heat shock protein 90 kDa using the TIGR human gene index, however, analysis of the transcript contig revealed that the last SAGE tag in the sequence did not correspond to the SAGE tag from EST 7. Therefore, the SAGE tag assigned to EST 7 does not originate from heat shock protein 90 kDa and remains unidentified. Finally, attempts were made to identify EST 8, but no 'hits' were identified.

The above EST analysis demonstrates that the SAGE tag assignments are putative identifications and are limited to the sequences that are present in the database. Care must be taken to ensure that the SAGE tag is identified from the full length sequence of a given transcript, in particular the 3' end as that is the section of the cDNA sequence that is used to identify the SAGE tags. However, the majority of cDNA libraries are generated from reverse transcribed RNA using an oligo d(T) primer and therefore are biased for the 3' end of the transcript. Therefore, the majority of SAGE tag assignments are usually correct.

4.2.5 Roles of differentially expressed genes in metastasis

4.2.5.1 Cytoskeletal and motility associated genes

The cytoskeleton in higher eukaryotes is composed of three major classes of proteins which were named according to diameter size and include: actin microfilaments (6nm), intermediate filaments (10nm) and microtubules (23nm) (Darnell *et al.*, 1986; Fuchs & Weber, 1994). The cytoskeletal structures play a central role in the control of cytoplasmic organisation and cell shape, cell motility and invasion, proliferation, differentiation and the transduction of extracellular signals (Sherbet & Lakshmi, 1997; Sutherland & Witke, 1999). Therefore, the observed altered expression of a number of cytoskeletal and motility associated gene transcripts between primary and metastatic cell lines was not surprising. These included Keratin K5, Keratin K18, α -actinin and Motility Related Protein-1 (MRP-1/CD9) which showed decreased expression in the metastatic cell line, and γ -actin, Profilin and Scar, all of which showed increased expression in the metastatic cell line. With the exception of Keratin K5 the relative difference in expression levels ranged from approximately two to five fold as determined from the SAGE tag abundances (see Tables 4.1, 4.4 and 4.5). Keratin K5 was barely detected in all of the metastatic cell lines examined (see Section 4.2.3).

Intermediate filaments

Keratin K5 and K18 are intermediate filament components of type II and type I respectively. Other categories of the intermediate filament superfamily include desmin, vimentin, neurofilament proteins (NF) and glial fibrillary acidic protein (GFAP) but the keratins are the largest and most complex group (Fuchs & Weber, 1994). Pairs of type II (K1-K8) and type I (K9-K20) keratins are usually expressed differentially in most epithelial cells at various stages of development and differentiation (Fuchs & Weber, 1994). Keratin K5 pairs with Keratin K14, while Keratin K8 usually pairs with K18. Tumour cells often retain the differentiation properties of the cells from which they are derived and therefore antibodies against

keratins are often used to identify carcinomas of epithelial cell origin (Darnell *et al.*, 1986). Other intermediate filaments detected in the SAGE libraries but which were not differentially expressed included Keratin K8 (SW480/SW620: 19/17) and K19 (SW480/SW620: 3/3). Interestingly, Keratin K14, which usually pairs with K5 was not detected in either SAGE library which indicates that it is not highly expressed in either cell line.

Several studies have associated mutations in either K5 or K14 with epidermolysis bullosa simplex (EBS) (Oshima, 1992; Fuchs & Weber, 1994; Sorensen *et al.*, 1999). EBS is subdivided into three major types depending on severity and includes Weber-Cockayne (mildest), Koebner (intermediate) and Dowling-Meara (severe). EBS is largely an autosomal dominant disorder and results in intraepidermal blistering due to cytolysis within the basal layer of the epidermis (Fuchs & Weber, 1994). Therefore, interruption of the intermediate filament network makes cells more fragile and prone to breakage upon mechanical stress. This emphasises the role keratin filaments play in maintaining the mechanical framework of the cell. The expression of Keratin K5 was shown to be drastically reduced in metastatic cell lines derived from primary tumours of the colon (see Table 4.3). This was the first indication that Keratin K5 may have a role to play in colon tumour progression. However, expression of Keratin K5 has been correlated both positively and negatively with tumour progression. Paine *et al.* (1995) showed by RT-PCR and *in situ* hybridization studies that spliced rat Keratin K5 message was absent in anaplastic cells and present in the benign keratin-producing cells from which they were derived. An opposite effect was found in head and neck squamous cell carcinoma, i.e. presence of a spliced transcript of Keratin K5 was correlated with metastatic spread to the lymph nodes (McDonald *et al.*, 1998). It has been proposed that metastatic potential is possibly related to defective differentiation of the tumour cell (Sherbet & Lakshmi, 1997) and this is substantiated by poorly differentiated colon tumours being frequently associated with poor prognosis (Deans *et al.*, 1992). Moreover, keratins are known to be expressed differentially during differentiation of

epithelial cells (Fuchs & Weber, 1994) and, thus, loss of Keratin K5 may affect the ability of the tumour cells to differentiate and could contribute to the metastatic phenotype.

Cytokeratins 8 and 18 have been found to be highly expressed in colon tumours (Frigerio *et al.*, 1995) and the SAGE analysis also found this to be the case. Moreover, the SAGE analysis also showed a slight decrease (<2 fold, tumour/metastatic: 21/11) in expression of cytokeratin 18 in the metastatic cell line, while cytokeratin 8 showed an approximate equal tag abundance (tumour/metastatic: 19/17). Down-regulation of a number of cytokeratins has been reported previously in tumour cells/tissue relative to matched normal, which suggests that alteration of cytokeratins can occur as an early or late event in tumour progression. Another type II keratin (T27795), as mentioned in Section 4.2.3, showed decreased expression in non-small cell lung cancer relative to normal cells (Hibi *et al.*, 1998). In addition, Keratin K20 showed decreased expression in colon tumour relative to normal tissue (Zhang *et al.*, 1997), while decreased expression of Keratin K19 has been correlated with increased invasiveness in oral squamous cell carcinoma cell lines (Crowe *et al.*, 1999). Thus, altered expression of intermediate filaments is associated with tumour progression and several studies have found a negative correlation of cytokeratin expression and tumour progression (Paine *et al.*, 1995; Zhang *et al.*, 1997; Hibi *et al.*, 1998; Crowe *et al.*, 1999). Differential expression of cytokeratins in relation to metastasis may be due to defective differentiation, as already suggested for Keratin K5, or may be due to alteration of many of the other functions associated with the cytoskeleton including control of cell motility and shape.

Actin microfilaments and motility

Microfilaments are composed of polymers of actin subunits (G-Actin) and are important for cell motility and chemotaxis and are also associated with a number of cellular functions such as cytokinesis, endocytosis and signal transduction (Sutherland & Witke, 1999). Motile cells assemble actin filaments at the leading

edge and crosslink them into a rigid, orthogonal network (Machesky *et al.*, 1999). The SAGE analysis has identified increased expression of γ -actin plus two actin-binding proteins: Profilin and Scar (see Table 4.4) in the metastatic cell line (SW620). Profilin is an essential, ubiquitous protein that binds monomeric (G) actin in a 1:1 molar ratio. Profilin has a major role in the assembly of actin filaments at the barbed ends. It has a dual function and acts as an inhibitor of actin polymerisation when the barbed ends of the filament are blocked by capping proteins and in motile regions of the filament, where uncapped barbed ends are actively elongating, the profilin-actin complex can actively participate in filament growth (Gutsche-Perelroizen *et al.*, 1999). The Arp2/3 complex creates new actin filaments by stabilising actin oligomers as they form spontaneously rather than binding actin monomers directly. The Scar protein is an activator of *de novo* actin filament assembly through the Arp2/3 complex and the new filament is assembled as a branch of an existing filament (Machesky *et al.*, 1999). The increased expression of γ -actin, Profilin and Scar in metastatic cells suggests increased actin polymerisation in these cells and therefore implies either an increase in motility or an alteration of cell shape.

Motility Related Protein-1 (MRP-1/CD9) and possibly α -actinin showed decreased expression in the metastatic cell line (SW620) relative to the primary tumour cell line (SW480) (see Tables 4.4 and 4.5). The differential expression of MRP-1/CD9 between the cell lines SW480 and SW620 had also been observed by Cajot *et al.* (1997) as described in Section 4.2.2. Several studies, in a number of different cancers including breast (Miyake *et al.*, 1995 & 1996), non-small cell lung (Higashiyama *et al.*, 1995), oesophageal (Uchida *et al.*, 1999) and colon cancer (Cajot *et al.*, 1997; Mori *et al.*, 1998), have reduced expression of MRP-1/CD9 associated with metastasis. The CD9 antigen was originally associated with B lymphocyte differentiation and was later found to be expressed in a wide range of many nonhaemopoietic tissues and is believed to have a role in platelet activation and aggregation (Jennings *et al.*, 1990). Cloning and sequencing of cDNAs of MRP-1 (Miyake *et al.*, 1991) and CD9 (Lanza *et al.*, 1991) and subsequent sequence

comparisons found MRP-1 and CD9 to be the same transcript. The MRP-1/CD9 protein is a member of the TM4SF superfamily, which bind to one another to form a cell surface-associated tetraspan network. The effect of expression of MRP-1/CD9 on motility has led to some conflicting observations. Overexpression of MRP-1/CD9 suppressed motility in a number of highly motile human cell lines (Ikeyama *et al.*, 1993) while higher levels of endogenous MRP-1/CD9 were correlated with increased motility in primary colon tumours relative to their metastatic derivatives (Cajot *et al.*, 1997). The latter study included the cell lines used in the SAGE study, i.e. SW480 and SW620, and the motility was suppressed by an antibody against MRP-1/CD9. Another study has also shown a higher spontaneous locomoting activity in the primary tumour cell line, SW480, relative to the lymph node metastatic cell line SW620 (Kubens & Zanker, 1998). Regardless of the effect MRP-1/CD9 has on motility the underlying consistent observation is that reduced expression of MRP-1/CD9 in primary tumours is correlated with poor prognosis (Higashiyama *et al.*, 1995; Miyake *et al.*, 1996; Mori *et al.*, 1998; Uchida *et al.*, 1999). However, Uchida *et al.* (1999) observed that expression of MRP-1/CD9 was inversely correlated with invasion of local lymph nodes but not with distant metastases. It is clear that tumour cells at various stages of malignancy will have different requirements. Therefore, genes that are differentially expressed at an early stage in tumourigenesis may be altered again at a later stage. This phenomenon was also recently observed for Heparin/heparan sulphate interacting protein (HIP) which was increased in expression in early colon tumourigenesis and then down-regulated at a later stage to favour metastasis (Wang *et al.*, 1999).

The expression of α -actinin may be down-regulated in the metastatic cell line SW620 (see Table 4.5) but this has yet to be verified by an independent method. However, in the event of decreased expression being confirmed the association of α -actinin with the cytoskeleton is significant. α -actinin is an actin binding protein that colocalises with integrins at focal adhesion plaques in response to an integrin binding a ligand (Kumar, 1998). Focal adhesions are protein complexes that link the integrins

to the cytoskeletal complexes. Integrins contain a large extracellular domain, a short transmembrane domain and a cytoplasmic carboxy terminal domain of variable length and are capable of binding a range of ligands some of which are in the extracellular matrix. Thus, integrins are important for adhesion and a number of signalling pathways which are involved in cytoskeletal organisation, cell proliferation and cell survival pathways (Kumar, 1998). Down-regulation of α -actinin may reflect the altered motility or adhesive properties of the metastatic cell in response to integrin signalling.

In summary, the expression of cytokeratins, MRP-1/CD9 and actin binding proteins has been shown to be altered in metastatic colon tumour cell lines. These differences in expression may result in alteration of cell motility, shape or cytoskeletal organisation which are necessary steps for tumour metastasis.

4.2.5.2 Cystatin S

The cystatins are members of a superfamily of cysteine proteinase inhibitors that are classified into three groups: family 1 (the stefins), family 2 (the cystatins), and family 3 (the kininogens) (Barrett *et al.*, 1986). Cystatin S strongly inhibits papain and ficin, and partially inhibits stem bromelain and bovine cathepsin C (Isemura *et al.*, 1984). The cDNA for Cystatin S was cloned by Bobek *et al.* (1991) and expression of Cystatin S mRNA seemed to be restricted to salivary tissues. The cystatins are secretory proteins and their physiological role is associated with regulation of protein metabolism or protection of cells and tissues against inappropriate proteolysis (Saitoh *et al.*, 1998). Loss of Cystatin S expression was observed in all of the colon metastatic cell lines examined (see Section 4.1.4). Increased proteolysis is a necessary step in metastatic progression (Cockett *et al.*, 1998) and therefore loss of expression of an inhibitor of proteolysis in metastatic cells is consistent with metastatic spread. In addition, recent studies have shown down-regulation of Cystatin M (Sotiropoulou *et al.*, 1997) and Cystatin B (Shiraishi *et al.*, 1998) in breast and oesophageal cancers respectively. Cystatin M was identified as a

novel cystatin by differential display cloning by comparison of primary and metastatic breast tumours from the same patient and showed the closest homology to Cystatin C (Sotiropoulou *et al.*, 1997). Moreover, overexpression of Cystatin C in B16F10 melanoma cells inhibited motility and *in vitro* metastasis by 50% (Sexton & Cox, 1997). Thus, down-regulation of three different cystatins has been correlated with metastasis in different tumour types. In addition, expression of Cystatin B has recently been associated with prevention of cerebellar apoptosis (Pennacchio *et al.*, 1998).

The differential expression of Cystatin S and the other cystatins mentioned above suggests a role for the cystatins in progression of primary tumours to malignancy. However, while Cystatins M, C and B are expressed in a variety of tissues (Sotiropoulou *et al.*, 1997; Saitoh *et al.*, 1998), expression of Cystatin S has mainly been reported in the salivary glands (Bobek *et al.*, 1991). However, Cystatin S expression has been detected in a colon tumour cDNA library as mentioned in Section 4.2.3 (Frigerio *et al.*, 1995). This begs the question as to how a salivary cystatin is involved in colon tumour progression. It is reasonable to state that tumour cells behave differently to the normal cells from which they were derived. Therefore, certain genes can be switched on or off in tumour cells that would not normally be expressed in their normal cell counterparts. It is possible that Cystatin S is switched on during the early stages of tumourigenesis and is then subsequently switched off again at a later stage as the characteristics of the tumour cells change to metastatic cells. This concept has already been mentioned in the last section in relation to MRP-1/CD9 and HIP, and may prove to be a common theme in tumour progression. This emphasises the need to analyse the expression of genes from as many stages of tumour progression as possible and not just between normal and tumour cells.

4.2.5.3 Serum Amyloid A (SAA)

SAA is an acute phase reactant and is found in the circulation as an apolipoprotein bound to HDL (High Density Lipoprotein) (apoSAA) (Strachan *et al.*,

1989). During an inflammatory response circulating levels of SAA increase by about 1000 fold, primarily by transcriptional induction (Lowell *et al.*, 1986). Thus, SAA is thought to have an important role in host defense during inflammation, however, the precise role which SAA plays in this process is still not clear. SAA is the serum precursor of amyloid A (AA) protein, which is the major component of the secondary amyloid plaques deposited in major organs as an occasional consequence of chronic inflammatory disease (Husebekk *et al.*, 1985). Although, SAA is primarily produced in the liver, widespread distribution of SAA mRNA has recently been described in a number of epithelial tissues including the stomach and the small and large intestines (Urieli-Shoval *et al.*, 1998).

The promoter of the SAA gene has multiple transcriptional regulatory elements. These include an NF- κ B DNA-binding element, two C/EBP DNA-binding elements and an SAA-activating sequence (SAS) (Ray & Ray, 1998). The NF- κ B and C/EBP DNA-binding elements are important for induction of SAA transcription in the liver (Ray *et al.*, 1995). Transcriptional induction of SAA in nonhepatic cells is believed to be largely regulated by the SAS element which has been shown to interact with a family of proteins termed SAS-binding factors (SAF) (Ray & Ray, 1998). Several other genes also possess the SAF sequence within their promoters and include several inflammation- or acute-phase responsive genes and *c-Myc* (Ray & Ray, 1998). In addition, the SAF proteins are likely to be activated by a phosphorylation event, which indicates that altered levels of kinases, phosphatases and their inhibitors during tumour development may alter the expression levels of a range of genes associated with metastasis. The involvement of protein phosphatases in the development of the metastatic phenotype is discussed further in Chapter 5.

SAA is a secretory protein and has been shown to interact with the extracellular matrix and to induce immune cell adhesion to the matrix (Preciado-Patt *et al.*, 1996). It is also implicated in playing a role in the migration of both mast cells (Olsson *et al.*, 1999) and phagocytes (Su *et al.*, 1999). Increased expression of SAA mRNA has also been described in response to p53-induced apoptosis (Polyak *et al.*,

1997). The potential role that SAA may have in metastasis is currently unclear but its interaction with the extracellular matrix may be significant.

4.2.5.4 p32 subunit of Splicing Factor 2 (SFp32)

The p32 protein was initially identified as a co-purified protein with pre-mRNA splicing factor SF2 from HeLa cells (Krainer *et al.*, 1991; Honore *et al.*, 1993). Cloning of an Hyaluronic Acid (HA)-binding protein found it to be identical to the p32 protein upon sequence analysis (Deb & Datta, 1996). Additional receptors for HA include CD44, RHAMM and ICAM-1 (Entwistle *et al.*, 1996). HA is a ubiquitous component of the extracellular matrix and occurs transiently in the cell nucleus and the cytoplasm (Entwistle *et al.*, 1996). HA acts as both an inhibitor and a stimulator of angiogenesis (Rooney *et al.*, 1995). In addition to interacting with HA, the p32 protein has been associated with a variety of molecules including the lamin B receptor (Simos & Georgatos, 1994), HIV Tat (Yu *et al.*, 1995) and Rev (Luo *et al.*, 1994; Tange *et al.*, 1996), herpes simplex virus 1 Orf-P protein (Bruni & Roizman, 1996), Epstein-Barr virus EBNA1 protein (Wang *et al.*, 1997) and adenovirus polypeptide V (Matthews & Russell, 1998).

Subcellular localisation studies on p32 have resulted in some controversy. It has most recently been reported to localise to the mitochondrial matrix (Muta *et al.*, 1997; Matthews & Russell, 1998) and the nucleus (Matthews & Russell, 1998). Two quite different roles for p32 have been proposed. The first is a role for p32 in maintaining mitochondrial oxidative phosphorylation (Muta *et al.*, 1997). A functional yeast homologue p30 was identified and p32 was found to be capable of suppressing the growth impairment in yeast cells with a p30 gene disruption (Muta *et al.*, 1997). However, a role for p32 in the regulation of Splicing Factor 2 (SF2) has also been described (Petersen-Mahrt *et al.*, 1999). The SF2 protein is a member of the SR family of splicing factors that enhance constitutive RNA splicing, but are also capable of regulating alternative splicing (Mayeda & Krainer, 1992; Manley & Tacke, 1996). The p32 protein is proposed to both prevent RNA binding and to block

phosphorylation of SF2 (Petersen-Mahrt *et al.*, 1999). The observed increase in expression of p32 in some of the metastatic cell lines examined here suggests a possible role for p32 in metastasis. This may involve changes in splicing patterns of target genes, as ratios of particular and general splicing factors play a role in alternative splicing (Bai *et al.*, 1999). Moreover, the phosphorylation status of SF2 affects its interaction with several protein targets in different ways (Xiao & Manley, 1998). Finally, the ability of p32 to bind HA may also be significant as increased levels of HA in colorectal cancer have been correlated with reduced survival (Ropponen *et al.*, 1998).

4.2.5.5 The ribosomal proteins

A total of 22% of genes expressed in colonic tissue are involved in the processes of gene and protein expression (Adams *et al.*, 1995). This includes a substantial portion of ribosomal proteins which are necessary for protein synthesis. The SAGE analysis has identified 43 ribosomal proteins among the most abundant transcripts in the cell lines SW480 and SW620 (see Chapter 3). Approximately, 20% of these ribosomal proteins showed altered levels of expression between the primary and metastatic cell lines and ranged from 2 to possibly 11 fold (see Table 4.4). Ribosomal proteins L12, L3, S19 and L28 showed decreased expression in the metastatic line relative to the primary colon tumour cell line based on SAGE tag abundances, while ribosomal proteins S28, S12, L10, HL23 homologue, phosphoprotein P0 and L35 all showed increased expression in the metastatic cell line relative to the primary tumour cell line. Ribosomal proteins form part of the ribonucleoparticle and therefore are an essential part of the protein translational machinery. However, evidence is gathering to suggest that a number of ribosomal proteins possess a second function, separate from their role in the particle (Wool, 1996).

A number of studies have previously reported altered expression of ribosomal proteins in relation to cancer progression (Ou *et al.*, 1987; Levy *et al.*, 1991; Pogue-

Geile *et al.*, 1991; Barnard *et al.*, 1992; Barnard *et al.*, 1993; Denis *et al.*, 1993). Barnard *et al.* (1993) identified increased expression of four ribosomal proteins in colon tumours in comparison to matched normal colonic mucosa. These included L18, L37, S6 and phosphoprotein P0, although overexpression of L18 and S6 did not correlate with Dukes' stage. Overexpression of L3 was identified in hepatocellular carcinoma (Ou *et al.*, 1987), while S3, S6, S8, S12, L5 and P0 were increased in both colonic polyps and tumours suggesting that overexpression of these ribosomal proteins occurred as an early event in tumourigenesis (Pogue-Geile *et al.*, 1991). Ribosomal protein S13 also showed overexpression in colon tumours and increased expression has been correlated with the proliferative rate of rat tumour cell lines (Denis *et al.*, 1993). An increase in expression of the human homologue of yeast ribosomal S28 was also described in colon tumours compared to normal tissue (Zhang *et al.*, 1997). Expression of the ribosomal proteins identified from the studies mentioned above was compared to the SAGE data. The SAGE analysis showed an approximately equal tag abundance for L18, S13 and S6 (see Chapter 3 Tables 3.2 & 3.3), while L37, S8 and S3 were not detected. Phosphoprotein P0, S12, L5¹ (Table 4.4) and S28 (Table 4.1) all showed over-expression in the metastatic cell relative to the primary tumour cell line, which also showed increased expression in tumour relative to normal tissue as mentioned above. Additional novel increases include L10, HL23 homologue and L35 (Table 4.4). The SAGE analysis also identified reduced expression of a number of ribosomal proteins in the metastatic cell line, notably L3, which was in contrast to its increased expression in hepatocellular carcinoma mentioned above. Other novel down-regulated ribosomal proteins identified here include L12, S19 and L28 (Table 4.4).

The majority of studies to date showed increased expression of a number of ribosomal proteins in tumour relative to normal cells. The SAGE analysis was based on comparisons of tumour to metastatic cells and showed both increased and decreased expression of ribosomal proteins. The increase in ribosomal protein

¹ L5 is not listed in either Table 4.4 or 4.5 as it showed tag abundances of 1/7 (tumour/metastatic cell line) and therefore may not prove to be a true difference in expression (see Section 4.4.1).

transcripts is thought to be a result of neoplasia and not a direct causal event in carcinogenesis (Pogue-Geile *et al.*, 1991). Thus, up-regulation of ribosomal proteins is a property of rapidly dividing cells and increased protein synthesis (Ou *et al.*, 1987; Pogue-Geile *et al.*, 1991; Barnard *et al.*, 1993). Discovery of alternative roles for specific ribosomal proteins may change this view. This is exemplified by ribosomal protein L5 which has been shown to form a complex with Mdm2 and p53 (Wool, 1996), both of which are associated with tumour progression (Sherbet & Lakshmi, 1997). In addition, L18 has recently been found to negatively regulate double-stranded RNA-activated protein kinase (PKR), which is necessary for regulation of protein synthesis initiation (Kumar *et al.*, 1999). Thus, ribosomal proteins are more than subunits in the ribonucleoparticle and may possess a variety of additional functions such as regulators of protein expression. The observation of both increased and decreased expression of several ribosomal proteins in the metastatic cell line may be significant, but a greater understanding of ribosomal proteins is required before a specific role in tumourigenesis can be inferred.

4.2.5.6 Other notable differentially expressed transcripts

A number of the genes listed in Table 4.5, if proved to be differentially expressed, possess functions that could contribute to metastasis. Some of these functions are mentioned briefly below. The cellular ligand of Annexin II, p11 (also known as S100 calcium-binding protein A10) may be decreased in expression in the metastatic cell line. The annexins are a family of calcium-dependent phospholipid-binding proteins and down-regulation of one of its members, Annexin VI, has been associated with melanoma metastasis (Francia *et al.*, 1996). The p11 ligand/S100A10 induces the dimerisation of Annexin II and is thought to function as a regulator of protein phosphorylation. In addition, Annexin II is a receptor for tPA (tissue Plasminogen Activator) and p11 is believed to play a key role in the stimulation of tPA-dependent plasminogen activation (Kassam *et al.*, 1998). This may result in alteration of proteolysis which is a necessary step in metastasis (Duffy, 1996).

Prohibitin, which may also be down-regulated in metastatic cells, was originally isolated as a candidate antiproliferating gene in rat liver cells (Sato *et al.*, 1992). Prohibitin is known to inhibit DNA synthesis (<http://www.ncbi.nlm.nih.gov/Schuler/Unigene/>) and has been proposed to be a tumour suppressor gene in breast cancer (Sato *et al.*, 1992).

The 47kDa subunit of translation initiation factor 3 (eIF3) may be increased in expression in metastatic cells. Translation initiation factor 3 is a multiprotein complex of approximately 600kDa and plays a central role in the initiation of protein translation (Asano *et al.*, 1997). The 47kDa subunit acts as a regulatory subunit of the 26S proteasome, which is involved in the ATP-dependent degradation of ubiquitinated proteins. Control of protein synthesis and protein turnover may prove to be important events in tumourigenesis.

4.2.5.7 Conclusions

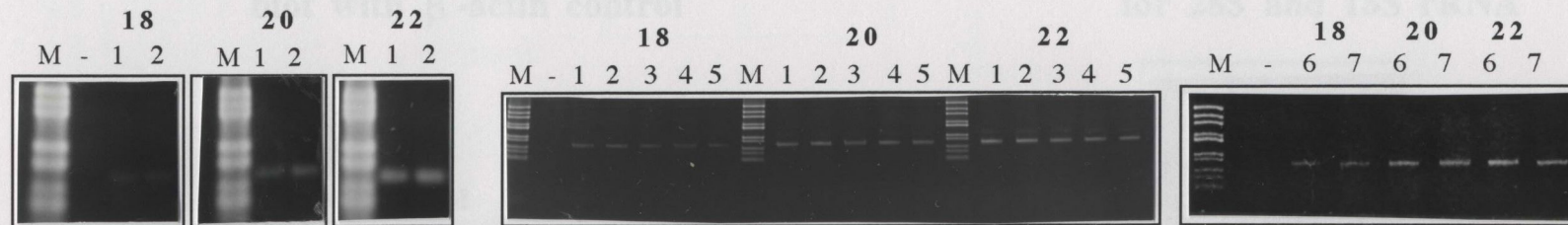
An analysis of a moderate number of SAGE tags (approximately 10,000 in total) has allowed an identification of 81 differentially expressed genes. However, as further analysis has shown, with a moderate level of SAGE tags, comparison of the gene expression profiles must be interpreted with caution as tag abundances <10 did not consistently represent true differences in expression. Thus, tag abundance differences <10 need to be verified by an independent method. The re-assessed SAGE data reduced the number of differentially expressed genes to 49. Of these, 5 were confirmed differences, 22 are likely differences and a further 22 are possibly differentially expressed. In addition, SAGE tag assignments for the most part are correct, but must be monitored carefully in the event of artefacts in the sequence databases or tag assignments to partial sequences including ESTs.

The five confirmed differentially expressed genes were examined in other metastatic cell lines and some showed a similar pattern of expression. The majority of differentially expressed transcripts possess functions which would suggest a possible role in colon tumour progression. A general expression pattern associated

with metastasis that seems to be emerging from the data includes alteration of RNA splicing, protein expression, the cytoskeleton and cell migration. Moreover, the possibility of expression levels changing in the same gene at different stages of tumourigenesis may prove to be significant and emphasises the importance of comparing gene expression patterns in matched samples from as many tumour stages as possible.

The analysis described here has provided a wealth of information on gene expression patterns in colon tumour metastatic cell lines. Data is provided on the most abundant transcripts and genes that are expressed at the same and different levels are described. This information now forms the basis for further study of colon tumour metastasis.

A. β -actin Semi-quantitative RT-PCR controls.



B. Genomic contamination test.

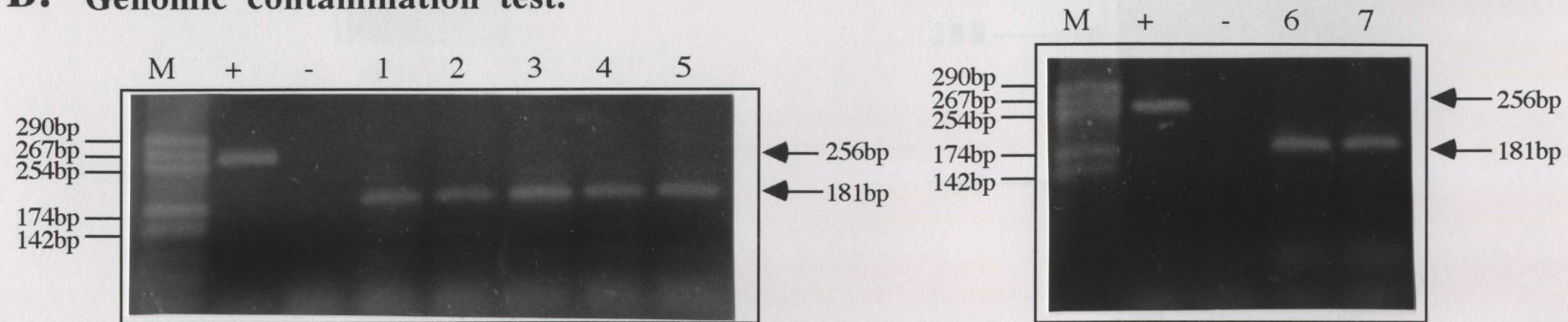
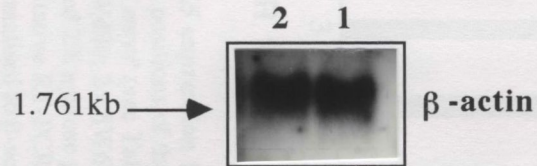


Figure 4.1 Controls for semi-quantitative RT-PCR and northern analysis. M= pBS-*Hae* III molecular weight marker and fragment sizes are described in base pairs (bp). - = negative control. The cell lines are numbered as follows: 1= SW480; 2= SW620; 3= LoVo; 4= Colo201; 5= T84; 6= SW1116; 7= 498LI. **A.** The semi-quantitative RT-PCR analyses were controlled by amplification of β -actin in parallel and PCR products (109bp) were resolved by electrophoresis on 2% agarose or 12% non-denaturing polyacrylamide gels and stained with ethidium bromide. A number of different PCR cycle numbers were sampled to ensure the exponential phase of the PCR reaction was being observed and are written above each panel. Similar β -actin levels between samples demonstrated an equivalent amount of input RNA. **B.** += positive control, i.e. genomic DNA. The presence of only the 181bp band in all of the reverse transcribed cDNAs from each cell line demonstrated that the samples were free from contaminating genomic DNA (256bp).

C. Hybridization of Northern blot with β -actin control



D. Estimation of distance migrated for 28S and 18S rRNA

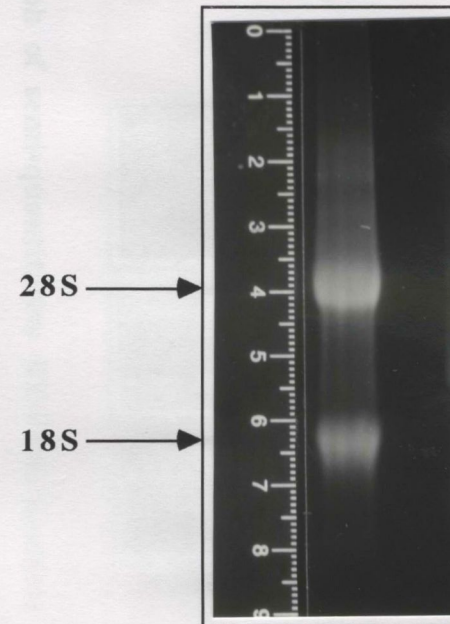
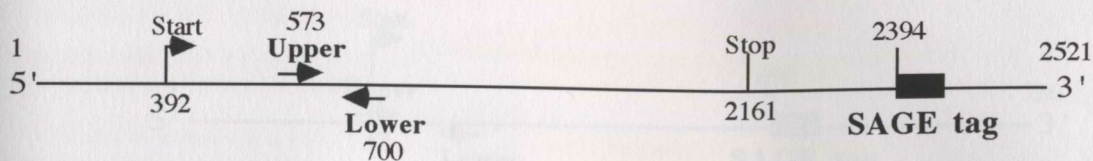
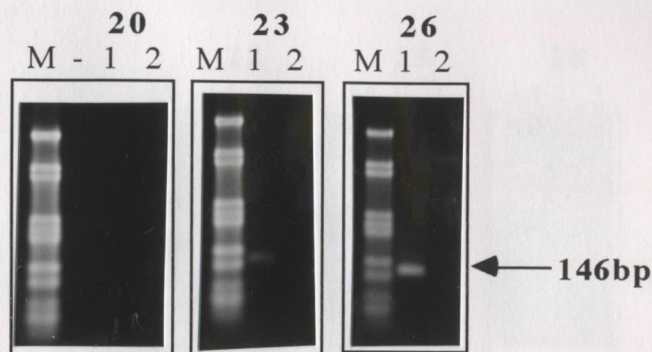


Figure 4.1 Controls for semi-quantitative RT-PCR and northern analysis (continued). C. The Northern blot of SW480=1 and SW620=2 RNA was hybridized with a β -actin cDNA probe. Equivalent signals were detected in both samples which indicated an equivalent amount of input RNA and efficiency of transfer. The same blot was used for all subsequent Northern blot hybridizations described in this chapter. D. Approximate transcript sizes were estimated by migration relative to 28S and 18S rRNA, which migrate as approximately 2 and 5kb bands respectively (Sambrook *et al.*, 1989). A sample of total RNA, which was resolved in parallel with the RNA for hybridization, was stained with ethidium bromide and the distance of the rRNA bands was estimated by comparison to a ruler.

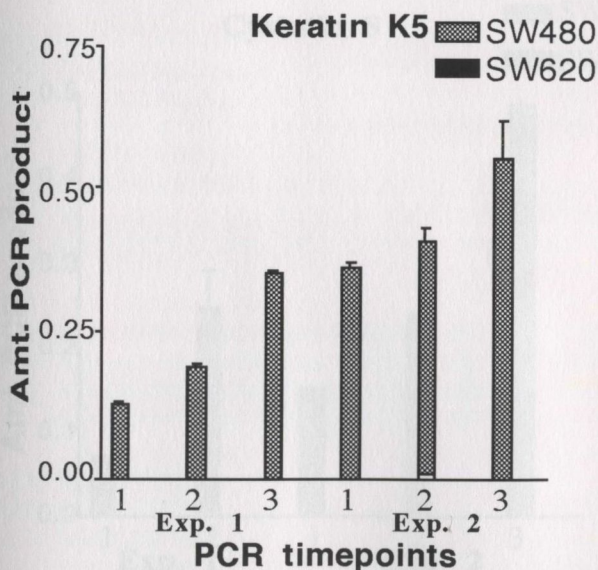
A. Keratin K5 mRNA



B. Semi-quantitative RT-PCR analysis



C. Graph of semi-quantitative RT-PCR analysis



D. Northern blot analysis

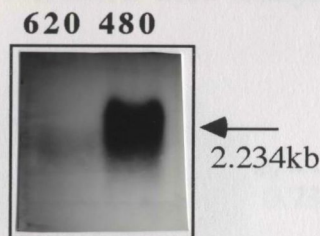
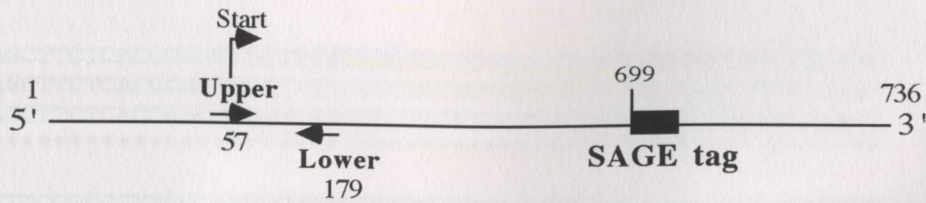
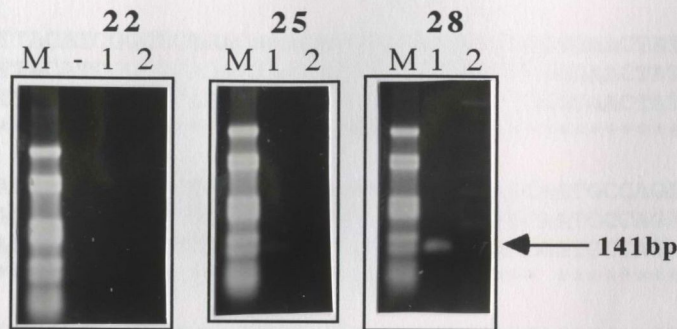


Figure 4.2 Keratin K5 expression in SW480 and SW620. **A.** An illustration of Keratin K5 mRNA transcript. The positions of the primers used in the semi-quantitative RT-PCR analysis are noted as 'Upper' and 'Lower' (see Table 2.1). **B.** M= pBS-*Hae* III molecular weight marker; -= negative control; 1= SW480; 2= SW620. Semi-quantitative RT-PCR products were visualised on ethidium bromide stained 2% agarose gels. The PCR cycle numbers sampled are shown above each panel. **C.** Semi-quantitative RT-PCR products were quantified on the ALFexpress™ automated sequencer and were normalised to β -actin levels. β -actin and Keratin K5 PCR products were loaded in the same lane in triplicate. The experiment was repeated several times over a range of cycle numbers and the results of two separate experiments (Exp. 1 & 2) are shown graphically above (GraphPad, Prism™, USA) and the data is given in Appendix 4.2. **D.** The Keratin K5 specific RT-PCR product was hybridized to a Northern blot of SW480 and SW620 and a transcript of expected size was detected. A high level of expression of Keratin K5 was detected in SW480 by the semi-quantitative RT-PCR analysis and northern blot. Expression in SW620 was not detected by semi-quantitative RT-PCR but a faint signal was detected on the Northern blot and this may be a homologue of Keratin K5.

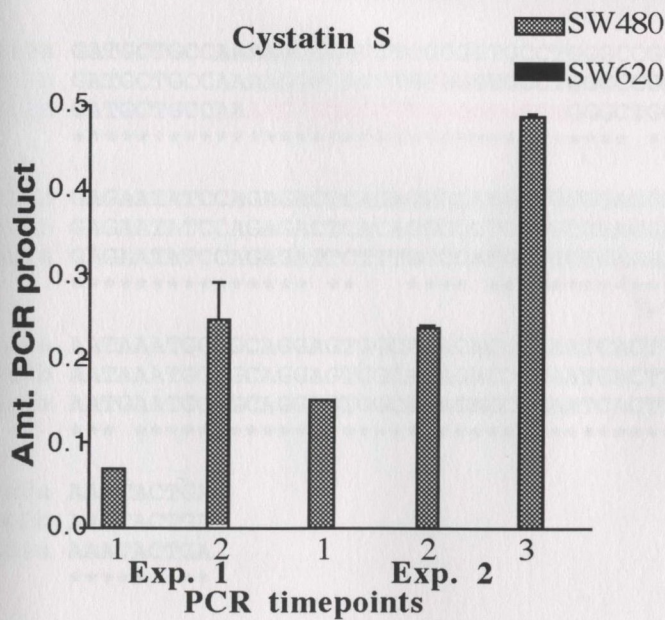
A. Cystatin S mRNA



B. Semi-quantitative RT-PCR analysis



C. Graph of semi-quantitative RT-PCR analysis



D. Northern blot analysis

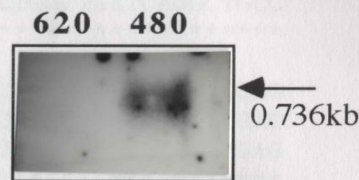


Figure 4.3 Cystatin S expression in SW480 and SW620. **A.** An illustration of Cystatin S mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. **B.** M= pBS-Hae III molecular weight marker; -= negative control; 1= SW480; 2= SW620. Cystatin S semi-quantitative RT-PCR products were visualised on an ethidium bromide stained 2% agarose gel and the cycle numbers sampled are shown above each panel. **C.** Semi-quantitative RT-PCR products were quantitated on the ALFexpress™ automated sequencer and normalised to β -actin. Cystatin S and β -actin PCR products were loaded in the same lane in triplicate. The results of repeated experiments over a range of cycle numbers are described in Appendix 4.3. The results of two experiments are shown above (GraphPad, Prism™, USA). **D.** The purified Cystatin S RT-PCR product hybridized to a single transcript of expected size as shown above. The northern blot and semi-quantitative RT-PCR analyses demonstrated a high level of expression of Cystatin S in SW480 and expression in SW620 was not detected by either method.


```

saa2a ATGAAGCTTCTCACGGGCCTGGTTTTCTGCTCCTTGGTCCTGAGTGTGAGCAGCCGAAGC
saa2b ATGAAGCTTCTCACGGGCCTGGTTTTCTGCTCCTTGGTCCTGAGTGTGAGCAGCCGAAGC
saa1a ATGAAGCTTCTCACGGGCCTGGTTTTCTGCTCCTTGGTCCTGGGTGTCAGCAGCCGAAGC
*****

saa2a TTCTTTTCGTTTCCTTGGCGAGGCTTTTGTATGGGGCTCGGGACATGTGGAGAGCCTACTCT
saa2b TTCTTTTCGTTTCCTTGGCGAGGCTTTTGTATGGGGCTCGGGACATGTGGAGAGCCTACTCT
saa1a TTCTTTTCGTTTCCTTGGCGAGGCTTTTGTATGGGGCTCGGGACATGTGGAGAGCCTACTCT
*****

saa2a GACATGAGAGAAGCCAATTACATCGGCTCAGACAAATACTTCCATGCTCGGGGGAACTAT
saa2b GACATGAGAGAAGCCAATTACATCGGCTCAGACAAATACTTCCATGCTCGGGGGAACTAT
saa1a GACATGAGAGAAGCCAATTACATCGGCTCAGACAAATACTTCCATGCTCGGGGGAACTAT
*****

saa2a GATGCTGCCAAAAGGGGACCTGGGGGTGCCTGGGCCGCAGAAAGTGATCAGCAATGCCAGA
saa2b GATGCTGCCAAAAGGGGACCTGGGGGTGCCTGGGCCGCAGAAAGTGATCAGCAATGCCAGA
saa1a GATGCTGCCAAAAGGGGACCTGGGGGTGTCTGGGCTGCAGAAAGCGATCAGCGATGCCAGA
*****

saa2a ATGAAGCTTCTCACGGGCCTGGTTTTCTGCTCCTTGGTCCTGAGTGTGAGCAGCCGAAGC
saa2b GACATGAGAGAAGCCAATTACATCGGCTCAGACAAATACTTCCATGCTCGGGGGAACTAT
saa1a GACATGAGAGAAGCCAATTACATCGGCTCAGACAAATACTTCCATGCTCGGGGGAACTAT
*****

saa2a GATGCTGCCAAAAGGGGACCTGGGGGTGCCTGGGCCGCAGAAAGTGATCAGCAATGCCAGA
saa2b GATGCTGCCAAAAGGGGACCTGGGGGTGCCTGGGCCGCAGAAAGTGATCAGCAATGCCAGA
saa1a GATGCTGCCAAAGGGGACCTGGGGGTGTCTGGGCTGCAGAAAGCGATCAGCGATGCCAGA
*****

saa2a GAGAATATCCAGAGACTCACAGGCATGGTGCGGAGGACTCGCTGGCCGATCAGGCTGCC
saa2b GAGAATATCCAGAGACTCACAGGCCGTGGTGCGGAGGACTCGCTGGCCGATCAGGCTGCC
saa1a GAGAATATCCAGAGATTCTTTGGCATGGTGCGGAGGACTCGCTGGCTGATCAGGCTGCC
*****

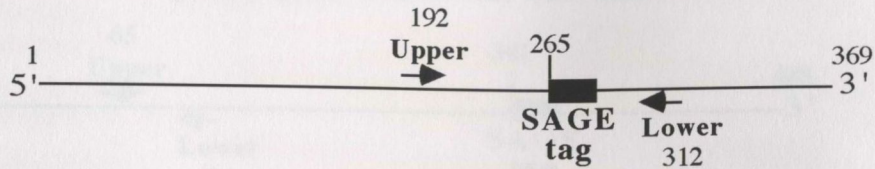
saa2a AATAAATGGGGCAGGAGTGGCAGAGACCCCAATCACTTCCGACCTGCTGGCCTGCCTGAG
saa2b AATAAATGGGGCAGGAGTGGCAGAGACCCCAATCACTTCCGACCTGCTGGCCTGCCTGAG
saa1a AATGAATGGGGCAGGAGTGGCAAGACCCCAATCACTTCCGACCTGCTGGCCTGCCTGAG
***

saa2a AAATACTGA
saa2b AAATACTGA
saa1a AAATACTGA
*****

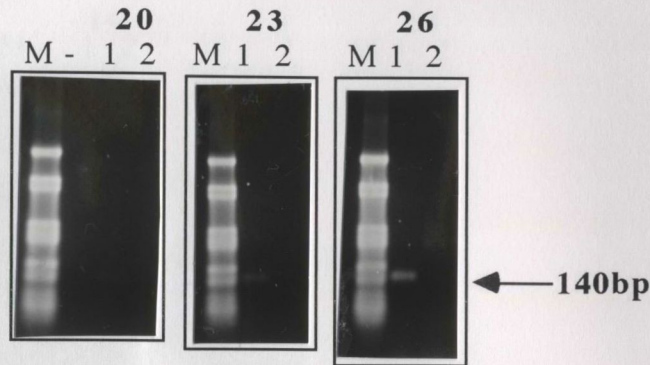
```

Figure 4.4 Multiple sequence alignment of Serum Amyloid A (SAA) cDNAs. The above alignment was generated using CLUSTAL W (1.7) provided by the Irish National Centre for Bioinformatics (<http://acer.gen.tcd.ie>). The tag from the SAGE analysis is shown in **green**. SAA2b does not contain the appropriate 'CATG' (the mismatch is shown in black) and therefore would not have been detected by the SAGE analysis. SAA1a and SAA2a contain identical SAGE tags and therefore would not have been distinguished in the SAGE analysis. The primers used in the semi-quantitative RT-PCR analysis are shown as **red (SAA.U)** and **blue (SAA.L)** and were specific for SAA1a but 19 out of 20 bases were shared with SAA2a and SAA2b.

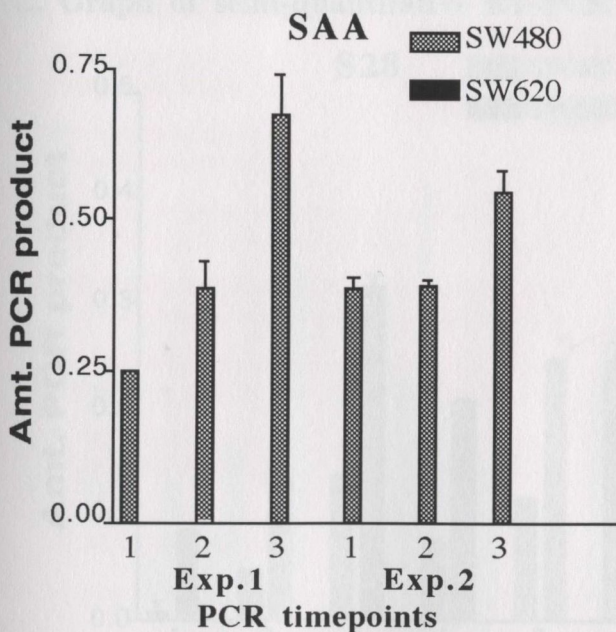
A. SAA mRNA



B. Semi-quantitative RT-PCR analysis



C. Graph of semi-quantitative RT-PCR analysis



D. Northern blot analysis

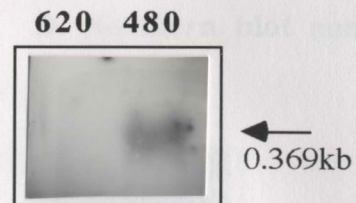
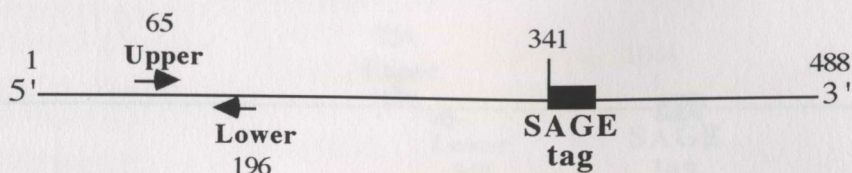
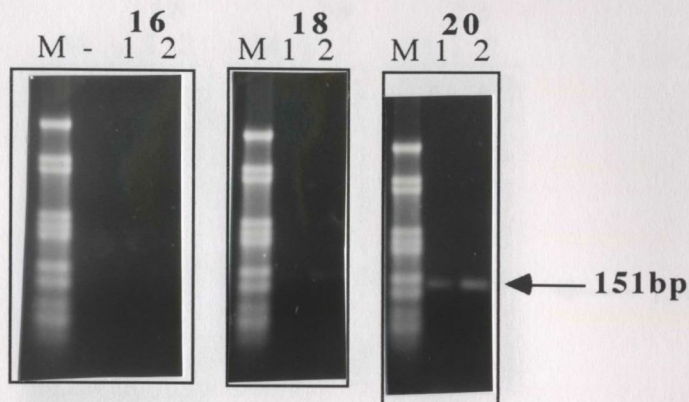


Figure 4.5 Expression of Serum Amyloid A (SAA) in SW480 and SW620. **A.** An illustration of SAA mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. **B.** M= pBS-Hae III molecular weight marker; -= negative control; 1= SW480; 2= SW620. SAA semi-quantitative RT-PCR products were resolved by electrophoresis on a 2% agarose gel and stained with ethidium bromide. The cycle numbers sampled are shown above each panel. **C.** The semi-quantitative RT-PCR products were quantified as relative amounts to β -actin using the ALFexpressTM automated sequencer and the data is described in Appendix 4.5. SAA and β -actin PCR products were loaded in the same lane in triplicate over a range of cycle numbers from different experiments. The data from two separate experiments (Exp. 1 & 2) are shown in the graph above (GraphPad, PrismTM, USA). **D.** Northern analysis with purified SAA RT-PCR product detected a single transcript of expected size in SW480 only. Semi-quantitative RT-PCR and Northern blot analyses showed a high level of expression of SAA in SW480 and expression in SW620 was not detected.

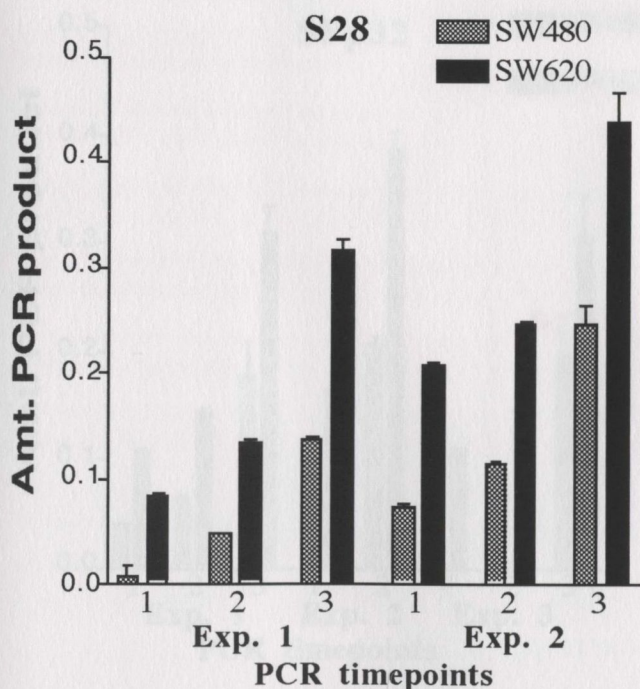
A. The human homologue of yeast ribosomal S28 mRNA



B. Semi-quantitative RT-PCR analysis



C. Graph of semi-quantitative RT-PCR analysis



D. Northern blot analysis

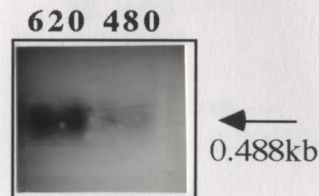
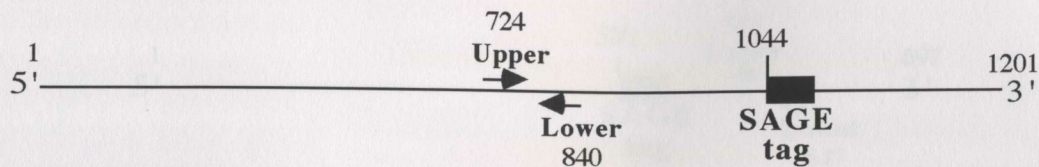
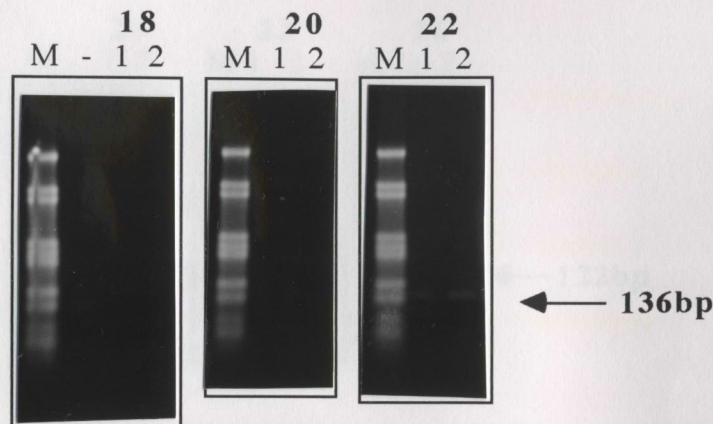


Figure 4.6 Expression of S28 in SW480 and SW620. **A.** An illustration of S28 mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. **B.** M= pBS-*Hae* III molecular weight marker; -= negative control; 1= SW480; 2= SW620. S28 semi-quantitative RT-PCR products were resolved by electrophoresis on a 2% agarose gel and stained with ethidium bromide. The cycle numbers sampled are shown above each panel. **C.** The semi-quantitative RT-PCR products were quantified as relative amounts to β -actin using the ALFexpressTM automated sequencer and the data is described in Appendix 4.6. S28 and β -actin PCR products were loaded in the same lane in triplicate over a range of cycle numbers from different experiments. The data from two separate experiments (Exp. 1 & 2) are shown in the graph above (GraphPad, PrismTM, USA). **D.** Northern analysis detected a single transcript of expected size in both cell lines with a higher level of expression in SW620 (>2 fold) relative to SW480. This agreed with the semi-quantitative RT-PCR analysis.

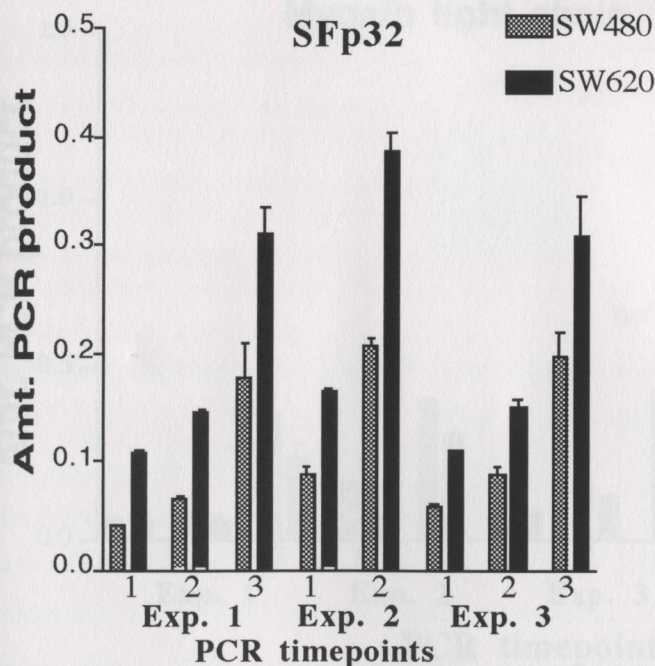
A. p32 subunit of Splicing Factor 2 (SFp32) mRNA



B. Semi-quantitative RT-PCR analysis



C. Graph of semi-quantitative RT-PCR analysis



D. Northern blot analysis

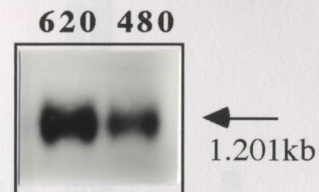
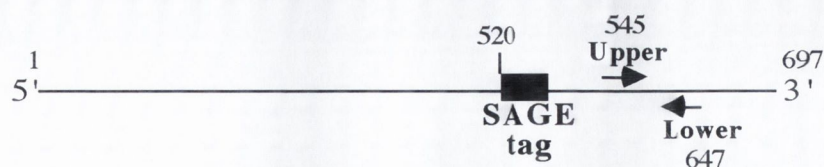
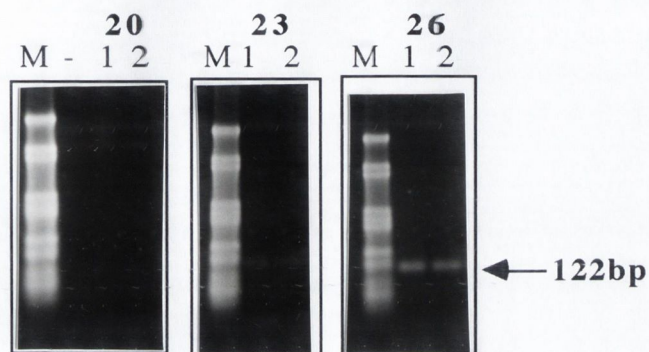


Figure 4.7 Expression of SFp32 in SW480 and SW620. **A.** An illustration of SFp32 mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. **B.** M= pBS-*Hae* III molecular weight marker; -= negative control; 1= SW480; 2= SW620. S28 semi-quantitative RT-PCR products were resolved by electrophoresis on a 2% agarose gel and stained with ethidium bromide. The cycle numbers sampled are shown above each panel. **C.** The semi-quantitative RT-PCR products were quantified as relative amounts to β -actin using the ALFexpressTM automated sequencer and the data is described in Appendix 4.7. SFp32 and β -actin PCR products were loaded in the same lane in triplicate over a range of cycle numbers from different experiments. The data from three separate experiments (Exp. 1, 2 & 3) are shown in the graph above (GraphPad, PrismTM, USA). **D.** Northern analysis detected a single transcript of expected size in both cell lines with a higher level of expression in SW620 (<2 fold) relative to SW480. This agreed with the semi-quantitative RT-PCR analysis.

A. Myosin light chain mRNA



B. Semi-quantitative RT-PCR analysis



C. Graph of semi-quantitative RT-PCR analysis

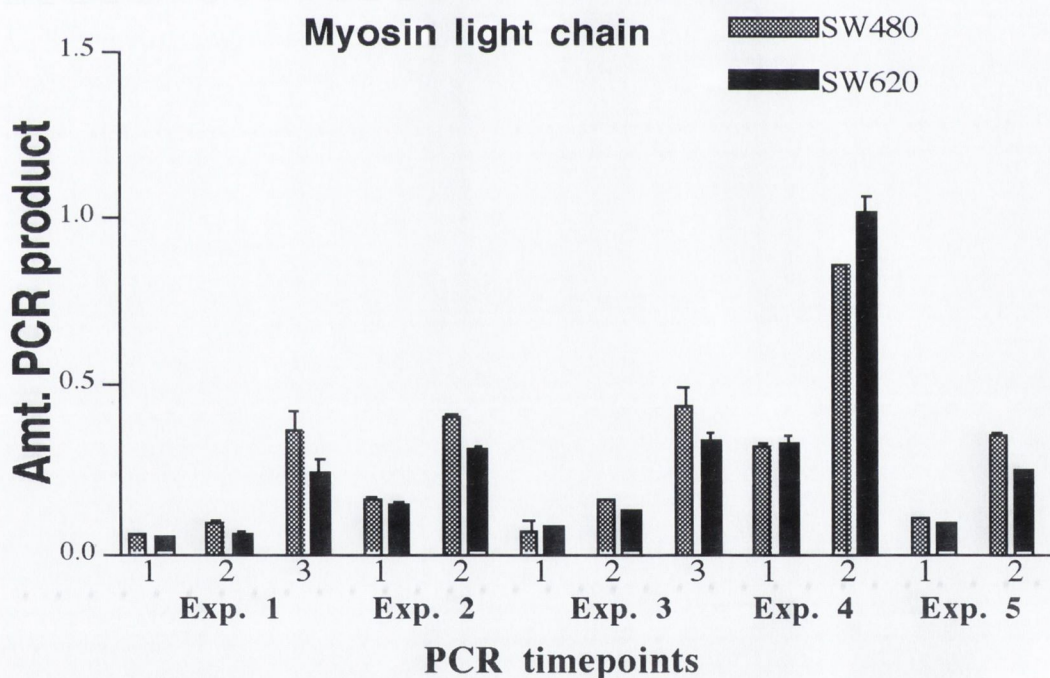
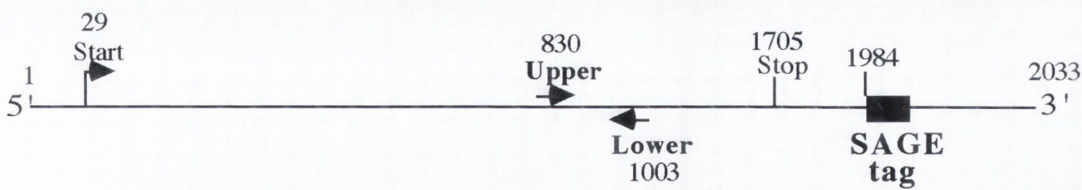
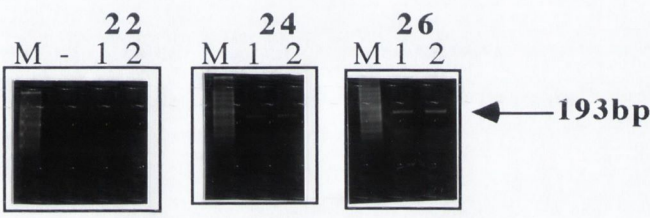


Figure 4.8 Expression of Myosin light chain in SW480 and SW620. **A.** An illustration of Myosin light chain mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. **B.** M= pBS-*Hae* III molecular weight marker; -= negative control; 1= SW480; 2= SW620. Myosin light chain semi-quantitative RT-PCR products were resolved by electrophoresis on a 2% agarose gel and stained with ethidium bromide. The cycle numbers sampled are shown above each panel. **C.** The semi-quantitative RT-PCR products were quantified as relative amounts to β -actin using the ALFexpressTM automated sequencer and the data is described in Appendix 4.8. Myosin light chain and β -actin PCR products were loaded in the same lane in triplicate over a range of cycle numbers from different experiments. The data from five separate experiments (Exp. 1, 2, 3, 4 & 5) are shown in the graph above (GraphPad, PrismTM, USA). Repeated experiments did not show a consistent difference in expression between SW480 and SW620.

A. hnRNP L mRNA



B. Semi-quantitative RT-PCR analysis



C. Northern blot analysis

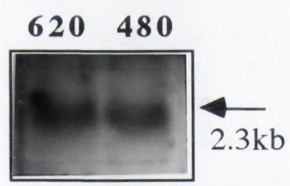
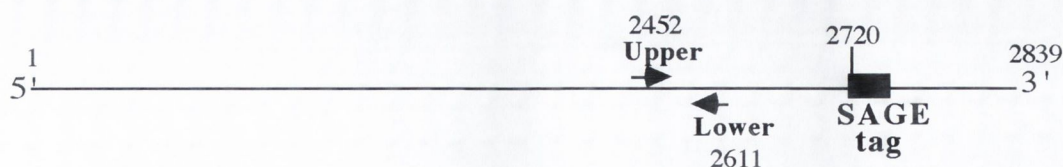
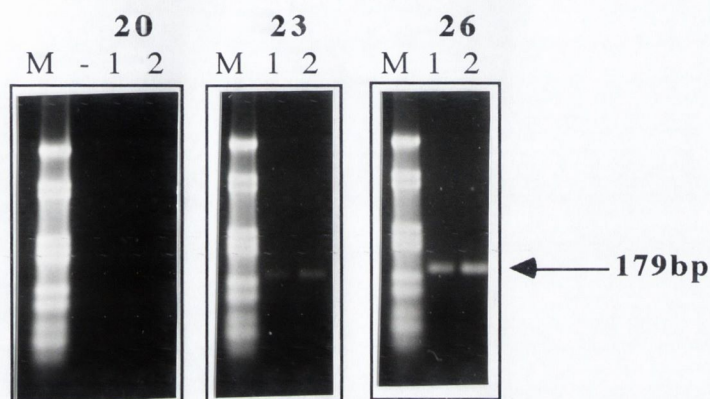


Figure 4.9 Expression of hnRNP L in SW480 and SW620. **A.** An illustration of hnRNP L mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. **B.** M= pBS-*Hae* III molecular weight marker; -= negative control; 1= SW480; 2= SW620. hnRNP L semi-quantitative RT-PCR products were resolved by electrophoresis on a 12% non-denaturing gel and stained with ethidium bromide. The cycle numbers sampled are shown above each panel. Repeated experiments did not show a definitive difference in expression between SW480 and SW620. **C.** Northern blot analysis detected a single transcript of approximately 2.3kb in size in both cell lines after hybridization with the purified hnRNP L RT-PCR product. The size was consistent with a previous report (Pinol-Roma *et al.*, 1989). Similarly to the semi-quantitative RT-PCR analysis, the northern blot analysis showed that hnRNP L was expressed at a similar level in SW480 and SW620.

A. SSRP1 mRNA



B. Semi-quantitative RT-PCR analysis



C. Graph of semi-quantitative RT-PCR

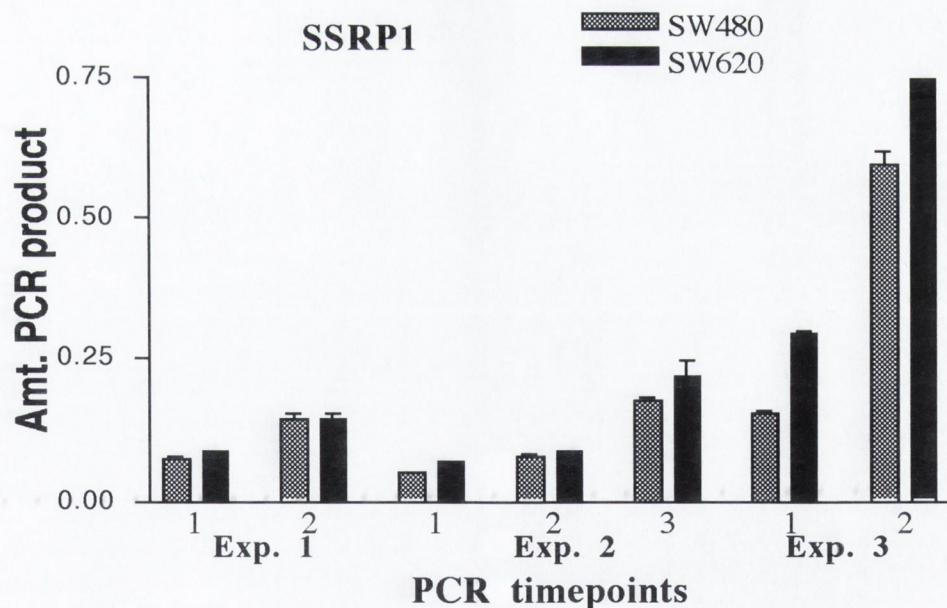
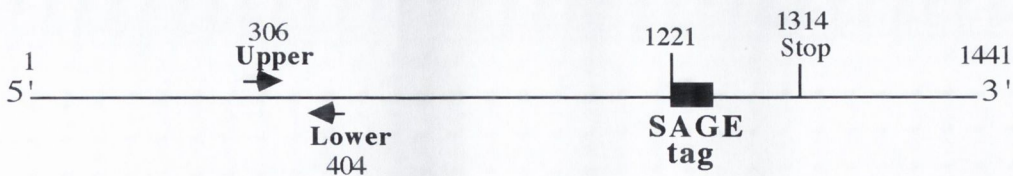
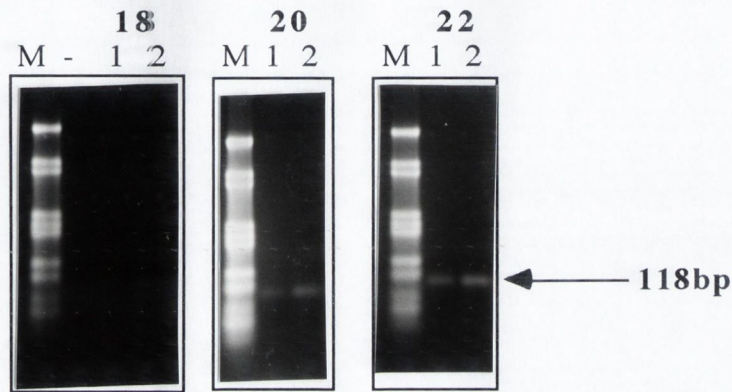


Figure 4.10 Expression of SSRP1 (High mobility group box protein) in SW480 and SW620. **A.** An illustration of SSRP1 mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. **B.** M= pBS-*Hae* III molecular weight marker; -= negative control; 1= SW480; 2= SW620. SSRP1 semi-quantitative RT-PCR products were resolved by electrophoresis on a 2% agarose gel and stained with ethidium bromide. The cycle numbers sampled are shown above each panel. **C.** The semi-quantitative RT-PCR products were quantified as relative amounts to β -actin using the ALFexpressTM automated sequencer and the data is described in Appendix 4.10. SSRP1 and β -actin PCR products were loaded in the same lane in triplicate over a range of cycle numbers from different experiments. The data from three separate experiments (Exp. 1, 2, & 3) are shown in the graph above (GraphPad, PrismTM, USA). Repeated experiments showed a slight increase in expression in SW620 which was in contrast to the SAGE results.

A. β -tubulin mRNA



B. Semi-quantitative RT-PCR analysis



C. Graph of semi-quantitative RT-PCR analysis

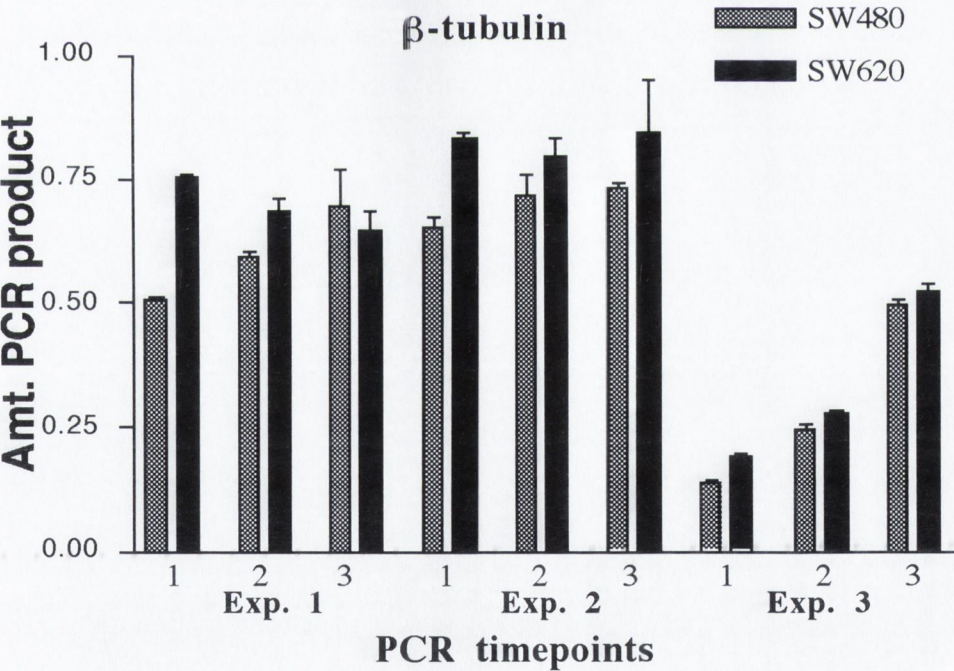
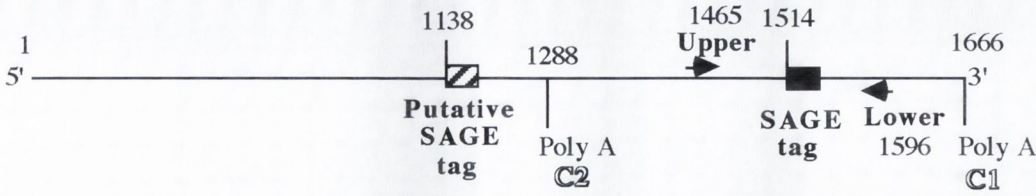
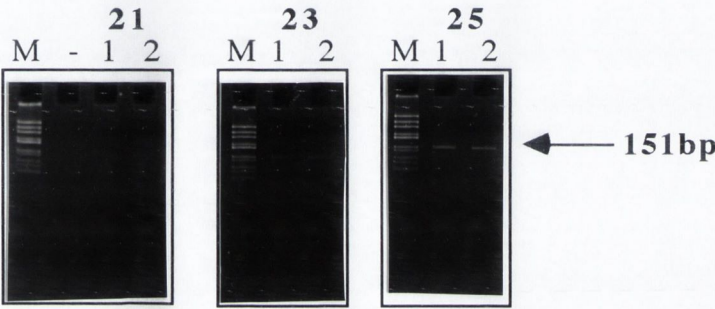


Figure 4.11 β -tubulin expression in SW480 and SW620. A. An illustration of β -tubulin mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. B. M= pBS-*Hae* III molecular weight marker; -= negative control; 1= SW480; 2= SW620. β -tubulin semi-quantitative RT-PCR products were resolved by electrophoresis on a 2% agarose gel and stained with ethidium bromide. The cycle numbers sampled are shown above each panel. C. The semi-quantitative RT-PCR products were quantified as relative amounts to β -actin using the ALFexpressTM automated sequencer and the data is described in Appendix 4.11. β -tubulin and β -actin PCR products were loaded in the same lane in triplicate over a range of cycle numbers from different experiments. The data from three separate experiments (Exp. 1, 2, & 3) are shown in the graph above (GraphPad, PrismTM, USA). Repeated experiments, which were normalised to β -actin levels, only showed a marginal increase in expression in SW620.

A. hnRNP C mRNA



B. Semi-quantitative RT-PCR analysis



C. Northern blot analysis



Figure 4.12 Expression of hnRNP C in SW480 and SW620. **A.** An illustration of hnRNP C mRNA. The positions of the primers used in the semi-quantitative RT-PCR analysis are shown as 'Upper' and 'Lower'. **B.** M= pBS-*Hae* III molecular weight marker; -= negative control; 1= SW480; 2= SW620. hnRNP C semi-quantitative RT-PCR products were resolved by electrophoresis on a 12% non-denaturing gel and stained with ethidium bromide. The cycle numbers sampled are shown above each panel. Repeated experiments did not show a definitive difference in expression between SW480 and SW620. **C.** Northern blot analysis detected a single transcript of approximately 5.0kb in size in both cell lines after hybridization with the purified hnRNP C RT-PCR product. The size expected was a transcript of 1.9kb as reported previously (Nakagawa *et al.*, 1986). Both the northern blot and the semi-quantitative RT-PCR analyses did not demonstrate differential expression of hnRNP C.

PACAP type-3/VIP type-2 receptor sequence

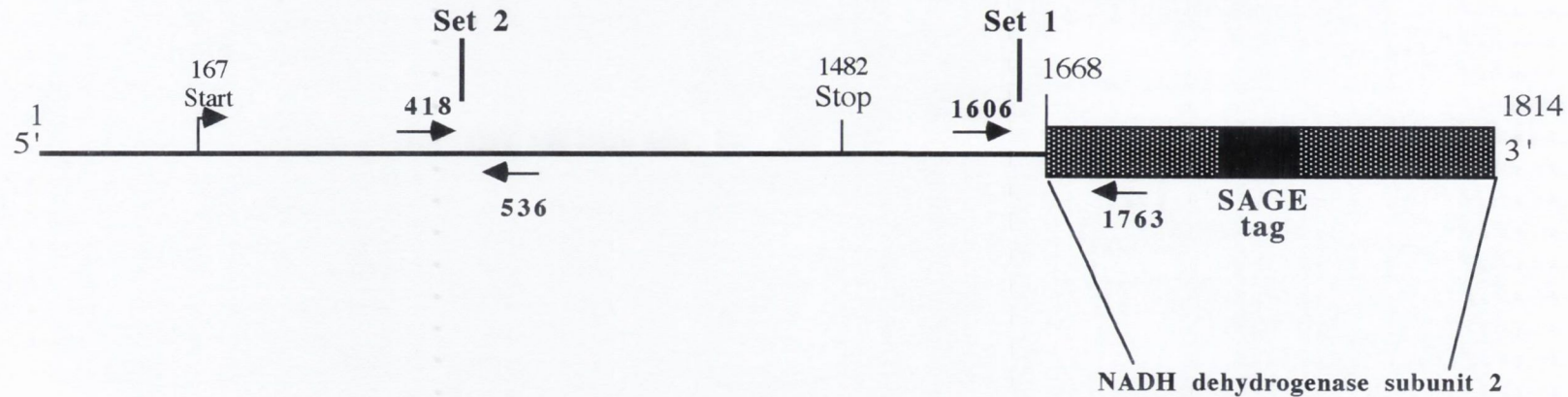


Figure 4.13 Illustration of PACAP type-3/VIP type-2 receptor sequence and region of identity with NADH dehydrogenase subunit 2 (N2). ■ = region which shares 100% identity with N2. ■ = location of SAGE tag sequence. The illustration shows the sequence region shared by PACAP type-3/VIP type-2 and N2 and is based on a BLAST (Altschul *et al.*, 1990) homology search which is detailed in Appendix 4.13. Two primer sets were designed and are shown above as Set 1 and Set 2. Neither set yielded specific PCR products of expected size based on the construct shown above. The sequence entry for PACAP type-3/VIP type-2 (U18810) may be an artefact.

Semi-quantitative RT-PCR analysis of Keratin K5

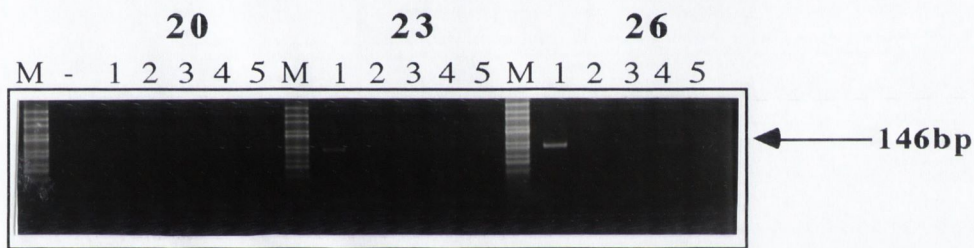


Figure 4.14 Expression of Keratin K5 in other metastatic cell lines. M= pBS-*Hae* III molecular weight marker; - = negative control. The cell lines are numbered as follows: 1= SW480; 2= SW620; 3= LoVo; 4= Colo201; 5= T84. Keratin K5 semi-quantitative RT-PCR products were visualised on ethidium bromide stained 12% non-denaturing polyacrylamide gels and the experiment was repeated at least three times. The PCR cycle numbers sampled are shown above each panel. Expression of Keratin K5 was not detected in any of the cell lines tested except SW480, even at high cycle numbers. Expression of Keratin K5 was also not detected in the second pair of isogenic cell lines, i.e. SW1116 and 498LI (data not shown).

Semi-quantitative RT-PCR analysis of Cystatin S

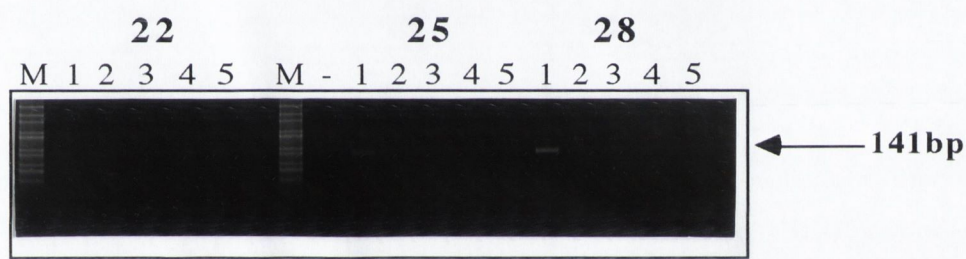
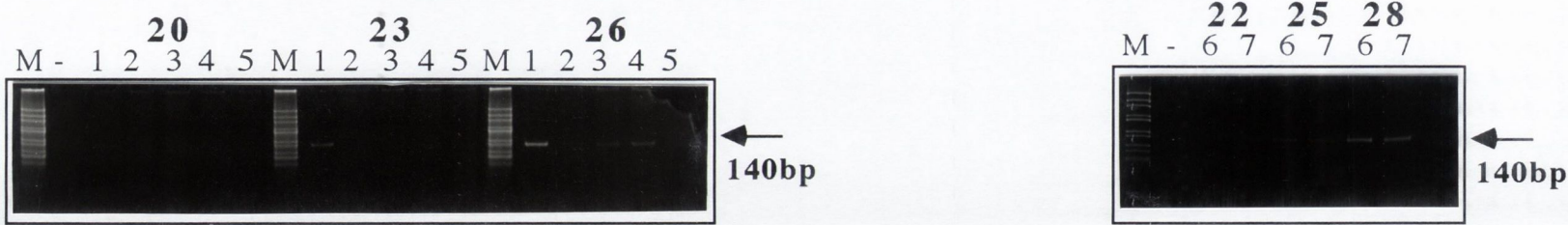


Figure 4.15 Expression of Cystatin S in other metastatic cell lines. M= pBS-*Hae* III molecular weight marker; - = negative control. The cell lines are numbered as follows: 1= SW480; 2= SW620; 3= LoVo; 4= Colo201; 5= T84. Keratin K5 semi-quantitative RT-PCR products were visualised on ethidium bromide stained 12% non-denaturing polyacrylamide gels and the experiment was repeated at least three times. The PCR cycle numbers sampled are shown above each panel. Expression of Cystatin S was not detected in any of the cell lines tested except SW480, even at high cycle numbers. Expression of Cystatin S was also not detected in the second pair of isogenic cell lines, i.e. SW1116 and 498LI (data not shown).

A. Semi-quantitative RT-PCR analysis of SAA



B. Graphs of semi-quantitative RT-PCR analysis

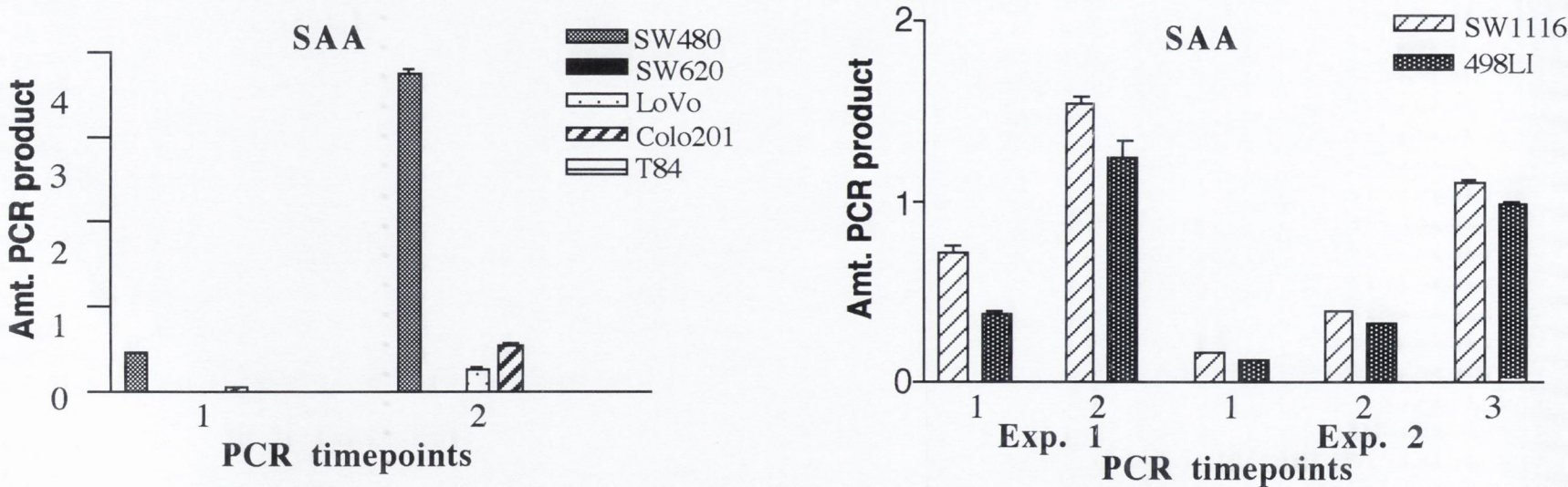
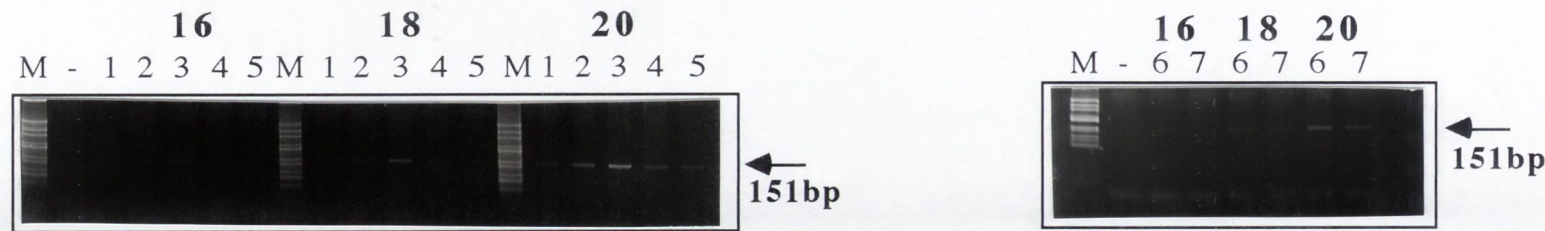


Figure 4.16 Expression of Serum Amyloid A (SAA) in other metastatic cell lines. A. M= pBS-*Hae* III molecular weight marker; - = negative control. The cell lines are numbered as follows: 1= SW480; 2= SW620; 3= LoVo; 4= Colo201; 5= T84; 6= SW1116; 7= 498LI. SAA semi-quantitative RT-PCR products were resolved by electrophoresis on 12% non-denaturing polyacrylamide gels and stained with ethidium bromide. The PCR cycle numbers sampled in each experiment are shown above each panel. B. SAA semi-quantitative RT-PCR products were quantified on the ALFexpress™ automated sequencer and normalised to β -actin levels. SAA and β -actin RT-PCR products were loaded in the same lane in triplicate and the ratios are described in Appendix 4.16. The results of two separate experiments (Exp. 1 & 2) are shown for SW1116 and 498LI. SAA expression was decreased in all metastatic cell lines examined and the fold differences are summarised in Table 4.3.

A. Semi-quantitative RT-PCR analysis of S28



B. Graphs of semi-quantitative RT-PCR analysis

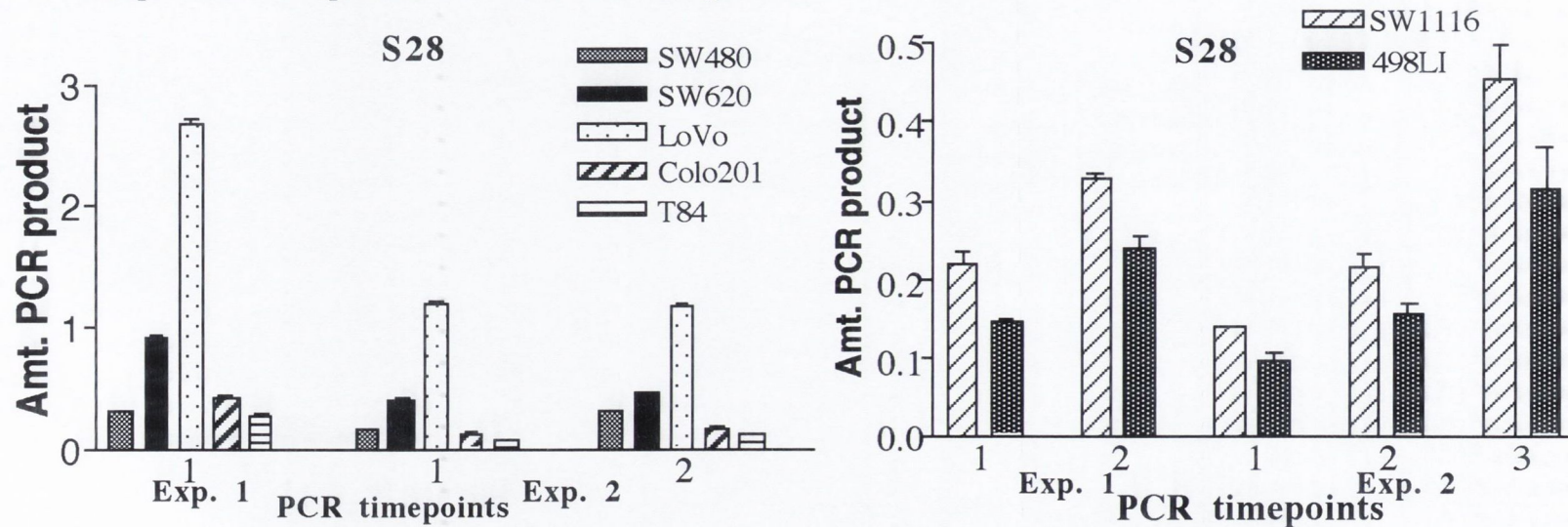
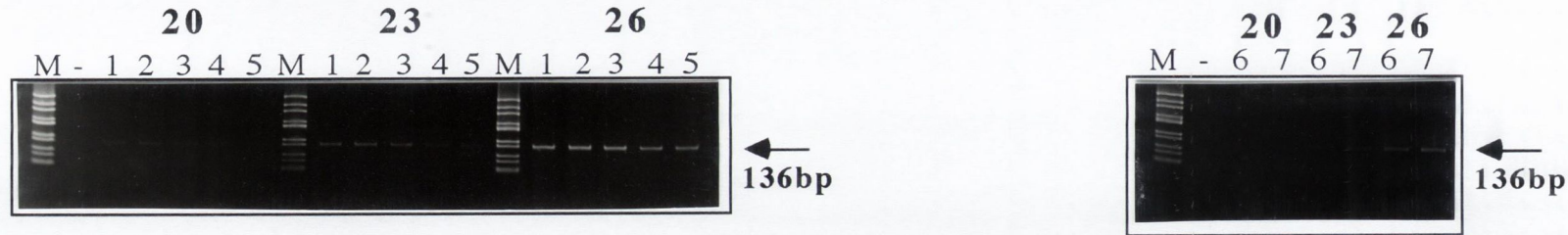


Figure 4.17 Expression of the human homologue of yeast ribosomal S28 (S28) in other metastatic cell lines. A. M= pBS-*Hae* III molecular weight marker; - = negative control. The cell lines are numbered as follows: 1= SW480; 2= SW620; 3= LoVo; 4= Colo201; 5= T84; 6= SW1116; 7= 498LI. S28 semi-quantitative RT-PCR products were resolved by electrophoresis on 12% non-denaturing polyacrylamide gels and stained with ethidium bromide. The PCR cycle numbers sampled in each experiment are shown above each panel. B. S28 semi-quantitative RT-PCR products were quantified on the ALFexpress™ automated sequencer and normalised to β -actin levels. S28 and β -actin RT-PCR products were loaded in the same lane in triplicate and the ratios are described in Appendix 4.17. The results of two separate experiments (Exp. 1 & 2) are shown for all cell lines. S28 showed increased expression in two of the five metastatic cell lines examined, i.e. SW620 and LoVo. Fold differences are summarised in Table 4.3.

A. Semi-quantitative RT-PCR analysis of SFp32



B. Graphs of semi-quantitative RT-PCR analysis

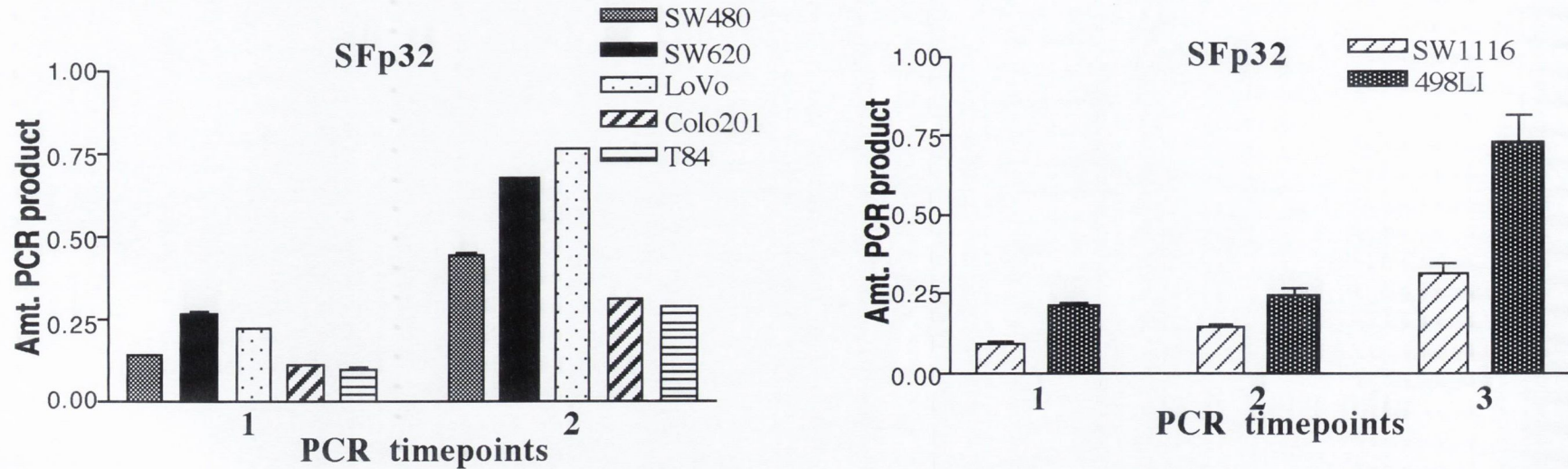


Figure 4.18 Expression of the p32 subunit of splicing factor (SFp32) in other metastatic cell lines. A. M= pBS-*Hae* III molecular weight marker; - = negative control. The cell lines are numbered as follows: 1= SW480; 2= SW620; 3= LoVo; 4= Colo201; 5= T84; 6= SW1116; 7= 498LI. SFp32 semi-quantitative RT-PCR products were resolved by electrophoresis on 12% non-denaturing polyacrylamide gels and stained with ethidium bromide. The PCR cycle numbers sampled in each experiment are shown above each panel. B. SFp32 semi-quantitative RT-PCR products were quantified on the ALFexpress™ automated sequencer and normalised to β -actin levels. SFp32 and β -actin RT-PCR products were loaded in the same lane in triplicate and the ratios are described in Appendix 4.18. SFp32 expression showed an increase in three of the five metastatic cell lines examined, i.e. SW620, LoVo and 498LI. Fold differences are summarised in Table 4.3.

Table 4.1 Subset of differentially expressed tags further analysed

SAGE Tag	T/M	Gene	Accession Number
ACTAACACC	3/18	PACAP type-3/VIP type-2 receptor	U18810
CTGTTGGTG	1/17	Human homologue of yeast ribosomal S28	D14530
GCCCCTGCT	18/1	Keratin type II K5	M21389
GTACACACA	15/1	Cystatin S	X54667
GTGCTGAAT	8/1	Myosin Light Chain mRNA	U02629
GTGCGGAGG	7/0	Serum Amyloid A	X51439
TACTCTTGG	0/7	hnRNPL	X16135
TGAGGCCAG	6/0	High mobility Group Box (SSRP1) mRNA	M86737
ATAGACATA	1/6	p32 subunit of Splicing Factor 2	M69039
AACGACCTC	1/6	β -tubulin	V00599
TCAAATGCA	0/5	hnRNP C	M16342
*TACCATCAA	32/26	Glyceraldehyde 3-phosphate dehydrogenase	J02642
*GCTTTATTT	9/8	β -actin	X00351

T/M is the ratio of tags from the primary tumour cell line SW480 (T) compared to the metastatic cell line SW620 (M). Genes written in bold were confirmed as differentially expressed by independent methods (see Section 4.1.2). * The 'housekeeping' genes GAPDH and β -actin showed similar tag abundances as expected.

Table 4.2 List of cell lines examined for altered expression of confirmed differentially expressed transcripts

Cell line	ATCC No.	Description	Reference
LoVo	CCL-229	Initiated from a metastatic tumour nodule in the left supraclavicular region of a patient with a histologically proven diagnosis of adenocarcinoma of the colon.	Drewinko <i>et al.</i> , 1976.
Colo 201	CCL-224	Isolated from ascitic fluid from a patient with adenocarcinoma of the colon.	Semple <i>et al.</i> , 1978.
T84	CCL-248	A transplantable human carcinoma cell line derived from a lung metastasis of a colon carcinoma. The original histological characteristics of the colon carcinoma were maintained throughout transplantation in nude mice.	American Type Culture Collection (ATCC).
SW1116	CCL-233	Derived from a Dukes' A/Grade II adenocarcinoma of the colon extending into the muscularis	Leibovitz <i>et al.</i> , 1976.
498LI	-	Established by serial subcapsular splenic injection of SW1116 cells and subsequent harvesting of hepatic nodules in nude mice.	Lieberman <i>et al.</i> , 1991 ¹ .

Cell lines written in **bold** are metastatic cell lines and SW1116 is a primary colon tumour cell line.

¹ 498LI was a kind gift of Dr. Meenhard Herlyn, Wistar Institute, PA, USA.

Table 4.3 Fold differences in expression between primary tumour and metastatic cell lines

Gene transcript	SW480	SW620	LoVo	Colo201	T84	SW1116	498LI
<i>Fold decreased expression in metastatic cell lines ^a</i>							
Keratin type II K5	1.0	nd	nd	nd	nd	nd	nd
Cystatin S	1.0	nd	nd	nd	nd	nd	nd
Serum Amyloid A	1.0	nd	16	8.6	nd	1.0	1.4
<i>Fold increased expression in metastatic cell lines ^b</i>							
p32 subunit of Splicing Factor 2	1.0	2.0	1.6	0.8	0.7	1.0	2.2
Human homologue of yeast ribosomal S28	1.0	2.2	6.2	0.9	0.6	1.0	0.7

Cell lines written in **bold** are metastatic cell lines and those written in plain text are the primary colon tumour cell lines. Fold differences were calculated based on β -actin ratios and relative to expression in primary tumour cell lines. The results are based on repeated experiments and are shown in Figures 4.14 to 4.18 . Cell lines SW620, LoVo, Colo201 and T84 are described relative to SW480, and 498LI is described relative to SW1116. nd, not detected.

^a Fold differences are described as decreased expression in the metastatic cell lines relative to the primary tumour cell lines.

^b Fold differences are described as increased expression in the metastatic cell lines relative to the primary tumour cell lines.

Table 4.4 Subset of SAGE tags which are potentially differentially expressed ^a

SAGE Tag	T/M	Gene	Accession Number
CTGGGTAA	25/11	S19 ribosomal protein	M81757
CAAACCATC	21/11	Keratin K18	X12883
AAGGTGGAG	19/10	ORF	X80822
TAGGTTGTC	18/7	Homologue of mouse p21	X64899
GCAGCCATC	16/8	L28 ribosomal protein	U14969
AAGATTGGT	13/4	Motility related protein-1 (MRP-1)	X60111
ACATCATCG	13/3	L12 ribosomal protein	L06505
GTTCCCTGG	13/6	Fau mRNA	X65923
GGACCACTG	11/1	L3 ribosomal protein	X73460
CTCAACATC	13/29	Po ribosomal phosphoprotein	M17885
CTAAGACTT	10/24	cDNA clone	D56531
TCAGATCTT	8/19	Scar protein	M22146
CGCCGCCGG	8/17	L35 ribosomal protein	U12465
TGATTTTAC	8/16	Autonomously replicating sequence	L08441
TGGGCAAAGCT	1/11	EST 1	N31199 ¹
ACCCTTGGC	4/16	EST 2	D52188
CTAGCCTCA	3/15	γ -actin	X04098
GCCGAGGAA	5/14	S12 ribosomal protein	X53505
GTGTAAACC	5/14	L10 ribosomal protein	L25899
GGCTGGGGG	4/12	Profilin	J03191
ACACAGCAA	1/10	EST 3	C03810
GGCAAGCCC	2/10	csa-19 cDNA clone	U12404

T/M is the ratio of tags from the primary tumour cell line (T) compared to the metastatic cell line (M). ^a Differential expression of the listed tags have not been confirmed but are likely to represent true differences in expression based on their tag abundances.

¹ Identified as HL23 ribosomal protein homologue using UniGene (<http://www.ncbi.nlm.nih.gov/UniGene/>)

Table 4.5 Subset of SAGE tags which are possibly differentially expressed ^a

SAGE Tag	T/M	Gene	Acc. No.
TCAACTTCT/GG	5/0	cDNA clone	N57419
AACTGCTTC	4/0	cDNA clone	W32468
AGCAGATCA	4/0	p11, cellular ligand of Annexin II	M38591
AGGAGCTGC/TG	4/0	NADH dehydrogenase-ubiquinone	U65579
ATTGGCTTA	3/0	Prohibitin mRNA	S85655
CATCTTCAC/CA	4/0	EST 4	177313
CCGGCGCGT/GT	4/0	unidentified	
GAGTGGCTA/TT	4/0	cDNA clone	N83778
GGCTGATGT/GG	4/0	Glycyl-tRNA synthetase	U09510
TCCTTCTCC	4/0	α -actinin	X15804
TGGGGTACC/TT	4/0	EST 5	F20308
TGTGTGTTT/GT	4/0	Histone H1(0)	X03473
TTGGGTTTT/CC	4/0	mRNA for KiAA0023 gene	D14689
AGGCATTGA/AA	0/6	Putatively transcribed sequence	Z20711
CTCAACAGC/AA	0/5	eIF3 47kDa subunit	U94855
GCCCCTCCG	0/5	EST 6	112496
GTGAGCCCA	0/5	EST 7	178726
TAGACTTAT	0/5	Aspartate aminotransferase mRNA	M63180
TCCGCGAGA/AG	0/5	EST 8	76286
AAAAGTGA/AA	0/4	mRNA for surface glycoprotein	Z50022
AAGCAGGGT	0/4	EST 9	15697
GCACCCGCC/TG	0/4	cDNA clone	T40544

T/M is the ratio of tags from the primary tumour cell line (T) compared to the metastatic cell line (M). Acc. No. refers to the Genbank accession number.

^a Differential expression of the listed tags have not been confirmed to date. All tag abundances are <10 and therefore these gene transcripts may or may not be differentially expressed and require verification by an independent method.

CHAPTER 5

Analysis of DDRT clone PM208

5.0 INTRODUCTION

A previous study was carried out by Dr. Peter McWilliam to identify differentially expressed gene products using the method of Differential Display Reverse Transcription-PCR (DDRT) (Liang & Pardee, 1992). This study was carried out prior to the SAGE analysis on the same pair of cell lines as described in Chapters 3 and 4, i.e. the primary colon tumour cell line SW480 and the isogenic metastatic cell line SW620. DDRT is a PCR-based method that allows the identification of differentially expressed genes between multiple different samples (Liang & Pardee, 1995). DDRT differs from SAGE in only identifying differences in expression whereas SAGE provides a gene expression profile from each sample of interest (described in Chapter 3). The DDRT method involves isolation of undegraded RNA, reverse transcription with a set of anchor primers to yield single stranded cDNAs, subsequent PCR amplification with the anchor primer and an upstream arbitrary primer. The PCR amplification step usually involves incorporation of a radioactive isotope (^{32}P or ^{33}P) and the PCR products are separated on a polyacrylamide gel and patterns compared between cells of different types. Any difference products identified are excised from the gel, reamplified, cloned and sequenced. The anchor primers are directed to the Poly(A)+ tail and are of the form $\text{d(T)}_{12}\{\text{ACG}\}\text{N}$ (where $\text{N}=\text{A, G, C or T}$). Thus, DDRT clones are biased to the extreme 3' end of the mRNA transcript. A schematic of the DDRT method is shown in Figure 5.1. By varying both the anchor and arbitrary primers, different subpopulations of cDNA products will be amplified (Liang & Pardee, 1992; Bauer *et al.*, 1993; Liang & Pardee, 1995; Dunican *et al.*, 1997; McWilliam *et al.*, unpublished observations). However, DDRT-PCR tends to generate a high false positive rate and a number of factors such as amount of starting material, primer concentration and contaminating genomic DNA all have to be carefully controlled (Liang & Pardee, 1995; McWilliam *et al.*, unpublished observations).

Several DDRT PCR products were isolated, cloned and sequenced (McWilliam *et al.*, 1999). One of these clones, PM208, showed increased expression

in the primary tumour cell line (SW480) relative to the metastatic cell line (SW620). Clone PM208 was sequenced in both directions and contains an insert size of 560bp. The sequence of this differentially expressed product has been deposited into Genbank and its designated accession number is Y13810 (McWilliam *et al.*, 1997). Differential expression was initially confirmed by semi-quantitative RT-PCR using the DDT primers and conditions described in Chapter 2. The DDRT analysis, isolation and sequencing of PM208 and subsequent confirmation of differential expression were all carried out by Dr. Peter McWilliam. Further analysis of clone PM208 included identification by extensive *in silico* analysis, northern blotting and examination of expression in a second pair of metastatic tumour cell lines. Expression of a related gene product was also analysed in SW480 and SW620 by semi-quantitative RT-PCR as described in Chapter 2, Section 2.3.7 and Chapter 4, Section 4.1.1.

5.1 RESULTS

5.1.1 Identification of DDRT clone PM208

An initial BLAST homology search of the PM208 sequence (Altschul *et al.*, 1990) revealed that it is >90% homologous to a pseudogene for Inhibitor 2 of Protein Phosphatase 1 (IPP-2) mapping to human chromosome 6 (X87344). The sequence of the expressed IPP-2 gene (Helps *et al.*, 1994; Sanseau *et al.*, 1994) did not show homology to PM208. However, multiple alignment between clone PM208, the published human IPP-2 cDNA (X78873) and the IPP-2 pseudogene revealed that the region of homology for PM208 lies 3' to the polyadenylation signal used in the transcription of IPP-2. In addition, two mRNA transcripts (approximately 2kb and 4kb in size), specific for IPP-2, had been detected on multiple-tissue Northern blot analysis (Sanseau *et al.*, 1994). The Genbank entry for the IPP-2 cDNA sequence appeared to originate from the smaller, 2kb transcript and therefore complete cDNA sequence data for the longer 4kb transcript was not available in the database. This raised the possibility that PM208 originated from the 4kb transcript of IPP-2 and this was subsequently confirmed by the following *in silico* and Northern blot analysis.

The BLAST search of PM208 also revealed that the sequence of clone PM208 overlapped three EST (Expressed Sequence Tag) clones (W58406; N52605; T65549). The latter EST clones were obtained from the I.M.A.G.E (Integrated Molecular Analysis of Genomes and their Expression) Consortium and were cycle sequenced in both directions using M13 Forward and Reverse primers as outlined in Chapter 2. These clones were also >90% homologous to the pseudogene for IPP-2. A number of subsequent BLAST searches, using the sequenced EST clones, revealed a series of overlapping sequences going in a 3' to 5' direction, all exhibiting >90% homology to the pseudogene for IPP-2. There was a gap (438bp) in the sequence between cDNA clone HSC0Ha041 (F08719) and the published IPP-2 cDNA sequence (Helps *et al.*, 1994; X78873). Primers were designed (GAPU & GAPL, Chapter 2 Table 2.1) to amplify, clone and sequence the missing cDNA ("GAP clone") and this was carried out by Ms. Orna Tighe. The overlapping sequences were assembled into a single

contig using the program 'Gel Assembly' (Wisconsin GCG package, Version 9.1) and this is illustrated in Figure 5.2. The contig sequence is shown in Appendix 5.2 and is a combination of sequences retrieved from Genbank¹, extended sequencing of the IMAGE clones and the sequence of clones PM208 and the "GAP" clone. The contig indicates that clone PM208 originated from the longer, 4kb IPP-2 transcript. The sequence of this additional 3' UTR sequence for IPP-2 has been deposited in Genbank (AJ133812).

To confirm the *in silico* analysis described above, total RNA was extracted from SW480 and SW620 cells and analysed by Northern blot as described in Chapter 2 Section 2.3.11. Two IPP-2 specific probes were hybridized to the blot, Ipp2-1 was specific for the "Gap" sequence, which overlaps the 3' untranslated region of the published 2kb IPP-2 cDNA (Helps *et al.*, 1994) and Ipp2-2 was specific for a sequence further downstream which overlaps DDRT clone PM208. The exact positions of the two probes are shown in Figure 5.2 and Appendix 5.2. The results of the Northern blot are shown in Figure 5.3. Equal loading and transfer was verified with the β -actin-1 probe as described in Chapter 2 Section 2.3.11. Level of expression was determined by densitometry (Bio1D v6.32 software) and described as a relative amount to β -actin expression and analysed using GraphPad (Prism™, USA) (Figure 5.4). Ipp2-1 hybridized to two transcripts of approximately 2kb and 4kb in size. Both transcripts showed increased expression in SW480 relative to SW620, demonstrating that both forms of IPP-2 mRNA are differentially expressed. However, the longer 4kb transcript showed a slightly larger difference in expression than the 2kb transcript, i.e. 2.8 fold compared to 2.3 fold (Figures 5.3 and 5.4). This agreed with previous semi-quantitative RT-PCR data with primer sets specific for PM208 (DDT primers Chapter 2 Table 2.1) and for the coding region of IPP-2 (nucleotides 294-420) (DDP primers Chapter 2 Table 2.1) (McWilliam *et al.*, In preparation). The Ipp2-2 probe hybridized to the 4kb transcript only, demonstrating

¹ Sequences were retrieved using either Entrez (<http://www3.ncbi.nlm.nih.gov/Entrez/>) or the ACNUC retrieval system provided by the Irish National Centre for Bioinformatics.

that DDRT clone PM208 represents the 4kb mRNA transcript of IPP-2. Moreover, the 4kb transcript was more abundant relative to the 2kb transcript within each cell line, i.e. 2:1 in SW480 and 1.8:1 in SW620.

In summary, clone PM208 was isolated as a confirmed differentially expressed DDRT product which showed decreased expression in the metastatic cell line (SW620) relative to the primary tumour cell line (SW480). Extensive sequence analysis has identified clone PM208 as the 4kb mRNA transcript of IPP-2 which was subsequently confirmed by Northern blot analysis. The analysis has also confirmed differential expression of the smaller, 2kb IPP-2 transcript.

5.1.2 Expression analysis of Hox11, an associated gene product

Inhibitor-2 (IPP-2) is a heat stable protein that acts as one of a number of regulatory subunits of Protein Phosphatase-1 (PP1) (Sanseau *et al.*, 1994). PP1 regulates a variety of cellular processes including cell cycle progression, cell metabolism, transcriptional regulation and neuronal function (Cohen, 1989 & 1997). A recent report has identified Hox11 as another possible inhibitor of PP1 and as an inhibitor of Protein Phosphatase 2A (PP2A) (Kawabe *et al.*, 1997). Hox11 is a homeobox-containing oncogene associated with T-cell leukaemias (Tang & Breitman, 1995). The interaction of Hox11 with PP1/PP2A is thought to disrupt a G2/M cell cycle checkpoint resulting in genomic instability and oncogenesis (Kawabe *et al.*, 1997). Expression of Hox11 was tested in the cell lines SW480 and SW620 to determine if another putative inhibitor of PP1 showed a difference in expression similar to IPP-2.

Expression of Hox11 was analysed by semi-quantitative RT-PCR. Experiments were performed at least three times using the conditions and primers (Hox11; primer positions 1006 and 1333 : M75952) (β -actin-L (254bp; see Table 2.2 for primer positions) described in Chapter 2 Section 2.3.7. PCR products, sampled from a number of timepoints, were resolved by electrophoresis on 12% polyacrylamide gels (Figure 5.5) and quantitated using the ALFexpress™ automated

sequencer as described in Chapter 2 Section 2.3.7. Expression levels were compared between samples as a ratio relative to the amount of β -actin PCR product and the data was analysed using GraphPad (Prism™, USA) (Figure 5.6). A 2-3 fold increase in expression of Hox11 was observed in the primary tumour cell line (SW480) relative to the metastatic cell line (SW620) (Parle-McDermott *et al.*, 1998). Thus, Hox11 showed a similar pattern of expression to IPP-2, i.e. increased expression in SW480. However, the RT-PCR data indicated that Hox11 is not a highly abundant transcript. High PCR cycle numbers (i.e. 29-35) and an increased amount of input cDNA were necessary to detect Hox11 even though the PCR conditions were optimal.

The Hox11 cDNA was cloned (pAPD.1) and sequenced to confirm identity, as outlined in Chapter 2. This was deemed necessary as extreme amplification conditions were necessary to detect Hox11 and the possibility of a non-specific PCR product had to be ruled out. Figure 5.7 shows a pairwise alignment of Hox11 (S38742) with the sequence of clone pAPD.1. A Hox11 specific probe was generated and hybridized to a Northern blot of total RNA from SW480 and SW620 as described in Chapter 2 Section 2.3.11. Hox11-specific mRNA was not detected on the blot even at reduced stringency (data not shown). This confirms that Hox11 is a low abundance transcript in SW480 and SW620 and thus can only be detected by RT-PCR.

In summary, two inhibitors of PP1, IPP-2 and Hox11, showed a decrease in expression in a metastatic cell line relative to a primary tumour cell line.

5.1.3 Expression of IPP-2 in SW1116 and 498LI

The expression of both transcripts of IPP-2 were examined in a second pair of primary colon and metastatic cell lines, SW1116 and 498LI respectively (described in Chapter 4 Table 4.2). Semi-quantitative RT-PCR was performed using primer sets DDT (PM208), DDP (ORF IPP-2) and β -actin-S under the conditions described in Chapter 2 Section 2.3.7. The PCR products were analysed as described above and the results are shown in Figure 5.8. Both sets of primers showed a decrease in expression

in the metastatic cell line 498LI relative to the primary colon tumour cell line SW1116. This is consistent with the expression pattern of IPP-2 in SW480 and SW620. Thus, both the 2kb and 4kb transcripts of IPP-2 show decreased expression in two separate pairs of metastatic cell lines relative to their primary tumour cell line counterparts.

5.2 DISCUSSION

The DDRT clone PM208 was identified as the 4kb transcript of inhibitor-2 (IPP-2) of Protein Phosphatase-1 (PP-1) by *in silico* and Northern blot analysis. IPP-2 is expressed as two transcripts of 2 and 4kb in size as reported previously (Sanseau et al., 1994). The Northern blot analysis showed that both forms of IPP-2 transcript were increased in expression in the primary colon tumour cell line (SW480) relative to the metastatic cell line (SW620). In addition, the 4kb transcript showed both a slightly increased fold difference in expression between the cell lines and a higher abundance within each cell line relative to the 2kb transcript. Moreover, decreased expression of both transcripts was observed in a second metastatic cell line (498LI) relative to the matched primary colon tumour cell line (SW1116).

5.2.1 Alternative polyadenylation signals

It is likely that the two transcripts for IPP-2 arise by use of alternative polyadenylation signals as they differ only at the 3' end and not in the coding region. During processing of pre-mRNAs in the nucleus, transcripts are cleaved at a specific site and approximately 200 adenosines are added at the 3' end (polyadenylation) by Poly (A) polymerase (Wahle & Keller, 1992). Polyadenylation signals are usually of the form AAUAAA, but variations of this sequence have been reported, and are found in the 3' UTR of the transcript. Cleavage of the transcript usually occurs approximately 11-23 nucleotides downstream of the hexanucleotide. In addition, a GU- or U-rich region is usually found 10-30 bases downstream of the cleavage site (Wahle & Keller, 1992; Edwalds-Gilbert *et al.*, 1997). Numerous genes have been reported to possess multiple Poly (A) sites within their 3' UTR (Edwalds-Gilbert *et al.*, 1997) similarly to IPP-2. Examination of the complete sequence of the 3' UTR of IPP-2 (Appendix 5.2) showed four potential Poly (A) signals within this region. The last poly (A) signal is 14 nucleotides away from the Poly (A) tail and therefore is the signal that is used for generation of the 4kb transcript. In addition, it also seems to be the preferred signal, at least in the tumour cell lines SW480 and SW620, based on the

Northern blot analysis. The Poly (A) signal closest to the 5' end of the UTR is probably used in the generation of the 2kb transcript based on the size difference of the two transcripts. The extra two Poly (A) signals indicates the possibility of additional transcripts. However, only the 2 and 4kb transcripts were detected on a multiple-tissue Northern blot (Sanseau *et al.*, 1994) and in the tumour cell lines SW480 and SW620.

The use of alternative tandem polyadenylation sites may have a regulatory function or may reflect the random use of signals with varying inherent strengths (Edwalds-Gilbert *et al.*, 1997). However, different forms of mRNA may exhibit greater stability or may be translated more efficiently. This is exemplified by the eukaryotic initiation factor 2 α (eIF-2 α) which contains multiple Poly (A) sites within the 3' UTR and produces two mRNAs of 1.6kb and 4.2kb. The 1.6kb transcript was found to be less stable than the 4.2kb transcript but was translated more efficiently. The ratio of the 1.6 to 4.2 kb mRNA ranged from 1:1 to 1:10 in different tissues and phases of the cell cycle. Thus, ratios favouring the smaller transcript produced more protein per transcript than ratios favouring the larger transcript. This was due to more efficient processing and transport of the more abundant smaller transcript, which was also translated more efficiently (Edwalds-Gilbert *et al.*, 1997). This suggests that the choice of Poly (A) site is not a random event but may be a response to the cellular environment through alteration of the polyadenylation machinery. Thus, in relation to the IPP-2 transcripts, the slight decrease in the ratio of the 4kb and 2kb mRNAs in the metastatic cell line SW620 may be significant and examination of the protein level of IPP-2 in these cell lines may prove interesting.

5.2.2 Roles for IPP-2 and Hox11 in metastasis

The decreased expression of IPP-2 and Hox11 in the metastatic cell line SW620 and of IPP-2 in a second metastatic cell line 498LI suggests a possible role for these genes in tumour progression. This may be through their role as inhibitors of Protein Phosphatase-1 (PP-1) (Sanseau *et al.*, 1994; Kawabe *et al.*, 1997). PP-1 is a

member of the Serine/Threonine phosphatase family which also includes PP2A, PP2B and PP2C (Barford, 1996). Reversible protein phosphorylation is an essential mechanism of signal transduction pathways in cells. This is achieved by the activities of a variety of protein kinases and phosphatases including the above mentioned Ser/Thr phosphatases (Barker *et al.*, 1994; Barford, 1996; Kwon *et al.*, 1997). In addition to IPP-2 and Hox11, PP-1 binds a wide range of regulatory subunits which includes p53BP2, kinesin-related protein, ribosomal proteins L5² and RIPP-1, RNA splicing factor, NIPP-1, inhibitor-1 and the retinoblastoma gene product (Barford, 1996). PP-1 and PP-2A have been implicated in a variety of roles including glycogen metabolism, muscle contraction, control of the cell cycle, RNA splicing (Barford, 1996) and also in the regulation of cytoskeletal organisation (Maier *et al.*, 1995; Shinoki *et al.*, 1995). The balance between the Serine/Threonine cAMP-dependent protein kinases (PKA) and PP-1/2A was found to influence the metastatic characteristics of Lewis Lung carcinoma cells by affecting the phosphorylation status of the cytoskeletal structures (Young *et al.*, 1993; Maier *et al.*, 1995). Inhibition of PP-1/2A by okadaic acid in non-metastatic cells resulted in a profound effect on both the cellular morphology and the distribution of cytoskeletal elements which resulted in increased motility and invasiveness (Maier *et al.*, 1995). This suggested that increased phosphorylation via PP-1 and PP2A is a feature of metastatic cells. This is further strengthened by the observation of increased expression of B56 γ , a regulatory subunit of PP2A, in metastatic melanoma cell lines relative to normal cells (Francia *et al.*, 1997). In addition, the B56 subunits of PP2A have recently been associated with regulation of β -catenin degradation in the Wnt signaling complex (Seeling *et al.*, 1999) (see Chapter 1). In contrast to these observations, the inhibitors IPP-2 and Hox11 both showed decreased expression in the metastatic cells, with a presumed increase in PP-1/2A activity. However, IPP-2 has recently been reported to function as a molecular chaperone and therefore plays a role in the folding of PP-1 protein rather than acting as an inhibitor (MacKintosh *et al.*, 1997). Thus, reduced expression

² Interestingly, ribosomal protein L5 may be increased in the metastatic cell line SW620 relative to the primary tumour cell line SW480 (see Chapter 4 Section 4.2.5.5).

of IPP-2 might not necessarily result in increased PP-1 activity and may result in misfolded PP-1. Expression of PP-1 is currently being examined in both pairs of metastatic colon tumour cell lines. The interaction of HOX11 with PP-1/2A has only been shown in *Xenopus* oocytes (Kawabe *et al.*, 1997) but its expression was examined in the tumour cell lines as Hox11 is a homeobox-containing oncogene associated with specific T-cell leukemias (Lichty *et al.*, 1995; Tang & Breitman, 1995). However, Hox11 is not normally expressed in adult tissues at levels detectable by routine Northern analysis (Dube *et al.*, 1991; Hatano *et al.*, 1991; Kennedy *et al.*, 1991) and its abundance in the cell lines SW480 and SW620 was also rare (see Section 5.1.2). Thus, the effect of HOX11 (if any) on PP-1/2A activity in these tumour cells is likely to be small and therefore the 2 to 3 fold decrease of this rare transcript in metastatic cells is not likely to be relevant.

5.2.3 Conclusions

The analysis of DDRT clone PM208 has clearly demonstrated the benefits of utilising the sequence databases. The approach described in Section 5.1.1 did not require cloning and sequencing of the full length 4kb cDNA. Instead, limited sequencing and cloning combined with extraction of the appropriate Genbank and dbEST entries allowed a sequence contig to be constructed which identified PM208 as the 4kb transcript of IPP-2.

IPP-2 and Hox11 interact with PP-1 and both showed decreased expression in metastatic cells but IPP-2 is by far the most interesting. Hox11 was difficult to detect and its interaction with human phosphatases has not been demonstrated to date. IPP-2 possesses multiple polyadenylation signals, two of which are used in the tumour cell lines. The association of IPP-2 with PP-1 and the possibility of altered stability and translatability of different forms of IPP-2 transcripts make it an interesting candidate for involvement in metastasis.

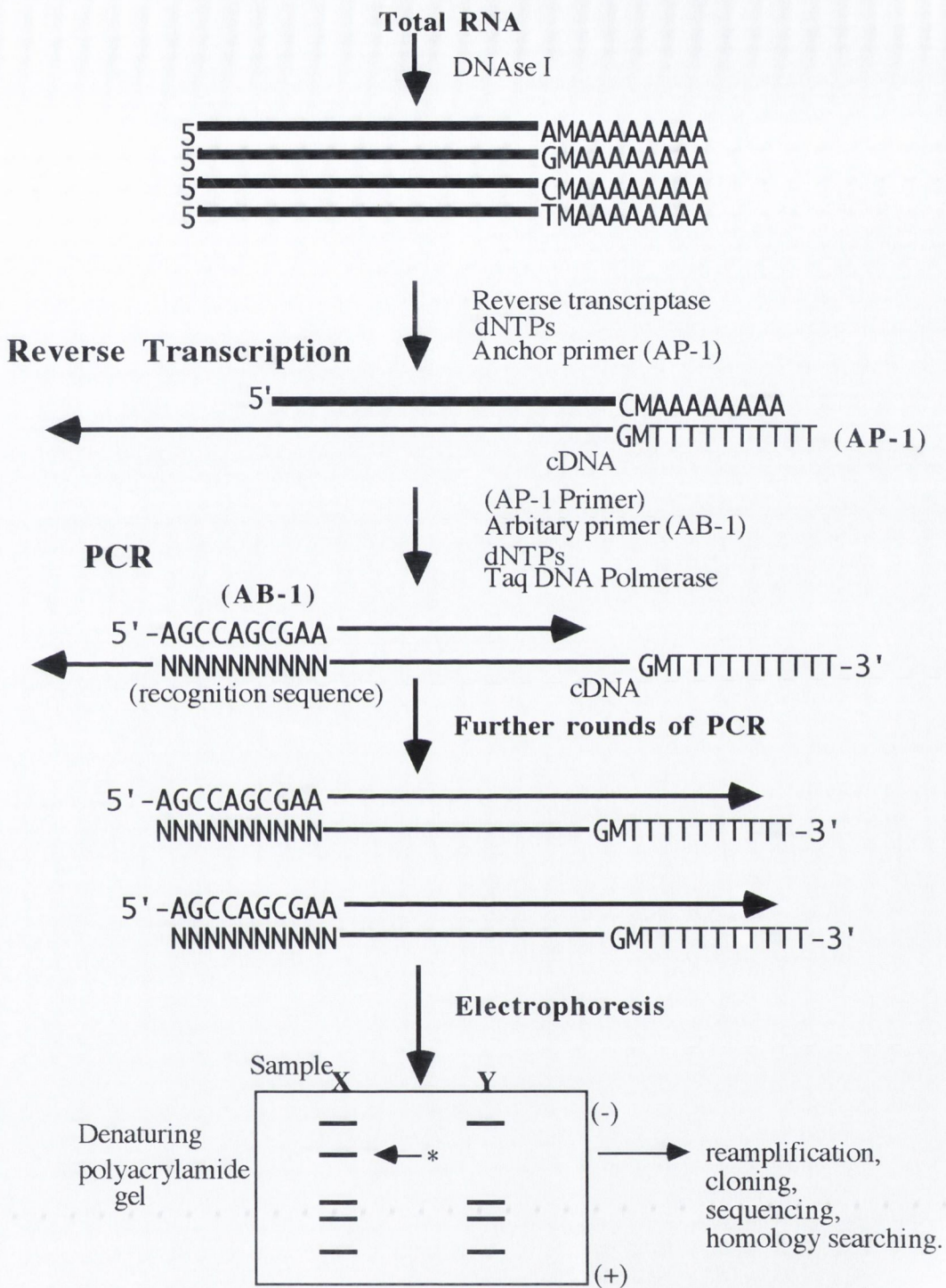


Figure 5.1 The method of Differential Display Reverse Transcription-PCR (DDRT) (based on Liang & Pardee, 1995). Total RNA is extracted from cells or tissues of interest, DNAase I treated to remove potentially contaminating chromosomal DNA and a subset of the mRNAs are reverse transcribed using an anchored oligo (dT) primer. A subset of the cDNAs are then amplified by PCR using an upstream primer of arbitrary sequence and the original anchor primer. Amplified PCR products from two or more samples are 'displayed' by electrophoresis in parallel on denaturing polyacrylamide gels. A radiolabelled dNTP is usually included in the PCR amplification step to allow visualisation by autoradiography. * Putative differentially expressed cDNA bands are excised, remplified, cloned and sequenced.

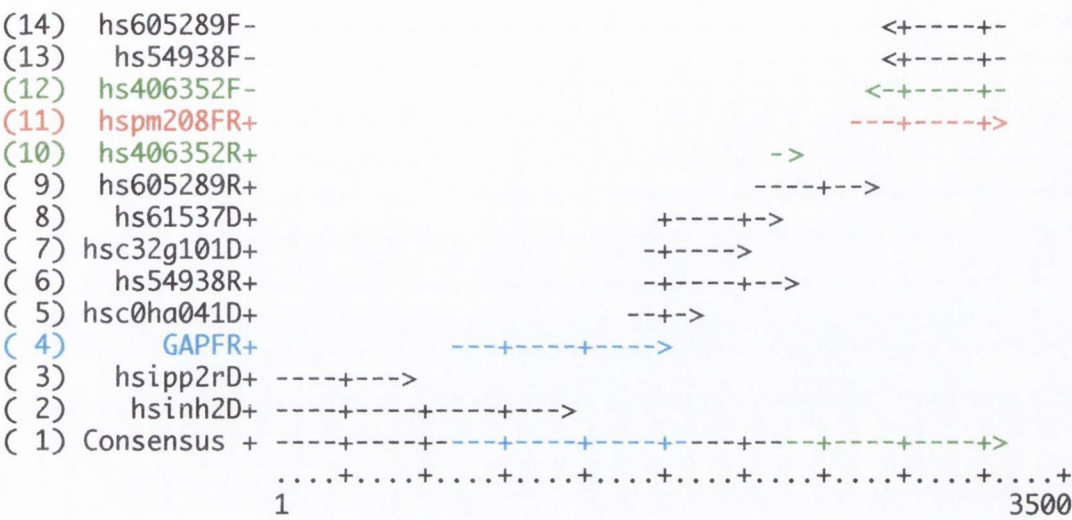


Figure 5.2 Contig construct of extended IPP-2 3' UTR. The above sequence contig was assembled using the Wisconsin GCG package, Version 9.1. The complete sequence is described in Appendix 5.2 and has been submitted to Genbank and assigned the accession number AJ133812. The contig consists of sequences from ESTs (Expressed Sequence Tags), IPP-2 submitted entries and clones PM208 and GAP. Clones were sequenced with M13 Forward (F) and Reverse (R) primers or the sequences were extracted directly from Genbank (D). The sequences are designated as F, R or D as appropriate. The contig clearly shows that DDRT clone **PM208 (hsmp208FR)** (shown in red) originated from a longer IPP-2 sequence that extends beyond the published sequences for IPP-2 (hsipp2rD and hsinh2D). Ipp-2 specific probes used in the Northern blot are highlighted in **cyan (Ipp2-1)** and **green (Ipp2-2)**.

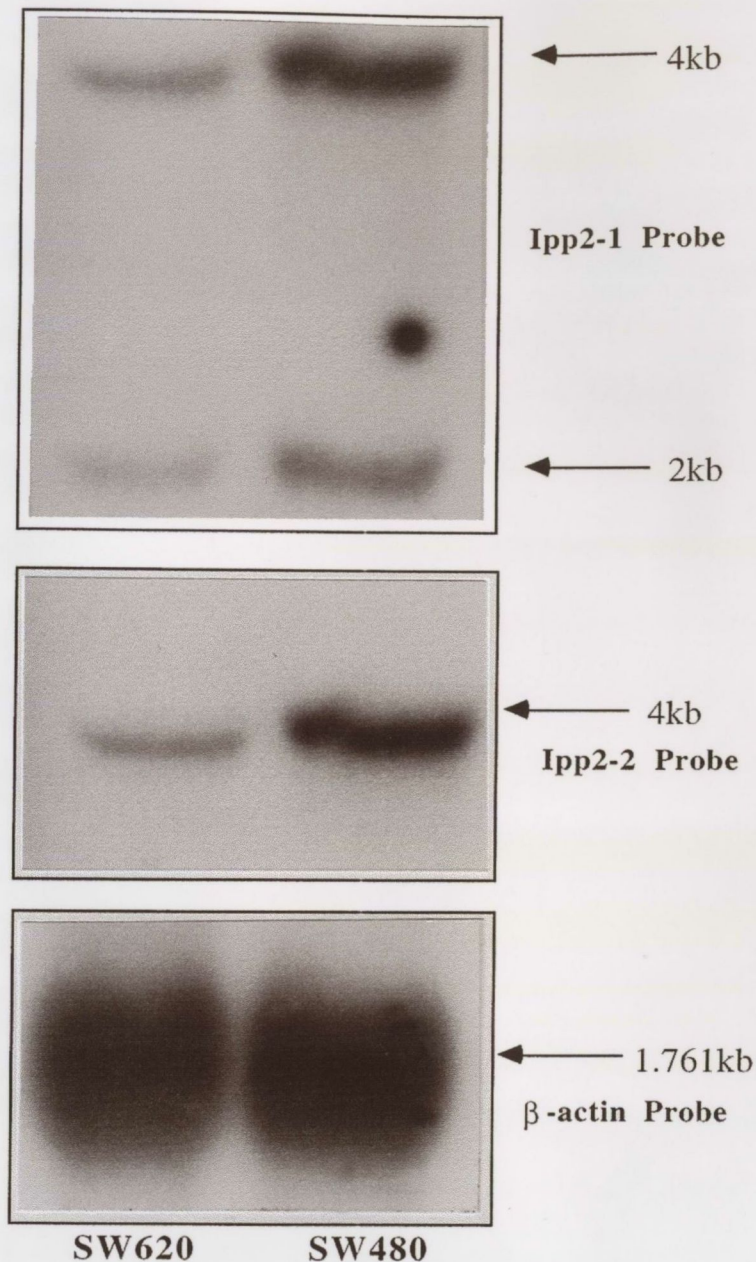


Figure 5.3 Northern blot analysis of IPP-2. A northern blot of SW480 and SW620 was hybridized with three probes: two *Ipp-2* specific probes (see Figure 5.2 and Appendix 5.2) and a β -actin specific probe. Ipp2-1 hybridized to two transcripts of approximately 2 and 4kb in size. Ipp2-2 hybridized to the 4kb transcript only. These results confirmed the *in silico* analysis and proved that the DDRT clone PM208 originated from the 4kb transcript of IPP-2. The results also confirmed decreased expression of both transcripts in SW620. The control probe β -actin showed an approximately equal level of expression which confirmed equivalent amounts of input RNA and efficiency of transfer.

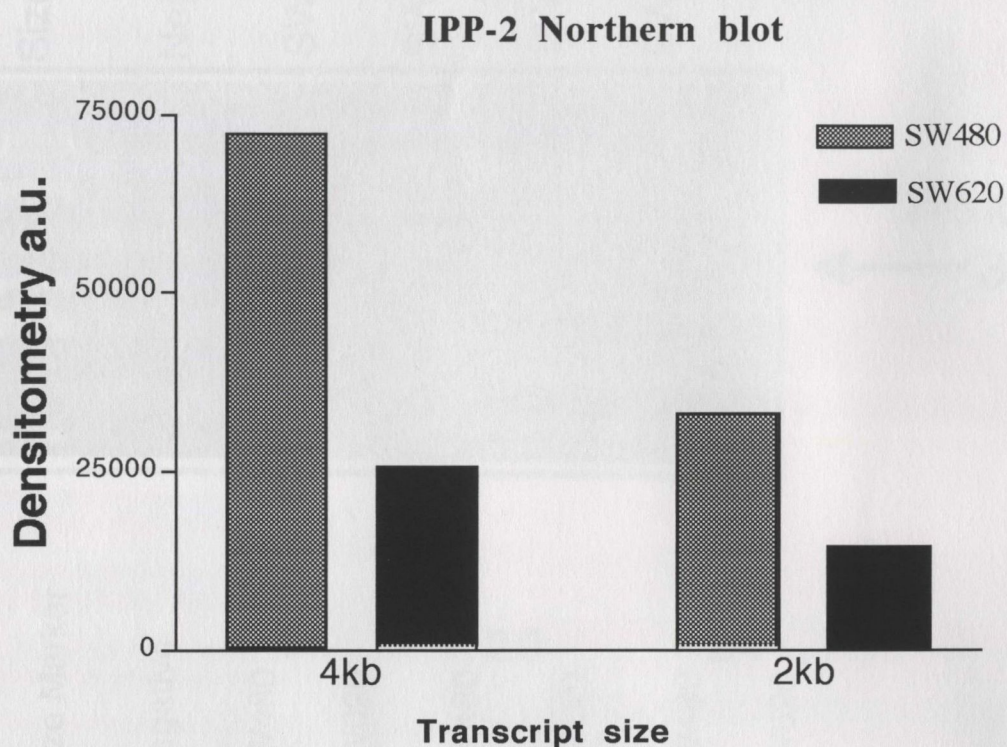


Figure 5.4 Graph of IPP-2 Northern blot analysis. Relative expression of both transcripts was determined by densitometry (Bio1D v6.32 software) and normalised to β -actin levels (see Appendix 5.4). Both transcripts showed approximately a 2 to 3 fold increase in expression in SW480 relative to SW620.

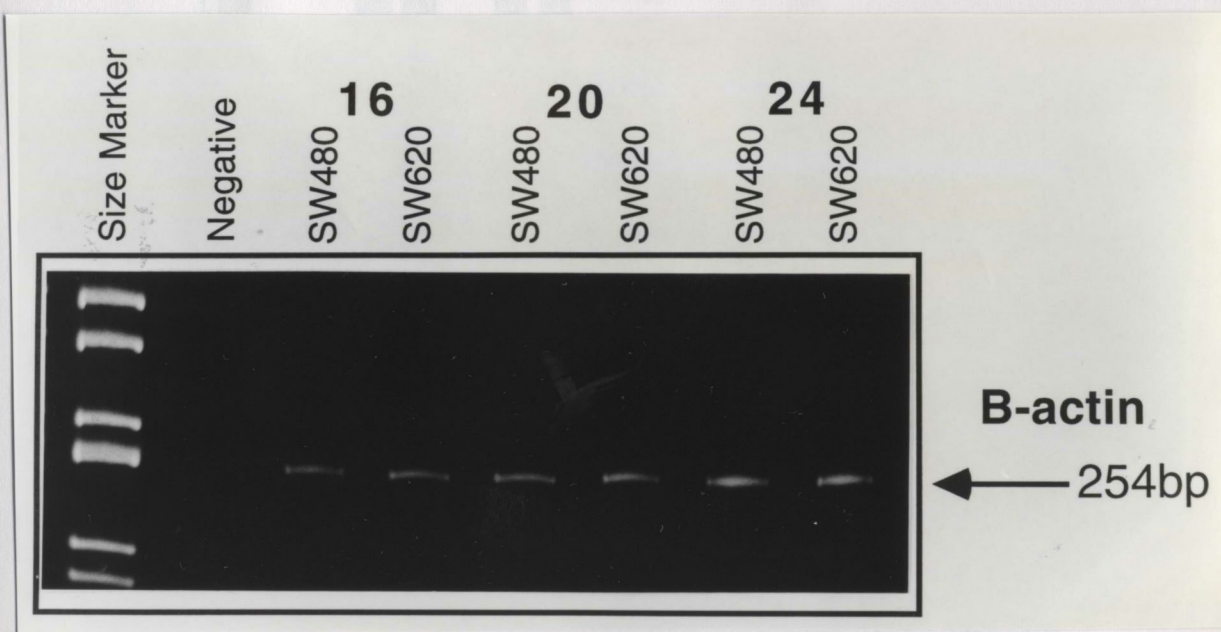
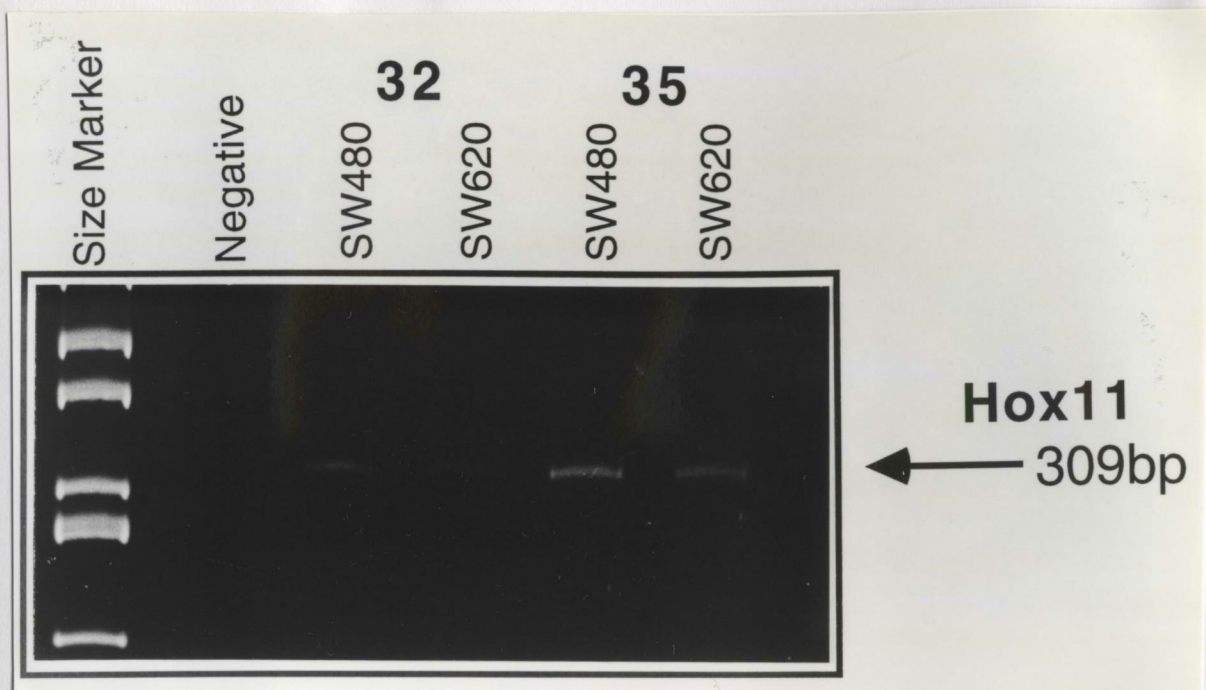


Figure 5.5 Expression of Hox11 in SW480 and SW620. Size Marker= pBS-*Hae* III molecular weight marker. Semi-quantitative RT-PCR products for β -actin and Hox11 were electrophoresed on 12% polyacrylamide gels and stained with ethidium bromide. The experiment was repeated >3 times and a range of cycle numbers were sampled to ensure the exponential phase of the PCR reaction was being sampled. The numbers above each pair of samples refers to the PCR cycle sampled. Hox11 showed increased expression in SW480 relative to SW620 while the control, β -actin, showed equivalent expression levels between the two cell lines.

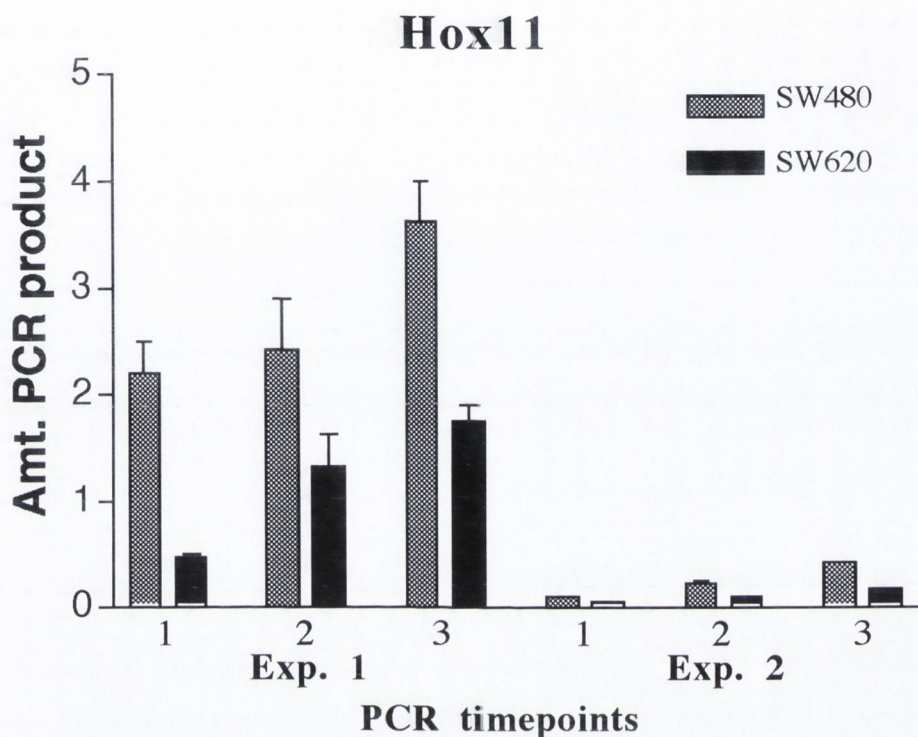
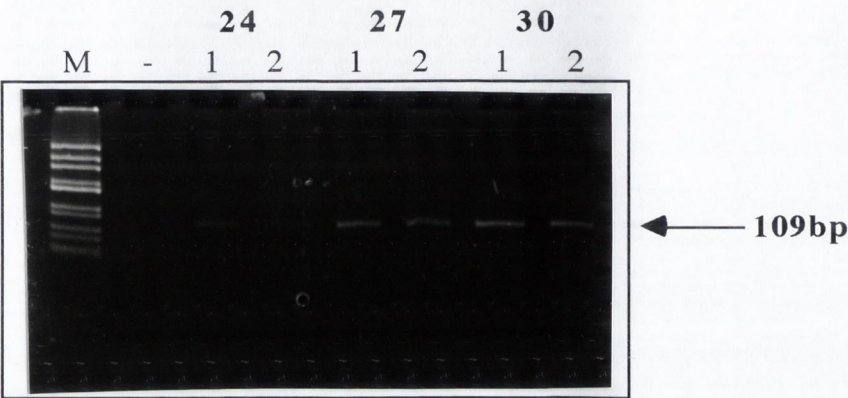


Figure 5.6 Graph of Hox11 semi-quantitative RT-PCR analysis in SW480 and SW620. The semi-quantitative RT-PCR products were quantified relative to β -actin levels using the ALFexpress™ automated sequencer and the data is described in Appendix 5.6. Hox11 and β -actin PCR products were loaded in the same lane in triplicate over a range of cycle numbers from different experiments. The data from two separate experiments (Exp. 1 & 2) are shown in the graph above (GraphPad, Prism™, USA). Hox11 showed a 2 to 3 fold increase in expression in SW480 relative to SW620.

A. Semi-quantitative RT-PCR analysis of IPP-2 with DDP primers



B. Semi-quantitative RT-PCR analysis of IPP-2 with DDT primers

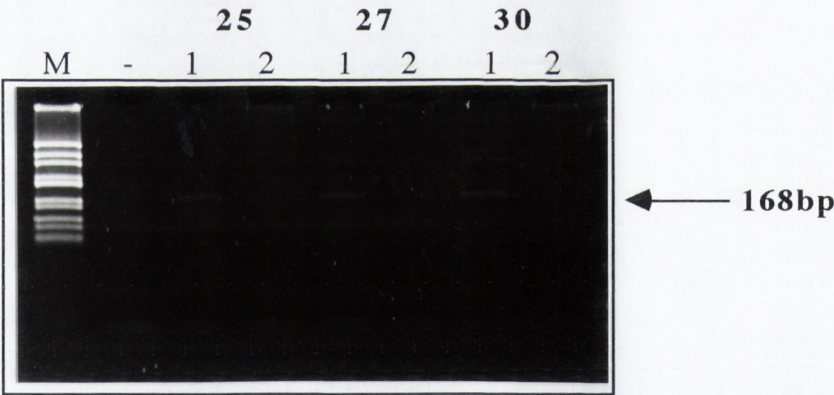


Figure 5.8 Expression of IPP-2 in SW1116 and 498li. M= pBS-*Hae* III molecular weight marker. -= negative control. 1= SW1116; 2= 498LI. Semi-quantitative RT-PCR products were visualised on 12% non-denaturing polyacrylamide gel and stained with ethidium bromide. **A.** Semi-quantitative RT-PCR analysis with DDP primers, which are specific for a region within the open reading frame of IPP-2 (see Chapter 2), showed decreased expression in the metastatic cell line 498LI relative to SW1116. **B.** Semi-quantitative RT-PCR analysis with DDT primers, which are specific for a region of the 3'UTR of IPP-2 where the DDRT clone PM208 originated, also showed decreased expression in 498LI relative to SW1116. This demonstrated that both transcripts of IPP-2 showed differential expression in a second pair of isogenic tumour cell lines.

CHAPTER 6

Analysis of protein expression in cell lines and tumours

6.0 INTRODUCTION

The gene expression studies described in the previous chapters were based on the use of cultured cell lines. The colon is a complex tissue made up of several cell types (Kondoh *et al.*, 1992), therefore the use of human colon carcinoma cell lines provided a way of examining a relatively pure population of carcinoma cells. Although the cell lines retain the morphological and physiological properties of their *in vivo* counterparts (Drewinko *et al.*, 1976; Leibovitz *et al.*, 1976; Semple *et al.*, 1978) alteration in gene expression patterns may occur during establishment of the cell lines *in vitro*. Thus, the possibility of a cell line artefact must be eliminated before a definitive association between gene expression levels and the development of metastasis can be made. This may be achieved by examination of candidate gene expression levels in a substantial number of tumour samples from different stages of tumour progression.

The limited number of metastatic tumours biopsied makes an analysis of late stage colon tumour progression difficult. In the vast majority of cases metastatic tumours are not biopsied as it offers no benefit to the patient. However, the lymph nodes of colon carcinoma patients are usually examined to establish the extent of tumour progression as lymph node involvement is a strong prognostic indicator (see Chapter 1). Thus, paraffin-embedded tumour material from different Dukes' stages, including lymph node metastases, were made available for analysis by the Department of Pathology, Beaumont Hospital and the Royal College of Surgeons in Ireland. The paraffin-embedded samples allowed an analysis of expression to be carried out by immunohistochemistry. Immunohistochemistry employs a polyclonal or monoclonal antibody to examine the distribution and level of expression of the tissue antigen (Rainbow, 1994). Prior to immunohistochemical analysis, the differential expression observed for mRNA transcript levels also has to be established at the protein level. Western blot analysis of protein fractions extracted from the cell lines allows protein levels to be examined. However, the Western and immunohistochemistry analyses is limited by the availability of antibodies to specific

proteins. In addition, this type of analysis is not amenable to secretory proteins and, in such cases, analysis of expression levels in clinical material would have to be restricted to mRNA levels.

Among the five confirmed differentially expressed genes described in Chapter 4 Section 4.1.2, commercial antibodies were only available for two. Keratin K5 was only available as a monoclonal antibody that also cross-reacted with Keratin 8. Thus, the proteins cannot be distinguished by immunohistochemical analysis but could be resolved on the basis of size on a Western blot. In addition, a monoclonal Keratin 8 specific antibody was received as a kind gift from Dr. David Easty, University College Dublin and would act as a control for both the Western blot and immunohistochemical analyses. An antibody produced against the Amyloid A (AA) component was also commercially available and was capable of cross-reacting with Serum amyloid A (SAA), which is the serum precursor of AA. However, SAA is a secretory protein, but reports of interaction with the extracellular matrix (Preciado-Patt *et al.*, 1996) meant that an immunohistochemical analysis was worth investigating.

The monoclonal antibodies for Keratin 5/8 (Neomarkers), Keratin 8 and Amyloid A (Dako) were used to analyse protein levels and localisation in the cell lines and in paraffin-embedded colon tumour material. Western blot analysis was carried out on subcellular fractionated protein samples from all 7 cell lines which included SW480, SW620, SW1116, 498LI, LoVo, Colo201 and T84 (described in Chapter 4). Immunohistochemical analysis was carried out on paraffin-embedded tumour sections ranging from Duke's B to Duke's C with matched lymph node metastases.

6.1 RESULTS

6.1.1 Subcellular fractionation and Western blot analysis

Subcellular fractionation was performed on all cell lines using a previously optimised method (Volkov *et al.*, 1998) and is described in Chapter 2 Section 2.4.2. Protein concentrations were estimated by the Markwell-Lowry assay (Markwell *et al.*, 1978) as described in Chapter 2 Section 2.4.3. Equivalent amounts of protein were Western blotted as outlined in Chapter 2 Section 2.4.4.

Keratin 5/8 and 8

The keratins are intermediate filaments and are components of the cytoskeletal structure of the cell (see Chapter 4 Section 4.2.5.1). Therefore, Keratins 5 and 8 were expected to localise in the cytoskeletal fraction. Although the main interest was in the expression of Keratin 5, the absence of an antibody specific for Keratin 5 meant that the level of Keratin 8 protein expression also had to be established. The SAGE analysis showed an approximately equal tag abundance for Keratin 8 between the primary colon tumour cell line SW480 and the lymph node metastatic cell line SW620 (see Chapter 4 Section 4.2.5.1). Thus, equal Keratin 8 protein levels were expected between the two cell lines. Whole cell and cytoskeletal fractions from SW480 and SW620 were Western blotted and incubated with the Keratin 8 antibody at a dilution of 1:400. A faint band of the expected size, i.e. 52.5 kDa, was detected in the cytoskeletal fractions in both cell lines (see Figure 6.1). Attempts were made to improve detection of Keratin 8 with this antibody by increasing the concentration but this resulted in no improvement. However, the Keratin 8 specific band, although faint, showed an approximately equal amount of protein in the two cell lines.

The Keratin 5/8 antibody detects proteins of 58kDa and 52.5kDa in size, i.e. Keratin 5 and Keratin 8 respectively. Expression of Keratin 5 at the mRNA level was relatively high in the primary tumour cell line SW480 and barely detectable in the metastatic tumour cell line SW620 (see Chapter 4 Section 4.1.2). Thus, expression of Keratin 5 protein was not expected in SW620. Subcellular fractions, including

membrane, cytosol and cytoskeletal fractions, from the cell lines SW480 and SW620 were resolved by electrophoresis on 10% SDS-PAGE in duplicate as described in Chapter 2 Section 2.4.1. One gel was Western blotted and incubated with Keratin 5/8 antibody at a dilution of 1:200 (Figure 6.2) and the second gel was stained with Coomassie blue (Figure 6.3). Figure 6.2 clearly shows two bands in the SW480 cytoskeletal fraction which are approximately the sizes expected for Keratin 5 and 8. A single band corresponding to Keratin 8 was detected in the SW620 cytoskeletal fraction. Keratin 5 was not detected in SW620 as expected. However, a relatively small amount of Keratin 8 was also detected in the membrane and cytosol fractions in both cell lines and was probably due to depolymerised keratin filaments (Dr. Aileen Long, pers. comm.). Figure 6.3 shows the relative amount of each protein fraction prior to Western blotting.

The cytoskeletal fractions from the second pair of primary and metastatic cell lines, i.e. SW1116 and 498LI, and the three other metastatic cell lines LoVo, Colo201 and T84 were also analysed by Western blot and incubated with the Keratin 5/8 antibody as described above. The cytoskeletal and whole cell fractions of SW480 and SW620 were also included as a control. A duplicate SDS-PAGE of the cytoskeletal fractions from the additional five cell lines was stained with Coomassie blue and is shown in Figure 6.4. Keratin 5 mRNA was not detected in all five cell lines as described in Chapter 4 Section 4.2.3. The Western blot analysis shown in Figure 6.5 detected a strong signal for Keratin 8 in all five cell lines as expected. A definitive Keratin 5 band was not detected in the cell lines (Figure 6.5a) but an extended exposure showed a band just above the Keratin 8 signal in all cell lines except T84 (Figure 6.5b). The band may be indicative of a small amount of Keratin 5 protein in these cell lines but is more likely to be a 'background' signal.

In summary, Keratin 5 protein was expressed only in the SW480 cell line and Keratin 8 was expressed in all of the cell lines examined as expected. Thus, the protein levels observed for Keratin 5 and 8 were similar to the mRNA levels and meant that the immunohistochemical analysis could proceed with greater confidence.

Serum Amyloid A

The Amyloid A antibody, which cross-reacts with Serum Amyloid A (SAA), would not normally detect SAA by Western blot analysis as it is a secreted protein. However, in tumour cells the characteristics of the SAA protein may be altered. To examine this possibility whole cell preparations of all cell lines were resolved on 15% SDS-PAGE and Western blotted. The blot was incubated with the Amyloid A antibody at a dilution of 1:200, but the 12.5 kDa SAA protein was not detected even after prolonged exposures. Thus, the SAA protein is most likely to be secreted directly out of tumour cells and may play an important role in the extracellular environment surrounding tumours.

6.1.2 Immunohistochemical analysis

Immunohistochemical analysis was performed as described in Chapter 2 Section 2.4.5. A total of 60 paraffin-embedded sections were analysed for Keratin 5/8 and Amyloid A staining which included 20 Duke's B, 20 Duke's C and 20 metastatic lymph node deposits from the Duke's C patients. The percentage staining was examined for both tumour and normal tissue and Keratin 5/8 staining within the tumour cells was graded as follows: Grade 1= weak cytoplasmic staining, Grade 2= strong membrane staining and weak cytoplasmic staining, Grade 3= strong membrane and strong cytoplasmic staining. The analysis included Keratin 5/8, Amyloid A and the corresponding negative controls. The Keratin 8 antibody was found to be unsuitable for immunohistochemistry as described in Chapter 2 Section 2.4.5. Therefore, Keratin 8 could not be used as a control for the Keratin 5/8 antibody. Moreover, any differences that might be observed between tumours of different stages would not be correlated to Keratin 5 alone as the expression pattern of Keratin 8 was not determined in the same sections.

Keratin 5/8

All slides were examined by two individuals separately and according to the grading system described above. The results for the Keratin 5/8 antibody are listed in Appendix 6.6/7 and the average percentage staining for each stage is described in Table 6.1. Examples of the staining patterns observed from each stage are shown in Figure 6.6 (normal (a) and Dukes' B (b)) and Figure 6.7 (Dukes' C (a) and lymph node metastasis (b)). There was an average increase ranging from 15 to 20% in the number of tumour cells stained relative to the number of normal cells stained in the Duke's B, C and the lymph node metastasis sections. However, the large standard deviations from the mean values meant that an overall correlation between increased Keratin 5/8 staining and primary colon tumours could not be made. In addition, comparison of average tumour staining patterns from each stage showed no difference in the percentage of staining. Moreover, the distribution of average percentage staining within each grade was approximately the same in each stage. Comparison of Duke's C tumours with the corresponding lymph node metastases from the same patient showed either an increase (10 out of 20 patients), a decrease (6 out of 20 patients) or no change (4 out of 20 patients) in the percentage of tumours cells stained (see Appendix 6.6/7). Thus, there was no obvious correlation between Keratin 5/8 staining and colon tumour progression.

Amyloid A

All slides were examined by two individuals and the percentage staining of tumour and normal cells was noted. The majority of slides did not show any staining within the tumour or normal cells and an example is shown in Figure 6.8a. However, staining around the tumour cells was noted in some cases but was difficult to distinguish from background staining (see Figure 6.8b). Amyloid staining in a section of the kidney acted as a positive control and is shown in Figure 6.9. Thus, the presence or absence of Serum Amyloid A (SAA) in the extracellular matrix and surrounding tissue could not be confirmed. Overall, SAA was not detected

intracellularly as expected and extracellular staining could not be distinguished from 'background' staining. Thus, analysis of SAA protein expression is not a useful method of investigating a correlation with colon tumour progression and an analysis of SAA mRNA levels is likely to be more informative.

6.2 DISCUSSION

Western blot analysis has shown that Keratin 5 is also differentially expressed at the protein level in metastatic colon tumour cell lines. Similarly to mRNA levels (see Chapter 4) Keratin 5 protein was only detected in the primary tumour cell line SW480. However, in contrast to the similar SAGE tag abundances observed for Keratin 5 and Keratin 8 in SW480 i.e. 17 and 19 respectively (see Chapter 3 Table 3.2), the amount of Keratin 8 detected at the protein level was at least three times that of Keratin 5 (see Figure 6.2). Thus, Keratin 8 mRNA may be translated more efficiently or the protein may have a longer half life. The putative role of Keratin 5 in colon tumour metastasis is discussed in Chapter 4 Section 4.2.5.1 and a definitive association with metastasis would require an observation of decreased expression of Keratin 5 in a substantial number of metastatic tumours of the colon. The Keratin 5/8 immunohistochemical results described above did not show an inverse correlation with advanced Dukes' stage. However, it is likely that the more highly abundant Keratin 8 protein masked any potential Keratin 5 expression differences. In addition, high expression of Keratin 8 in colon tumours has been reported previously (Frigerio *et al.*, 1995) and the observed increase in staining of Keratin 5/8 in tumour relative to normal cells in some of the sections may be due to Keratin 8 protein expression only. Therefore, the immunohistochemical analysis with Keratin 5/8 antibody proved to be inconclusive and an inverse correlation of Keratin 5 with colon tumour metastasis still has to be tested. This could be achieved by generating an antibody specific for Keratin 5 and repeating the immunohistochemical analysis or analysing mRNA expression levels in fresh frozen metastatic tumour specimens when available.

The Western and immunohistochemical analyses with the Amyloid A antibody did not yield any useful data. Thus, an association of SAA with colon tumour metastasis still requires further analysis and should be restricted to an analysis of mRNA levels only.

In summary, differential expression of Keratin 5 protein in metastatic colon tumour cell lines has been clearly demonstrated. In addition, an inverse correlation of

Keratin 5 and SAA with colon tumour metastasis has neither been proved or disproved based on the above analyses and further experimentation is required before a definite association can be made.

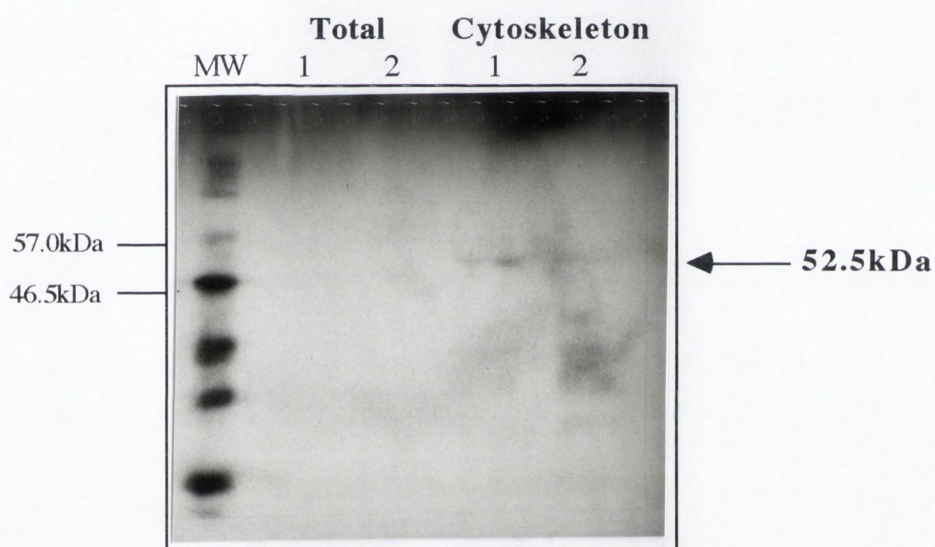


Figure 6.1 Keratin 8 Western blot analysis in SW480 and SW620. MW= Broad range biotinylated protein marker (NEB, UK). 1= SW480; 2= SW620. Total= total cell lysates. Cytoskeleton= cytoskeletal fraction isolated by subcellular fractionation. A Western blot of whole cell and cytoskeletal fractions from SW480 and SW620 were incubated with Keratin 8 antibody (1:400) and a faint band of expected size was detected at approximately the same intensity in the cytoskeletal fractions only.

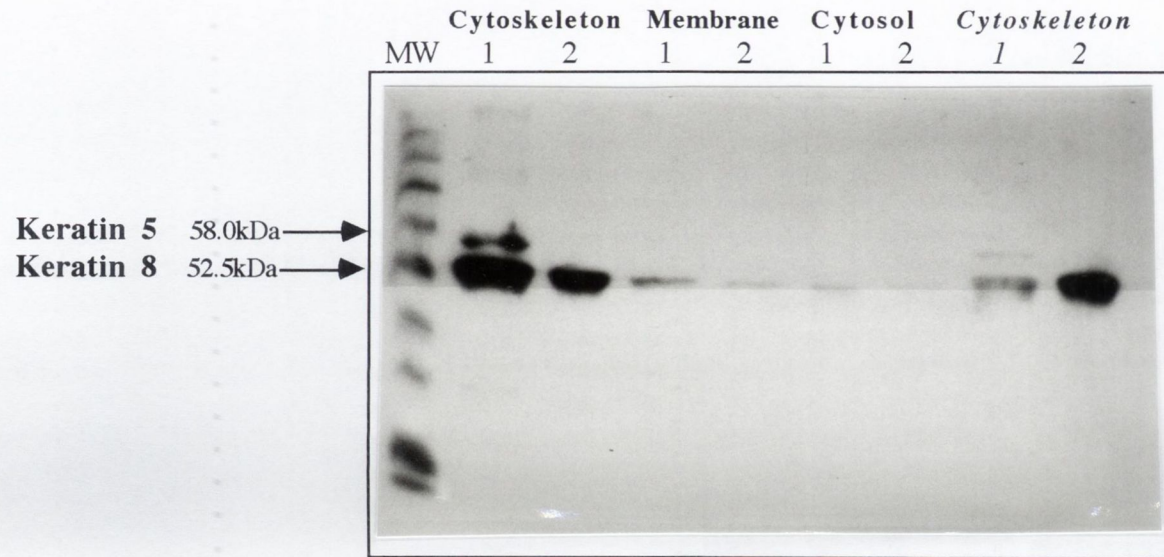


Figure 6.2 Keratin 5/8 Western blot analysis in SW480 and SW620. MW= Broad range biotinylated protein marker (NEB, UK). 1= SW480; 2=SW620. Equivalent amounts of subcellular fractions from SW480 and SW620 were resolved on 10% SDS PAGE and Western blotted and included cytoskeletal, membrane and cytosolic fractions. The cytoskeletal fractions written *in italics* are the same as the ones written in plain text except approximately three times more protein was loaded in the SW620 lane relative to the SW480 lane. Incubation with Keratin 5/8 antibody (1:200) clearly showed the majority of Keratin 5 and 8 was expressed in the cytoskeletal fraction as expected. Keratin 5 protein was only detected in SW480 and increasing the amount of SW620 protein (*cytoskeleton 2*) and prolonged exposure did not result in detection of Keratin 5 in SW620.

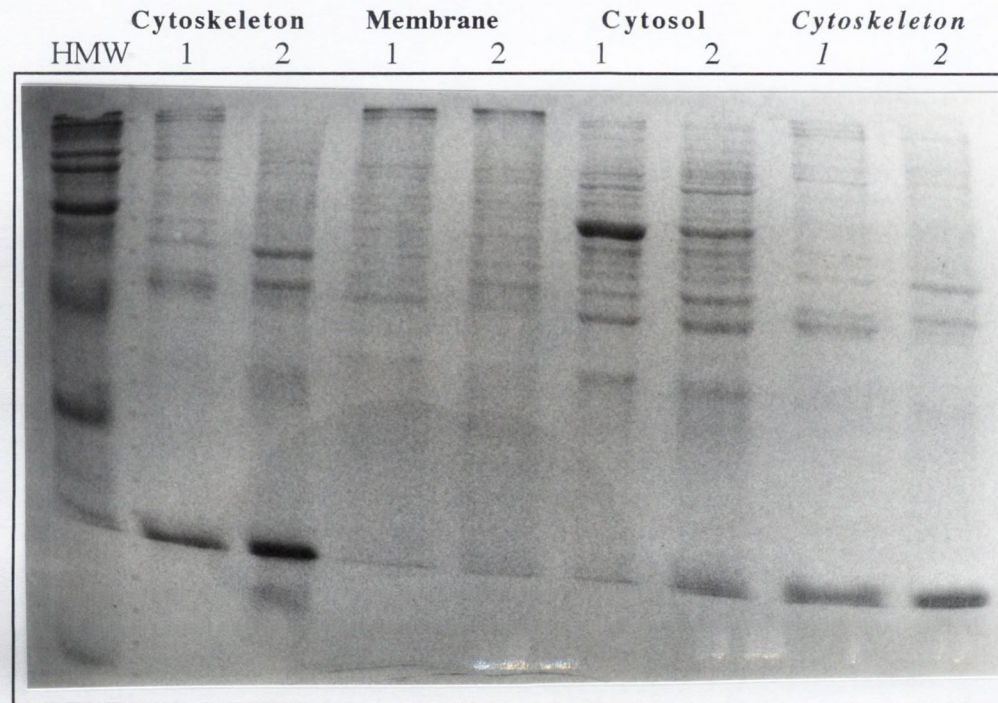


Figure 6.3 Amount of protein loaded in Keratin 5/8 Western blot analysis of SW480 and SW620. HMW= high molecular weight marker (sizes: 205, 116, 97.4, 66, 45, and 29kDa). Duplicate amounts of protein from each fraction used in the Western blot analysis (Figure 6.2) were resolved on a 10% SDS-PAGE and stained with Coomassie blue. The gel shows that each pair of fractions contained a similar amount of protein.

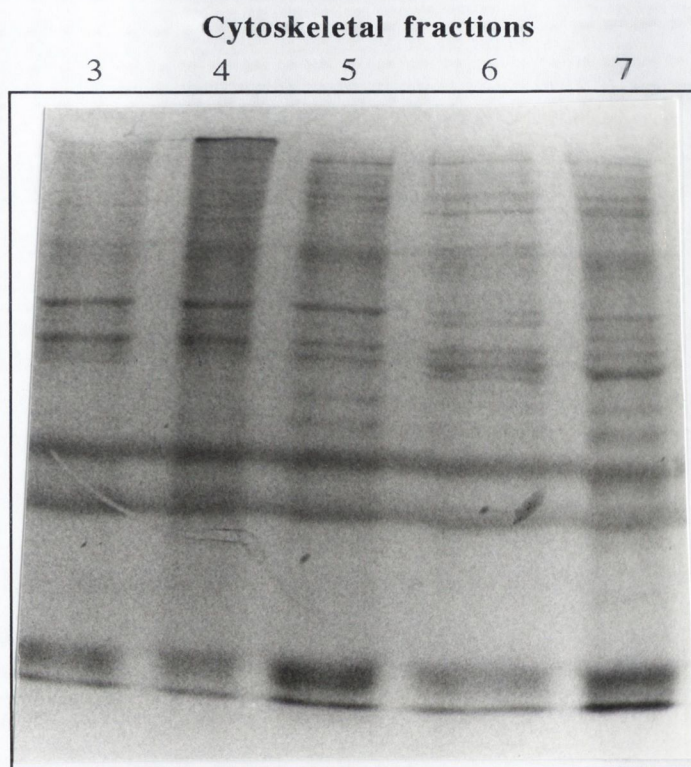
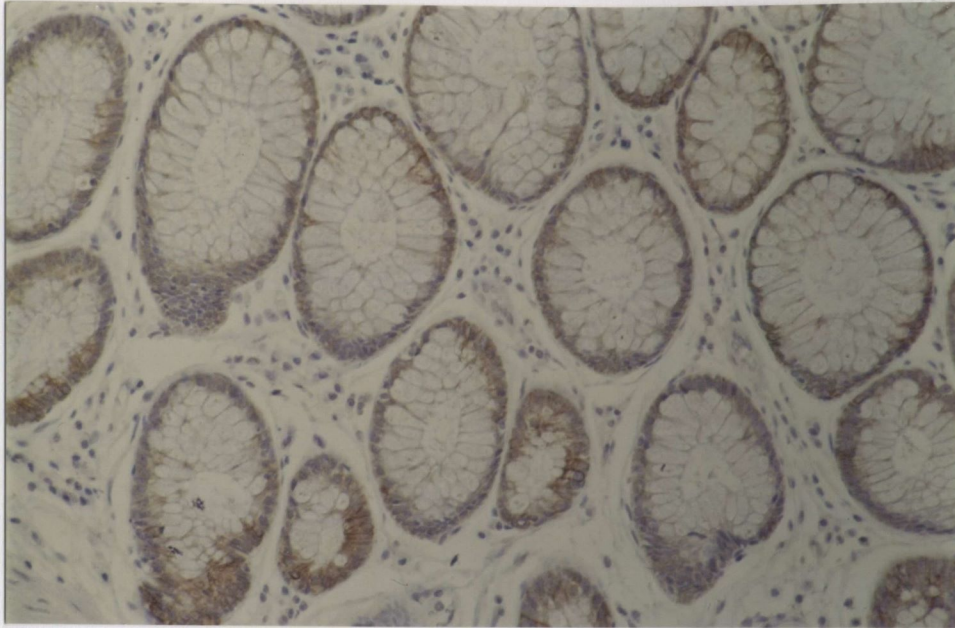


Figure 6.4 Amount of protein loaded in Keratin 5/8 Western blot of LoVo, Colo201, T84, SW1116 and 498LI. 3= LoVo, 4= Colo201, 5= T84, 6=SW1116, 7=498LI. Duplicate amounts of the cytoskeletal fractions from the cell lines used in the Western blot (see Figure 6.5) were electrophoresed on a 10% SDS-PAGE and stained with Coomassie blue as shown above.

A. Normal



B. Dukes' B tumour

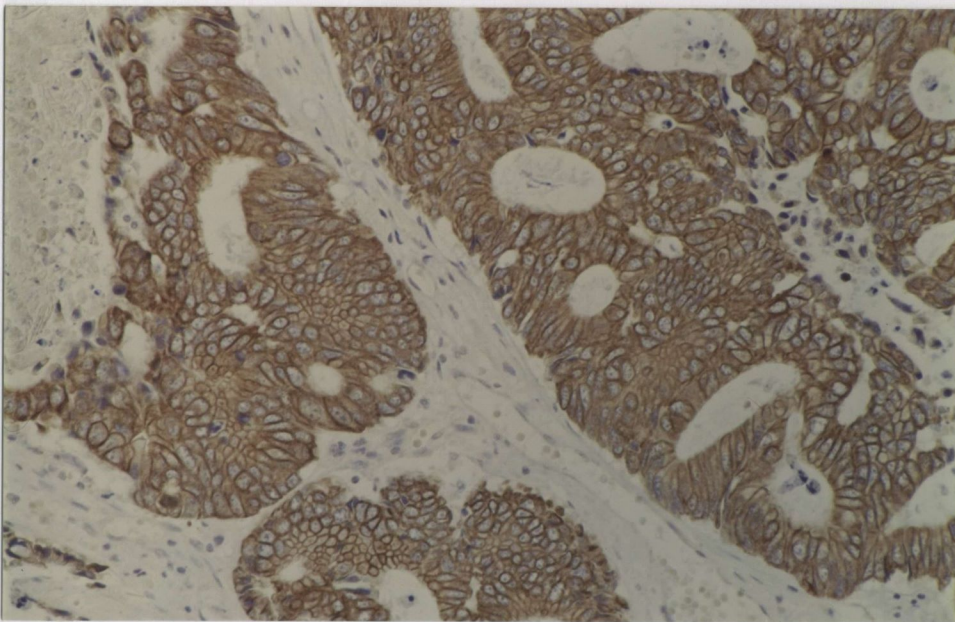
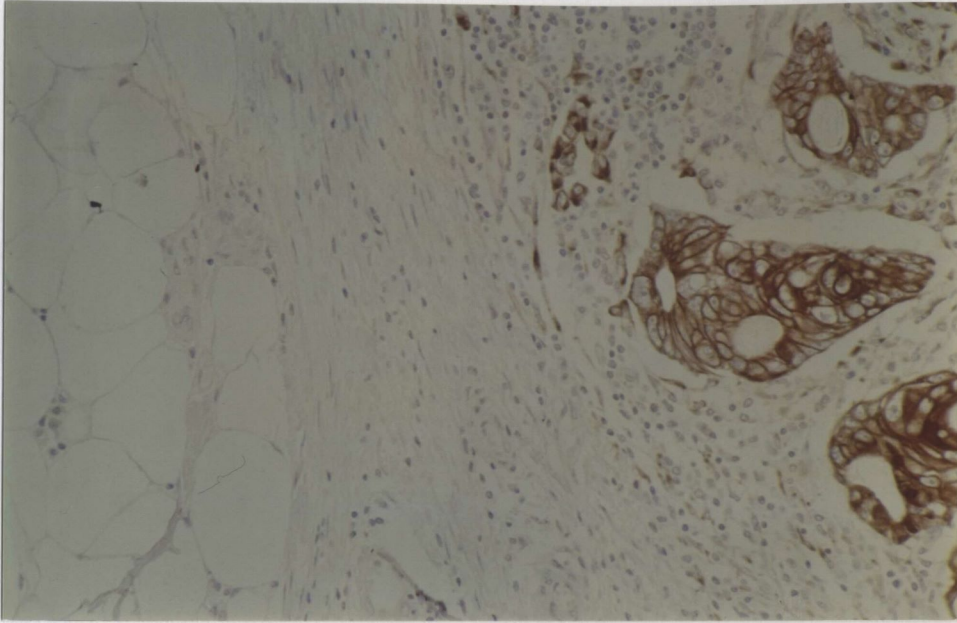


Figure 6.6 Keratin 5/8 immunohistochemical staining of normal colon and Dukes' B tumour. **A.** Keratin 5/8 staining of normal colon tissue. A significant percentage of normal cells in this section were negative. **B.** Keratin 5/8 staining of a Dukes' B tumour from the same patient. All tumour cells in this section showed positive staining of varying grades.

A. Dukes' C tumour



B. Lymph node metastasis

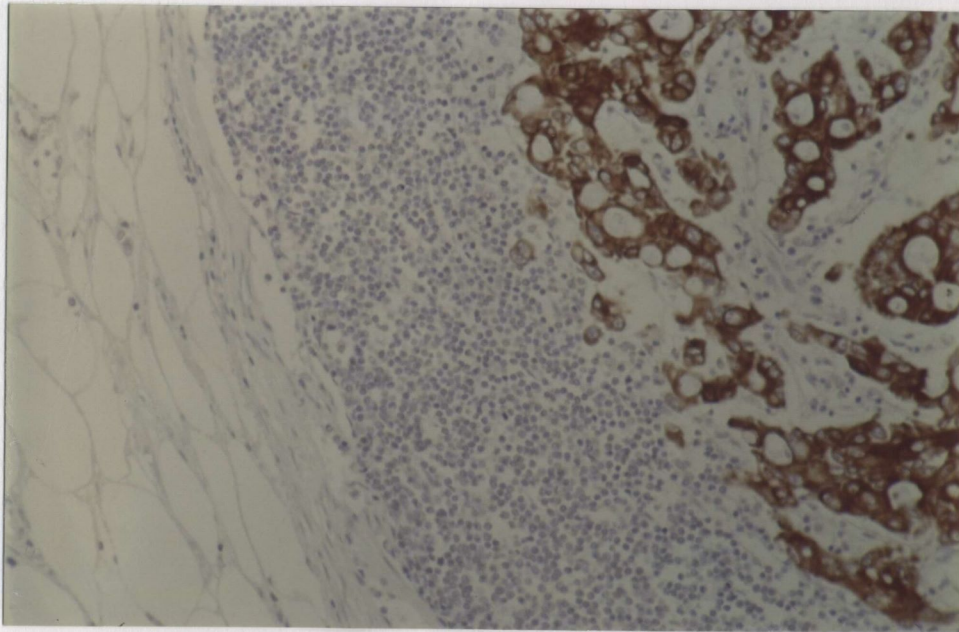
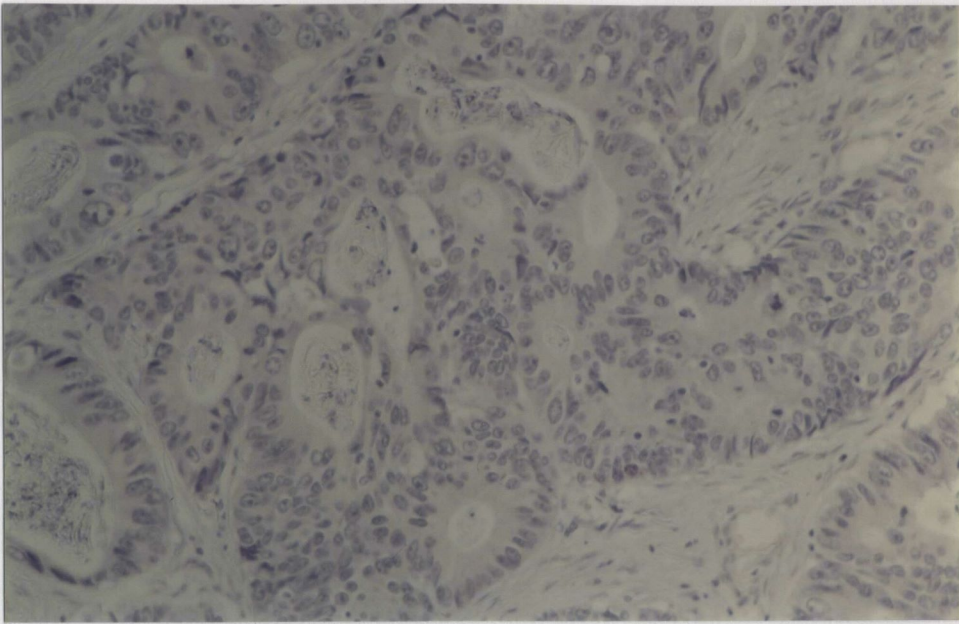


Figure 6.7 Keratin 5/8 immunohistochemical staining of a Dukes' C tumour and a lymph node metastasis. **A.** Keratin 5/8 staining of a Dukes' C tumour. The majority of tumour cells in this section showed strong staining for Keratin 5/8 of varying grades. **B.** Keratin 5/8 staining of a lymph node metastasis of the same patient. All tumour cells in this section showed strong staining for Keratin 5/8.

A.



B.

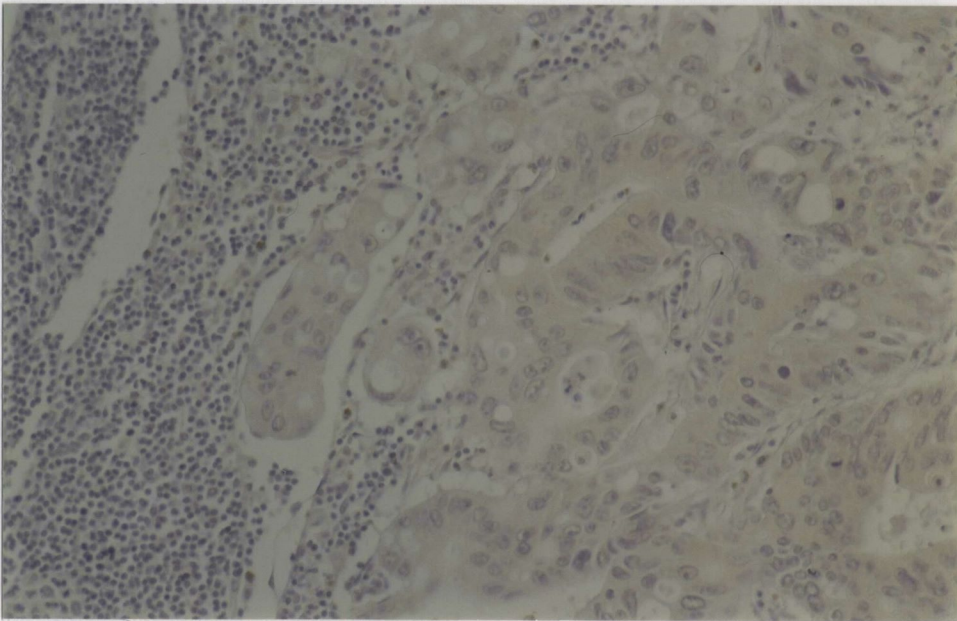


Figure 6.8 Amyloid A immunohistochemical staining of Dukes' B tumours. A. Typical Amyloid A negative staining pattern observed for the majority of sections including the Dukes' C and lymph node metastases. **B.** Staining around the tumour cells was observed in some of the sections but could not be distinguished from the background.

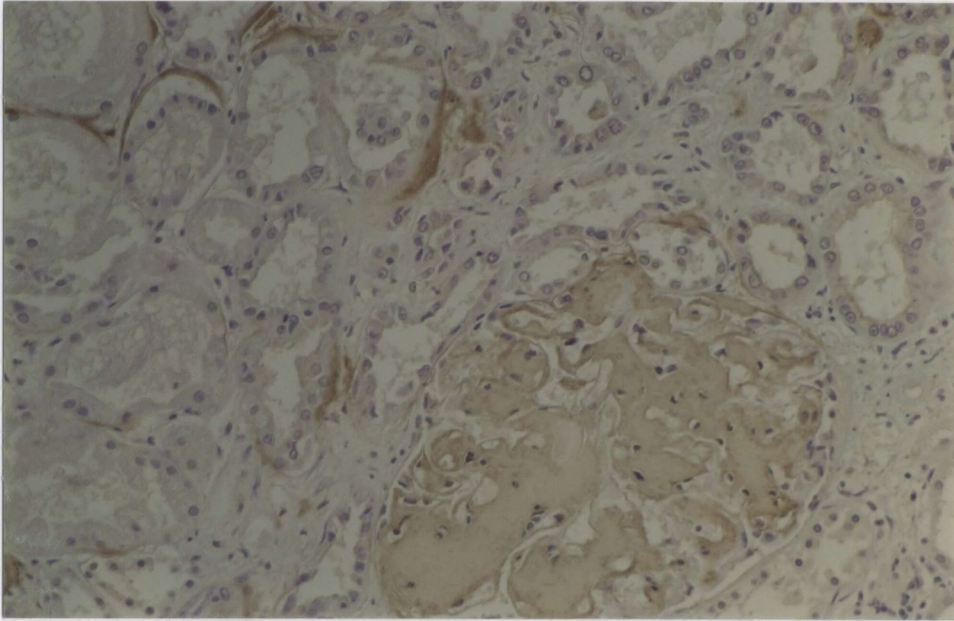


Figure 6.9 Amyloid A positive control. A section of a kidney showing amyloid deposits in the glomerulus and the collecting ducts.

Table 6.1 Average Keratin 5/8 percentage staining for each Dukes' stage

The percentage staining from each stage, which is detailed in Appendix 6.6/7, was analysed using GraphPad (Prism™, USA). n= number of sections. S.D.= standard deviation. S.E.= standard error.

	% Normal positive	% Tumour positive	Grade 1	Grade 2	Grade 3
Dukes' B					
n	12	20	20	20	20
Mean %	61.25	81.6	34.95	41	24.05
S.D.	21.65	15.97	28.4	24.51	28.62
S.E.	6.25	3.57	6.35	5.48	6.40
Dukes' C					
n	13	20	20	20	20
Mean %	73.92	89.6	24.75	50.35	23.15
S.D.	25.48	9.74	21.18	25.42	25.48
S.E.	7.07	2.18	4.74	5.68	5.70
Lymph node metastases					
n	4	20	20	20	20
Mean %	67.5	87.9	30.53	43.58	25.89
S.D.	5	20.67	26.44	24.44	31.85
S.E.	2.5	4.62	6.07	5.61	7.31

CHAPTER 7

General Discussion

7.1 SUMMARY OF RESULTS

The main aim of this study was to identify altered gene expression patterns that occur during the process of colon tumour metastasis. This was achieved by comparing gene expression profiles generated for the pair of isogenic cell lines SW480 and SW620 (Leibovitz *et al.*, 1976) using the method of Serial Analysis of Gene Expression (SAGE) (Velculescu *et al.*, 1995). SW480 and SW620 were derived from a primary colon tumour and a lymph node metastasis of the same patient respectively (Leibovitz *et al.*, 1976). These cell lines provided a model system for examining late stage colon tumourigenesis and allowed identification of a panel of differentially expressed genes that may be associated with metastasis. In addition to the SAGE data, the panel of differentially expressed genes also included IPP-2 and Hox11 which were identified from DDRT and semi-quantitative RT-PCR analyses on the same pair of cell lines (described in Chapter 5). This panel of genes formed the basis of further analysis.

The SAGE protocol is a recently introduced method (Velculescu *et al.*, 1995) and therefore implementation required a certain amount of optimisation as outlined in Chapter 3. The moderate SAGE analysis described here included a total of over 10,000 tags which resulted from sequencing at least 150kb of cDNA. This analysis identified the most highly expressed genes in SW480 and SW620, thus providing new information on the expression patterns of abundant genes in these cell lines. Prior to this work, studies on SW480 and SW620 or similar isogenic cell lines identified only differences in gene expression (Kondoh *et al.*, 1992; Yamamoto *et al.*, 1996; Cajot *et al.*, 1997; Riley *et al.*, 1997). Therefore, the data produced from the SAGE analysis allowed, for the first time, comparison of the most abundant transcripts between colon tumour cell lines. This analysis not only allowed differences to be identified, but transcripts which did not alter in expression between the two cell lines were also identified. Gene transcripts which are not altered in expression during metastasis are as important to identify as transcripts that are differentially expressed. Moreover, the identification of differentially expressed genes provides clues to the metastatic

process but have to be considered in the context of a complete gene expression profile. Therefore, identification of altered and unaltered expression of specific genes during tumourigenesis contributes to a greater understanding of metastasis.

The gene expression profiles from the isogenic cell lines were categorised into functional groups and the proportions of each category, which included cell protein synthesis/turnover, signalling/communication, cell structure/motility, homeostasis/defense, RNA synthesis and metabolism, were comparable between SW480 and SW620 with minor fluctuations. The dominant category by far was protein synthesis/turnover which demonstrated that a substantial portion of the cell's energy is devoted to this process. This was expected in cells that are actively growing and dividing. Approximately 20% of the most abundant transcripts were cDNA clones or ESTs and therefore still have to be characterised. Future identification of these transcripts will be aided by the expanding EST database and the completion of the entire human genome sequence.

Comparison of the gene expression profiles between SW480 and SW620 identified 81 gene transcripts that showed at least a three-fold alteration in expression in one cell line compared to the other (Chapter 3, Table 3.5). Subsequent confirmatory analyses led to a re-assessment of the SAGE data and this reduced the panel of differentially expressed SAGE tags from 81 to 49 (Chapter 4). A total of eleven differentially expressed tags were analysed by semi-quantitative RT-PCR and/or Northern blot and only five of these were confirmed as real differences in expression. Thus, the type of moderate SAGE analysis carried out here and by at least one other group (deWaard *et al.*, 1999) showed that comparison of tag abundances <10 may not accurately represent real differences in expression and should be confirmed by an independent method. This resulted from an insufficient amount of tags sampled and has not proved to be a problem when $>50,000$ tags were sequenced per compared sample (Hermeking *et al.*, 1997; Madden *et al.*, 1997; Polyak *et al.*, 1997; Velculescu *et al.*, 1997; Zhang *et al.*, 1997; Hibi *et al.*, 1998). The final SAGE data analysis identified a total of 49 differentially expressed tags

which included 5 independently confirmed differences, 22 likely differences and a further 22 which may be differentially expressed. The final panel of differentially expressed genes included 15 ESTs or cDNA clones. One of these clones was successfully identified using the Unigene website (<http://www.ncbi.nlm.nih.gov/UniGene/>). A second EST clone did not contain the last *Nla* III site when assembled as part of a putative transcript (HT1323) using TIGR (<http://www.tigr.org/tdb/hgi/hgi.html>) and therefore the SAGE tag was incorrectly assigned. Moreover, identification of a possible cloning artefact in the PACAP type-3/VIP type-2 receptor sequence (Chapter 4, Section 4.1.3.6) re-emphasised the need for caution when making gene to tag assignments. There are several problems in using such a short sequence tag for identifying and counting transcripts. Some SAGE tags will be shared by more than one gene and therefore a definitive assignment cannot be made. Sequencing error rates will also have an effect on tag assignments and this includes errors in SAGE tag sequencing and errors within the Genbank database. Errors introduced when sequencing the SAGE clones cannot be identified and therefore SAGE tag counts of one may have arisen due to a sequencing error and are usually ignored. However, errors within the Genbank database and the absence of a complete set of sequences may result in an incorrect tag to gene assignment. The main problem arises with ESTs which are usually single-pass sequences and therefore will contain a certain amount of sequencing errors. On the other hand, well-characterised gene transcripts are usually error-free with minor differences due to polymorphisms. Thus, the high throughput nature of SAGE results in some sacrifice in accuracy and therefore tag to gene assignments have to be assessed critically especially in relation to ESTs and incomplete gene transcript entries.

The sequence databases were also utilised to identify DDRT clone PM208 (see Chapter 5). Clone PM208 was identified as an alternative polyadenylation transcript of IPP-2 largely by utilisation of the EST and Genbank databases. The sequence of the additional 3' UTR region of IPP-2 was assembled as described in Chapter 5 and submitted to Genbank (AJ133812). Subsequent Northern blot analysis

confirmed that both transcripts of IPP-2 were differentially expressed in the cell lines SW480 and SW620. Hox11, an associated gene product, was also found to be differentially expressed in these cell lines. This analysis brought the total panel of identified differentially expressed genes to 51. Expression of a subset of six genes from this panel was examined in other metastatic cell lines (Chapters 4 & 5) and two out of the six were further analysed at the protein level and in clinical material by immunohistochemistry (Chapter 6). Expression of Keratin K5, Serum Amyloid A, Cystatin S and both forms of IPP-2 mRNA transcript showed an increase in the primary tumour cell line SW480 relative to the metastatic cell line SW620 and this was demonstrated by three independent methods, i.e. SAGE or DDRT, semi-quantitative RT-PCR and Northern blot. Decreased or complete absence of expression of these genes was also found in all the additional metastatic cell lines examined by semi-quantitative RT-PCR which included LoVo, Colo201, T84 and 498LI (see Chapter 4, Tables 4.2 & 4.3). Differential expression of Keratin K5 was also demonstrated at the protein level and its expression was found exclusively in the cytoskeletal fraction. The similar transcript levels of Keratin K5 to K8 shown in the SAGE analysis was found to differ from the protein ratios which showed at least three times more Keratin K8 relative to Keratin K5. The immunohistochemical analysis was inconclusive for Keratin K5 due to the cross-reactivity of the antibody to Keratin K8 and was also inconclusive for SAA as possible detection in the extracellular matrix could not be distinguished from background staining (see Chapter 6). The human homologue of yeast ribosomal S28 (S28) and the p32 subunit of Splicing Factor 2 (SFp32) showed increased expression in the metastatic cell line SW620 relative to the primary tumour cell line SW480 and again this was shown by three independent methods as stated above. Increased expression of S28 and SFp32 was also observed in at least one other metastatic cell line (see Chapter 4, Table 4.3).

7.2 CONTRIBUTION TO COLON TUMOUR METASTASIS RESEARCH

Cancer metastasis is a highly complex process that must involve the coordinate induction and repression of a range genes (Sherbet & Lakshmi, 1997). The metastatic process may be divided into a number of different steps which are described in detail in Chapter 1 but, briefly, includes vascularisation of the tumour, detachment of metastatic cells from the primary tumour, local invasion of the extracellular matrix, entry into the vascular system, survival in the circulation, extravasation into secondary organ tissue and proliferation at the secondary site (Ahmad & Hart, 1997). Thus, metastasis involves gene products associated with angiogenesis, degradation of the extracellular matrix, adhesion, motility, proliferation and avoidance of apoptosis. A large number of studies have analysed expression of gene products previously associated with metastasis in a range of colon tumours of different stages and found that altered expression had a correlation with metastasis. Examples of some recent studies include analyses of CD44 variants (Sherbet & Lakshmi, 1997; Weg-Remers *et al.*, 1998; Yamane *et al.*, 1999), the DCC protein (Saito *et al.*, 1999), VEGF mRNA (Ishigami *et al.*, 1998), metalloproteases and their inhibitors (Lottaz *et al.*, 1999; Masuda & Aoki, 1999), α -catenin and E-cadherin (Gofuku *et al.*, 1999), MTA1 (Toh *et al.*, 1997), cyclooxygenase-2 (Tsuji *et al.*, 1997) and c-Myc (Kakisako *et al.*, 1998).

In an effort to reach a greater understanding of the metastatic process, identification of new genes whose expression is altered during tumourigenesis has been the subject of intense study over the past number of years. Investigation into colon tumour metastasis is no exception and the introduction of methods such as DDRT (Liang & Pardee, 1992), PCR-based subtraction methods (Diguid *et al.*, 1989; Wieland *et al.*, 1990; Hara *et al.*, 1991; Hubank & Schatz, 1994) and more recently SAGE (Velculescu *et al.*, 1995) and cDNA microarray (Schena *et al.*, 1995) have resulted in reports of a variety of differentially expressed genes associated with colon tumourigenesis. One of the largest studies to date, which was published by the Kinzler group at Johns Hopkins University, focused on identifying gene expression

differences between normal colon and tumour tissue/cell lines (Zhang *et al.*, 1997). The study consisted of a large scale SAGE analysis which identified a total of 289 transcripts that were expressed at significantly different levels between normal colon and tumour tissue (Zhang *et al.*, 1997). An additional important finding from this study was that a significant portion of the gene expression differences were preserved during *in vitro* culture demonstrating the usefulness of cultured cell lines. The Kinzler group has since completed extensive SAGE analysis on 14 additional tumour cell lines, the majority of which were derived from colon tumours. Limited public access to this information is available through the Unigene website (<http://www.ncbi.nlm.nih.gov/UniGene/>). A comprehensive analysis examining the differences between primary colon and metastatic tumours had not been described prior to the study described here. However, methods such as DDRT-PCR have been applied to such isogenic cell lines and resulted in the association of acylphosphatase (Riley *et al.*, 1997), MRP-1/CD9 (Cajot *et al.*, 1997) and IPP-2 (Chapter 5) expression with colon tumour metastasis. The SAGE analysis described in this work produced a comprehensive examination of the most abundant transcripts in the primary colon and metastatic tumour cell lines derived from the same patient. Thus, these results are the first to compare the expression profiles of highly abundant transcripts in primary colon tumour and metastatic cell lines and has resulted in identification of genes which are expressed at both the same and different levels. The roles that some of the identified differentially expressed genes might play in metastasis have already been discussed in Chapters 4 and 5. Cytoskeletal/motility and ribosomal associated transcripts were the main two categories that showed altered expression. However, the mechanisms behind the differences in expression observed are not known but when complete loss of expression of a specific gene is observed, chromosomal loss may be the cause. Thus, the apparently complete loss of expression of Serum Amyloid A and Cystatin S in the metastatic cell line SW620 may be due to chromosomal instability. Chromosomal instability is a characteristic of most colorectal tumours and involves gains and losses of whole chromosomes

(Lengauer *et al.*, 1998) and evidence has shown that a defect in a well defined mitotic checkpoint may be one of the underlying causes (Cahill *et al.*, 1998). The primary colon tumour cell line SW480 is characterised by chromosomal instability (Cahill *et al.*, 1998) and a recent comparative genomic hybridization analysis of primary and metastatic colon tumours identified specific chromosomal alterations, some of which were shared between primary and metastatic tumours and others which were altered in metastatic tumours only (Paredes-Zaglul *et al.*, 1998). Alterations that were unique to metastatic tumours included loss of chromosomes 9q, 11q, and 17q. Interestingly, Serum amyloid A (SAA) is encoded on chromosome 11 which may explain the complete loss of expression of SAA in the metastatic cell line SW620. Although chromosomal instability occurs early in tumour evolution, the genetic variations resulting from these instabilities increase with tumour progression (Lengauer *et al.*, 1998). Thus, complete loss of expression of genes in metastatic cells may be due to further increased chromosomal instability. A comparative genomic hybridization analysis of SW480 and SW620 would resolve this issue.

In addition to the differentially expressed genes identified by others as mentioned above, the data described here increases our understanding of colon tumour metastasis and complements the data that already exists. Comprehensive expression profiles from every stage of colon tumourigenesis will complete the picture of metastatic spread and the analysis presented here contributes a significant piece to the puzzle and will form the basis of further research. However, recent technical advances in laser capture microdissection (LCM) (Emmert-Buck *et al.*, 1996) and expression profiling methods would now allow a more significant study to be carried out on colon tumour metastasis. Although the isogenic cell lines used in this study provided a useful and convenient model system, direct examination of clinical material is more desirable. The problem of cellular heterogeneity within a tissue can now be overcome by advances in tissue microdissection. LCM is a fast and dependable method that allows distinct subpopulations of cells within a tissue to be captured and preserved under direct microscopic visualisation (Simone *et al.*, 1998).

Specific cells within a tumour can now be isolated and improved extraction methods means that RNA can be efficiently isolated and a gene expression profile generated with a relatively small amount of RNA using a cDNA microarray (DeRisi *et al.*, 1996; Lockhart *et al.*, 1996) or microSAGE (Datson *et al.*, 1999). This would allow a comparison of gene expression profiles amongst the subpopulations of cells that exist within a tumour which possess varying metastatic capabilities. Thus, the gene expression patterns of normal epithelium, premalignant lesions, invasive cancer foci and host response cells all from the same tissue of a single patient can be sampled and compared as described above (Simone *et al.*, 1998). This would allow identification of gene expression changes during the process of tumour metastasis and not after it has already happened and therefore is likely to be more clinically relevant. This type of study is just beginning to be developed and will have a major role in understanding the changes tumour cells undergo in order to become metastatic.

7.3 FUTURE WORK

7.3.1 Selection of candidate genes for further analysis

The panel of 51 differentially expressed genes identified in this work will be expanded by a Suppression Subtractive Hybridization (SSH) analysis on the same cell lines, i.e. SW480 and SW620, which is currently underway. This work will complement the SAGE analysis by isolating moderate and low abundance transcripts. The basis of SSH includes the isolation of mRNA from the two cell populations being compared and subsequent conversion to double-stranded cDNA. The isolation of transcripts expressed in one cell population and not the other is achieved by first treating one population as the 'tester' cDNA and the second population as the 'driver' cDNA and then *vice versa*. Both the driver and tester cDNA populations are digested with a frequently cutting restriction endonuclease. The tester cDNA is divided in half and two different adaptors are added to each population of tester. Excess driver cDNA is then added to each population of tester cDNA and the cDNA mixture is heat denatured and allowed to reanneal. This step results in a normalisation (or equalisation) of high and low abundance single stranded tester cDNAs due to second order hybridisation kinetics. The two populations are then mixed and are subjected to a second round of hybridisation in the presence of more driver cDNA. The products of this step are then amplified by PCR with primers complementary to the adaptors. The adaptors are designed so that only differentially expressed cDNAs are exponentially amplified (Diatchenko *et al.*, 1996; Dunican *et al.*, 1997). Thus, the subtracted cDNAs are then subcloned and identified by subsequent sequencing and homology searching. The subtracted cDNAs from each cell line and a second isogenic pair of cell lines (SW1116 and 498LI) may also be applied to the commercial cDNA microarray filters from Genome Systems (Gene Discovery Arrays) which contains 18,000 cDNAs per filter. Selected genes from this panel will be analysed further as candidate metastasis-associated genes once differential expression has been confirmed and a similar pattern of expression is observed in additional metastatic colon tumour cell lines.

Candidate genes should be analysed initially in a range of clinical material including normal, primary and metastatic tumour samples. Due to the complexity of tumour tissue use of microdissected tumour samples, *in situ* hybridization or immunohistochemistry would be necessary to accurately assess differential expression of candidate genes in clinical material. Unfortunately, the immunohistochemical analysis described in this work was inconclusive due to the high expression and cross-reactivity of the antibody with Keratin K8 (see Chapter 6). A correlation of expression with Dukes' stage would strongly implicate that gene in the development of colon tumour metastasis and would warrant further analysis.

7.3.2 Further analysis of candidate genes

Demonstration of altered expression of candidate genes having a direct effect on the metastatic properties of tumour cells would provide further evidence for a role in the development of metastasis. This may be achieved by either overexpressing or inhibiting expression of the gene and observing the effects on the invasive properties of the tumour cells. Thus, genes that are decreased in expression in the metastatic cell line SW620 might be overexpressed in SW620 and inhibited in the primary tumour cell line SW480 and vice versa. Overexpression of candidate genes requires isolation of the full length cDNA and cloning into a suitable expression vector. Full length cDNA clones may be obtained from the Genbank depositor or may be commercially available from companies such as ATCC (American Type Culture Collection), TIGR (The Institute of Genomic Research) and Invitrogen. Alternatively, isolation of the full length clone can be obtained by screening a cDNA library if necessary. The cloned cDNA of interest is then transfected into either SW480 and SW620 and the effects on the metastatic properties of the cells are assessed by a modified Boyden Chamber assay (Albini *et al.*, 1987), now available commercially (Biocoat® Matrigel® Invasion Chamber, Becton Dickinson). Similarly, inhibition of specific genes may be achieved by antisense or ribozyme technologies and the effects on the cells observed as described above. Gene products whose increased expression correlates with a more invasive phenotype would be considered enhancers of

metastasis. Similarly, gene products whose increased expression correlates with a less invasive phenotype would be considered inhibitors of metastasis.

Once a particular gene has been demonstrated to have a definitive role in metastasis extensive functional studies on the encoded protein could be carried out such as localisation studies, effect of the major signalling pathways on its expression, identification of associated proteins and effects of its expression in other cell lines with different metastatic abilities. It is unlikely that single genes are responsible for metastasis but rather an accumulation of genetic changes is the likely scenario. However, altered expression of certain genes may have a strong influence on metastasis by administering 'downstream' effects on the expression of other genes when their own expression is altered.

7.4 CONCLUSIONS

Differences in gene expression determine the behaviour and characteristics of cells of different types. Therefore, an understanding of the differences between cell types will be aided by identifying alterations in their gene expression patterns. This decade has seen a revolution in techniques for identifying gene expression differences (Dunican *et al.*, 1997) where formerly they were laborious, required a large amount of starting material and yielded little reward (Benton & Davis, 1977; Grunstein & Hogness, 1977; St. John & Davis, 1979; Timberlake, 1980; Zimmerman *et al.*, 1980). The 'holy grail' of any differential expression analysis is to identify the specific genes that are involved in the phenomenon of interest. However, genes may be altered in expression for reasons other than involvement in a particular process. Thus, deciphering which changes are important and which are not is the real challenge.

The application of SAGE to an examination of colon tumour metastasis has generated a large amount of novel data. A range of genes have been identified as candidates for involvement in metastasis and the gene expression profiles generated provide a reference database for future use. Although SAGE has provided more information than any of its predecessors it is unlikely to become a routine method in most molecular biology laboratories due to the large amount of sequencing involved. The recent arrival of the commercial cDNA filters provides a quick and relatively easy alternative for examining gene expression. Further improvements in the microarray technology and completion of the human genome sequence will allow extensive gene expression analyses to be carried out in a relatively short period of time. This will allow more time and energy to be focused on identified differentially expressed gene products and their role in a particular process.

REFERENCES

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

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., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, S.M., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle Jr., P.S., Kelley, J.M., Klimek, K.M., Kelley, J.C., Liu, L.I., Marmaros, S.M., Merrick, J.M., Moreno-Palanques, R.F., McDonald, L.S., Nguyen, D.T., Pellegrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Barthlow, R., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Huddleston, K.A., Greene, J.M., Gruber, J., Hudson, P., Kim, A., Kozak, D.L., Kunsch, C., Ji, H., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. & Venter, C.J. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature Suppl.* **28**: 3-146.

Agarwal, M.L., Agarwal, A., Taylor, W.R. & Stark, G.R. (1995). p53 controls both the G2/M and the G1 cell cycle checkpoints and mediates reversible growth arrest in human fibroblasts. *Proc. Natl. Acad. Sci. USA* **92**: 8493-8497.

Ahmad, A & Hart, I.R. (1996). Biology of Tumor Micrometastasis. *J. Hematother.* **5**: 525-535.

Ahmad, A & Hart, I.R. (1997). Mechanisms of metastasis. *Crit. Rev. Oncol. Hematol.* **26**: 163-173.

Albini, A., Iwamoto, Y., Kleinman, H.K., Martin, G.R., Aaronson, S.A., Kozlowski, J.M. & McEwan, R.N. (1987). A rapid in vitro assay for quantitating the invasive potential of tumor cells. *Cancer Res.* **47**: 3239-3245.

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

Asano, K., Vornlocher, H.P., Richter-Cook, N.J., Merrick, W.C., Hinnebusch, A.G. & Hershey, J.W.B. (1997). Structure of cDNAs encoding human eukaryotic initiation factor 3 subunits. *J. Biol. Chem.* **272**: 27042-27052.

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

Aznavoorian, S., Murphy, A.N., Stetler-Stevenson, W.G. & Liotta L. A. (1993). Molecular Aspects of Tumor Cell Invasion and Metastasis. *Cancer* **71**: 1368-1383.

Baba, S., Masago, S.A., Takahashi, T., Kasama, T., Sugimura, H., Tsugane, S., Tsutsui, Y. & Shirasawa, H. (1995). A novel allelic variant of serum amyloid A, SAA1 gamma: genomic evidence, evolution, frequency, and implication as a risk factor for reactive systemic AA- amyloidosis. *Hum. Mol. Genet.* **4**: 1083-1087.

Bai., Y., Lee, D., Yu, T. and Chasin, L. A. (1999). Control of 3' splice site choice by ASF/SF2 and hnRNP A1. *Nucleic Acids Res.* **27**: 1126-1134.

Barford, D. (1996). Molecular mechanisms of the protein serine/threonine phosphatases. *Trends Biochem. Sci.* **21**: 407-412.

Barker, H.M., Brewis, N.D., Street, A.J., Spurr, N.K. & Cohen, P.T.W. (1994). Three genes for protein phosphatase 1 map to different human chromosomes: sequence, expression and gene localisation of protein serine/threonine phosphatase 1 beta (PPP1CB). *Biochim. Biophys. Acta.* **1220**: 212-218.

Barnard, G.F., Staniunas, R.J., Bao, S., Mafune, K.I., Steele, G.D., Gollan, J.L. & Chen, L.B. (1992). Increased expression of human ribosomal phosphoprotein P0 in messenger RNA in hepatocellular carcinoma in colon cancer. *Cancer Res.* **52**: 3067-3072.

Barnard, G.F., Staniunas, R.J., Mori, M., Puder, M., Jessup, M.J. Steele, G.D. & Chen, L.B. (1993). Gastric and hepatocellular carcinomas do not overexpress the same ribosomal protein messenger RNAs as colonic carcinoma. *Cancer Res.* **53**: 4048-4052.

Barrett, A.J., Fritz, H., Grubb, A., Isemura, S., Jarvinen, M., Katunuma, N., Machleidt, W., Muller-Esterl, W., Sasaki, M., & Turk, V. (1986). Nomenclature and classification of the protein homologous with the cysteine-proteinase inhibitor chicken cystatin. *Biochem. J.* **236**: 312.

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

Beach, C.M., deBeer, M.C., Sipe, J.D., Loose, L.D. & deBeer, F.C. (1992). Human serum amyloid A protein: Complete amino acid sequence of a new variant. *Biochem.J.* **282**: 615-620.

Beavon, I.R.G. (1999). Regulation of E-cadherin: does hypoxia initiate the metastatic cascade? *J. Clin. Pathol: Mol. Pathol.* **52**: 179-188.

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

Behrens, J., Boris-Alexander, J., Wurtele, M., Grimm, J., Asbrand, C., Wirtz, R., Kuhl, M., Wedlich, S. & Birchmeier, W. (1998). Functional interaction of an axin homolog, conductin, with β -catenin, APC, and GSK3 β . *Science* **280**: 596-599.

Bennett, M.A., Kay, E.W., Mulcahy, H., O'Flaherty, L., O'Donoghue, D.P., Leader, M. & Croke, D.T. (1995). ras and p53 in the prediction of survival in Dukes' stage B colorectal carcinoma. *J. Clin. Pathol: Mol. Pathol.* **48**: M310-M315.

Benton, W.D. & Davis, R.W. (1977). Screening λ gt recombinant clones by hybridization to single plaques in situ. *Science* **196**: 180-182.

Betts, J.C., Edbrooke, M.R., Thakker, R.V. and Woo, P. (1991). The human acute-phase serum amyloid A gene family: Structure, evolution and expression in hepatoma cells. *Scand. J. Immunol.*, **34**: 471-482.

Bobek, L.A., Aguirre, A. & Levine, M.J. (1991). Human salivary cystatin S. *Biochem. J.* **278**: 627-635.

Bodmer, W.F., Bailey, C.J., Bodmer, J., Bussey, H.J.R., Ellis, A., Gorman, P., Lucibello, F.C., Murday, V.A., Rider, S.H., Scambler, P., Sheer, D., Solomon, E. & Spurr, N.K. (1987). Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature (London)* **328**: 614-616.

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

Borsani, G., Ballabio, A. & Banfi, S. (1998). A practical guide to orient yourself in the labyrinth of genome databases. *Hum. Molec. Genet.* **7(10)**: 1641-1648.

Braga, V.M.M. (1999). Small GTPases and regulation of cadherin dependent cell-cell adhesion. *J. Clin. Pathol.: Mol. Pathol.* **52**: 197-202.

Bruni, R. & Roizman, B. (1996). Open reading frame P- a herpes simplex virus gene repressed during productive infection encodes a protein that binds a splicing factor and reduces synthesis of viral proteins made from spliced mRNA. *Proc. Natl. Acad. Sci. (USA)* **93**: 10423-10427.

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

Cajot, J., Sordat, I., Silvestre, T. and Sordat, B. (1997). Differential display cloning identifies Motility-related Protein (MRP1/CD9) as highly expressed in primary compared to metastatic human colon carcinoma cells. *Cancer Res.* **57**: 2593-2597.

Cao, Y., O'Reilly, M.S., Marshall, B., Flynn, E., Ji, R.W. & Folkman, J. (1998). Expression of angiostatin cDNA in a murine fibrosarcoma suppresses primary tumour growth and produces long-term dormancy of metastases. *J. Clin. Invest.* **101**: 1055-1063.

Chan, E.F., Gat, U., McNiff, J.M. & Fuchs, E. (1999). A common human skin tumour is caused by activating mutations in β -catenin. *Nat. Genet.* **21**: 410-413.

Cho, K.R., Oliner, J.D., Simons, J.W., Hedrick, L., Fearon, E.R., Preisinger, A., Hedge, P., Silverman, G.A. & Vogelstein B. (1994). The DCC gene: Structural analysis and mutations in colorectal carcinoma. *Genomics* **19**: 525-531.

Cleveland, D.W., Lopata, M.A., MacDonald, R.J., Cowan, N.J., Rutter, W.J. & Kirschner, M.W. (1980). Number and evolutionary conservation of α & β -tubulin and cytoplasmic β - & γ - genes using specific cloned cDNA probes. *Cell* **20**: 95.

Cockett, M.I., Murphy, G., Birch, M.L., O'Connell, J.P., Crabbe, T., Millican, A.T., Hart, I.R. & Docherty, A.J.P. (1998). Matrix metalloproteinases and metastatic cancer. *Biochem. Soc. Symp.* **63**: 295-313.

Cohen, P. (1989). The structure and regulation of protein phosphatases. *Annu. Rev. Biochem.* **58**: 453-508.

Cohen, P.T.W. (1997). Novel protein serine/threonine phosphatases: variety is the spice of life. *Trends Biochem. Sci.* **22**: 245-251.

Crowe, D.L., Milo, G.E. & Shuler, C.F. (1999). Keratin 19 downregulation by oral squamous cell carcinoma lines increases invasive potential. *J. Dent. Res.* **78**: 1256-1263.

Darnell, J., Lodish, H. & Baltimore, D. The cytoskeleton and cellular movements: Actin-Myosin. *In: Molecular Cell Biology*, pp. 771-814. New York: Scientific American Books, Inc., 1986.

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

Deans, G.T., Parks, T.G., Rowlands, B.J. & Spence, R.A.J. (1992). Prognostic factors in colorectal cancer. *Br. J. Surg.* **79**: 608-613.

Deb, T.B. & Datta, K. (1996). Molecular Cloning of Human Fibroblast Hyaluronic Acid-binding Protein Confirms Its Identity with P-32, a Protein Co-purified with Splicing Factor SF2. *J. Biol. Chem.* **271**: 2206-2212.

DeLeeuw, W.J.F., Slagboom, P.E. and Vijg, J. (1989). Quantitative comparison of mRNA levels in mammalian tissues: 28S ribosomal RNA level as an accurate internal control. *NucleicAcids Res.* **17**: 10137-10138.

Denis, M.G., Chadeneau, C., Lecabelle, M.T., LeMoullac, B., LeMevel, B., Meflah, K. & Lustenberger, P. (1993). *Int. J. Cancer* **55**: 275-280.

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

DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**: 680-686.

de Waard, V., van den Berg, B.M.M., Veken, J., Schultz-Heienbrok, R., Pannekoek, H. and van Zonneveld, A. (1999). Serial analysis of gene expression to assess the endothelial cell response to an atherogenic stimulus. *Gene* **226**: 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. (1996). Suppression Subtractive hybridisation: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* **93**: 6025-6030.

Dube, I.D., Kamel-Reid, S., Yuan, C.C., Lu, M., Wu, X., Corpus, G., Raimondi, S.C., Crist, W.M., Carroll, A.J., Minowada, J. & Baker, J.B. (1991). A novel human homeobox gene lies at the chromosome 10 breakpoint in lymphoid neoplasias with chromosomal translocation t(10;14). *Blood* **78**: 2996-3003.

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

Dunican, D., Parle-McDermott, A. & Croke, D.T. (1997). Gene expression profiling of cells & tissues. *Curr. Diagn. Path.* **4**: 170-175.

Duffy M J. The Biochemistry of Metastasis. (1996). *Adv. Clin. Chem.* **32**: 135-166.

Duguid, J.R. & Dinauer, M.C. (1989). Library subtraction of in vitro cDNA libraries to identify differentially expressed genes in scrapie infection. *Nucleic Acids Res.*, **18**: 2889-2792.

Drewinko, B., Romsdahl, M., Yang, L.Y., Ahearn, M.J. and Trujillo, J.M. (1976). Establishment of a human carcinoembryonic antigen-producing colon adenocarcinoma cell line. *Cancer Res.* **36**: 467-475.

Edwards-Gilbert, G., Veraldi, K. & Milcarek, C. (1997). Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* **25**: 2547-2561.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A. & Liotta, L.A. (1996). Laser Capture Microdissection. *Science* **274**: 998-1001.

Entwistle, J., Hall, C.L. and Turley, E.A. (1996). HA receptors: regulators of signalling to the cytoskeleton. *J. Cell. Biochem.* **61**: 569-577.

Fearon, E.R. & Vogelstein, B. (1990). A Genetic Model for Colorectal Tumorigenesis. *Cell* **61**: 759-767.

Fisher, E.R., Robinsky, B., Sass, R., Fisher, B. (1989). Relative prognostic value of the Dukes and Jass systems in rectal cancer: findings from the National Surgical Adjuvant Breast and Bowel Projects. *Dis. Colon Rectum* **32**: 944-949.

Fort P, Marty L, Piechaczyk M., el Sabrouty, S., Dani, C., Jeanteur, P. & Blanchard, J.M. (1985). Various rat adult tissues express only one major mRNA species from the glyceraldehyde-3-phosphate-dehydrogenase multigenic family. *Nucleic Acids Res.* **13**: 5.

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

Francia, G.F., Poulson, R., Mitchell, S.D., Marshall, J.F. & Hart, I.R. (1997). Identification of a protein phosphatase 2A regulatory subunit differentially expressed in malignant cells. *Br. J. Ca.* **75**: 1237.

Frigerio, J., Berthezene, P., Garrido, P., Ortiz, E., Barthelémy, S., Vasseur, S., Sastre, B., Seleznieff, I., Dagorn, J. and Iovanna, J.L. (1995). Analysis of 2166 clones from a human colorectal cancer cDNA library by partial sequencing. *Hum. Molec. Genet.* **4**: 37-43.

Fuchs, E. and Weber, K. (1994). Intermediate filaments: Structure, Dynamics, Function, and Disease. *Annu. Rev. Biochem.* **63**: 345-382.

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

Ghigna, C., Moroni, M., Porta, C., Riva, S. & Biamonti, G. (1998). Altered expression of heterogeneous nuclear ribonucleoproteins and SR factors in human colon adenocarcinomas. *Cancer Res.* **58**: 5818-5824.

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

Giovannucci, E. & Willett, W.C. (1994). Dietary factors and risk of colon cancer. *Ann. Med.* **26**: 443-452.

Gofuku, J., Shiozaki, H., Tsujinaka, T., Inoue, M., Tamura, S., Doki, Y., Matsui, S., Kikkawa, N. & Monden, M. (1999). Expression of E-cadherin and alpha-catenin in patients with colorectal carcinoma. Correlation with cancer invasion and metastasis. *Am. J. Clin. Pathol.* **111**: 29-37.

Goldsworthy, S.M., Goldsworthy, T.L., Sprankle, C.S. & Butterworth, B.E. (1993). Variation in expression of genes used for normalization of Northern blots after induction of cell proliferation. *Cell Prolif.* **26**: 511-517.

Gouy, M., Gautier, C., Attimonelli, M., Lanave, C., and di Paola, G. (1985) ACNUC- a portable retrieval system for nucleic acid sequence databases: logical and physical designs and usage. *CABIOS* **1**: 167-172.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M. *et al.* (1991). Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* **66**: 589-600.

Grunstein, M. & Hogness, D.S. (1977). Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* **72**: 3961-3965.

Gutsche-Perelroizen, I., Lepault, J., Ott, A. & Carlier, M.F. (1999). Filament assembly from Profilin-Actin. *J. Biol. Chem.* **274**: 6234-6243.

Hatano, M., Roberts, C.W.M., Minden, M., Crist, W.M. & Korsmeyer, S.J. (1991). Deregulation of a homeobox gene, HOX11 by the t(10;14) in T-cell leukemia. *Science* **253**: 79-82.

Hara, E., Kato, T., Nakada, S., Sekiaya, S., Oda, K. (1991). Subtractive cDNA cloning using oligo(dT)₃₀-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells. *Nucleic Acids Res.* **19**: 7097-7104.

He, M., Liu, E. & Conway, K. Semi-quantitative reverse transcriptase-polymerase chain reaction for the detection of low abundance transcripts in human cells. In "Reverse Transcriptase PCR" (J.W. Larrick & P.D. Siebert, eds.), pp. 289-299. Ellis Horwood, London, 1995.

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

Hedrick, S.M., Cohen, D.I., Nielsen, E.A., Davis, M.M. (1984). Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature*, **398**: 149-153.

Heller, R.A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E. and Davis, R.W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. USA*, **94**: 2150-2155.

Helps, N.R., Street, A.J., Elledge, S.J. & Cohen, P.T. (1994). Cloning of the complete coding region for human protein phosphatase inhibitor 2 using the two hybrid system and expression of inhibitor 2 in *E.coli*. *FEBS Lett.* **340**: 93-98.

Hermeking, H., Lengauer, C., Polyak, K., He, T.C., Zhang, L., Thiagalingam, S., Kinzler, K.W. & Vogelstein, B. (1997). 14-3-3 σ is a p53-Regulated Inhibitor of G2/M Progression. *Molec. Cell* **1**: 3-11.

Hibi, K., Liu, Q., Beaudry, G.A., Madden, S.L., Westra, W.H., Wehage, S.L., Yang, S.C., Heitmiller, R.F., Bertelsen, A.H., Sidransky, D. and Jen, J. (1998) Serial Analysis of Gene Expression in Non-Small Cell Lung Cancer. *Cancer Res.* **58**: 5690-5694.

Higashiyama, M., Taki, T., Ieki, Y., Adachi, M., Huang, C.I., Koh, T., Kodama, K., Doi, O. & Miyake, M. (1995). Reduced Motility Related Protein-1 (MRP-1/CD9) gene expression as a factor of poor prognosis in non-small lung cancer. *Cancer Res.* **55**: 6040-6044.

Hong, W.K. & Sporn, M.B. (1997). Recent advances in Chemoprevention of Cancer. *Science* **278**: 1073-1078.

Honore, B., Madsen, P., Rasmussen, H.H., Vandekerckhove, J., Celis, J.E. & Leffers, H. (1993). Cloning and expression of a cDNA covering a complete coding region of the P32 subunit of human pre-mRNA splicing factor SF2. *Gene* **134**: 283-287.

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

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

Hussebakk, A., Skogen, B., Husby, G. & Marhaug, G. (1985). Transformation of amyloid precursor SAA to protein AA and incorporation in amyloid fibrils in vivo. *Scand. J. Immunol.* **21**: 283-287.

Ikeyama, S., Koyama, M., Yamaoko, M., Sasada, R. & Miyake, M. (1993). Suppression of cell motility and metastasis by transfection with human Motility-related Protein (MRP-1/CD9) DNA. *J. Exp. Med.* **177**: 1231-1237.

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D. & Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*. **363**: 558-61.

Inomata, M., Ochiai, A., Akimoto, S., Kitano, S. & Hirohashi, S. (1996). Alteration of β -catenin expression in Colonic Epithelial Cells of Familial Adenomatous Polyposis Patients. *Cancer Res.* **56**: 2213-2217.

Isemura, S., Saitoh, E. & Sanada, K. (1984). Isolation and amino acid sequence of SAP-1, an acidic protein of human whole saliva. *J. Biochem.* **96**: 489-498.

Ishigami, S.I., Arie, S., Furutani, M., Niwano, M., Harada, T., Mizumoto, M., Mori, A., Onodera, H. & Imamura, M. (1998). Predictive value of vascular endothelial growth factor (VEGF) in metastasis and prognosis of human colorectal cancer. *Br. J. Ca.* **78**: 1379-1384.

Jennings, L.K., Fox, C.F., Kouns, W.C., McKay, C.P., Ballou, L.R. & Schultz, H.E. (1990). The activation of human platelets mediated by anti-human platelet p24/CD9 monoclonal antibodies. *J. Biol. Chem.* **265**: 3815.

Kakisako, K., Miyahara, M., Uchino, S., Adachi, Y. & Kitano, S. (1998). Prognostic significance of c-myc mRNA expression assessed by semi-quantitative RT-PCR in patients with colorectal cancer. *Oncology Reports* **5**: 441-445.

Kassam, G., Le, B.H., Choi, K.S., Kang, H.M., Fitzpatrick, S.L., Louie, P. & Waisman, D.M. (1998). The p11 subunit of the annexin II tetramer plays a key role in the stimulation of t-PA-dependent plasminogen activation. *Biochemistry* **37**: 16958-16966.

Kawabe, T., Muslin, A.J. & Korsmeyer, S.J. (1997). Hox11 interacts with protein phosphatases PP2A & PP1 & disrupts a G2/M cell-cycle checkpoint. *Nature*. **385**: 454-458.

Kennedy, M.A., Gonzalez-Sarmiento, R., Kees, U.R., Lampert, F., Dear, N., Boehm, T. & Rabbitts, T.H. (1991). HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24. *Proc. Natl. Acad. Sci. USA* **88**: 8900-8904.

Kenzelmann, M. & Muhlemann, K. (1999). Substantially enhanced cloning efficiency of SAGE (Serial Analysis of Gene Expression) by adding a heating step to the original protocol. *Nucleic Acids Res.* **27**: 917-918.

Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S.B., Pohida, T., Smith, P.D., Jiang, Y., Gooden, G.C., Trent, J.M. and Meltzer, P.S. (1998). Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.*, **58**: 5009-5013.

Kinzler, K.W. & Vogelstein, B. (1996). Lessons from Hereditary Colorectal Cancer. *Cell* **87**: 159-170.

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

Kluve-Beckermann, B., Dwulet, F.E. & Benson, M.D. (1988). Human serum amyloid A: Three hepatic mRNAs and the corresponding proteins in one person. *J. Clin. Invest.*, **82**: 1670-1675.

Kluve-Beckermann, B. & Song, M. (1995). Genes encoding human serum amyloid A proteins SAA1 and SAA2 are located 18kb apart in opposite transcriptional orientations. *Gene* **169**: 289-290.

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., Momizu, S., Konishi, F., Utsunomiya, J. & Miyaki, M. (1996). Molecular Nature of Colon Tumours in Hereditary Nonpolyposis Colon Cancer, Familial Polyposis, and Sporadic Colon Cancer. *Gastroenterology* **111**: 307-317.

Kondoh, N., Schweinfest, C.W., Henderson, K.W. and Papas, T.S. (1992). Differential expression of S19 ribosomal protein, Laminin-binding protein, and human lymphocyte antigen class I messenger RNAs as associated with colon cancer progression and differentiation. *Cancer Res.* **52**: 791-796.

Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B. & Clevers, H. (1997). Constitutive Transcriptional Activation by a β -Catenin-Tcf Complex in APC^{-/-} Colon Carcinoma. *Science* **275**: 1784-1787.

Krainer, A.R., Mayeda, A., Kozak, D. & Binns, G. (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding- proteins, U1 70K, and *Drosophila* splicing regulators. *Cell* **66**: 383-394.

Kubens, B.S. & Zanker, K.S. (1998). Differences in the migration capacity of primary human colon carcinoma cells (SW480) and their lymph node metastatic derivatives (SW620). *Cancer Lett.* **131**: 55-64.

Kumar, C.C. (1998). Signaling by integrin receptors. *Oncogene* **17**: 1365-1373.

Kumar, K.U., Srivastava, S.P. & Kaufman, R.J. (1999). Double-stranded RNA-activated protein kinase (PKR) is negatively regulated by 60s ribosomal subunit protein L18. *Mol. Cell. Biol.* **19**: 1116-125.

Kwon, Y.G., Huang, H.G., Desdouits, F., Girault, J.A., Greengard, P. & Nairn, A.C. (1997). Characterization of the interaction between DARPP-32 and protein phosphatase 1 (PP-1): DARPP-32 peptides antagonize the interaction of PP-1 with binding proteins. *Proc. Natl. Acad. Sci. USA* **94**: 3536-3541.

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

Lanza, F., Wolf, D., Fox, C.F., Kieffer, N., Seyer, J.M. Fried, V.A., Coughlin, S.R., Phillips, D.R., & Jennings, L.K. (1991). cDNA cloning and expression of platelet p24/CD9. *J. Biol. Chem.* **266**: 10638-10645.

Leibovitz, A., Stinson, J.C., McCombs, W.B., McCoy, C.E., Mazur, K.C. & Mabry, N.D. (1976). Classification of Human Colorectal Adenocarcinoma cell lines. *Cancer Res.* **36**: 4562-4569.

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

Leppert, M., Dobbs, M., Scambler, P., O'Connell, P., Nakamura, Y., Stauffer, D., Woodward, S., Burt, R., Hughes, J., Gardner, E., & White, R. (1987). The gene for familial polyposis coli maps to the long arm of chromosome 5. *Science* **238**: 1411-1443.

Levine, A.J. (1997). p53, the cellular gatekeeper for growth and division. *Cell* **88**: 323-331.

Levy, A.T., Cioce, V., Sobel, M.E., Barbisa, S., Grigioni, W.F., Liotta, L.A. & Stetlet-Stevenson, W.G. (1991). Increased expression of of the M_r 72,000 type IVcollagenase in human colonic adenocarcinoma. *Cancer Res.* **51**: 439-444.

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

Liang, P. & Pardee, A.B. (1995). Recent advances in differential display. *Curr. Opin. Immun.* **7**: 274-280..

Lichty, B.D., Ackland-Snow, J., Noble, L, Kamel-Reid, S. & Dube, I.D. (1995). Dysregulation of *HOX11* by Chromosome Translocations in T-cell Acute Lymphoblastic Leukemia: A Paradigm for Homeobox Gene Involvement in Human Cancer. *Leuk. Lymphoma* **16**: 209-215.

Lieberman, M.D., Sigal, R.K., Williams, N.N. and Daly, J.M. (1991). Natural killer cell stimulatory factor (NKSF) augments natural killer cell and antibody-dependent tumoricidal response against colon carcinomal cell lines. *J. Surg. Res.* **50**: 410-415.

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

Liu, B., Parsons, R., Papdopoulos, 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. (1996). Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nat. Med.* **2**: 169-174.

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

Lothe, R.A. (1997). Microsatellite instability in human solid tumors. *Molecul. Med. Today* **February**: 61-68.

Lottaz, D., Maurer, C.A., Hahn, D., Buchler, M.W. & Sterchi, E.E. (1999). Nonpolarized secretion of human meprin alpha in colorectal cancer generates an increased proteolytic potential in the stroma. *Cancer Res.* **59**: 1127-1133.

Lowell, C.A., Potter, D.A., Stearman, R.S. & Morrow, J.F. (1986). Transcriptional regulation of serum amyloid A gene expression. *J. Biol. Chem.* **261**: 8453-8461.

Lu, S.L., Kawabata, M., Imamura, T., Akiyama, Y., Nomizu, T., Miyazono, K. & Yuasa, Y. (1998). HNPCC associated with germline mutation in the TGF- β type II receptor gene. *Nat. Genet.* **19**: 17-18.

Luo, Y., Yu, H. and Peterlin, B.M. (1994). Cellular protein modulates effects of human immunodeficiency virus type 1 Rev. *J. Virol.* **68**: 3850-3856.

Machesky, L. M., Mullins, R.D., Higgs, H.N., Kaiser, D.A., Blanchoin, L., May, R.C., Hall, M.E. & Pollard, T.D. (1999). Scar, a WASp-related protein, activates nucleation of actin filaments by the Arp2/3 complex. *Proc. Natl. Acad. Sci. USA* **96**: 3739-3744.

MacKintosh, C., Garton, A.J., McDonnell, A., Cohen, P.T., Tonks, N.K. & Cohen, P. (1997). Further evidence that inhibitor-2 acts like a chaperone to fold PP1 into its native conformation. *FEBS Lett.* **397**: 235-238.

Madden, S.L., Galella, E.A., Zhu, J., Bertelsen, A.H. & Beaudry, G.A. (1997). SAGE transcript profiles for p53-dependent growth regulation. *Oncogene* **15**: 1079-1085.

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

Manley, J.L. & Tacke, R. (1996). SR proteins and splicing control. *Genes Dev.* **10**: 1569-1579.

Mann, B., Gelos, M., Siedow, A., Hanski, M.L., Gratchev, A., Ilyas, M., Bodmer, W.F., Moyer, M.P., Riecken, E.O., Buhr, H.J. & Hanski, C. (1999). Target genes of β -catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. *Proc. Natl. Acad. Sci. USA* **96**: 1603-1608.

Markwell, M.A.K., Haas, S.M., Bieber, L.L. & Tolbert, N.E. (1978). A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. *Anal. Biochem.* **87**: 206-210.

Masuda, H. & Aoki, H. (1999). Host expression of matrix metalloproteinase-2 and tissue inhibitor of metalloproteinase-2 in normal colon tissue affects metastatic potential of colorectal cancer. *Dis. Colon Rectum* **42**: 393-397.

Matthews, D.A. and Russell, W.C. (1998). Adenovirus core protein Vinteracts with p32- a protein which is associated with both the mitochondria and the nucleus. *J. Gen. Virol.* **79**: 1677-1685.

Mayeda, A. & Krainer, A.R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and Splicing Factor SF2. *Cell* **68**: 365-375.

McDonald, L.A., Walker, D.M. and Gibbins, J.R. (1998). Cervical lymph node involvement in head and neck cancer detectable as expression of a spliced transcript of type II keratin K5. *Oral Oncol.* **34**: 276-283.

McDonnell, S.E., Navre, M., Coffey, R.J., Matrisian, L.M. (1991). Expression and localisation of the matrix metalloproteinase pump-1 (MMP-7) in human gastric and colon carcinomas. *Molec. Carcinogen.* **4**: 527-533.

McWilliam, P., Parle-McDermott, A., Dunican, D., Tighe, O. & Croke, D.T. (1999). Characterisation of clones isolated by Differential Display RT-PCR from colon cancer cell lines SW480 and SW620. *Abstract. Ir. J. Med. Sci.* **168 suppl. 1**: 44.

McWilliam, P., Parle-McDermott, A., Dunican, D., Tighe, O., Troy, A., Sexton, D.M., O'Donoghue, D.P., Sheahan, K. & Croke, D.T. Inhibitor-2 of Protein Phosphatase 1 is differentially expressed between the colon cancer cell lines SW480 & SW620. (In Preparation).

Mehlen, P., Rabizadeh, S., Snipas, S.J., Assa-Munt, N., Salvesen, G.S. & Bredesen, D.E. (1998). The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis. *Nature* **395**: 801-804.

Miyake, M., Koyama, M., Seno, M. & Ikeyama, S. (1991). Identification of the motility-related protein (MRP-1), recognized by monoclonal antibody M31-15, which inhibits cell motility. *J. Exp. Med.* **174**: 1347-1354.

Miyake, M., Nakano, K., Ieki, Y., Adachi, M., Huang, C.L., Itoi, S.I., Koh, T & Taki, T. (1995). Motility Related Protein 1 (MRP-1/CD9) expression: inverse correlation with metastases in breast cancer. *Cancer Res.* **55**: 4127-4131.

Miyake, M., Nakano, K., Itoi, S.I., Koh, T. & Taki, T. (1996). Motility-related Protein-1 (MRP-1/CD9) reduction as a factor of poor prognosis in breast cancer. *Cancer Res.* **56**: 1244-1249.

Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996). XTcf3 transcription factor mediates beta-catenin induced axis formation in *Xenopus* embryos. *Cell* **86**: 391-399.

Mori, M., Mimori, K., Shiraishi, T., Haraguchi, M., Ueo, H., Barnard, G.F. & Akiyoshi, T. (1998). Motility related protein 1 (MRP-1/CD9) expression in colon cancer. *Clin.Cancer Res.* **4**: 1507-1510.

Morin, J.P., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., Kinzler, K.W. (1997). Activation of β -Catenin-Tcf Signaling in Colon Cancer by Mutations in β -Catenin or APC. *Science* **275**: 1787-1789.

Morson, B.C. & Dawson, I.M.P. "Gastrointestinal Pathology." Blackwell Scientific Publications, England, 1990.

Mou, L., Miller, H., Li, J., Wang, E. & Chalifour, L. (1994). Improvements of the differential display method for gene analysis. *Biochem. Biophys. Res. Commun.* **199**: 564-569.

Muta, T., Kang, D., Kitajima, S., Fujiwara, T. and Hamasaki, N. (1997). p32 protein, a Splicing Factor 2-associated protein, is localized in mitochondrial matrix and is functionally important in maintaining oxidative phosphorylation. *J. Biol. Chem.* **272**: 24363-24370.

Nakagawa, T.Y., Swanson, M.S., Wold, B.J., & Dreyfuss, G. (1986) Molecular cloning of cDNA for the nuclear ribonucleoprotein particle C proteins: A conserved gene family. *Proc. Natl. Acad. Sci. USA* **83**: 2007-2011.

Namikawa, R. & Shtivelman, E. (1999). SCID-hu mice for the study of human cancer metastasis. *Cancer Chemother. Pharmacol.* **43 Suppl**: S37-41.

Nathke, I.S. (1999). The adenomatous polyposis coli protein. *J. Clin. Pathol: Mol. Pathol.* **52**: 169-173.

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. & Roninson, I.B. (1990). Quantitative analysis of MDR1 (multidrug resistance) gene expression in human tumours by polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **87**: 7160-7164.

Nowak, R. (1995). Entering the postgenome era. *Science* **270**: 368-371.

Olsson, N., Siegbahn, A. and Nilsson, G. (1999). Serum amyloid A induces chemotaxis of human mast cells by activating a pertussis toxin-sensitive signal transduction pathway. *Biochem. Biophys. Res. Commun.* **254**: 143-146.

O'Reilly, M.S., Homgren, L., Shing, Y., Chen, C., Rosenthal, R.A., Moses, M., Lane, W.S., Cao, Y., Sage, E.H. & Folkman, J. (1994). Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by Lewis lung carcinoma. *Cell* **79**: 315-328.

Oshima, R.G. (1992). Intermediate filament molecular biology. *Curr. Opin. Cell Biol.* **4**: 110-116.

Ou, J.H., Yen, T.S.B., Wang, Y.F., Kam, W.K. & Rutter, W.J. (1987). Cloning and characterisation of a human ribosomal-protein genes with enhanced expression in fetal and neoplastic cells. *Nucleic Acids Res.* **15**: 8919-8934.

Paine, M.L., Gibbins, J.R., Choi, J.K., McDonald, D.A., Manthey, A.M., Walker, D.M. and Kefford, R.F. (1995). Intracellular post-transcriptional down-regulation responsible for loss of a keratin differentiation marker in tumour progression. *Anticancer Res.* **15**: 2145-2154.

Paredes-Zaglul, A., Kang, J.J., Essig, Y.P., Mao, W., Irby, R., Wloch, M. & Yeatman, T.J. (1998). Analysis of colorectal cancer by comparative genomic hybridization: evidence for induction of the metastatic phenotype by loss of tumour suppressor genes. *Clin. Cancer Res.* **4**: 879-886.

Parle-McDermott, A., McWilliam, P., Tighe, O., Dunican, D. & Croke, D.T. (1998). Hox11 is differentially expressed between the colon cancer cell lines SW480 and SW620. *Br. J. Ca.* **78 suppl. 1**: 54.

Pennacchio, L.A., Bouley, D.M., Higgins, K.M., Scott, M.P., Noebels, J.L. & Myers, R.M. (1998). Progressive ataxia, myoclonic epilepsy and cerebellar apoptosis in cystatin B-deficient mice. *Nat. Genet.* **20**: 251-258.

Perry, I., Hardy, R., Jones, T. & Jankowski, J. (1999). A colorectal cell line with alterations in E-cadherin and epithelial biology may be an in vitro model of colitis. *J. Clin. Pathol: Mol. Pathol.* **52**: 231-242.

Petersen-Mahrt, S.K., Estmer, C., Ohrmalm, C., Matthews, D.A., Russell, W.C., and Akusjarvi G. (1999). The splicing factor-associated protein, p32, regulates RNA splicing by inhibiting ASF/SF2 RNA binding and phosphorylation. *EMBO J.* **18**: 1014-1024.

Pinol-Roma, S., Swanson, M.S., Gall, J.G. & Dreyfuss, G. (1989). A Novel Heterogeneous Nuclear RNP Protein with a Unique Distribution on Nascent Transcripts. *J. Cell Biol.* **109** : 2575-2587.

Pogue-Geile, K., Geiser, J.R., Shu, M., Miller, C., Wool, I.G., Meisler, A.I & Pipas, J.M. (1991). Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. *Mol. Cell Biol.* **11**: 3842-3849.

Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W, & Vogelstein, B. (1997). A model for p53-induced apoptosis. *Nature* **389**: 300-305.

Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J.K.V., Markowitz, S.D., Trush, M.A., Kinzler, K.W. & Vogelstein, B. (1998). Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat. Genet.* **20**: 291-293.

Powell, J. (1998). Enhanced concatamer cloning- a modification to the SAGE (Serial Analysis of Gene Expression) technique. *Nucleic Acids Res.* **26**: 3445-3446.

Prolla, T.A., Baker, S.M., Harris, A.C., Tsao, J.L., Yao, X., Bronner, C.E., Zheng, B., Gordon, M., Reneker, J., Arnheim, N., Shibata, D., Bradley, A. & Liskay, R.M. (1998). Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. *Nat. Genet.* **18**: 276-279.

Preciado-Patt, L., Hershkovich, R., Fridkin, M. and Lider, O. (1996). Serum Amyloid A Binds Specific Extracellular Matrix Glycoproteins and Induces the Adhesion of Resting CD4 T Cells. *J. Immun.* **156**: 1189-1195.

Rainbow, R.D. Immunohistochemistry. In "Laboratory Histopathology: A complete reference." (Woods, A.E. & Ellis, R.C., eds.). Churchill Livingstone, London, Longman Group Limited, 1994.

Ray, A., Hannink, M. & Ray, B.K. (1995). Concerted participation of NF- κ B and C/EBP heteromer in lipopolysaccharide induction of serum amyloid A gene expression in liver. *J. Biol. Chem.* **270**: 7365-7374.

Ray, A. & Ray, B.K. (1998). Isolation and functional characterization of cDNA of Serum Amyloid A-Activating Factor that binds to the Serum Amyloid A promoter. *Mol. Cell. Biol.* **18**: 7327-7335.

Riley, H.D., Macnab, J., Farrell, T.J. & Cohn, K. (1997). The expression of acylphosphatase is associated with the metastatic phenotype in human colorectal tumors. *Carcinogenesis* **18**: 2453-2455.

Rooney, P., Kumar, S., Ponting, J. & Wang, M. (1995). The role of Hyaluronan in tumour neovascularization (Review). *Int. J. Cancer* **60**: 632-636.

Roose, J., Huls, G., van Beest, M., Moerer, P., van der Horn, K., Goldschmeding, R., Logtenberg, T. & Clevers, H. (1999). Synergy between tumor suppressor APC and the b-catenin-Tcf4 target Tcf1. *Science* **285**: 1923-1926.

Ropponen, K., Tammi, M., Parkkinen, J., Eskelinen, M., Tammi, R., Lipponen, P., Agren, U., Alhava, E. and Kosma, V.M. (1998). Tumor cell-associated hyaluronan as an unfavorable prognostic factor in colorectal cancer. *Cancer Res.* **58**: 342-347.

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

Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, E. & Polakis, P. (1997). Stabilization of b-Catenin by Genetic Defects in Melanoma Cell Lines. *Science.* **275**: 1790-1792.

Saito, M., Yamaguchi, A., Goi, T., Tsuchiyama, T., Nakagawara, G., Urano, t., Shiku, H. & Furukawa, K. (1999). Expression of DCC protein in colorectal tumors and its relationship to tumor progression and metastasis. *Oncology* **56**: 134-141.

Saitoh, E., Minaguchi, K. & Ishibashi, O. (1998). Production and characterisation of two variants of human cystatin SA encoded by two alleles at the CST2 locus of the type 2 cystatin gene family. *Arch. Biochem. Biophys.* **352**: 199-206.

Sambrook, J., Fritsh, 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. (1996). Colonization of human lung grafts in SCID-hu mice by human colon carcinoma cells. *Int. J. Cancer* **65**: 864-869.

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

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

Sato, T., Saito, H., Swensen, J., Olifant, A., Wood, C., Danner, D., Sakamoto, T., Takita, K., Kasumi, F. & Miki, Y. (1992). The human prohibitin gene located on chromosome 17q21 is mutated in sporadic breast cancer. *Cancer Res.* **52**: 1643-1646.

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

Schwartz, D. & Rotter, V. (1998). p53-dependent cell cycle control: response to genotoxic stress. *Semin. Cancer Biol.* **8**: 325-336.

Scott, N. & Quirke, P. (1993). Molecular biology of colorectal neoplasia. *Gut.* **34**: 289-292.

Seeling, J.M., Miller, J.R., Gil, R., Moon, R.T., White, R., Virshup, D.M. (1999). Regulation of β -catenin signaling by the B56 subunit of Protein Phosphatase 2A. *Science* **283**: 2089-2091.

Semple, T.U., Quinn, L.A., Woods, L.K. and Moore, G.E. (1978). Tumor and Lymphoid cell lines from a patient with carcinoma of the colon for a cytotoxicity model. *Cancer Res.* **38**: 1345-1355.

Sexton, P.S. & Cox, J.L. (1997). Inhibition of motility and invasion of B16 melanoma by the overexpression of cystatin C. *Melanoma Res.* **7**: 97-101.

Sheng, H., Shao, J., Williams, C.S., Pereira, M.A., Taketo, M.M., Oshima, M., Reynolds, A.B., Washington, M.K., DuBois, R.N. & Beauchamp, R.D. (1998). Nuclear translocation of β -catenin in hereditary and carcinogen-induced intestinal adenomas. *Carcinogenesis* **19**: 543-549.

Sherbet, G.V. & Lakshmi, M.S. "The Genetics of Cancer: genes associated with cancer invasion, metastasis, and cell proliferation." Academic Press, UK, 1997.

Shinoki, N., Sakon, M., Kambayashi, J., Ikeda, M., Oiki, E., Okuyama, M., Fujitani, K., Yano, Y., Kawasaki, T. & Monden, M. (1995). Involvement of protein phosphatase-1 in cytoskeletal organisation of cultured endothelial cells. *J. Cell. Biochem.* **59**: 368-375.

Shiraishi, T., Mori, M., Tanaka, S., Sugimachi, K., Akiyoshi, T. (1998). Identification of cystatin B in human esophageal carcinoma, using differential displays in which the gene expression is related to lymph-node metastasis. *Int. J. Cancer* **79**: 175-178.

Simone, N.L., Bonner, R.F., Gillespie, J.W., Emmert-Buck, M.R. & Liotta, L.A. (1998). Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet.* **14**: 272-276.

Simos, G. and Georgatos, S.D. (1994). The lamin B receptor-associated protein p34 shares sequence homology and antigenic determinants with the splicing factor-2 associated protein p32. *FEBS Lett.* **346**: 225-228.

Simpson, W.C. & Mayo, C. W. (1939). The mural penetration of the carcinoma cell in the colon: anatomic and clinical study. *Surg. Gynecol. Obstet.* **68**: 872-877.

Sipe, J.D., Colten, H.R., Goldberger, G., Edge, M.D., Tack, B.F., Cohen, A.S. and Whitehead, A.S. (1985). Serum amyloid A (SAA): biosynthesis and post-synthetic processing of preSAA and structural variants defined by complementary DNA. *Biochemistry* **24**: 2931-2936.

Smith, K.J., Johnson, K.A., Bryan, T.M., Hill, D.E., Markowitz, S., Willson, J.K.V., Paraskeva, C., Petersen, G.M., Hamilton, S.R., Vogelstein, B. & Kinzler, K.W. (1993). The APC gene product in normal and tumour cells. *Proc. Natl. Acad. Sci. USA* **90**: 2846-2850.

Smith, R.G. (1994). Southwestern Internal Medicine Conference: Hereditary Predisposition to Colorectal Cancer: New Insights. *Am. J. Med.Sci.* **308**: 295-306.

Sobel, M.E. (1990). Metastasis Suppressor Genes. *J. Natl. Cancer Instit. (USA)* **82**: 267-276.

Sobin, L.H., Hermanek, P. & Hutter, R.V.P. (1988). TNM Classification of Malignant Tumors. *Cancer*. **61**: 2310-2314.

Solomon, E., Voss, R., Hall, V., Bodmer, W.F., Jass, J.R., Jeffreys, A.J., Lucibello, F.C., Patel, I. & Rider, S.H. (1987). Chromosome 5 allele loss in human colorectal carcinomas. *Nature*. **328**: 616-619.

Sorensen, C.B., Ladekjaer-Mikkelsen, A.S., Andresen, B.S., Brandrup, F., Veien, N.K., Buus, S.K., Anton-Lamprecht, I., Kruse, T.A., Jensen, P.K., Eiberg, H., Bolund, L. and Gregersen, N. (1999). Identification of novel and known mutations in the genes for keratin 5 and 14 in Danish patients with epidermolysis bullosa simplex: correlation between genotype and phenotype. *J. Invest. Dermatol.* **112**: 184-90.

Sotiropoulou, G., Anisowicz, A. & Sager, R. (1997). Identification, cloning and characterization of Cystatin M, a novel cysteine proteinase inhibitor, down-regulated in breast cancer. *J. Biol. Chem.* **272**: 903-910.

Spirio, L.N., Samowitz, W., Robertson, J., Robertson, M., Burt, R.W., Leppert, M. & White, R. (1998). Alleles of APC modulate the frequency and classes of mutations that lead to colon polyps. *Nat. Genet.* **20**: 385-388.

Steeg, S.P., Bevilacqua, G., Pozzatti, R., Liotta, L.A. & Sobel, M.E. (1988). Altered Expression of *NM23*, a Gene Associated with Low Tumor Metastatic Potential, during Adenovirus 2 *Ela* Inhibition of Experimental Metastasis. *Cancer Res.* **48**: 6550-6554.

Steel, D.M., Sellar, G.C., Uhlar, C.M., Simon, S., DeBeer, F.C. & Whitehead, A.S. (1993). A constitutively expressed serum amyloid A protein gene (SAA4) is closely linked to, and shares structural similarities with an acute phase serum amyloid A protein gene (SAA2). *Genomics* **16**: 447-454.

St. John, T.P. & Davis, R.W. (1979). Isolation of galactose-inducible DNA sequences from *Saccharomyces cerevisiae* by differential plaque filter hybridisation. *Cell* **16**: 443-452.

Strachan, A.F., Shephard, E.G., Bellstedt, D.U., Coetzee, G.A., van der Westhuyzen, D.R. & de Beer, F.C. (1989). Human serum amyloid A protein. *Biochem. J.* **263**: 365-370.

Strachan, T., Abitbol, M., Davidson, D. & Beckmann, J.S. (1997). A new dimension for the human genome project: towards comprehensive expression maps. *Nat. Genet.* **16**: 126-132.

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

Stryer, L. "Biochemistry." 3rd edition. ,W.H. Freeman & co., New York, 1988.

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

Su, S.B., Gong, W., Gao, J.L., Shen, W., Murphy, P.M., Oppenheim, J.J. and Wang, J.M. (1999). A seven-transmembrane G protein-coupled receptor, FPRL1, mediates the chemotactic activity of serum amyloid A for human phagocytic cells. *J. Exp. Med.* **189**: 395-402.

Sutherland, J.D. & Witke, W. (1999). Molecular genetic approaches to understanding the actin cytoskeleton. *Curr.Opin.in Cell Biol.***11**: 142-151.

Swanson, M.S., Nakagawa, T.Y., LeVan, K. & Dreyfuss, G. (1987). Primary Structure of Human Nuclear Ribonucleoprotein Particle C Proteins: Conservation of Sequence and Domain Structures in Heterogeneous Nuclear RNA, mRNA, and Pre-rRNA-Binding Proteins. *Mol. Cell. Biol.* **7**: 1731-1739.

Takeichi, M. (1990). Cadherins: a molecular family important in selective cell-cell adhesion. *Annu. Rev. Biochem.* **59**: 237-52.

Tang, S. & Breitman, M.L. (1995). The optimal binding sequence of the Hox11 protein contains a predicted recognition core motif. *Nucleic Acids Res* **23**: 1928-1935.

Tange, T.O., Jensen, T.H. and Kjems, J. (1996). In vitro interaction between human immunodeficiency virus type 1 rev protein and splicing factor ASF/SF2-associated protein, p32. *J. Biol. Chem.* **271**: 10066-10072.

Tanigawa, N., Amaya, H., Matsumura, M., Lu, C., Kitaoka, A., Matsuyama, K. & Muraoka, R. (1997). Tumor angiogenesis and mode of metastasis in patients with colorectal cancer. *Cancer Res.* **57**: 1043-1046.

The European Polycystic Kidney Disease Consortium. (1994). The Polycystic Kidney Disease 1 Gene Encodes a 14kb Transcript & Lies within a Duplicated Region on Chromosome 16. *Cell* **77**: 881-894.

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

Thibodeau, S.N., Bren, G. & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* **260**: 816-9.

Timberlake, W.E. (1980). Developmental gene regulation in *Aspergillus nidulans*. *Develop. Biol.*, **78**: 497-510.

Toh, Y., Oki, E., Oda, S., Tokunaga, E., Ohno, S., Maehara, Y., Nicolson, G.L. & Sugimachi, K. (1997). Overexpression of the *MTAI* gene in gastrointestinal carcinomas: correlation with invasion & metastasis. *Int. J. Cancer (Pred. Oncol.)* **74**: 459-463.

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

Uchida, S., Shimada, Y., Watanabe, G., Gang Li, Z., Hong, T. Miyake, M. & Imamura, M. (1999). Motility-related protein (MRP-1/CD9) and KAI1/CD82 expression inversely correlate with lymph node metastasis in oesophageal squamous cell carcinoma. *Br. J. Cancer* **79**: 1168-1173.

Urieli-Shoval, S., Cohen, P., Eisenberg, S. & Matzner, Y. (1998). Widespread expression of serum amyloid A in histoologically normal human tissues. Predominant localization to the epithelium. *J. Histochem. Cytochem.* **46**: 1377-1384.

Vasen, H.F.A., Mecklin, J.P., Khan, P.M. & Lynch, H.T. (1990). The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (HNPCC). *Dis. Colon Rect.* **34**: 424-425.

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

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1997a). Serial analysis of gene expression: detailed protocol (Version 1.0b), November 1995. Available from Johns Hopkins Oncology Centre and Howard Hughes Medical Institute, 424 North Bond Street, Baltimore, MD 21231, USA; Fax: +1 410 955 0548.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1997b). Serial analysis of gene expression: detailed protocol (Version 1.0c), September 1997. Available from Johns Hopkins Oncology Centre and Howard Hughes Medical Institute, 424 North Bond Street, Baltimore, MD 21231, USA; Fax: +1 410 955 0548.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E, Jr., Hieter, P., Vogelstein, B. & Kinzler, K.W. (1997c). Characterization of the Yeast Transcriptome. *Cell* **88**: 243-251.

Volkov, Y., Long, A. & Kelleher, D. (1998). Inside the crawling T cell: Leukocyte function-associated Antigen-1 cross-linking is associated with microtubule-directed translocation of Protein Kinase C isoenzymes β (I) and δ . *J. Immunol.* **161**: 6487-6495.

Wahle, E. & Keller, W. (1992). The biochemistry of 3' -end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.*, **61**: 419-440.

Waldman, T., Kinzler, K.W. & Vogelstein, B. (1995). p21 is necessary for the p53-mediated G1 arrest in human cancer cells. *Cancer Res.* **55**: 5187-5190.

Wang, Y.L., Finan, J.E, Middeldorp, J.M. and Hayward, S.D. (1997). p32/TAP, a cellular protein that interacts with EBNA-1 of Epstein-Barr virus. *Virology* **236**: 18-29.

Wang, Y., Cheong, D., Chan, S. & Hooi, S.C. (1999). Heparin/heparan sulfate interacting protein gene expression is up-regulated in human colorectal carcinoma and correlated with differentiation status and metastasis. *Cancer Res.* **59**: 2989-2994.

Weg-Remers, S., Anders, M., von Lampe, B., Riecken, E.Ol, Schuder, G., Feifel, G., Zeitz, M. & Stallmach, A. (1998). Decreased expression of CD44 splicing variants in advanced colorectal carcinomas. *Eur. J. Cancer* **34**: 1607-1611.

Wei, Y. & Mojsov, S. (1996). Tissue specific expression of different human receptor types for Pituitary Adenylate Cyclase Activating Polypeptide and Vasoactive Intestinal Polypeptide: implications for their role in human physiology. *J. Neuroendocrinol.* **8**: 811-817.

Wieland, I., Bolger, G., Asouline, G. & Wigler, M. (1990). A method for difference cloning: Gene amplification following subtractive hybridization. *Proc. Natl. Acad. Sci. USA*, **87**: 2720-2724.

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

Woo, P., Sipe, J., Dinarello, C.A. & Colten, H.R. (1987). Structure of a human serum amyloid A gene and modulation of its expression in transfected L cells. *J. Biol. Chem.* **262**: 15790-15795.

Wool, I.G. (1996). Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* **21**: 164-165.

Wynford-Thomas, D. (1997). Proliferative lifespan checkpoints: cell-type specificity and influence on tumour biology. *Eur. J. Cancer* **33**: 716-726.

Xiao, S.H. and Manley, J. L. (1998). Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J.* **17**: 6359-6367.

Yamamoto, H., Itoh, F., Hinoda, Y. & Imai, K. (1996). Inverse Association of Cell Adhesion Regulator Messenger RNA Expression with Metastasis in Human Colorectal Cancer. *Cancer Res.* **56**: 3605-3609.

Yamane, N., Tsujitani, S., Makino, M., Maeta, M. & Kaibara, N. (1999). Soluble CD44 variant 6 as a prognostic indicator in patients with colorectal cancer. *Oncology* **56**: 232-238.

Young, M.R.I., Lozano, Y., Djordjevic, A. & Maier, G.D. (1993). Protein phosphatases limit tumor motility. *Int. J. Cancer* **54**: 1036-1041.

Yu, L., Loewenstein, P.M., Zhang, Z. and Green, M. (1995). *In vitro* interaction of the human immunodeficiency virus type 1 Tat transactivator and the general transcription factor TFIIB with the cellular protein TAP. *J. Virol.* **69**: 3017-3023.

Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B. & Kinzler, K.W. (1997). Gene Expression Profiles in Normal and Cancer Cells. *Science* **276**: 1268-1272.

Zimmerman, C.R., Orr, W.C., Leclerc, R.F., Barnard, E.C. & Timberlake, W.E. (1980). Molecular cloning and selection of genes regulated in *Aspergillus* development. *Cell* **21**: 709-715.

APPENDICES

The appendices are arranged in order of chapter and are given the same number as the figure they are referring to.

CHAPTER 4

APPENDIX 4.2 Keratin K5 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

APPENDIX 4.3 Cystatin S semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

APPENDIX 4.5 SAA semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

APPENDIX 4.6 S28 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

APPENDIX 4.7 SFp32 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

APPENDIX 4.8 Myosin light chain semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

APPENDIX 4.10 SSRP1 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

APPENDIX 4.11 β -tubulin semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

APPENDIX 4.13 BLAST search of PACAP type-3/VIP type-2 receptor sequence (U18810).

APPENDIX 4.16 SAA semi-quantitative RT-PCR data normalised to β -actin in all seven cell lines.

APPENDIX 4.17 S28 semi-quantitative RT-PCR data normalised to β -actin in all seven cell lines.

APPENDIX 4.18 SFp32 semi-quantitative RT-PCR data normalised to β -actin in all seven cell lines.

CHAPTER 5

APPENDIX 5.2 Nucleotide sequence contig of the extended 3' UTR of IPP-2.

APPENDIX 5.4 IPP-2 Northern blot densitometry data normalised to β -actin in SW480 and SW620.

APPENDIX 5.6 Hox11 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620.

CHAPTER 6

APPENDIX 6.6/7 Keratin 5/8 immunohistochemical data.

APPENDIX 4.2 Keratin K5 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 4.2].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= Keratin K5 PCR cycle number sampled. Mean= average β -actin/Keratin K5 ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
18	0.127	0.0033	3	0.000	-	3
20	0.193	0.0033	3	0.000	-	3
22	0.357	0.0033	3	0.000	-	3
20	0.363	0.0088	3	0.000	-	3
22	0.407	0.0219	3	0.000	-	3
24	0.550	0.666	3	0.000	-	3

APPENDIX 4.3 Cystatin S semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 4.3].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpressTM automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= Cystatin S PCR cycle number sampled. Mean= average β -actin/Cystatin S ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
24	0.067	0.0033	3	0.000	-	3
26	0.243	0.0467	3	0.000	-	3
18	0.150	0.0000	3	0.000	-	3
20	0.237	0.0033	3	0.000	-	3
22	0.487	0.0033	3	0.000	-	3

APPENDIX 4.5 SAA semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 4.5].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= SAA PCR cycle number sampled. Mean= average β -actin/SAA ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
20	0.250	0.000	3	0.000	-	3
22	0.383	0.045	3	0.000	-	3
24	0.673	0.067	3	0.000	-	3
20	0.387	0.018	3	0.000	-	3
22	0.393	0.007	3	0.000	-	3
24	0.547	0.033	3	0.000	-	3
18	0.177	0.007	3	0.000	-	3

APPENDIX 4.6 S28 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 4.6].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= S28 PCR cycle number sampled. Mean= average β -actin/S28 ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
16	0.010	0.010	3	0.083	0.0033	3
18	0.050	0.000	3	0.133	0.0033	3
20	0.137	0.003	3	0.317	0.0088	3
16	0.073	0.003	3	0.207	0.0033	3
18	0.113	0.003	3	0.247	0.0033	3
20	0.247	0.018	3	0.437	0.0291	3

APPENDIX 4.7 SFp32 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 4.7].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= SFp32 PCR cycle number sampled. Mean= average β -actin/SFp32 ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
18	0.040	0.001	3	0.107	0.003	3
20	0.063	0.000	3	0.143	0.003	3
22	0.177	0.032	3	0.310	0.025	3
20	0.087	0.007	3	0.163	0.003	3
22	0.207	0.007	3	0.387	0.018	3
18	0.057	0.003	3	0.110	0.000	3
20	0.087	.007	3	0.150	0.006	3
22	0.197	0.022	3	0.307	0.037	3

APPENDIX 4.8 Myosin light chain semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 4.8].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= Myosin light chain PCR cycle number sampled. Mean= average β -actin/Myosin light chain ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
18	0.063	0.0033	3	0.047	0.0033	3
20	0.097	0.0067	3	0.067	0.0067	3
22	0.373	0.0484	3	0.247	0.0367	3
20	0.160	0.0100	3	0.147	0.0088	3
22	0.413	0.0033	3	0.310	0.0100	3
18	0.070	0.0306	3	0.087	0.0033	3
20	0.160	0.0000	3	0.130	0.0000	3
22	0.443	0.0491	3	0.337	0.0203	3
23	0.320	0.0058	3	0.330	0.0200	3
26	0.857	0.0033	3	1.017	0.0441	3
20	0.110	0.0000	3	0.093	0.0033	3
23	0.350	0.0100	3	0.250	0.0058	3

APPENDIX 4.10 SSRP1 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 4.10].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= SSRP1 PCR cycle number sampled. Mean= average β -actin/SSRP1 ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
20	0.073	0.0033	3	0.087	0.0033	3
22	0.143	0.0088	3	0.147	0.0067	3
18	0.050	0.0000	3	0.067	0.0033	3
20	0.077	0.0033	3	0.090	0.0000	3
22	0.177	0.0033	3	0.220	0.0265	3
23	0.153	0.0033	3	0.297	0.0033	3
26	0.593	0.0240	3	0.743	0.0033	3

APPENDIX 4.11 β -tubulin semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 4.11].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= β -tubulin PCR cycle number sampled. Mean= average β -actin/ β -tubulin ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
18	0.503	0.0088	3	0.753	0.0033	3
20	0.590	0.0153	3	0.683	0.0240	3
22	0.693	0.0731	3	0.643	0.0384	3
18	0.650	0.0252	3	0.833	0.0120	3
20	0.713	0.0433	3	0.793	0.0384	3
22	0.727	0.0120	3	0.843	0.1033	3
18	0.137	0.0088	3	0.187	0.0088	3
20	0.243	0.0088	3	0.280	0.0058	3
22	0.493	0.0120	3	0.523	0.0133	3

Appendix 4.13 [Figure 4.13] BLAST search of PACAP type-3/VIP type-2 receptor sequence (U18810)

gblAF014893.1|AF014893 Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
gene encoding mitochondrial protein, partial cds
Length = 1041

Score = 252 bits (127), Expect = 3e-64
Identities = 146/151 (96%), Gaps = 1/151 (0%)
Strand = Plus / Plus

```
Query: 1668 actcaacttaaactccagcaccacgaccctgctgctatctcgcacctgaaacaagctaac 1727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 654  actcaacttaaactccagcaccacgaccctactactatctcgcacctgaaacaagctaac 713
```

```
Query: 1728 atgactaacacccttaattccatccaccctcctctccctaggaggcctgcgcccgctaac 1787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 714  atgactaacacccttaattccatccaccctcctctccctaggaggcctgcccccgctaac 773
```

```
Query: 1788 cggctttttt-cgcaaattgggccattatcgaa 1817
          ||||||||| | ||||||||||||||||
Sbjct: 774  cggcttttttgcccaaattgggccattatcgaa 804
```


APPENDIX 4.16 SAA semi-quantitative RT-PCR data normalised to β -actin in all seven cell lines [Figure 4.16].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle= SAA PCR cycle number sampled. Mean= average β -actin/SAA ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle	SW480			SW620			LoVo			Colo 201			T84		
	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n
20	0.457	0.0033	3	0.000	-	3	0.000	-	3	0.057	0.0033	3	0.000	-	3
25	3.743	0.0521	3	0.000	-	3	0.263	0.007	3	0.533	0.0333	3	0.000	-	3

Cycle	SW1116			498LI		
	Mean	SEM	n	Mean	SEM	n
23	0.720	0.032	3	0.380	0.0058	3
26	1.540	0.023	3	1.237	0.0939	3
22	0.157	0.003	3	0.117	0.0033	3
25	0.390	0.000	3	0.320	0.000	3
28	1.110	0.010	3	0.973	0.0186	3

APPENDIX 4.17 S28 semi-quantitative RT-PCR data normalised to β -actin in all seven cell lines [Figure 4.17].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle= S28 PCR cycle number sampled. Mean= average β -actin/S28 ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle	SW480			SW620			LoVo			Colo 201			T84		
	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n
18	0.323	0.0067	3	0.930	0.015	3	2.660	0.055	3	0.437	0.0088	3	0.287	0.012	3
18	0.180	0.0000	3	0.403	0.007	3	1.200	0.015	3	0.123	0.0267	3	0.080	0.000	3
20	0.310	0.1000	3	0.473	0.007	3	1.173	0.022	3	0.180	0.0058	3	0.130	0.006	3

Cycle	SW1116			498LI		
	Mean	SEM	n	Mean	SEM	n
18	0.220	0.015	3	0.143	0.0033	3
20	0.327	0.007	3	0.240	0.0153	3
16	0.140	0.000	3	0.097	0.0088	3
18	0.213	0.019	3	0.153	0.0133	3
20	0.458	0.043	3	0.313	0.549	3

APPENDIX 4.18 SFp32 semi-quantitative RT-PCR data normalised to β -actin in all seven cell lines [Figure 4.18].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle= SFp32 PCR cycle number sampled. Mean= average β -actin/SFp32 ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle	SW480			SW620			LoVo			Colo 201			T84		
	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n	Mean	SEM	n
20	0.137	0.0067	3	0.263	0.0088	3	0.220	0.0058	3	0.107	0.0033	3	0.100	0.0058	3
22	0.443	0.0033	3	0.680	0.0000	3	0.763	0.0033	3	0.310	0.0000	3	0.290	0.000	3

Cycle	SW1116			498LI		
	Mean	SEM	n	Mean	SEM	n
16	0.087	0.012	3	0.210	0.010	3
18	0.137	0.012	3	0.243	0.024	3
20	0.310	0.035	3	0.730	0.084	3

APPENDIX 5.2 Nucleotide sequence contig of the extended 3' UTR of IPP-2 [Figure 5.2]

The **start** codon (**ATG**) is written in **magenta** and the **stop** codon (**TAG**) is written in **red**. The four potential **polyadenylation signals** (**AATAAA**) are written in **blue**. The fragments used in the northern blot are highlighted in **cyan** (**Ipp2-1**) and **green** (**Ipp2-2**). The sequence of DDRT clone PM208 is underlined. This sequence has been assigned the following Genbank accession number: AJ133812.

```
1  CCAATGGCGG CCTCGACGGC CTCGCACCGG CCCATCAAGG GGATCTTGAA
51  GAACAAGACC TCTACGACTT CCTCTATGGT GGCGTCGGCC GAACAGCCCC
101 GCGGGAATGT CGACGAGGAG CTGAGCAAAA AATCCCAGAA GTGGGATGAA
151 ATGAACATCT TGGCGACGTA TCATCCAGCA GACAAAGACT ATGGTTTAAT
201 GAAAATAGAT GAACCAAGCA CTCCTTACCA TAGTATGATG GGGGATGATG
251 AAGATGCCTG TAGTGACACC GAGGCCACTG AAGCCATGGC GCCAGACATC
301 TTAGCCAGGA AATTAGCTGC AGCTGAAGGC TTGGAGCCAA AGTATCGGAT
351 TCAGGAACAA GAAAGCAGTG GAGAGGAGGA TAGTGACCTC TCACCTGAAG
401 AACGAGAAAA AAAGCGACAA TTTGAAATGA AAAGGAAGCT TCACTACAAT
451 GAAGGACTCA ATATCAAAC T AGCCAGACAA TTAATTTCAA AAGACCTACA
501 TGATGATGAT GAAGATGAAG AAATGTTAGA GACTGCAGAT GGAGAAAGCA
551 TGAATACGGA AGAATCAAAT CAAGGATCTA CTCCAAGTGA CCAACAGCAA
601 AACAAATTAC GAAGTTCATA GACGAGATTT GTTCAACACT GCAATTGTTT
651 GTTAGATGTA AACCTGTGA CTATAGTACG TTGCTTCTTG TTCTTCACAA
701 TTCATGACTT AAGTACCAA ATGCATACCA GTTATTATAT ATTGCCAAGA
751 ATTAAATGAT AAACCTAGAG ACTGATTAGA CTGAAAATGC CTAATCGATA
801 TATATATTCT TGTGCCTAGT ACTTTACCAC AAATACAGTG TAATATCATC
```


851 AGTCCAAAAC TGCATTACTT TTGTAAAAAC ACTGGTTAAT TTGTATAAGA
901 TATTATAGAG CTTTTTATGC TTTAGAAGTT AAACAATATC TTTGGGGGGG
951 AACTAATTTA TTTTCATCAC TTGAAATGTG GTAGCTCTTA CAAAGTTTAT
1001 TGATTTGATT TTTTAAAAA TCAAAAGCCA ATTGAACAAC AGGATATATA
1051 GACTGATAAA TATTTAGGCT GAATAGTATT TTAACACTTG TCTTCAACTT
1101 GATTTGTCTG TTTAATTGAA AAGAATTATA AGAGTTACTG TTGCATTTTC
1151 TGACCTACTA TTTTAAAAAT TCCTGTTGAG TTTCTTTGTG TTTACAAGGA
1201 AAGGACTGAA CTTTTTCTCA TCAAACTAG CTTTTTCCC CACAAATAAA
1251 TTATCAGGTT AAACTTTCAC CAATGTCTGC TCTGTTTTT GTTTTTGGGG
1301 GGTTTTTGTT TGTTTTTGTT TTTTAATGT TTTTGGTACA CTGGGCAGAC
1351 TTCAGAGCAG TTTTTTAAAA **AAATAAATAT** TCTAATGTAG CTATCTCGCC
1401 ATTCCCTTTA AATACCTGTC TTAACCTCCT GCTTTTCTTT CCTACTCCTT
1451 TCCACACATA CGCACACAAT CTTCTACCTT TTAAAGGATC ATTAAGACTG
1501 TCACCACATT AGGAGCTCTT TCTCTCACTC TTCTGTCATT TGCTGCTATG
1551 TTGAAATTCT TATTTTGACC ATCAATGCCT ATGAATTCTT CTAATACGTG
1601 AAGAAAATAG TTTGAGTAGC AGCAGTGCTA TAGGTGGGAA ATACAGTTTA
1651 GCTGCTGAAT TTCTATACCT CTCTGATTTA CAGATTGCTA **ATTAAATGGC**
1701 TATTATTAGT TTGGATTAAT TCGACTTAAG AAAACAATA ACTGAGGGTT
1751 TTTTTTGTT GTTTTTTGAG GGTTTTCTTT GCATGAGAAT TGTATGTAAC
1801 CAGTGATATG ATTATTCCTG AATGTACAGA CAGAAGTAAG CCTGGACATT
1851 GTTAAATAGT CCCTGCTTTA AGGGACTACG ATAATGTGTA CTATGACAAA
1901 CGTGCTTTAT TCTTCTAACG CAGTAAGAAT TAGGTGGAAT TTTTTCCTTC

1951 AACCAAGTGC AGGAAAGCCC TGTGTGTCTT GGTTTAGTTA TGGTTTCATT
2001 TCTAGCCATA CAATTGATGA ATTGTGTACA ACTTTTGTTA GTACCAAAAT
2051 AATCTGTTAT ATGAACAGAC TTCTAAAATA ACGTCTGTAT ATTTTATATA
2101 TAGATACATA TATGAAAGAA GGCTTTTATT GAACAGCTTA TCTTCCACTT
2151 GCAGGTTTAT GGAAACAGCA GTATTTGAAA ATA**AATAAAA** AGTGGGAGAA
2201 TTCCTTGCTG TTAGAAGAAT GTGGCCATTA TTTTGATTTT TTGAATGAGA
2251 TATATAATCA AAGTACTGCT GAACTGTGAG TGCAGTATTC TAAACATTTT
2301 AGCTAGGAAT ACCACTGATT TAGAAACAAA ACTGTTTGTC TCTGGTTTCT
2351 GAATTTAAAA TGTTGGGATT ACCTGTTTAA ATCTGTCTTG GGGGATGTAG
2401 AGATTAAGTC TGTACATGTG CGTGACATA TATTCATGCA CCCTCTGATT
2451 TTGGTTTTCT TGTTTCTGAG TTCTTAGAAA GTACCCACAT ACTCTTTTTT
2501 TTTTTTTTTG TAGAAGTAGC TGTCATAGAG TGAAGAAAAG GATAAGACTT
2551 TAAACAGTTG ATTCTTTTTG TGTTTTCTAC AACTATTTT TTGAAATTTA
2601 AATCACAAAC TCATTTTCTG GTTTTATAGAA AGTAGATGAT GATTCAGAG
2651 GAGTAAGGCA TGCCAAGCAG CATGCTCAGT GGGGTTTTAG GCTGTCACAT
2701 GCAGCTGAGA AAAGGTATGT TCAAGTCATA AGTAGGTAAT TGATAGGGTA
2751 TGAACTAGTC AAAATATGAA CCATTATGAT TCAAGTTAGA TTTTCCTCTG
2801 GAGAGACAGA TCTGAATGTT CAGTTCTAGC CAAGGTAGAT TTTACTTTCA
2851 ACTTTTTAAT CAGTATCACT TTCTGTGCTT AACTATTTGG TGTTACCTTG
2901 TCTGTTTTCA TTTGTCTAAA ATTCTGCAGA GATGACTAAA ATTTGACATA
2951 ATGGTGTAAG TGGATTGTGA AGGTAACTTT CAGTTGAAGA TTAAAGCAAG

3001 ATGCCATTTT CCCCATGACT TTGGTTTTGT TTACATTTT CCCTTTCAGT
3051 TAGTATACAC TACACGTACT GTAATAAAAT ACAATAATAT GACAAAAAAA
3101 AAA

APPENDIX 5.4 IPP-2 Northern blot densitometry data normalised to β -actin in SW480 and SW620 [Figure 5.4]. Exposed blots were quantitated by densitometry using the Bio1D v6.32 software.

Transcript size	SW480	SW620
4kb	72420	25583
2kb	33161	14503

APPENDIX 5.6 Hox11 semi-quantitative RT-PCR data normalised to β -actin in SW480 and SW620 [Figure 5.6].

The ratios were calculated from the amount of Cy5 labelled PCR product detected by the ALFexpress™ automated sequencer and interpreted by the 'Allele links' software (Pharmacia). Cycle No.= Hox11 PCR cycle number sampled. Mean= average β -actin/Hox11 ratio from three separate loadings (n) of the same sample. SEM= standard error from the mean. The results from different experiments are separated by the dotted line.

Cycle No.	SW480			SW620		
	Mean	SEM	n	Mean	SEM	n
27	2.180	0.310	3	0.460	0.038	3
29	2.423	0.469	3	1.317	0.311	3
32	3.613	0.391	3	1.753	0.147	3
27	0.100	0.000	3	0.030	0.000	3
29	0.227	0.003	3	0.087	0.003	3
32	0.410	0.006	3	0.157	0.003	3

APPENDIX 6.6/7 Keratin 5/8 immunohistochemical data [Figures 6.6 & 6.7 and Table 6.1]

'Block number' refers to the sections cut from a paraffin embedded tumour sample from a particular patient. '% Normal Positive' is the percentage of normal cells which stained positively. '% Tumour Positive' is the percentage of tumour cells which stained positively. The stained tumour cells were also further subdivided into Grade 1= weak cytoplasmic staining, Grade 2= strong membrane staining & weak cytoplasmic staining, Grade 3= strong membrane & cytoplasmic staining. The results are divided into sections derived from Duke's B patients and Duke's C patients with matched lymph node metastases. The Duke's C and lymph node metastases from the same patient are listed sequentially (Duke's C tumour first and lymph node metastases second) and samples from different patients are separated by a dotted line. ' - ' refers to the absence of normal tissue.

Duke's B

Block Number	% Normal positive	% Tumour positive	Grade 1	Grade 2	Grade 3
588/94	-	70	2	28	70
1109/94	-	100	5	25	70
1546/94	65	98	15	70	15
4722/96	-	98	75	12	13
4922/96	75	98	90	7	3
60/97	-	90	83	15	2
2221/97	60	70	65	30	5
3165/98	60	100	2	2	96
3283/98	-	60	60	38	2
4435/98	100	98	30	60	10
4519/98	-	90	40	50	10
4542/98	10	80	15	55	30
5039/98	40	80	10	80	10
5111/98	65	100	2	28	70
5518/98	-	60	40	55	5
5617/98	60	50	10	90	0
527/99	60	70	50	20	30
611/99	80	70	25	60	15
626/99	60	80	50	45	5
872/99 A9	-	70	30	50	20

APPENDIX 6.6/7 Keratin 5/8 immunohistochemical data [Figures 6.6 & 6.7 and Table 6.1]

Duke's C and Lymph node metastases

Block Number	% Normal positive	% Tumour positive	Grade 1	Grade 2	Grade 3
4076/93 1H	98	95	40	40	5
4076/93 1L	-	80	55	40	5
4937/93 F	90	90	60	2	38
4937/93 G	70	90	35	60	5
5371/93 D	98	95	10	10	80
5371/93 F	-	100	0	5	95
6226/93 D	-	90	5	90	5
6226/93 M	-	96	0	90	10
6437/93 1E	-	98	5	75	10
6437/93 1K	-	100	0	10	90
6714/93 8D	100	98	2	10	78
6714/93 8G	-	95	60	35	5
535/94 D	100	90	5	65	30
535/94 J	-	90	60	20	20
875/94 J	50	98	30	60	10
875/94 Q	-	100	2	53	45
1219/94 1H	-	60	38	60	2
1219/94 1K	60	70	45	45	10
1624/94 2G	80	90	10	45	45
1624/94 2Q	70	100	0	80	20
1981/94 E	-	80	45	50	5
1981/94 L	-	10	95	0	5
2541/94 1B	80	85	10	40	50
2541/94 1G	-	70	10	50	40
4391/96 E	30	98	20	80	<1
4391/96 E*	-	100			
191/97 1C	40	95	35	60	5
191/97 1J	-	95	50	45	5
2262/97 1C	40	90	10	70	20
2262/97 1G	-	95	20	60	20

* Lymph nodes in the same section as the tumour.

APPENDIX 6.6/7 Keratin 5/8 immunohistochemical data [Figures 6.6 & 6.7 and Table 6.1]

Duke's C and Lymph node metastases continued

Block Number	% Normal positive	% Tumour positive	Grade 1	Grade 2	Grade 3
2481/97 1F	-	95	10	40	50
2481/97 1P	-	90	30	70	0
1324/98 A	-	100	10	70	20
1324/98 P	-	95	40	50	10
5176/98 A5	-	80	40	40	10
5176/98 A9	-	98	20	55	25
7509/98 A5	70	90	80	20	0
7509/98 A11	-	80	40	60	0
727/99 B6	85	75	30	70	0
727/99 B14	70	99	20	80	0