Alu-Polymerase Chain Reaction Genomic Fingerprinting: a technique to identify novel genetic lesions in pancreatic cancer

Submitted for the examination of PhD. within the Imperial College School of Medicine (Faculty of Science)

by

Arthur Bennett McKie B.Sc., M.Sc.

Supervisor: Professor Nicholas R. Lemoine

RIBL
LONDON
UNIV.

Declaration

I declare that the work presented in this PhD thesis was carried out by myself unless otherwise stated below.

This work has not already been accepted for any degree, and is not being concurrently submitted in candidature for any other degree.

The Fluorescent *in situ* hybridisation (FISH) was carried out by the ICRF Cytogenetics Laboratory at Lincoln's Inn Fields, and the interphase-FISH mapping was carried out for us by Miss Erika Mitchell at the CRC Paterson Laboratory.

The tissue *in situ* hybridisation was carried out by Dr. Hing -Y Leung.

Automatic sequencing of genomic shotgun clones was carried partly by the Royal Postgraduate Medical School-Dept. of Virology sequencing service and by Mr. Robert McFarlane and Dr. Donny Black at the CRC Beatson Institute.

Results of a preliminary immunoprecipitation experiment carried out as part of ongoing studies on the Lrp130 protein are exhibited with the kind permission of Dr. Iman El-Hariry.

Arthur B. McKie, BSc., MSc.

Acknowledgements

There are many people I would like to acknowledge for their contribution or encouragement during my pursuit of this PhD, including family, friends and work colleagues.

My gratitude and thanks must must begin with my supervisor Nicholas Lemoine, whose confidence and continued support made this project possible. I joined Nick's laboratory in 1991 with little confidence and a severely bruised ego. During seven years of enjoyable and often hard work, we have achieved many things, not least my refinement as a scientist.

I would also like to thank Dr. Helen Hurst who has always made time to listen to my problems, theories and occasional sporting yarns. Likewise I would like to thank Bill Gullick who has always offered words of encouragement and Donny Black for his gems of wisdom over the years.

I have to thank many of the staff at the ICRF Molecular Oncology Unit for their friendship and support. Notably Hing Leung, whose encouragement and advice progressed to a valued friendship. Additionally, Graham Carter, Claire Barton, Andres Guitierez, Klaus Prenzel, Vorapan Sirivatanauksorn, Hardev Pandha, Lesley-Ann Martin and "the therapists". A last special mention goes to Ann who has shared many burdens over "a few" drinks and my other drinking buddy Tom, whose encouragement and faith never fade.

Lastly, but by no means least my thanks and love are extended to my wife Joan, who has put up with my moods, obsessive behaviour and tantrums during the last 4 years and without whose help this thesis would never have been completed.

Abstract

A fingerprinting strategy has been developed based on polymerase chain reaction (PCR) amplification of genomic DNA with primers specific for the Alu repeat element. This has been applied to DNA from tumour and normal tissue from pancreatic cancer cases and cell lines showing different metastatic capabilities to isolate and identify fragments of genomic DNA rearranged in the malignant cells. These fragments have been sequenced and used as molecular probes to isolate hybridising clones from gridded bacteriophage P1, phage artificial chromosomes (PACs) and cosmid libraries. Once isolated, these large genomic clones have been localised by fluorescent *in situ* hybridisation (FISH) and later adjacent expressed sequences were identified with Northern analysis.

Further characterisation of one Alu-derived fragment on chromosome 2p22 has identified an adjacent novel transcript that appears to be up-regulated specifically in ductal adenocarcinoma of the pancreas. A European Pancreatic Cancer Reference Library System (EPCRLS) has been utilised to isolate a complete cDNA for this transcript. A transcript map of the region involved has begun to refine the location of five genes previously localised to chromosome 2 band p22, identifying the upregulated transcript as a leucine-rich lectin like molecule.

A second area of genetic loss has been identified at 17p13.3 at a region frequently deleted in breast and ovarian carcinoma as well as other tumour types. Alumorph-specific cosmid and PAC clones have been mapped relative to existing Genethon and Whitehead genome databases, placing them in a 500 kb interval between D17S5 (YNZ22) and D17S379. Multiplex PCR has defined an area of minimal genetic loss on a small number of microdissected cases and a candidate gene has been analysed for mutation. In conclusion we

suggest that Alu-PCR fingerprinting may be a useful technique for the identification of genes involved in tumourigenesis.

Contents

Page No.

List of figures and tables 13

List of abbreviations 16

Chapter 1: Introduction

1.1: Pancreatic cancer, a major clinical problem 18

1.2: Familial and genetic risk factors for pancreatic adenocarcinoma 20

1.3: The molecular genetic analysis of cancer 21

1.4: The involvement of known cancer genes in pancreatic adenocarcinoma 23

1.4.1: The KRAS activated oncogene 23

1.4.2: KRAS mutations in pancreatic adenocarcinoma 25

1.4.3: Possibilities for earlier diagnosis and therapy 27

1.4.4: The TP53 tumour suppressor gene 27

1.4.5: Mutations of TP53 in cancer 31

1.4.6: P16/INK4A/CDKN2/MTS1 34

1.4.7: TGFβ and the Smad4 tumour suppressor gene 35

1.4.8: Growth factor systems implicated in pancreatic adenocarcinoma 36

1.4.9: Chromosome 7 and pancreatic cancer 38

1.4.10:Possible remaining molecular lesions associated with pancreatic adenocarcinoma 39

1.5: Tumour invasion and metastasis 41

1.6: New technology dissects molecular events in pancreatic adenocarcinoma 42

1.6.1: Representational Difference Analysis (RDA) 43

1.6.2: Differential Display (DD) 45

1.6.3: The European Pancreatic Cancer Reference Library System (EPCRLS) 46
1.6.4: Comparative genome hybridisation (CGH) 47

1.7: Alu-PCR DNA fingerprinting 48
1.7.1: SINEs and the Alu family 48
1.7.2: SINEs as endogenous retroelements 51
1.7.3: The evolution of the Alu element and possible roles in the human genome 53

1.8: Alu-PCR generated genomic fingerprints to identify genetic imbalance 54
1.8.1: The generation of sequence-tagged sites by Alu-PCR 54
1.8.2: Alu-PCR identifies microsatellite instability in poly(dA) runs 55
1.8.3: Alu-PCR Vs AP-PCR; an ongoing comparison 56
1.8.4: The fingerprint 57
1.8.5: Southern hybridisation with "Alu-morphs" 57
1.8.6: Genomic gridded libraries: a vital source of cloned material 58

1.9: Research aims 59

Chapter 2: Materials and methods

2.1: Clinical specimens 60

2.2: Established tumour cell lines and SUIT2 sub-lines 60

2.3: Restriction digestion 61

2.4: Alu-directed Polymerase chain reaction (Alu-PCR) 62

2.5: Polynucleotide Kinase reaction 62

2.6: Polyacrylamide gel electrophoresis 63

2.7: Autoradiography and excision of bands from gels 63

2.8: Cloning and sequencing of PCR fragments 64

2.9: Southern hybridisations 66

2.10: Screening of genomic gridded libraries 67

2.10.1: P1 bacteriophage library 68

2.10.2: P1 artificial chromosome (PAC) library 70

2.10.3: Cosmid libraries 70

2.10.4: Varied cDNA libraries from pancreas and other tissues 71

2.10.5: Electroporation of cDNA libraries into *E. Coli* cells 71

2.10.6: Preparation of P1 bacteriophage, PAC and plasmid DNA 72

2.11: Northern Hybridisation 73

2.12: Tissue in-situ hybridisation (TISH) 74

2.12.1: Preparation of RNA riboprobes 74

2.12.2: Preparation of sections from clinical specimens 75

2.12.3: Prehybridisation protease treatment 75

2.12.4: hybridisation 76

2.12.5: Post-hybridisation washes and detection of signal 76

2.13: Pulse-field gel electrophoresis (PFGE) 77

2.14: Multiplex PCR analysis 77

2.15: RT-PCR with Marathon™ cDNA amplification 78

2.16: Screening for mutations in candidate genes with SSCP 80

2.17: PCR-bases methylation detection system - CpGenome 83

2.17.1: The mechanism 83

2.17.2: Primer design 84

Chapter 3: Results

3.1: Establishment of the Alu-fingerprinting technique 85

3.1.1: Restriction enzyme predigestion 85

3.1.2: Comparison of different primers 88

3.1.3: Length of Alu-specific primers 90

3.1.4: Restriction enzyme cutting frequency and double digests 93

3.1.5: Magnesium ion concentration and sources of Taq plymerase 94

3.1.6: Method used to label/visualise bands 95

3.1.7: Reproducibility of the fingerprint patterns 96

3.2: Alu-PCR on paired tumour and normal samples 96

3.3: Characterisation of the Alumorph AB1 and associated transcript 101

3.3.1: Southern hybridisation 101

3.3.2: Sequence analysis 103

3.3.3: Isolation of genomic clones 103

3.3.4: Analysis of P1 clones 105

3.3.5: Chromosomal localisation 109

3.3.6: Northern analysis and cDNA clone retrieval 112

3.3.7: Tissue *in-situ* hybridisation of chromosome 2 transcript 120

3.3.8: Methylation status of AB1 sequences 124

3.3.9: The CpGenome modification kit 129
3.3.10: CpG island library 133

3.4: The complete sequence analysis of the ART1 transcript 134
3.4.1: Isolation of cDNA clones 134
3.4.2: cDNA cloneGr-5 and accumulation of additional cDNAs 135
3.4.3: I.M.A.G.E. cDNA clones 137
3.4.4: Shotgun cloning and genomic sequencing 138
3.4.5: Genomic sequence similarities identify CYP1B1 gene 139
3.4.6: A 2p22 transcript map 140
3.4.7: Lrp130: an upregulated transcript cloned from hepatoblastoma 141
3.4.8: Identification of upstream sequence homologies 142
3.4.9: RT-PCR experiments 143
3.4.10: Chromosome 2p22 physical map is refined 148
3.4.11: Summary of chromosome 2p22 findings 150

3.5: Characterisation of the IP-1 Alumorph 151
3.5.1: Interphase-FISH analysis 151
3.5.2: Positional information 154
3.5.3: I.M.A.G.E. cDNAs 156
3.5.4: The mouse vibrator/*PITP* locus 156
3.5.5: Multiplex-PCR reveals homozygously deleted area between D17S5 and D17S379 157
3.5.6: The mouse *Vibrator* critical region and SSCP analysis of PITPn 160

3.6: Analysis of Alumorphs from SUIT2 metastatic cell lines 162
3.6.1: Met6 Alumorph 162
3.6.2: Southern and Northern with Met6 162
3.6.3: Met6 localised to short arm of chromosome 11 164
3.6.4: Met-L and Met-U Alumorphs 164

3.7: Isolation of Alumorph AB4 166
3.7.1: Southern and Northern analysis 166

Chapter 4: Discussion

4.1: The isolation of pancreatic cancer-specific lesions using Alu-PCR genomic fingerprinting 169

4.1.1: Alu-PCR - analysis of fingerprint profiles 169

4.1.2: Analysis of Alumorph sequences 173

4.2: The characterisation of the chromosome 2 lesions 175

4.2.1: Alumorph AB1 exhibits loss in three tumour cases 175

4.2.2: Isolation of 2p22 transcript over-expressed in pancreatic tumour cell lines 177

4.2.3: The Lrp130 gene discovery 179

4.2.4: ART1/Lrp130 may be part of large carbohydrate-lectin family 180

4.2.5: Cell adhesion and tumour metastasis 181

4.2.6: Lrp130-antibody co-localisation experiments 182

4.3: The possible role of the hypomethylated sequences 184

4.4: DNA methylation: its role in development and cancer 186

4.4.1: Methylation in development 186

4.4.2: Histone acetylation 188

4.4.3: Aberrant methylation in cancer 188

4.4.4: Evidence of AB1/Sp-0.3-23 hypomethylation releasing Lrp130 from transcriptional repression 191

4.4.5: PCR-based methylation detection system 191

4.4.6: Possible experiments to prove the hypomethylation is causatively linked to Lrp130 increased expression 193

4.5: Future research objectives for the 2p22 lesion 194

4.6: The analysis of the 17p13.3 lesion 196

4.6.1: Alumorph IP-1 localised to 17p13.3 196

4.6.2: Allelic losses on 17p13.3 196

4.6.3: PITP and its role in inositol lipid hydrolysis and signal transduction 198
4.6.4: The mouse Vibrator critical region and SSCP analysis of PITP 199
4.6.5: Future work 199

4.7: Analysis of other Alumorphs 201
4.7.1: BACH-1-specific Alumorph 201
4.7.2: 4.7.2 Met6 Alumorph 202

4.8: Future research 203
4.8.1: Est projects give us a molecular inventory of pancreatic tumours 203
4.8.2 The further development of the Alu-directed fingerprinting 205

4.9: Aims of the thesis fulfilled? 208

References 212

Appendices 239

List of tables and figures

		Page No.
Figure 1:	The RASactivation cycle	25
Figure 2:	The p53/p16 cell regulatory pathways	30
Figure 3:	Alu structure	51
Figure 4:	Examples of Alu fingerprints demonstrating predigestion of DNA with different restriction enzymes	86
Figure 5:	Example of a polymorphic fingerprint	87
Figure 6:	Examples of different Alu fingerprints generated with different Alu-specific primers	89
Table 1:	Alu-specific primers used	91
Figure 7a:	Fingerprints generated from truncated primers from the PDJ-33 Alu consensus primer	92
Figure 7b:	Fingerprints generated from double digestion of genomic DNAs	92
Figure 8:	Examples of Alumorphs	99
Table 2:	Fragments isolated with details of original observations	100
Figure 9:	Schematic of general approach using the AB1 alumorph as an example	102
Figure 10a & c:	Photographs of P1, PAC and cosmid digests	106
Figure 10b & d:	Tables of sizes	106
Figure 11:	Analysis of P1 bacteriophages with Alu-PCR	108
Figure 12:	Localisation of AB1 PCR fragment with monochromosomal somatic cell hybrid panel	110
Figure 13:	FISH localisation of AB1-specific P1s and YAC clone	111
Figure 14:	Northern analysis with alumorph-specific P1 fragments	113
Figure 15:	Northern analysis with AB1-specific cDNA	114

Table 3:	Details of Alu-generated fragments	115
Figure 16:	Analysis of expression levels with Bio-Rad phosphoimager	117
Figure 17:	Breast Northern	119
Figure 18:	Tissue *in situ* hybridisation of ART1 on pancreatic tumour specimens	121
Figure 19:	Tissue *in situ* hybridisation of ART1 on duodenum and bladder normal tissue sections	122
Table 4:	ART1 tissue distribution	123
Figure 20:	Analysis of methylation status of AB1 sequences by Hpa II/Msp I isoschizomer Southerns	125
Figure 21:	Methylation status of different genes	128
Figure 22:	Sequence of bisulphite modified DNA from tumour and normal samples	131
Table 5:	Examples of deaminated C-T residues in tumour cell lines	132
Figure 23:	Representatives of gridded libraries and details of ART1 cDNA clones isolated	136
Figure 24:	ART1 RT-PCR products	147
Figure 25:	Physical map of the 2p22 region	149
Figure 26:	17p13.3 Fluorescent in-situ hybridisation (FISH) and interphase-FISH	152
Figure 27:	Examples of LOH on 17p13.3	153
Figure 28:	17p13.3 Positional map	155
Figure 29:	PITP cDNA Southern	159
Figure 30:	Multiplex-PCR analysis of 17p13.3 region	160
Figure 31:	SSCP analysis of exons 3 and 4 of the PITP gene	162
Figure 32:	Metastasis-specific alumorph Met6 fingerprint and Southern blot	164
Figure 33:	Southern & Northern of alumorph AB4	168
Table 6:	Numbers and sizes of bands generated	171

Figure 34: AB1 amplification on tumour and normal DNA 176
Figure 35: Interaction of Lrp130 with E-cadherin and catenins in pancreatic cell lines 183
Table 7: Methylation of various CpG Islands in various tumours 190
Table 8: Sequence alignment of seven Alu-specific primers 206

Appendices

1: Patient details 239
2a: AB1/Sp. 03.-23 sequence homology 240
2b: IP-1/PITP sequence homology 240
2c: HER4 sequence homology 241
2d: BACH1 sequence homology 242
2e: Human cytochrome P450 IIA sequence homology 243
2f: Human cytochrome P450 IIIA sequence homology 243
2g: Lrp130/cDNA sequence homologies 244
3: Lrp130 sequence showing upstream homologies 245
4: Restriction maps of cDNA plasmids 260
5: Table of STS and EST primers 261
6: Sequence details of Bisulphite-modified samples 262
7: Alumorph sequences 263
8: Details of shotgun clones sequenced 265
9: Alu-PCR paper 268

List of abbreviations

AP-PCR	Arbitrary-Primed Polymerase Chain Reaction
ATP	Adenosine Tri-Phosphate
AT	Ataxia Telangiectasia
BAC	Bacterial Artificial Chromosome
bp	Base-pair
BCR	Breakpoint Cluster Region
BSA	Bovine Serum Albumin
cDNA	Complementary Deoxyribonulceic Acid
CEN	Centromere
CGH	Comparative Genome Hybridisation
cM	Centi-Morgan
cpm	Counts Per Minute
CTP	Cytosine Tri-Phosphate
DD	Differential display
ddNTP	Dideoxynucleotide triphosphate
dNTP	Deoxynucleotide triphosphate
DNA	Deoxyribo-Nucleic Acid
DTT	Di-Thio-Threotol
ddTTP	dideoxy Thymidine Tri-phosphate
EMBL	European Molecular Biology Laboratory
EPCRLS	European Pancreatic Cancer Reference Library System
ERCP	Endoscopic Retrograde Cholangio Pancreaticography
EST	Expressed Sequence Tag
Et Br	Ethidium Bromide
FAP	Familial Adenomatous Polyposis
FISH	Fluorescent In situ Hybridization
GDB	Genome Data Base
GTPase	Guanosine-TriPhosphatase
HGMP	Human Genomce Mapping Project
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
IPTG	Isopropyl-ß-D-thiogalactopyranoside
kb	Kilobase
KRAS	Kirsten-ras
LINE	Long Interspersed Nuclear Element
LOH	Loss Of Heterozygosity

LTR	Long Terminal Repeat
Mb	Mega-base
mg	Milligram
μg	Microgram
ml	Millilitre
μl	Microlitre
mRNA	Messenger Ribonucleic Acid
MTS	Multiple tumour-suppressor gene
OD	Optical density
ORF	Open Reading Frame
PAC	P1 phage Artificial Chromosome
PCR	Polymerase Chain Reaction
PFGE	Pulse Field Gel Electrophoresis
RACE	Rapid amplification of cDNA ends
RDA	Representational Difference Analysis
rDNA	Ribosomal deoxyribonucleic Acid
RF	Replicative Form
RFLP	Restriction Fragment Length Polymorphism
RH	Radiation Hybrid
RLDB	Reference Library Database
RNA	Ribo-Nucleic Acid
rRNA	Ribosomal ribonucleic acid
RT	Reverse transcriptase
SDS	Sodium Dodecyl Sulphate
SINE	Small Interspersed Nuclear Element
SSC	Standard Saline Citrate
SSCP	Single-Stranded Conformational Polymorphism
STS	Sequence Tagged Site
TBE	Tris-Borate EDTA
TEL	Telomere
TIGR	The Institute of Genome Research
Tm	Melting temperature
tRNA	Transfer ribonucleic Acid
UTR	Untranslated Region
VNTR	Varying Number of Tandem Repeat
YAC	Yeast Artificial Chromosome

Chapter One : Introduction

1.1 Pancreatic cancer, a major clinical problem

Pancreatic cancer is a major health problem in the western world. It is the fifth most frequent cause of cancer death in the USA and Europe (Neiderhuber *et al.*, 1995 ; Poston *et al.*, 1991) and its incidence has risen steadily over the past fifty years. The aggressive and almost uniformly lethal nature of pancreatic cancer has earned it the epithet "the dismal disease". Primary carcinoma of the pancreas has a ductal phenotype in the vast majority of cases although occasional cases can be classified as acinar cell carcinomas, squamous cell carcinomas, cystadenomas and other rare tumours. The tumour itself is frequently small, often accounting for up to one third of the gross lesion. The remainder consists of extreme fibrous connective tissue/stromal proliferation surrounded by a zone of pancreatitis. Pancreatic adenocarcinomas are most frequently found within the head of the pancreas and compress the pancreatic duct and common bile duct leading to jaundice in 80-90% of patients. Often by the time of presentation, adjacent structures such as the portal vein, vena cava, stomach and duodenum are invaded by the tumour.

The pancreas is composed of three major elements: endocrine cells forming the Islets of Langerhans, acinar cells that produce secretory enzymes and a ductal system that drains the enzymatic secretions into the small intestine. Tumours of the endocrine and acinar cells are rare, the former usually give rise to symptoms referable to excess hormone production. Tumours arising in the ductal epithelium are the common type of pancreatic cancer accounting for 90% of cases. These tumours of the exocrine pancreas are difficult to diagnose as the clinical signs arise late in the course of the disease and symptoms tend to be non-specific, the cardinal features being pain, weight loss and jaundice (Flanders & Foulkes, 1996). Upon diagnosis, the therapeutic

options are few, with little evidence of response to chemotherapeutic or radiotherapeutic regimens. For 90% of cases palliative care is the only option with a five year survival rate of 0.4%. Radical and palliative resection as appropriate represents the primary treatment for the minority of patients who are young enough and fit enough with sufficiently localized disease but the perioperative mortality remains significant (5-10%) (Bottger *et al.*, 1993). Surgical palliations are often performed to relieve jaundice, maintain gastric emptying and control pain. Stenting the biliary ducts can be achieved endoscopically, or by a combination of percutaneous-endoscopic techniques.

The past twenty years has seen a period of considerable refinement and polishing in the diagnosis, staging and treatment of pancreatic cancer. There have been no major breakthroughs but incremental gains through (a little) earlier and more accurate diagnosis, better imaging techniques and more treatment options including non-operative stenting of the biliary tract, and markedly reduced morbidity and mortality from surgical procedures. The development of specialist units has shown that resection can be undertaken in 10-30% of cases, with a surgical mortality rate of less than 5% compared with general surgical units (Neoptolemos & Lemoine, 1995). In spite of these improvements, few patients are truly cured. There are no valid screening tests and therefore the possibility of diagnosis continues to rely on the recognition of symptoms. By this time 98% of patients have cancer that cannot be cured, and the majority are even beyond significant palliation. Metastasis has already occurred in most patients even those with small tumours, suggestive of an intrinsic biological behaviour characterised by the aggressive tumour growth and cell migration.

The epidemiology of pancreatic cancer offers few clues to its aetiology or pathogenesis. Its incidence increases with age, and it appears more common in men than women and in blacks more than whites (at least in north America). Cigarette smoking is a well established risk factor, resulting in a 4-fold increase in incidence and previous history of pancreatitis is a risk factor

also, though the strength of this association is uncertain. Recent multinational studies have shown that a history of pancreatitis for seven years or more before diagnosis of cancer is associated with a relative risk of 2.04 (95% confidence interval of 1.53-2.72) (Bansal & Sonnenberg, 1995). Diabetes mellitus has long been associated with pancreatic cancer. Recently, Evans *et al* reported a large pedigree with nine cases of pancreatic cancer inherited in an autosomal dominant fashion over three generations (Evans *et al.*, 1995). All nine family members diagnosed with pancreatic cancer had diabetes and five had diabetes for more than 10 years before diagnosis of the cancer. There has always been some uncertainty as to whether diabetes, often found at presentation with pancreatic cancer cases is a causative or a symptomatic phenomenon and the relationship between the two diseases has been particularly difficult to determine when the diabetes is diagnosed along with the cancer.

1.2 Familial and genetic risk factors for pancreatic cancer

There are few clearly defined familial cancer families, although pancreatic cancer may occur with increased frequency in several inherited cancer predisposing syndromes. Predispositions are apparent within hereditary non-polyposis coli families (Lynch *et al.*, 1985), Von Hippel Lindau syndrome (Fill *et al.*, 1979), the Familial Atypical Multiple Mole Melanoma (FAMMM) syndrome (Lynch & Fusaro, 1991), the autosomal dominant disorder ataxia telangiectasia (Vorechovsky *et al.*, 1996) and within certain populations carrying mutations in the recently identified BRCA2 gene discussed in more detail below. The association with the FAMMM syndrome is particularly interesting with pancreatic adenocarcinoma being the second most common cancer in FAMMM pedigrees. The observed/expected (O/E) ratio for the frequency of pancreatic cancer among 200 members of nine FAMMM families was 13.4 ($p < 0.001$) (Bergman *et al.*, 1990). Mutations in the cell cycle inhibitor p16/INK4A has been found to cosegregate with both melanoma and pancreatic adenocarcinoma in these families. In other studies however, there

is no bias towards pancreatic tumours in 370 cases from 15 families (Schutte *et al.*, 1997). There may be a biological explanation for this apparent subset of FAMMM kindreds who have an excess of pancreatic cancer. In families where the p16/INK4A mutation disrupts protein function by initiating premature truncation of the protein in *in vitro* assays, the risk of pancreatic cancer was increased 13-fold (95% CI 1.5-47.4), whereas no cases of pancreatic cancer were found in families with p16/INK4A mutations that did not totally abrogate the protein function (Goldstein *et al.*, 1995).

There have been a number of case reports in the literature which suggest that there is a familial form of pancreatic cancer. For example, Lynch *et al* describe 47 individuals with pancreatic cancer in 18 families in which multiple family members had pancreatic cancer (Lynch *et al*, 1990). Anecdotal reports and several case-control studies (Fernandez *et al*, 1994) suggest that between 3 and 10% of pancreatic cancers are caused by inherited factors. Reports of autosomal dominant pancreatic cancer families are beginning to be reported, but as yet they are restricted to centres in the USA with well documented pancreatic cancer registries, (for example The Johns Hopkins pancreatic cancer web site: http://www.jhu.edu/pancreas).

1.3 The molecular genetic analysis of cancer

Tumour development involves a dysregulation of the cellular processes which govern the generation of cells, their differentiation and their death. Tumour progression is known to involve multiple alterations of genes with key roles in cell homeostasis. The genetic basis for tumour progression was demonstrated in the early 1980s when it was shown that transfection of human tumour DNA into NIH 3T3 cells could result in their neoplastic transformation (Krontiris & M., 1981; Shih *et al.*, 1981). The DNA sequences responsible were soon identified, and since these early investigations the list of positively transforming (proto)oncogenes has grown to an overwhelming length. Their dominant action in changing the cellular phenotype was

relatively easy to demonstrate, in contrast to the recessive nature of action of the tumour-suppressor genes. Early experiments (Bregula *et al.*, 1971; Harris *et al.*, 1969), which fused normal and tumour cells to produce hybrid cells lacking the tumorigenic phenotype were received with scepticism. The normal cells had presumably provided the genetic information that was absent in the tumour cells, suggesting the existence of DNA sequences which normally function to suppress uncontrolled growth and division. With an explosion of interest later in the 1980's in the molecular genetics of cancer, searches for direct evidence of alterations in tumour tissues compounded the theories laid by the somatic cell hybrid experiments.

The vast majority of mutations in cancer genes are somatic and are found only in the individuals cancer cells. There are however, a tiny minority of individuals (about 1%) who exhibit clear hereditary transmitted cancer. These familial cases carry germline mutations in every cell of their body. Although relatively rare, the familial cancer cases are of major biological importance since studies of these specific germline mutations responsible for the inherited form of the cancer may provide vital insights into the more sporadic cancers, e.g. Familial adenomatous polyposis (FAP). Detection of these gene alterations which include deletions, rearrangements, amplifications and point mutations is often the end result of work over many years in determining the location of these vital genes (positional cloning) (Collins, 1991). Often the cancer geneticist has important clues to follow, such as translocations or microscopically visible deletions which act as pointers suggesting the proximity of gene(s) involved in the pathogenesis of the particular tumour type. For example, the observation of a large interstitial deletion of chromosome 5q21 in a patient with hundreds of intestinal polyps helped identify the APC gene (Varesco *et al.*, 1989). The Neurofibromatosis type 1 (NF1) gene was isolated through the observation of balanced translocations on chromosome 17, band q11 in individuals with this disease (Yagle *et al*, 1990). In most cases, linkage analysis remains the mainstay of efforts to map inherited cancer genes. The research takes years of

accumulating data from studies of families with inherited predispositions to the particular disease. Of critical importance is the discovery of these predisposing families and the conformation of clear diagnostic criteria to recognise such families. Genetic studies have a greater likelihood of success in families in which multiple affected and unaffected individuals in two or more generations are available for analysis rather than in families in which only a few individuals can be studied (Fearon, 1997). Unfortunately, pancreatic cancer usually falls into this second category as relatively few reports of familial occurrences are known and most of the reported instances possess only small pedigrees with uncertain family history and lack of patient material to analyse due to the rapid demise of the affected members. A complicating factor in genetic studies is that cancer is not a single disease, even when it arises in the same organ site. Rather, it is a spectrum of many diseases, some of which are very common and others extremely rare. Thus, families in which multiple members develop a rare form of cancer, such as retinoblastoma are much more likely to have a segregating mutation with the disease than families affected by more common cancers such as adenocarcinoma of the lung, breast, colon or pancreas (Fearon, 1997). Nonetheless, an inherited cancer syndrome should be considered when numerous family members develop tumours at an especially young age or affected individuals develop multiple primary tumours.

1.4 The involvement of known cancer genes in pancreatic adenocarcinoma

1.4.1 The KRAS activated oncogene

Ras is a small membrane bound 21kDa guanine nucleotide-binding protein that acts as a molecular switch linking receptor tyrosine kinases to downstream cytoplasmic signalling cascades. Ras cycles between an active GTP-bound state and an inactive GDP-bound state. The activity of Ras is regulated by the opposing activities of guanine nucleotide exchange factors (GNEFs) and GTPase activating proteins (GAPs) (figure 1). GNEFs catalyse the activation of Ras proteins by exchanging Ras bound GDP for GTP. To date,

five potential mammalian Ras GNEFs have been described, including Sos (Son of sevenless) (Bar-Sagi, 1994). GAPs accelerate the normally slow intrinsic GTPase activity of the Ras proteins, thereby inactivating them. Three mammalian Ras GAPs have been described, including the p120GAP (Boguski *et al.*, 1993), neurofibromin (the product of the von Reckinghausen neurofibromatosis type I gene), (Bollag *et al.*, 1991) and GAP1 (Maekawa *et al.*, 1994). The ras genes were first characterised as transforming agents of highly oncogenic retroviruses. Three mammalian ras genes were identified as cellular oncogenes: *HRAS*, *KRAS* and *NRAS*. Oncogenically activated forms of all three genes, encoding proteins with point mutations at amino acid residues 12, 13 and 61, have been detected in numerous human malignancies, although the incidence of the specific ras gene that is mutated in a particular tumour varies (Bos, 1989).

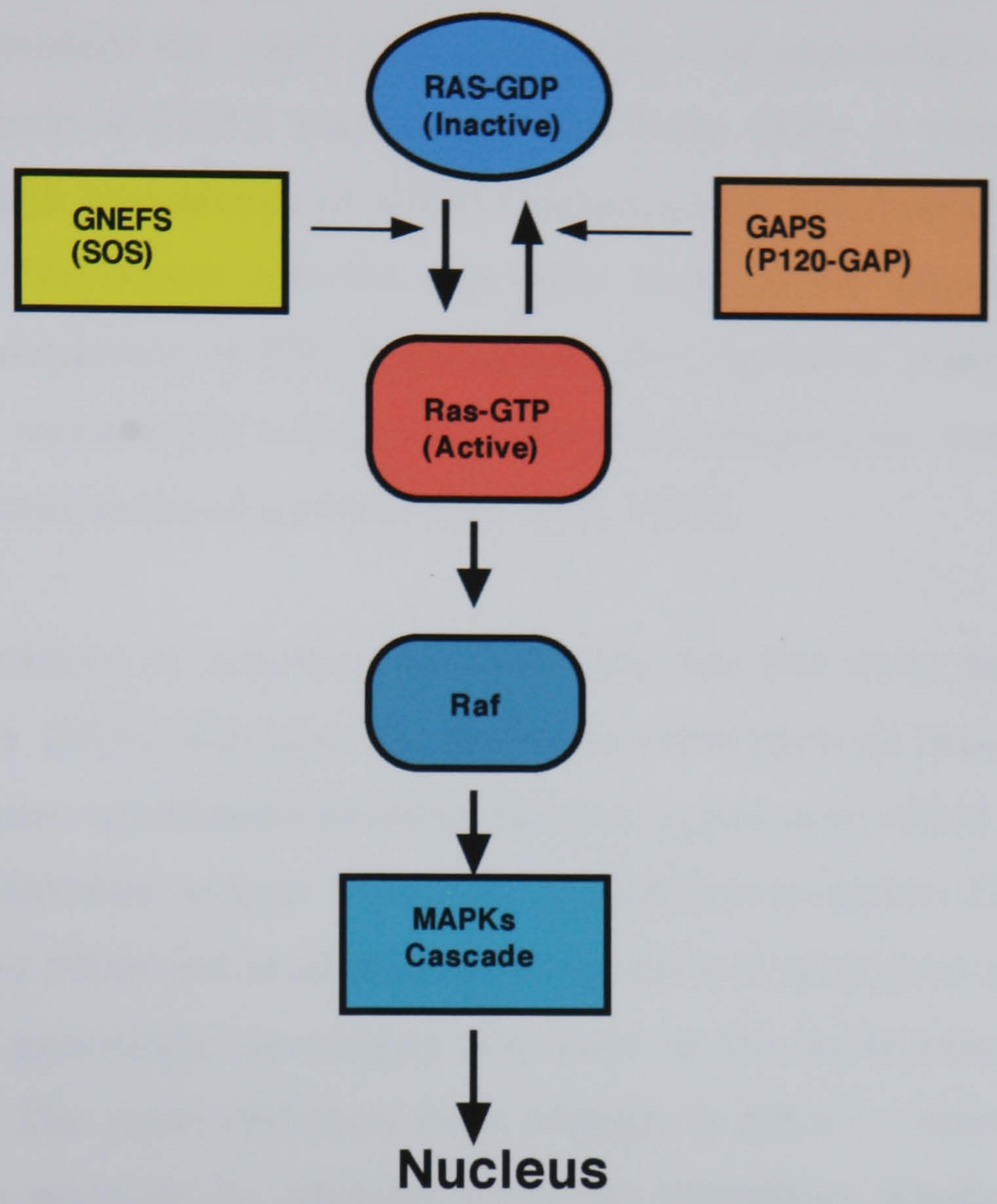


Figure 1: The Ras cycle, showing how Ras can be regulated by the opposing activities of GNEFs and GAPs

1.4.2 KRAS mutations in pancreatic cancer

Mutations in the *KRAS* gene have been identified in 75-85% of pancreatic carcinomas in numerous studies (Almoguera *et al.*, 1988, Bos, 1988, Mariyama *et al.*, 1989, Hruban *et al.*, 1993b). The mutations are invariably activating point mutations confined to codons 12 or 13 of the gene. *KRAS* mutations have also been identified in "hyperplastic" lesions of the pancreatic duct, in about 70%

of reported cases (Yanagisawa *et al.*, 1993). Ductal hyperplasias with evidence of *KRAS* mutations were derived from patients with pancreatic adenocarcinoma. It is suggested that these hyperplasia are neoplastic and are likely to represent the early precursor lesions of pancreatic adenocarcinoma. The occurrence of *KRAS* mutations in an early stage of tumour progression, as well as high prevalence of *KRAS* mutations in pancreatic carcinoma cases, suggests an important role for this gene early in the tumorigenic pathway. This high frequency of KRAS mutations distinguishes pancreatic carcinoma from other carcinomas where the mutation frequencies vary from virtually absent to a maximum of around 50% (Bos, 1989).

Pancreatic cancer is notable also not only for the very high frequency of activation of *KRAS* but also for the tight restriction of mutational events to codon 12 alone (mutations in other tumour types may affect *KRAS*, *HRAS* or *NRAS* and involve codons 13 and 61 in addition to codon 12). The pattern of base changes observed is also interesting since it may give some insight into the type of genotoxic carcinogen involved in the molecular pathogenesis of this cancer. The most common base change at *KRAS* codon 12 in pancreatic cancers is a guanine to adenine (G to A) transition, similar to the type of mutations seen in colon cancer. This shows a marked contrast to the pattern observed in lung tumours, in which guanine to thymine transversions are the most common lesions (G to T). This is surprising, since the epidemiology of pancreatic cancer suggests that cigarette smoking is a major aetiological factor in the disease, as it is of course in lung cancer. Hence one might have expected that the genes involved in both tumour types would show a similar "molecular fingerprint". The high prevalence of G to T transversions in lung cancer can be ascribed to the action of the carcinogen benzpyrene, which forms adducts with guanine residues and causes misrepair. The prevalence of G to A transitions in pancreatic cancer could be due to the action of a different class of carcinogens found in cigarette smoke such as the aromatic amines and N-nitrosamines. There is evidence that the prevalence of *KRAS* point mutations is higher in patients who have smoked cigarettes than in those

who have never smoked (Hruban *et al.*, 1993a). In addition, experimental models in the Syrian hamster show that tobacco-specific nitrosamines such as (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone) can induce pancreatic cancer (Rivenson *et al.*, 1988) and that tumours induced by nitrosamines like this contain *KRAS* mutations at codon 12 (Fujii *et al.*, 1990).

1.4.3 Possibilities for earlier diagnosis and therapy

Clearly, the high frequency and specificity of *KRAS* mutations in early pancreatic cancer makes their detection an attractive approach to early diagnosis and possibly even to screening in populations at risk (for example, elderly smokers and those with chronic pancreatitis). Several groups have shown that PCR-based techniques can successfully detect *KRAS* mutations in exfoliated tumour cells present in the pancreatic juice collected at endoscopic retrograde cholangiopancreatography (ERCP) of patients suspected of having pancreatic cancer. It is also possible to detect for the presence of mutant *KRAS* alleles in DNA extracted from exfoliated cells in faecal material (Caldas *et al.*, 1994b).

Mutant *KRAS* may also present a target for specific therapy. To be active in signal transduction and oncogenic transformation, RAS protein must be localised to the cell membrane, a process that requires lipid modification of the carboxy terminus. This lipid modification could be inhibited by drugs that interfere with HMG-CoA reductase or other enzymes involved in cholesterol metabolism. Lovastatin is an example of this class of agent, and it has been shown to inhibit the growth and tumourigenicity of pancreatic cancer cells (Mikulski *et al.*, 1992).

1.4.4 The *TP53* tumour suppressor gene

The p53 protein is found in all tissues of the body and has been highly conserved through evolution. It is not essential for life but plays a crucial role in maintaining the genetic stability of cells (Lane, 1992). This is demonstrated by the mouse transgenic model, where homozygote p53 knock-outs develop

normally, however, develop cancers relatively early in life (Donehower *et al.*, 1992). The *TP53* tumour suppressor gene is mutated or lost in almost half of all human cancers, indicating that functional inactivation of this gene has a central role in the progression of many human tumours (Greenblatt *et al.*, 1994). In addition to the role of *TP53* in somatic tumours, inherited mutations in the germline *TP53* gene have been shown to cause a familial cancer predisposition, Li-Fraumeni syndrome (Malkin *et al.*, 1990 ; Srivastava *et al.*, 1990).

The p53 protein is activated in response to DNA damage and then operates as a transcription factor, up-regulating the expression of target genes that promote growth arrest or apoptosis (figure 2-(1)). The specific response elicited is context dependant and can correlate with the functional status of the retinoblastoma protein (pRB). In pRB-competent cells, p53 induces G1-check-point arrest by induction of the cell cycle regulator p21/WAF1/CIP1, whereas pRB-deficient cells bypass the G1-checkpoint response and undergo p53-dependant apoptosis (Levine, 1997). These biological endpoints establish the basis of the anti-oncogenic potency of p53. As would be expected for such an important regulator of diverse processes, p53-mediated actions are subject to stringent multilevel control (Lane & Hall, 1997). The cellular concentration of p53 protein increases in response to DNA damage. This effect is thought to be achieved through the increased translation of pre-existing p53 mRNA, although the exact mechanisms are still not fully understood. Protein stabilisation may also contribute to the rise in p53 concentration as it is normally a very short lived protein with a cellular half-life of 5-20 minutes (Chowdary *et al.*, 1994). A major aspect of p53 regulation involves a negative feedback loop with MDM2, itself a cellular oncoprotein (the Mdm-2 gene was originally isolated as an oncogene amplified in a mouse double minute chromosome). In cells undergoing DNA damage or responding to certain forms of stress, the ordinarily short-lived p53 protein is stabilised, rapidly accumulates and undergoes several activating post-translational modifications (Levine, 1997). The activation of p53 leads to the upregulation

of many gene targets including the *MDM2* gene. The MDM2 protein can bind to the p53 transactivation domain and downregulate its transcriptional activity (Chen *et al.*, 1993). In addition, MDM2 can also target p53 for ubiquitin-mediated proteasomal degradation (Haupt *et al.*, 1997). Not surprisingly, *MDM2* amplification and overexpression have been observed in some cancers, rendering them functionally deficient for the p53 protein (mutations in the *TP53* gene are not usually observed in these tumours).

Communication between p53 and pRB pathways occurs at a number of levels. For example, the general cyclin kinase inhibitor p21/WAF1/CIP1 is a target for transcriptional activation by p53 which in turn inhibits the cyclin-dependent protein kinases (CDK)-mediated phosphorylation of pRB leading to G1-checkpoint arrest (Sherr, 1996). Studies have shown that this induction of the p21 protein is prolonged, lasting for at least four days and has been proposed to be permanent. (Di Leonardo *et al.*, 1994). It is becoming apparent however, that different cell types develop a variety of responses to DNA damage involving the same critical proteins such as p53 and p21. Studies to examine the responses to actinomycin D induced DNA double strand breaks in normal fibroblasts suggest that after an initial induction of p53 and p21 the levels of these proteins return to baseline levels after 8 days. By this time the concentration of p21 protein (which goes up as high as 100-fold shortly after treatment) is down to just 2-4-fold over baseline. Following the drop in p21 concentration a large increase in the expression level of the tumour suppressor gene p16/INK4a (detailed below) is observed. It is proposed that the transient increase in p21 is followed by a more prolonged p16 induction leading to permanent senescence of the cell (Robles & Adami, 1998). It has also been demonstrated that p21WAF1/CIP1 also plays a role in the exit of cells from the cell cycle during terminal differentiation. Here the expression of p21WAF1/CIP1 is independent of p53 and is consistent with the normal tissue development in p53-null mice (Halevy *et al.*, 1995 ; Parker *et al.*, 1995).

Adenovirus can disrupt this p53-TBP interaction through competitive binding and so relieve the p53-mediated repression of transcription (Horikoshi *et-al* ., 1995)

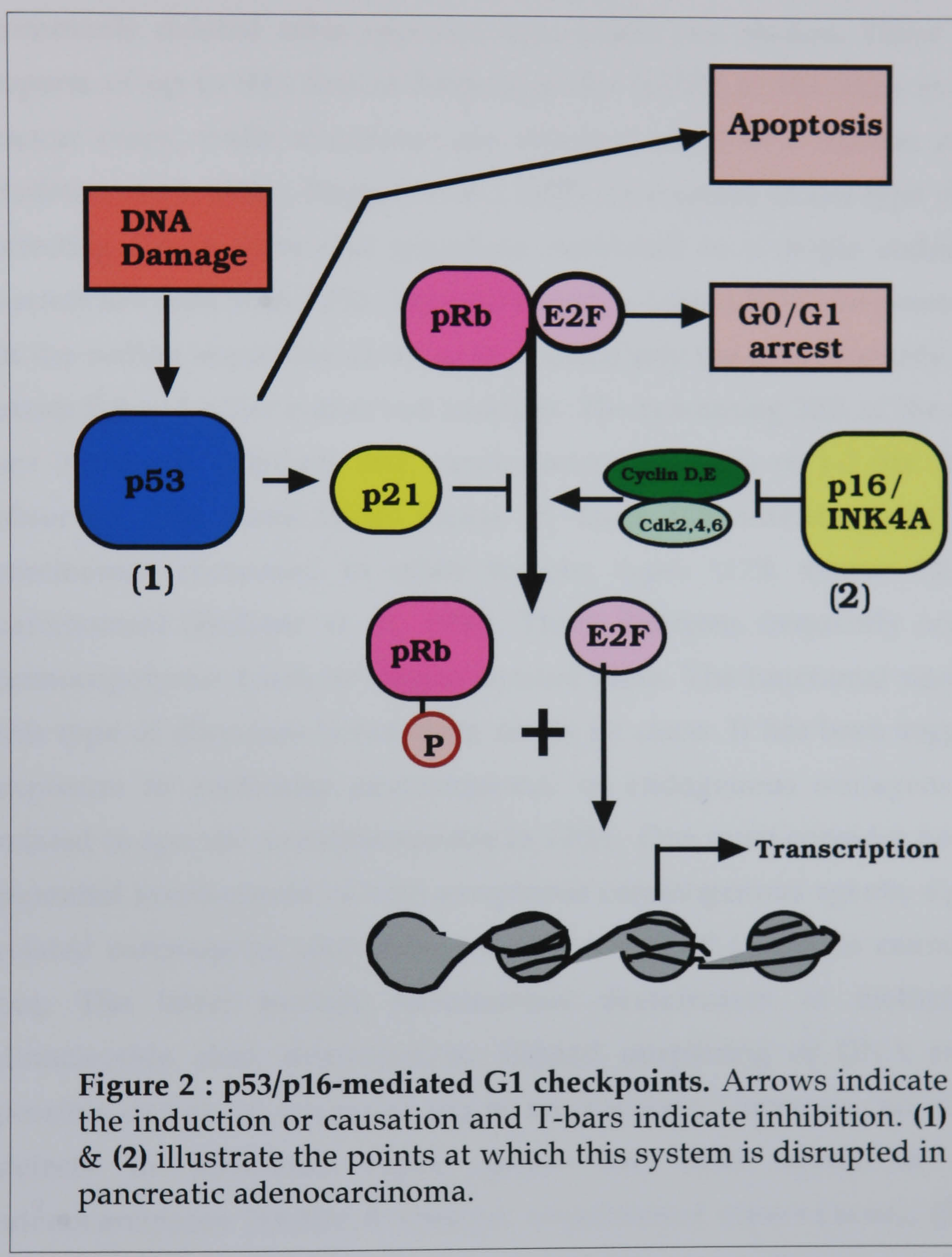


Figure 2 : p53/p16-mediated G1 checkpoints. Arrows indicate the induction or causation and T-bars indicate inhibition. **(1)** & **(2)** illustrate the points at which this system is disrupted in pancreatic adenocarcinoma.

1.4.5 Mutations of *TP53* in cancer

The *TP53* tumour suppressor gene may be the most commonly mutated gene in human tumours and this locus at chromosome 17p13.1 is one of the most commonly deleted areas reported from allele loss studies. There have been reports of up to 80% loss of heterozygosity (LOH) at this locus in pancreatic cancer cases, while mutations are observed in 55-60% (Barton *et al.*, 1991; Redston *et al.*, 1994 ; Ruggeri *et al.*, 1997). In contrast to the type of mutation affecting *KRAS* (missense mutations restricted to a single codon), a wide variety are seen with *TP53*. The vast majority (about 80%) are point mutations in the coding sequences of the gene, particularly the evolutionarily conserved exons 5-9 and other conserved residues. The remaining 22% of the alterations are intragenic deletions and, rarely, micro-deletions of 1-2 bp. It has been observed that there is an excess of these micro-deletions in pancreatic carcinomas compared to other tumour types (17% versus 4% in colon carcinomas) (Redston *et al.*, 1994). These deletions frequently occur within homocopolymer tracts or simple iterated bases. The functional significance of this type of alteration is not clear, nor is its cause. It has been suggested that exposure to particular environmental or endogenous mutagens might be related to specific mutation spectra in *TP53*, One must consider however, the potential involvement of both exogenous carcinogenous agents, e.g. smoking related carcinogens, and endogenous biochemical processes cannot be ruled out. The latter include spontaneous deamination at methylated CpG dinucleotide sites, depurination, slipped mispairing of DNA strands and possible defects in mismatch repair mechanisms. Currently, no evidence for defects in mismatch repair genes has been found in pancreatic adenocarcinoma (McKie & Olianni: unpublished observations), (Seymour *et al.*, 1994). It has also been observed that different mutations are favoured in colorectal, lung or U.V.-induced skin carcinomas. Knowledge of the causative agents or the mechanisms of genetic alterations might provide clues to the processes involved in neoplastic progression.

In a study by Ruggeri *et al*, a large series of frozen and archival non-neoplastic,

neoplastic and metastatic ductal pancreatic lesions (136) were analysed for abnormalities in the p53 gene/protein, mdm-2 and p21WAF1/CIP1 (Ruggeri *et al.*, 1997). Moreover, they attempted to relate the abnormalities to the tumour stage and grade of tumour. Previous reports had been small by comparison, looking at only a handful of cases and only one previous study included data comparing p53 mutations in primary and metastatic lesions from the same case. Their immunohistochemical staining data revealed abnormal p53 protein levels in 56% of cases of sporadic ductal adenocarcinoma of the pancreas in agreement with earlier studies including those published by Barton *et al* in this laboratory. Moreover, an identical percentage of metastatic pancreatic lesions revealed abnormal p53 staining and approximately 60% of primary tumours and metastases from the same patient also exhibited positive abnormal protein. These results (including analysis of 17 cases of paired primary and metastatic material) were confirmed by SSCP and DNA sequence analysis. This data supports the hypothesis that p53 alterations in primary pancreatic adenocarcinomas are not temporarily associated with metastatic progression but occur earlier in the tumourigenic pathway. In support of this Barton *et al* had found abnormal p53 staining in carcinomas *in situ* and fine needle aspirates (Barton *et al.*, 1991). These findings prompted Ruggeri *et al* to examine the p53 status of non-neoplastic pancreata exhibiting dysplasia and atypical nuclei (Ruggeri *et al.*, 1997). None of these samples revealed abnormal p53 immunostaining. Mutant p53's role in initiating adenocarcinoma of the pancreas cannot be ruled out and is further complicated by the fact that the histology of ductal pancreatic cancer and the identity of unequivocal precursor lesions is still under debate.

1.4.6 *p16/INK4A/CDKN2/MTS 1*

D-type cyclins induced in response to mitogenic signals act as growth factor sensors, progressively assembling into active haloenzyme complexes with their catalytic partners the cyclin-dependant kinases cdk4 and cdk6 during G1 phase. The primary function of cyclin D-dependent kinases is to phosphorylate the retinoblastoma tumour suppressor protein (Rb), allowing

progression of the cell cycle. Phosphorylation of Rb is a molecular hallmark of passage through the G1 restriction point, where cells commit to enter S phase and can complete the cell cycle in the absence of mitogens. A family of cdk inhibitors, the so called INK4 proteins specifically inhibit the activities of the cyclin D-dependent kinases, cdk4 and cdk6 (figure 2-(2)). without affecting other known cdks. Four such proteins have now been identified, (p16/INK4A, p15/INK4B, p18/INK4C and p19/INK4D) which act in a biochemically indistinguishable manner in binding to and inhibiting cdk4 and cdk6.

The convergence of the fields of cancer genetics and cell-cycle biochemistry brought the*p16/INK4A* gene to the forefront of cancer biology. Genetic analysis of familial melanoma kindreds established strong linkage to a tumour suppressor at the 9p21 chromosomal hotspot (Cannon-Albright *et al*, 1992). The *p16/INK4A* gene was proposed as a target of mutations and deletions at chromosome 9p21, exhibiting frequent allelic loss in a variety of tumour types (Kamb *et al.*, 1994 ; Nobori *et al.*, 1994). Point mutations and small deletions occur in spontaneously arising oesophageal carcinomas, biliary tract cancers, and pancreatic adenocarcinomas as well as in families with inherited predisposition to melanoma and pancreatic adenocarcinoma (Pollock *et al.*, 1996). In pancreatic carcinoma, the *p16/INK4A* gene was found to be inactivated in 80% of cases (Caldas *et al.*, 1994a). Nearly half of these were nucleotide sequence changes, all being accompanied by loss of the wild-type allele. The remainder involved homozygous deletion of both copies of the gene. The p16-cyclin D/cdk4-Rb pathway is a target for genetic disruption in the tumourigenic process, tumours lacking p16 invariably have a functional pRb protein and visa-versa, whereas tumours exhibiting cyclin D amplification retain functional p16 and pRb. Mice lacking *p16/INK4A* develop tumours early in life, whereas those lacking other *INK4* family members do not. The large homozygous deletions seen in pancreatic cancer commonly include the adjacent candidate tumour suppressor gene p15/*INK4B*, (another of the cyclin D/cdk inhibitors), it is possible that this gene is also important in the pathogenesis of pancreatic cancer. In contrast to

p53, little is known about the regulation of p16 expression or activity, or the circumstances in which p16 might act physiologically to promote cell cycle arrest.

The *p16/INK4A* gene story was further complicated by the discovery that a portion of the gene has the capacity to encode a second growth-inhibitory protein p14/ARF, (ARF stands for alternative-reading frame protein - the mouse homologue is called p19/ARF)) (Quelle *et al*, 1995). A distinct promoter drives the expression of *p14/ARF* exclusively, incorporating the upstream *p14/ARF*-specific 1β-exon encoding the biologically active portion of the p14/ARF protein. Point mutations in the second exon of the *INK4A/ARF* gene do not appear to compromise the p14/ARF function. Structurally p14/ARF shares no homology with proteins in the databases, but its overexpression induces G1 arrest through a mechanism that does not involve CDK inhibition. Despite the growth-arresting activity of the p14/ARF protein, its relevence to tumour suppression was initially doubted, because of the low incidence of tumour-specific alterations that exclusively targeted the 1β-exon. Recently however, a critical region that is deleted in glioblastomas maps to the region containing the 1β-exon (Larsen. 1996). Most significantly, a recent report identified *p14/ARF* rather than *p16/INK4A* as the crucial target in T-cell acute lymphocytic leukemia (ALL) that exhibit 9p21 lesions (Gardie *et al*, 1998). Moreover, although germline *p16/INK4A*-specific mutations are associated with cancer susceptibility, the majority of tumour-specific *INK4A/ARF* alterations are exon 2 deletions in humans and therefore will affect both the *p16/INK4A* and *p14/ARF* ORFs. These observations imply a role for p16/INK4A and p14/ARF in tumour suppression.

After submission of this thesis, Bates *et al*, showed that the E2F-1 transcription factor released from pRB sequestration directly activates expression of the human tumour suppressor gene *p14/ARF*, which binds to the MDM2-p53 complex and prevents p53 degradation. (Bates *et al.*, 1998). These results complete the pathway linking abnormal proliferative signals such as loss of

RB, with the activation of a p53 response, through E2F-1 and *p14/ARF*. They suggest that E2F-1, a protein inherently activated by cell-cycle progression, is part of a fail-safe mechanism to protect against aberrant cell growth.

1.4.7 TGFß and the Smad4 tumour suppressor gene

The transforming growth factor-Beta (TGFß) family of signalling molecules can regulate the proliferation of many cell types. Until recently, analysis of signalling by TGFß concentrated on the activation of serine/threonine kinase receptors at the cell surface. TGFß binds a type-II receptor kinase, which phosphorylates (and activates) a type-I receptor kinase that then initiates downstream signalling. However, the discovery of the MAD-related family of signal transducers has increased our understanding of intracellular events in the TGFß pathway. The family now includes the prototypic Drosophila gene, *Mothers against dpp* (Mad), three genes from *Caenorhabditis elegans* (sma-2, sma-3 and sma-4) and seven vertebrate members (sma1-7) (Padgett *et al.*, 1998). The Smad2 and Smad4 proteins are tumour suppressors and are involved in TGFß -mediated signalling. Recent investigations of these genes has uncovered a general framework of how the pathway functions. The activated TGFß receptor interacts with Smad2 and probably Smad3 phosphorylating them at a conserves SSXS motif at the extreme C-terminus (Zhang *et-al.*, 1996). Smad4 is not phosphorylated upon TGFß stimulation and in fact lacks the C-terminal SSXS motif required for phosphorylation by thee type I receptor. Smad4 is thought to modulate transport to the nucleus after binding to the phosphorylated Smad2/3 (Padgett *et al.*, 1998). Phosphorylated Smad2 then forms a heterodimeric complex with Smad4 and accumulates in the nucleus (Lagna *et al.*, 1996).

In the nucleus, it is thought Smad2/Smad4 complexes can regulate transcriptional responses by specifically interacting with DNA-binding proteins, such as FAST1 (Chen *et al.*, 1996). In later studies to determine whether other DNA sequences might bind to the Smad proteins, Kern *et-al* tested the binding of random pools of oligonucleotides (Zawel *et al.*, 1998).

The DNA-binding activity of *Drosophila* Mad was found to reside in its amino-terminal MH1 domain. To evaluate whether the analogous domain of the human Smad proteins could bind DNA, GST fusion proteins containing the MH1 domain of Smads 2,3 and 4 were constructed and expressed in *E. Coli.* Electrophoretic mobiltiy shift assays (EMSA) were performed using the purified proteins and labelled oligonucleotides. Specific binding of an octamer sequence - 5′-GTCTAGNC-3′ or its complement to Smad4 and Smad3 was identified, though not to Smad2. The 8 bp palindrome was called the Smad binding element (SBE).

1.4.8 Growth factor systems implicated in pancreatic cancer

Many of the polypeptide growth factors that control growth and differentiation of mammalian cells bind to cell surface receptors with cytoplasmic protein tyrosine kinase domains. These growth factors induce their receptors to undergo dimerisation and subsequent auto-phosphorylation on tyrosine residues. Receptor autophosphorylation stimulates binding of cytoplasmic signalling proteins that contain SH2 (or Src homology region 2) domains, which bind with high affinity to specific phosphotyrosine-containing motifs on stimulated receptors.

Epidermal growth factor receptor (*EGFR*) (also known as ErbB-1) is a transmembrane protein with intrinsic tyrosine kinase activity which is involved in normal cellular growth and differentiation (Wiley *et al.*, 1995). The *EGFR* gene is usually overexpressed at high levels in pancreatic adenocarcinoma due to increased gene transcription (Korc *et al.*, 1986 ; Lemoine *et al.*, 1992a). EGFR belongs to a family of closely related transmembrane proteins that include HER-2/neu (also known as ErbB-2), HER3 (ErbB-3) and HER4 (ErbB-4) (Kraus et al., 1989 , Plowman et al., 1993). HER-2/*neu* appears to bind to a family of ligands, whereas HER3 and HER4 bind to the family of heregulins only. In addition HER4 also binds to betacellulin, a potent mitogen produced by the β cells of the islets of Langerhans. It has been observed that HER-2/*neu* and HER3 are frequently

overexpressed in pancreatic cancers (Hall *et al.*, 1990 , Lemoine *et al.*, 1992b) and amplification of *HER-2/neu* has been found in a significant number of cell lines of other origin. A recent study has shown that the pattern of HER-2/*neu* expression was related to glandular differentiation and early oncogenesis of pancreatic cancer, but not related to survival differences. HER3 however, has been found to be associated with advanced tumour stage and shorter postoperative survival (Friess *et al.*, 1995).

To date at least 15 EGF-like ligands have been identified including TGF α, amphiregulin (AR), heparin-binding EGF-like growth factor (HB-EGF) and epiregulin. These ligands form a complex series of interactions with different members of the EGF receptor family that promote either homo- or hetero-dimerisation (Carraway & Cantley, 1994). Some of these ligands such as amphiregulin have the potential to act as an autocrine growth factor in pancreatic cancer (Funatomi *et al.*, 1997).

The fibroblast growth factors (FGF) are a family of structurally related polypeptides that are mitogenic for a broad range of cell type as well as mediators of a wide spectrum of developmental and pathophysiological processes. The FGF family presently comprises ten members (FGF1-10) with FGF1 (also known as acidic-FGF) and FGF2 (also known as basic-FGF) typical of the group being chemotactic towards fibroblasts, participating in tissue repair and promoting angiogenesis (Gospodarowicz *et al.*, 1986). They are detected at low levels in normal pancreas and exhibit increased expression at the mRNA level in chronic pancreatitis. There is increased expression of FGF1 and FGF2 in pancreatic carcinomas (60% and 50% respectively), which appears to correlate with advanced tumour stage (Yamanaka *et al.*, 1993). FGFs stimulate cellular responses by binding to and activating a family of membrane-bound tyrosine kinases, the fibroblast growth factor receptors (FGFR). FGFR1 and FGFR2 are expressed in a wide variety of cell types, while FGFR3 and FGFR4 have a more restricted pattern of expression. All four FGFRs are expressed in normal pancreas tissue at relatively low levels (Leung

et al., 1994b) whereas significantly upregulated expression of FGFR1,3 and 4 is observed in 60% of pancreatic cancers (Leung *et al.*, 1994a, Leung *et al.*, 1994b).

1.4.9 Chromosome 7 and pancreatic cancer

The presence of tumour suppressor gene(s) on 7q is suggested by several lines of evidence, including *in vitro* experiments and the detection of frequent 7q deletions by either cytogenetic or molecular techniques in different types of epithelial neoplasms. *In vitro* experiments showed that the introduction of a normal chromosome 7 in a carcinoma cell line resulted in inhibition of tumourigenicity of those cells when injected into nude mice, and that the insertion of intact chromosome 7 to immortalised human fibroblast cell lines restored senescence properties (Ogata *et al.*, 1993; Zenklusen *et al.*, 1994b). Deletions in 7q with breakpoints varying from q11 to q34 have been described by cytogenetic analysis in many different malignancies, including oesophageal and gastric cancers. Recent studies using microsatellite markers reported high frequencies of 7q LOH in gastric, colon, (Zenklusen *et al.*, 1995) head and neck, prostate, (Zenklusen *et al.*, 1994c) ovary and breast cancers (Zenklusen *et al.*, 1994a). In these five studies the minimal area of loss was localised at 7q31.1-q32, suggesting this region may harbour possible tumour suppressor genes. Although the extensive allelotyping (Hahn *et al*, 1995) only detected around 20% LOH on 7q, cytogenetic studies of 66 pancreatic cancers with abnormal karyotypes have shown that chromosome 7 is among the chromosomes most frequently affected by structural abnormalities (32 of 66 cases, 48%) (Griffin *et al.*, 1995; Johansson *et al.*, 1992). Recently, an apparent homozygous deletion has been observed in pancreatic cell lines using Arbitrarily Primed PCR (AP-PCR). These investigators also observed 55-80% LOH with a limited number of cases using 12 microsatellite markers (Achille *et al.*, 1996). The area of loss reported here overlaps with the minimal area of loss delineated in the above studies (Zenklusen *et al.*, 1994c)

1.4.10 Possible remaining molecular lesions associated with pancreatic adenocarcinoma

The profile of genetic mutations in pancreatic cancer is distinct from other neoplasms. Ninety percent or more of pancreatic cancers harbour activating point mutations in codon 12 of *KRAS*, the *p16/INK4A* tumour suppressor gene is inactivated in 80% of pancreatic cancers and p53 in 50-75% of cases. Inactivation of the DPC4 (Smad4) gene may be more specific for pancreatic cancer. Smad4 is inactivated in less than 15% of colorectal cancers and less than 10% of other major malignancies, however is inactivated in 50% of pancreatic cancers (Schutte *et al*, 1996).

Rozenblum *et al* summarised their analysis of 42 pancreatic adenocarcinomas for *KRAS*, *p16/INK4A*, *TP53* and *DPC4* mutations and have carried out investigations into the probability of associations between some of the common pancreatic cancer gene mutations and disease pathogenesis (Rozenblum *et al*., 1997). Associations were evaluated between the genetic status of the *p16/INK4A*, *TP53* and *DPC4* tumour suppressor genes and tumour diameter, node status, differentiation age at surgical resection survival and death of the disease. DPC4 inactivation always appeared accompanied by p16 inactivation ($p = 0.007$), however the converse was not true as p16 inactivation was not always accompanied by *DPC4* inactivation. Inactivation of p53 was not associated with *p16/INK4A* or *DPC4* mutations ($p = 1.0$). A trend for mutation of all three tumour suppressor genes and poor prognosis was also suggested (not surprisingly).

The vast majority of the tumour series Rozenblum and colleagues studied were found to harbour accumulated mutations in three or four of the genes analysed. Mutational analysis of some tumours in the series revealed the coexistence of *BRCA2* mutations with three of these genes and in one case all four genes outlined above. It is yet unclear whether this is a feature of adenocarcinoma of the pancreas, or whether such a high prevalence of mutation events have not yet been identified in other cancers. The evaluation

of genetic alterations as they naturally occur in human tumours allows the formulation of hypotheses concerning the biological processes that are involved in the tumourigenic pathway. A central belief of tumourigenesis is that positive selection is exerted upon those tumour cells that alter rate-limiting regulatory pathways, implies that mutation of one gene abrogates the need for mutation of another gene in the same tumour-suppressive pathway. It follows therefore in the genetic profile of these genes in pancreatic adenocarcinoma that *KRAS*, *TP53*, *p16/INK4A*, *DPC4* and *BRCA2* must each belong to a different tumour suppressive pathway. We should not overlook the observation that these rate-limiting pathways may interact. *DPC4* inactivation was observed to be dependant on prior *p16/INK4A* inactivation or indeed *p14/ARF* inactivation, therefore *p16/INK4A* (*p14/ARF*) inactivation may contribute a selective advantage for cells also carrying *DPC4* inactivating mutations to thrive.

Pancreatic cancer is likely to harbour changes in additional, yet uncharacterised genes. It has been hypothesised that up to seven genes might be altered in most human cancers and cytogenetic studies carried out on primary cultures and xenograft explants are also beginning to give some clues to the locations of genes which may be important in pancreatic cancer (Johansson *et al.*, 1992 and Bardi *et al.*, 1993). Microscopic abherations combined with allelotype analyses of pancreatic adenocarcinoma (Seymour *et al*, 1994 and Hahn *et al*, 1995) would indeed suggest the involvement of several additional tumour suppressor genes in this tumour type. Losses on chromosome arms 1p, 6q, 7q, 11p and 17p have encouraged our laboratory and others to investigate known tumour suppressor genes at these loci and in some cases initiate more detailed allele loss studies with microsatellite markers.

1.5 Tumour invasion and metastasis

As mentioned above, at the time of diagnosis approximately 85% of symptomatic patients have metastatic disease (Cubilla *et al.*, 1978). Even in tumours of 2cm or smaller, regional lymph node metastases are present in half of these cases. The aggressive spread of a pancreatic tumour is one of the main reasons for the abismal outlook for the disease. Neural invasion is a common feature of pancreatic cancer spread. Although invasion of neural tissue occurs in some other cancers, including cancer of the prostate, the prevalence of neural involvement in pancreatic cancer is much higher than in other cancers (Pour *et al.*, 1994). Therefore, recognition of neural invasion of pancreatic cancer and factors which may influence it are of great clinical interest.

The immunoglobulin (Ig) supergene family are cell surface adhesion molecules which possess Ig-like folds in the extracellular domains. Some Ig family adhesion molecules may be involved in tumour cell invasion and cancer metastasis. Two gene products, the carcinoembryonic antigen (CEA) and "deleted in colon carcinoma" (DCC) appear to function as dominant and recessive metastasis-related oncogenes respectively (Birchmeier *et al.*, 1991). CEA is a widely used tumour marker which is thought to mediate Ca2+ - independant, homotypic aggregation of cultured human colon carcinoma cells. A direct positive correlation has been observed between serum levels of CEA and the tumorigenicity of excised human colorectal carcinoma cells in nude mice (Benchmol *et al.*, 1989). DCC is a putative tumour suppressor gene located on human chromosome 18q21 and ecodes a protein with homology to neural-cell adhesion molecule, (N-CAM) (Fearon *et al.*, 1990). It could therefore be inferred that loss of this gene could induce a loss of cell adhesive properties leading to tumour metastasis. As it's name suggests, DCC was first identified as being present within large genomic deletions of chromosome 18q21 in sporadic colorectal tumours. More recent data discussed below has implicated similar large deletions in a great number of pancreatic tumours

which would make DCC a particularly interesting target when considering the rapid metastatic nature of this tumour with neural invasion. With the dicovery of the DPC4 (*Smad4*) gene adjacent to DCC in the deleted region of 18q (Hahn *et al.*, 1996b), the exact role, if any, of DCC in pancreatic tumorigenesis remains unclear.

1.6 New technology dissects molecular events in pancreatic adenocarcinoma

In an effort to circumvent the physiological problem of the host desmoplastic response Kern and colleagues undertook the development of a xenograft model of pancreatic adenocarcinoma (Schutte & Kern, 1995). At the time of surgery, 2 mm pieces of primary tissue are implanted into severe combined immunodeficient (SCID) mice. Only the neoplastic cells expand in the murine environment so that when the xenografts are harvested at a size of 1.0 cm^3, the residual "contamination" of non-neoplastic human cells is believed to be below the threshold of most molecular detection methods. The integrity of the xenograft explants were analysed by comparison of allele loss and p16/*INK4A* mutation analysis, showing consistent results. Using these methods, a complete allelotype of pancreatic adenocarcinoma was undertaken (Hahn *et al.*, 1995). Eighteen xenograft cases were used in a PCR-based allelotype using 283 polymorphic markers with an average coverage of 4.1 informative markers per chromosome arm. Highly frequent allelic loss (>60%) was observed at chromosomes 1p, 9p, 17p and 18q. Moderately frequent allelic loss (40-60%) was seen at 3p, 6p, 6q, 10q, 12q, 13q, 18p, 21q, and 22q. These results complement cytogenetic findings (Johansson *et al.*, 1992 and Bardi *et al.*, 1993) and allelotyping studies of our own carried out on primary tumour tissue where the pathology has been confirmed and only specimens exhibiting 70% carcinoma morphology were included in the analysis. We observe 35-40% allele loss at chromosomes 1p32-36, 6q26-27, and also 11p and 11q. (Ding, Perry & McKie - unpublished data) Initial LOH studies carried out in this laboratory (Barton *et al,* 1991) on a small series of cases observed no losses at 17p or 18q. Analyses on an equally small series of microdissected cases

detailed in this thesis has observed modest levels of allele loss around 17p13.3. Markers around the p16/*INK4A* multiple tumour suppressor gene highlight homozygous deletion of tumour cell lines in 30% of cases as well as LOH in a smaller percentage of primary tumour cases (McKie & Prenzel: unpublished data). Convincing LOH data was also accumulated for the long arm of chromosome 6, particularly band q26-27. Cytogenetic and molecular studies have shown that non-random deletions and LOH within chromosome 6q are frequent events in many human neoplasms including breast and ovarian carcinomas (Shelling et al, 1995), melanomas (Trent *et al.*, 1990), and other tumour types. Chromosome transfer experiments indicate that chromosome 6 can suppress the tumourigenic activity of melanoma cells (Trent *et al.*, 1990). Recently, a publication by Queimado *et al* refined the localisation of two deleted regions of 6q associated with salivary gland carcinoma (Queimado *et al.*, 1998). Salivary glands share many physiological similarities to pancreatic ductal structure and the more distal peak of LOH between D6S297 and D6S37 is common with the area of loss identified in our studies (McKie & Perry: unpublished data).

Over the past few years, positional cloning has been responsible for the identification of mutant genes causing some of the most common cancers. Although this formidable list is continuing to expand, there are reasons to believe that positional cloning may soon have to (at least) share the limelight with alternative methods. Mainly based around the Polymerase Chain Reaction (PCR), several "genome scanning" techniques have been recently developed and used by a number of laboratories trying to identify candidate genes involved in various tumour types.

1.6.1 Representational Difference Analysis (RDA)

Representational Difference Analysis (RDA) is a means of isolating DNA fragments that are present in only one of two nearly identical genomes (e.g. tumour and normal) (Lisitsyn *et al.*, 1993). It utilizes a subtractive DNA hybridisation method but differs from the conventional methods by using

"representations" of the genomes that have reduced complexity. Representations are generated by a PCR-based size selection applied to the restriction fragments of both genomes. RDA takes advantage of both subtractive hybridisation as well as DNA reassociation kinetics to favour the reiterated PCR amplified products. DNA from the tumour must represent a subset of the normal DNA of the patient, with the differences being those sequences that have been deleted from the tumour genome during tumourigenesis. Kern and colleagues have applied RDA successfully to identify DNA fragments that are deleted in pancreatic tumour tissues leading to the identification of a new tumour suppressor gene DPC4 (Deleted in Pancreatic Cancer 4) (Hahn *et al.*, 1996b), later termed Smad4. An area of homozygous deletion occurring in 30% of pancreatic carcinomas was identified at 18q21.1 (Hahn *et al.*, 1996a), excluding the previously cloned DCC (Deleted in Colon Cancer) gene (Fearon *et al.*, 1990). The deletion included Smad4, a gene similar in sequence to a Drosophila gene (mad) implicated in a transforming growth factor-beta (TGFß)-like signalling pathway. Mutations of the Smad4 gene were reported in 22% of pancreatic carcinomas not exhibiting homozygous deletions at 18q21.1. Furthermore, this gene has been implicated as being mutated in other tumour types such as colon (Thiagalingam *et al.*, 1996) and oesophageal tumours (Barrett *et al.*, 1996), thus it representing a potential multiple tumour suppressor gene.

Another fragment isolated by Kern and his colleagues with the RDA approach (DPC1) mapped within the minimal area of loss thought to harbour the second breast cancer susceptibility gene (BRCA2) (Schutte *et al.*, 1995). In an interesting story of fortuitous cloning, the breast cancer consortium narrowed their search onto the much smaller area of homozygous deletion around this RDA fragment where they found the BRCA2 gene. Paradoxically, subsequent BRCA2 mutation analyses carried out on breast cancer families led to the identification of families (particularly within the Ashkenazi Jewish population) who had germ line BRCA2 mutations, but often had increased incidence of pancreatic cancer also (Phelan *et al.*, 1996). Within these

populations germline BRCA2 mutations were identified in 7.3% of patients with carcinoma of the pancreas, comparing well with the prevalence of mutations in apparently sporadic breast (<4%) and ovarian carcinoma (<4%). Remarkably, a family history clinically suggestive of an inherited predisposition to cancer was not seen among the BRCA2 carriers with pancreatic carcinoma. Only one BRCA2 carrier with pancreatic carcinoma had a family history of breast cancer, which was present in only one family member. This despite the fact that the particular mutation was strongly implicated in the development of breast cancer, and the two other mutations observed would produce truncated protein biochemically similar to those reported from well defined breast cancer families. The authors of these studies did point out that the small family size (particularly with small numbers of female members) may explain the paucity of pancreas cancers observed. Other studies carried out have shown the number of pancreatic cancer cases to be roughly a tenth of the number of breast cancer cases (Phelan. *et al.*, 1996) (Thorlacius S *et al.*, 1996).

1.6.2 Differential display

In 1992 Liang and Pardee had described another PCR based screening approach to detect imbalances of transcribed sequences in targeted populations of cells such as tumours - Differential Display (DD) (Liang & Pardee, 1992). Until then, methods to distinguish mRNAs in comparative studies relied largely on subtractive hybridisation techniques. This technology had several disadvantages in that it is difficult to establish, is poorly reproducible, requires large amounts of RNA and is time consuming. The basic principle of differential display is to reverse-transcribe the two populations of mRNA being compared using degenerate anchored oligodeoxythymidine (oligo-dT) primers which anneal to the beginning of a subpopulation of the poly(A) tails of mRNAs. The anchored oligo-dT primers comprise of 11 to 12 thymines plus two additional 3′ bases (T-12 VN) where V can be guanine, adenine or cytosine and N is guanine, adenine, cytosine or thymine which provide specificity. Each degenerate anchored oligo-dT primer set will, in theory, direct

the synthesis of one quarter of the total mRNA population. The differential display of cDNA species has been successfully applied to pancreatic adenocarcinoma in house, with the identification of mRNA species that are specifically up-regulated in tumour cell lines. If we consider however, that these tumour cell lines have been raised from the culture of tumour explants, then we are again probably addressing molecular changes which are occurring late in tumour development. It is believed that by looking at pancreatic tumours at the DNA level, genomic changes which may have occurred at any stage in the tumourigenic pathway may be identified. One method that may be employed for such genomic screening is arbitrarily primed PCR (AP-PCR) (Welsh & McClelland, 1990), In this instance, arbitrary primers are used in a PCR to generate fingerprint profiles of genomic DNA. This technique is the subject of an active study in our group also and some comparisons of the two methods will be discussed later in this thesis.

1.6.3 The European Pancreatic Cancer Reference Library System (EPCRLS)

Taking into account recent estimates, the total number of human genes ranges between sixty thousand and one hundred thousand, of which only a fraction (12-20%) are transcribed at any one time depending on the cell type and differentiation status. Any mammalian cell (with the exception of brain) should therefore have an average expressed gene number of ten to twenty thousand. Thomas Gress and his colleagues at the University of Ulm have developed automated library gridding and picking robotics with advanced image analysis technology and have constructed high-density gridded cDNA libraries from primary cultures of normal pancreas and pancreas tumour. The regular gridded arrays of cDNAs display their main advantage when using complex probes hybridising with hundreds or thousands of the library clones. Hybridisations with labelled probes of DNA complementary to the complete mRNA of a cell or tissue (a cDNA pool) allows the identification of clones containing sequences abundantly expressed in the tissue used to generate the probe. Using cDNA pool probes of pancreatic tissue (normal, chronic pancreatitis and carcinoma) or cell lines, the libraries can deliver data about

the expression profile of many hundreds of transcripts simultaneously. Additionally, by using cell lines before and after treatment with carcinogenic reagents or cytokines and other drugs, it may be possible to identify which transcripts are being activated in the cells in response to these stimuli.

In summary, the use of gridded cDNA filter arrays allows pooling of all data produced by a variety of different approaches and laboratories. By making cDNA library resources from pancreas tissues publicly available, all users will have the possibility of working with the same set of cDNA clones and expression profile data can be collated from the different participating laboratories.

1.6.4 Comparative genome hybridisation (CGH)

Recent years have also witnessed the development of FISH-based techniques to circumvent the need to use a great many polymorphic markers in the allelotype approach. Comparative genomic hybridisation (CGH) put simply is a subtractive hybridisation process, where genomic DNA from both tumour and normal samples is hybridised onto a normal metaphase chromosome spread. The two "DNA samples being used as probes are labelled using nick-translation, each one with a nucleotide attached to different fluorescent reagents. It is then possible to detect areas on the chromosome spread where one of the colours may be missing or over-represented relative to the other. As with all FISH-based approaches, the success of this technique then depends heavily on sophisticated image analysis systems which are highly specialized and expensive. Despite its obvious attractions, relatively few reports have appeared in the literature, usually the analyses being carried out by experienced cytogenetics laboratories. Recent analysis of this kind has been carried out on DNA from primary pancreatic adenocarcinomas and results generated agree with allelotype data that is now compiled for this tumour type (Fukushige *et al.*, 1997). Losses are observed on chromosomes 1p, 3p, 4q, 6q, 9p, 13q, 17p and 18q, whereas gains were observed on chromosomes 1q, 3q, 13q, 14q, 17q, 7, 8q, 11q, 12p, 19q, and 20. Although the authors refined their

analysis further to frequent-moderate losses/gains the resulting areas defined still cover immense regions of chromosome banding, representing huge physical distances.

1.7 Alu-PCR DNA fingerprinting

Since the first report of specific DNA amplification using PCR in 1985, the number of different applications has grown almost as exponentially as PCR amplification itself. However, the fundamental prerequisite for the amplification of any DNA is knowledge of sequences flanking the region of interest which usually limits applications to regions of known DNA sequence. Nelson and colleagues first sought to amplify human DNA fragments of unknown sequence from complex mixtures of human and other species DNAs (Nelson *et al.*, 1989). In particular, they applied the PCR to isolate human DNA specifically from somatic cell hybrids retaining human chromosome fragments in rodent cell backgrounds. By employing oligonucleotide PCR primers specific for the abundant primate-specific short interspersed repeat element Alu, they were able to generate ladders of human DNA fragments or fingerprints from the human chromosomal DNA in the complex mixture. Furthermore this method proved to be invaluable in the rapid analysis of large inserts within YACs, preferable to the alternative of gel isolation of large DNA fragments. The technique, known as Alu-PCR or IRS-PCR (interspersed repetitive sequence-PCR) has also been extended to specific amplification of genomic DNAs from other mammals including mouse, rat, cow and pig. In human genome research, the approach has flourished over the past decade and is now an accepted tool for the analysis of the human DNA content of somatic cell hybrids, YACs, BACs, (Bacterial Artificial Chromosomes) and P1 bacteriophages.

1.7.1 SINEs and the Alu family

Two major classes of mammalian interspersed repetitive DNA families have been characterised on the basis of repeat unit length (Singer M, 1991). SINEs

are short interspersed nuclear elements. The most conspicuous human SINE is the Alu repeat family (so called because of early attempts at characterising the sequence using the restriction endonuclease Alu I). The Alu repeat is primate specific and has attained a very high copy number in the human genome. Other mammals have other types of related repeats such as the mouse B1 family. The other major class of mammalian repeat element are the LINEs : long interspersed nuclear elements. Human LINEs are exemplified by the LINE-1 or L1 element (also called the Kpn repeat because of early attempts at characterising this family using the restriction endonuclease Kpn I). The L1 element is however also found in other mammals such as the mouse, so is not of practical use to amplify human DNA sequences from a complex mixture such as a somatic cell hybrid or YACs.

The Alu repeat is the most abundant sequence in the human genome, with a copy number of around 7.5 x 10^5 copies per haploid genome which represents 7% of the 3 x 10^9 nucleotides. A more recent extrapolation of the amount of total retro-elements in the human genome, based on GenBank entries (over 40kb) suggests that a figure of 35% is more realistic (Smit, 1996). Alu repeats have a relatively high GC content, and although dispersed mainly throughout the euchromatic regions of the genome, have been reported to be preferentially located in R chromosome bands. These correspond to the pale chromosomal bands seen when using standard Giemsa staining and represent the most transcriptionally active regions of the genome. Regions of the genome such as the HLA-DR have been found to be more densely populated with retroelements like Alu (Andersson *et al.*, 1998), although this may merely reflect the high density of transcribed genes within this region.

The full-length Alu repeat is about 280 bp long and is usually flanked by short (6-18 bp) direct repeats. The typical Alu sequence is a tandemly repeated dimer, with the repeats sharing approximately 120 bp of sequence followed by a short sequence which is rich in A residues on one strand and T residues on the complementary strand. There is asymmetry between the two tandem repeated

BIBL LONDON UNIV.

units however; one repeat unit has an internal 32-bp sequence absent in the other (figure 3) (Strachan *et al.*, 1996).

Any function of the Alu repeat is unknown, however it has been suggested that their ubiquitous presence may promote unequal recombination which could be evolutionary advantageous by leading to gene duplication. In support of this, observations that regions of the HLA locus are particularly enriched for the SINEs have recently been made. The HLA loci are probably the most polymorphic in the human genome which is a necessary prerequisite to its function in antigen presentation. The two tandemly repeated units of the Alu sequence show a striking resemblance to the sequence for 7SL RNA, a component of the signal recognition particle which facilitates transport of proteins across the membrane of the endoplasmic reticulum. Because of this and the fact they are flanked by direct repeats, it is assumed widely that the Alu repeats have been propagated by retroviruses from 7SL RNA and therefore represent processed 7SL pseudogenes (Wilkinson *et al.*, 1994).

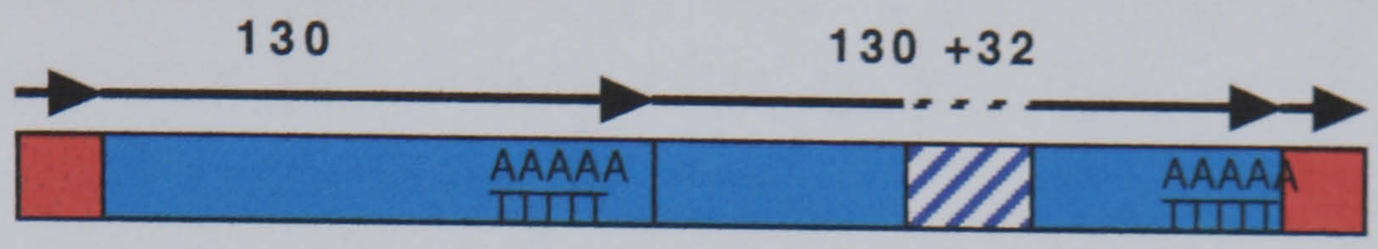


Figure 3: Structure of the full-length Alu dimer

1.7.2 SINEs as endogenous retroelements

The genomes of all eukaryotes contain multiple copies of DNA sequences that are related to sequences found in retroviruses. These elements are transmitted through the germ line as stable Mendelian genes, yet they exhibit structural and sequence similarities to exogenous retroviruses. These similarities have led investigators to speculate that endogenous retroelements are remnants of prior infections with exogenous retroviruses and with evolutionary time, changes have occurred to make them no longer infectious or pathogenic (Wilkinson *et al.*, 1994). These questions and hypotheses are far ranging and outwith the specific criteria of this PhD thesis however a concise overview of how SINEs, LINEs and other retroelements are related may serve to present clearer ideas of possible roles in the human genome.

A retroelement is a transposable or mobile genetic element which utilises an RNA intermediate in its replicative cycle. These elements are widespread in both vertebrate and invertebrate kingdoms and utilise a mechanism of transposition that involves the synthesis of an RNA intermediate which then acts as a template for an RNA- dependant DNA polymerase, commonly referred to as Reverse Transcriptase (Temin & Mizutani, 1970). All retroelements use reverse transcriptase to generate a DNA copy of their

sequence which can then be stably integrated back into the genome and complete the transposition process. This function appears to be very important in nature since it has been conserved throughout evolution. As a consequence of their common transposition mechanism, all retroelements share certain features. Firstly, because reverse transcriptase generates a new copy of the element, transposition increases the elements copy number within the genome. Further, as the site of integration is not restricted, there are multiple copies dispersed throughout the genome. Finally, during the integration steps, the target site is opened by staggered nicks to each DNA strand. This creates a short direct repeat of the target DNA sequence which immediately flanks the element on each end. These three features are characteristic of the many transposable elements. In addition, many retroelements contain an indication of their RNA intermediate state such as the presence of a 3 poly-adenosine tract downstream of a polyadenylation signal. Retroelements have been subdivided into two categories based on the sequence structure of the element. The primary distinction made among retroelements is whether the sequence is bound by long terminal direct repeats (LTRs) or not. Elements that contain LTRs are designated as retrotransposons, while those that do not contain LTRs are designated retroposons (Hull & Will, 1989). The retroposon group is a more heterogenous one and further distinctions can be drawn to subdivide this group. Those elements that do not encode reverse transcriptase (e.g. SINEs) are thought to be able to serve as efficient templates for the enzyme when it is supplied by another element (L1 element).

In humans and other mammals, retroposons were originally identified as repetitive sequences present at very high copy numbers within the genome. As observed above, these elements were originally classified on the basis of conserved restriction endonuclease sites. Coincidentally, these designations of SINEs and LINEs also correspond with the capacity of the element to encode reverse transcriptase (LINEs) or not (SINEs). The most abundant and well characterised example of this second class of retroposon are the Alu elements.

Alu elements are present in 5 x10 ^{5} - 10 ^{6} copies in the human genome. They do not contain an open reading frame (ORF) rather their nucleotide sequence indicates that they were derived from the gene coding for the 7SL structural RNA (Ullu & Tschudi, 1984). Because a cellular gene has been identified as the source of Alu elements, they can be regarded as a form of processed pseudogene. Processed pseudogene is the term used to designate a dispersed copy of a cellular gene, which by the absence of introns or the presence of a polyadenosine tract and flanking direct repeats, appears to be a reversed transcribed and reintegrated copy of that gene. Several important differences set Alu elements apart from most processed pseudogenes. Their extremely high number of copies indicates that they serve as efficient templates for reverse transcriptase. In contrast, the copy number of most processed pseudogenes is in the range of single copy to no more than a few hundred.

1.7.3. The evolution of the Alu element and possible roles in the human genome

Alu repeats occur only in primate species and can be grouped into sub-families which differ in their respective consensus sequences at a number of positions. These distinct Alu subfamilies are thought to be relics of periods of intense amplification that occurred in primates from the time of eutherian radiation until the present (Deininger *et al.*, 1992). For example, the average age of members of the human subfamily Alu-Sx, which accounts for nearly half of all Alu repeat copies in the human genome, is estimated to be about 40 million years old. This pre-dates the divergence of the New World monkeys. Members of the Alu-J subfamily are thought to be about 55 million years old dating from the time of prosimian divergence. Interestingly, Serine/Threonine kinases are known to contain Alu repeats of the Alu-J subfamily which suggests these genes have been around for at least 55 million years.and are more similar to the 7SL DNA at specific points (Jurka & Smith, 1988) As expected, the older repeats are common to all primates, while those amplified relatively recently are restricted to closely related species. Examples of the more recently amplified repeats are the Alu-S (intermediate) and Alu-Y

(young) subfamilies found only in humans and the great apes (Batzer *et al.*, 1996). The younger subfamilies of Alu sequences that are restricted to the human genome, some of which are polymorphic, have arisen through insertions in the NF-1 gene (Wallace *et al.*, 1991) and factor IX gene (Viddaud *et al.*, 1993).

In practise, Alu elements are found in the introns of almost all known protein-encoding genes. There is evidence to suggest all retroposons have played a role in evolution of the human genome by serving as a source of genetic variation, either through recombination or by providing a greater effective target size for possible mutation events. Of course one can say the converse is true in that by increasing the size of intronic sequences as targets for mutation, strictly conserved protein-encoding sequences are less likely to be mutated. As an example of this, the human Collagen I alpha1 gene (COLI A1) is a prime candidate for study. Small exons from 51- 306 bp are separated by kilobases of intronic sequence rich in repetitive sequences and may contribute to the conserved gene structure of fibrillar collagen genes. Insertions and splicing of intragenic Alu sequences into mRNA might contribute positively to the protein evolution if they are advantageous or neutral. However most of these events cause genetic disease due to the premature truncation of the peptide as Alu contains numerous stop codons.

1.8 Alu-PCR generated genomic fingerprints to identify genetic imbalance

1.8.1 The generation of sequence-tagged sites by Alu-PCR

Repetitive element fingerprinting in mammals has focused around Alu and LINE sequences, though pilot studies to look at more intermediate repetitive sequences (MIRs) and specific sequences motifs such as transcription factor-binding motifs have been described (Birkenmeier *et al.*, 1992). As highlighted in this introduction, Alu-based PCR strategies have been used as a tool since 1989 (Nelson *et al.*, 1989). Developed chiefly as a means of amplifying human DNA from a complex mixture such as that encountered

with a monochromosomal somatic cell hybrid, this application has now encompassed analysis and screening of large cloned DNA fragments such as YACs and BACs. Inter-Alu-PCR has been applied to the manufacture of region-specific libraries of gene sequences later (Klein *et al.*, 1993). In applying the Alu-PCR to generate region-specific STSs, Klein *et al* aimed to isolate transcribed sequence-tags from specific human chromosome sites (chromosome 22) (Klien *et al.*, 1993). Starting with immature heteronuclear RNA (hnRNA), the liklihood of identifying a cDNA through the inter-Alu-PCR would increase since Alu sequences are predominant in introns and 3 untranslated regions (3 UTRs). Upon achieving a smear of PCR products, fragments were cloned with the help of linker restriction sites being attached to the Alu-primers. The Alu-PCR products were column-purified from excess primers then digested with *Sal I*, phosphorylated and cloned into Bluescript™. Screening of this library of Alu-derived clones was achieved by random sequencing to generate sequence tags and localisation of these to chromosome 22. The authors concede that the percentage of isolated clones containing exonic sequence would be small (small genes). A possible refinement to the technique however, could be the application of long-range PCR which might result in the amplification of inter-Alu fragments with exons in the centre.

1.8.2 Alu-PCR identifies microsatellite instability in poly(dA) runs

The phenomenon of genomic instability plays an important role in the development of malignancies where increased rates of replication error phenotype (RER) make the alterations of critical cancer-related genes more probable. This phenotype was originally studied by typing a number of polymorphic microsatellites used otherwise as tools in linkage mapping studies (Ionov *et al.*, 1993). Inter-Alu PCR because of its multiplex character, represents a more general and less biased approach to amplifying a variety of genomic fragments with the potential to reveal the RER phenotype (Krajinovic *et al.*, 1996). In this approach, Krajinovic *et al* observe the most common band alteration to be changes in electrophoretic mobility. The most

frequent band shifts which were characterised were all due to contractions of the poly(A) tails of the Alu elements. This observation suggests that simple sequence repeats such as stretches of poly (A)n are the most likely microsatellites to express the RER phenotype. Diagnosis of the disease and the prognosis for the treatment may be tightly linked to the detection of the RER phenotype. In colorectal cancer, where surgery rather than chemotherapy is a major form of treatment, the presence of microsatellite instability is usually associated with a good prognosis (Thibodeau *et al.*, 1993). In contrast, in children with T-cell acute lymphatic disease, a mutator phenotype is associated with early relapse and a refractoriness to chemotherapy (Speicher, 1995). If such a pattern reflects resistance of mismatch repair-deficient cells to alkylating agents used in chemotherapy (Branch *et al.*, 1995), the detection of RER instabilities in patients that qualify for chemotherapy could be of crucial importance.

1.8.3 Alu-PCR Vs AP-PCR; an ongoing comparison

With the development of AP-PCR (Welsh *et al.*, 1990) in the laboratory it is proving interesting to compare and contrast the volume and quality of results so far being achieved. At an early stage in the AP-PCR analysis of pancreatic tumour/normal pairs, it seemed that more differences were apparently being picked up with the AP-PCR technique. Upon further analysis however, less of the AP-PCR fragments isolated appear to hybridise to specific bands on Southern blots and a higher proportion seem to contain repeated sequences. Difficulty in localising some of the AP-PCR fragments has been experienced in comparison to the Alu-PCR fragments we have had localised, particularly with FISH analyses. These are puzzling findings since we would expect the Alu-PCR species to be more likely to cross react/hybridise to Alu repeats at different chromosomal sites, especially when one considers the abundance of Alu in the human genome. Thus far the AP-PCR has been carried out with primers from literature (Achille *et al.*, 1996), it may prove fruitful to look at the design of these further and investigate whether there is perhaps a bias towards some simple repeat.

1.8.4 The fingerprint

Alu-generated DNA fingerprinting can be used to generate ladders or fingerprints of anonymous DNA fragments from human genomic DNA and we have extended this to compare the profile of bands achieved between tumour and normal DNA extracts from the same patient. By including only one Alu-specific primer in a PCR amplification, it is possible to amplify sequences between two Alu repeats in an inverted orientation relative to each other, providing they are sufficiently close for an amplified fragment to be generated (100-2,000 bases). A fingerprint of random amplified fragments can be displayed by gel electrophoresis. The abundance of the Alu repeat element will normally result in such reactions yielding a smear of Alu-generated intervening fragments. This pattern can be simplified however by predigesting the genomic DNA with frequent-cutting restriction endonucleases which cleave the DNA template at a proportion of the sites between inverted Alu repeats. The number of fragments generated is thus reduced to perhaps only a few dozen thus achieving a simpler, more easily interpreted band pattern. By repeating the amplification with a particular primer on the same DNA samples cut with different restriction enzymes, different Alu fingerprint patterns can be generated (see Figure 4). When the fingerprint obtained from a tumour DNA sample is compared with that of a normal DNA sample from the same individual, differences between the two indicate genomic damage or change. For example, fragments observed as being absent in the tumour track compared with the normal may indicate that part or all of the sequence is deleted in the tumour cells.

1.8.5 Southern hybridisation with Alu-morphs

It is accepted that a great many of the inter-individual differences found using this scanning technology may result from neutral polymorphisms in the DNA sequence acquired through mutations at or near the restriction enzyme recognition sites. Nevertheless, identification of these RFLP fragments may be of value to other scientists as markers should they prove to be in an interesting area of the genome and be significantly polymorphic. Therefore,

by hybridizing any fragments isolated first over Southern blots of tumour/normal pairs it is possible to assess whether the fragments contain unique sequences and which fragment changes are tumour-related and how frequently potential losses occur.

1.8.6 Genomic gridded libraries: a vital source of cloned material

Once tumour-specific sequence rearrangements are identified, gridded genomic libraries can be used to isolate larger fragments suitable for fluorescent *in situ* Hybridization (FISH), to assign the fragments to specific loci. Access to the Reference Library DataBase (RLDB) set up originally within the ICRF, means high-density gridded YAC, P1 and chromosome-enriched cosmid libraries are a rich resource. Further characterisation of these genomic clones through restriction mapping, subcloning and shotgun sequencing may identify similarities with existing gene sequences in the databases. Additionally, the generation of sequence-tagged sites (STSs) can be employed to refine the chromosomal location. STS markers have become important mapping tools simply because the existence of that sequence can be assayed very conveniently by PCR. Most STSs are nonpolymorphic and can be assigned to specific chromosomal locations by typing a panel of somatic cell hybrids. Once this is achieved a subchromosomal location can be ascertained by carrying out PCR on other reagents such as a radiation-reduced hybrid mapping panel or a YAC contig panel. STSs therefore have come to provide a critical link between genetic and physical mapping and as the density of STS markers (including polymorphic microsatellite markers) increases the need for specialist pulse-field gel electrophoresis (PFGE) has become minimal.

With the isolation of subclones encompassing the original Alu-derived fragments from the P1 and YAC clones, Northern hybridisation may then allow us to identify transcribed gene sequences. Again gridded libraries of cDNAs such as those available through the EPCRLS may provide a valuable resource in the construction of cDNA contigs and identification of transcripts.

1.9 Research Aims

(1) To develop DNA fingerprinting techniques and identify sequences which are altered in adenocarcinoma of the pancreas. Such sequences will be isolated and compared across additional tumour samples to assess the frequency of the mutation event.

(2) To identify candidate oncogenes and tumour suppressor genes associated with these genetic rearrangements by the using gridded genomic and cDNA libraries and subsequent generation of contiguous maps with the ultimate elucidation of gene structure and sequence.

(3) To analyse the pattern of expression of such isolated genes by Northern blotting, RT-PCR and in-situ hybridisation.

(4) To investigate the prevalence of such genetic rearrangements in human pancreatic cancers by the design of a simple PCR based diagnostic assay, which may be carried out on clinical samples including archival paraffin-embedded tissue specimens.

Chapter Two: Materials and methods.

2.1 Clinical specimens

Several surgical units provided tissues for investigations, including Professor R.C.N. Williamson, RPMS, London, Mr R.C.G.Russell, The Middlesex Hospital, London and Professor J.P. Neoptolemos, Queen Elizabeth Hospital, Birmingham. Details of the specimens can be found in Appendix 1. Samples of the resected tissue were snap frozen in liquid nitrogen immediately after surgery, and DNA was prepared from tumour-rich tissue blocks as described previously (McKie *et al.*, 1993).

2.2 Established tumour cell lines and SUIT 2 sub-lines

A number of human pancreatic tumour cell lines were used in this study. The pancreatic cell lines Panc1, CaPa-1, CaPan-2, BxPc3, MiaPaCa-2, HPAF, CFPac1, Hs766T and HuPac1 were obtained from the American Culture Collection (ATCC), Rockville, MD, USA. The cell lines PT45, SW850, SW979, 818.1, 818.4 and 818.7 were donated by Dr. H. Kalthoff and Professor W. Schmiegel, Department of Immunology, University Hospital Eppendorf, Hamburg, Germany, T3M4 was provided by Dr. T Okabe, Tokyo, Japan. The SUIT 2 cell line and the S2-sublines (described below) were kindly donated by Dr. T. Iwamura and Dr. M.A. Hollingsworth, The Epply Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Centre, Omaha, Nebraska, USA.

The human pancreatic tumour cell line SUIT 2, which was derived from a liver metastasis, shows spontaneous metastasis to lung and regional lymph nodes after subcutaneous injection of the tumour cells into the lateral flank of nude mice. Twenty eight sublines were cloned from SUIT 2 by culture in soft

agar and were characterised for their morphological and metastatic properties (Iwamura *et al.*, 1987;Taniguchi *et al.*, 1992) . Four of the sublines were selected for further study: S2-007 (a moderately differentiated tubular adenocarcinoma, highly metastatic), S2-020 (a poorly differentiated tubular adenocarcinoma, moderately metastatic), S2-013 (well-differentiated tubular adenocarcinoma, moderately metastatic) and S2-028 (papillotubular adenocarcinoma, rarely metastatic). Two additional cell lines were established by sequential selection either of lung colonies arising in nude mice following tail vein injection (SP-VP10, moderately differentiated tubular adenocarcinoma, high potential for lung colonisation) or of spontaneous metastases in the lung following subcutaneous injection (S2-CP9, moderately differentiated tubular adenocarcinoma, very highly metastatic to lungs). S2-VP10 and S2-CP9 are cell lines that were established after ten and nine cycles of selection respectively. All of these cell lines have distinct metastatic and morphological properties. They were maintained in Dulbecco's modified minimal essential medium (DMEM) supplemented with 5% foetal calf serum (FCS). DNA and Northern filters from these cell lines were a gift from Takeshi Iwamura, they were not maintained in culture in our laboratory.

2.3 Restriction digestion

Restriction digests of tumour/normal DNA samples from the same individuals was performed overnight as recommended by the enzyme manufacturer (Boehringer Mannheim). For DNA predigestion prior to Alu-PCR, ten units of restriction enzyme (1μl) was used in a 20μl reaction volume containing 1-5 μg of DNA (5μl). In the preparation of gels for Southern transfer, ten micrograms of DNA from paired tumour and normal samples as well as pancreatic cell lines was digested in 60-80 µl volumes with the addition of 10% the volume of restriction enzyme (60-80 units).

2.4 Alu-directed Polymerase Chain Reaction (Alu-PCR)

The conditions for PCR amplification were standardised; 50mM KCl, 10mM Tris-Cl pH 8.3, 1.5mM $MgCl_2$, 0.1% Triton X-100, 0.2mM dNTPs, 1 unit of BioTaq™ polymerase (Bioline Ltd) in a 50ul reaction volume. Bioline later only sold their BioTaq™ polymerase with the NH_4 based buffer. The amount of DNA template was equalised in the PCR reaction mixture by diluting the DNA stocks to comparable concentrations. The Alu primer concentration was a little higher (at 2 μM) than normally used in an attempt to favour annealing of the primer to the repeat elements by displacing any rapidly forming duplexes in the template by virtue of this high concentration. PCR cycle parameters were; 94^{0}C for 3 minutes, followed by 10-15 cycles consisting of 57^{0}C for 30 seconds, followed by 72^{0}C for 60 seconds, followed by 94^{0}C for 45 seconds. A final "chase" cycle of 72^{0}C for 5 minutes was added to allow complete elongation of all products.

Radioactive labelling of PCR products was achieved by the addition of 1-2 μCi of ^{32}P α-dATP (and the reduction of the unlabelled dATP in the buffer to 0.01 mM) followed by 20 more cycles of the above parameters. Alternatively, the PCR products were labelled by first tagging the primer with ^{32}P using γ-ATP and the enzyme T4 Polynucleotide Kinase (T4 PNK), see below. PCR amplification was carried out using a the Hybaid Thermal cycler ™.

2.5 Polynucleotide Kinase reaction

This procedure is useful for radioactive labelling of the 5' end of oligonucleotide primers. The Kinase enzyme requires that the 5′ end of the oligonucleotide has been previously dephosphorylated with alkaline phosphatase. This is not necessary with oligonucleotides made from our "in-house" machine as they are synthesized from the 5′ end on the column matrix and do not have a terminal phosphate group unless requested. The forward kinase reaction catalyses the exchange of the terminal γ-phosphate

which is labelled with ^{32}P from the ATP to the terminal phosphate on the oligonucleotide. Conditions are as follows; 50mM Tris-Cl pH 7.5, 10mM $MgCl_2$, 5mM DTT, 0.5mM spermidine, 100 pmoles of oligonucleotide, 50 μCi of ^{32}P γ-ATP, and 10 units of T4 PNK. The reaction is allowed to proceed for 1 hour at 37^{0}C.

2.6 Polyacrylamide gel electrophoresis

Polyacrylamide gels are made from acrylamide and N'N'-methylene bisacrylamide mixtures dissolved in electrolyte and polymerized by the addition of a chemical catalyst. The physical properties and resolving power of the gels are determined by the acrylamide concentration and the proportion of bisacrylamide added as cross-linker. The wide range of pore sizes which can be obtained by varying the ratio of acrylamide: bisacrylamide makes this a very popular choice and gels can be made for many different purposes. For the purposes of displaying the Alu-generated PCR ladders, a non-denaturing 6% acrylamide gel matrix was used with an acrylamide: bisacrylamide ratio of 29:1 in Tris-Borate-EDTA electrolyte (TBE). For the most part the gels were run relatively slowly overnight at 400V constant voltage supply. The apparatus used was the large Bio-Rad gel format which are 50cm long and we found an overnight run meant the Alu-PCR fingerprint was well displayed under these conditions. For sequencing, a denaturing 6% acrylamide gel matrix was used with an acrylamide: bisacrylamide ratio of 19:1 in TBE buffer equilibrated in 8M Urea.

2.7 Autoradiography & excision of bands from gels

After electrophoresis, the apparatus was dismantled and the gel mounted wet on a piece of old X-ray film or Whatman chromatography paper, covered with Saran wrap™ then exposed to X-OMAT™ X-ray film (Kodak) within a cassette at -70 ^{0}C overnight or for 3-4 days depending on the activity in the gel. After developing the X-ray film, any interesting bands were cut out carefully

from the gel with a scalpel after precise orientation with the autoradiograph. The fragments of gel were placed into 100μl of a buffer containing 1x T.E. (10mM Tris-Cl pH7.5, 1mM EDTA) supplemented with 0.1M NaCl and left overnight at room temperature allowing the DNA fragment to elute out of the gel matrix. The bands could then be re-amplified to provide a useful amount of DNA fragment by using 10 μl of the eluate in a PCR with the same oligonucleotide primers that generated them.

With the sequencing gels, all the gel solutions were equilibrated in 8 M Urea to ensure denaturation of DNA strands. Another consideration is that because ^{35}S or ^{33}P have a much lower energy of ß emission, the gels have to be dried down on a slab dryer with a vacuum pump drawing off the moisture. This ensures the lower energy ß particles are not quenched by the gel. These gels are then exposed at room temperature overnight or for 2-3 days as required. We later replaced ^{35}S with ^{33}P for this application. The ^{33}P has the properties of being a soft β particle emitter which has 2-3 times the energy of ^{35}S but still little enough compared to ^{32}P that it results in the same tight resolution for sequencing or Tissue *In Situ* Hybridisation (TISH). It has replaced ^{35}S as the isotope of choice for these applications in our unit, having a shorter half-life than ^{35}S it has important safety and radioactive waste storage/disposal implications also.

2.8 Cloning and sequencing of PCR fragments

Most of the routine cloning in our lab is carried out in Bluescript™ phagemid vector (Stratagene). The cloning of PCR fragments into the Bluescript™ plasmid was carried out by the method of Holton and Graham (1990), based on the observation that Taq polymerase adds a single non-template-directed deoxyadenosine residue to the 3′ end of duplex PCR products (Holton *et al.*, 1990). A blunt-ended vector is prepared by digestion with EcoRV, which is then tailed with dideoxythymidine triphosphate (ddTTP) using the enzyme terminal transferase. The use of ddTTP ensures the addition of only one T

residue. The vector ends thus generated have a single 3′ overhanging T residue which lacks a 3′-hydroxyl (OH) group and is therefore incapable of forming a phosphodiester bond. By contrast, the other strand can ligate to other DNAs since it contains a 5′- phosphate. A ddTTP-tailed vector can ligate directly to a PCR product that contains a 3′-overhanging A by forming a phosphodiester bond between the vector's 5′-phosphate and the PCR product's 3′-OH group from the overhanging A. Dideoxythymidine triphosphate tailing was performed using a terminal transferase kit (Boehringer Mannhiem). The linearised DNA was tailed in a 50ul reaction in 25 mM Tris-Cl pH6.6, 200mM Potassium cacodylate, 250mg/ml BSA, 1.5mM $CoCl_2$, 10mM ddTTP (Pharmacia), using 100 units of terminal transferase for 1 hour at 37°C. The DNA was then purified by phenol/chloroform extraction followed by ethanol precipitation.

Ligations were set up with a five to ten fold excess of the PCR fragment over the ddT-tailed plasmid vector DNA. Typically 50 ml ligation volumes were set up with 10µl of isolated PCR product, which had been purified from excess primers and dNTPs by Geneclean™ purification or the Qiaquick™ PCR-purification columns depending on the size of the amplicon. 50ng of ddT-tailed plasmid vector were included with 10 µl of a 5x ligation buffer (150 mM Tris-HCl, pH7.8, 50 mM $MgCl_2$, 50mM DTT, 10% PEG 4000,10 mM ATP) and 1 µl (400 units) of T4 DNA ligase from New England Biolabs.

Sequencing of interesting clones was carried out using the Sequenase kit™ (Amersham-United States Biochemicals) and ^{33}P α-dATP as the radionucleotide. Double-stranded DNA templates work quite well with the Sequenase kit, using the alkaline denaturation step outlined in the Sequenase handbook. Two to three micrograms of DNA template is denatured in 0.2N NaOH (5 mins at room temperature), the mixture is then neutralised by adding 0.4 vols of 5M ammonium acetate (pH7.5) and the DNA is the precipitated with 4 volumes of ethanol, (on dry ice for 10 mins). After washing the pelleted DNA with 70% ethanol, it is redissolved in 7µl of

distilled water, and 2 μl of the Sequenase buffer then 1μl of the T3 or T7 primers are added. For good results the Sequenase handbook recommends using purified, RNA-free plasmid DNA which has been maxi-prepped on a CsCl gradient. We find however, that mini-prepped DNA can be used, giving consistent results if it is first cleaned using the Geneclean™ kit (Bio 101 in La Jolla USA). This kit relies on the binding of DNA fragments onto a "glass milk" suspension and rapid washing and elution from the glass matrix. Latterly, the laboratory has switched to sequencing with the new ^{33}P labelled ddNTP terminator nucleotides with the new *Thermostable-Sequenase*™ kit from USB-Amersham. This new kit launched in January 1997 combines Amersham's new thermostable Sequenase with a cycle-sequencing protocol and the ingenious concept of labelling the dideoxy-nucleotide mixes so that only when an amplification product is terminated with one of these isotopic ddNTPs does the fragment get labelled. Therefore in contrast with internal labelling with a deoxynucleotide, the occurrence of parallel terminations due to secondary structure are minimised. The unit has automated sequencing services which require the DNA templates prepared with Qiagen spin-preps or tips to give high quality DNA for the Applied Biosystems sequencer.

2.9 Southern hybridisation

Filters carrying paired tumour/normal DNA samples digested with various restriction enzymes were prepared. The digested DNA samples of each case were electrophoresed on a 20 cm long 0.8% agarose gel for 16-20 hours. After depurination of the gel with 0.25N HCl, alkali blotting was carried out to transfer the DNA fragments onto Hybond N+™ nylon filters (Amersham-Pharmacia Biotech), according to the manufacturer's instructions. The probes were labelled either by radioactive PCR, (for small fragments) or by the use of random hexamers (Feinberg *et al.*, 1984). For the smaller fragments such as AB1 (270 bp) dual labelling was achieved by adding both α-dATP and α-dCTP in the form of ^{32}P labeled nuclides in the PCR reaction and once again the amounts of unlabeled dATP and dCTP were reduced to 0.01mM. After

competition with 0.1ug/ul sonicated human placental DNA (Sigma chemicals) for 2 hours at 65 ^{0}C following an established regime (Sealey *et al.*, 1985), filters were hybridised with the probes in 5x SSC, 5x Denhardt's, 0.5% sodium dodecyl sulphate and 1mg/ml yeast total RNA in bags at 65^{0}C overnight in a shaking water bath. The filters were then washed in four successive washes of 2x SSC:0.1% SDS, 1x SSC:0.1%SDS, 0.5x SSC:0.1%SDS and finally 0.2x SSC:0.1%SDS at 65^{0}C in the shaking water bath. The gradually decreasing concentration of salt has the effect of gradually increasing the stringency of wash thus removing non-specific hybridisation. Filters were exposed to Kodak X-OMAT™ film for 20-72 hours with intensifying screens.

2.10 Screening of genomic gridded libraries

Using current technologies, physical mapping of complex genomes is arduous and is hampered by various practical difficulties. Chimeric clones mean that cells contain two non-contiguous pieces of DNA from different parts of a desired genome. This may occur as a result of co-ligation of two different restriction fragments. Arranging clones in a physical order means relying on some clone-specific features. DNA sequences that contain repetitive DNA can make the task of ordering clones difficult. A major requirement for the efficient physical mapping of complex genomes is the need for gridding of DNA libraries. Following transformation of the host cells, individual bacterial or yeast colonies are picked into individual wells of microtitre dishes and stored in 96 to 384 well dishes. Each clone can then be assigned an identifying co-ordinate, comprising dish number, row number and column number. When libraries need to be screened by hybridisation with labelled DNA probes, high density gridded colony filters are robotically made in which each colony has been spotted on to the appropriate membrane on a grid with a uniform number of rows and columns. This simplifies the identification of hybridisation-positive colonies. These automated procedures allow the production of an unlimited number of identical membranes for each library. After an overnight incubation on agar plates, an *in situ* colony forms on the

membrane for each spotted clone.

An alternative to hybridisation of gridded nylon filters with radiolabelled nucleic acid probes is provided by the development of "pools" of clones which may be used as a template for specific PCR assays. This technology was available originally for the analysis of the ICI and ICRF YAC pools but has now been extended to the human PAC library. As an example of PCR-pools, the PAC library consists of approximately 120,000 clones in 315 384-well microtitre plates. The 21 primary pools each consist of the contents of 15 such plates. A primary screen involves PCR analysis of each pool with the sts amplicon perhaps resulting in 3 positive pools. The secondary PCR screen is carried out on the the 15 individual plate pools, then on the individual 384-well plates in the third round of screening.

2.10.1 P1 Bacteriophage library

A gridded P1 genomic library was obtained from the Genome Analysis Laboratory at ICRF Lincoln's Inn Fields. The bacteriophage P1 system allows the cloning of genomic DNA fragments that are as large as 95kb in size, (Pierce *et al.*, 1992) This capability significantly increases the repertoire of tools that are available for cloning and analysing complex eukaryotic genomes. Prior to the development of this cloning system, the two other systems used were cosmid libraries and YAC libraries, having a maximum capacity for foreign DNA of 40kb and 1000kb respectively. The YAC libraries have several limitations including low cloning efficiency, difficulty in isolating the cloned DNA and chimeric cloning artifacts. The P1 cloning system complements cosmid and YAC cloning systems by offering an intermediate range of clonable genomic fragment sizes along with the ability to isolate from *E. coli* large amounts of plasmid DNA by standard plasmid DNA preparation techniques.

The gridded library used the second generation P1 cloning vector pAd10-sacBII (Pierce *et al.*, 1992) which is currently considered the vector of choice for generating libraries. This vector features positive selection for clones with

inserts while restricting the growth of clones without inserts, also contains rare-cutting restriction sites and both T7 and Sp6 phage promoters directly bordering the *BamHI* cloning site to allow sequencing of both ends of a P1 clone to generate sequence tagged sites. The ability to produce RNA probes from both ends of a cloned genomic insert facilitates mapping efforts that involve chromosome walking strategies and contig assembly.

Positive selection of clones that contain DNA inserts is achieved by use of the sacBII cassette present in the pAd-sacBII P1 cloning vector. The sacBII gene is from *Bacillus amyloliquifaciens* and codes for the enzyme levansucrose which converts sucrose to levan. The accumulation of levan is lethal to *E. coli*, so expression of the sacBII gene will kill the bacterial host when they are grown in the presence of 2% sucrose. Cloning insert DNA into the cloning site within the sacBII cassette interrupts sacB expression and allows the the bacteria to grow on media containing sucrose. Single colonies from selector plates are picked and inoculated into the wells of 96-well microtitre dishes and grown for a period of time, then an aliquot of the culture is spotted from the microtitre plates onto 20x20 cm nylon membranes with a "spotting robot" These automated procedures allow the production of an unlimited number of identical *in situ* membranes for each library. Seven membranes containing 20,000 separate clones had been gridded out by robotic arm (representing three-fold coverage of the genome) and were screened with specific probes, then individual positive colonies could be isolated by sampling from the corresponding well in the original microtitre plate. Screening of the gridded membranes with specific probes was carried out after labelling the PCR fragments by including a labelled nucleotide in a further PCR reaction. Alu repeat sequences present in the probe are competed out in the presence of human genomic DNA from placenta for 2-3 hours at 65^{0}C before hybridisation is carried out in bags overnight again at 65^{0}C (Sealey *et al.*, 1985).

2.10.2 P1 artificial chromosome (PAC) library

The features of the P1 and F-factor systems have been combined to produce the P1-derived artificial chromosome (PAC) cloning sytem (Iounnou *et al.*, 1994). The new P1-derived vector, cCYPAC-1 has been constructed for the cloning of large DNA fragments by electroporation procedures as demonstrated for the F-factor based vectors (Shizuya*et al.*, 1992). The PAC vector is able to handle inserts in the 100 - 300 kb range and is reported to be more stable than the P1 bacterophage system. The new vector was derived from pAD10SacBII by deletion of the adenovirus stuffer fragment and insertion of a modified pUC19 plasmid into the *Bam HI* cloning site.

2.10.3 Cosmid Libraries

Cosmid vectors are hybrids derived from plasmids and the bacteriophage Lambda l. The l phage contains at each end single-stranded complementary stretches of DNA, the so-called *cos* sites which act as *cohesive termini* for concatenating multiple copies of the l genome. The l packaging enzymes then cleave this concatomer into the l-sized peices of DNA by recognising two *cos* sites approximately 48 kb apart. Thus, the *cos* sites are all thats necessary for packaging DNA into phage l. Therefore, plasmids containing the *cos* sites can accept up to 40 kb of foreign DNA between the two *cos* sites, then be "packaged" into *E. Coli* with the aid of *in vitro* l packaging mixes. Cosmids were developed in the late seventies as an improvement in cloning capacity from the bacteriophage l system. They have since been superceded by the P1, PAC and BAC libraries offering a more realistic link between smaller l based sytems (now exclusively used for cDNA cloning) and the large YAC vectors.

Nevertheless, cosmids are still prefered for the manufacture of chromosome-enriched libraries and for FISH analysis. The cosmid library used in this work was kindly provided by Dr. Anna-Marie Frischauf, but was not however gridded. Approximately 20,000 colonies were produced on a 22 X 22 cm square plate by suitable dilution of four separate aliquots of the cosmid library. Each aliquot had been calculated as allowing 2-3 times the coverage of the total

human genome. The average insert size of the cosmid library was given as 34 kb, cloned into the vector pCOS2-EMBL. Primary, secondary and tertiary screens had to be carried out to isolate a pure clonal culture. Additionally we had obtained three chromosome-enriched cosmid libraries (for chromosomes 1, 6 and 11) from the RLDB which had been gridded on nylon 22 X 22 cm squares. Each library was spotted onto 1-7 filters depending on the number of clones needed to cover the particular chromosome.

2.10.4 Varied cDNA libraries from pancreas and other tissues

The pancreatic cDNA libraries were prepared by Dr. T.M. Gress and his colleagues from the tumour cell lines PATU8988s and 8988t and from two pancreatic ductal carcinomas and a control normal pancreas tissue provided by the Department of Surgery, University of Ulm, Germany. The cDNA libraries were synthesized using oligo-deoxythymidine (oligo-dT) primers and reverse transcriptase according to standard protocols and clones into the plasmid vector pSPORT1™ (Life Technologies, Paisley, UK). For automated analysis, the cDNA libraries were presented in a regular gridded arrangement allowing computerised data analysis and processing as the P1 gridded library outlined above. Additional libraries were acquired from the MRC Human Genome Mapping Project (HGMP) at the Sanger Centre, Hinxton. They were supplied as aliquots of the cloned cDNA libary which then had to be transformed into the relevant *E.Coli* bacterial strain. To achieve high enough transformation efficiencies so that we attained a complete representation of the whole library, electroporation was employed. Libraries made from the colon tumour cell line HT-29 and the lung carcinoma (ref - 937218) by Stratagene. A cDNA library from normal colon tissue was also screened.

2.10.7 Electroporation of cDNA libraries into *E. Coli* cells

Electrocompetent MC106/P3 or DH5 α cells were either bought from Invitrogen or prepared in the laboratory. A fresh colony from an L-agar plate was picked and grown overnight shaking in 10mls of L-broth at 37 ^{0}C. The 10 mls overnight culture is then used to seed 1 litre of L-broth and the culture is

then grown to an OD_{600nm} of 0.5-0.7. The flask is then chilled on ice before the cells are pelleted in a pre-cooled centrifuge at 4,000rpm (2,800x g) for 15 minutes. The bacterial cell pellet is then gently resuspended in 500mls if ice-cold sterile deionised water supplemented with 10% glycerol. The centrifugation is repeated as above followed by the cell pellet resuspension in 250 mls of the ice-cold water/10% glycerol mixture. This procedure is repeated again resuspending in 100 mls of water/10% glycerol, and finally 4.0 mls of water/10% glycerol. The resulting cell density should be approximately $3x\ 10^{10}$ cells/ml, which is then split into 100ml aliquots and frozen immediately on dry ice and stored at -70°C for future use.

Electroporation was carried out on a Bio-Rad Gene-pulser™ following the instruction manual and detailed parameters laid out for these cell strains. MC106/P3 and DH5 α cells were electroporated with the following parameters:- 25mF capacity, 250 Ω resistance and 2.5Volts. Efficiencies of 10^9 cfu/ml were expected from the aliquoted library stocks received from the MRC, however values of 10^7 cfu/ml were achieved which was enough to give a high density coverage of 3-5,000 bacterial colonies on 8 - 10 133 mm plates.

2.10.6 Preparation of P1 bacteriophage, PAC and plasmid DNA

Extraction of DNA from plasmid constructs and cDNA clones was carried out using the Qiaprep™ spin miniprep kit protocol following the QIAGEN instruction booklet. This employs a modified alkaline lysis approach where the lysed *E.Coli* cellular extract is bound to a silica support in the presence of high salt on a spin-column insert provided. The plasmid DNA can then be washed with an ethanol/low salt wash, before being eluted from the column in deionized water. The Qiaprep™ spin miniprep kit is believed to be equally suitable for the preparation of cosmid DNA, however, we have found the yield and quality of cosmid DNA isolated this way to be poor.

Bacteriophage P1 DNA was extracted using the established alkaline lysis method, but extra care was needed to ensure all the solutions were relatively

fresh and particularly the solution 2 (0.2N NaOH/1.0% SDS) was made fresh each time. Additionally, care was needed to prevent mechanical shearing of these large plasmids during cell lysis and pipetting. As detailed in the text, the yields of P1 bacteriophage DNA were always very small (typically 20-60 mg of DNA from 500 mls of culture). Better yields were achieved later with the adsorption of the Wizard™ Maxi-prep kit from Promega Biochemicals, however 500mls of culture were still required to yield 150-200 mg. The Wizard™ Maxi-prep kit was also used to prepare cosmid and PAC DNA, however greater amounts were obtained often 400-500 mg, which is not dissimilar to a regular high-copy plasmid.

2.11 Northern Hybridisation

For many experimental purposes, unambiguous data and accurate measurement of RNA molcular size can only be obtained if all the secondary structure of the RNA is abolished during electrophoresis. The method most frequently used for fractionating RNA is by electrophoresis through agarose in horizontal slab gels, following denaturation of the RNA at high temperatures in the presence of glyoxal or formamide. Formaldehyde-containing gels are the method of choice in most laboratories now as maintainance of a strict pH level during electrophoresis is crucial to the glyoxal denaturation method. For electrophoresis in a 12 x 16 cm 1% agarose gel, mix 0.5g of agarose, 10 mls of 5 x MOPS running buffer, and 31 mls of water. Heat the mixture to dissolve the agarose, adjusting the final volume if necessary, then allow to cool to 55-60°C. In a fume cabinet, add 9 ml of 37% (12.3M) formaldehyde (final concentration 2.2M) and mix by swirling quickly, pour the gel into the gel-forming tray set up in the fume cabinet and allow it to set for at least 60 mins. After cooling, the gel is set up in the submarine reservoir with 1 x MOPS running buffer and the RNA samples are denatured by the addition of formamide to 50% final volume, 15% formaldehyde and 1 x MOPS buffer and heated to 55°C for 15 minutes. The samples are then snap cooled on ice and 2 mls of sterile loading dye mix is added: 50% glycerol, 1mM EDTA, 0.4% Bromophenol-blue, 0.4%

Xylene-Cyanol. Buffer recirculation is not necessary, as best results are achieved with a relatively short electrophoretic run of three-four hours at 70 V.

The Northern transfer is set up in the same way as the Southern, except the solvent used should be 10 x SSC and the process should be left overnight. In addition, it has been found that the binding of RNA fragments to the Hybond™ membrane is severely reduced by the presence of the formaldehyde, especially smaller RNA fragments. Therefore it is advisable to rinse the gel in 10 x SSC two-three times before setting up the transfer to wash out the formaldehyde. The hybridisation of Northern filters is carried out following exactly the same protocols.

2.12 Tissue *in-situ* hybridisation (TISH)

2.12.1 Preparation of RNA riboprobes

Plasmid M0982 containing the smallest cDNA insert of 1.6 kb had been cloned into the plasmid vector pSPORT1. The polylinker of pSPORT1 was flanked by the Sp6 and T7 promoters. Digestion with *Eco RI* lead to cleavage of the cDNA insert in the centre and within the polylinker at the T7 end, meaning a riboprobe could be generated from the Sp6 promoter of 800 bp which would prove ideal for TISH. Alternatively, digestion with *Hind III* again cut the insert around the middle of the cDNA and within the polylinker at the Sp6 end, meaning a riboprobe could be generated from the opposite strand using the T7 promoter, generating a probe size of 600 bp. Eventual hybridisation of each probe to duplicate slides resulted in a positive hybridisation from the T7 generated riboprobe which must represent the antisense strand.

In order to obtain high quality riboprobes, complete linearisation of the plasmid template was critical, therefore following restriction digestion linearisation was confirmed by gel electrophoresis. In some instances the plasmid digestion had not gone to completion, in which case the fraction of

linearised template was excised and extracted from the gel with Geneclean™ (Bio-101 INC. La Jolla). Subsequently, *in vitro* transcription was performed incorporating digoxigenin with a digoxigenin labelling kit (Boehringer Mannheim). As described in the manufacturer's instructions, transcription was performed for one hour at 37 ^{0}C in a twenty microlitre reaction mixture as follows: 1 μg of linearised template, 2 mls NTP labelling mixture, 2 mls 10x transcription buffer and 2 mls of polymerase enzyme (Sp6 or T7). The quality of full length transcripts was initially checked on denaturing polyacrylamide gels. The riboprobes were then precipitated in absolute ethanol to remove excess free nucleotides.

2.12.2 Preparation of sections from clinical specimens

Formalin-fixed, paraffin wax-embedded blocks from sixty resected pancreatic cancers were used in these and other studies which had been acquired from the Academic Hospital, Jette, Brussels and the Hammersmith Hospital, London. Sections of 5-6 mm were prepared and floated onto autoclaved DEPC-treated water for mounting onto 3-aminopropyltriethoxysilane (TESPA) - coated slides (Sigma, UK). After drying at 37^{0}C, the sections were de-waxed in xylene and rehydrated sequentially in 100%, 70%, 30% alcohol and finally in DEPC-treated deionised water. The sections were then incudated in a 0.02M hydrochloric acid solution for 10 minutes, followed by two five minute washes in phosphate-buffered saline (PBS) to remove the acid.

2.12.3 Prehybridisation protease treatment

Optimal permeablisation of the tissues was achieved by employing an incubation in PBS containing 0.01% (v/v) Triton X-100 for 90 seconds before digestion with proteinase K at 1 mg/ml for 10 minutes at 37^{0}C. The protease digestion was then terminated by incubation in PBS containing 2 mg/ml glycine for five minutes at room temperature, following which glycine was removed by a single wash in PBS. Endogenous alkaline phosphatase activity was quenched by treatment in 20% (v/v) aqueous acetic acid at 40^{0}C for 30 seconds. The sections were then post-fixed in a freshly prepared 4% (w/v)

paraformaldehyde solution in PBS for 10 minutes. This post fixation following protease digestion appeared crucial to avoid losses in signal when the mRNAs were disassociated and freed from the overlying proteins. Excess paraformaldehyde was then washed off with PBS and the sections were dehydrated in graded DEPC-treated alcohol, from 0% through 30% 70% and finally absolute. Following air drying, the sections were then ready for hybridisation overnight.

2.12.4 Hybridisation

Immediately before hybridisation, each digoxigenin-labelled riboprobe was diluted 1 in 20 in a hybridisation buffer containing 50% formamide, 5% (w/v) dextran sulphate, and 2x SSC, then denatured at 80^0C for 5 minutes. Sufficient hybridisation mix (15-30 ml) to cover the tissue sections was applied to the slides and coverslips put on. Hybridisation was continued overnight for 16-18 hours at 42^0C

2.12.5 Post-hybridisation washes and detection of signal

To remove unbound probes after the overnight hybridisation, the slides were incubated sequentially as follows at 42^0C : twice in 2x SSC for 15 minutes each, once in 1x SSC for 15 minutes and finally twice in 0.2x SSC for 15 mins each. Non-specific binding of anti-digoxigenin antibody to protein sites was blocked with a 3% bovine serum albumin (Sigma UK) solution containing 100 mM Tris-HCl, 150 mM NaCl, pH 7.5 at room temperature for 30 minutes. Incubation with 1:1000 diluted alkaline phophatase-conjugated anti-digoxigenin serum, (sheep polyclonal antibody, Boehringer Mannheim) for 2 hours at room temperature was used to report bound digoxigenin-labelled probes following rinsing in the above buffer. Signals were visualised with nitroblue tetrazolium salt (NBT) and 5-bromo-4-chloro-3-indoyl phosphate toluidinium salt (BCIP) in an overnight reaction in darkness. For each 10 mls of developing buffer (containing 100mMTris-HCl, 100 mM NaCl and 50 mM MgCl2, pH9.5, 40 mls of a 75 mg/ml NBT and 35 mls of a 50 mg/ml BCIP were added.

2.13 Pulse-Field gel electrophoresis (PFGE)

PFGE was carried out on a Bio-Rad CHEF-DR II™ apparatus following the guidlines included in the manual for the adequate separation of fragment sizes required (cosmids, P1 bacteriophage and PACs). Typically, 1.0% agarose gels were run in 0.5% TBE for 11 hours with an initial switch time of 1.0 second and a final switch time of 6 seconds at a voltage gradient of 6.0 volts/cm.

2.14 Multiplex PCR analysis

For the most part, multiplex-PCR reactions involved the combining of two sets of STS-specific primers which shared annealing temperatures and PCR conditions (or almost) in one reaction tube. Thus infrequent problems were encountered when amplifying two separate STS markers in the same reaction tube. Problems were encountered when trying to amplify STS markers of significantly different amplicon sizes. Such was the case with the M0982 cDNA (1,100 bp) and GAPDH (174 bp) STSs, where additional dNTPs and lengthier extension times had to be employed to amplify the larger cDNA amplicon. The multiplex reactions with GAPDH included seemed to be more dependant on template concentrations, with higher amounts of template resulting in smeared or even multiple bands with this primer set. Therefore, half the amount of DNA was routinely used in these multiplex reactions which prompted the dilution of all the microdissected DNA samples by four to try to achieve a more reproducible PCR from experiment to experiment. The relative intensity of control and test bands were compared visually, with the products produced from constituent normal DNA samples.

2.15 RT-PCR with Marathon™ cDNA amplification kit

Marathon™ cDNA amplification (Clontech laboratories Inc.) is a unified method for performing both 5′ and 3′ RACE (rapid amplification of cDNA ends) from the same template. Marathon cDNA amplification is a flexible tool - replacing conventional RACE kits to amplify just the 5′ or 3′ end of a particular cDNA, Alternatively, both 5′ and 3′ RACE may be undertaken to clone full-length cDNAs using the methods described in the Marathon cDNA cloning manual.

First strand synthesis uses a modified lock-docking oligo (dT0 primer which contains two degenerate nucleotide positions at the 3′ end. These nucleotides position the primer at the start of the poly-A tail and thus eliminate the 3′ heterogeneity inherent with conventional oligo (dT) priming.
The following were combined in a sterile 0.5 ml microfuge tube.

1 μg PolyA+ RNA or 20 mg total RNA
Incubate at 70 ^{0}C for 2 minutes, then snap cool on ice.
add the following to each reaction tube:
2 μl 5X 1st strand synthesis buffer
1 μl dNTP mix (10 mM)
1 μl α-^{32}P dATP (1 mCi/ml)
1 μl MMLV reverse transcriptase (100 units/ml)
10 μl Total volume

The α-^{32}P dATP is optional, addition of the isotope facilitates tracking the cDNA through subsequent steps and trouble-shooting when using the kit for the first time
The contents were mixed gently by pipetting and incubated at 42 ^{0}C for 1 hour, then the tubes were placed on ice to terminate the first strand synthesis.
Second strand synthesis is performed with a cocktail of *E. Coli* DNA polymeraseI, RNase H and *E. Coli* Ligase. These enzymes degrade the RNA

and synthesize the second cDNA strand.
The following components were combines in the reaction tube from the 1st strand synthesis.

10 µl 1st strand reaction volume
48.4 µl sterile H_2O
16 µl 5x 2nd strand buffer
1.6 µl dNTP mix (10mM)
4 µl 20x 2nd strand enzyme cocktail
80 µl total volume

The components were mixed thoroughly by gentle pipetting then incubated for 1 hour at 16^0C for 90 minutes.

Then 2 µl of T4 DNA polymerase was added and mixed gently and incubation was carried on at 16^0C for a further 45 minutes.
The T4 DNA polymerase creates blunt ends on the double stranded cDNA fragments. The second strand synthesis is terminated by the addition of 4 µl of EDTA.

The reaction mixture is then extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with 0.4 volumes of 4M ammonium acetate and 2.5 volumes of room-temperature ethanol. It is at these stages that inclusion of the α-^{32}P dATP is useful to keep track of the incorporated activity also giving an accurate interpretation of yield.
Following creation of blunt ends with T4 DNA polymerase, the double-stranded (ds) cDNA is ligated to the Marathon™ cDNA adaptor. This adaptor is partially double stranded and is phosphorylated at the 5′ end to facilitate blunt-end ligation of the adaptor to both ends of the ds cDNA by T4 DNA ligase.

Half the ds cDNA reaction volume then has the adaptor ligated onto it .

5 µl ds cDNA

2 µl Marathon cDNA adaptor (10 mM)

2 µl 5x DNA ligation buffer

1 µl T4 DNA ligase (1 unit/ml)

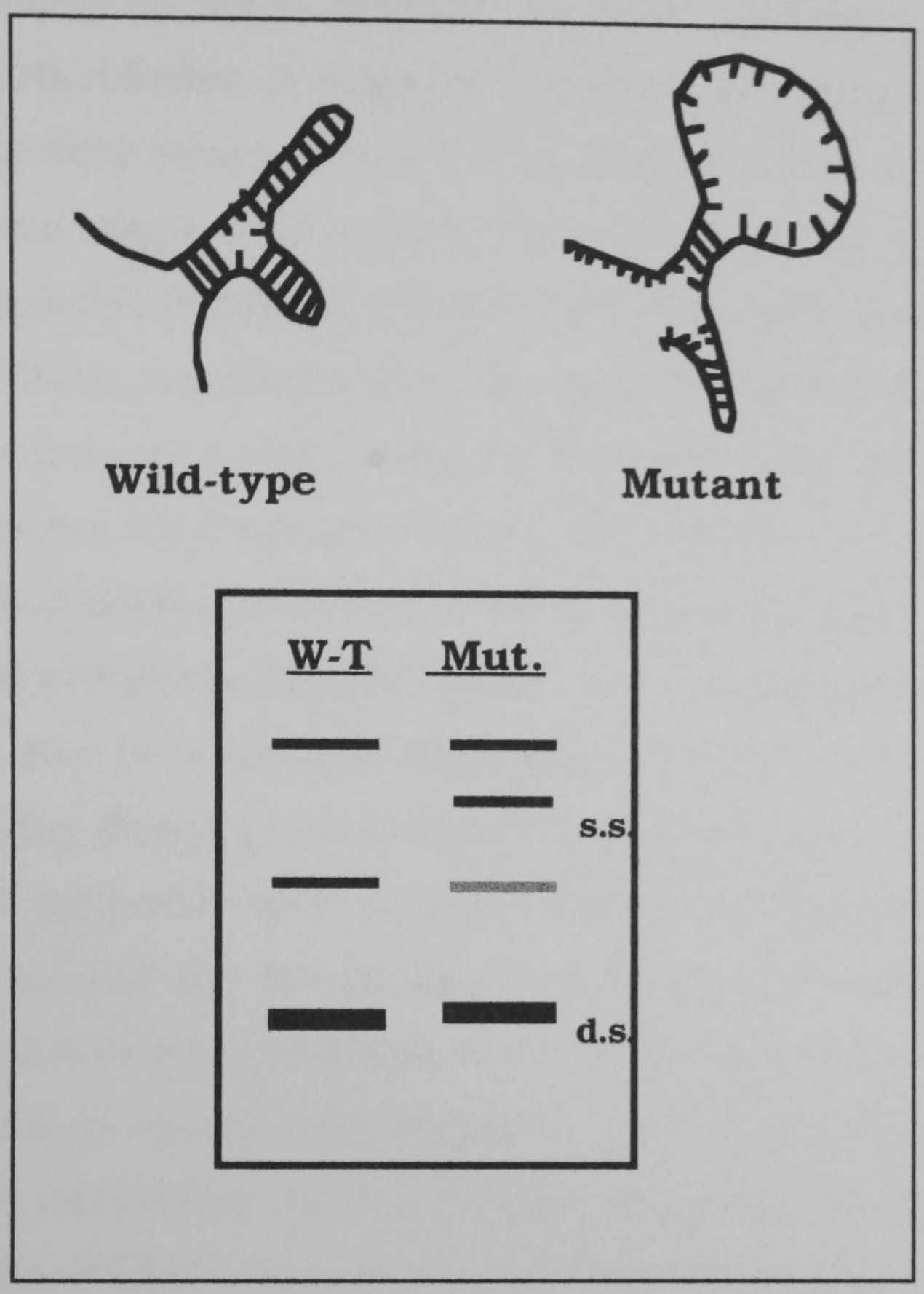
10 µl total volume

The ligations are carried out at 16^{0}C overnight. the the tubes are heated to 70 ^{0}C to inactivate the ligase.

The resulting library of non-cloned adaptor-ligated cDNA fragments is then diluted 1:10 with Tris/EDTA. The RACE reactions use only a fraction of the material generated for each RNA amplification, therefore there is enough material to carry out 20 individual amplifications.

2.16 Screening for mutations in candidate genes with SSCP

Single-Strand Conformational Polymorphism (SSCP) is one of the simplest, and more sensitive methods for screening multiple samples for the presence of unknown mutations (Orita, *et al.*, 1989). This technique relies on mutation-induced changes of the tertiary structure of single-stranded DNA molecules and the effects these conformational changes have on the mobility of the DNA molcules within a non-denaturing gel system (see figure below). The target sequence is amplified by PCR, then denatured by heat in the presence of 50% formamide before separation as single-stranded molecules by electrophoresis in a non-denaturing polyacrylamide gel. Sequence variants usually show differences in mobility, and the presence of mutations is revealed as the appearance of new bands in autoradiograms or silver stained gel.



Principle of SSCP, demonstrating how secondary nucleic acid folding influences conformer migration in the acrylamide gel matrix. As well as some residual normal single-stranded conformer, there may be the appearance of an alternative single-stranded conformer.

The tertiary structure of single stranded DNA changes under different physical conditions, e.g. temperature or ionic environment. Accordingly, the sensitivity of SSCP depends on these conditions. Fragment size is also an important factor. A single base mutation is going to have less consequence to the tertiary structure of a 400 bp fragment than a 200 bp fragment. Although one can adopt some empirical rules to screen for mutations within a gene, it is not possible to predict whether certain mutations can be detected under given conditions, especially when the mutation is in a new sequence. The mutation detection sensitivity however is believed to be high, being more than 80% in a single run for fragments shorter than 300 bp. As stated above, the sensitivity of SSCP decreases with increasing fragment length, therefore it is desirable to select primers so that the length of the amplifies product is less than 300 bp. This may be achievable when the genomic structure of a gene is known and one may design primers which flank individual exons. Large exons or longer DNA molecules such as whole cDNAs can be screened by designing primer sets so that the entire region is covered by overlapping short amplicons. Alternatively larger amplicons may be cut into several smaller fragments with the use of strategically identified restriction endonuclease cleavage sites. The SSCP undertaken in this thesis was on the Phosphatidyl-Inositol-Transport Protein (PI-TP) which fortunately consists of large introns and small exons.

Individual exons were amplified using standard PCR conditions of 50mM KCl, 10mM Tris-Cl pH 8.3, 1.5mM MgCl2, 0.1% Triton X-100, 0.2mM dNTPs, 1 pmol of each primer and 1 unit of BioTaq™ polymerase (Bioline Ltd) in a 20-50ul reaction volume. Following an initial denaturation at 94^0 C for 180 seconds, thirty cycles of amplification consisting of 57-59^0C for 30 seconds, 72^0 C for 60 seconds and 94^0 C for 45 seconds with a final "chase" cycle of 72^0 C for 300 seconds to allow complete elongation of all products. The amount of DNA template was equalised in the PCR reaction mixture by diluting the DNA stocks to comparable concentrations. Upon satisfactory amplification of the products, they were diluted 20-fold in 1x T.E. and 2-5 µl was used to seed a second "hot synthesis" reaction where 1 µCi ^{32}P or ^{33}P α-dATP was added to a

reaction and the dNTPs added were low in dATP (0.2mM dTTP, dGTP, dCTP, 0.02mM dATP). A further 10-15 cycles of PCR employing the same parameters as above were undertaken, then 5 mls of the final product was added to 5 mls of the formamide loading buffer for analysis.

It has been found in the laboratory that high detection sensitivity (single base point mutations) can be achieved over a wide range of templates using large a gel apparatus from Bio-RAD with 6% acrylamide gels (49:1 monomer:bis-acrylamide ratio) in 0.5x T.B.E. supplemented with 10% glycerol and run overnight (15 W constant power) at room temperature.

2.17 PCR-based Methylation detection system - CpGenome™

2.17.1 The mechanism

The CpGenome™ DNA modification kit from Oncor Inc. utilizes technologies exclusively licensed from The Johns Hopkins University School of Medicine. The kit contains the reagents required to perform a bisulphite modification on DNA samples. In this bisulphite reaction, all unmethylated cytosines are deaminated and converted to uracils while modified 5-methylcytosines remain unaltered (Yuanxiang *et al.*, 1997). Thus the sequence of the treated DNA will differ if the DNA was originally methylated versus unmethylated. Primers can be designed to specifically amplify each of the sequences based upon these chemically-induced changes. PCR products can further be analysed on a gel, or by sequencing.

(The DNA modification protocol was followed using the kit manual supplied by Oncor Inc.)

2.17.2 Primer design

The methylation-specific PCR analysis consists of the chemical modification of test DNA with the bisulphite followed by PCR amplification. The chemical modification creates the differences between the methylated and

unmethylated DNA, converting unmethylated cytosine to uracil through deamination. Two PCR primers should be designed to anneal to the alternative sequence differences. One primer set (U) will anneal to unmethylated DNA that has undergone a chemical deamination. A second primer set (M) will anneal to the methylated DNA that has not undergone the chemical modification due the the methyl group. The standard rules for primer design apply for the creation of the methylation-sensitive PCR (MSP) primers. They should be approximately 20-22 bp in length with similar melting temperatures and the product produced should be 100-200 bp in length. To minimize primer-dimer formation, primers should not be complementary, especially at the 3′ ends. Since most of the CpG islands are located within the 5′upstream sequences of genes which may be 70-90% GC-rich, discrimination between the methylated and unmethylated templates is best achieved when the last base on the 3′ end of the primers is complementary to the base under analysis in the test sequence.

One of the primers designed to amplify the AB1 fragment sequence originally, had the second cytosine nucleotide of a *Hpa II* recognition sequence as its last 3′ nucleotide. In order to function as an MSP primer however, primers complementary to this strand would have to be synthesized with either G or A as the last base at their 3′ end.

```
TGGTTAGAAAATCACTCCTCA CcGGG----------
                      ------- gGCCCTTGTGTGAATGACCATG
                              tGCCCTTGTGTGAATGACCATG
```

In practise, no amplification was achieved substituting either (M or U) primers for the original in conjunction with an "upstream" primer. As an alternative, the whole area was amplified with two flanking primers which amplified 212 bp of sequence around this *Hpa II* site and the resultant PCR fragments were sequenced automatically after purifying them from excess primers and nucleotide.

Chapter Three : Results

3.1 Establishment of the Alu-fingerprint technique

3.1.1 Restriction enzyme predigestion

Development of this technique followed pilot experiments carried out in the laboratory of Dr. Finbarr Cotter at the Institute of Child Health which showed promise in generating reproducible profiles from human DNA samples. The first experiments were set up in our laboratory using the Alu-specific primer BK-33 on undigested and EcoRI cut DNA from four paired tumour and normal samples. The results of this first experiment were disappointing in that both digested and undigested DNA samples produced similar continuous smears of amplification products which appeared more abundant higher up the acrylamide gel at the limit of its resolving capacity (figure 4a). It was postulated that the *EcoRI* restriction enzyme might be inappropriate and was not cutting the human DNA frequently enough to reduce the number of amplified products in the range which could be resolved by acrylamide gel electrophoresis. A second experiment proved more successful with the same primer and PCR conditions after digestion with the restriction enzyme *Alu I*. Over forty discrete bands were resolved on this second gel with a constant band pattern not only between tumour and normal DNA from the same case, but between different cases (compare figures 4a & 4c). This result was surprising as with 7.5×10^5 possible priming sites and the added complication of cleaving the DNA template with a restriction enzyme added to slightly differing DNA/salt concentrations variable profiles for each sample were anticipated. In contrast, almost perfect reproducibility was observed between all samples with differences in band mobility in only a minority of bands from individual to individual reflecting natural polymorphism in the population around *Alu I* restriction enzyme sites (see figure 5 for example of polymorphism).

Figure 4: Comparison of fingerprint profiles with different restriction enzyme digests

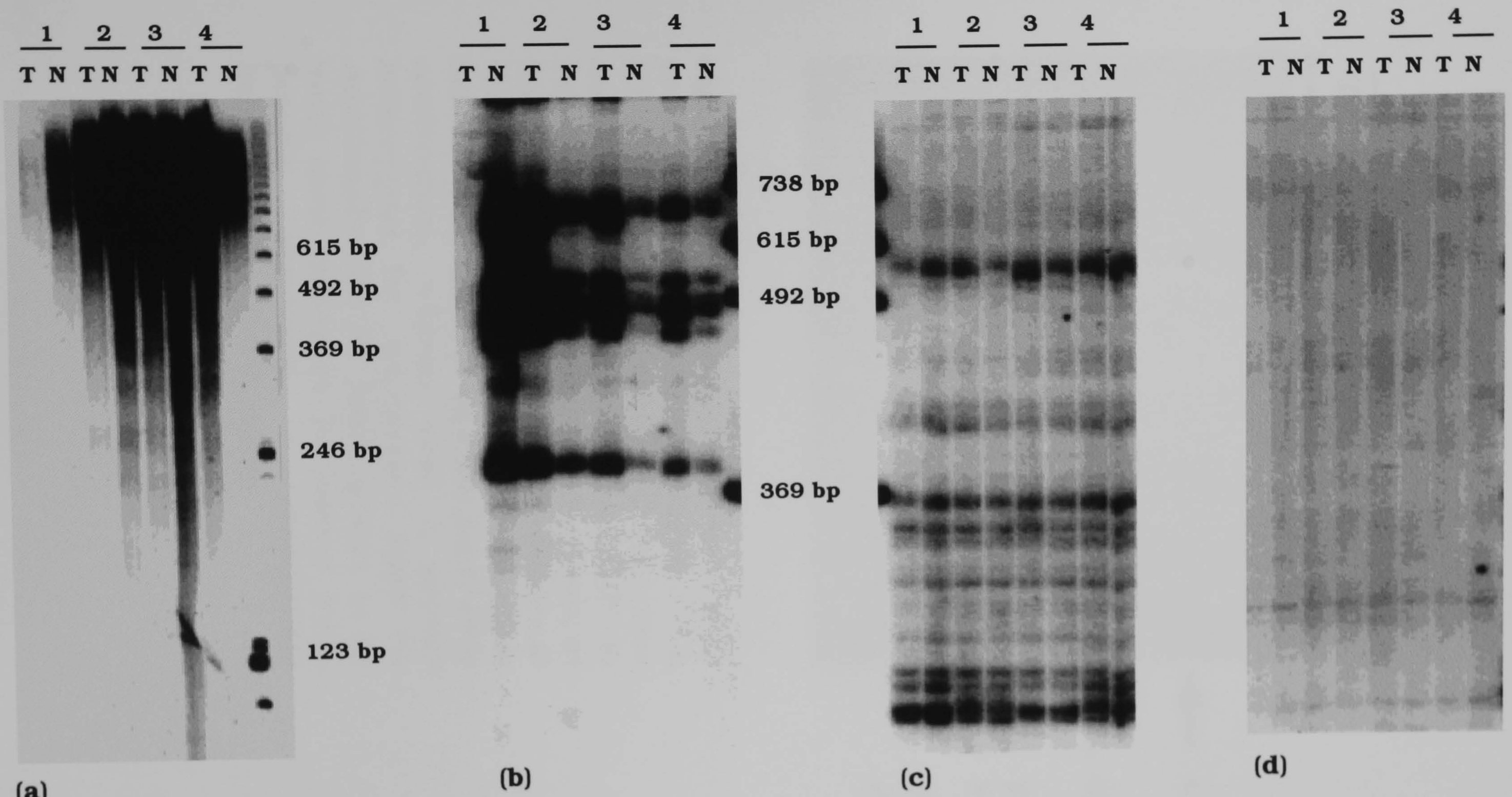


Figure 4: Autoradiograph illustrating Alu-fingerprint profiles achieved with the Alu-specific primer BK-33 following digestion of four paired tumour/normal cases with different restriction enzymes; (a) Eco RI, (b) Hinf I, (c) Alu I and (d) Rsa I. The autoradiograph (a) was produced seperately, and accordingly has a seperate molecular size marker

Figure 5: identification of a polymorphic locus with Alu-PCR fingerprinting

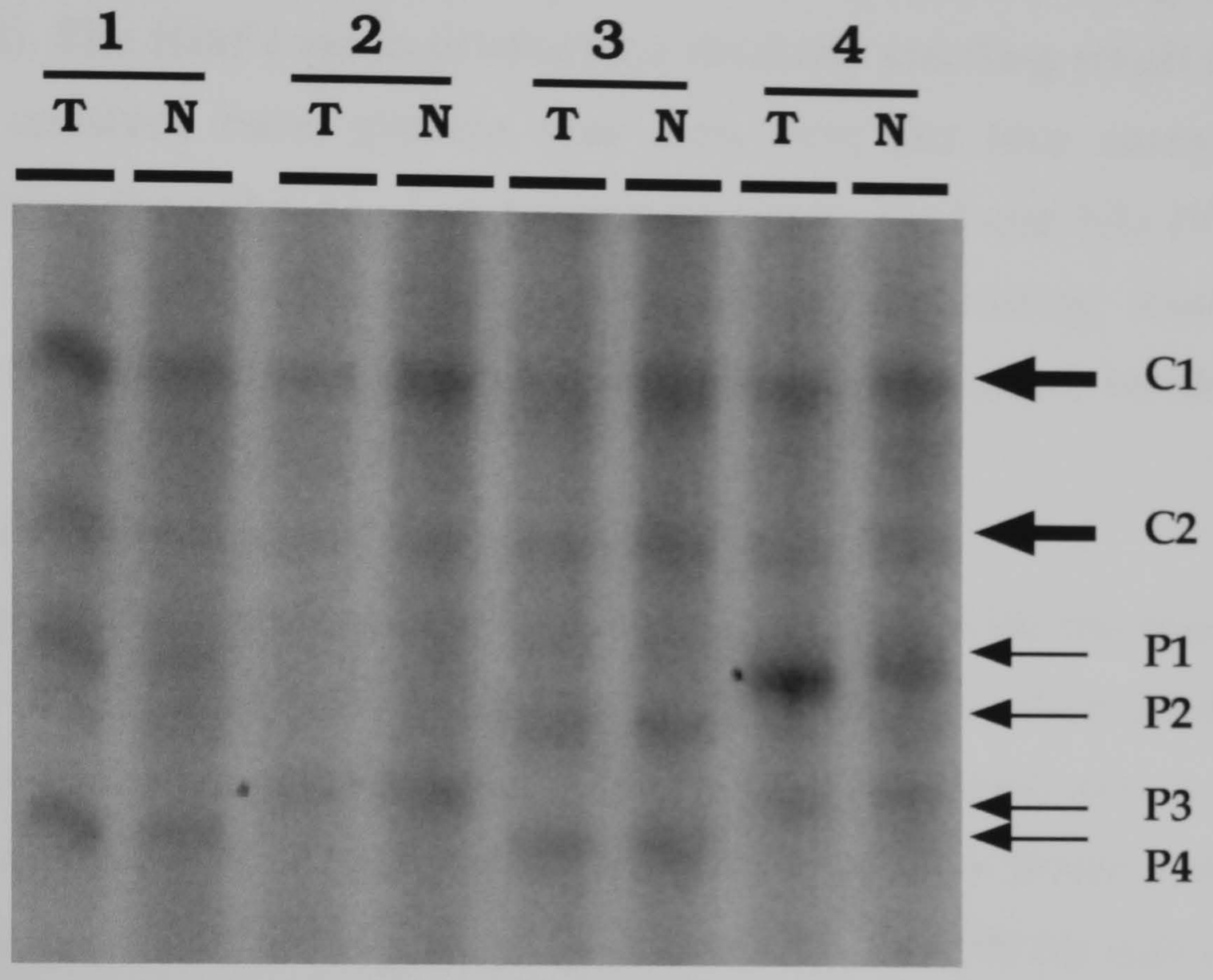


Figure 5 : This small area from a fingerprint profile illustrates two constant bands (C1 & C2) across tumour/normal DNA from four cases and four polymorphic bands (P1 to P4). From the fingerprint it is clear none of these individuals share the same pattern of bands/alleles. Case 1,3 & 4 appear heterozygous for three separate band/allele combinations whereas case 2 appears homozygous for polymorphic band P3. Additionally, the P1 band appears to be much more intense in the tumour track, possibly suggesting a chromosome imbalance due to ploidy status.

In conclusion, *Alu I* is cleaving the vast majority of potential templates for Alu-directed PCR and there are around forty remaining intact template sites that are all the same and equally accessible to priming and PCR. The restriction enzymes *Hinf I* and *Rsa I* were compared alongside the *Alu I* (figure 4). The *Hinf I* again produced a similarly startling result in that a more or less constant band pattern was seen over the four cases, though less informative than the *Alu I* restriction enzyme. *Taq I* and *Nla IV* proved to be equally successful, though the more frequently cutting *RsaI* and *Sau3A* produced less informative patterns of fewer, smaller sized bands (figure 4d).

3.1.2 Comparison of different primers

To compare the different band patterns obtainable with the four Alu-specific primers in the laboratory at this time, each primer was then used in PCR with four tumour/normal pairs. The Cotter IV primer produced a cleaner band pattern with less of a background smear and slightly fewer bands (25-30). A higher molecular size range of fragments from 400-1500 bp was observed. The primer BK-33 produced 40-50 bands in the fingerprint profile, with a much wider range of fragment size from 150 bp to 700 bp although the bands were perhaps not as tightly resolved. A third primer Alu-V, produced the widest range of band sizes from 200 bp -1,000 bp, though perhaps its upper range was near that achieved by the Cotter IV primer. Alu-V did however have a more smeared background signal compare with the other two primers so far mentioned (figure 6a-d). Interestingly when we used TC65, one of the original Alu-specific primers described, (Nelson *et al.*, 1989) a much less informative fingerprint pattern was achieved with two to four bands of a mid-range molecular size. This primer was by far the longest of the four so far tested, while BK33 is the shortest. Therefore was the size of the primer recognition sequence the only factor which governed the size range and complexity of banding pattern? If so, what would be the smallest primer one could employ before recognition fidelity was compromised. Did the primer sequence location within the Alu repeat element have any effect on pattern complexity? Some of these questions were addressed later.

Figure 6: Comparison of different Alu-specific primers

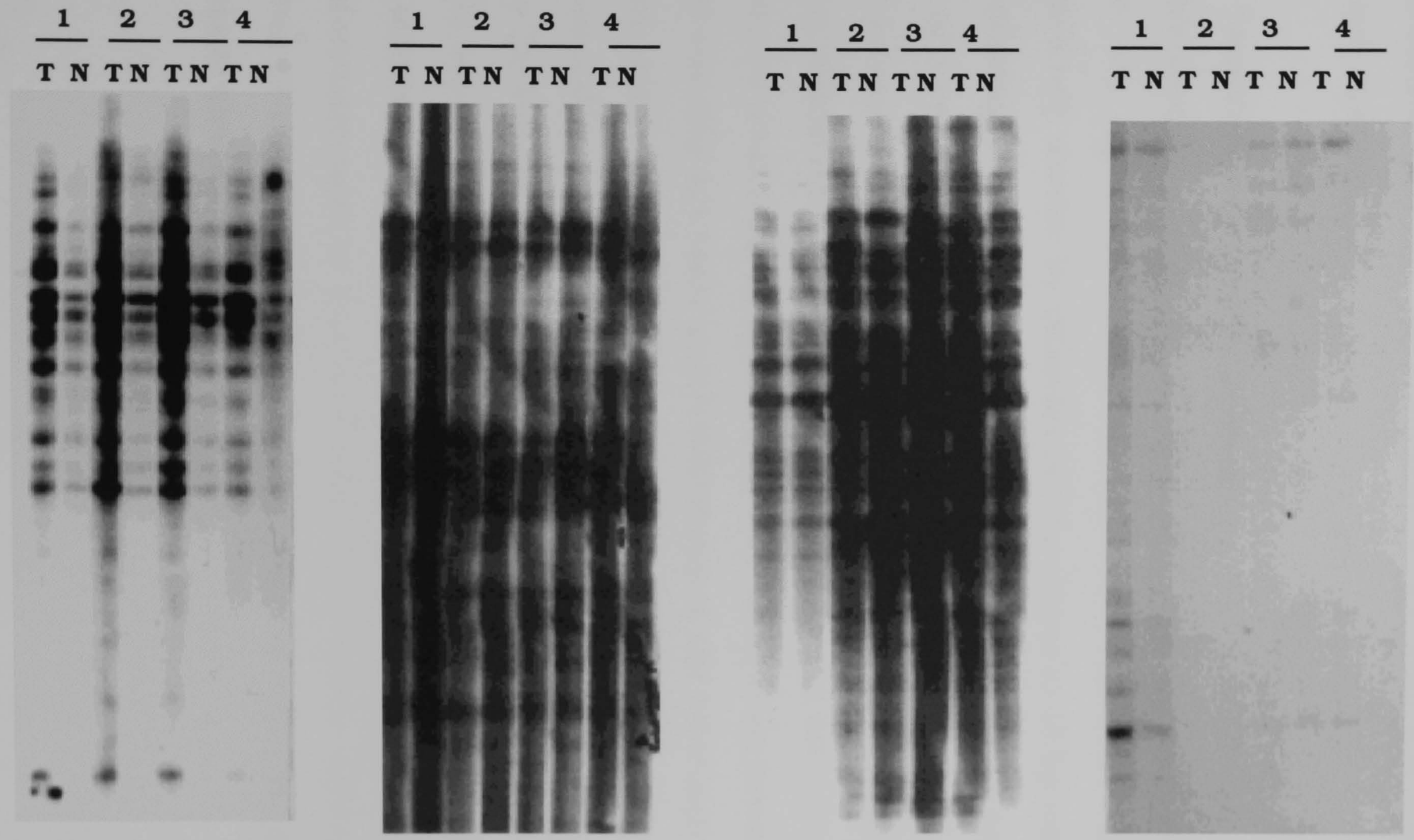


Figure 6: Autoradiograph illustrating the fingerprint profiles achieved with the Alu-specific primers (a) Cotter IV, (b) Alu V, (c) BK-33 and (d) TC65 on four paired tumour & normal cases digested with the Alu I restriction enzyme.

Primarily we were interested to see how viable this technique would prove to be in identifying molecular lesions associated with pancreas cancer. With this objective, work concentrated on identifying fingerprint abnormalities, isolating and characterising them to differing degrees until we were satisfied we had three or four clear cancer-associated DNA rearrangements. Experiments carried out later to investigate the effects of primer length and Alu-consensus position are described in the following sections of the text to maintain continuity.

3.1.3 Length of Alu-specific primers

The Alu-specific primer PDJ-33 which is 35 bp long and specific for the 5′ end of the Alu repeat consensus (47-13) was examined. Primers which were progressively shorter by intervals of 5 bp down to 15 bp were designed (Table 1) truncating the primer from the 3′ end. Alu-PCR was carried out on three cases of tumour and normal DNA, labelling internally with ^{32}P α-dATP to avoid kinase end-labelling five separate primers. The results demonstrated that amplification with the full-sized primer was not very informative with the enzyme used (*Alu I*), resulting in only three to six bands. The reduction of primer size did not make much difference to the informativeness of the band pattern, until one reached the 20-mer and subsequently 15-mer primers. We then see a dramatic increase in the number of bands from 3-6 to 12-15 bands (figure 7a) We carried out the same experiment of truncation at the 3′ end of the Alu repeat element with the Cotter IV primer, but this primer appears more informative in its standard form and no discernible improvement on the informativeness of the band pattern is evident. Table 1 summarises the sequence of the Alu-specific primers employed and their location within the Alu repeat element.

Table 1: Alu-specific primers

Primer	sequence (5'- 3')	Length	Alu-homol
1. BK-33	5'- CTGGGATTAC AGGCGTGAGC C -3'	21	34-13
2. COTTER IV	5'- CAGAATTCGC GACAGAGCGA GACTCCGTCT C -3' (*ECO RI*)	31	263-282
3. Alu V	5'- GTGACCACCG CGCCCAGGCC -3'	20	20-1
4. TC65	5'- AAGTCGCGGC CGCTTGCAGT GAGCCGAGAT -3'	30	220-236
5. Alu 1	5'- GAATTCACGC CACTGCACTC CAGCCTG -3'	27	232-259
6. PDJ33	5'- GCCTCCCAAA GTCCTGGGAT TACAGGCGTG AGCCA	35	47-13
7. PDJ33/25	5'- GCCTCCCAAA GTCCTGGGAT TACAG	25	47-23
8. PDJ33/20	5'- GCCTCCCAAA GTCCTGGGAT	20	47-27
9. PDJ33/15	5'- GCCTCCCAAA GTCCT	15	47-32
10. COTT-IV/25	5'- GCAACAGAGC GAGACTCCGT CTCAA	25	259-284
11. COTT-IV/20	5'- GCAACAGAGC GAGACTCCGT	20	259-279
12. COTT-IV/15	5'- GCAACAGAGC GAGAC	15	259-274

Figure 7: Investigating some of the parameters involved with Alu-PCR fingerprinting

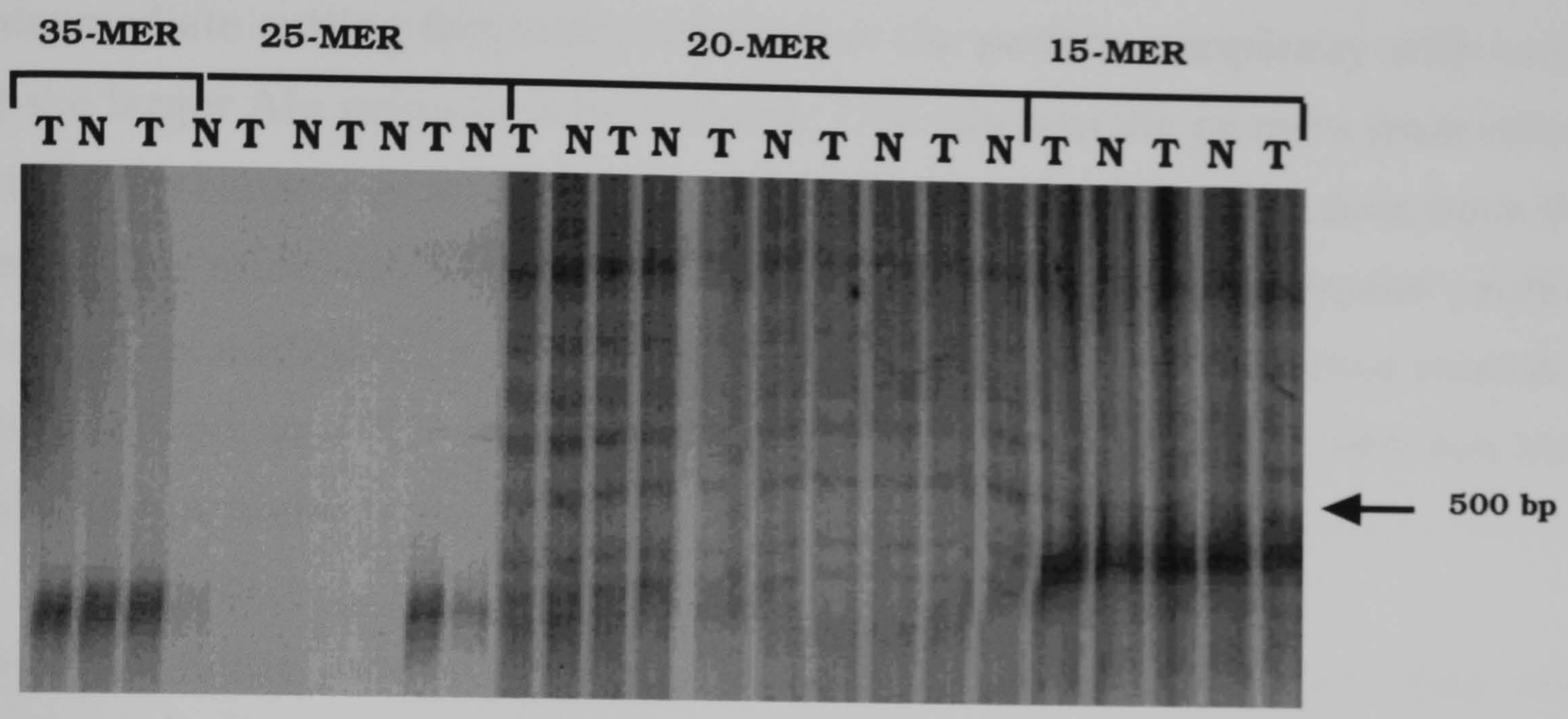


Figure 7a: Alu-PCR carried out on the same cases with progressively shorter primers specific for the PDJ 33 Alu consensus sequence

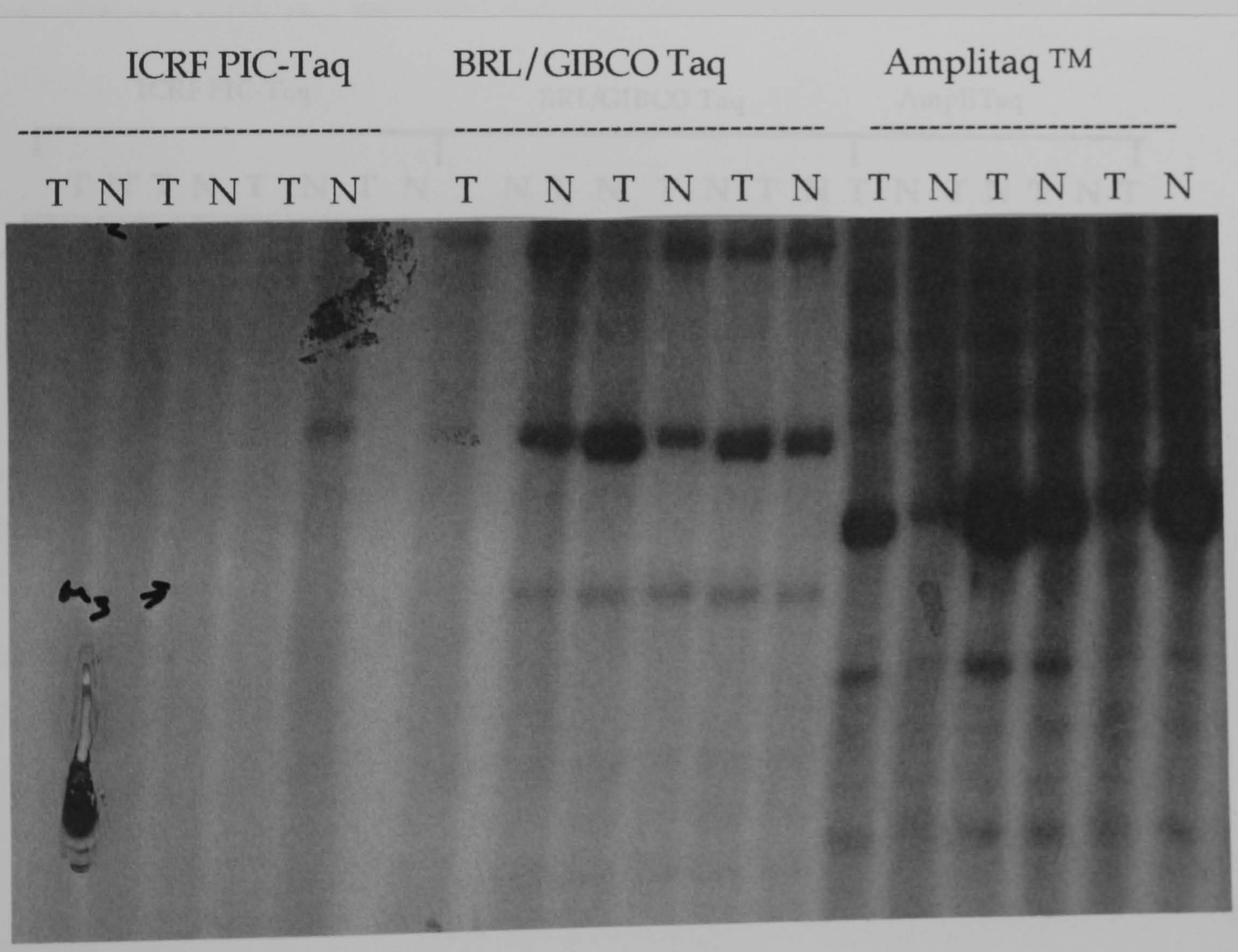


Figure 7b: Alu-PCR carried out on the same cases with the same PCR conditions, but with three seperate sources of Taq polymerase: ICRF PIC-Taq, BRL/GIBCO Taq and AmpliTaq from Perkin-Elmer.

3.1.4 Restriction enzyme frequency and double digests

Another variable which could be employed to increase the informativeness of a band pattern was the use of restriction enzymes which had an intermediate cutting frequency to improve the pattern complexity with some of the larger Alu primers. Alternatively, two Alu-specific primers from either end of the consensus sequence could be included in a PCR. This does have the immediate effect of increasing the number of bands in the fingerprint profile. Analysis of additional fragments produced to ascertain whether they were as a result of one or other primer in the pair, or hybrid fragments between two different primers were not carried out.

The possibility of predigesting the sample DNA with more than one restriction enzyme was quickly eliminated as a helpful strategy. For this, the same 2 tumour/normal pairs were digested with (i) *EcoRI+Pst I*, (ii) *EcoRI* only and (iii) *Pst I* only, then amplified the samples under exactly the same conditions with the BK-33 primer. *EcoRI* and *Pst I* were chosen as they have similar buffer requirements, therefore both enzymes should be anticipated to cut equally well. One would expect to see a similar banding pattern achieved for the two sets of tumour and normal DNAs predigested with *EcoRI + Pst I*, a similar similarity in pattern between the four samples of the next two pairs predigested with *EcoRI* alone and for *Pst I*. In contrast however, there seemed to be no reproducible pattern between any of the three sets of four samples. In fact, differences between the tumour and normal samples from the same case were observed (data not shown). Clearly, despite the buffer for both enzymes being the same, slight differences resulting in inhibition or only partial cleavage from sample to sample were evident. Alternatively, partial digests due to one DNA sample having slightly different salt concentration, or residues from the extraction process could produce differences in the band pattern compromising the whole strategy. *EcoRI* is an enzyme that is particularly prone to STAR activity also, again compromising the usefulness of the procedure.

Another possibility could be mechanical fragmentation of the DNA into smaller pieces to reduce the number of potential templates. This was not considered as a feasible option since there would be no control over the position of the mechanical breakage. Thus two samples from the same individual (perhaps tumour and normal DNA) would not be predicted to produce the same band profile. Later in the course of these studies however, a collaborating worker achieved reproducible fingerprint profiles from DNA extracted from archival paraffin embedded tissue which was known to be highly fragmented due to the fixation and embedding processes. This suggests that sufficient intact DNA template molecules survive the tissue fixation to offer satisfactory primer binding sites. Clearly this will greatly extend the potential applications of this approach.

3.1.5 Magnesium ion concentration and sources of Taq polymerase

Magnesium ion concentration [Mg^{2+}] in the Alu-PCR reaction was also investigated as to whether it has an effect on band pattern complexity. Higher magnesium concentrations often have the effect of lowering the specificity of binding of the primers, therefore we predicted that the higher [Mg^{2+}] the more complex and perhaps random the band pattern. In fact what we observed was quite different in that from 1.0 mM Mg^{2+} to 2.5 mM Mg^{2+} we see little change in overall complexity and number of bands achieved with the primer BK-33 (30-40 bands from 200-1,000 bp), though at 2.5 mM we see a slightly more pronounced background smear. At 5.0 mM Mg^{2+} we see a bias towards the upper 700-1,000 bp products over the smaller band sizes (data not shown).
Slightly cleaner, more informative band profiles were produced with potassium (K^+) buffers compared to ammonium buffers (NH_4^+) which have been adopted in preference by Taq polymerase suppliers.

What did prove to be an important consideration was the source of Taq polymerase enzyme used. Initial experiments were carried out with BioTaq™ enzyme from Bioline laboratories, however during the course of these early experiments an "in-house" preparation of Taq was made available to ICRF

laboratories which proved unsuccessful. With the ICRF "PIC-Taq" very few bands were produced using the most informative enzyme/primer pairing. A second commercial preparation of Taq polymerase purchased from Life Technologies (BRL/GIBCO) proved just as poor, with Perkin Elmer's "Amplitaq" giving just as good results as the Biotaq™ enzyme but at significantly greater cost, see figure 7b. For this experiment, all the enzymes were used with the Bioline buffer preparation which has proved to be the most successful with the in-house PIC-Taq enzyme in conventional PCR.

3.1.6 Method used to label/visualise bands

The initial label of choice when carrying out the Alu-PCR fingerprinting was ^{32}P since it represents a fast, cheap and efficient way to visualise all bands on an acrylamide gel - even ones which amplify less efficiently. We found that the use of ^{35}S in particular resulted in higher background smearing with less definition of some of the fainter amplification products which are often the ones exhibiting variation. The introduction of ^{33}P into the laboratory has been a big success for many applications and it is now the isotope of choice for carrying out AP-PCR. Although ^{33}P was an improvement on ^{35}S , the need to dry gels and more extended periods of exposure to visualise some of these faint PCR products was a disadvantage. Time was also spent investigating the possibility of applying silver staining methodologies to the Alu-PCR fingerprinting gels. This proved the least successful with only the main products being visualised after some 20-30 mins with the background competing out any possibility of visualising even the intermediate-intensity bands. An added obstacle to this type of signal development is the requirement to handle the gel more in the silver staining procedures. To maximize the extent of the fingerprint resolved and to examine six to eight tumour samples simultaneously under two or three conditions, a large sequencing gel rig was used (Bio-Rad Sequigen II) which made handling difficult and increased the background caused by mechanical blemishes on the gel surface.

3.1.7 Reproducibilty of the fingerprint patterns

The reproducibility of the fingerprint patterns across many individuals was good within any experiment. When it came to trying to reproduce a particular fingerprint band patterns in subsequent experiments, the exact pattern could often not be repeated. Such is the case with many of these PCR-based techniques, but probably more so with this type of approach where amplification of products is dependent on the random annealing of consensus primers to a heterogenous target. Analysing some of the Alu-PCR products has revealed all manner of degenerate Alu sequences, as well as intermediate repetitive sequences (MIRs) and Line L1 sequences (discussed in chapter 4). Subtle differences in nucleic acid composition, primer annealing conditions or reaction constituents will always introduce a variation to each experiment.

3.2 Alu-PCR on paired tumour and normal samples

In order to identify sequences which were altered in DNA from tumour tissue relative to DNA from constitutive normal tissue sources, twelve pairs of samples were taken from pancreatic cancer cases we had in the laboratory. The eight cases had all been histologically proven to represent cases of ductal adenocarcinoma of the pancreas and were being used in the laboratory for allelotype experiments as well as looking for mutations in other cancer-associated genes such as *APC*, *p16/INK4A* and *Smad4/DPC4*. The tumour samples had been examined under the microscope and areas representing greater than 50% tumour tissue had been delineated on hematoxylin and eosin stained slides then cut from the mounted tissue block. These samples had previously been concentration-matched by their optical density readings and by judging the relative intensities of the smears produced when digesting them for Southern hybridisation.

Altogether, fifteen PCR fragments were cut from the acrylamide gels and re-amplified with the same primer which generated them as described in the

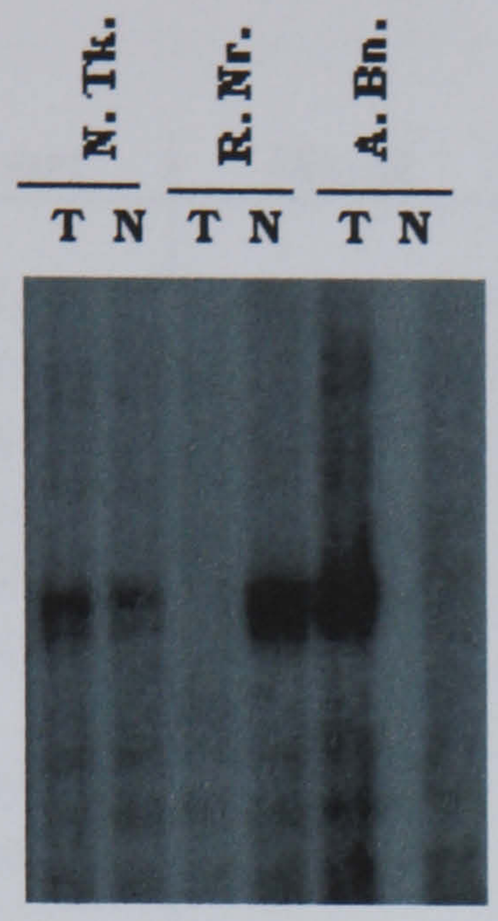
Materials and Methods. Two fragments failed to re-amplify after repeated attempts (AB6 and Met-7) and since the acrylamide gels were not being fixed or dried down at this stage, this could not be attributed to PCR inhibition due to fixative. Each one had demonstrated differences between tumour and normal tissue DNA from the same case(s). Table 2 summarises the fragments isolated with the original observation and conditions used. Nearly all of the fragments isolated were examples of bands which had been lost in the tumour lane in comparison with its constituent normal lane. One fragment (AB2) was first observed as an apparent band mobility shift, however on closer inspection the observation of faint bands co-migrating with the two stronger bands may indicate a preferential amplification of alternative alleles. Alu-PCR has been described as a method to identify the replicative error phenotype (RER) in colorectal cancer (Krajinovic *et al.*, 1996). Such replicative errors are identified as band shift in Alu-PCR profiles, however this laboratory and others have found the RER phenotype not to be a frequent contributory factor in adenocarcinoma of the pancreas. This may explain why there is a scarcity of band shift anomalies seen in this analysis carried out on pancreas tumour/normal pairs. Differences in the intensity of bands between tumour and normal were also observed in only two or three instances, however it was felt we should concentrate on definite loss of fragments to begin with so to maximize our chances of identifying possible loss of tumour suppressor genes. Three examples of band rearrangements observed in tumour samples compared with their constituent normal samples are illustrated in figure 8.

Fragments were labelled and hybridised to Southern blots of tumour/normal pairs to see if the fragment generated would hybridise to a unique site in the genome, or result in a smear of hybridisation suggesting it was merely as a result of polymorphism within the Alu repeat itself which was creating the differential fingerprint pattern. These mainly small PCR fragments often required dual labelling to achieve high enough incorporation of radionuclide to carry out Southern analysis (see Materials & methods). Some of the

fragments isolated gave smears of hybridisation even after competition with total human DNA (Sealey *et al.*, 1985). Four fragments gave only constant bands across each track of Southern blots and two exhibited clear polymorphisms, one of which showed tumour-specific loss of bands.

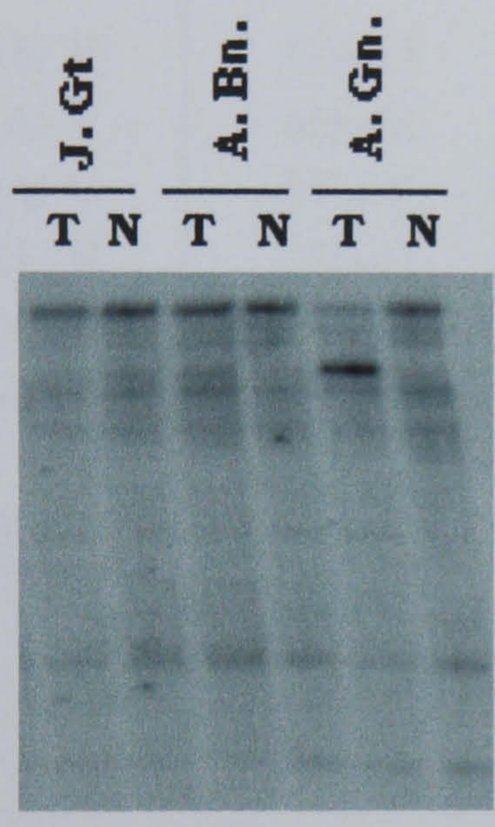
Figure 8: Examples of Alu-PCR fingerprint rearrangements identified in tumour samples/cell lines

(a)



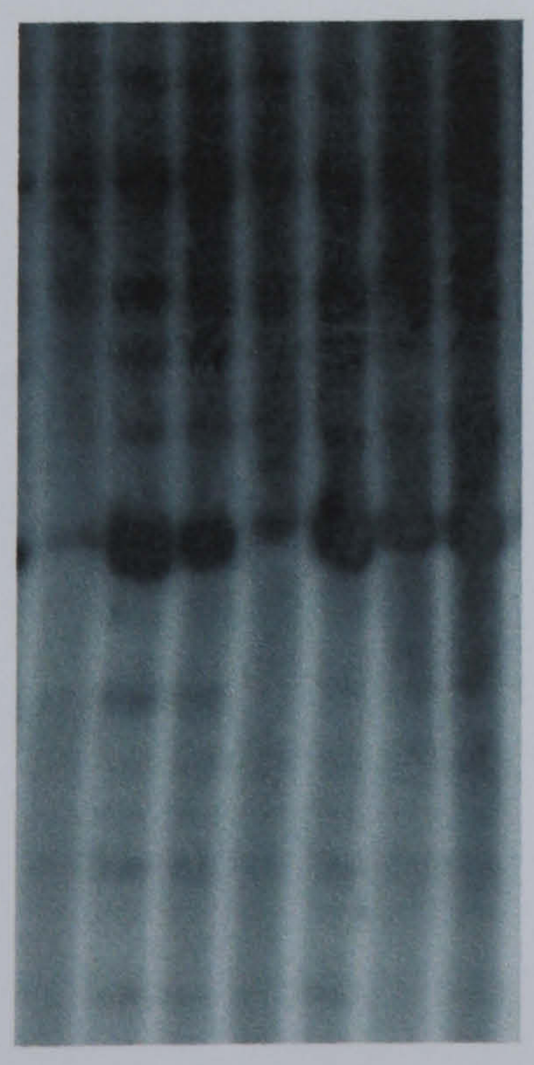
IP-1: A 350 bp fragment lost in two separate tumour samples compared with their constitiuent normal samples (only one shown here).

(a)



AB4: A 700 bp band lost in one tumour lane compared with its constituent normal lane (A.Gn). The appearance of an additional strongly amplified band at approx. 650 bp in the same lane may suggest this Alumorph is in the border of a deletion

(b)



Met-U: A 295 bp band lost in the SUIT-2 parent cell lines and S2-020 sub-line isolated by T. Iwamura

Met-L: A 235 bp band lost in the SUIT-2 parent cell lines as well as S2-020, S2-CP9 and S2-VP10 sub-lineS isolated by T. Iwamura

(c)

Table 2: Alu-PCR fragments isolated with initial observation

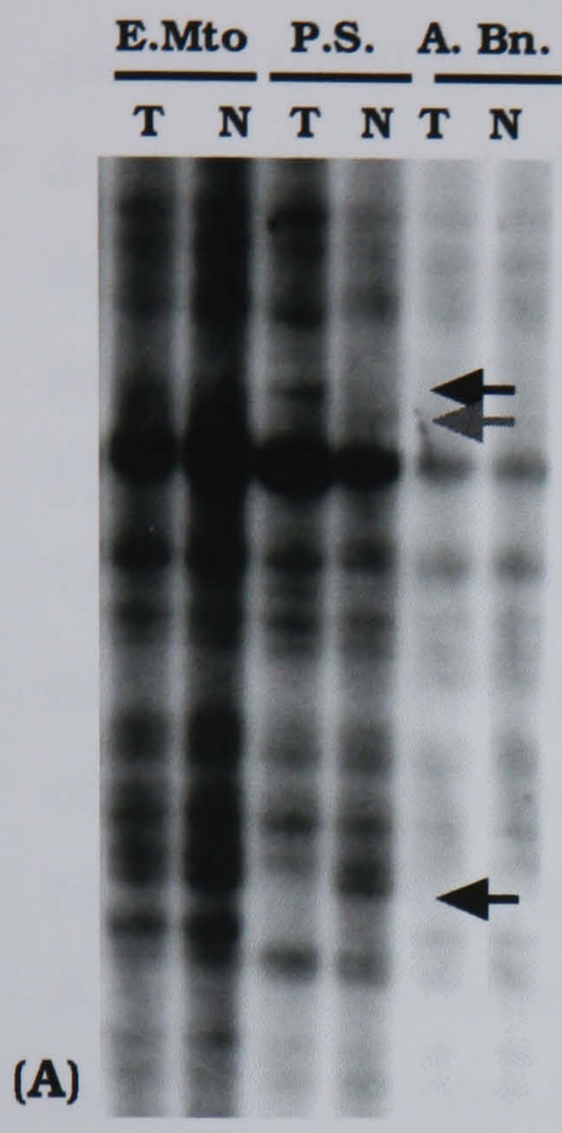
Alumorph	Primer	Digest	Case(s)	Observation	Size (bp)
AB1	BK-33	Alu I	P.S.	Loss of band - 1 case	270
AB2	BK-33	Alu I	P.S.	Loss/shift of band - 1 case	410
AB3	BK-33	Alu I	S.Atk.	Loss of band - 1 case	290
AB4	BK-33	Alu I	R. Tnr.	Loss of band - 1 case	700
AB5	BK-33	Alu I	D.Bir.	additional band - 1 case	370
AB6	BK-33	Alu I	R.Tnr.	Loss of band - 1 case	275
IP-1	Cotter IV	Alu I	R.Nr.	Loss of band - 2 cases	350
IP-2	Cotter IV	Alu I	?	Loss of band - 1 case	280
Met-U	ALU V	Hinf I	S2-lines	Loss of band - 2 cell lines	294
Met-L	ALU V	Hinf I	S2-lines	Loss of band - 4 cell lines	235
Met-2	ALU V	Hinf I	S2-lines		285
Met-3	ALU V	Hinf I	S2-lines		190
Met-6	Cotter IV	Alu I	A.Gn.	Loss/shift of band - 1 case	700
Met-7	Cotter IV	Alu I	A.Gn.	Gain/shift of band - 1 case	650
V1	BK-34	Alu I	10409	Loss of band - 1 case	225

3.3 Characterisation of the Alumorph AB1 and its associated transcript up-regulated in pancreatic adenocarcinoma

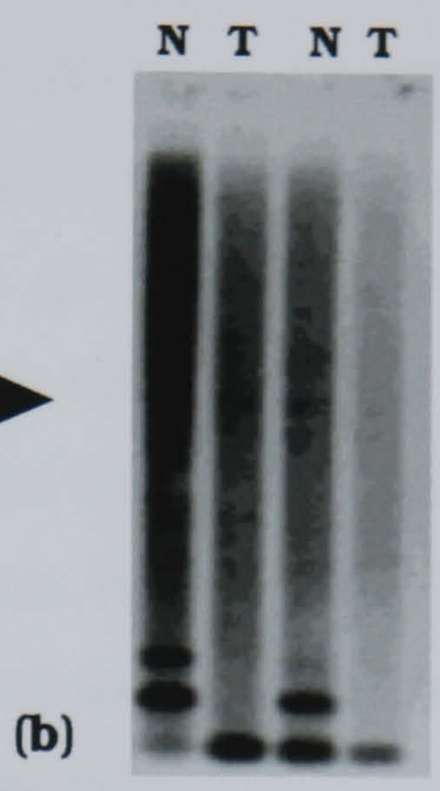
3.3.1 Southern Hybridisation

A polymorphism observed on one of the first Alu-PCR gels (with BK33 on *Alu I* digested DNA) as a band of around 270 bp which was missing in the tumour DNA of sample 2 (figure 9a) gave an interesting Southern blot result. On hybridisation to a filter containing nine tumour/normal pairs digested with *Msp I*, three small bands were observed to hybridise with the probe on the autoradiograph. One very small band (approx. 70 bp) was observed to be constant over the nine cases, one slightly larger band (approx. 100 bp) was present in four of the cases, and a larger band again, (200 bp) was present in only one of the pairs (exhibiting all three bands) (figure 9b). Interestingly, the middle sized band was missing from the tumour track in three cases, one of which had the larger band which was also missing. In summary, this 270 bp fragment (now called Alumorph AB1) was isolated as a band present in the normal but not the tumour track of a paired case (P.S.) and exhibited an MspI polymorphism on Southern hybridisation with alleles missing in two further tumour samples when compared with their normal partner sample (J.Gt. & M.Ch.). Clearly these two larger alleles are less frequent but identify a genetic imbalance between the tumour and normal material with loss of all or part of a chromosome arm. A second Alu-PCR fragment which was isolated as a mobility shift between tumour and normal lanes of the same case as AB1 was larger and thus proved easier to amplify and label. Upon hybridisation with a *Hind III* Southern blot this band produced an interesting pattern of bands with two constant and two polymorphic bands however, no losses were observed between tumour and normal samples.

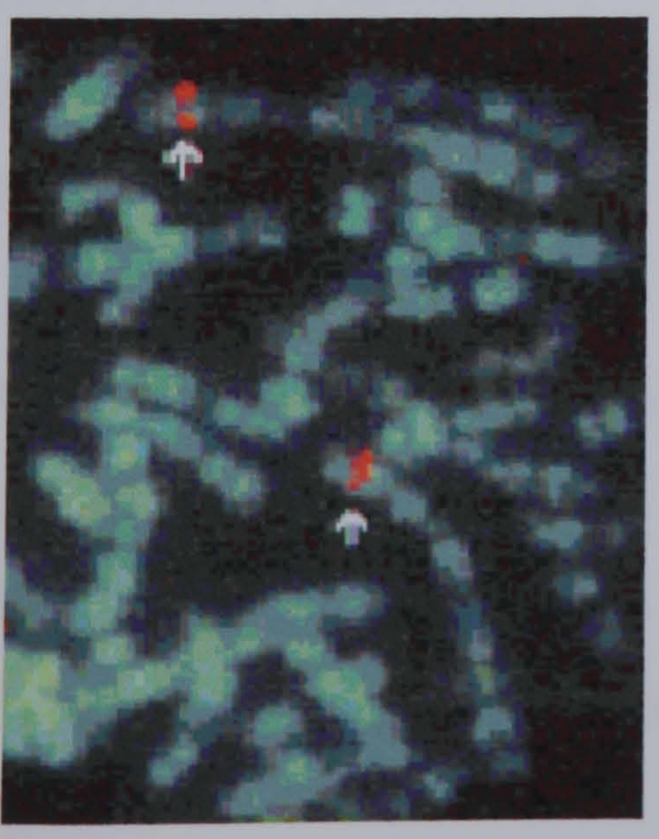
Figure 9: Schematic representing each stage in the overall strategy to isolate tumour-specific genes by Alu-PCR using the Alumorph AB1 as an example



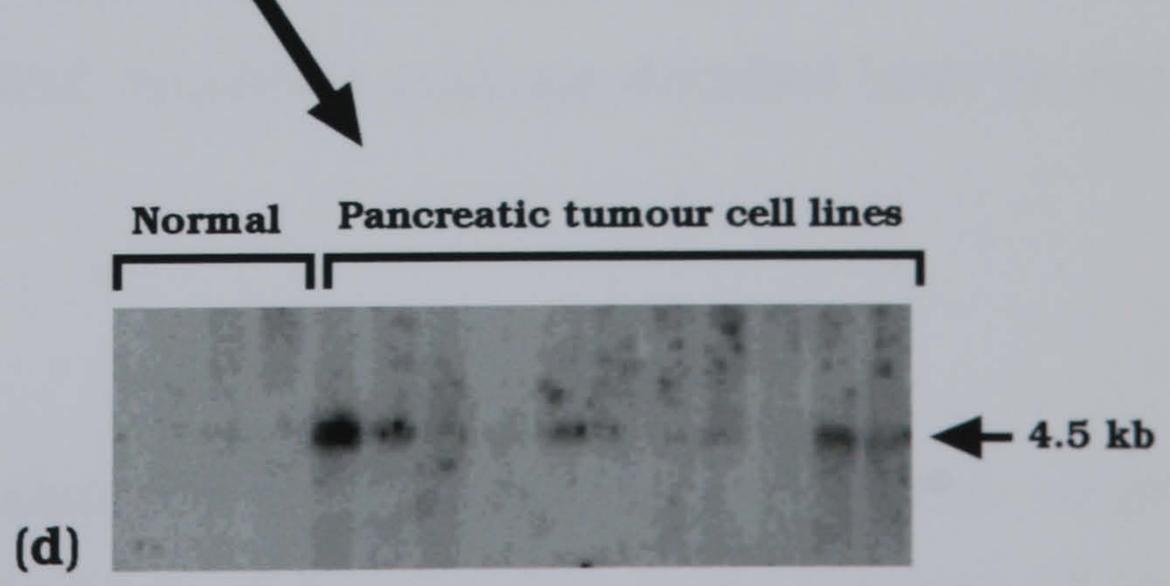
Southern hybridisation



Northern hybridisation



Large genomic clones used in FISH localisation position AB1 on chromosome 2p22



Expressed sequences identified from the genomic clones with cDNAs isolated exhibit a tumour-specific expression

3.3.2 Sequence analysis

Ten of the sixteen excised Alu-generated fragments showing some variation termed "Alumorphs," were cloned into Bluescript™ by the T-A cloning method (Holton & Graham, 1990). Sequencing was initially carried out on the double-stranded plasmid templates with flanking T7 and T3 primers using the standard Sequenase™ kit. The complete sequence for the 270 bp insert was easily obtained for the AB1 clone in both directions on sequencing with T7 & T3 primers. Five tandemly repeated copies of a tetranucleotide sequence "A-G-T-T" were observed in the centre of the 270 bp sequence. It was originally postulated that expansion of this simple repeat may be the source of the polymorphic bands observed on the Southern, but later experiments ruled out this explanation (see discussion later). The fragment was 50% AT/GC, with stretches of thymidine bases as well as stretches of guanine bases indicative of the Alu repeat element (appendix 2a). Three *MspI* restriction enzyme recognition site could be identified within the sequence which would result in fragment sizes of 24 bp, 64 bp, 72 bp and two flanking fragments of undetermined size. The 3' end fragment did appear to be GC rich and accordingly appeared to have more *Msp I* sites, therefore its likely, the 5' flanking end was the hybridising 200 bp upper band on the Southern. This 270 bp sequence (like the ten of the other fragments sequenced) was rather short for a meaningful search of the databases of cloned sequences, but we did pick up quite a strong similarity (70%) between Alumorph AB1 and a relatively low copy repeat sequence (Sp-0.3-23) which demonstrates germ cell-specific hypomethylation as well as hypermethylation in placental tissue and embryonal carcinoma cells (Appendix 2a). This is an observation which prompted further investigations and sequence analysis detailed later in this thesis.

3.3.3 Isolation of genomic clones

In an attempt to gain more sequence information around these small fragments isolated from Alu-fingerprints and to localise them through FISH, they were used as probes on gridded genomic libraries acquired from the

Reference Library Database (RLDB) (Zehetner & Lehrach, 1994). We initially acquired a human genomic P1-bacteriophage gridded library (Francis *et al.*, 1994) and later, a cosmid library (not gridded) made by Dr. Anna-Maria Frischauf at ICRF. With the development of the MRC human genome mapping group (MRC-HGMP) resources at Hinxton in Cambridge, many more resources were freely available to registered members. A human genomic PAC library made by P. de-Jong is distributed on gridded filter arrays (Iounnou *et al.*, 1994) as well as PCR pools. Mega-YAC (YACs with inserts of >1Mb) pools from CEPH and the ICRF. Additionally, the facility to obtain individual YAC clones from the CEPH library which one could identify with DNA markers from a databases such as the Whitehead Genome Database. Gridded library filters are suitable for hybridisations with a wide range of different probes and with 20-36,000 individual clones spotted robotically on a 20x20 cm filter represent a very powerful resource to any gene mapping/hunting project. The MRC-HGMP and RLDB also distribute chromosome-enriched libraries. The HGMP distribute the monochromosomal libraries created at the Lawrence Livermore National Laboratory (LLNL) and the RLDB distribute libraries they have created themselves, some at ICRF.

Initially, the 270 bp AB1 fragment was used as a probe to hybridise against the P1 bacteriophage library after employing the recommended conditions for optimal competition. The resulting autoradiograph had a high background level of signal to all the clones on the grid, even after stringent post-hybridisation washes down to 0.1 x SSC, 0.1% SDS at 65^{0}C. This could of course be attributed to the presence of Alu or other repeat motifs within our small probe. Twelve P1 clones were selected from the autoradiograph, picked by the RLDB laboratory and supplied as stabs on agar slopes with kanamycin. These stabs represented pure cultures, corresponding to the clone number which was spotted on that grid-reference on the filter, removing the need for secondary or tertiary screening.

3.3.4 Analysis of the P1 clones

To confirm the positive hybridisation on the gridded filters, *Bgl II* restriction digests were carried out on the P1 clones followed by electrophoresis overnight on a 0.7% agarose gel. The P1 bacteriophage exists as single copy autonomous prophage in the *E. Coli* cell, thus the amount of DNA isolated from ten millitres of culture is minimal. To see the products of any restriction digest analysis, half the amount of isolated DNA from a 10 ml culture was loaded onto a 0.7% agarose gel and run slowly overnight. The *Bgl II* enzyme cuts the vector sequence twice resulting in the generation of a 6 kb band and so any subsequent fragments have been generated through cuts in the insert DNA. The fragments displayed on the submarine agarose gel were numerous, ranging from 5-16 for each clone (gel photograph; figure 10a). Two of the P1 clones were estimated to be around 80-90 kb in total, but most were 50-60 kb (table-figure 10b). Analysis of the sizes of some of the cosmids and PACs was made by PFGE at a later date (gel photograph; figure 10c & d).

After blotting the gel and hybridizing with the original 270 bp fragment, five of the twelve P1 clones hybridised to the fragment. There was no common hybridising pattern to the eight positive clones however, and some of the large hybridising fragments could not be contiguous with such a small DNA probe. It is therefore probable that Alu cross hybridisation was accounting for a lot of the positivity, despite the labelled fragment being competed with Cot DNA prior to hybridisation with the filter. The stability of the P1 clones had been drawn into question, in that repeated propagation of these clones could result in deletion of large fragments of the insert DNA. Also as the P1 vector molecule exists as a single autonomous circle in the *E. coli* cell, the DNA yield is always poor. With these limitations in mind, an attempt was made to fingerprint the eight potentially positive P1 clones and demonstrate whether there was any overlap/continuity between the clones. PCR was carried out with three Alu-specific primers in conjunction with either T7 or Sp6 primers homologous to the T7 and Sp6 promoters flanking the polylinker of the P1 vector.

Figure 10: Agarose gel electrophoresis of digested genomic clones with estimation of their size

(a)

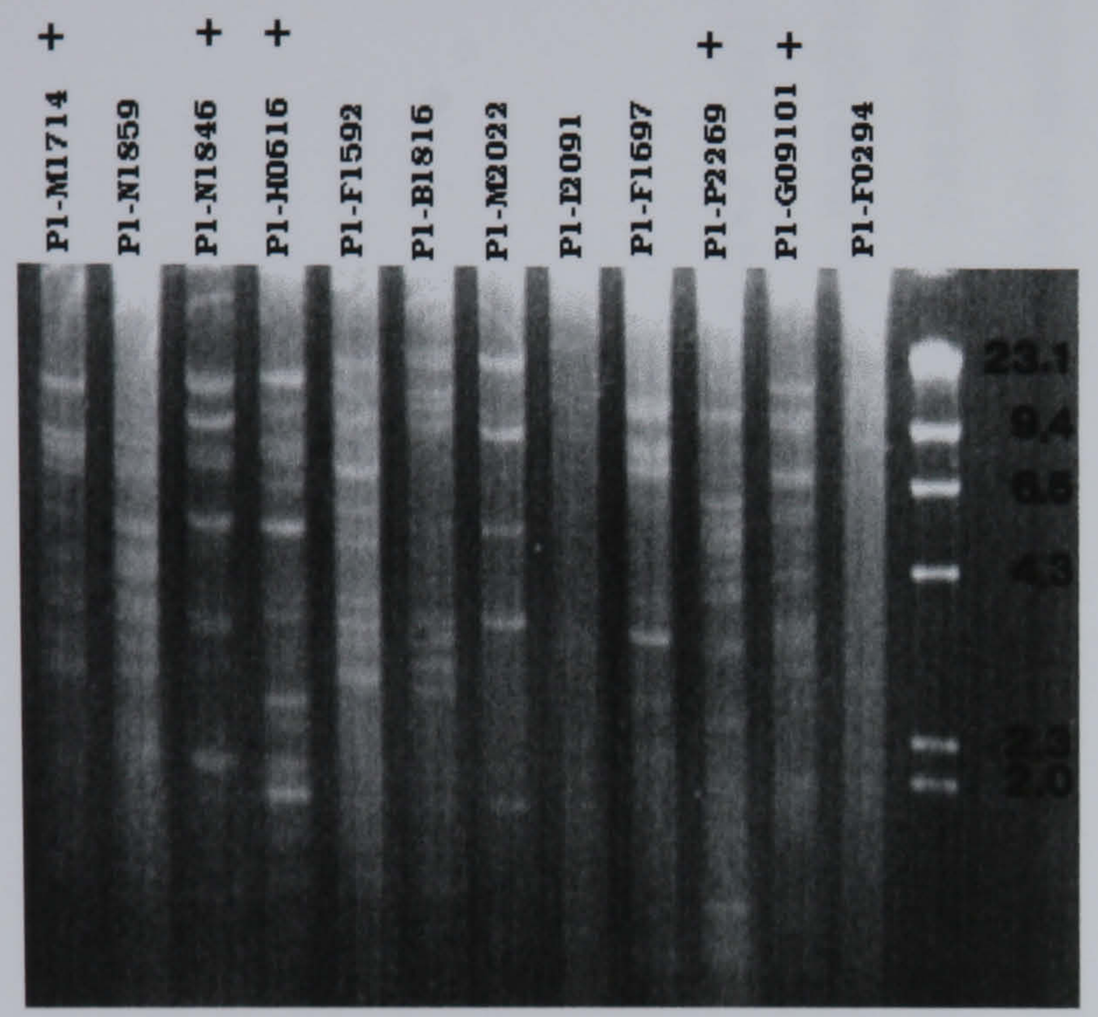


Figure 10 (a): Regular 0.7% agarose gel electrophoresis of 12 isolated P1 bacteriophage clones from gridded library after digestion with Bgl II. Clones specific for the AB1 fragment are identified by (+) after the gel was blotted and hybridised.
(b): The approximate size estimation of the AB1-positive P1 clones are detailed in the table below.

(b)

P1-M1714	P1-N1846	P1-HO616	P1-P2269	P1-G09101
13.0	14.0	15.0	9.7	17.0
8.5	9.0	10.0	6.0	9.7
8.0		8.3	5.7	9.4
7.5	7.0	7.5	5.5	7.0
6.5	6.5	6.3	5.3	6.0
5.5	5.0	5.2	5.0	5.7
4.4			4.3	5.2
4.0		3.6	4.0	4.3
3.9	3.7		3.9	3.8
3.5	3.5		3.5	3.5
2.8		2.7	2.5	3.2
2.7		2.4	2.2	
2.1	2.0	2.0	2.1	2.0
		1.8	2.0	1.9
		1.6	1.9	1.7
	1.0	1.4	1.7	
			1.5	
68.4	51.7	67.8	66.9	80.4

(c)

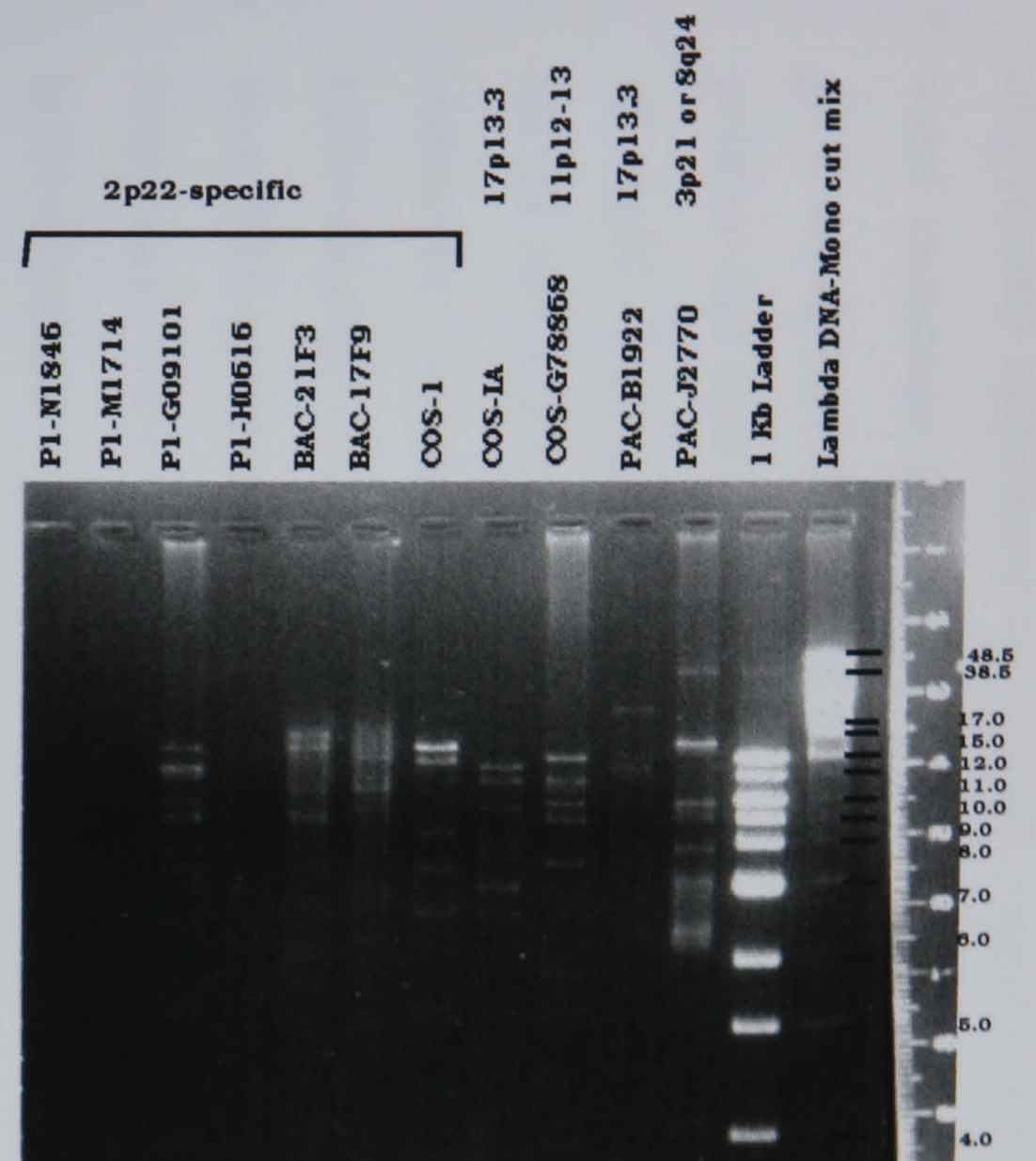


Figure 10 (c): Pulse-Field-Gel Electrophoresis of 2p22 and 17p13-specific cosmids, P1 bacteriophage clones, (only one with enough DNA to stain) PACs and two 2p22-specific BACs after digestion with EcoRI restriction enzyme
(d): A more accurate table of band sizes (kb) of the cosmid and PAC/BAC fragments

(d)

G09101	21F3	17F9	COS-1	COS-IA	G788	B1922	J2770
12.4	14.4	12.5	13.0			24.5	38.5
11.5	13.0	12.0	12.0	12.0		12.5	14.0
11.0	12.0	11.3		10.5	11.5	11.5	10.0
10.5	10.5	10.5		10.0	10.5	11.0	
9.7	9.0	9.0		9.5		10.0	9.5
9.3	8.8	8.0	8.2	7.5	8.0	9.0	8.0
9.0	7.8	7.8	7.3		7.7	7.5	
7.5	7.2	6.9	6.7		7.5		7.0
7.3	6.1	6.8			7.0		6.8
7.0	5.8	6.7			6.5		6.4
6.8	5.5	6.3		5.7			6.2
6.5	5.2	5.8		5.0		5.7	5.7
5.3		5.0	5.3			5.0	5.0
5.0	4.0	4.7					
4.0	3.8	4.5			4.2		
126.5	113.5	111.5	52.5	68.0	62.8	96.7	114.1

Four sets of PCR were set up for each P1 clone combining the Alu-specific primers TC65 with T7 and Sp6 as well as Alu-A1 with T7 and Sp6 (figure 11). Interestingly, these Alu-PCR results did demonstrate that each combination of primers appeared to exhibit regular band patterns, or at least bands which were common to the eight P1 clones. The A1/Sp6 combination appeared to give an almost identical band pattern across the eight clones. A1 with the T7 primer did also show one band at approximately 1.2 kb which was common to all eight clones, and another at approximately 1.5 kb which was common to six of the eight clones. Thus there was evidence that some DNA sequence similarity existed between the eight P1 clones despite any Alu cross hybridisation. Upon the isolation of five cDNA clones described later in the text, STS primer sets specific for sequences within these cDNAs and the AB1 sequence were used to confirm their presence within the P1 clones. DNA hybridisation was also employed to verify that the cDNA clones in particular lay within the P1 clones (discussed later).

The P1 phage library was used to furnish probes for only two of the Alumorph fragments as the library was soon to be superseded by the gridded PAC library and later PAC PCR pools. The PACs proved much easier to propagate, yielding higher amounts of DNA from smaller volumes of culture and reported to be more resistant to genetic recombination (discussed in Materials and Methods chapter). In effect, they could be treated like cosmids as far as DNA extraction was concerned.

Figure 11: Alu-PCR fingerprinting of putatively positive P1 clones to assess overlap

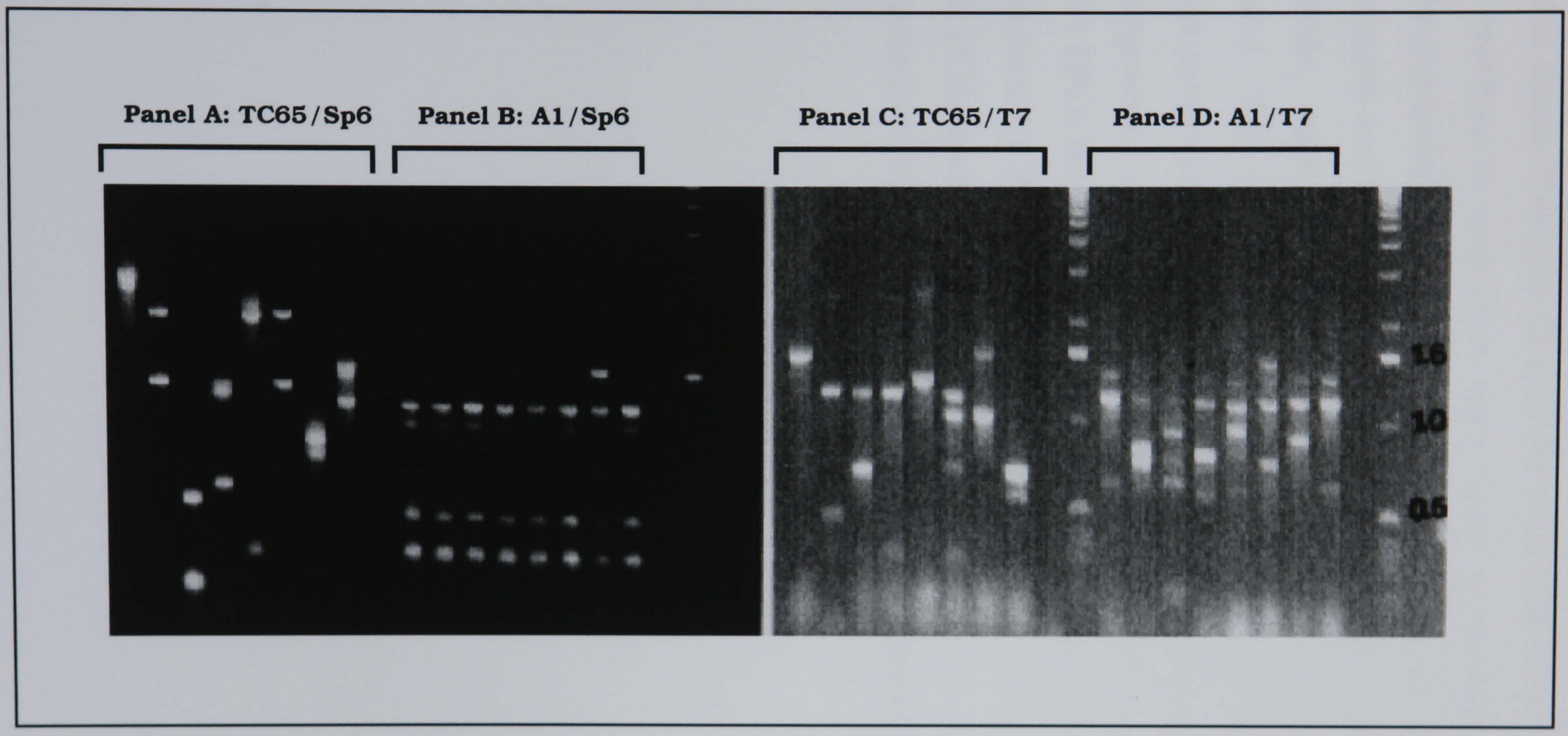


Figure 11: Four panels of PCR products amplified from the same eight positive P1 clones, (M1714, N1846, H0616, F1592, B1816, P2269 and G09101 in that order) with (A) TC65/Sp6, (B) A1/Sp6, (C) TC65/T7 and (D) A1/T7. Each panel also has a water negative control PCR sample loaded 9th from left to right. Although the experiment does not exclude the possibility of inter-Alu PCR occurring, the main objective of trying to assess the degree of similarity between the P1 clones has been achieved. Clearly, the A1 Alu primer demonstrates great similarity in band pattern between the eight P1 clones, particularly with the Sp6 primer which might suggest we are getting specific vector-Alu amplification with some combinations. The PCR carried out with A1 & T7 is obviously not so achievable, and therefore one would suppose a great deal of the bands in this pairing will be inter-Alu, however there is still a degree of overlapping products with the eight clones.

3.3.4 Chromosomal localisation

The large genomic clones specific for the Alumorphs were localised by FISH analysis in collaboration with the ICRF-Cytogenetics Laboratory. In addition two cosmid clones (cosmids IA and IB) were sent to collaborators at the CRC Paterson Institute to map them more finely by interphase-FISH. With some of the Alu-generated fragments for which we had sequence, we were able to design primers and localize them by PCR on a panel of twenty four somatic cell hybrids, each containing one human chromosome on a rodent background. The results obtained from the monochromosomal somatic cell-hybrid panel were often verified by sending a larger genomic clone to the ICRF Cytogenetics Laboratory for FISH analysis. Initial difficulty in preparing enough P1 DNA from the AB1-specific clones to carry out FISH necessitated the localisation by PCR on the monochromosomal somatic cell hybrid panel. From this PCR-based localisation it appeared the AB1 fragment lay on either chromosome 2 or 10, a result we could not refine further (figure 12).

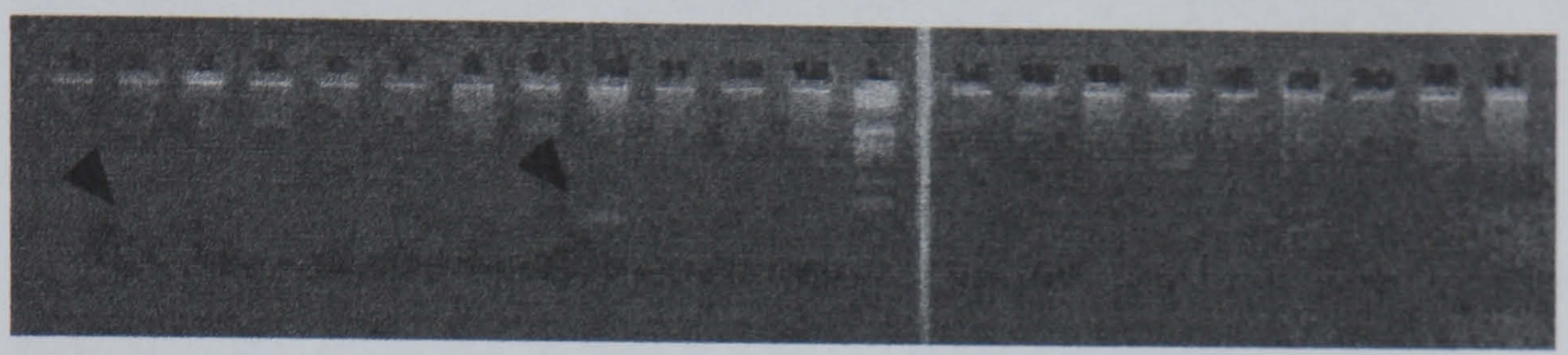


Figure 12: Illustrates the result we obtained carrying out the AB1-specific PCR on the ICRF's mono-chromosome somatic cell hybrid mapping panel. Although we observe a positive band on DNA from the hybrids containing chromosomes 2 and 10 (**Arrows**), later AB1-specific P1s and a YAC were localised conclusively to 2p22 (see figure 9)

Four AB1-specific P1 clones (N1846, M1714, H0616 & G09101) proved difficult to localise in FISH experiments as the signals generated from them were very weak. Again this was attributed to the lack of DNA as the amounts passed onto the Cytogenetics laboratory were the minimum amounts they felt they could carry out effective FISH analysis. Eventually the P1 clone M1714 produced satisfactory signals hybridising to chromosome 2 band p22 in eleven metaphase spreads. A YAC clone specific for the AB1 fragment was subsequently isolated in an effort to verify this weak FISH result, and was also mapped to 2p22 (figure 13). Another of the P1 clones (G09101) had given less convincing signals to 2p22, but also 2q34 in 4/15 spreads,15q12-13 and 15q23 (data not obtained). Interestingly, the hybridisation to 15q may have a direct bearing in relation to gene sequences identified to be contiguous to our P1 clones (see section 3.4.6).

Figure 13 : Localisation of alumorph AB1-specific YAC (1c) to chromosome 2 band p22

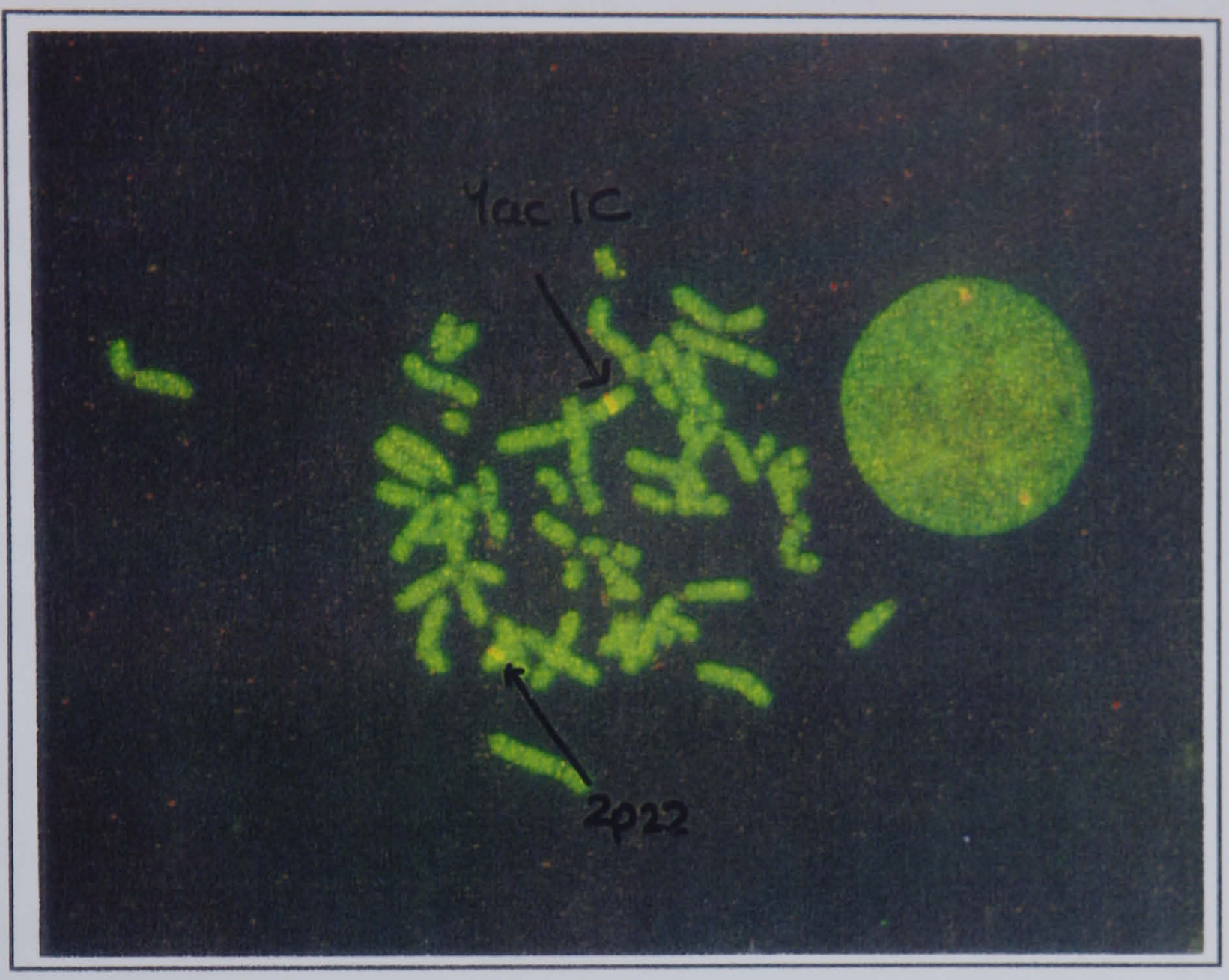
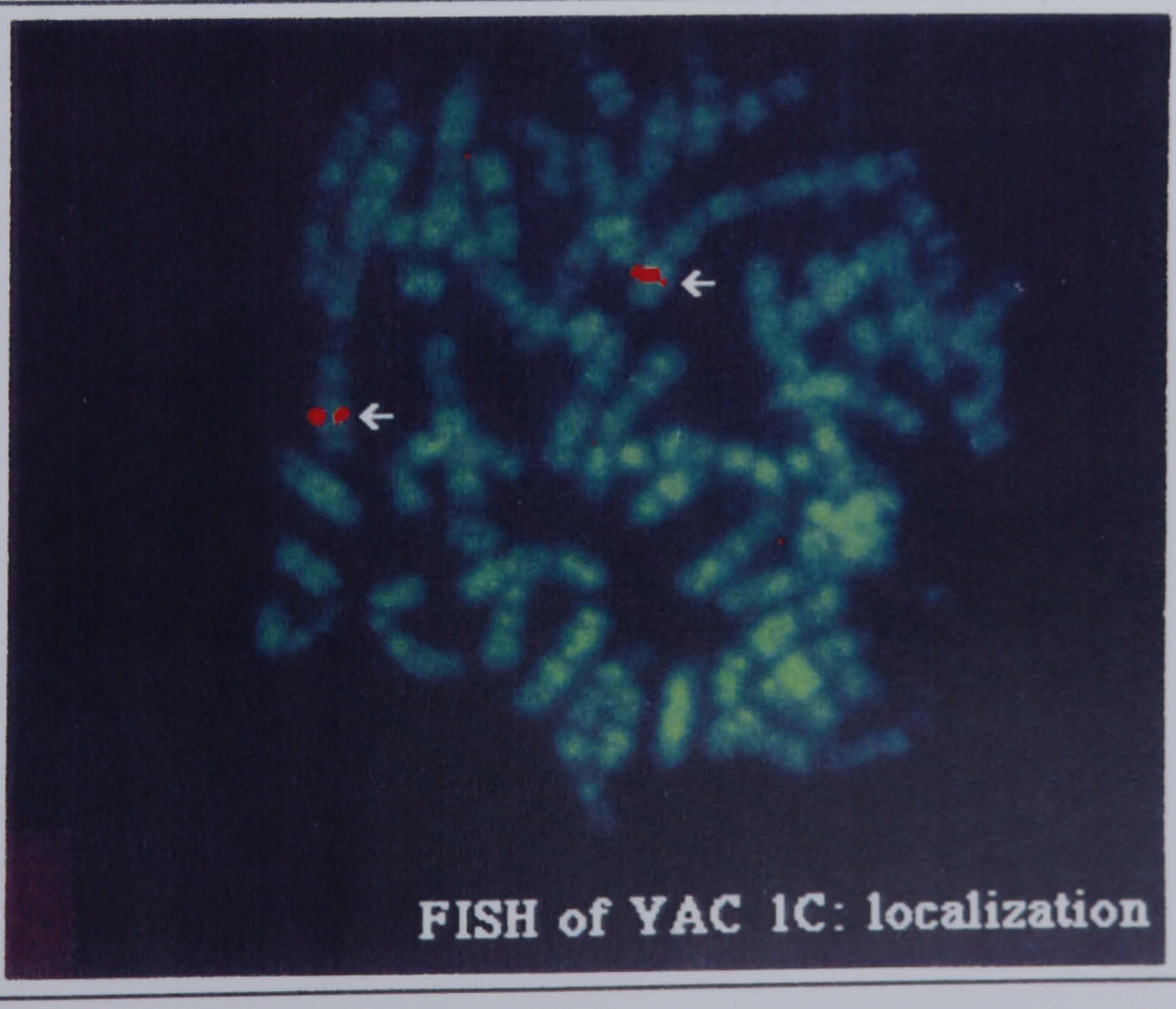


Figure 13a: Hybridisation of AB1-specific YAC 1C to chromosome 2 band p22 on the metaphase spread in the centre of the photograph. Hybridising signals are also evident within the normal interphase nucleus to the right of the chromosome spread

Figure 13b: CCD-enhanced image of a metaphase spread showing the localisation to 2p22.



3.3.6 Northern analysis and cDNA clone retrieval

With the isolation of fragment-specific genomic clones, hybridizing bands were then identified, subcloned and used as probes with northern blots. A summary of the transcript sizes (where identified) is given in table 3 and examples of northern transcript bands are illustrated in figure 14. Alumorph AB1 hybridised to a 12 kb *Eco RI* fragment of the P1 clone G09101 which identified a faint 4.5 kb transcript found in mRNA extracted from established pancreatic tumour cell lines (figure 14a). The transcript appeared to be specific to pancreatic tumour as total RNA from normal pancreas exhibited only very low levels of the transcript. Furthermore, there was no evidence of any transcript on a multiple tissue northern blot from Clontech (not shown).

Utilising this 12 kb fragment as a probe, we subsequently isolated two overlapping cDNA clones from a gridded human foetal brain cDNA library and later three more cDNAs from gridded pancreatic cDNA libraries. Hybridization of these to further northern blots containing tumour cell line RNA as well as three RNA samples from normal tissue demonstrated a transcript which appeared to be expressed at low levels in normal tissues, but exhibited elevated level of expression in tumour cell lines which varied in different lines (figure 15a). Further northern analysis was carried out on more pancreatic tumour cell lines with the cDNA M0982 (see later) to assess what percentage were over-expressing the AB1-specific transcript (called **ART1** for **A**lu-PCR **R**ecovered **T**ranscript no.1). Two further Northerns were made with pancreatic tumour cell lines and hybridisation revealed twelve cell lines from twenty four exhibited elevated levels of ART1 expression compared to the normal levels (barely detectable) in the total RNA from primary cultures of normal pancreas tissue (Figure 15a). A visual interpretation of relative signal intensity for the 24 pancreatic tumour cell lines is given in (Figure 15b) .

Figure 14 : Northern analysis of Alumorph-specific fragments and cDNAs

Figure 14a: Northern analysis of ART1 expression levels using total RNA from one primary culture sample of normal pancreas tissue (N1) and from twelve pancreatic tumour cell lines, as indicated. The transcript appears to migrate at approx. 4.4 kbp and demonstrates varied expression with little or no expression in the normal samples to three or four-fold greater expression in the cell lines T3M4, Colo357, Patu2 and PSN-1.

Figure 14b: Northern analysis using a Met-L specific P1 fragment exhibiting expression in the Suit-2 parent tumour cell line and the three clonal sub-lines which show increasing metastatic properties (S2-007, S2-013 & S2-020)

Figure 14c : Northern analysis using the AB4 Alumorph fragment exhibiting faint expression of a 3.0 kb transcript in the Suit-2 tumour cell line and its clonal derivatives which have increased metastatic potential, however, more abundant expression in the derived cell line S2-028 which is rarely metastatic

Figure 15a: Northern analysis of ART1 transcript on RNA from pancreatic tumour cell lines

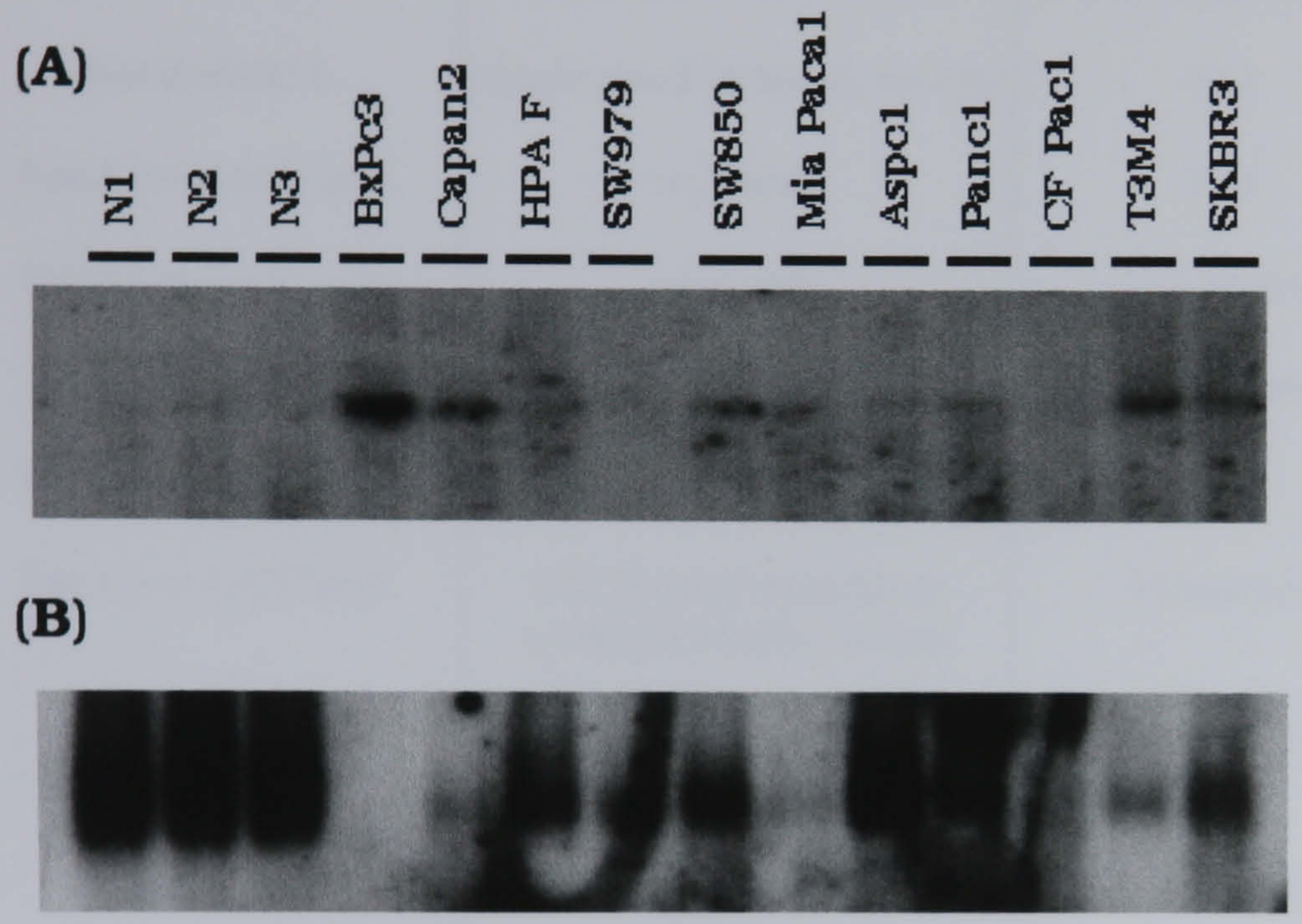


Figure 15 A: Northern analysis of ART1 expression levels using total RNA from three primary cultures of pancreas tissue (N1-N3) and from ten pancreatic tumour cell lines, as indicated. The transcript appears to migrate at approximately 4.5 kbp and demonstrates varied expression levels with little or no expression in the normal samples and up to five or ten-fold greater expression in cell lines BxPc3, Capan2 and T3M4.

B: Control hybridisation with the same filter probed with a Beta-actin cDNA to demonstrate that the differences in apparent expression are not due to differences in the amount of RNA loaded onto the gel. The loading of RNA on each lane is not uniform, however there is more of the normal tissue RNA loaded (N1-N3) as well as the Aspc1 and Panc1 cell lines which all show low level expression of the transcript.

Figure 15b: ART1 band intensity expressed as multiples of that observed (visually) with normal tissue (*)

Tumour cell line	Band intensity (*)
Normal	*
BxPc3	*******
T3M4	*****
CaPan 2	****
HAPF	**
SW979	*
SW850	***
Mia-PaCa1	*
AsPc1	*
Panc1	**
Panc3	***
CF Pac1	*
PSN-1	*
PT45	*
818.1	*
818.4	*
818.7	*********
PaTu1	*
PaTu2	***
Colo 357	**
SUIT2	**
PaCa3	*
PaCa44	**
Hu Pac1	***
(SKBR3)	***

Table 3: Details of Alu-generated fingerprint fragments (Alumorphs) examined

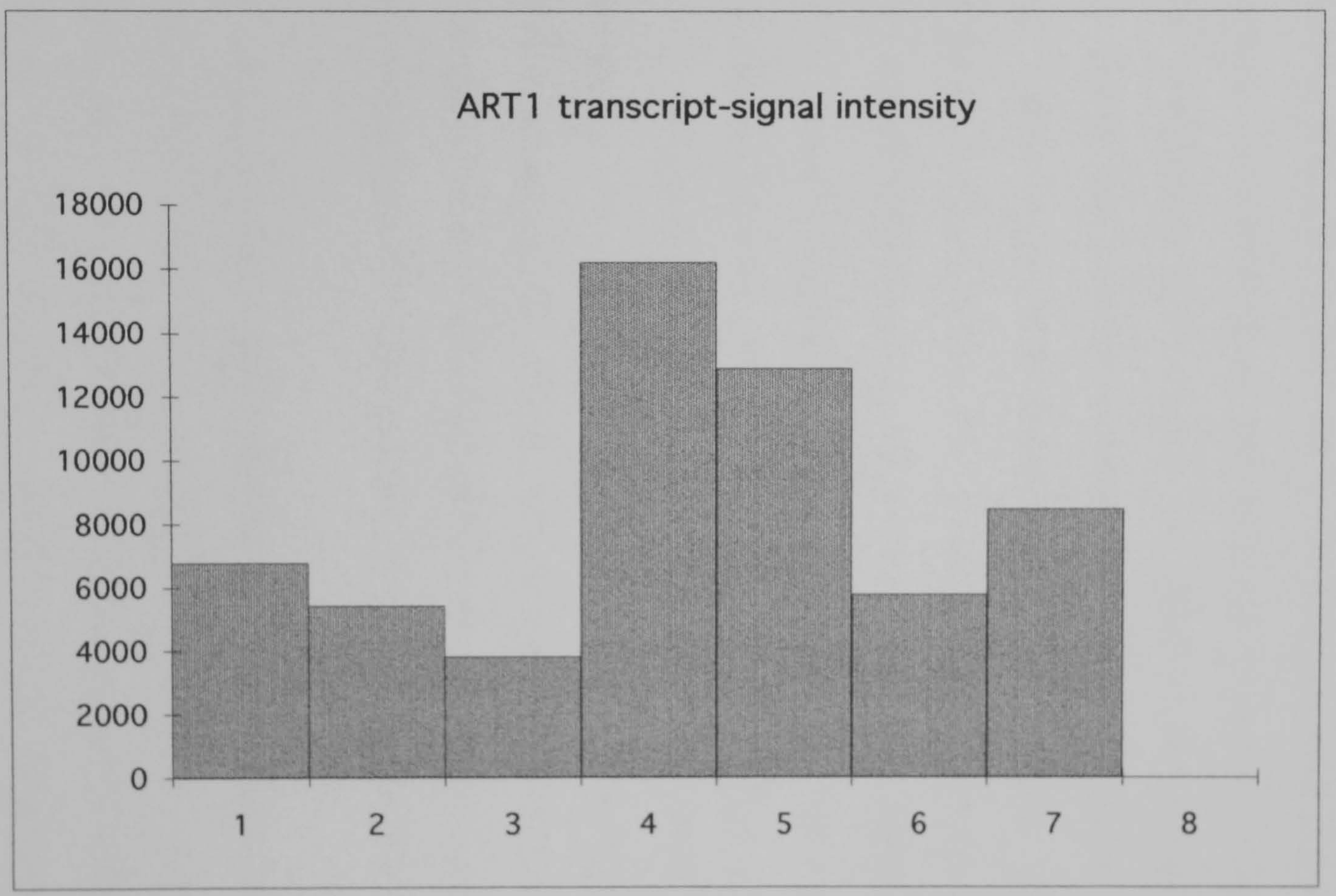
Fragment/(Sizebp)	Observation/no. cases	Southern	genomic clones	chromosomal location	Northern	cDNA clones	Seq. homol
AB1/(270)	band loss /1	Msp I polymorphism 2/18 cases exhibit loss	4P1 clones, 1YAC, 2 Cosmids, 2 BACS	2p22	4.0 - 4.5kb transcript	8-10 clones	Panc. ESTs LEU-rich gene-2p22 CYP1B1
AB2/410	band shift/1	single Hind III band, no loss	ND	19 (SCHP)	ND	ND	TATA-less T.F. (Alu)
MET-U(294)	band loss/cell lines	no bands	ND	ND	ND	ND	Rb intron (Alu)
MET-L(235)	band loss/cell lines	no bands	3 P1 clones	3p21 & 8q24	3.0 kb transcript	ND	Melanocyte EST
MET-2 /(285)	band loss/cell lines	single Bam HI band, no loss	No clones identified	ND	ND	ND	Alu
MET-3 /(190)	band loss/cell lines	no bands	ND	ND	ND	ND	ND
MET-6 (500)	band loss/cell lines	Eco RI polymorphism multiple bands-no loss	3 Cosmids	11p12-13	large transcript(>8kb)	none	Alu, ESTs
IP-1/(350)	band loss/2	single Hind III band, no loss	2 Cosmids, 1 PAC	17p13.3	2.0kb transcript	2 + Est clones	POU** PI-TP
IP-2/(280)	band loss/1	no bands	1 Cosmid	1q41	none	2	none
AB4/(700)	band loss/1	3 constant Hind III bands No loss	2 PACs	12 OR 22	3.0kb transcript*	none	BACH1-BZIP
AB5/(370)	Additional band/1	2 constant EcoRI bands No loss	ND	ND	ND	ND	Panc. ESTs transcription factor
V1 /(225)	band loss/1	ND	ND	19 (SCHP)	ND	ND	

ND = Not Defined
* - Transcript detected on Northern with Alumorph fragment
** - Homology determined from sequence of original Alumorph fragment
SCHP = Somatic cell hybrid panel

Later, attempts were made to quantify this increased expression levels using a Bio-RAD Phosphoimager with Bioanalyst™ software. The levels of the 4.5 kb transcript were quantified by integrating the signal intensities measured (figure 16b). The cell line Panc1 was the weakest signal visible therefore this was taken as an intermediate to basal level expresser of the ART1 transcript (though it does appear to be up-regulated when compared with normal pancreas tissue expression). The values for the higher expressing cell lines for volume adjusted signal intensity was then divided by that achieved from the Panc1 cell line to give conservative estimates of the levels of increased transcript expression. BxPc3 exhibited 1.5 - 2 times the level of expression compared to Panc1 as did T3M4. The highest expressing cell line was 818.7 which expressed the transcript 4.2 times the level of Panc1.

Figure16: Analysis of ART1 signal intensity with Bio-Rad molecular analyst

Sample number	Sample	Mean (PD)	Area (mm*mm)	Volume (PD*mm*mm)	Adj. Volume (- Bckgnd)	Adj. Volume/Panc 1
1	BxPc3	1291.94	13.52	17467.0288	6781.7672	1.775484921
2	T3M4	1191.38	13.52	16107.4576	5422.196	1.419545519
3	Panc 1	1072.85	13.52	14504.932	3819.6704	1
4	818.7 upper	1989.57	13.52	26898.9864	16213.7248	4.244796829
5	818.7 lower	1317.15	13.52	17807.868	12895.5112	3.37607957
6	BxPc3 (2)	1217.32	13.52	16458.1664	5772.9048	1.511362027
7	T3M4 (2)	1417.95	13.52	19170.684	8485.4224	2.221506442
8	Background	790.33	13.52	10685.2616	0	-



Included in one of the Northerns (Figure 15a) was one lane of RNA from the breast tumour cell line SKBR3. This prompted us to examine breast tumour cell lines with regard to the expression levels of this transcript. A Northern filter containing ten breast tumour cell lines as well as two transformed, non-tumourigenic cell lines (kindly provided by Helen Hurst) was probed with the cDNA M0982. Again elevated expression was identified with some of the tumour cell lines (Figure 17a). With this experiment there was no "normal" cell population to compare to as a basal expression level. It appeared the two transformed, non-tumourigenic cell lines had the highest expression level with the tumour cell lines exhibiting differential expression from barely visible to three-fold higher. From the β-actin control hybridisation however (Figure 17b), it is evident these two SV40-transformed cell lines were overloaded by comparison to the other lanes. The tumour cell lines on the Northern had been loaded in order of increasing ERB-B2 expression (from left to right), which appeared to be matched to a degree in that these cell lines also seemed to show increasing expression of this 4.5 kb transcript in the same order. This is not observed in pancreatic cell lines, since HPAF and PT45 are believed to express ERB-B2 highly yet are two of the cell lines expressing the ART1 transcript at basal levels. Conversely, T3M4 is believed to be negative for ERB-B2 expression but exhibits one of the highest levels of expression of this transcript. Further work to explore the observation was not undertaken.

With the exception of the cell line 818.7, the tumour cell lines expressing higher levels of the ART1 transcript were of a moderate to well-differentiated morphology (BxPc3, T3M4, CaPan2, Colo357). Other cell lines demonstrating a well differentiated morphology did not however express equally high levels of the ART1 transcript however (HPAF and CFPac1).

Figure 17: Northern analysis of ART1 expression with breast tumour and non-transformed cell lines

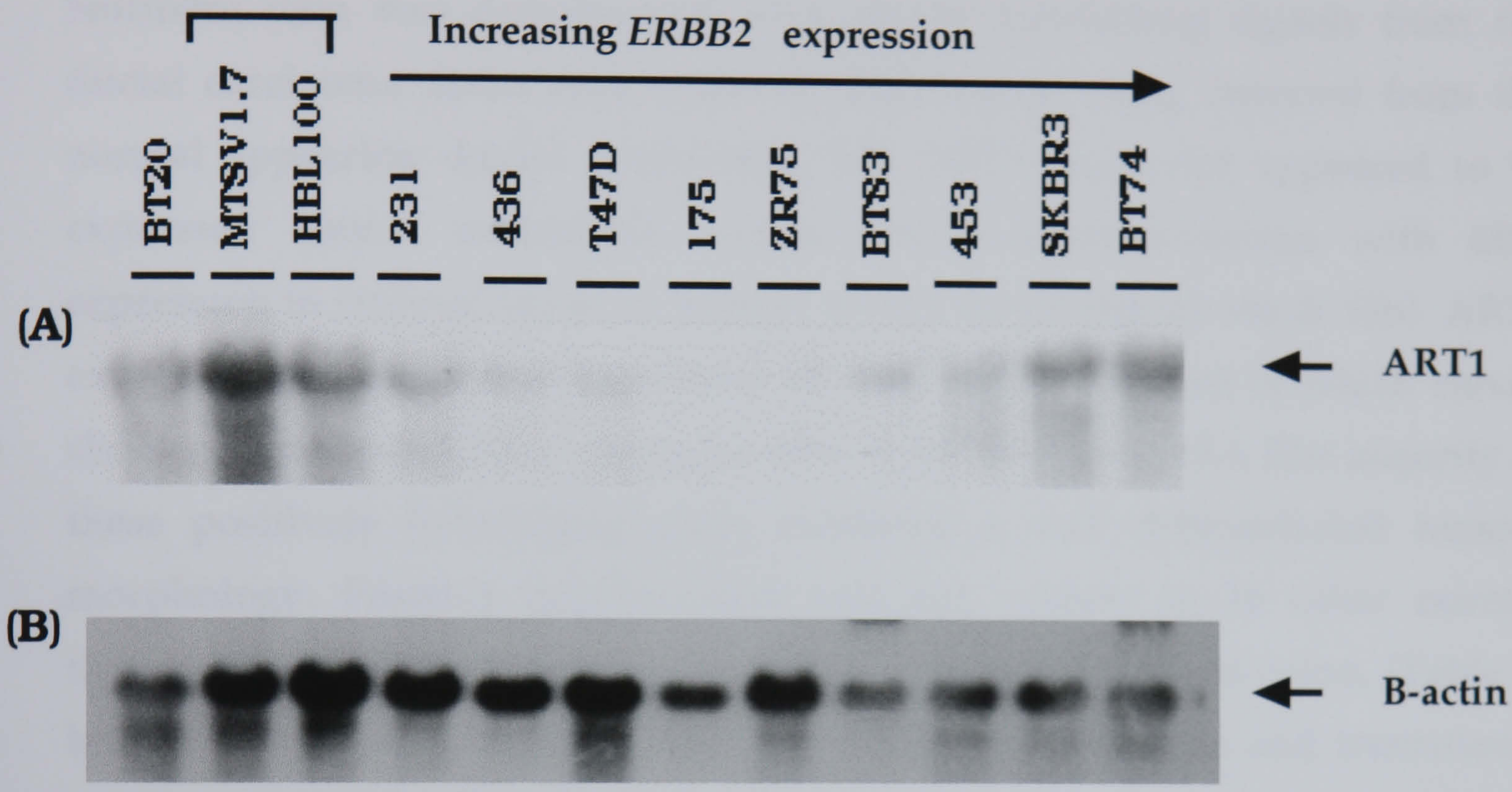


Figure 17 A: Northern analysis of ART1 expression using total RNA from ten breast tumour cell lines and two transformed but non-tumourigenic breast cell lines, (MTSV1.7 & HBL100). The 4.5 kb transcript shows differential expression with highest levels in the two transformed cell lines, tumour lines BT20, 231, SKBR3 and BT74. The high expression in the non-tumorigenic cell lines may in part be accounted for by the fact these three cell lines appear overloaded relative to the tumour lines
B: Control hybridisation of Beta-actin with the same filter

3.3.7 Tissue *in situ* hybridisation of chromosome 2 transcript

Dr. Leung carried out TISH with one of the cDNAs isolated. The smallest cDNA clone isolated from the gridded foetal brain library (discussed later-M0982) when cleaved with either *Not I* or *Sal I* gave optimum riboprobe fragments with T7 and Sp6 respectively to carry out *in situ* hybridisation. The Northern data was corroborated with strong hybridising signals from the ductal carcinoma tissue and relatively little signal being detected from the normal appearing ductal epithelium. The ART1 transcript appeared to be expressed almost exclusively within ductal adenocarcinoma with little expression in stromal tissue or normal ductal tissue (figure 18a & 18b). ART1 expression was examined in a total of sixty frozen sections of which eleven slides proved to exhibit a strong positive hybridisation (18%). The majority of these positively hybridising cases exhibited a well differentiated tumour morphology. Positive hybridisation was not evident in 16 other normal tissues investigated including normal oesophagus, liver and colon, (Table 4) but we did show some expression in duodenal epithelium and transitional bladder epithelium (figures 19a & 19b). There was evidence of hybridisation to the parietal cells of stomach, however the level of signal was significantly less.

In summary, these results so far identify a gene up-regulated in pancreatic cancer normally expressed at very low levels in most tissues. The transcript may be associated with genomic losses around a 270 bp fragment, both of which reside on the short arm of chromosome 2, band p22. Additionally, the increased expression of this transcript appears to occur in breast tumours also, suggesting the up-regulation may not be limited to pancreas cancer.

Figure 18: Tissue *In situ* Hybridisation (TISH) of ART1 transcript to Ductal adenocarcinoma of the pancreas

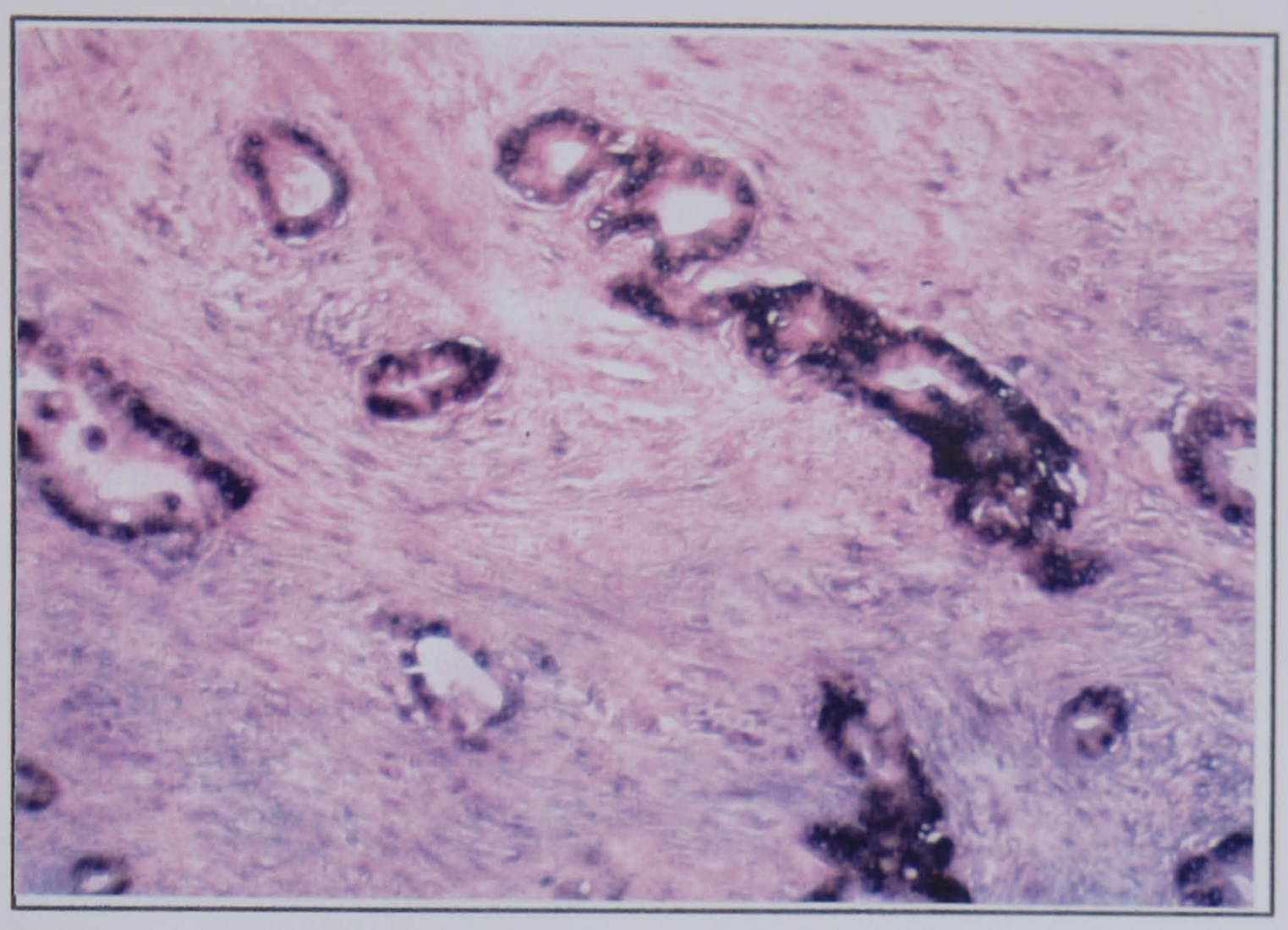


Figure 18a: Tissue *in situ* hybridisation of ART1 riboprobe to ductal adenocarcinoma of the pancreas. Strong hybridisation is evident to the cancerous duct tissue, not however to the surrounding stromal tissue.

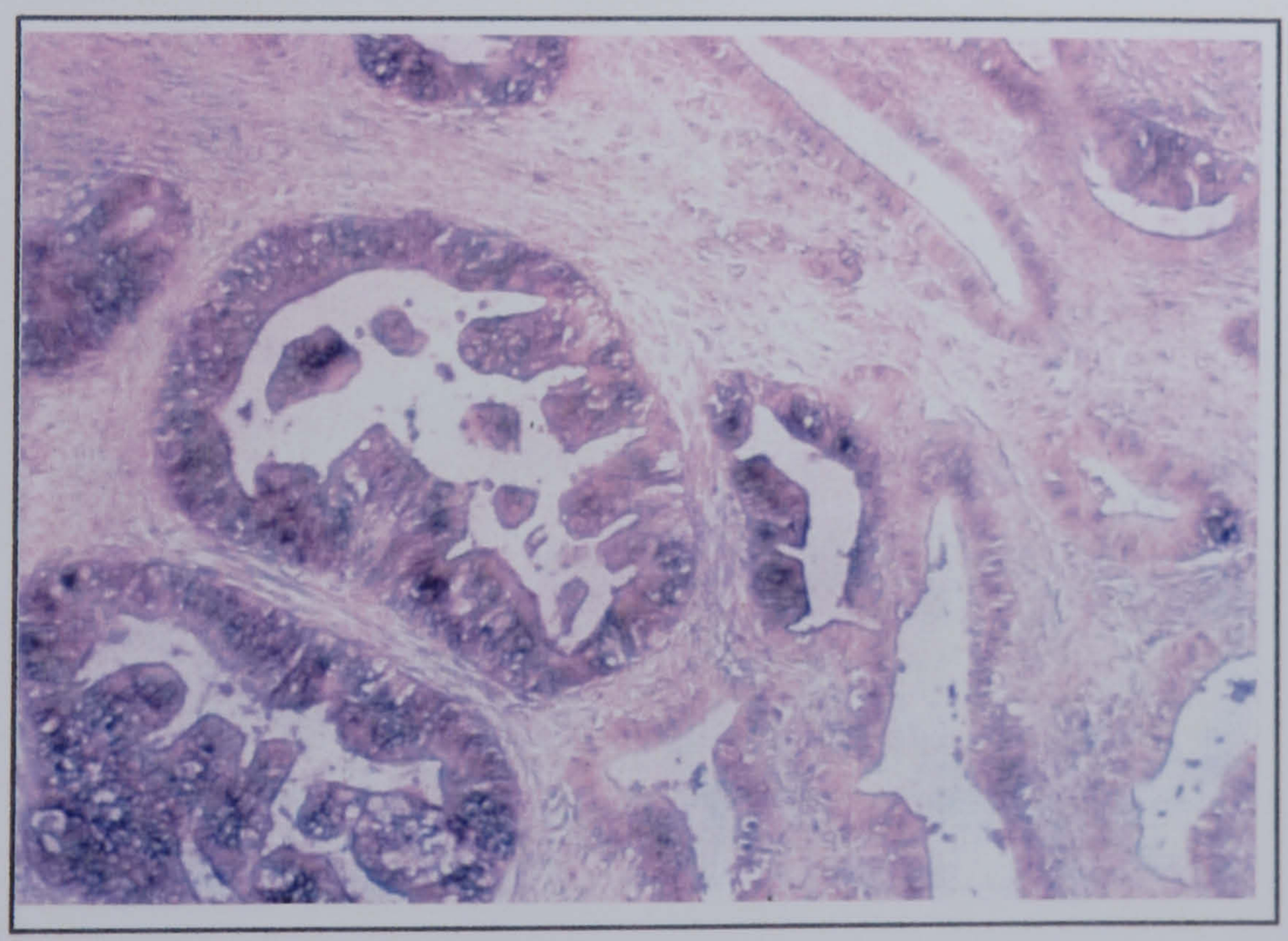


Figure 18b: At higher magnification, the hybridisation is clearly associated with the irregular epithelium of the cancerous ducts. In contrast, the regular columnar epithelium of the normal duct tissue shows relatively little evidence of ART1 expression.

Figure 19: Tissue *In situ* Hybridisation (TISH) of ART1 transcript to duodenum and bladder .

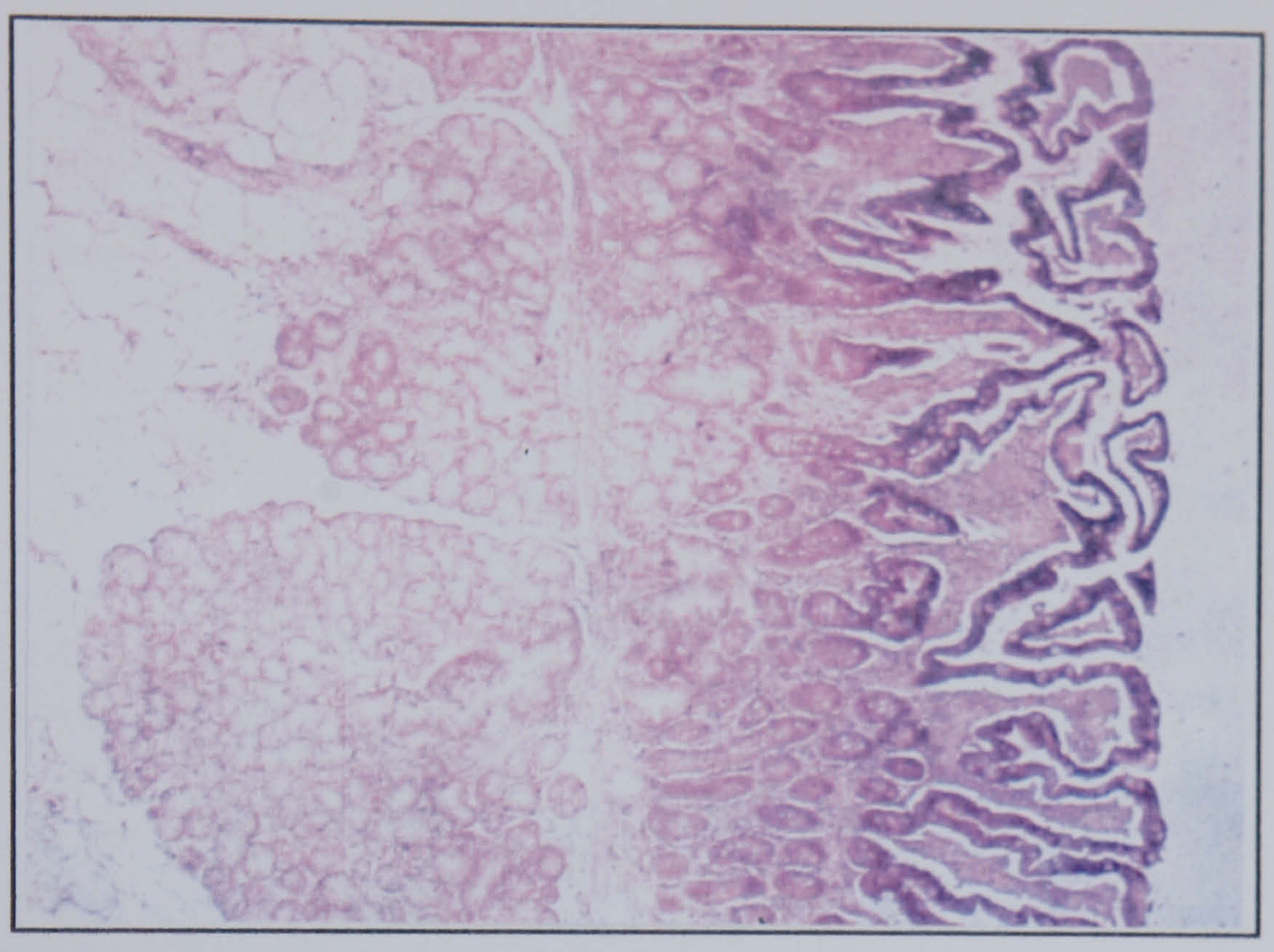


Figure 19a : Tissue *in situ* hybridisation of ART1 riboprobe to a cross section of duodenal tissue showing clearly ART1 hybridises to the surface epithelial layer and not to sub-mucosal layer or Brunner's glands.

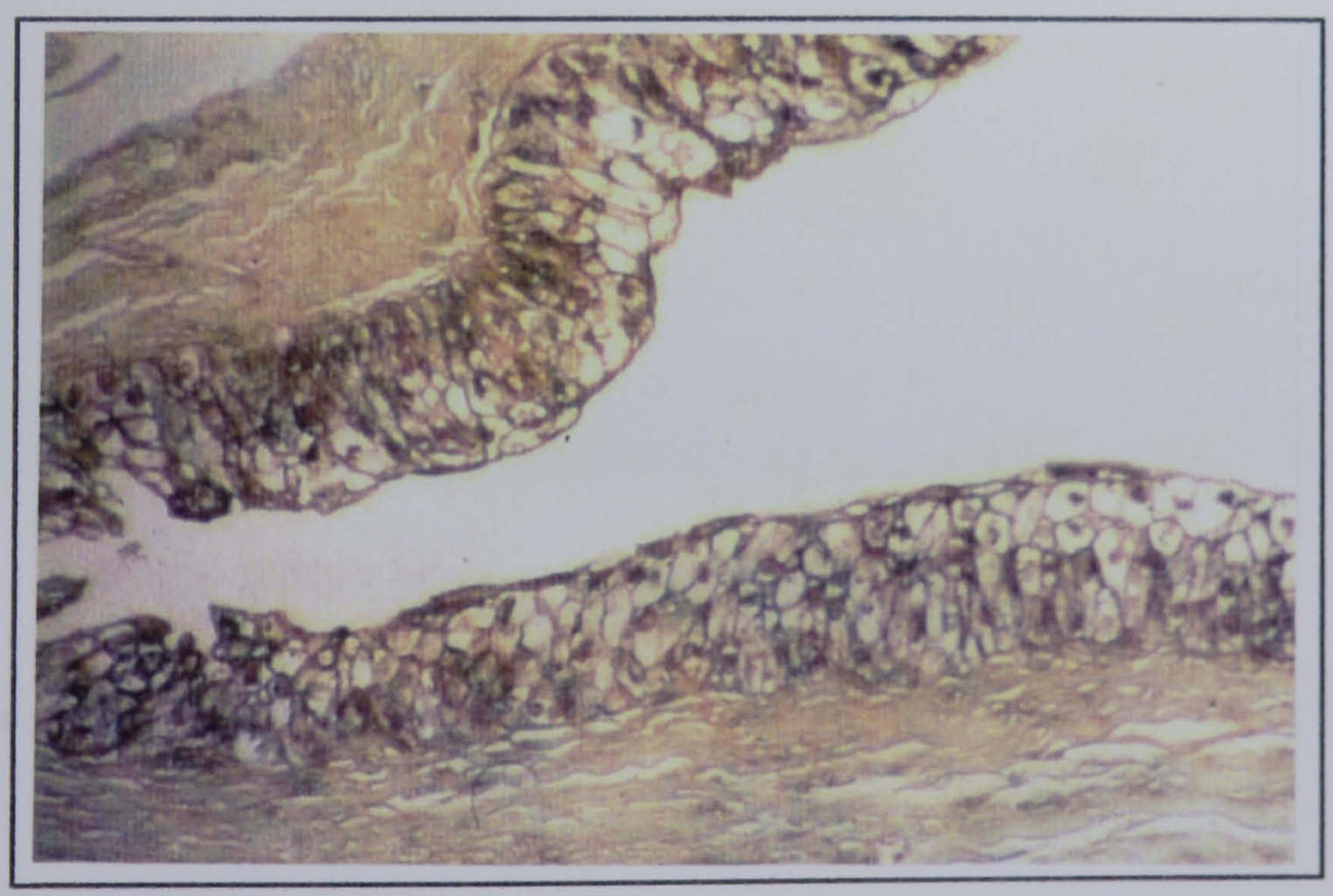


Figure 19b : A cross section of bladder wall showing ART1 expression within the transitional epithelium.

Table 4: Tissue distribution of ART 1
(*In situ* / Northern analysis)

Pancreatic ductal adenocarcinoma		++++
Normal Tissue		
Normal pancreas	Ductal	–
	Acinar	+/-
	Islets	++
Oesophagus		–
Stomach (parietal cells)		+
Liver		–
Duodenum (epithelium)		++
Colon		–
Prostate		–
Spleen		–
Artery		–
Lung		–
Thyroid		–
Bladder (epithelium)		++
Cerebellum		–
Thymus (Adult)		–
Bone marrow		–
Salivary gland		–
Testis		–

3.3.8 Methylation status of AB1 sequences

The observation of significant sequence homology between the Alumorph AB1 and a low copy repeat sequence which demonstrates a germ cell-specific hypomethylation and embryonal tissue hypermethylation prompted the investigation of the methylation status of sequences around the AB1 fragment. Starting with the established pancreatic tumour cell lines as a source of uncontaminated tumour cells we compared the hybridisation patterns obtained when this DNA and DNA from normal tissue samples were cut with the methylation-sensitive restriction enzyme *Hpa II*. Five cell lines were picked which exhibited a range of ART1 transcript expression from very high in the case of BxPc3 to virtually nothing with CF Pac1 (figure 20a). Interestingly, two of the three cell lines which exhibited high expression levels of the ART1 transcript demonstrated the hybridisation of the 270 bp probe to a small *Hpa II* fragment. The cell lines which exhibited little or no expression of the ART1 transcript showed no such hybridizing fragment. Furthermore, five normal tissue DNA samples showed no such hybridizing fragment. Upon digesting the five cell lines with *Msp I* (an isoschizomer of *Hpa II* insensitive to methylation), all five demonstrated a hybridizing band at the same size as the *Hpa II* fragment thus indicating two of the three cell lines exhibiting ART1 up-regulation also exhibit hypomethylation (figure 20b). Additionally, the cell line 818.7 also exhibited a hybridising *Hpa II* fragment and proved also to express the ART1 transcript at elevated levels, perhaps the highest level yet observed.

Figure 20: Analysis of methylation status of AB1 sequences by Hpa II/Msp I isoschizomer Southern blots

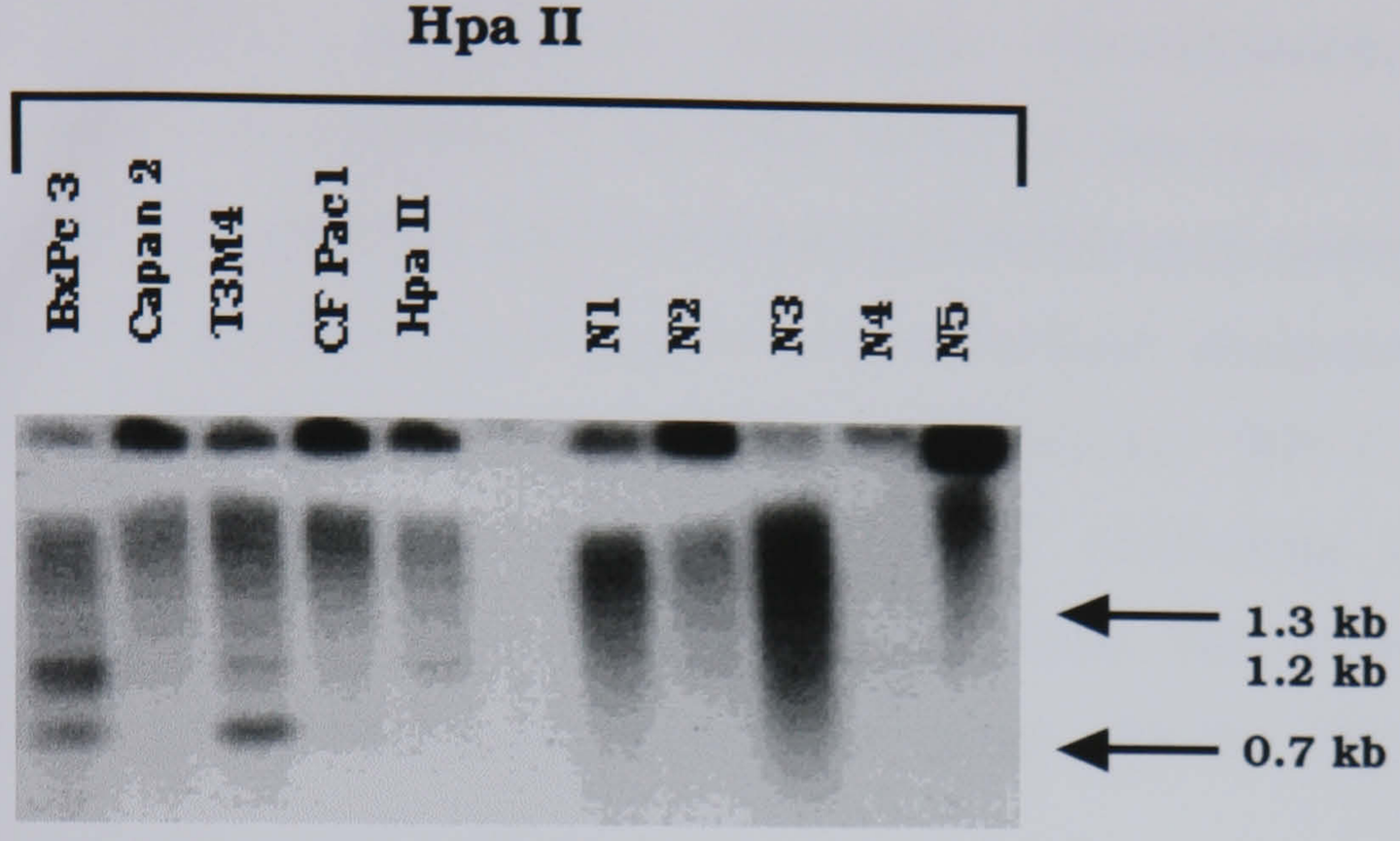


Figure 20a: Hpa II Southern of high-intermediate-low ART 1 expressing tumour cell lines probed with AB1 Alumorph fragment

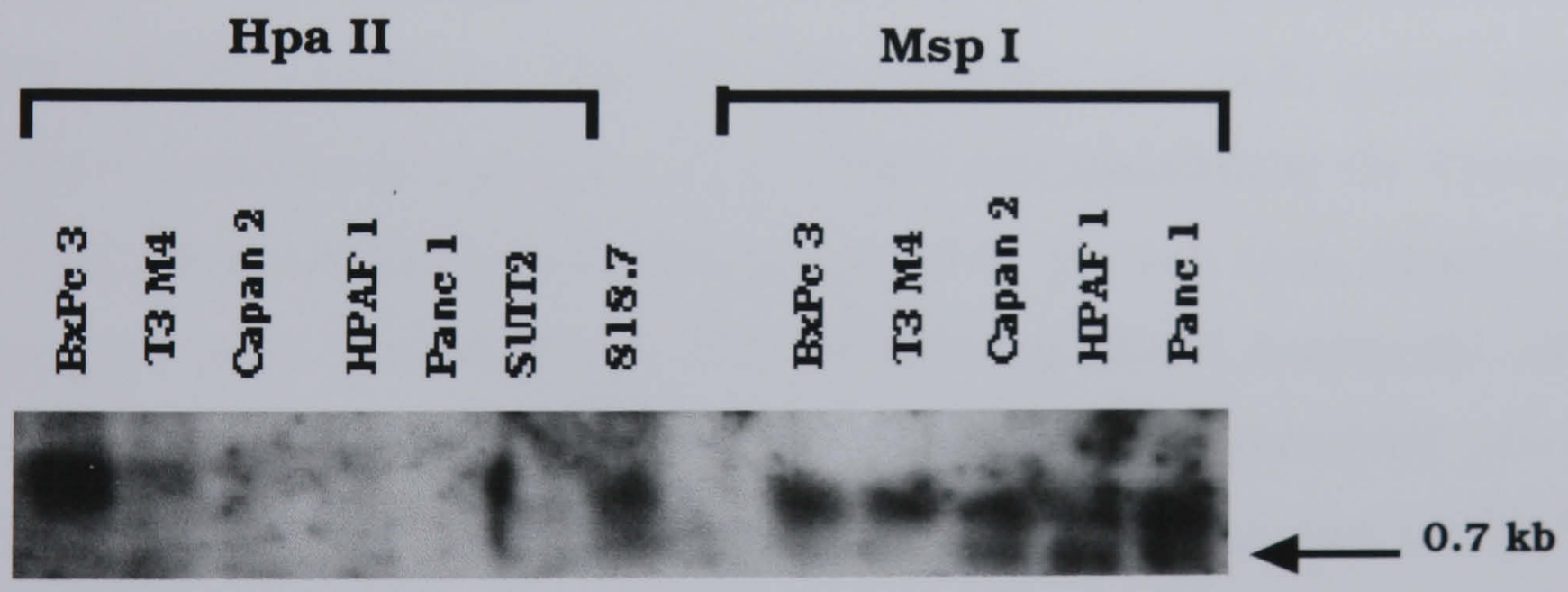


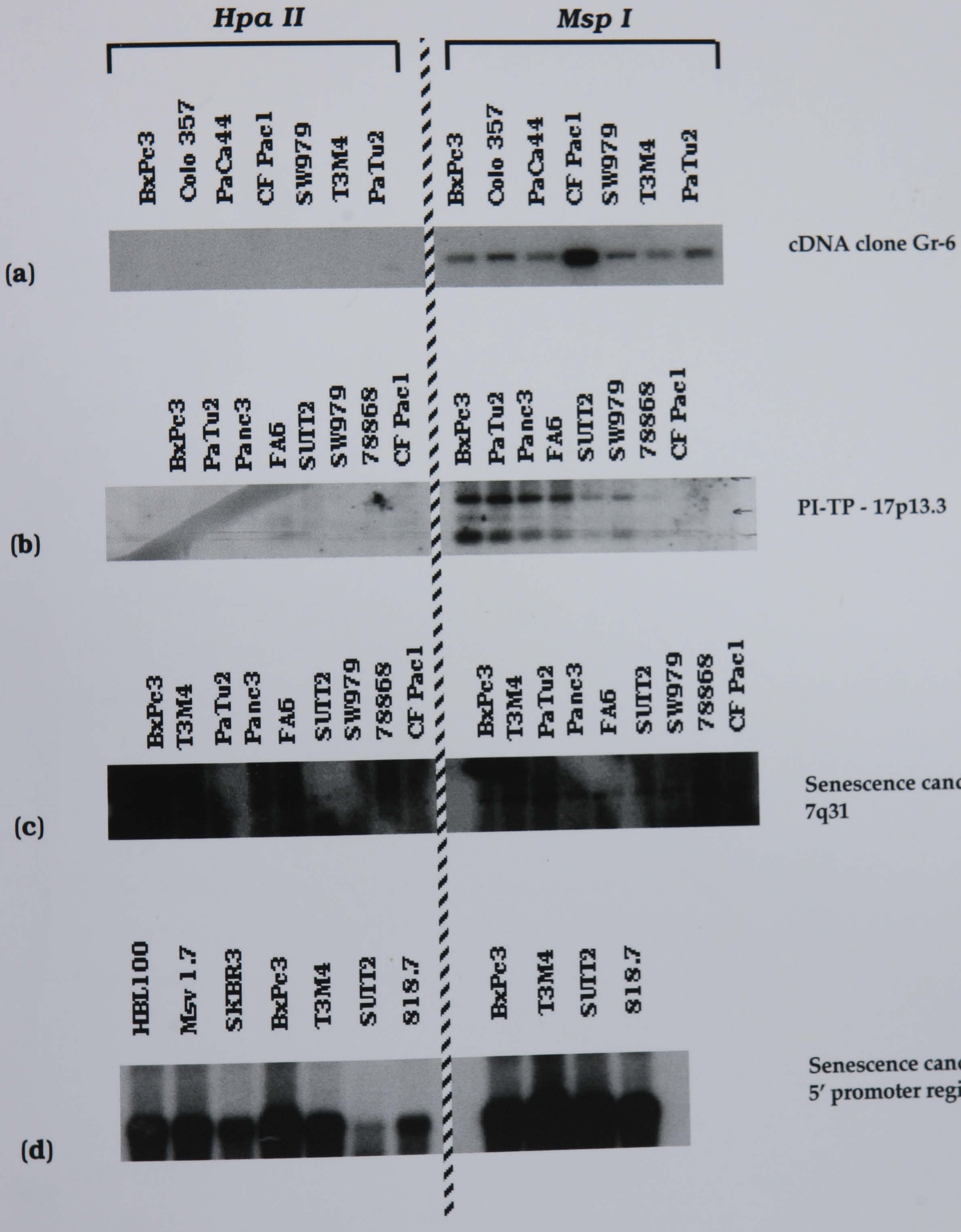
Figure 20b: Hpa II & Msp I Southern of 7 ART 1 expressing tumour cell lines probed with the AB1 fragment, demonstrating hybridisation of an Msp I band in all cases, but only BxPc3, T3M4 and 818.7 giving clear Hpa II bands

These experiments were repeated with DNA from additional tumour cell lines as well as tumour and normal tissue from individual cases. This type of analysis may be confounded by the abundant normal stromal element of any tissue explant, in order to achieve an enriched population of greater than 70% tumour cells, we have exploited microdissection of pancreatic adenocarcinoma lesions. A limiting factor in this type of approach however, is the quantity of DNA isolated from microdissected material and any analysis is limited to PCR-based approaches. Further analyses of the additional tumour cell lines exhibiting ART1 upregulation (818.7, Panc3, PaTu2 and SUIT2) revealed only 818.7 and SUIT2 exhibiting a correlation with hypomethylation and upregulated expression. The transformed breast cell line HBL100 also exhibited a *Hpa II* fragment with the AB1 probe .

To investigate whether this hypomethylation phenomenon is limited to the 2p22 sequences or is a global occurrence, the methylation status of several gene sequences was ascertained by probing *HpaII/Msp I* Southern blots of digested pancreatic cancer cell lines with cDNAs from different genes throughout the genome. These included the phosphatidylinositol-transfer protein (PITP) gene on 17p13.3, (figure 21b), the DCC gene and a novel (putative senescence) gene on 7q31 which was isolated by Dr. Donny Black at the CRC Beatson Institute (figure 21c & 21d). In each case, hybridising *Hpa II* fragments were not observed, but hybridising *Msp I* fragments were clearly visible suggesting the methylation at 2p22 was a localised phenomenon. When a DNA fragment from the putative promoter region of the 7q31 gene was hybridised to the *Hpa II/Msp I* filters however, *Hpa II* fragments were observed hybridising to the DNA of every pancreatic tumour cell line (4 in total) as well as three breast cell lines (Figure 21d). This hypomethylation is the normal situation for this promoter. Further analysis of this kind was carried out with upstream clones from two additonal genes, ERBB2 and DCC (localised to 17q21 and 18q21.1 respectively). Again there appeared no evidence of *Hpa II* fragments co-migrating with the *Msp I* fragments.

Another probe used in this approach was a 1.6 kb fragment of the Gr-6 cDNA (see below) from the middle of the ART1 transcript. In the cell lines which were high expressers of ART1 and had shown *Hpa II* tiny fragments (HTFs) when using the AB1 genomic fragment, no evidence of any HTFs was found with the cDNA clone. The *Msp I* half of the Southern however, did show hybridisation to small bands of the expected size (Figure 21a). It is thought hypomethylation associated with transcriptional activity usually extends into the first exon/intron only of a regulated gene (Helen Hurst-personal communication).

Figure 21: Methylation status of Pancreatic cell lines at specific gene loci compared to AB1 with Hpa II/Msp I Southerns



3.3.9 The CpGenome Modification kit

The methylation-specific PCR analysis consists of chemical modification of test DNA with bisulphite followed by PCR amplification. The chemical modification creates the differences between the methylated and unmethylated DNA, converting unmethylated cytosine to uracil through deamination. Two PCR primers should be designed to anneal to the alternative sequence differences. One primer set (U) will anneal to unmethylated DNA that has undergone a chemical deamination. A second primer set (M) will anneal to the methylated DNA that has not undergone the chemical modification due the the methyl group. The standard rules for primer design apply for the creation of the methylation-sensitive PCR (MSP) primers. They should be approximately 20-22 bp in length with similar melting temperatures and the product should be 100-200 bp in length. To minimize primer-dimer formation, primers should not be complementary, especially at the 3′ ends. Since most of the CpG islands are located within the 5′ upstream sequences of genes which may be 70-90% GC-rich, discrimination between the methylated and unmethylated templates is best achieved when the last base on the 3′ end of the primers is complementary to the base under analysis in the test sequence.

One of the primers designed to amplify the AB1 fragment sequence originally had the second cytosine nucleotide of a Hpa II recognition sequence as its last 3′ nucleotide. In order to function as an MSP primer however, primers complementary to this strand would have to be synthesised with either G or A as the last base at their 3′ end (see below).

TGGTTAGAAAATCACTCCTCACcGGG----------

primer-M ------- gGCCCTTGTGTGAATGACCATG

primer-U tGCCCTTGTGTGAATGACCATG

The unmethylated cytosines will react with the bisulphite chemical to deaminate cytosine to uracil, which will then be incorporated as thymine in the PCR amplification. These PCR products should serve as a template specifically for primer-U. Alternatively, if the DNA is methylated, then bisulphite will not react with the 5-methylcytosine and the PCR product should serve as a specific template for primer-M. In practice, no amplification was achieved substituting either (M or U) primers for the original in conjunction with an "upstream" primer. As an alternative, the whole area was amplified with two flanking primers which amplified 212 bp of sequence around this *Hpa II* site and the resultant PCR fragments were sequenced automatically after purifying away from excess primers and nucleotide.

Both manual and automatic sequencing of these small PCR templates gave inconsistent results. The purification of the 212 bp PCR fragments from excess primers and nucleotide proved to be a limiting step, as even when employing the most efficient methods to achieve this (Qiagen PCR-purification kit), losses of 30-50% of the PCR fragment were experienced. Some of the sequence derived is detailed in Table 5 below and figure 22.

Figure 22: Bisulphite DNA modification and sequencing of tumour and normal samples to identify hypomethylated sequences

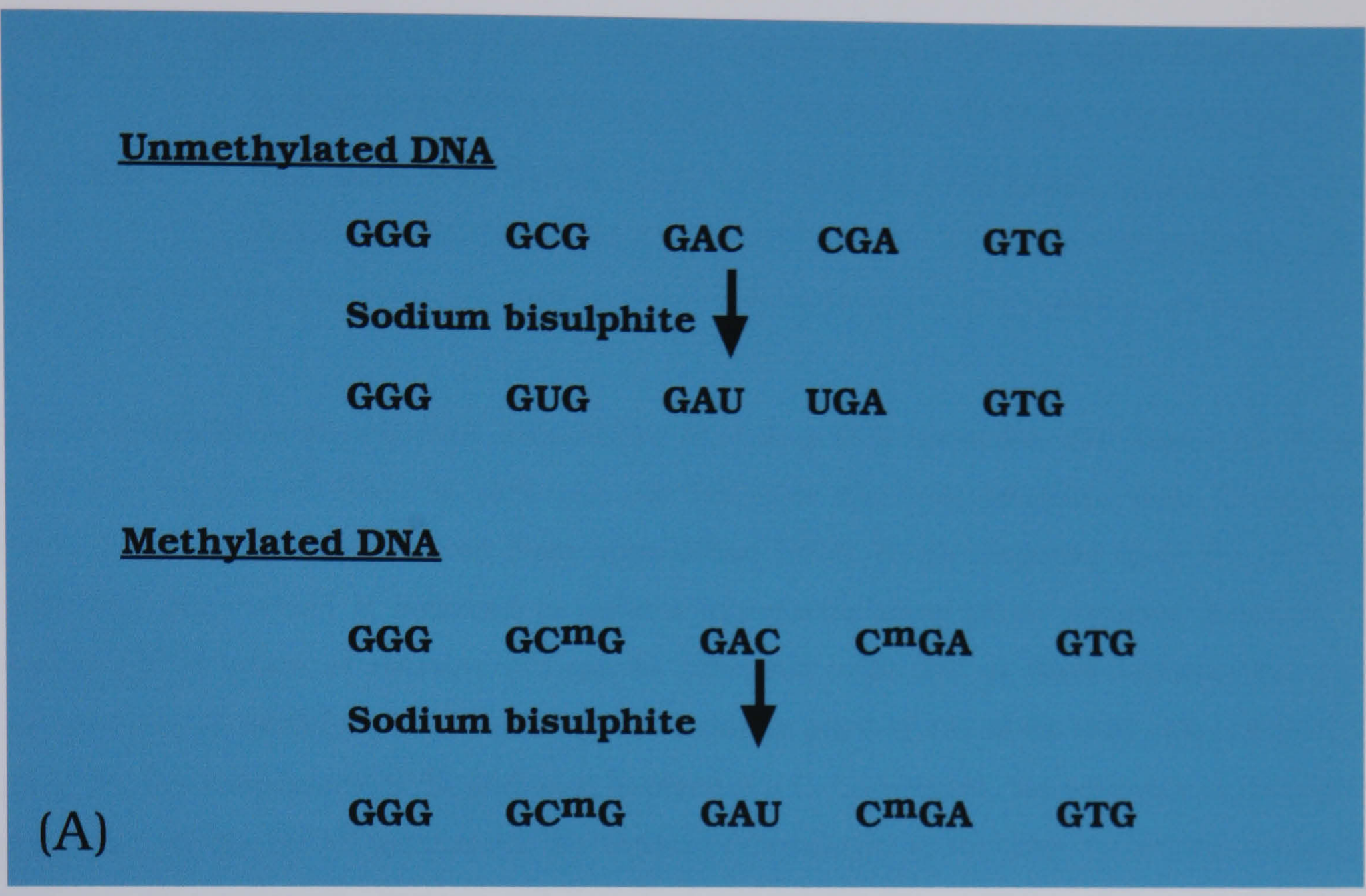
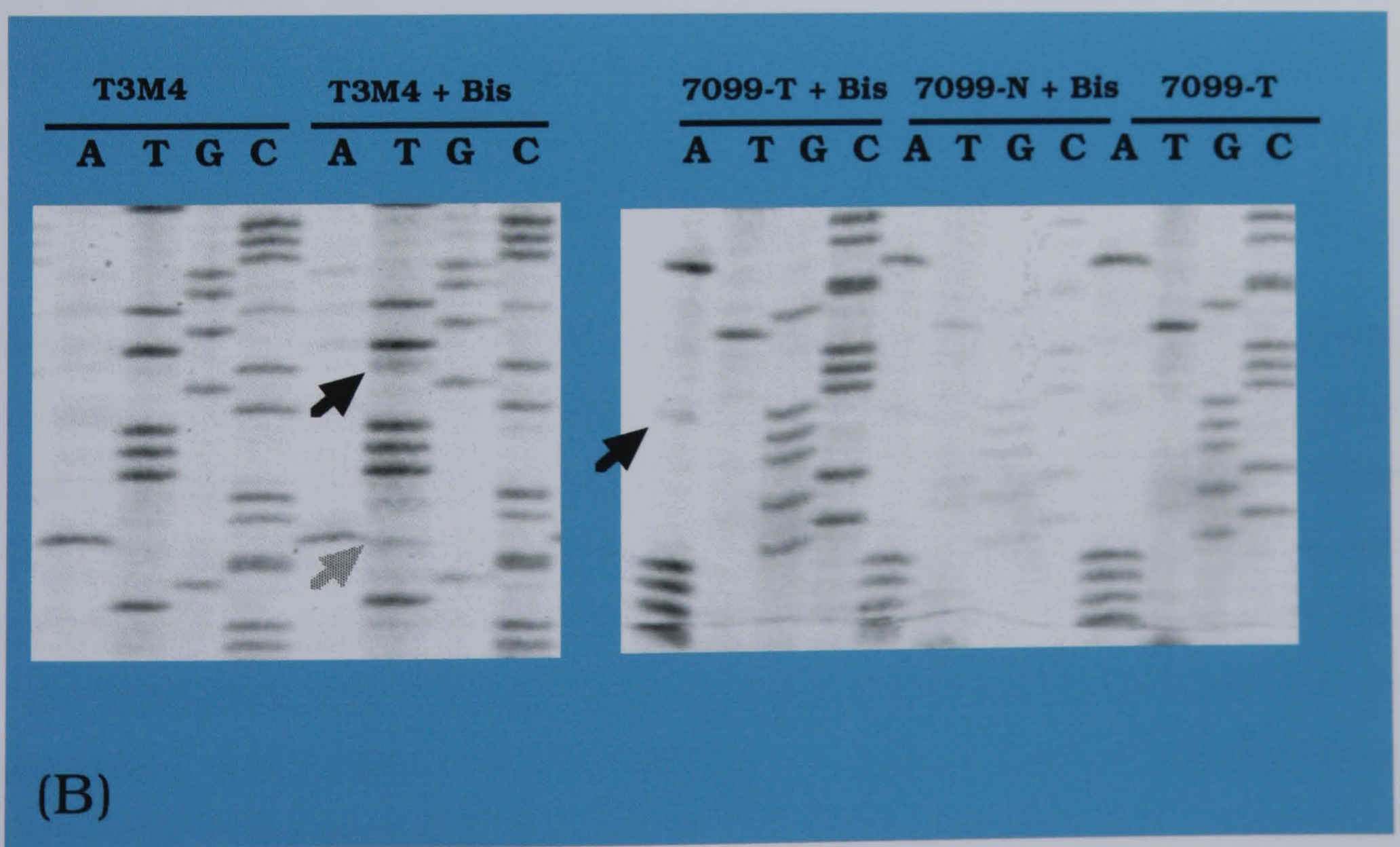


Figure 22A: Illustrates the theory of the bisulphite reaction - where all unmethylated cytosines are deaminated and converted to uracils, while 5-methylcytosines remain unaltered. **Figure 22B:** Examples of sequenced PCR products generated from treated and untreated DNA from the pancreatic tumour cell line T3M4 and a tumour/normal pair-7099. Arrows indicate the occurrence of thymine nucleotide co-migrating with cytosine nucleotide perhaps reflecting the heterogenous cell population of tumours.



T3M4-Bis	GGAGTATGTNCTGACCGTTTNGATNTCAGCCCTTTNAAAGCGCGGGCCCTG
T3M4	AGCAGGGT-CTGACCGCCGGGATTTCA-CC-TTTTNAC-CGCGGGCC-TG
HPAF-BIS	AGCAGGGT-CTGACCGCCGGGATTTCA-CC-TTTTNAC-CGCGGGCC-TG
HPAF	AGCAGGGT-CTGACCGCCGGGATTTCA-CC-TTTTAAC-CGCGGGCC-TG

Table 5: illustrates some sequence information from four PCR templates, two from each of two pancreatic tumour cell lines. In each case one template has been amplified from Bisulphite-modified DNA, the other derived from unmodified DNA. In the sequence from the cell line T3M4 (top line) which is believed to exhibit hypomethylation three separate instances of thymine (T) in place of cytosine (C) can be observed with two of these representing CpG dinucleotides. In addition, there is also T nucleotide in place of a G at the next adjacent base to two of the three replacement Ts, and an A in place of a C. Conversely, the sequence of templates generated from the HPAF tumour cell line showed no evidence of thymine nucleotides replacing cytosines. This pancreatic tumour cell line does not exhibit any evidence of hypomethylation and expresses the ART1/Lrp130 transcript at low levels.

Two instances of thymine replacing cytosine at CpG dinucleotide sequences from the bisulphite-modified T3M4 DNA suggests hypomethylation in this cell line. Further the cell line HPAF which has not exhibited evidence of hypomethylation and expresses the ART1/Lrp130 transcript at lower levels compared to T3M4. Sequence of the PCR product from the other cell line exhibiting hypomethylation (BxPc3) did not give a similar result however.

These results had to be repeated, and PCR reactions carried out in duplicate for each modified and unmodified DNA template. The DNA from four tumour samples were also modified as well as their constituent normal DNA. The second round of sequencing carried out on bisulphite treated cell line and tumour/normal case DNA was done manually. A concern was that if the bisulphite reactions were not going to completion, perhaps due to incomplete denaturation of the template DNA, as well as heterogenous results due to the probable presence of DNA from normal cells, a number of

instances of deamination may be missed. This did appear to be the case particularly with the tumour cases. In figure 22 the cell line T3M4 as well as tumour 7099, three to four instances of a cytosine base co-migrating with a fainter (more diffuse) thymine base suggested (a) we may have a mixed population of methylated and unmethylated templates, particularly from isolated tumour specimens and (b) the reactions were not optimized. In support of the suggestion that the bisulphite reaction was not going to completion, it appeared as if selective cytosine residues in the template amplified had been modified. Local areas of the template composed of Gs and Cs only appeared less likely to be modified, whereas A-T base-pairs in the vicinity of a CpG dinucleotide may have contributed to the efficient denaturation due to reduced bond energy of the A-T base-pairing.

3.3.10 CpG Island Library

The MRC-HGMP distribute a CpG island library (Cross *et al.*, 1994). The originators constructed an affinity matrix that contained the methyl-CpG binding domain of the rat chromosomal protein MeCP2 attached to its surface. Employing a column containing the matrix, they collected fractionated DNA fragments according to the degree of CpG methylation, strongly enriching for those sequences that are highly methylated. They have developed procedures for bulk isolation of CpG islands from human genomic DNA. The resulting library of CpG fragments can be used directly to isolate the 5′ end of transcripts and place genes on genomic maps. The library was supplied as aliquots of transformed bacteria which could then be plated out in high density 132 mm plates, (see Materials & methods chapter 2). Upon screening the CpG library with an IMAGE-cDNA clone specific for the 5′ end of the ART1 transcript, a single clone was eventually isolated which proved very similar to the AB1/Sp 0.3-23 sequences. This was reassuring and suggested that we were dealing with the 5′ upstream sequences of a gene.

3.4 The complete sequence analysis of the ART1 transcript

3.4.1 Isolation of cDNA clones

The isolation of two overlapping cDNA clones from a gridded foetal brain cDNA library obtained from the RLDB, and three cDNAs from gridded libraries from pancreatic tumour and normal material (Gress *et al.*, 1995), facilitated the construction of a sequence contig of the ART1 transcript. Initially the two cDNA clones from the foetal brain cDNA library were thought to be homologous at one end representing the polyadenylation site with a $A_{(30)}$ poly-A tail. The clone M0892 was 1.6 kb and the clone M1314 was 0.6 kb. Three more cDNA clones were isolated from gridded pancreatic libraries, clones Gr-5, Gr-6 and Gr-9 were 5.0kb, 2.4 kb and 3.4 kb respectively. It was predicted that the whole transcript should be represented within these five cDNAs. The gridded library filters of pancreatic cDNA showing the three positively hybridising clones (figure 23b) appeared remarkably clean and subsequent challenges of the same filters with cDNA and genomic fragments of the ART1 transcript yielded no further clones. If this is considered with the Northern and tissue *in situ* data, it may suggest that we have indeed a transcript expressed at low levels normally and therefore ill-represented in cDNA libraries available particularly from normal tissue. This observation became pertinent later in the investigations when trying to isolate more cDNA clones. With the accumulation of more sequence data, clones M1314 and Gr-9 were ruled out, since they were non-homologous with the others. All five were similar at the 3′ putative polyadenylation sites, however more pronounced differences in sequence generated from the opposite ends of the M1314 and Gr-9 inserts indicated they did not represent the same transcript and subtle differences were subsequently identified near the putative 3′ end of the sequence.

The clone M0982 proved straightforward and the complete sequence of this clone was achieved in eight weeks of manual sequencing. The cDNA clone was partially restriction-mapped and two separate subclones were generated at

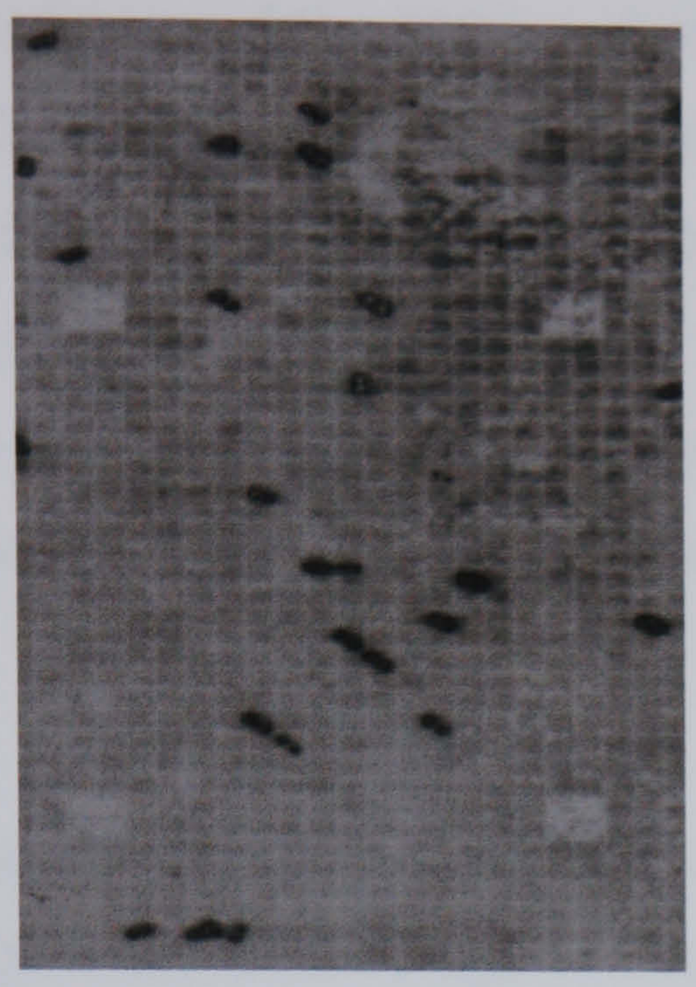
600 bp and 850 bp after digestion with *Eco RI* and *Hind III* respectively. The small *Hind III/Eco RI* fragment sequence was derived from the full size cDNA clone once flanking sequence was generated and primers could be ordered (see M0982 restriction map-appendix 4). The 2.4 kb Gr-6 clone was completely contiguous with M0982, extending our sequence contig by 800 bp, which was completed some weeks later. The large 5 kb Gr-5 clone however proved more difficult to sequence completely. It was quickly ascertained that Gr-5 and the other two cDNA clones were homologous at their putative 3′ ends and therefore we could concentrate on the other half of the cDNA clone.

3.4.2 cDNA clone Gr-5 and accumulation of additional cDNAs

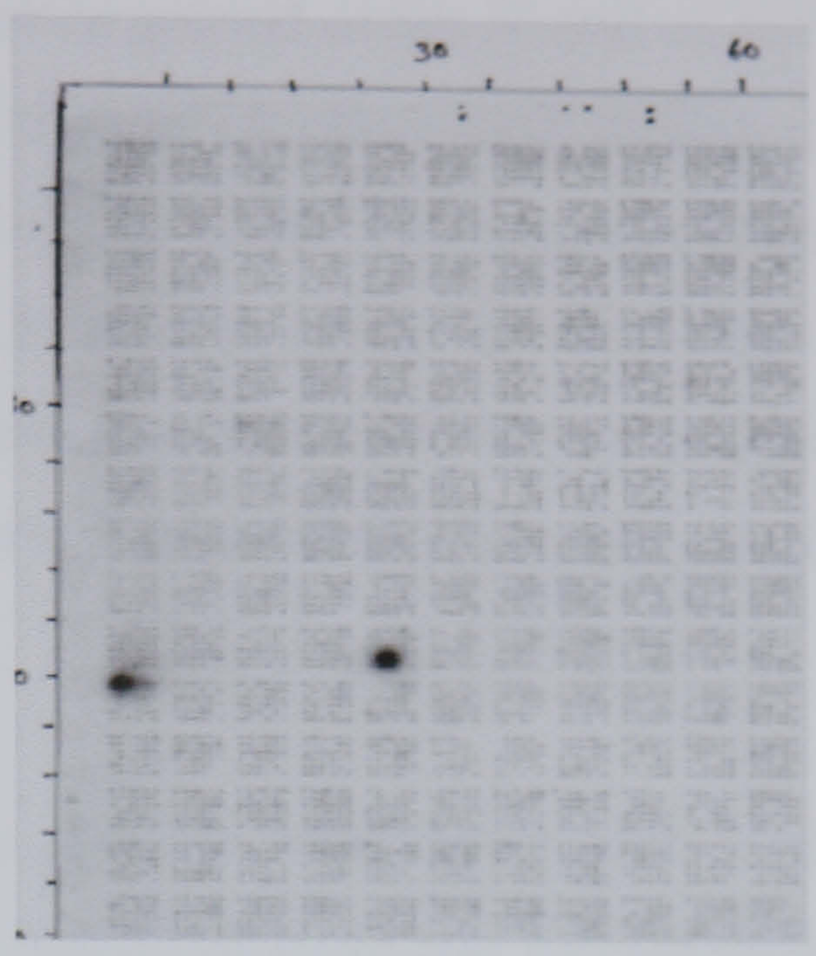
Progress was made sequencing the 5′ ends of the Gr-6 clone simultaneously with the Gr-5 clone. However, the 5′ end of the Gr-5 clone was more G-C rich and four primer-walk sequencing reactions produced only 730 bp of contiguous sequence. This sequence represented a putative open-reading-frame (ORF) on analysis by the programme GeneJockey™. Detailed BLAST and FASTA searches with this ORF and the 3′ sequence data generated so far did not demonstrate any significant homologies with known genes in the databases.

The size of clone Gr-5 represented a large template for the T7-based *Sequenase* ™ kits in use in the laboratory at this time and efforts to try Taq-based sequencing systems to overcome this and the problems with a G-C rich template were undertaken. A definite advantage was gained with the introduction of the ^{33}P labelled ddNTP terminator nucleotides with the new *Thermostable-Sequenase*™ kit from Amersham-Pharmacia Biotech (described in chapter 2).

Figure 23: Examples of gridded library technology identifying two cDNA clones specific for the upregulated transcript ART1

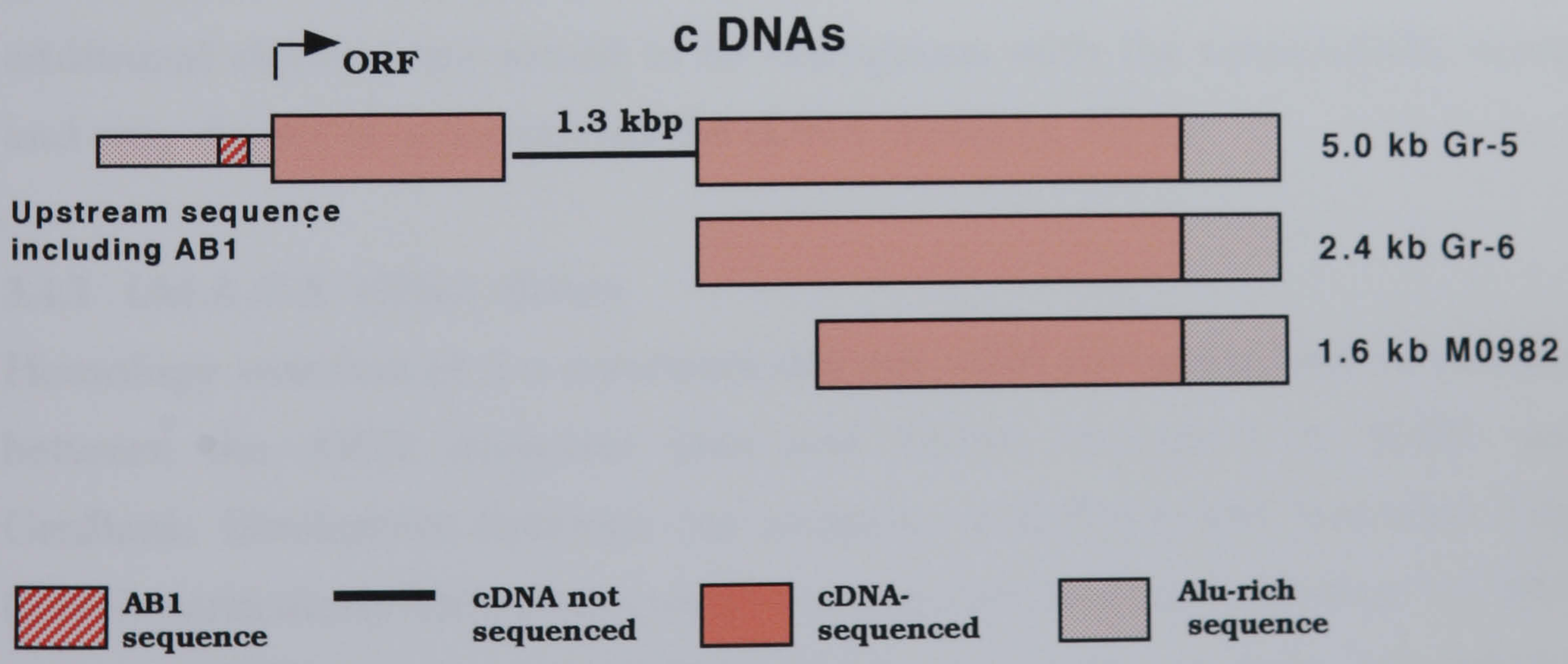


23a: A section of a dual-gridded human PAC library. Each clone is spotted in duplicate, therefore unless a positive signal has two spots in the correct orientation it is ignored as a false positive



23b: Section of a gridded cDNA library produced by Gress *et-al.* The cDNA libraries were only singularly gridded, however, the signal to noise ratio is not as much of a problem. The result above highlights cDNAs Gr-6 and Gr-5 probed with the brain cDNA insert M0982

23C: A representation of the cDNA sequence gained from the isolation of clones from the EPCRLS. The 5.0 kb cDNA clone Gr-5 proved difficult to sequence completely, eventually discarded as a product of co-ligation in the library manufacture. It did nevertheless provide us with 5' sequence contiguous with the upstream sequences containing AB1



Protracted experiments were carried out to close the gap between the 5′ end of the Gr-6 clone and the 730 bp of sequence gained at the 5′ end of the Gr-5 clone (figure 23C). It became apparent that the cDNA clone Gr-5 included a block of repeated sequence within the 5′ half of the clone. The sequence derived from this region also exhibited areas of homology with *E. Coli* and *C. Elegans* genome sequencing projects. The conclusion was that some contaminating genomic material (human or bacterial) had been included in the ligation stages.

The Gress pancreatic cDNA libraries and the foetal brain cDNA library from the RLDB were probed again with a fragment from the 730 bp of sequence as well as the Gr-6 clone with no further cDNAs forthcoming. Other cDNA libraries created from tumour tissues such as colon and lung were obtained from the MRC-HGMP and two additional clones (Co.Ca-5 and Co.Ca-7) were isolated from the colon tumour library using a restriction fragment from the 5′ end of the Gr-6 clone. No additional clones were isolated from the lung tumour library, the normal colon library or a placental library obtained from Dr. Helen Hurst. A further cDNA clone was however isolated from a mouse teratoma cDNA library obtained from Dr. Anna-Marie Frischauf again using the Gr-6 restriction fragment as a probe. Upon sequencing, all three of these additional clones were found to be contiguous with the Gr-6/M0982 contig and only clone CoCa-5 extended the cDNA contig by 200 bp.

3.4.3 I.M.A.G.E cDNA clones

Homology searches of the databases did not yield any significant similarities between the ART1 sequence data and known sequences in EMBL and GenBank. Similarities between our sequence and Expressed Sequence Tags (ESTs), particularly ESTs from pancreatic islet-cell libraries (Takeda *et al.*, 1993) and the EPCRLS (Gress *et al*, 1995) were however encouraging. The IMAGE cDNA clones from the Washington University/Merck Est sequencing project were a rich source of cDNA clones for this project and the 17p13.3 project described below. This genome sequencing initiative is based on the random

sequencing of cDNA clones from libraries generated from different tissues. These cDNA clones are available through a consortium (the I.M.A.G.E. consortium) free of royalty. Upon acquiring four IMAGE cDNA clones from the MRC-HGMP, sequence derivation from both ends was quickly produced and cDNA contig assembly with the existing 5′ sequence of the Gr-5 clone was achieved. On searching the databases again repetitive element homology was a recurring hit, in addition similarity was observed with repeated sequences within ribosomal RNA (rRNA) intergenic spacer elements (>90%). Subsequent database searches with the rRNA spacer identified complete homology with the hypomethylated sequence Sp-0.3-23 and the AB1 sequence originally identified by Alu-PCR fingerprinting. Additionally, genomic sequence from P1 shotgun clones (see later) which encompassed the AB1 sequence motif also exhibited homology with the rRNA intergenic spacer (appendix 8).

3.4.4 Shotgun cloning and Genomic sequencing

In an attempt to generate more sequence information quickly and to work towards facilitating derivation of the eventual gene structure of the ART1 transcript, simple shotgun libraries of two of the P1 clones were generated.

The P1 clones chosen were those used for the original FISH localization (N1846 and G09101). The P1 clone G09101 was firstly used as a template to generate a simple shotgun library by double-digesting the whole clone with *Eco RI* and *Bam HI*. No attempt was made to purify the P1 arms away from the insert DNA as we would employ a positive screening regime to isolate shotgun cloned fragments which hybridised to the cDNA clones. Of course some of these positive clones may have vector DNA at either end of the fragment, but again this was seen as a minor consideration as BLAST searches would identify such vector sequence. The primary objective was the derivation of as much genomic sequence information as quickly as possible, particularly at the putative 5′ end of the transcript. The *Bam HI/Eco RI* double digests were incubated for only 2 hours, thus minimizing the possibility of any contaminating exonuclease degradation and having a bias towards only

partial digestion which of course is beneficial in shotgun library construction. The digests were heat inactivated at 65^{0}C for 15 minutes then allowed to cool on ice before setting up ligation reactions with phosphatased *Bam HI/Eco RI* cut Bluescript™ KS+. Incubated overnight at 15^{0}C, large ligation volumes (50 µl) were adopted to minimize the influence of contaminating impurities co-isolated with the fragments. Ratios of 10:1 (Vol/Vol) in favour of the shotgun fragments over vector molecules were set up with approximately 100 ng of vector DNA included (see Materials & Methods chapter).

The second P1 shotgun library was created by first digesting the N1846 clone with the frequent cutting restriction enzyme *Rsa I* over a time course, thus generating partial digests at the shorter time points. The resulting blunt-end fragments could then be ligated into *Sma I* cut Bluescript™ KS+ vector. The *Rsa I* digests were set up and four 10 ml aliquots were removed at thirty minute intervals, added to phenol/chloroform and vortexed to maximise the enzyme inactivation. The five pools of phenol/DNA extracts were then pooled and precipitated together resulting in a mixture of DNA fragments cut to various degrees of completion (see materials & methods chapter). Ligations were set up in a large volume of 100 µl with approximately 100 ng of Bluescript™ KS+ DNA to approximately 1 µg of shotgun fragments

3.4.5 Genomic sequence similarities identify *CYP1B1* gene

Altogether, 176 of the shotgun clones were isolated after hybridising to the Gr-6, M0982 and 5′ ORF cDNA fragments. Many of them represented duplicate clones and many of them overlapped as contigs, the largest of which was 5.1 kb. Appendix 8 gives a summary of the clone numbers, origin, size and the extent to which they were sequenced. A contiguous block of five shotgun clones spanning 2.3 kb of yielded an interesting sequence similarity upon searching with the more sensitive FASTA programme. Additionally, sequences from another shotgun clone (P1sh19) which proved contiguous with the original AB1 sequence also demonstrated the same sequence similarity with a Dioxin-induced Cytochrome P450 gene (*CYP1B1*). The

genomic sequence highlighting this *CYP1B1* similarity was also demonstrating high level homology with the intergenic spacer sequences from *E.Coli* and *C.Elegans*. Further sequence-tags (200-300 bp) from the T7 end of the P1 clone N1846 and the BAC clone B21-F3 demonstrated reasonable sequence similarity (39%) with human *CYPIIIA* (appendix 2e) as well as a human *CYP3A5*. (41% - appendix 2f). The *CYP1B1* gene was already localised to bands p21-22 of chromosome 2 and the cDNA sequence accumulated for the ART1 transcript was quickly compared to the *CYP1B1* gene sequence for any homology. No similarity was evident. It therefore suggests the ART1 gene must reside in close proximity to the *CYP1B1* gene. Further analysis of the *CYP1B1* gene sequence (accession number: M38457) and in-house genomic and cDNA sequences with the ABI-sequence assembly package AutoAssembler™, established that the rRNA-like repetitive sequence encompassing the AB1/hypomethylated sequence was contiguous to the 5′ end of the *CYP1B1* gene sequence.

CYP1B1 is the third and newest member of this P450-1 family, identified in 1994. Although sharing 40% sequence homology with the other family members across species, *CYP1B1* appeared to have a less complex gene structure with the coding sequence contained within three exons instead of seven as in *CYP1A1* and *CYP1A2*. Interestingly, both *CYP1A1* and *CYP1A2* have previously been localised to chromosome 15q. This perhaps furnishes us with a possible explanation as to why the cytogenetics laboratory were observing dual localisation of one of the P1 clones (G09101) to 15q.

3.4.6 A 2p22 transcript map

This represented the first time it was known exactly where on the short arm of chromosome 2 our transcript of interest was located (apart from the FISH location). Having a marker in the *CYP1B1* gene and details of its map position relative to known sts/est markers could be gained from the Whitehead physical mapping database. It was quickly established on which contig map *CYP1B1* could be located on chromosome 2 (WC2.3) and which CEPH YACs

lay on that contig. By identifying two closely flanking markers on that contig,(D2S177 & D2S2328) it was quickly established what transcripts had been localised there. The marker D2S177 lay just 5′ to the *CYP1B1* gene and D2S2328 lay 2 cM proximally from D2S177. Within this interval thirty seven cDNA markers of which eight were known genes (including *CYP1B1*) had been localised. The majority of these cDNA markers were single-read ests generated either by the Washington Univ./Merck or the TIGRE Est sequencing projects. In addition to *CYP1B1*, there were seven other known genes or homologues of known genes including a 130 KD Leucine-rich protein (130 KD Leu-R), a second drug oxygenase-type protein 3-Hydroxyanthranilate-3,4-Dioxygenase (HAO), a hypothetical Nuclear-Ribonucleoprotein Intergenic Region similar to an *S.Cerevisiae* gene (Het-Ribo) and a Sodium/Calcium exchanger precursor protein (Na/Ca-Ex).

The sequences for these genes were fetched from the EMBL/GenBank databases and compared against the cDNA sequence data using the ABI-AutoAssembler™ package. Sequence similarity was evident between the Gr-6/Co.Ca.5 sequence and the Leu-rich 130 KD protein transcript. Upon trying a sequence alignment with the 2.4 kb Gr-6/M0982 cDNA contig however, the sequence similarity became less significant apart from the two polyadenosine tails. There was similarity between the Gr-6/Co.Ca.5 contig and the 130 KD transcript, and when aligned as separate entities using the AutoAssembler™ programme, both the Gr-6 and the Co.Ca.5 cDNAs were very similar with the 130 KD transcript. The M0982 cDNA itself again did not appear to exhibit any homology with the 130 KD transcript, however when broken down into the *Hind III* and *EcoRI* fragments which were sequenced separately, the *HindIII* fragment and a *Bam HI/EcoRI* fragment both proved homologous to the 3′ region of the 103 KD Leucine-rich mRNA sequence.

3.4.7 Lrp130: an upregulated transcript cloned from Hepatoblastoma

The 130 KD Leucine-rich transcript was originally isolated in 1994 as part of a study to isolate FGF-associated genes from the hepatoblastoma cell line

HepG2 (Hou *et al.*, 1994). The transcript is 4.7 kb in length encoding 3.8 kb of coding sequence which appears strongly expressed specifically in the HepG2 cells but no evidence of the transcript was observed in any of seven other tissues examined (brain, heart, liver, kidney, lung, prostate and spleen). Thus the 130 KD transcript shares a similar expression profile with the ART1 transcript in terms of size and its low expression levels in multiple tissues. Subsequently, the authors disclose that the gene did appear to be expressed in normal small intestine and that the 130 KD protein (designated Lrp130) had no FGF or EGF binding activity and so was subsequently set aside (Wallace McKeehan, personal communication). Leucine residues account for 13% of the Lrp130 peptide sequence, the 167 leucines appear about every 15 to 19 amino acids. This is also a feature in common with a hypothetical peptide sequence derived from the whole M0982/Gr-6 cDNA transcribed three separate reading frames. Leucine residues make up 10-12% of the M0982/Gr-6 peptide sequence appearing 6-17 amino acids apart. Neither hydrophobic amino-terminal secretory signal peptides sequences nor hydrophobic transmembrane domains were apparent from the putative peptide sequence of Lrp130 suggesting it is unlikely to be secreted as a transmembrane molecule. Their deduced amino acid sequence for Lrp130 contains four potential N-linked glycosylation sites in the amino-terminal end of the molecule.

3.4.8 Identification of upstream sequence homologies

Additional upstream sequence from the Lrp130 gene proved to be homologous to the rDNA repeated sequence we had identified from cDNA clone Gr-5 and the homologous IMAGE cDNA clones sequenced. A contig of cDNA clones and Lrp130 sequence extended the upstream sequence information to approximately 1,800 bp, incorporating the rDNA repeat element and the original AB1 alumorph which appeared 130 bp upstream of the Lrp130 start site (appendix 3). Upon carrying out promoter scans and motif searches with the Nix programme (see Materials and Methods), an adjacent CpG island which overlapped with the AB1 sequence was predicted with the

Grail package. Additionally, Line1 sequence homology (only 240 bp) was detected 300 bp upstream from AB1 and Alu sequence homology 52 bp downstream from AB1 (only 60 bp between AB1 and the start site). This information furnishes us with possible target sequence to explain how the alumorph AB1 was generated. A possible CAAT enhancer element was present approximately 320 bp upstream of the start site and TATAA box around the predictable 30 bp upstream of the transcription start site. Additional scanning software (Nix) available from the HGMP website was employed to identify possible sequence motifs within this 1,800 bp area. Sequences for numerous transcription factor recognition motifs were highlighted including AP-1, NF-k and Metallothioneine heavy metal response elements. It remains to be seem however if these apparent sites really exist and will depend on *in vitro* footprinting studies to identify any such interactions.

3.4.9 RT-PCR experiments

Reverse transcriptase-PCR experiments were initially carried out to try and amplify the sequence within the gap between the Gr-5 ORF sequence and the downstream Gr-6/Co. Ca 5/M0982 2.4 kb cDNA contig. Primers were designed from the 5 end of the Gr-6 clone in both sense and antisense orientations. Two sets of such primers were designed 200bp from the 5 end of the sequence we had and nested primers 100 bp from the end so we could verify we were amplifying a credible fragment by carrying out a second round of nested PCR. Similarly, primers were designed in the same way from the 730 bp of ORF sequence we had at the 5 end of the Gr-5 clone. Total RNA was made from the two pancreatic tumour cell lines which exhibited the highest expression levels of the ART1 transcript (BxPc3 and T3M4) and one cell line which exhibited the lowest expression (CF PAC1). Reverse-transcription was carried out using 20 μg of total RNA in the presence of 1.0 pmole of upstream and downstream primers, 4 units of RNAsin and 100 units of the MMLV enzyme from Life Technologies as described in the Materials and Methods chapter.

Three RT fragments were generated with three separate experiments to try and amplify the intervening gap sequence. The first experiment yielded a 2.8 kb fragment with the BxPc3 cell line RNA only. The fragment was cloned by T-tailed cloning described in the Materials and Methods chapter (Holton & Graham, 1990) and the ends were sequenced. The sequence generated was not homologous to the upstream or downstream cDNA sequence already known and the primer sequences could not be identified within the ends of the 2.8 kb fragment. It was reasoned that perhaps because no attempt to purify the random hexamers from the second strand cDNA reaction before PCR was carried out contaminating arbitrarily primed fragments may dominate the resultant PCR population of fragments.

A Marathon™ cDNA cloning kit was also purchased from Clontech which allowed the rapid amplification of both 5 or 3 ends of transcripts when only limited sequence was available. Basically, one generates a library of cDNA fragments from the tissue/cells of choice using cDNA synthesis from random hexamers. Once synthesised, the cDNA population of fragments have double-stranded adaptor primers ligated onto each end which serve as a template for the amplification of unknown sequences using an adaptor-specific primer (AP1) and a primer from the known sequence one already has for the gene of interest (see Materials and Methods). Employing the Gr-6 downstream sequence primer (anti-sense strand) with the adaptor primer 1 (AP1) resulted in bands at approximately 2.5 and 1.2 kb. The 2.5 kb band amplified in the BxPc3 and T3M4 cell line RNAs, though at a much lower yield in T3M4 track (Figure 24a). RNA from the cell line SW979 produced a barely visible band of the same molecular size, and all three cell line extracts produced a lower band at 1.2 kb, again exhibiting the same differential expression from high (BxPc3) to low (SW979). The control mRNA sample did not demonstrate any bands with these primers, however the TFR positive control primers did give a band with the control mRNA at 5 kb as prescribed.

The PCR products produced above were taken a stage further, and an aliquot of the BxPc3, T3M4 and SW979 reactions were first diluted 100-fold before 1ml was included as a template in a second round of PCR with the Gr-6 anti-sense primer and the adaptor primer 2 (AP2). This yielded only one band in the BxPc3 track only at approximately 1.5 kb with no lower band evident and no bands now appearing in the T3M4 or SW979 cell line RNA tracks. The 1.5 kb band was cloned into Bluescript™ using the the T-tailed method described in the Materials and Methods chapter (Holton & Graham, 1990). Sequence analysis this time confirmed we had homology to the downstream Gr-6 cDNA clone and the anti-sense primer sequence could be identified within the 1.5 kb fragment. There was no sequence similarity however with the upstream sequences (730 bp ORF) we had thus far identified.

An additional 800 bp of sequence extending from the existing 5 end (T7-end) of the 2.4 kb Gr-6/M0982 contig was generated which proved homologous to a previously isolated cDNA clone (c7-vii) which had been discarded due to lack of homology with the existing sequence. This cDNA clone and the 800 bp of sequence from the cloned RT product exhibited (39%) sequence similarity to a metastasis associated transcript (mta-1) isolated from hepatocellular carcinoma cell lines by differential display (Yasushi *et al.*. 1994). This sequence similarity between our RT-PCR generated fragment and the mta-1 gene extended over 300 bp of sequence with gaps however. The sequence in this region of the mta-1 gene was believed by the authors to exhibit homology to a protein kinase C domain, however they had also localised this gene to chromosome eleven.

Once the Lrp130 transcript homology was established, primers were designed from the cDNA Gr-6 sequence (homologous to Lrp130) and from the 5 end of the Lrp130 mRNA sequence in an attempt to amplify the remaining part of this transcript which had eluded us. This experiment yielded a corresponding fragment of 1.8 kb in the BxPc3 cell line RNA only (figure 24b) which represented the expected size of fragment. Cloning and sequence analysis of

the ends of this fragment confirmed it to be homologous to the Lrp130 sequence in the database, though there still remained some ambiguity at the 5 end.

Figure 24 : RT-PCR of ART1/Lrp130 cDNA utilizing Marathon cDNA synthesis and cloning kit

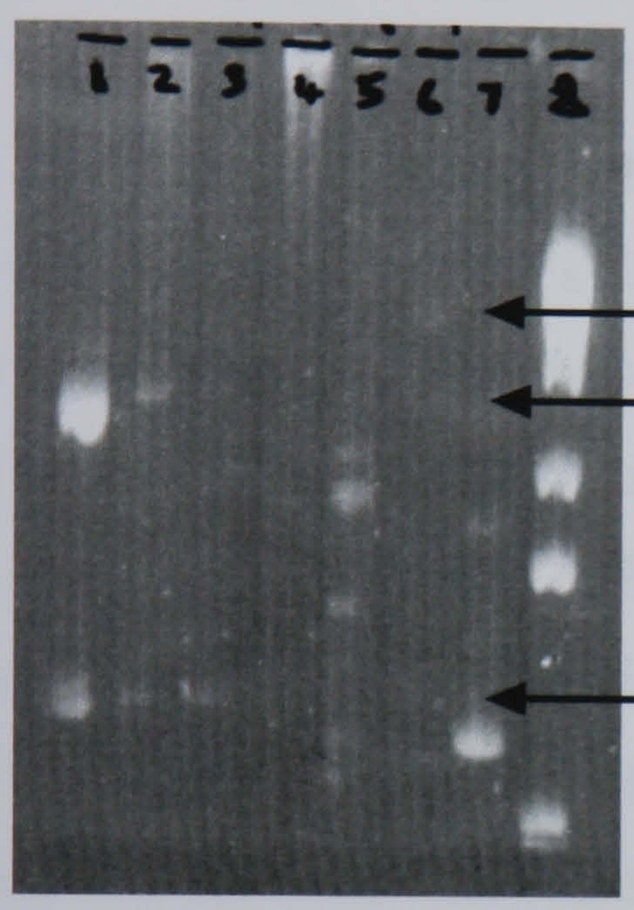


Figure 24a : RT-PCR gel illustrating the products formed after generating the cDNA library of randomly amplified fragments ligated to adaptor molecules generated from the Marathon cDNA cloning kit followed by PCR with an adaptor-primer and a ART1/Lrp130 gene-specific primer

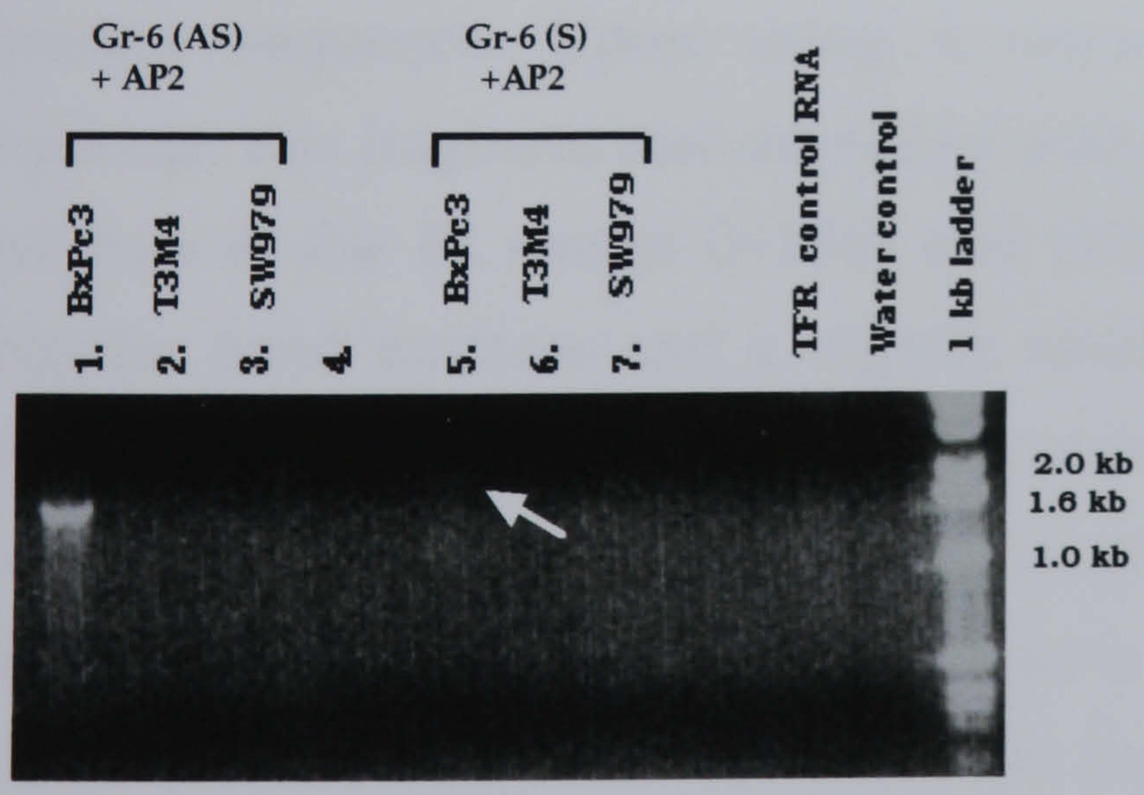


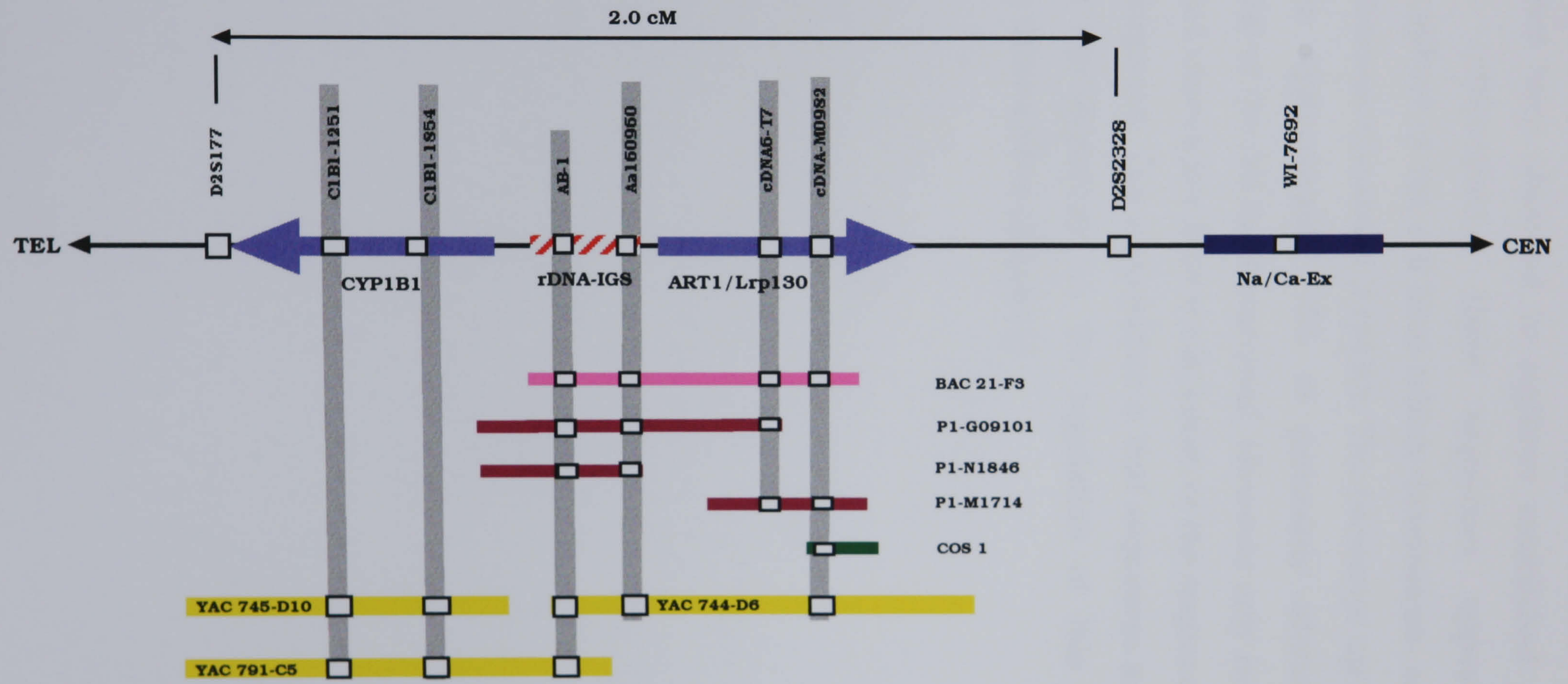
Figure 24b : RT-PCR gel illustrating the product formed after second rounds of PCR with the Gr-6 antisense and sense primers with the adaptor primer 2(AP2). Only the BxPc3 sample yielded a product with Gr-6/AP2 primers at approx. 1.5 kb. No lower molecular sized bands were evident, neither did any of the other samples yield a product following this secondary PCR.

3.4.10 Chromosome 2p22 physical map is refined

The position of the known transcripts on chromosome 2 band p22 were refined relative to the P1 bacteriophage, BAC and cosmid clones isolated in the laboratory with WC2.3 contig-specific STS amplification (Figure 25). STS markers generated from clones isolated in the laboratory were also positioned on the CEPH YACs from contig WC2.3. It was known from hybridisation experiments and PCR localisation that the large 2p22-specific BAC clone (B21-F3) contained both the AB1 sequences and an STS marker from the cDNA M0982 which amplified equally well on genomic DNA. An STS marker from the *CYP1B1* sequence did not amplify with DNA from the BAC or any of four P1 clones. It did however amplify in three of the contiguous CEPH-YACs from contig WC2.3, including YAC 745-D10, known to be positive for other *CYP1B1* markers. The 1.3 kb sequence-tag from the cDNA clone M0982 amplified cleanly in two of the P1 clones (M1714 and G09101) as well as B21-F3, a cosmid clone and the three original cDNAs (M0982, Gr-6 and Gr-5). The M0982-STS did not amplify with any of the YAC DNA templates. (details of sts sequences employed are give in appendix 5).

It has already been established that the AB1 sequence and subclones from two P1 clones (N1846 and G09101) are contiguous with the 3 end of the CYP1B1 genomic sequence. Upon using a sequence-tag encompassing the AB1 sequence, this fragment also amplified with DNA from the three CEPH-YACs, and two of the P1 clones (N1846 and G09101) as well as giving additional complex band patterns and a slightly smaller fragment with the BAC clone and a cosmid clone. A second STS was created from sequence 400 bp upstream to the AB1 fragment (STS-16096). This fragment amplifies in the same two P1 clones (N1846 and G09101), the BAC clone and the CEPH-YACs from contig WC2.3 (745-D10 and 791-C5). We could therefore assume the two genes are divergent from a point at the 5 end of the CYP1B1 gene and its overlap with the AB1/repetitive rRNA spacer. Additional CEPH-YACs were ordered from the HGMP which were contiguous with 745-D10 in the opposite direction, the sequence-tags 160960 and M0982 amplified with the YAC 744-D6.

Figure 25: Physical map of the 2p22 region with CYP1B1 and Lrp130 genes, localising the STSs onto the YAC, BAC, P1 and cosmid clones isolated



3.4.11 Summary of chromosome 2p22 findings

Sequences highlighted by the original AB1 Alumorph appear deleted in DNA from three primary tumour cases when compared to their constituent normal DNA on Southern blot analysis (16% of eighteen cases examined). Such deletion had not been observed in eighteen established pancreatic tumour cell lines. Alternatively, these sequences appear to be hypomethylated in a subset of the cell lines which demonstrate an elevated expression of a downstream transcript (Lrp130). The transcript up-regulation was demonstrated in approximately 15% of pancreatic adenocarcinoma tissues and 50% (12/24) of the cell lines examined. However only four of these cell lines. demonstrated observable hypomethylation of the sequences around the AB1 Alumorph fragment. The supposition is that sequences around this Alumorph fragment are important to the regulation of this transcript (possible mechanisms discussed in chapter 4).

3.5 Characterisation of the IP-1 Alumorph

Alumorph IP-1 was a 350 bp fragment which was observed to be lost in two pancreatic tumour DNA samples compared with their constituent normal DNA samples (figure 8a). The fragment was used as a probe in Southern hybridisations but did not detect any discrete bands originally, only repetitive smearing down the lanes despite extensive competition regimens. Sequence analysis revealed significant similarity to the human, mouse, rat and xenopus POU domain sequence. This sequence similarity prompted us to isolate two cosmid clones specific for this fragment, both of which were co-localised to chromosome 17p13 by FISH analysis in the ICRF Cytogenetics laboratory. Neither of the cosmids appeared to hybridise to the *TP53* cDNA sequence, though the gene was not ruled out for some time due to the lack of genomic clones for *TP53* at our disposal.

3.5.1 Interphase-FISH analysis

To refine the localisation of these two cosmids we collaborated with a group at the CRC Paterson Institute who had an interest in the small arm of chromosome 17 and expertise with interphase-FISH (Stack *et al.*, 1995). Triple-probe FISH and interphase-FISH carried out by them positioned our cosmids telomeric to the TP53 gene, but almost co-localising with the marker YNZ22 (D17S5) (Figure 26). The conclusion was drawn that the IP-1 cosmid used (1A) was so close to YNZ22 that it must be less than 1 Mb away.

Allelic loss of the short arm of chromosome 17 is among the most commonly recognised genetic alterations in pancreatic cancer well as breast, ovarian, bladder, lung, prostate and neuroectodermal cancers (discussed in chapter 4). In a small series of precisely microdissected tumours (n=14) carried out in this laboratory, there is evidence of up to 25% Loss of Heterozygosity (LOH) around the marker D17S28 (figure 27). With the convincing LOH data from Hahn et-al, and the localisation of IP-1 to 17p13.3 in the vicinity of D17S28 it was decided this locus merited further investigation.

Figure 26: Localisation of IP-1-specific cosmids using Fluorescence *In Situ* Hybridisation (FISH).

Figure 26a: Using fluorescence *in situ* hybridisation (FISH) on normal metaphase spreads, cosmid IA was mapped to chromosome band 17p13.

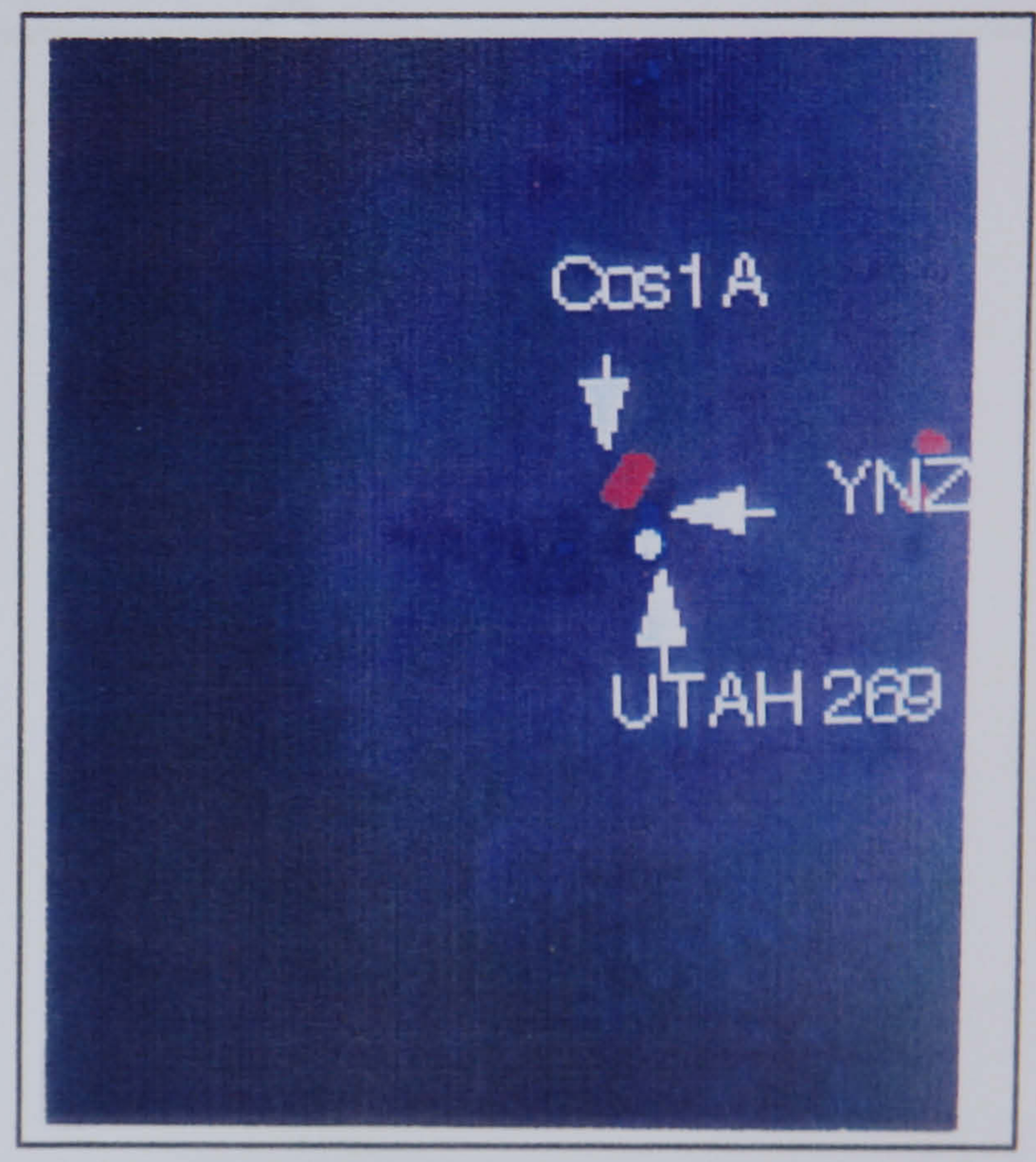
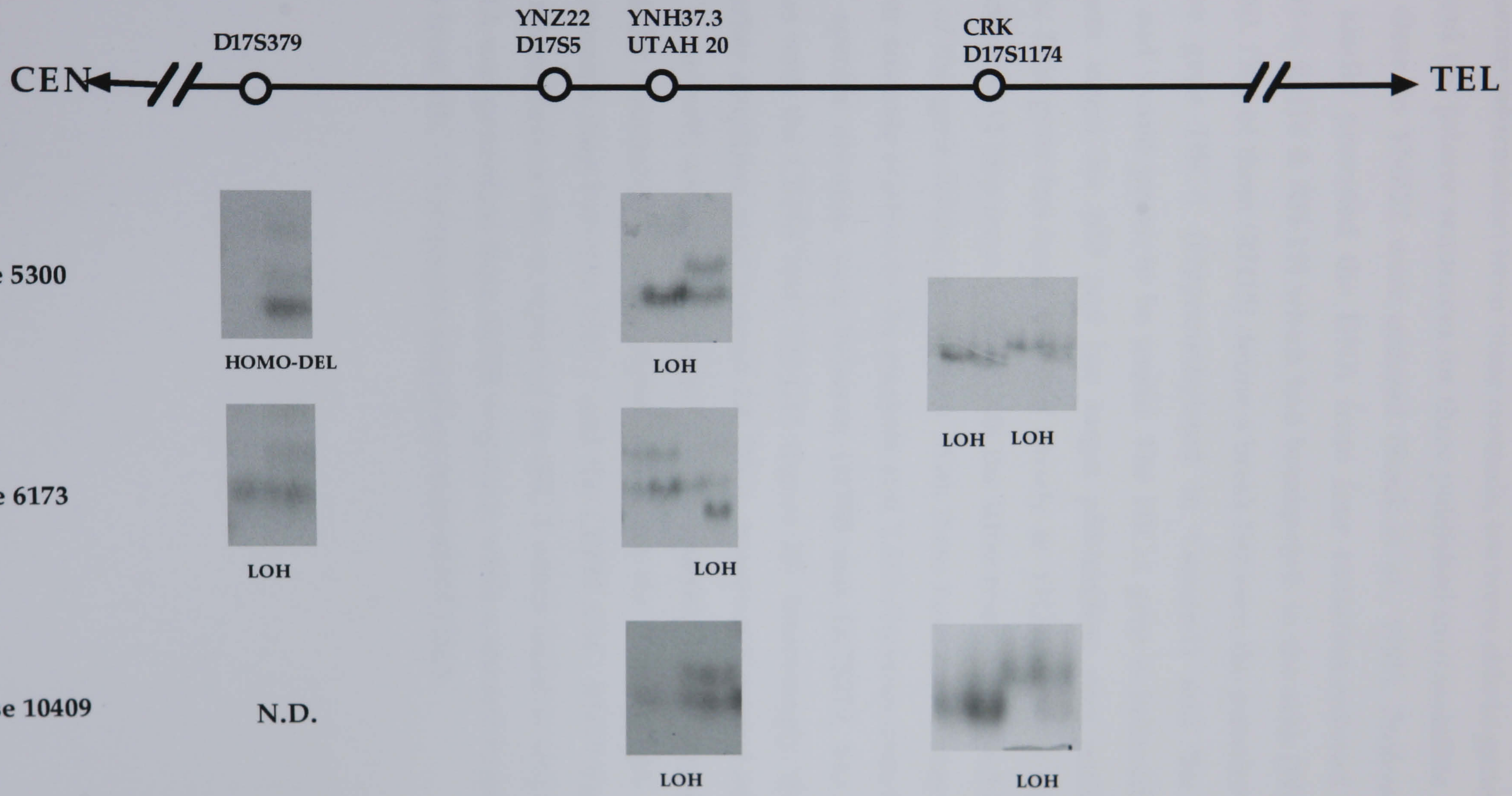


Figure 26b: Interphase FISH carried out with two YACs specific for D17S5 and Utah 269 markers and the IP-1-specific cosmid IA

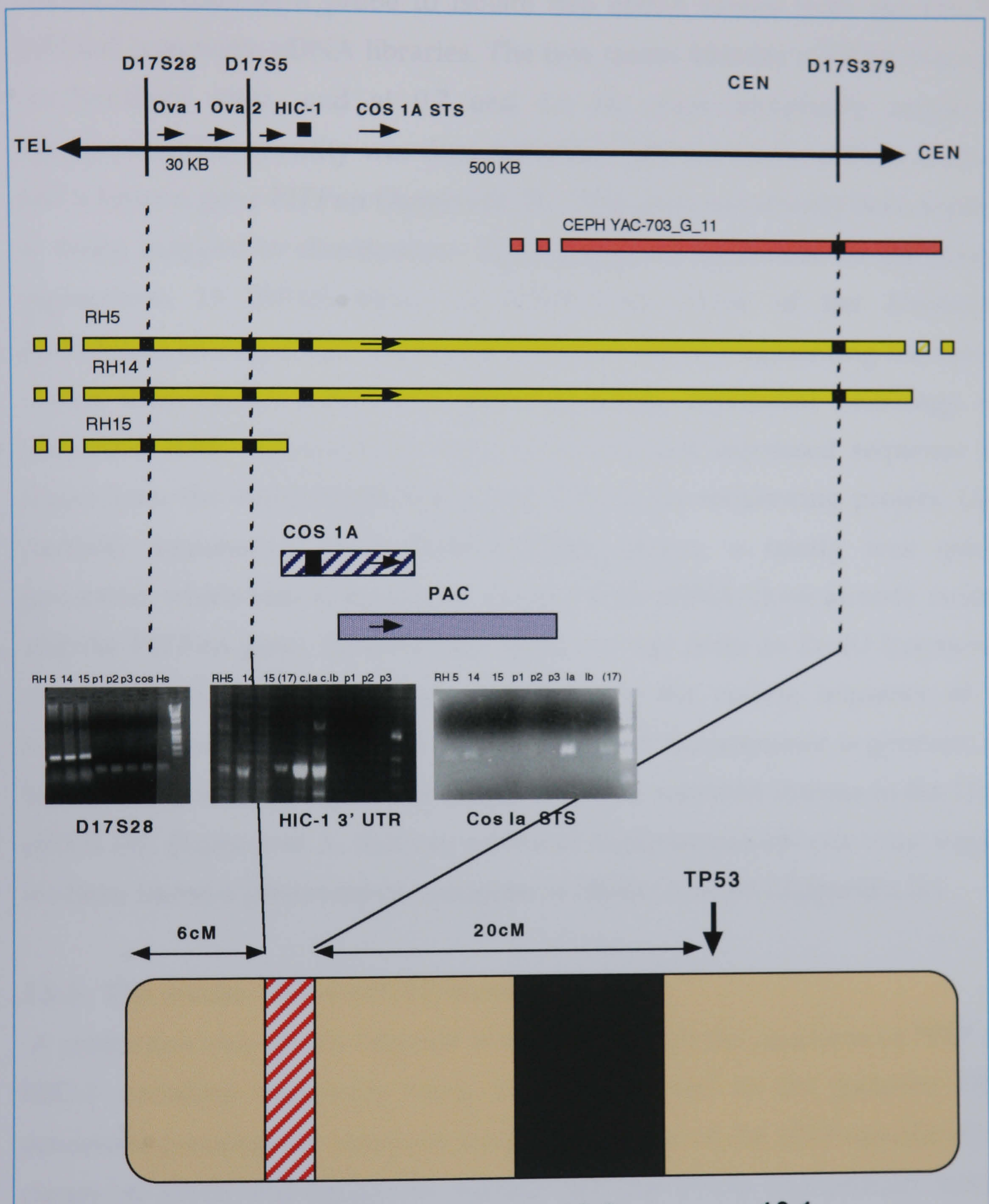
Figure 27: Loss of heterozygosity analysis on chromosome 17p13.3 in pancreatic adenocarcinoma



3.5.2 Positional information

With sequence information from these cosmids, we were able to generate STS markers and the primer sequences for three published microsatellite markers mapping close to YNZ22 were ordered (Stack *et al.*, 1995). Professor Ellen Solomon kindly provided the DNA from four radiation-reduced hybrids (RH-5, RH-6, RH14 & RH-15) which had breakpoints in the area (Wilgenbus *et al.*, 1996). One of them (RH15) define a break between the putative tumour suppressor gene HIC-1 (Hypomethylated in Cancer-1) and the marker D17S379, and would prove to be useful. The HIC-1 gene is believed to be a downstream target for p53 and has target p53-binding sites within its 5 sequences. This gene has been mapped closely to YNZ22, and is quoted as being as close as 11 kbp from YNZ22 with the latter residing in the 5 flanking sequence of the gene (Wales *et al.*, 1995). With these resources it was possible to map our cosmids relative to the markers and YACs/hybrids from the area. The IP-1 specific cosmids map between D17S5 and D17S379, but are not contiguous with the CEPH YAC 703-G11 (figure 28). Interestingly, the HIC-1 3UTR marker amplifies with cosmid 1A DNA, however 5 primer sequences designed by myself, amplify the correct sized fragment from chromosome 17 DNA and the 3 radiation-reduced hybrids, but not the IP-1 cosmids. Clearly, the IP-1 cosmids map between HIC-1 and the CEPH YAC, RH15 may prove useful to researchers as the sts reported for HIC-1 which failed to amplify with RH15 DNA was generated from 3UTR sequence, while a second sts generated in our lab from HIC-1 5 sequence amplified from RH15 DNA.

Figure 28: Physical map of region between D17S28 & D17S379 incorporating Cos1a and PAC



RH = Radiation hybrid CEPH-YAC

3.5.3 IMAGE cDNAs

A cosmid fragment hybridising to the original IP-1 alumorph was isolated, cloned and used as a probe to isolate two cDNA clones from the EPCRI.S gridded pancreatic cDNA libraries. The two clones identify a 2.0 kb transcript on Northern blots, and at 0.7 and 1.6 kb, were completely sequenced. Significant (97%) identity was demonstrated with one of the cDNAs (clone 7) and a known gene *PITPnα* (Appendix 2b). This gene has already been reported as being mapped to chromosome 17p13.3 and is a candidate for the retinitis pigmentosa 13 (RP13) locus on 17p13.3 by virtue of the *Drosophila* homologue having a role in eye development. Upon fragmenting the cosmid clones with *EcoRI*, subcloning and sequencing, significant homology was identified with one of these *EcoRI* subclones and expressed sequence tags (Ests) from the Washington University/Merck Est sequencing project. Upon partially sequencing these IMAGE-cDNA clones, a contig was quickly generated which was homologous to the 1.6 kb cDNA clone already isolated and the *PITPnα* gene. Additionally, sequence tags from an *EcoRI* fragment of cosmid IA exhibited significant homology to the coding sequence of the receptor tyrosine kinase HER4/*ERB B4*. The fact our sequence is genomic and the area of homology spans the position of two reported introns in the HER4 cDNA (H. Hurst and A. Skinner personal communication) one may suggest we have found a processed pseudogene of HER4/*ERB B4* (Appendix 2c)

3.5.4 The mouse vibrator/PITP locus

A restriction map of the cosmids is underway with the positions of PITP and HIC-1 encoding sequences being mapped as well as the putative HER4 processed pseudogene. Southern analysis with one of the PITP-specific cDNA clones on *EcoRI* -cut pancreatic tumour cell line DNAs and primary tumour and normal DNA pairs revealed a constant band at approx. 6.0 kb in all of the samples except tumour and normal DNA from two individuals (LaF and CrDn.) which exhibited a 3.0kb band instead (figure 29). Clearly, the fact normal DNA samples exhibit this shift rules out a tumour-specific deletion, a more likely explanation being a rare neutral polymorphism.

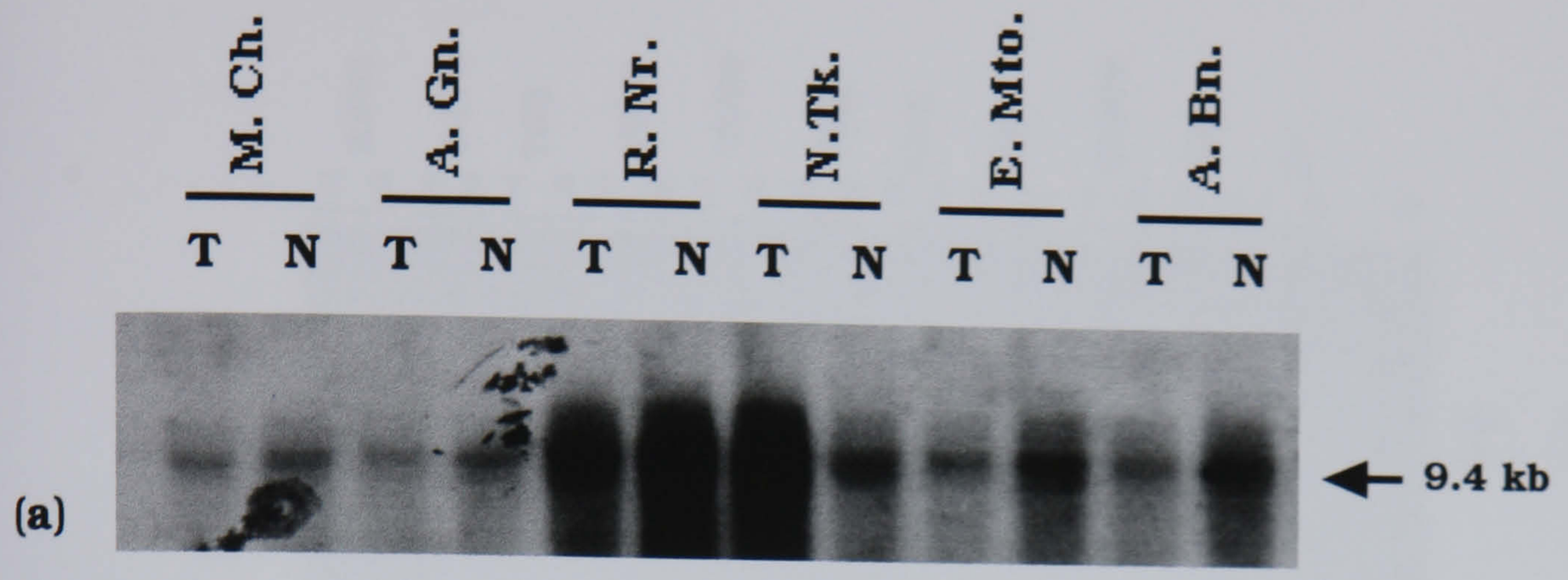
3.5.5 Multiplex-PCR reveals homozygously deleted area between D17S5 and D17S379

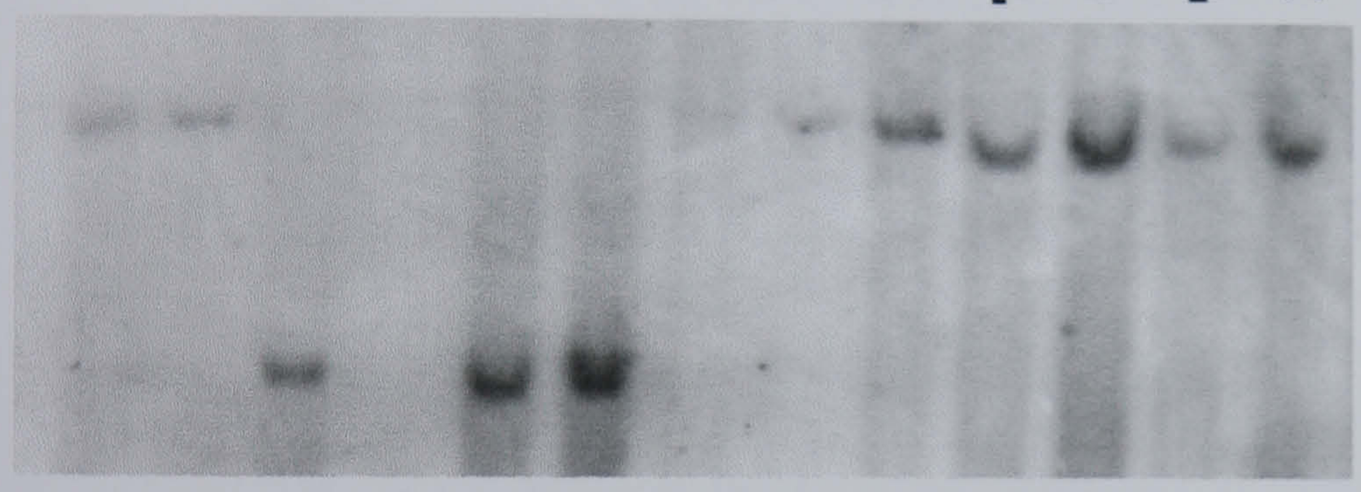
Previously, we had demonstrated LOH on 17p13.3 employing microsatellite markers around YNZ22 (D17S5). The marker Utah-20 and D17S28 both exhibited LOH (Figure 27) in three out of fourteen informative cases (21%). Additionally, two tumour samples from cases (1083 & Biggs) exhibited consistent failure to amplify with more than one marker in this region. The sample 1083 was however being amplified successfully in separate AP-PCR experiments ongoing in the laboratory (Vorapan Sirivatanauksorn) which rules out the possibility that these two samples were merely contaminated with something inhibitory to the Taq enzyme. Problems had been experienced in amplifying fragments from DNA extracted from paraffin-embedded archival material (particularly normal samples from tumour/normal pairs). It was possible that because thicker sections had been cut from the normal tissue material to speed up the process, the paraffin content could be causing inhibition.

It was thus decided to carry out multiplex-PCR reactions with the 17p13.3 markers and a GAPDH amplicon. Initially primers for exon 5 of the *TP53* were employed as *TP53* had not previously been found exhibit deletions or LOH in fifteen pancreatic tumour cell lines when used to screen for mutations (Barton *et al*, 1991).. The *GAPDH* primer pair were later preferred to *TP53*-specific primers. The relative intensity of control and test bands were compared visually with the products produced from constituent normal DNA samples. In addition to the three cases exhibiting LOH above, case 1083, 763 and Bggs (see figure30) exhibited homozygous deletion of the 17.13.3 -specific PCR fragments D17S379, 25HB6-RE YAC sts, and the 3UTR sts from HIC-1. Analysis with the D17S1174 marker and 38B8-RE YAC sts revealed no such loss (figure 30). We may therefore assume that we are observing homozygous deletion of an area including the markers D17S1174 and D17S379. approximately 500 kb in these three cases. Together, the LOH data and cases of homozygously deleted sequences would suggest we are observing

allelic imbalance with these markers in 6/14 pancreatic cancer cases which represents 43 %.

Figure 29: Southern and Northern analysis of the PITP cDNA clone7 in pancreatic cell lines and tumour/normal pairs





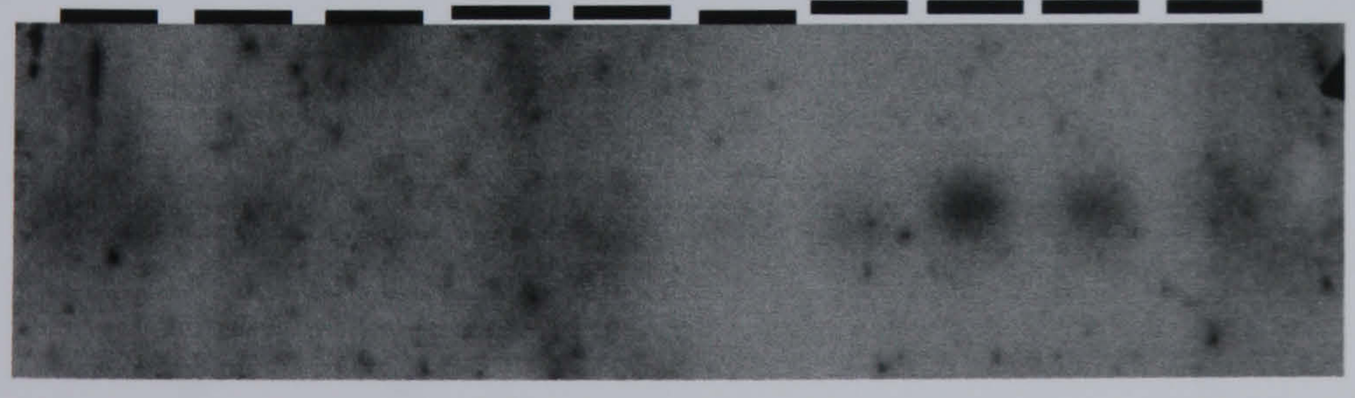
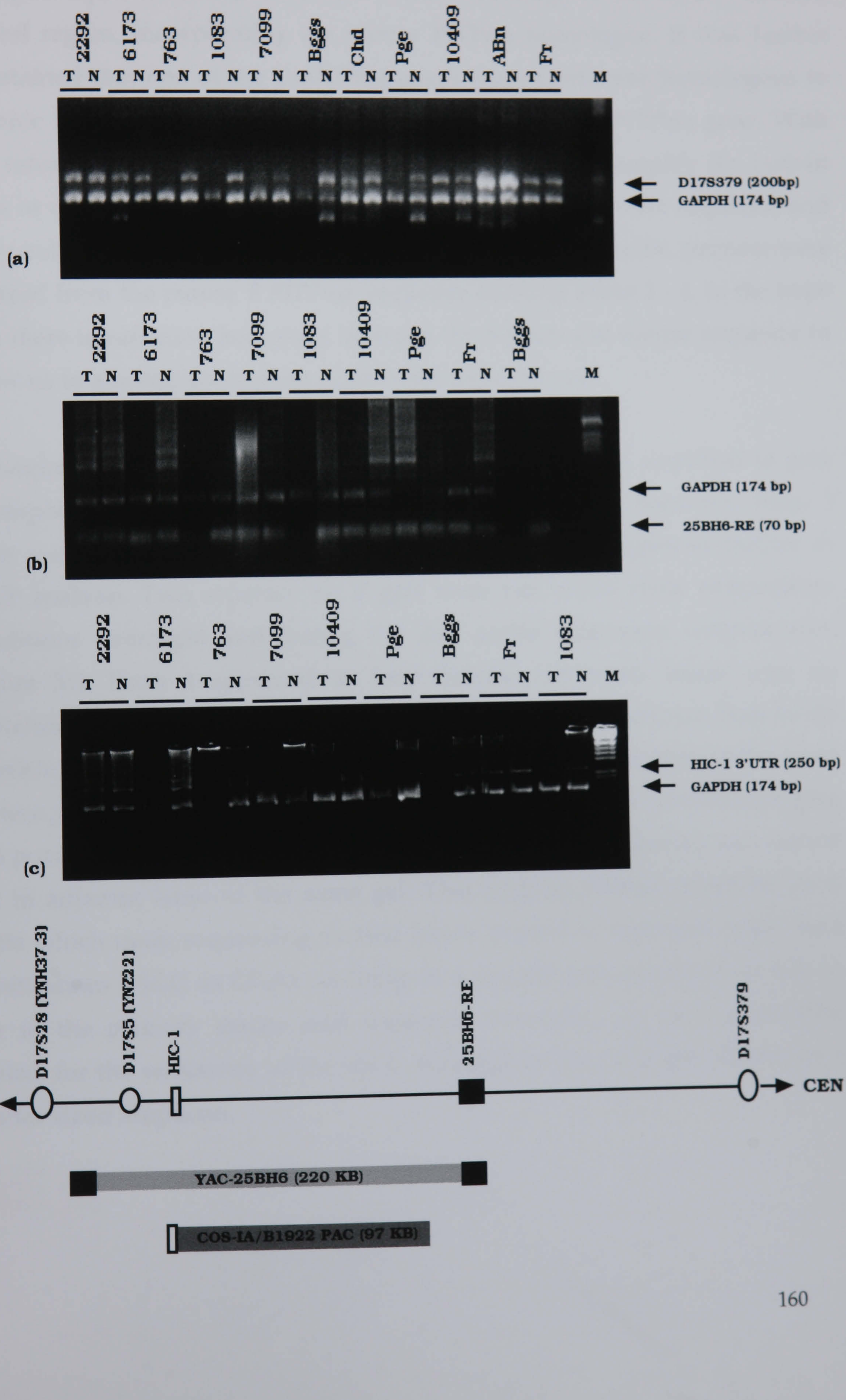


Figure 29: **(a)** A Hind III Southern blot with 6 tumour/normal pairs probed with the PITP cDNA clone 7. A 9.4 kb constant band hybridises exhibiting no evidence of rearrangement. **(b)** An Eco RI Southern with 7 pancreatic tumour cell lines and three tumour/normal pairs probed with the cDNA clone 7. Here, two of the pancreatic tumour/normal pairs exhibit a shift in the hybridising band mobility from 4.5 kb to 2.0 kb.(both tumour and normal lanes) suggesting the occurence of a neutral polymorphism. **(c)** Northern blot containing 10 pancreatic cell line RNAs exhibits varied expression of a 2.0 kb transcript with strongest signal in the 3 moderately to well differentiated cell lines AsPc1, Mia-PaCa2 and CaPa-2

Figure 30: Multiplex PCR analysis of 17p13.3 deletion area

Three examples of STS markers from this region used in analysis (a,b & c), with line diagram (d) depicting physical map and distances.



3.5.6 The mouse *Vibrator* critical region and SSCP analysis of *PITPnα*

As well as homology to the human *PITPnα* gene sequence the cDNA and genomic sequence we had exhibited similar homology to the mouse vibrator critical region, incorporating the mouse PITPnα homologue. It was further ascertained that one of the *EcoRI* fragments from CosIa was homologous to intronic sequences between exons 3 and 4 of the mouse *PITPnα* gene. With this information it was felt it was worthwhile to try and amplify the human gene or exons from it with primers designed from the mouse sequence and carry out SSCP analysis to look for *PITPnα* mutations. Specific, primers were ordered from the mouse *P PITPnα* sequence flanking exons 1 - 4, in the hope that there is sufficient homology between the human and mouse sequence to allow us to amplify the human exons with these primers.

Subsequently, only the primers specific for exons 3 and 4 amplified to give corresponding amplicon sizes predicted from the mouse sequence. Exon 3 gave a sufficiently clean single band which could be employed further in SSCP analysis. Two separate SSCP gels were run under room temperature conditions overnight and during the day under cold-room temperatures. (figure 31). Exon 4 appeared as four definite conformer bands with no deviation the band pattern in any of the eight cell lines analysed. Four to six possible conformers existed for exon 4, though again no alteration to the band pattern was evident. SSCP analysis of an amplified exonic fragment belonging to a putative tumour suppressor gene localised to chromosome 6q was carried out in adjacent lanes of the same gel. This analysis yielded observed band shifts which upon sequencing excised bands proved to represent single base substitutions (GGC to GGA) resulting in a neutral polymorphism of Gly to Gly in the primary amino acid sequence. Therefore, we had an internal control for the sensitivity of the SSCP included on the same gel which was a similar sized amplicon.

Figure 31: SSCP analysis of exons 2 and 3 of the PITP gene

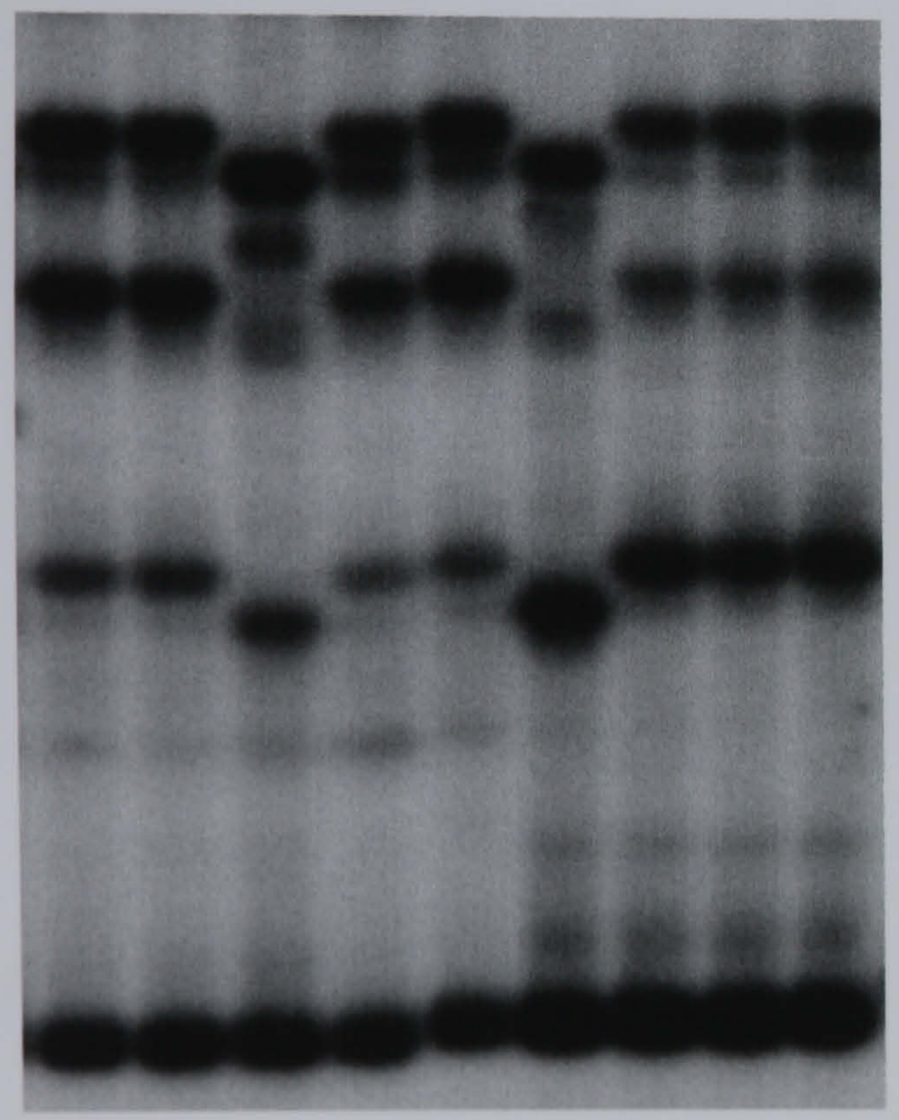


Figure 31A: Example of the conformational change induced by a neutral polymorphism (GGC - GGA) in exon 8 of novel ovarian cancer gene on 6q27, illustrating the sensitivity of the SSCP run in the laboratory

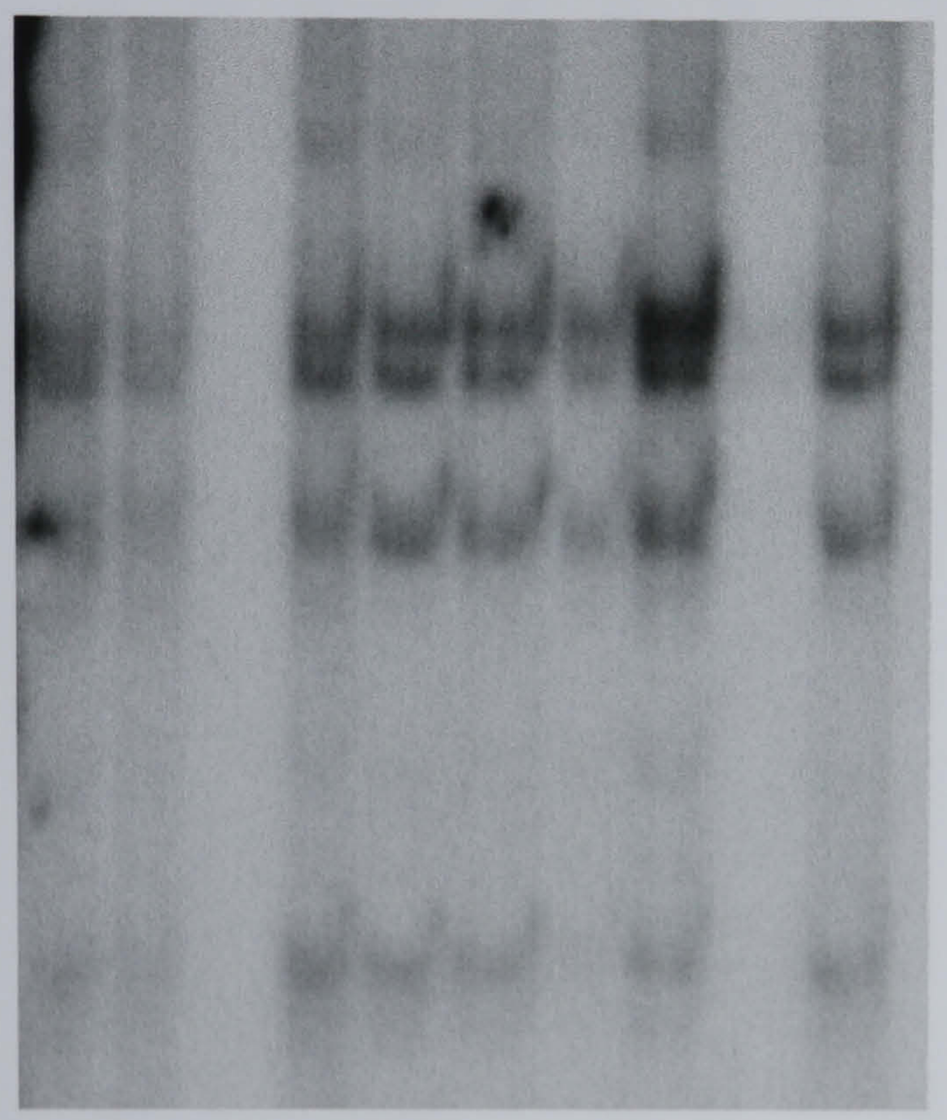


Figure 31B: Exon 3 of PITP amplified from pancreatic tumour cell lines and electrophoresed on the same gel as samples in panel **31A,** showing no signs of conformational polymorphism

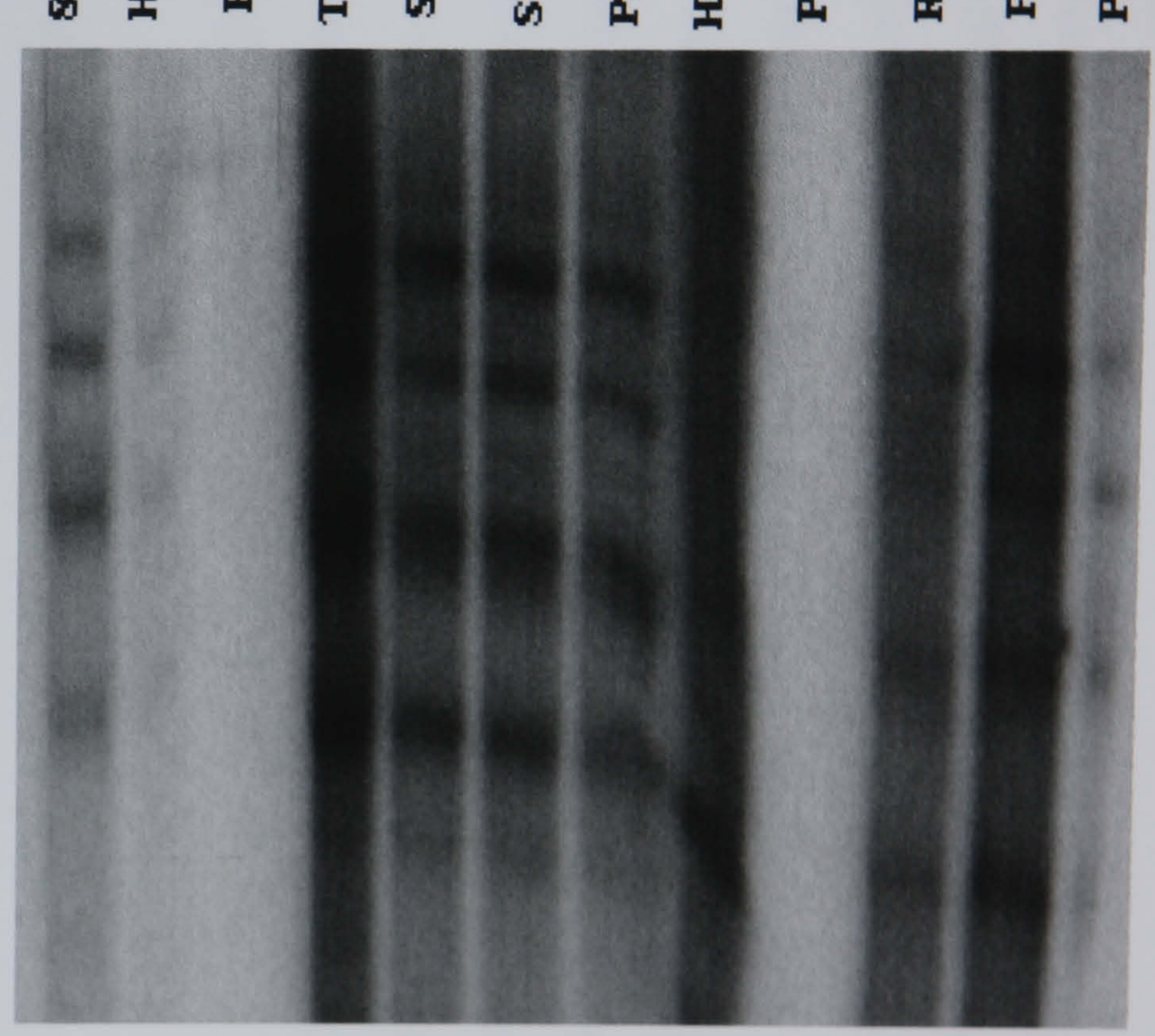


Figure 31C: Exon 2 of PITP amplified from pancreatic tumour cell lines showing no signs of conformational polymorphism, The distortion of band pattern with sample Hs700T was due to the gel stretching when taking the plates apart.

3.6 Analysis of Alumorphs from SUIT2 metastatic cell lines

3.6.1 Met6 Alumorph

A 500 bp fragment termed Met6 (the sixth Alumorph identified from a series of cell lines which differed in metastatic potential (Taniguchi *et al.*, 1992)) had originally been identified as being present in the parent cell line (SUIT2), lost in the two most metastatic sub-lines (S2-007 and S2-013), but also present in the remaining four sub-lines. The sub-lines S2-007 and S2-013 did however have a larger band at approximately 600 bp which was only present in these two lines (figure 32a). Both bands were excised, cloned and sequenced. Neither fragment exhibited any interesting sequence homologies, only similarity to Alu-Sx and LINE/L1 repeat elements and I.M.A.G.E cDNA Ests reported to contain Alu sequence also.

3.6.2 Southern and Northern with Met6

When used as a radiolabelled probe, the Met6 fragment hybridised to six to eight bands on a Southern blot of *EcoRI* digested DNA samples from paired tumour and normal tissues (Figure 32b). The probe not only hybridised to numerous bands, there was also a background smear evident down each track, suggesting the presence of repetitive sequences. Subtle polymorphisms were evident between two bands at approximately 1.6 kb and 1.8 kb, but no losses were observed between tumour and normal tracks of the same cases. There were probably other polymorphic bands masked by the repetitive smearing, despite stringent washes at 0.1x SSC/0.1%SDS at 65°C. The larger (Met7) fragment also hybridised to an *EcoRI* filter giving faint bands masked by strong background repetitive smears, however, we did not observe the same discrete sharp bands or any evidence of polymorphism, therefore the fragment was not pursued. Met6 hybridised to a multiple-tissue Northern (Clontech) revealed the presence of a large 8-10 kb transcript in two of the tissues with additional heavy smearing on each lane down the length of the filter.

Figure 32: Metastasis-specific Alumorph Met6 fingerprint and Southern blot

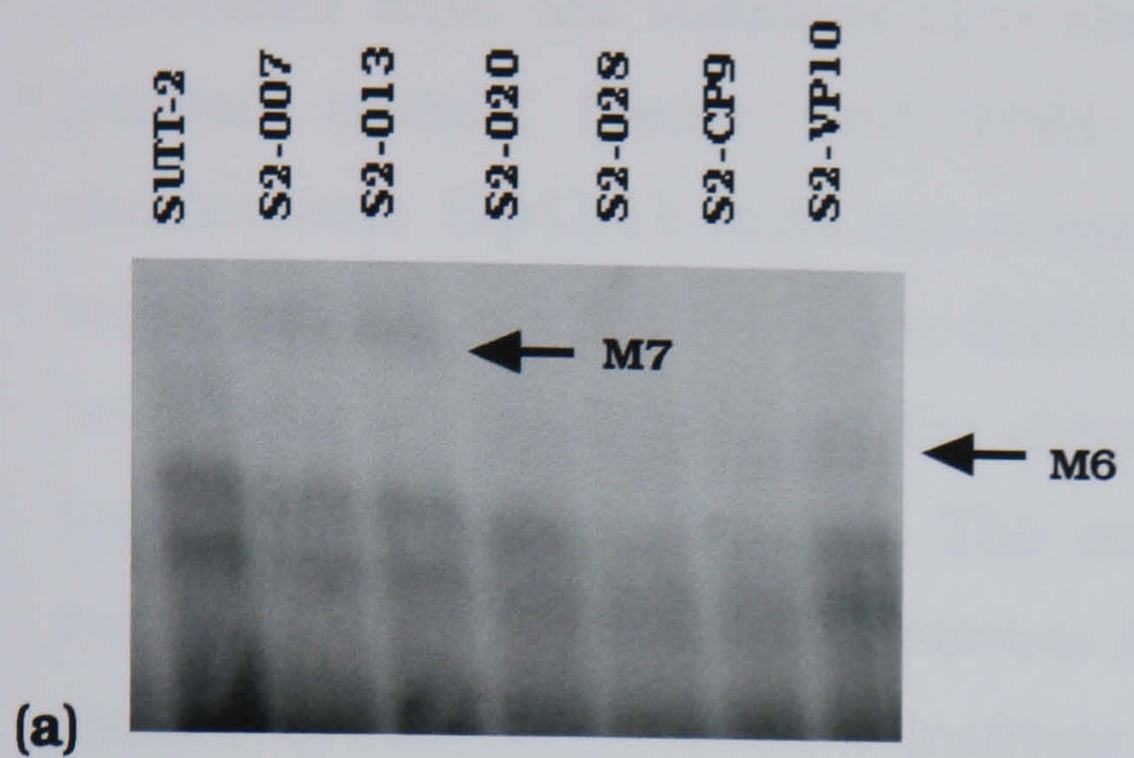


Figure 32(a): The 500 and 600 bp (Met6 & Met7) fragments lost/gained in S2-007 and S2-013 sub-lines respectively.

(b): Hybridisation with the Met6 fragment as a probe on an *Eco RI* Southern. At least six bands (from 0.75-3.0 kb) hybridise to the fragment with a limited amount of polymorphism evident

3.6.3 Met6 localised to the short arm of chromosome 11

Southern analysis of the ICRF somatic cell hybrid panel (cut with *Hind III*) localised this fragment to chromosome 11p, a result which was later confirmed with the isolation of a clone from a chromosome 11-enriched gridded cosmid library and FISH analysis, mapping the cosmid to chromosome 11p12-13 (data not shown). The FISH result was reported as being difficult to interpret by the ICRF Cytogenetics laboratory, and that there were weaker signals all over the chromosome 11. The strongest signal did localise to 11p12-13 however. The cosmid (G08878) had previously been requested and localised by numerous laboratories. Groups within the ICRF, the ICR at Fulham Rd and the Institute of Child Health (ICH) had placed the cosmid on the long arm of chromosome 11 at band q23. Equally, researchers at the MRC and CRC laboratories in Edinburgh and Cambridge respectively have placed this cosmid on the short arm of chromosome 11 at band p13-15. On contacting both the ICRF and MRC laboratories involved, it was established that this cosmid is believed to be particularly rich in repeat sequences which is the contributory factor in the confusion over its location. In addition, the cosmid is believed to contain zinc-finger sequence motifs which may also compound the confusion as chromosome 11, (particularly the q arm) is very rich in zinc finger containing genes. Researchers at the MRC Human Genetics Unit (HGU) believe the cosmid to reside within the Beckwith-Weidermann Syndrome (BWS) region which also ofcourse harbours the Wilms Tumour-suppressor gene WT-1. Two further cosmids have since been isolated with the Met6 Alumorph as a probe (G19798 and H01210).

3.6.4 Met-L and Met-U Alumorphs

Two further metastatic cell line-specific Alumorphs, Met-L and Met-U, were isolated as fragments which exhibited loss in some of the SUIT-2 derived cell lines after digestion with *HinfI* and amplification with the BK-33 primer (figure 8c). Met-L was a fragment of 235 bp which appeared lost in the two SUIT2-derived cell lines S2-CP9 and S2-VP10. These cell lines had been

isolated from lung metastases in mice after xenograft implants of SUIT2 cells, (see Materials & Methods). Sequence analysis, revealed no significant similarity to sequences in the EMBL/GenBank databases. A P1 clone was isolated and was used to demonstrate localisation to chromosome bands 3p21 and 8q24 by FISH (data not obtained) which could not be refined further. We hybridised a Met-L specific P1 fragment to a Northern blot of total RNA form the SUIT-2 derived cell lines and observed a 3.0 kb transcript that exhibited differential expression (figure 14b). High expression of the transcript was observed in the SUIT-2 parent cell line and in three clonal derivatives S2-007, S2-013 and S2-020 which show a poorly differentiated phenotype and are moderately to highly metastatic (S2-007>S2-020). Clearly, this fragment may represent an interesting gene which seems to be expressed in the more moderately metastatic cell lines, whereas loss of expression correlating with the deletion of this fragment are seen in the two cell lines derived from lung metastases.

The Alumorph fragment Met-U was approximately 295 bp and had no similarity to sequences in the EMBL/GenBank databases.

3.7 Isolation of Alumorph AB4

Alumorph AB4 was also identified as a fragment (approximately 700 bp) lost in the tumour fingerprint profile compared with its paired normal sample. Interestingly, there appeared to be an additional band in the same tumour track migrating at approximately 650 bp which did not appear in the adjacent tumour or normal tracks. Believing this smaller fragment to be a truncated product of the larger 700 bp fragment, both this 650 bp product and the 700 bp fragment (from the constituent normal lane) were excised and attempts were made to amplify the two. The 700 bp fragment amplified using the same Alu-specific primer (Cotter IV) and conditions which generated it, however the 650 bp fragment did not. Further attempts were made to amplify the 650 bp putative truncation product using the smaller Alu primer BK-33 proved

unsuccessful. Additionally, a random 15mer was employed in an improvised AP-PCR approach to try and amplify the 650 bp band, but this too proved unsuccessful with the production of a much smaller smeared fragment at approximately 150 bp.

3.7.1 Southern and Northern analysis

The AB4 fragment is the largest to be isolated thus far and has been used directly as a probe in Southern and Northern hybridisations. AB4 identifies three *Hind III* bands on Southern hybridisation (figure 33a), but no losses between tumour and normal lanes are observed. A transcript band of 3.0 kb is seen on Northern filters of the SUIT-2 derived cell lines exhibiting differential expression (figure 14c). High expression of the transcript was observed in the SUIT-2 clonal derivative S2-028 which has a papillotubular phenotype with low metastatic potential. In comparison, the parental cell line SUIT-2 and its derivatives S2-007, S2-013 and S2-020 with higher metastatic properties all show the same band, though much reduced in intensity. The fragment was then hybridised to a second Northern containing total RNA from ten pancreatic tumour cell lines we had in the laboratory. Following a three day exposure to X-ray film (figure 33b), only two of the ten cell lines have shown to express the 3.0 kb transcript signal (PaCa3 and T3M4). Exposure of a second film for ten days (figure 33c) showed much weaker signals from the cell lines PaTu2, PaCa44, PaCa3, T3M4, PSN-1 at an additional higher molecular size (4 to 5 kb) and in Panc1 a slightly larger, much stronger hybridising band (around 5 kb+). This filter had been stringently washed down to 0.5x SSC, 0.1% SDS at 65 ^{0}C, therefore either we were observing some splice variation between cell lines, or close sequence similarity between other transcripts, perhaps belonging to a family of genes was being observed.

Figure 33: Southern and Northern analysis of the AB4 Alumorph fragment

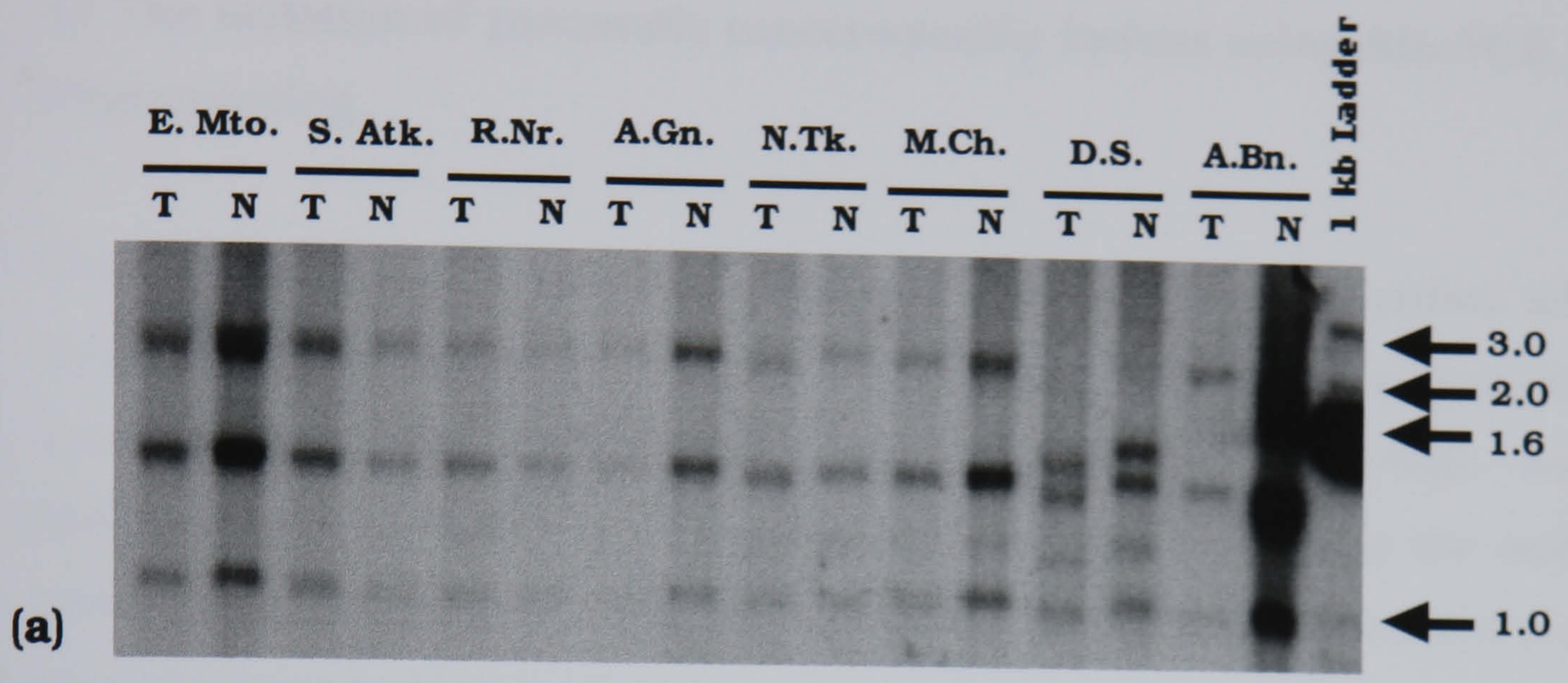


Figure 33 (a): Hind III Southern blot hybridised with the 700 bp AB4 Alumorph fragment as a probe - identifying three bands at approximately 1.0 kb, 1.4 kb and 2.2 kb. Case D.S. clearly identifies a polymorphism where the 2.2 kb band is missing and bands of approximately 1.25 and 1.55 kb appear. There is no evidence of loss between the tumour and normal lanes however. The last (A.Bn. Normal lane) is grossly overloaded leading to a slight difference in mobility.

Figure 33 (b & c): illustrate 3 and 10 days exposures respectively, from the hybridisation of the same 700 bp fragment to a Northern blot. As well as a 3.0 kb transcript band appearing in the PaCa3 and T3M4 lines, a higher band is evident after the 10 day exposure, particularly in the Panc1 lane.

Chapter four: Discussion

4.1 The isolation of pancreatic cancer-specific lesions using Alu-PCR genomic fingerprinting

The Alu-specific PCR genomic fingerprinting technique described allows the scanning of the human genome for genetic rearrangements associated with a cancer phenotype. With the generation of sequence-tags for these rearrangements, their use as molecular probes may facilitate the isolation of gene sequences involved. Further analysis of such genes associated with these rearrangements may lead to an understanding of the mutational mechanism and its contribution to the molecular pathology of pancreatic and perhaps other cancers.

4.1.1 Alu-PCR - analysis of fingerprint profiles

The purpose of this project was to use the Alu-PCR fingerprinting technique developed in the laboratory to identify tumour-specific lesions in pancreatic adenocarcinoma. The highly ubiquitous distribution of the Alu repeat means the coverage of the genome afforded by Alu sequence primed amplification could be potentially high. Combinations of single primers homologous to different regions of the Alu element and different restriction endonucleases employed to digest the genomic DNA can vary the complexity of the fingerprint and coverage of the genome. The fractionation of the template DNA by restriction digestion in this approach is analogous to the preparation of "representations" of the tester DNA necessary in RDA. In both approaches, it is designed to simplify the analysis of paired genomes by taking defined subsets for comparison. In table 6 it can be seen that the initial screen was a total of (approximately) 397 bands which amplified consistently from various samples of normal and tumour DNA. Fifteen of those bands showed differences between tumour and constituent normal DNA samples and some

of the bands were aberrant in more than one case (e.g., Alumorph IP-1). Hence, abnormalities were detected in approximately 4% of the genomic fragments that were sampled in this analysis.

Clearly, only a fraction of the genome is sampled in each of the Alu primer/restriction digest combinations and some proved more informative than others (e.g. BK-33/*Alu I* combination). Frequent-cutting restriction enzymes with four base recognition sequences such as *AluI, HinfI, MspI* and *RsaI* gave optimum fingerprint profiles which did not appear to correlate with how many recognition sites were within the Alu repeat consensus sequence. There appeared to be no correlation with the restriction enzyme cleaved genomic DNA and fingerprint pattern complexity. Looking at the genomic sequences of three genes as examples (TP53, Lrp130 and APC) both *AluI* and *Hinf I* cleave genomic DNA four to eight times more frequently than *RsaI*, yet produce a more informative fingerprint profile. The enzymes *Sau 3A* and *RsaI* have very similar four base recognition sequences containing all four bases and they exhibited similarly poor fingerprint profiles with few low molecular sized bands evident. The restriction enzyme *AluI* similarly recognises a four base sequence with all four bases present, yet here we observe highly informative fingerprint profiles with most of the Alu-specific primers employed. The distribution of recognition sites within gene sequences could be an important factor. Restriction enzymes with only G and C bases in their recognition sequence appeared to give less informative profiles compared with *AluI* and *Hinf I*. The Alu repeat element is GC-rich (approx. 60%), thus will be cleaved to a higher degree with restriction enzymes recognising GC-rich sequences.

Table 6: Numbers and size-range of bands generated

res. enzy. / primer	Alu I	Hinf I	Hae III	Rsa I	Hinc II
BK33	**43 bands 100-750 bp**	**9 bands 400-1500 bp**	**no bands**	**7 bands 300-600 bp**	**36 bands 400-1,300**
Cotter IV	**52 bands 250-1400 bp**	**17 bands 600-1500 bp**	**3 bands 250-400 bp**	**10 bands 600-1400 bp**	**18 bands 600-1,500**
Alu V	**31 bands 100-900 bp**	**26 bands 300-1,000**	**ND**	**14 bands 300-600 bp**	**12 bands 600-1,500**
BK34	**18 bands 300-600 bp**	**ND**	**ND**	**ND**	**ND**
517	**37 bands 100-600 bp**	**ND**	**ND**	**ND**	**ND**
TC 65	**2 bands**	**ND**	**ND**	**ND**	**62 bands 200-800**

ND: Not Determined

res. enzy. = restriction enzyme

Primers specific for the less GC-rich 3′ tail of the Alu repeat element may still be able to anneal to such fragmented templates providing the inter-Alu sequences were also relatively AT-rich. If we look at table 6 although the numbers are small, it is clear that the Cotter IV primer (bases 260-282 Alu consensus sequence) appears more successful at providing informative fingerprint profiles than the BK-33 primer (bases 36-17 Alu consensus sequence) with the enzymes *Hae III* and *Msp I*. Employing the Cotter IV primer in preference to BK-33 could mean the difference between bands and no bands. The Cotter IV primer also results in a larger spectrum of fragment size which might also reflect the influence in GC content of adjacent sequences. One would also expect these restriction enzymes to cleave genomic DNA more frequently at the 5′ end of genes associated with CpG islands. It has been calculated that there are approximately 45,000 CpG islands in the human genome (Antequera & Bird, 1993a) and that approximately 56% of these may be associated with genes. Therefore if a very rough estimate of average gene size is made of 10-15 kb, and assume there is no overlap between gene sequences, then about 25-35% of the genome would be occupied by genes (Strachan & Read, 1996). Cleaving genomic DNA predominantly at CpG islands means at least 12-17% of the genome is being analysed, however, since its also probably knocking out the majority of Alu priming, enzymes with AT-rich recognition sites will be more informative.

Localisation of eight of these sixteen isolated Alumorphs has indicated that a wide distribution of fragments are being generated in this analysis. The human genome has potential Alu-containing "hot spots" such as the HLA region on the short arm of chromosome 6 where one would suspect preferential amplification. In fact we have as yet not observed any duplication of fragments localised to the same chromosome arm. The sixteen fragments isolated represented 4% of the total number of fragments amplified, though this is probably an overestimate as some bands will have been overlooked. If an average band size (S) is taken for each spectrum of band sizes obtained in each fingerprint then we multiply this by the number of bands observed (N), a

very rough estimate of the total number of base-pairs generated for each profile will be generated (T). Employing this very crude calculation it is estimated that 252 kb of DNA was amplified by the Alu-PCR which represents 0.008% of the haploid genome.

4.1.2 Analysis of alumorph sequences

In the sequence analysis of some of the isolated Alumorph fragments, it was apparent that incomplete Alu-like sequence elements had been serving as target sequences for PCR. Not only more degenerate Alu sequences, but L1 sequences and probably other families of repeat elements (for example MERs) had served as templates for initial primer extension and PCR. This makes Alu more accessible as a target for PCR amplification in the context employed here, but would make it less attractive as a technique to amplify a specific stretch of the genome, knowing a particular gene(s) resided there. It must be remembered however that we are employing shorter Alu-specific primers than would normally be used to generate fingerprints of YAC clones for example. One would expect some of the primers to be human-specific however, if we look at table 7 comparing the sequence of five well documented Alu primers including BK-33, it can be observed that BK-33 shares shares 100% homology over its length with Alu-5′. This primer does have 3 additional nucleotides at the 5′ end which may confer additional specificity.

The generation of fingerprint data from YACs to compile overlapping contigs does however rely on a degree of specificity in that the same band pattern may assign a group of YACs to the same contig or not depending on slight differences in template annealing. The question of reproducibility would be an issue here as well. These studies may foster a certain degree of scepticism regarding contig mapping data generated from fingerprint analysis in the future. Sequence analysis of the alumorph fragments and any adjacent sequence revealed only fragments AB2, AB3, Met-L and V1 exhibited any significant sequence homology to the Alu consensus element (figure below).

Additionally, only with fragment AB3 could we align the Alu-specific primer which generated it

Alumorph AB3 and Met-L sequences with Alu homology underlined and Cotter-IV primer sequence in italics

AB3

AACTTTTCTT TTTTTTTTTT GAGACAGAGT CTTGCTCTGT CGCCCCAGGC

TGGAGTGCAG TGGCGCGATC TCGGCTCACT GCCACCTCAC TGCCACCTCT

GCCTCCCGGT TCACGCCATT CTCTCCTGCC TCAGCCTCCC CGAGTAGCTG

GCACTACAGG CGCCCCGCCA CCACGCCCGG CTAATTTTTT TGTATTTTTT

TAGTAGAGAT GGGGGTTTCA CCGTGTTCAC CAGGATGGTC TCGATCTCCT

GACCTCATGA TCCGCCCGCC TTGGCCTCCC AAAAT*GCTGG GATTACACGT*

GTG AGGC AAG AACTTTTTAA AAGTGCATCT TGCGCACGGG TAGATCCAT

AAACC

Met-L

CTGGGATTAC AGGCGTGACG GCTGTGCCCG GCTGCACAAA CACTTTCAGA

AAGCAACAGA GGACGAAACA TTTTCTAACT CATTTATAAG CCAGCATAAC

CTTGATCCAA AGCCTAAGGA GATTAAATAA AATTGAGGAC CGAAACCTAG

TCCAATACGG ACCGAATTAT TTACTCAATC TTTTACAAAG CANGAGACAA

CGAAGACTTT CACAAACACG CTGGCCGTTC GCCAG

Characterization of some of the alumorphs was taken further in some cases compared to most, particularly the AB1 alumorph fragment discussed below.

4.2 The characterisation of the chromosome 2 lesion

4.2.1 Alumorph AB1 exhibits loss in three tumour cases

The first Alumorph fragment (AB1) which demonstrated hybridisation to unique sequences on Southern blotting also exhibited a tumour-specific loss of such sequences. Three bands hybridising to the AB1 fragment on an *Msp I* Southern blot corresponded to three fragments generated by cleaving AB1 with *Msp I*. The two larger fragments at 75 and 200 bp appear lost in two tumour samples compared to their constituent normal sample which were in addition to the original loss demonstrated with the Alu-PCR fingerprint. The AB1 fragment was subsequently localised to the short arm of chromosome 2 band p22 with the isolation of P1 bacteriophage clones and a single YAC clone from the RLDB gridded libraries created at the ICRF.

This 2p22 localisation would present a daunting challenge since at this time, few genes and disease loci were mapped to the short arm of the second largest chromosome. The q arm is reported as being more gene-rich with many ion-channel and signal transducing genes localised there. The short arm of chromosome 2 represents ninety three million base-pairs of DNA which amounts to 4% of the total genome (Strachan 1996). The significance of this is realised when one considers the size of the entire chromosome 17 for example, at eighty million base pairs chromosome 17 only represents 2.7% of the genome. As discussed later, the resources available on chromosome 17 are more abundant than exists for some of the larger chromosomes.

Characterisation of sequences around the original alumorph AB1 were finally achieved through shotgun cloning of the AB1-specific P1 clones as well as vectorette PCR and identification of IMAGE cDNAs. The AB1 sequence resides 135 bp upstream of the putative start codon of the 2p22 transcript, overlapping a putative CpG island clone isolated from the CpG library distributed by the MRC-HGMP Centre. The sequence between the AB1 fragment and the transcript start codon is similar to Alu repetitive elements (Alu Sg/Sx),

similarly 300 bp further upstream is a second area of tighter Alu Sg/Sx homology (bases -544 to -776). An area of Line L1 similarity extends from bases -986 to -1250 upstream from the putative start codon (appendix 3). This sequence information may present us with a suitable mechanism whereby direct repeats of the Alu sequence used to generate the fingerprints may have created a deleted PCR product by annealing to target sequence which is repeated in a tandem fashion. Such a mechanism was proposed by investigators in 1994 to explain a 7.8 kb deletion generated in PCR assays (Weizhen et al., 1994). It was further realised that this Alu-directed deletion occurred somatically in the low density lipoprotein receptor (LDLR) gene the group were studying. Evidence to support such a mechanism being responsible for the observed deletion here suggested by the presence of a complete copy of the BK-33 primer sequence within the Alu Sg/Sx element(-620 to -598) and an incomplete homology to the BK-33 sequence within the 3' end of the AB1 sequence (-158 to -142). Attempts to PCR amplify sequences employing primers derived from this sequence has always resulted in a heterogenous set of fragments, often with bands missing from the both tumour and normal samples (figure 35). Additionally, band(s) at approximately 500 bp (or larger) is often apparent in the normal lane of cases (3) and often with accompanying bands in the tumour lane as well (2 cases).

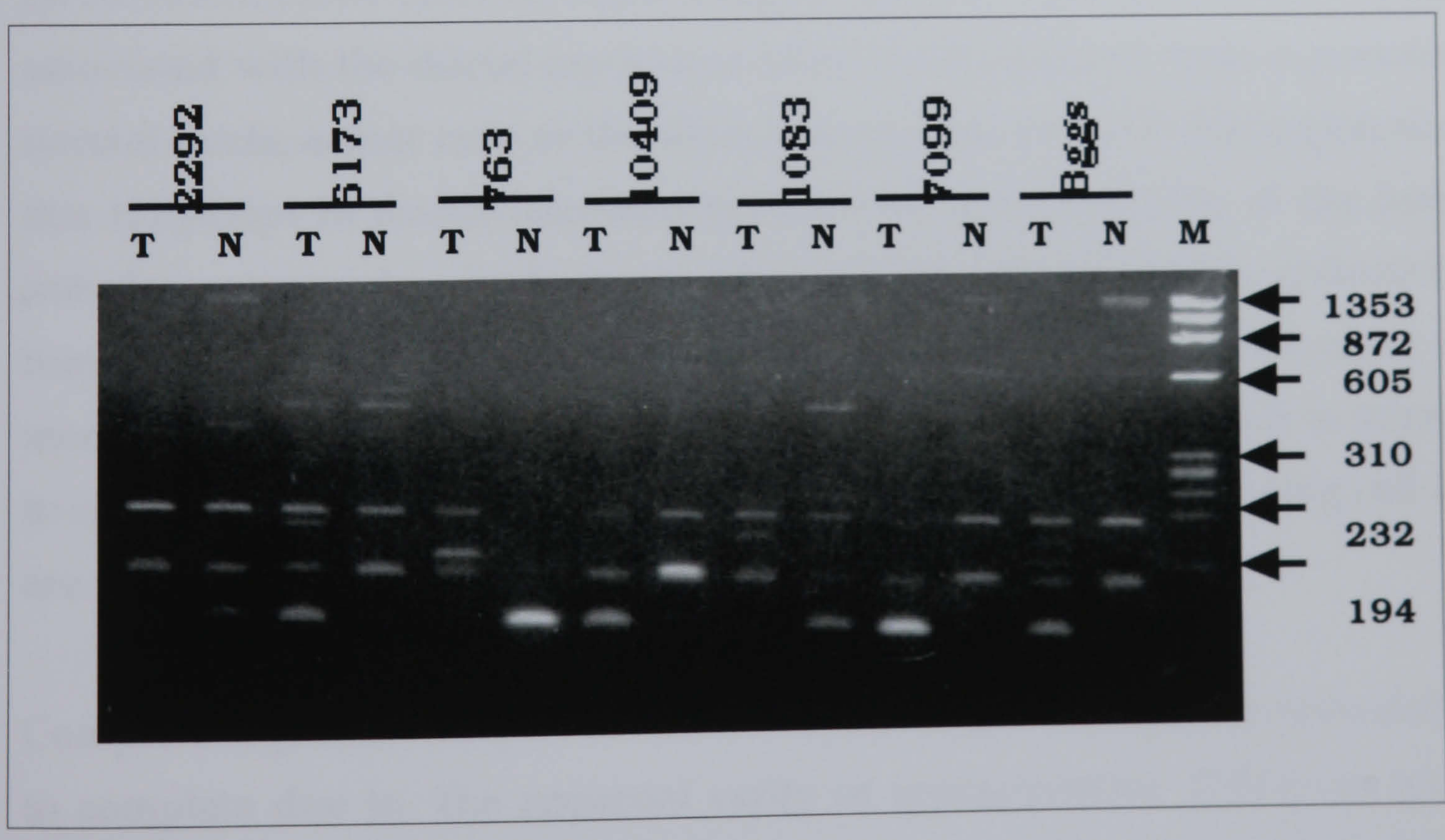


Figure 34: AB1-specific PCR carried out on pancreatic tumour and normal pairs. The variation in band pattern produced, particularly "shadow" bands would suggest an unstable target sequence.

Chromosome 2 losses have not been one of the most frequent finding associated with pancreatic cancer. However in a karyotypic analyses of xenograft-enriched tissue, 4/14 tumours show loss of the short arm of chromosome 2 and two further tumours show specific deletion of a region between 2p23-p21 (Seymour *et al* 1994). In accompanying LOH analysis 14/14 of the xenograft tumours exhibit loss of markers from chromosome 17p.

4.2.2 Isolation of 2p22 transcript overexpressed in pancreatic tumour cell lines

Isolation of contiguous P1 bacteriophage clones and subsequent subcloned fragments identified a 4.5 kb transcript adjacent to the deleted fragment which appeared to be expressed only in adenocarcinoma of the pancreas. The transcript is expressed at almost undetectable low levels in most tissues examined (assessed by multiple tissue northern and TISH of normal tissue blocks) including normal pancreas epithelial cells. Differential expression was observed however in the established pancreatic tumour cell lines, with certain cell lines exhibiting as much as five to ten fold increased expression compared to the levels seen in primary cultures of normal pancreatic cells. Altogether, 12/24 tumour cell lines exhibited clear increased expression of the transcript on Northern blots. TISH revealed a highly specific expression of the transcript associated with the ductal carcinoma tissue with relatively little expression in normal ducts, acinar cells or the abundant stromal element. The expression of this transcript in ductal adenocarcinoma was seen in 18-20% of the tumour samples examined, which represented a more heterogenous collection of tumour morphology. The expression did appear to be associated more with moderately to well-differentiated tumours an observation which is mirrored to some extent in the tumour cell lines in that the higher expressing cell lines are from well-differentiated tumours.

Complete sequence analysis of the chromosome 2 transcript has been difficult to complete due to the apparent rarity of representative cDNAs in normal tissue libraries. Accordingly, maximal use was made of pancreatic gridded cDNA libraries and other cDNA libraries derived from tumour material. The

characterisation of flanking sequences revealed the deleted AB1 fragment to be in a 5′ orientation to a putative ORF, therefore analysis of potential upstream regulatory elements was of prime importance. Achieving a more precise physical location with respect to other markers and transcripts was also difficult to achieve since the short arm of chromosome 2 represents 93 Mb of DNA sequence which is not so rich in known gene sequences or physical mapping data (cf Chromosome 17). Sequence similarity between genomic shotgun fragments generated in the laboratory and the CYP1B1 gene already reported to map to bands 2p21-22 eventually meant we could identify flanking markers and ESTs from that area. Sequence comparisons with known transcripts localised to chromosome 2p22 revealed homology with a previously reported unknown transcript encoding a 130 kDa leucine-rich protein (Lrp130). Lrp130 had been isolated by virtue of its up-regulation in a hepatoblastoma cell line (HepG2) and its potential association with fibroblast growth factors. Our original genomic fragment AB1 deleted in the fingerprint of a tumour sample compared with its constituent normal sample resides upstream of this transcript and within the 5′ flanking sequences of the *CYP1B1* transcript also. The possibility of these two genes having overlapping upstream promoter areas might suggest transcriptional regulation of both genes may be governed by the same cofactors. Researchers are now realising that DNA secondary conformational changes with its chromatin scaffold have a strong influence on transcriptional regulation and silencing.

The sequence derivation of this upstream region has offered some clues to possible mechanisms by which adjacent transcripts may be inappropriately expressed. Initial sequence homology to a low copy repeat family of sequences found to be hypomethylated specifically in germ cells and germ cell tumours led us to investigate whether the AB1 sequence was hypomethylated in pancreatic tumour cell lines. Preliminary experiments showed a correlation between hypomethylation of the upstream AB1 sequence and up-regulation of the Lrp130 transcript in three of the highest expressing pancreatic tumour cell lines. We have attempted to ascertain whether this methylation is

genome-wide, or restricted to this locus by carrying out Southern analyses with cDNAs dispersed throughout the genome on *HpaII/MspI* isoschizomer blots. The experiments suggest this is a local phenomenon since cDNA fragments representing sequence from the middle of the ART1 transcript as well as three other genes and their promoter areas show no *HpaII* tiny fragments (HTFs). One gene however, believed to be a candidate cell-senescence gene on chromosome 7q appears hypomethylated with a probe representing the 5′ promoter region of the transcript. This same probe appears normally methylated in breast tumour specimens and is believed to be hypermethylated in ovarian tumours (Donny Black - personal communication). It is unlikely that the hypomethylation is a cell culture-induced phenomenon, as it is reported that cell lines grown in culture tend to undergo genome-wide *de novo* methylation rather than hypomethylation (Antequera & Bird, 1993b).

Recent evidence would suggest however, that cytosine methylation and demethylation is not a static process and that local areas of hypomethylation may arise through the influence of DNA conformation perhaps in addition to the influences of binding factors like Sp1 which is known to protect housekeeping genes from inactivation through methylation during development (Jones, 1996). Additionally, areas of hypermethylation and gene silencing are thought to diverge outwards, influencing adjacent gene sequences inappropriately, a process which might be associated with the natural ageing process (Issa *et al.*, 1994).

4.2.3 The Lrp130 gene discovery

In attempting to identify surface markers and receptors associated with the FGFs and their expression in liver, researchers isolated a detergent-soluble protein with wheat germ lectin (WGL)-binding activity of molecular mass 130 kDa (Hou *et al.*, 1994). Using an antibody they screened a lambda bacteriophage expression library, isolating a 4734 bp cDNA which codes for a 130 kDa leucine-rich protein highly expressed in HepG2 hepatoblastoma cells. The

protein called Lrp130 is 1272 amino acids long with leucine residues accounting for 13% of the total sequence. Neither hydrophobic NH2-terminal secretory signal sequences nor a hydrophobic transmembrane domain were evident which suggests that Lrp130 is unlikely to be secreted, nor a transmembrane molecule. Multiple tissue Northern analysis demonstrated the Lrp130 transcript expression in normal pancreas and duodenum tissue only (as ART1). Western analysis led the authors to believe the protein was glycosylated, though further analysis of this gene was not carried out since it proved not to have any FGF-binding activity.

4.2.4 ART1/Lrp130 may be part of large carbohydrate-lectin family

Sequence similarity had already been observed between the ART1 transcript and another lectin-like molecule overexpressed in primary liver cancer. Differential screening of a human primary liver cancer cDNA library identified a novel gene named *HIP*, the expression of which was markedly increased in 25% of human primary liver cancers (Lasserre *et al.*, 1994). The *HIP* protein was appears much smaller in molecular size to the Lrp130, however maintains the leucine-rich component to primary amino acid structure. The authors painstakingly localised this gene to chromosome 2 bands pter-q32 through hybridisation to a specific somatic cell hybrid panel. *HIP* mRNA normal expression is restricted to the pancreas and small intestine and is believed to be a member of the family of calcium-dependant (C-type) lectins. These proteins participate in many cell-surface recognition events mediated by protein/protein carbohydrate interactions.

The C-type lectin family comprises multifunctional proteins with carbohydrate recognition domains (CRD) found associated to a variety of other protein domains (Drickamer, 1993). Other members have reported to have been isolated from pancreas including the *Reg* protein identified in the rat after differential screening of a cDNA library derived from a **reg**enerating pancreas (Terazono *et al.*, 1988). The human *Reg* gene has been isolated and demonstrates 49% homology to the *HIP* gene. It has been shown that the *Reg*

protein stimulates pancreatic B-cell regeneration after surgical removal. The *HIP* gene was independently isolated from pancreas after inflammation induced by pancreatic transplantation and was named pancreatitis-associated protein (*PAP*) (Keim *et al.*, 1992). Sequence identity searches carried out by the authors reveal *HIP* and *PAP* are identical and are further referred to tentatively as *HIP/PAP*.

The C-type lectin family includes the selectin cell adhesion molecules, endocytic receptors and secreted molecules found in the extracellular matrix and serum. Several lines of evidence support the idea that lectins play a role in cell recognition and thus act as determinants of cell behaviour and multicellular organisation. In particular, a role has been postulated for lectins in cell migration and metastasis formation. As mentioned above, the *Reg* protein is a pancreatic islet cell regenerating factor, however *Reg* mRNA has been observed expressed in colon and rectal tumours suggesting a possible expression associated with cell proliferation. It would thus be interesting to examine colorectal tumours for the expression of the Lrp130 transcript.

4.2.5 Cell adhesion and tumour metastasis

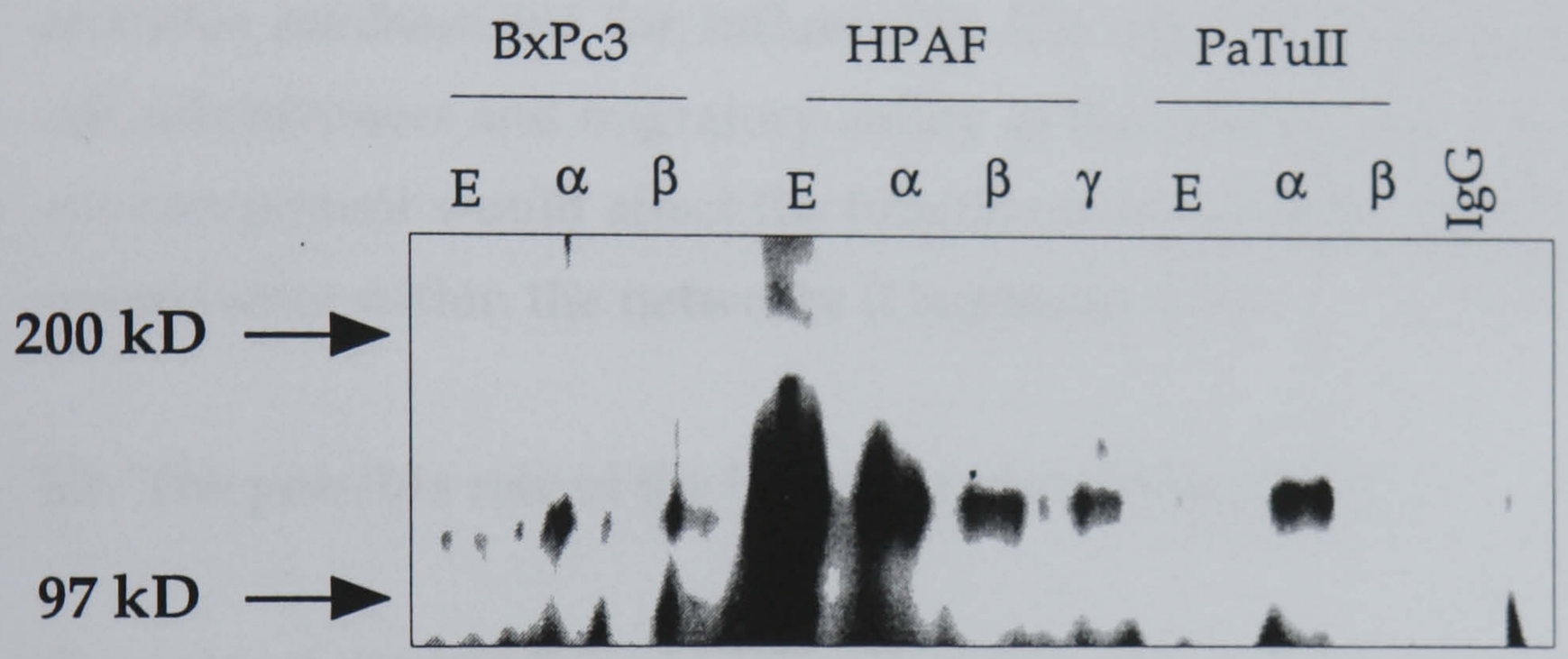
Cell adhesion is an important biological process that is implicated in cell homeostasis, inflammation, wound healing, tissue pattern formation and organogenesis. Recently however, increased interest has focussed on cell adhesion molecules and their possible involvement in promoting and sustaining tumourigenesis and metastasis. It has been realized that cell adhesion, like angiogenesis, proteolysis, motility and invasion, is also a key component in the metastasis cascade. Indeed, cell adhesion may play a governing role in each of these processes also. The cell adhesion molecules comprise at least four prominent families, the integrins, cadherins, immunoglobulins and selectins. In addition, there is a large array of other molecules which serve adhesive functions that defy categorisation as yet; typical examples being the endogenous lectins, proteoglycans, and sphingophospholipids.

4.2.6 Lrp130-Antibody co-localisation experiments

A *Xenopus* homologue of the Lrp130 protein has been reported to interact with C-cadherin stimulated us to examine such interactions in mammalian cells (Wallace McKeehan:unpublished data). Upon receiving an aliquot of Lrp130 antiserum from Wallace McKeehan's laboratory, (Hou *et al.*, 1994), experiments were carried out in the laboratory to investigate possible associations Lrp130 might have with other cell surface adhesion molecules. Various immunoprecipitates of E-cadherin and catenins which had been previously Western blotted, were challenged with the Lrp130 antiserum (figure 35). Interestingly, the antiserum recognised a 130 kDal protein in a complex precipitated with the E-cadherin antibody from the HPAF cell lysate (which expresses Lrp130 at a low level), but hardly at all from the BxPc3 lysate (shows upregulated Lrp130 expression). The α-catenin immunoprecipitated complex exhibited the almost the same result, however, the β and γ-catenin precipitates exhibit a much reduced signal in HPAF and BxPc3 lysates. This might suggest that the Lrp130 protein binds to E-cadherin only when catenins are not already bound. A possible function could be as a molecular chaperone to the small cytoplasmic tail of the E-cadherin molecule which binds to catenins, or alternatively it may function in some competitive way with catenins for this binding site. Such a competitive relationship with the catenins could be the basis for regulatory step in the transduction of signal to the cytoskeleton.

This preliminary experiment must be repeated however, and data on the relative expression levels of the catenins and cadherins in pancreatic tumour cell lines evaluated.

Figure 35: Interaction of lrp130 with E-cadherin and catenins in pancreatic tumour cell lines



E-cadherin and different catenins were immunoprecipitated from total cell lysates of BxPc3, HPAF and PaTuII cells with E-cadherin (lanes 1,4,8), α-catenin (lanes 2,5,9), β-catenin (lanes 3,6,10), γ-catenin (lanes 7) and IgG (lane 11) antibodies. This was followed by Western blotting using lrp130 antibody (1:500). Molecular weight markers of 200 kD, 97 kD are indicated by arrows.

B-catenin has the ability to bind to the APC protein, the product of the tumour suppressor gene that is mutated in Familial adenomatous polyposis and sporadic colorectal cancer. APC competes with E-cadherin for interaction with β-catenin and associates with microtubules. Thus through its regulated interactions with APC or α-catenin, cytoplasmic β-catenin may impinge on two of the major filamentous structural elements of cells. Additional interplay between β-catenin and the actin cytoskeleton is suggested by the finding that β-catenin is a binding partner for the actin-bundling protein fascin. It can be imagined that the summation of all these interactions provides mechanisms for influencing the signalling functions of β-catenin, cell adhesiveness and migratory ability in that alterations in the availability of one component would affect the functional interactions and locations of other components within the networks (Gumbiner, 1996).

4.3 The possible role of the hypomethylated sequences

The role of the homologous Sp-0.3.23 sequences is unclear, the evidence for regional *de novo* methylation in a large number and variety of sequences during early mammalian embryogenesis and demethylation in the germ line suggests an important functional significance which is discussed further below. Hypomethylation in sperm DNA and hypermethylation in embryonal and somatic tissue may indicate an imprinting phenomenon. Homology of the AB1/Hypomethylated species with flanking cDNA and genomic fragments generated in the laboratory show tight (98%) homology with repeated sequences derived from the intergenic spacer elements found between tandem arrays of rRNA genes. These rRNA tandem arrays can be up to 2.0 Mb in size containing hundreds of copies of many rRNA genes. The intergenic spacer sequences are rich in Alu and Line-1 (L1) repeat sequences hinting that the evolution of rRNAs as tandem arrays was perhaps due to duplication as a result of transposition through the L1 machinery. Accordingly, sequence similarity to L1 repeat elements has been observed on searching databases with the rRNA-like sequences isolated here.

The first step in the L1 mediated transposition is transcription of the L1 sequences from their internal promoter. L1 transcription is thought to be limited to undifferentiated early germ cells and dedifferentiated tumour cells as a result of undermethylation at CpG residues in these undifferentiated cells. This would suggest that L1 transcription relies upon the cellular Polymerase II enzyme. It has been suggested that one of the possible roles of CpG methylation, might be to reduce the expression of mobile genetic elements in differentiated cells (Yoder *et al.*, 1997). The authors here suggest that experiments to examine methylation patterns, usually employing the *HpaII/MspI* isoschizomers to digest the 5′ flanking sequences of genes are infact also characterising the 5′ flanking retroelements. They use as examples the house-keeping gene *HPRT,* (responsible for nucleotide scavenger pathways) and the *BRCA1* gene. With *HPRT,* a cluster of *HpaII/MspI* sites is found around the transcription start site, but nearly all (25 of 26) of the *HpaII/MspI* sites in the body of the gene lie in recognisable retroelements. The same arrangement is true for *BRCA1.*

The 28S, 18S and 5.8S rRNA genes are clustered together in large arrays containing about 60 copies of each gene. Such clustered arrays span about 2Mb of DNA (so called ribosomal DNA or rDNA) and are located on the short arms of each of the five acrocentric chromosomes. The other major RNA constituent of ribosomes, the 5S rRNA, is transcribed from genes which constitute part of a very large gene family which is clustered on the long arm of chromosome 1. One would assume the potential for recombination and subsequent cross-over between chromosome 1 and 2 is quite high. These are the two largest chromosomes, accounting for approximately 20% of the total human genome between them with chromosome 1 being particularly rich in CpG islands (Craig & Bickmore, 1994). Including spacer units, the rRNA gene families account for about 0.4% of the total DNA in the human genome and include a considerable number of pseudogenes and truncated genes/gene fragments. If we are proposing a recombination event has occurred by virtue of Alu or L1, inappropriately positioning the rRNA sequence between the

Lrp130 and *CYP1B1* genes, one would expect to observe gross rearrangement adjacent genes on Southern hybridisation. One may have to screen many individuals before such a recombination event might be observed. Alternatively, an Alu insertion event such as described for the *APC* and *BRCA 1* genes may be difficult to observed without ascertaining large chunks of upstream sequence for both tumour and normal cases.

4.4 DNA methylation: its role in development and cancer

4.4.1 Methylation in development

Jaenisch proposes a model for developmental regulation of methylation where initial methylation of germ line and embryonal cell lineages is described as nothing more than "epigenetic noise". In the development of the zygote from the two cell stage to blastocyst ready to implant, global demethylation occurs. With evolution the human endogenous retroelements may have seized this widow of opportunity to transpose throughout the genome taking adjacent genes with them. Upon implantation, global *de novo* methylation restores methyl-C to all cellulary expressed genes a process which Jaenisch believes "resets" the methylation pattern (Jaenisch, 1997b). The housekeeping genes remain under-methylated at their CpG islands due to the binding of Sp1 elements protecting the islands from *de novo* methylation (Brandeis *et al.*, 1994). Thus DNA methylation turns off permanently the transcription of those genes whose activity is not required in a particular cell type, but the housekeeping genes which need to be expressed at constitutively low levels in all cell types are protected from methylation. While the basic methylation pattern is generally preserved throughout development, additional local demethylation events take place on individual genes associated with their tissue-specific expression. The end result is that each cell type-specific gene is modified and repressed in most tissues, but is unmethylated in the tissue of expression.

Very little is known about the actual biochemical mechanism of

demethylation. Experiments in erythroleukemia cells showed that 5-methylcytosine residues are replaced with labelled cytosine during the demethylation reaction, suggesting that the target of this reaction is either the base alone or the entire nucleotide (Razin *et al.*, 1986). Weiss *et al* developed an *in vitro* system for studying demethylation (Weiss *et al.*, 1996). Their results show that this reaction is at least partly mediated by RNA molecules and that demethylation takes place through removal of DNA nucleotides and their conversion to an RNase-sensitive form.

The most direct mechanism by which DNA methylation could interfere with transcription would be preventing the binding of the basal transcription machinery to the naked DNA. It seems unlikely that DNA methylation would function to repress transcription globally by modifying the majority of CpGs in any chromosome if the only sites of action are a limited set of recognised elements for individual transcription factors (Kass *et al.*, 1997). The second possibility is that specific transcriptional repressors recognise the methyl-CpG and either independently or together with other components of chromatin, turn off transcription. This mechanism has the advantage of being independent of the DNA sequence itself, thereby offering a simple means of global transcriptional control. Moreover, because chromatin assembly represses transcription anyway, methylation-dependant repression would add to the repression already in place. Bird and colleagues have identified two repressors *MeCP1* and *MeCP2* that bind to methyl-CpG without apparent sequence specificity (Meehan *et al.*, 1992). Like DNA methylation itself, *MeCP2* is dispensable for the viability of embryonic stem cells, however it is essential for normal embryonic development. Recent studies indicate that *MeCP2* is a chromosomal protein with the capacity to displace histone H1 from the nucleosome (Nan *et al.*, 1996). Additionally, *MeCP2* contains a methyl-CpG DNA-binding domain, which might alter chromatin structure directly and a repressor domain which might function indirectly to confer long-range repression *in vivo*. More recently Nan and colleagues have established that *MeCP2* exists as a complex with histone deacetylase (Nan et al, 1998). The

complex also contains Sin3A, a co-repressor in other deacetylation-dependant silencing processes, as well as several other unidentified proteins. Further the authors show that methylation-dependant transcriptional silencing can be overcome by adding trichostatin A, which specifically inhibits histone deacetylase. Therefore these articles show that histone deacetylation is guided to specific chromatin domains by underlying genomic methylation patterns, and that methylation-dependant transcriptional silencing relies on histone acetylation.

4.4.2 Histone acetylation

The influence of acetylation on chromatin structure is not well defined. Hyperacetylation of the histone tails leads to subtle changes in nucleosome conformation involving histone protein-protein interactions between core histones and histone H1 and other proteins (Oliva *et al.*, 1990). Induction of hyperacetylation with the deacetylase inhibitor sodium butyrate increases the accessibility of DNA in chromatin to DNase I (Riggs *et al.*, 1977). High levels of histone acetylation also improve chromatin solubility, suggesting a reduced tendency to aggregate (Perry & Chalkley, 1981). Therefore these observations point to the possibility that deacetylation of histone proteins, particularly the basic lysine-rich N-terminal tails will increase the ionic interaction between lysine and negatively charged DNA and the compaction of histone proteins leading to a tighter packing of the nucleosome assembly structure, making accessibility for other proteins difficult.

4.4.3 Aberrant methylation in cancer

There is clear evidence that deamination of methyl-CpG dinucleotides is directly responsible for mutations that promote tumourigenesis. In the case of the *TP53* gene which is non-functional in many tumours, about one third of the mutations that destroy activity of the gene are at CpG dinucleotides (Jones *et al.*, 1992). Genomic sequencing confirms that the affected sites are indeed methylated in the genome. Unmethylated cytosine can also undergo deamination to yield uracil, rather than thymine. Despite the eventual

transition of the uracil to thymine upon replication, the C to T transition rate of unmethylated cytosine residues is much lower than that of 5-methylcytosine (Cooper & Youssoufian, 1988) Additionally, the repair mechanism for U-G mismatches is much more efficient than that for T-G mismatches (Laird & Jaenisch, 1996).

The changes in DNA methylation that accompany carcinogenesis have been summarised in many reviews. Decreases in overall content of 5-methylcytosine (Feinberg *et al.*, 1988), demethylation of specific loci (Feinberg & Vogelstein, 1983), *de novo* methylation of CpG islands (Baylin *et al.*, 1987) and increased levels of DNA methyltransferase enzymes (Kautiainen & Jones, 1986) have all been observed. All these observations show that methylation changes are a highly consistent feature of carcinogenesis, but have not provided evidence for a causal relationship between the two phenomena. Recent experiments with mice deficient in methyltransferase have, however, clearly established that decreased methylation caused by genetic or pharmacological manipulation leads to decreased colon polyp formation in mice carrying predisposing cancer-causing mutations (Laird *et al.*, 1995). This result provides us with direct evidence that methylation and carcinogenesis are linked. Other findings strongly supporting a role for aberrant methylation in carcinogenesis come from many observations that the CpG islands of growth regulatory genes and particularly those which demonstrate tumour-suppressor function are frequently hypermethylated (or *de novo* methylated) in tumours and cell lines (Table 7). Thus the *de novo* methylation seen in tumours is deactivating those copies of the gene sequences which are not inactivated through point mutation or deletion.

Table 7: Methylation of CpG islands in various tumours

Tumour type	CpG islands methylated (frequency)		References
Retinoblastoma	Rb	(6/77)	Sakai et-al 1991 Am. J. Hum. Genet. 48: 880-888
Renal cell cancer	VHL	(5/26)	Herman et-al 1994 PNAS 91: 9700-9704
Various cancers	p16/INK4A	(23/67)	Merlo et-al 1995 Nat. Med. 1: 686-692
Bladder cancer	Myf3	(3/3)	Rideout et-al 1994 Mol. Cell Biol. 14: 5487-5494

It seems likely that the *de novo* methylation seem in tumours is not driven by an active wave of methylation such as that seen in embryonic development, but rather local *de novo* methylation of CpG islands non-specifically. Tumour suppressor genes containing CpG islands may be susceptible to this *de novo* methylation, resulting in down regulation of growth-suppressing genes. There is direct evidence that methylation errors in the oestrogen receptor CpG islands in apparently normal colonic epithelium occurs in older individuals (Issa *et al.*, 1994). Since the incidence of most solid tumours increase with age, perhaps there is a causal relationship between age, increased methylation and formation of particularly tumours.

Conversely, studies have documented reduced levels of proto-oncogene methylation, suggesting a role for DNA hypomethylation in the activation of proto-oncogenes. Several studies have been carried out in rodent liver model systems where reduced levels of DNA methylation of *raf, c-myc, c-fos, c-H-ras* and *c-K-ras* has been associated with neoplasia (Ray *et al.*, 1994 ; Vorce & Goodman, 1989). There are also reports which dismiss such events as irrelevant and that it is specific mutation of codons which are the primary activating mechanisms (Menegazzi *et al.*, 1988). A good inverse correlation between hypomethylation status and gene expression has been reported in the

Bcl-2 gene in human B-cell chronic lymphocytic leukemias (Hanada *et al.*, 1993) however, it remains difficult to prove a cause and effect of such a relationship.

4.4.4 Evidence of AB1/Sp-0.3-23 hypomethylation releasing Lrp130 from transcriptional repression

In the results chapter, preliminary evidence is presented which suggests (at least in the three pancreatic cancer cell lines expressing the highest levels of Lrp130) that the upregulated expression of the transcript correlates with hypomethylation at *HpaII* sites within the AB1/Sp-0.3-23 sequences. Since the AB1/Sp-0.3-23 sequence resides within promoter area of the Lrp130 gene, this correlation is not surprising. The real issue would be to ascertain whether this Lrp130 upregulation is solely caused by hypomethylation, or indeed are activating point mutations present in other tumours or cell lines. More fundamentally perhaps, is this gene upregulation seen in other tumours such as breast (suggested from Northern data) and does such expression have a biological consequence to the tumour.

Unfortunately, RNA isolation from frozen tissue specimens has proven difficult and therefore we are restricted to established tumour cell lines to make assumptions about the regulatory significance of the hypomethylation. Further TISH of archival material may extend this study with a PCR-based methylation analysis to give more statistically relevant numbers. Additionally., the analysis of other tumour types as well as cell lines might present a more amenable subject tumour to study.

4.4.5 PCR-based Methylation detection system - CpGenome™

The CpGenome reactions were carried out with additional tumour and normal DNAs from microdissected pancreatic cancer cases which were then sequenced manually with the Dideoxy-γ-NTP terminator kit (Amersham-USB). The rationale being that if the reactions were not going to completion, or particularly with the isolated cases, there were two possible

species - the modified thymine nucleotides from hypomethylated tumour cells and the methyl-cytosines from normal stromal cells, there would be a greater possibility of observing it on a conventional ^{33}-P sequence ladder on a polyacrylamide gel with four lanes. In fact that is what was observed, in most cases, even the tumour cell lines which represent a pure population of tumour cells cytosine residues were present with thymine at the same position. Only the cell line T3M4 did in fact demonstrate a clear case of a cytosine residue having been replaced with thymine. The positions of the bases which appeared to have been modified appeared in areas of the sequence where all four bases were represented, rather than GC-rich stretches, which may reflect a measure of how efficient the denaturation step was because of the higher bond energy of the G-C stretches. One instance was observed of a cytosine base from a CpGpC trinucleotide deaminating to a thymine, both appearing to co-migrate at the same base position in the nucleotide sequence. In addition, a microdissected tumour case (7099) as well as demonstrating cytosine deamination randomly within poly-C tracts, one instance of a cytosine deamination at a CpG dinucleotide was observed with the sequence CCCGGG (figure 22).

Therefore it did appear (at least in one case - 7099) that the primary tumour samples may be hypomethylated also. We restricted our analysis to the microdissected DNA samples isolated from frozen sections as the possibility of detecting modified (tumour-specific) thymine nucleotides in the more grossly dissected tumour cases with at least 30% normal stromal cells there was not worth trying. The department's recent acquisition of a laser-dissecting microscope should prove of beneficial use in projects such as this where very precise microdissection of cell clumps can be isolated. The possibilities to investigate CpG methylation status and other phenomenon in putative early premalignant lesions as well as at different stages of tumour progression could prove invaluable.

4.4.6 Possible experiments to prove the hypomethylation is causatively linked to Lrp130 increased expression

Proving that there is a direct relationship between the hypomethylation and the transcript up-regulation is still going to be difficult however. Hypermethylation influences in cancer cells are predictably resolved by growing the cells in the presence of 5-azacytidine, an analogue of cytidine in which a nitrogen atom is substituted for a carbon in the number 5 position of the cytosine. The analogue is incorporated into newly synthesized DNA, this DNA then becomes unmethylated as the methyltransferase does not recognise the nitrogen group. Subsequently, any transcript which is suspected to be repressed by hypermethylation will then be released from such repression and transcriptionally activated. There is no converse system to induce hypermethylation and subsequent transcriptional repression in an otherwise transcriptionally active hypomethylated gene.

Transient transfection experiments using a cell line which had a constitutively low level of Lrp130 expression (CF Pac1) and hypo/hyper-methylated Lrp130 mini-gene constructs might furnish an indication of whether methylation status of the transgene influences cell morphology/growth. Of course, to start with a cell line which is tumour derived initially may camouflage any subtle changes in cell surface or behaviour. This type of experiment would be best carried out on primary cultures of normal pancreatic epithelial cells. *In vitro* methylation of an Lrp130 transgene with upstream sequences and subsequent gene knock-out through homologous recombination could be a line of investigation. This approach often leads to lethal mutants early in embryological development however, with no information gained about the specific role of the transgene being manipulated. Such has been the case for the type I epidermal growth factors such as EGF and *ERBB2*.

An *in vitro* demethylation assay system developed by Weiss *et al* however may provide the means to study an Lrp130 transgene further. In their article

describing the assay, they use 810 bp of the fragment of upstream sequences from the rat α-actin gene (Weiss *et al.*, 1996). They made sure it was methylated at three *HpaII* sites using a *HpaII* methylase, end-labelled the fragment then incubated it with their cell-free extract. When methylated the fragment could not be cut by the restriction enzymes *HpaII* or *Sma I*, but following incubation with their cell-free extract, the sites became cleavable. Subsequently they describe experiments which strongly suggest this demethylation is mediated by RNA molecules. The cell-free extract appears resistant to high concentrations of proteinase K, pronase E, a 70 °C heat denaturation, or when the reaction itself was incubated at 65 °C. In contrast incubation with RNase A, RNase T or EDTA completely abolished the activity. This cell-free extract could prove valuable to assess what difference hypo- and hypermethylated templates could have to the expression of the downstream Lrp130 cDNA in transient transcription assays.

4.5 Future research objectives for the 2p22 lesion

Following identification of sequence homology between the ART1 cDNA and the 130 kDa leucine-rich lectin-like cDNA, the possibilites of studying this gene further both at the DNA and protein level are numerous. Since the predicted amino acid sequence of this cDNA has no apparent transmembrane domains nor signal peptide for export from the Golgi, we assume it has some intracellular function. Therefore perhaps more meaningful data could be generated through expression of this protein in cells with subsequent localisation experiments carried out utilising fluorescent tags. With the complete transcript, the possibility of creating a mini-gene construct with the full length cDNA under the control of an inducible promoter is one which is easily achievable. Studies can then be carried out to examine possible phenotypic changes to normal pancreatic cell morphology on induction of the transcript. Similarily, *invitro* gene-knockout experiments using antisense oligonucleotides or full length antisense RNA may enable us to investigate any phenotypic changes induced by reduction of the endogenous expression

seen in tumour cell lines such as BxPc3 and 818.7. This gene appears upregulated in 50% of established pancreatic tumour cell lines and 18% of primary pancreatic tumours. The difference in expression levels seen here may distinguish the contribution of normal cell components of the complex stromal element in primary tumour samples. Preliminary data suggests hypomethylation of upstream sequences may induce inappropriate expression of this transcript and experiments to examine this further have already been described above.

The full length cDNA also allows the design of peptides which may precipitate an immunogenic response in rabbits giving us a polyclonal antiserum to use in immunocytochemical experiments. With such an antiserum, it may be possible to study the expression of this gene more widely quicker. A concern might be however that such antisera will be immunogenic with other lectin-type moeties. Immunoprecipitation experiments may also enable us to identify associated proteins which bind to this gene product. Indeed upon contacting the researchers in Houston who first reported this gene, ensuing dialogue identified a third group who had identified a Xenopus homologue of this protein immunoprecipitating with C-cadherin.

The analysis of Lrp130 with respect to its prognostic value in cases of adenocarcinoma of the pancreas may be carried out by looking at a larger number of cases with the antisera.

4.6 The analysis of the 17p13.3 lesion

4.6.1 Alumorph IP-1 localized to 17p13.3

The IP-1 alumorph on chromosome 17p13.3 presented us with a more readily achievable mapping project and with no shortage of compelling evidence in support of a tumour suppressor gene at this region. The short arm of chromosome 17 is of particular medical interest because of its frequent involvement in many different tumour types, as well as conditions such as the Miller-Dieker syndrome (Chong *et al.*, 1997) and other disease associated genes such as deafness (Kelsell et al 1997), cataract and several forms of degenerative retinal disorders (Inglehearn, 1998). In addition to the well characterised p53 gene (17p13.1) there is mounting evidence that a second tumour-suppressor gene mapping towards the telomere might be involved in the progression of many types of solid tumour including breast (Coles *et al*, 1990; Cornelis *et al* ., 1994: Merlo *et al* ., 1994: Nagai *et al.*, 1994; Stack *et al.*, 1995), ovarian (Phillips *et al.*, 1996; Schultz *et al.*, 1996) pancreatic (Hahn*et al.*, 1995; Seymour*et al.* 1994) bladder, (Williams *et al.*, 1994), hepatocellular carcinoma (Fujimori *et al.*, 1991)and prostatic tumours (Rinker-Schaeffer*et al.*, 1994). Additionally, other malignancies such as astrocytomas (Chattopadhyay *et al.*, 1997), medulloblastomas (Cogen *et al.*, 1992), small cell lung carcinoma (Konishi *et al.*, 1998) and uterine cervical cancer (Park *et al.*, 1995) have also been described.

4.6.2 Allelic losses on 17p13.3

Allelic loss of the short arm of chromosome 17 is among the most commonly recognised genetic alterations in pancreatic cancer well as breast and ovarian cancers. The chromosome 17p losses often coincide with mutations in the *TP53* gene at 17p13.1, however in some tumour types, 17p allelic loss occurs at a high frequency in regions distal to *TP53* and in the absence of *TP53* mutations. In some cases of breast cancer, the 17p allelic loss is limited to a telomeric portion of the chromosome that excludes the *TP53* locus (Coles *et al.*, 1990; Nagai *et al.*, 1994; White *et al.*, 1997). In a cytogenetic study of 62

primary pancreatic adenocarcinomas, (Griffin *et al.*, 1995) frequent loss of chromosome 17 is observed (13/62; 21%) as well as structural abnormalities with breakpoints on 17p occurring in 5% of cases. While reports of 100% loss of heterozygosity (LOH) on 17p in pancreatic adenocarcinoma using xenograft enrichment intensified our interest (Hahn *et al.*, 1995). The observed losses here always involve markers around the p53 locus though details of the markers used were not included. Donis-Keller and her colleagues report 80% of 39 ovarian tumours informative for the marker D17S28 (17p13.3) exhibiting allele loss (Phillips *et al.*, 1996), 7 of which were of low malignant potential. More moderate levels of 36-44% loss with markers around and including D17S28 are reported by the group at the Paterson institute in a series of sporadic breast tumours (Stack *et al.*, 1995). A figure of 43% is quoted (Nagai *et al.*, 1994) from a sample of 42 informative Brazilian cases of breast cancer. Additionally, they also demonstrate a correlation between distal 17p13.3 loss and higher risk of recurrence compared with more proximal losses around *TP53* (p=0.007). Additionally, in a recent report, (Konishi *et al.*, 1998) it was demonstrated that while 89% of SCLCs containing *TP53* mutations also exhibited LOH at 17p13.3, 44% of the overall number (99) demonstrated 17p13.3 LOH with no *TP53* mutation or LOH at 17p13.1.

Periodic assessment of LOH status at 17p and other loci has been carried out in our laboratory (McKie, Sirivatanauksorn, Ding, Perry-unpublished data) (Barton *et al.*, 1991). In an originally small series of pancreatic tumour normal pairs (n=22), Barton *et al* did not find evidence of allele loss with markers on 17p including YNZ22 (D17S5) and an intragenic p53 VNTR. In an equally small series of more precisely microdissected tumours (n=14), there is evidence of LOH with markers D17S1174, D17S5, and D17S379. The flanking markers Utah-20 (YNH37.3) and D17S1174 both exhibited LOH (Figure 27) in three out of fourteen informative cases (21%) delineating an area of approximately 500 kb. Obviously we do not observed such convincing LOH around 17p13 as the Johns Hopkins investigators which may reflect the efficiency of the two procedures in trying to achieve a homogenous tumour

sample. There is also the possibilty we are looking at a genetically different disease poulation. Nevertheless, the convincing LOH data from Hahn *et al*, and the localisation of IP-1 to 17p13.3 in the vicinity of D17S28 it was decided this locus merited further investigation.

4.6.3 PITPn and its role in inositol lipid hydrolysis and signal transduction

Phosphatidylinositol transport proteins were originally identified and named because of their ability to transport phosphatidylinositol through the aqueous phase from one membrane compartment to another. Recent data however, indicates roles for PITPn in the coupling of Phosphatidylinositol (4,5) bisphosphate (PIP_2) synthesis to signal transduction reactions and to membrane traffic in the mammalian cells (de Camilli *et al.*, 1996). Inositol lipids have specialised functions in eukaryotic organisms providing a source of second messengers in signal transduction mechanisms, but also are now recognised as acting as signalling molecules themselves. Phosphatidylinositol (4,5) bisphosphate (PIP_2) was initially recognised as a substrate for phospholipase C (PLC) in the generation of two second messengers inositol (1,4,5)triphosphate (IP_3) and diacylglycerol (DG) (Berridge, 1993). A regulatory role for PIP_2 as an intact lipid is believed to exist in the activation of phospholipase D and cytoskeletal reorganisation and intracellular vesicular traffic (de Camilli *et al.*, 1996).

Extracellular signals that stimulate PIP_2 hydrolysis include polypeptide growth factors (EGF, FGF and PDGF) that bind to receptors with intrinsic tyrosine kinase activity. The phosphorylated tyrosine residues of autophosphrylated receptors provides high affinity binding sites for several proteins containing SH2 domains including Phospholipase C (PLC). The activated receptor tyrosine kinase also phosphorylate PLC. The two second messengers generated from PIP_2 hydrolysis have been implicated in regulation of processes as diverse as cell proliferation, contraction, secretion and phototransduction. Diacylglycerol generated by hydrolysis of PIP_2 by PLC binds to and activates many members of the protein kinase C family.

Experiments with permeabilised HL60 cells in culture have proven that inositol lipid hydrolysis in the complete absence of PITPn can be increased eight-fold with the addition of PITPn to the cytosol (Thomas *et al.*, 1993). Thus PITPn acts not only as inositol lipid carrier, but dictates rates of hydrolysis of inositol lipids in a substrate-limiting fashion by providing the inositol lipid substrate at remote sites (cell membrane) from the Golgi apparatus. Thus one may be able to appreciate the possible far reaching consequences of this gene being deleted or mutated in tumour cells. The permeabilisation experiments suggest inositol hydrolysis would continue though at a reduced rate, thus perhaps compromising the efficacy of extracellular signalling mechanisms. Point mutation may also reduce the affinity for inositol lipid substrate or cause the irreversible binding of such substrates, again compromising the signal transduction process from receptor autophosphorylation to protein kinase cascade activation.

4.6.4 The mouse *Vibrator* critical region and SSCP analysis of *PITPn*

As well as homology to the human*PITPnα* gene sequence the cDNA and genomic sequence we had exhibited similar homology to the mouse *Vibrator* critical region, incorporating the mouse PITPnα homologue. The *Vibrator* mutation in mice causes an early-onset progressive action tremor and neurodegeneration of the brain stem and spinal chord leading to juvenile death A retrotranspositional insertion into intron 4 of this locus was found to be the main mutational event in reducing *PITPnα*mRNA levels.

4.6.5. Future work

The *PITPnα* gene, as the 17p13.3 candidate tumour suppressor gene needs to be characterised further and the exon/intron structure of the human gene defined. This gene represents our first candidate gene in the deleted region by virtue of its clear proximity to the original IP-1 alumorph. Clearly, experiments to look for mutations within pancreatic tumour samples are at an early stage, and a candidate gene approach at this point may have been presumptive, however a worthy starting point. The17p13.3 deletion area

represents around 500 kb of sequence with two known genes (*HIC-1* and *PITPnα*) as well as a putative processed pseudogene (sHER4) in proximity to the IP-1 alumorph. Further genes may reside within this deleted area since chromosome 17 is particularly gene rich. In the analysis of the NCBI-Unigene cluster database which assignes "clusters" of EST clones sharing sequence homology or encode the same gene, I have estimated that approximately sixty four transcripts are defined between markers D17S849 (telomere) and D17S379.

A sensible approach may be to firstly isolate more cosmid, PAC and YAC clones from this region then identify expressed sequences through hybridisation or PCR of known ESTs from the Unigene clusters. Once a subset of transcripts known to reside within a physical contig, the original EST clones can be used as probes in Southern blots to look for tumour-specific rearrangemnts, and/or candidate genes may be analysed for mutation. The cDNA clone 2 has not demonstrated sequence similarity with the *PITPnα* cDNA sequence, or any other similarity with a known gene, therefore another gene candidate may exist around cosmid IA for analysis.

An added complication in the analysis of genes at this locus may be the abundant reports of hypermethylation around the HIC-1 gene, particularly in ovarian, lung and basal cell carcinomas. Perhaps a large proportion of the tumour samples reported to be exhibiting LOH on 17p13 have their other chromosome 17 arm inactivated through hypermethylation.

4.7 Analysis of other Alumorphs

4.7.1 BACH1-specific Alumorph

A sixth Alumorph, AB4, was also identified as a fragment lost in the tumour fingerprint profile compared with its paired normal sample. The fragment is the largest to be isolated thus far (approximately 700 bp) and has been used directly as a probe in Southern and Northern hybridisations. No losses between tumour and normal lanes are observed on Southern blots, but expression of the transcript was observed in the SUIT-2 clonal derivatives with higher metastatic properties (figure 14c).

Recently, we carried out further searches on the databases and found a significant sequence similarity (91.3%) between bases 270 and 600 of the AB4 fragment and bases 185-515 of a bZIP transcription factor BACH1 (Appendix 2d). This similarity extended to the mouse homologue. The "bZIP" class of transcription factors contain a leucine zipper as a common structural motif required for protein dimerisation. All of these factors bind to specific DNA sequences as homo- or heterodimers with the DNA interaction surface being provided by a region of basic amino acids that immediately precedes the leucine zipper domain. Hence, many authors use the term "bZIP" to refer to this particular class of transcription factor. PCR primers have been synthesized from either end of the AB4 fragment, however amplification from genomic DNA has proved difficult. Multiple bands have been evident irrespective of conditions employed, and thus localisation with the ICRF somatic cell hybrid mapping panel eludes us for the moment. In view of the recent sequence search data we have now designed alternative primers which we hope may be more specific. Obviously if we have a sequence encoding something as abundant as a leucine zipper domain problems might be encountered amplifying other leucine zipper-containing genes unless primers are designed which are outwith this sequence motif. Additionally, searches may provide us with IMAGE cDNA clones specific for the sequence. We have acquired primary PCR-pools from the human PAC library we have already in the

gridded version, thus it may be possible to isolate an AB4-specific PAC with this resource.

4.7.2 Met6 Alumorph

Southern analysis of the ICRF somatic cell hybrid panel localised this fragment to chromosome 11p, a result which was later confirmed with the isolation of a clone from a chromosome 11 gridded cosmid library with FISH analysis confirming this result mapping the cosmid to chromosome 11p12-13 The chromosome 11p12-13 cosmid (G08878) had previously been requested and localised by numerous laboratories. Groups within the ICRF, the CRC-Fulham Rd and the Institute of child health (ICH) had placed the cosmid on the long arm of chromosome 11 at band q23. Equally, researchers at the MRC and CRC laboratories in Edinburgh and Cambridge respectively have placed this cosmid on the short arm of chromosome 11 at band p13-15. On contacting both the ICRF and MRC laboratories involved, it was established that this cosmid is believed to be particularly rich in repeat sequences which is the contributory factor in the confusion over its location. In addition, the cosmid is believed to contain zinc-finger sequence motifs which also may compound the confusion as chromosome 11, particularly the q arm is believed to be very rich in zinc finger containing genes. Researchers at the MRC Human Genetics Unit (HGU) believe the cosmid to reside within the Beckwith Weidermann Syndrome (BWS) region which also ofcourse harbours the Wilm's Tumour-suppressor gene *WT-1*. Chromosome 11p is an area of the genome known to be imprinted in the developmental abnormality BWS and Wilm's tumour. Two further cosmids have since been isolated again with the Met6 Alumorph as a probe and again the same list of researcher have previously requested and localised these Further analysis is needed to characterise this Alumorph more fully and a possible collaborative study with researchers within the ICRF laboratories in Edinburgh is being explored.

4.8 Future research

It has become clear that to characterise all the potentially interesting bands I have thus far isolated is a major undertaking. I have discussed some of the other findings to demonstrate the success with which the Alu-PCR technique has enabled us to isolate these species. I would emphasise however, these alumorphs are going to be characterised more fully, particularly AB4 with its interesting homology to the BACH1 bZIP transcription factor. Interestingly, another investigator in the laboratory working on differential display as a method to identify genes up-regulated in pancreatic cancer identified a transcript which also exhibited homology with the BACH1 protein. The DD-fragment exhibited differential expression of a 5 kb transcript up-regulated in 4 or more established pancreatic tumour cell lines (Helen Hurst-personal communication). Potential problems exist when isolating any gene which encodes an abundant structural motif or is a member of a gene superfamily. It becomes apparent when trying to isolate specific clones for that gene or family member that it is often difficult to distinguish close homologues when carrying out hybridisations. Often DNA probes from the 3UTR of such family member genes show sufficient sequence divergence to allow the successful isolation of the required gene.

4.8.1 EST projects give us a molecular inventory of pancreatic tumours

Recently, a pancreatic cancer-specific expression profile was generated using gridded arrays of pancreatic cancer cDNA libraries and differential hybridisations (Gress *et al.*, 1996). Of the 408 ESTs generated which appeared specifically overexpressed in pancreatic cancer, 269 (65%) proved to be homologous, or partly homologous to known genes or sequence tags from the GenBank databases. These 269 ESTs included 24 structural and cytoskeletal proteins (6% of 408 ESTs), 13 cell surface/adhesion molecules (3%), 13 known signal-transducing receptors and 9 other enzymes or cofactors (2.2%) which may have roles in metastasis such as plasminogen-activator. In an equally substantive inventory of expressed cDNAs (1,000) from pancreatic islets

(Takeda *et al.*, 1993), Bell *et-al* identify 603 sequence-tags as recognised genes and 397 unknown transcripts. These 603 transcripts included 23 cell surface/adhesion molecules (3.8%), 12 (2%) cytoskeletal-related transcripts and 4 (0.6%) proteolytic enzymes which may be assocated with ECM degradation. Encouragingly, the *Reg/HIP* gene product was found to be expressed at relatively high levels.

This technology is already becoming out of date, with the advent of DD and SAGE as well as the prospect of micro-chip arrays of cDNA tags. It will soon be possible to assess exactly what genes are being expressed specifically in a disease state such as pancreatic cancer or chronic pancreatitis by hybrising tissue cell lysates over these micro-chips. The last three years has seen our laboratory on the periphery of such technology. The development of PCR-based techniques such as DD have produced modest studies with the production of dozens rather than hundreds of expressed tags. There are many laboratories like us addressing such methods to their own disease phenotype in the same modest way. The limitations are often man-power, computing capacity/expertise to identify and categorise the identified tags and applied resources. The information one can gain from experiments of this kind perhaps undertaken on a specific tumour cell line or group of cell lines may be unrepresentative of human cancer and its progressive pathway. We have tried in the laboratory to apply such technology to clinical material comparing pure cancer isolates with normal tissue even in primary cultures in the case of DD.

How does genomic fingerprinting contribute to the identification of the tumourigenic pathway in pancreas cancer? The Alu-PCR technique was launched four years ago to assess the robustness of such technology and as a prelude to DD and AP-PCR technique development. We were looking to isolate specific DNA lesions which might point to the presence of a novel tumour suppressor gene. As discussed below, it is felt this has been achieved in part with the identification of a deleted region on 17p13.3 and the

identification of an upregulated gene, possibly as a result of local epigenetic influences on chromosome 2p22. The isolation of three further partially characterised rearranged genomic fragments has proven the techniques worth. With the isolation of associated transcripts, focus on differential expression of such transcripts in 3/5 instances has demonstrated how the two observations may be linked. A limitation in such analysis may be the assumptions made by observations in established tumour cell lines with reagents generated from primary tumour tissue. The isolation of intact RNA from frozen tissue specimens is difficult to achieve and with the bulk of pancreas tumour tissue representing the normal stromal element, meaningful experiments have to be carried out with micodissected tissue samples. Nevertheless, with the chromosome 2p22 locus, the increased expression seen in half the tumour cell lines has been observed also in 20% of primary tumour blocks.

During the course of these studies the group at Johns Hopkins have identified a new tumour suppressor gene (Smad4) with the RDA technique and identified a second smaller deletion which led to the identification of the BRCA2 gene. As discussed in the introduction, the use of tissue explants in mouse xenografts have circumvented the problems of stromal element contribution outlined above and provided a more abundant source of tumour tissue with maintainance of molecular integrity. Clearly, a marriage of such technologies would be of immense value in any disease phenotype.

4.8.2 The further development of the Alu-directed fingerprinting

Experiments have been carried out to examine the mechanics of generating these fingerprint profiles. Truncated oligonucleotide primers were manufactured from the Alu-specific primer PDJ-33 (Alu consensus:.nucleotides 47-13) which were progressively shorter by intervals of 5 bp from the full size of 35 bp to 15 bp. A survey of the Alu-specific primers available in literature with their specificity in amplifying only human sequences gives an indication of how complex the whole issue may be. It can

be observed from the table below first of all, that the PDJ-33 primer is the largest and yet is reported to be not human specific. The two primers which are reported to be human specific are the Alu-5′ and Alu-1 primers which both overlap almost completely with BK-33. Therefore we can assume BK-33 is probably human-specific in what it amplifies. Primer 278, like primer PDJ-33 contains 12 bp of 5′ sequence in addition to the Alu-1 sequence which we can assume confers the broader range specificity on these primers.

Table 8: Sequence alignment of seven Alu-specific primers

Primer name	Sequence (5' - 3')	Human specific?
Alu-V	GTGAGCCACCGCGCCCGGCC	**?**
PDJ-33	GCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA	**NO**
BK-33	CTGGGATTACAGGCGTGAGCC	**?**
278	CCTCCCAAAGTGCTGGGATTACAG	**NO**
CL1	TCCCAAAGTGCTGGGATTACAG	**?**
Alu-5'	GGATTACAGGCGTGAGCCAC	**YES**
Alu-1	GGGATTACAGGCGTGAGCCAC	**YES**

Of course whether an Alu primer is human-specific is not important in the context employed here, however such considerations identified in this survey serve to indicate how such PCR-based assays may be and the contribution made by short stretches of adjacent sequence.

Bassam *et al* have recently published a review article on the AP-PCR fingerprinting technology (Bassam & Bentley, 1994) trying to address some of the issues of how such small primer lengths used in this technology can result in highly complex band profiles at annealing temperatures often way above the melting temperature of the oligomers used. They observe that as one reduces the primer size, the resulting fingerprint profile decreases in informity, that is one observes fewer bands. This is the opposite to what one would expect for a completely random oligo annealing process and opposite to our observations with Alu, though clearly we are dealing with a different system and we have not dealt with primers below 15 nucleotides in length. These investigators also observe that smaller primers of 8-10mer size appear to anneal and give complex fingerprint profiles at 55-60^{0}C when they have a melting temperature of 30^{0}C. We also observe this with AP-PCR and results with Alu-PCR would suggest similar findings though the Alu-primers are G/C-rich and usually have higher annealing temperatures. They attribute this phenomenon to mismatch annealing events, however consideration should be given to networking the intricate branching of oligonucleotide primers partially annealing to each other, the template and hairpin loops formed by the single-stranded template and the PCR products already produced. It might prove interesting to carry out annealing experiments with end labelled oligomers and DNA templates to assess what kind of structures we attain with respect to mobility on PAGE and susceptibility to single strand-nuclease attack.

4.9 Aims of the thesis fulfilled?

The aims for this thesis were to first of all develop the DNA fingerprinting technology utilising Alu repeat sequences as high copy, ubiquitous targets as a means of amplifying random inter-Alu stretches of the genome. Using such technology to compare generated fingerprints from tumour and normal DNA samples isolated from the same individual, identify sequences which are altered in adenocarcinoma of the pancreas. Such sequences once isolated, could be used as molecular probes to examine additional tumour samples then assessing the frequency of the particular mutation event.

Clearly we have achieved the first part of this aim in developing an Alu-directed fingerprinting technique which has further been extended by collaboration with other groups and has now been published (McKie *et al.*, 1995; Leung *et al.*, submitted for publication 1999) I have achieved the second half of the aims outlined above by isolating at least two, perhaps five regions of genetic imbalance. Further analysis to ascertain the frequency of the chromosome 2 abnormality have been carried out (see below) and similarly with the associated up-regulated transcript.

The mutated or deleted sequences, (isolated from constituent normal tissue DNA if deleted), will then be used as radiolabelled probes to isolate clones from gridded genomic libraries, subsequently generating contiguous maps and sequence data. The pattern of expression of such genes isolated will be studied by Northern blotting, RT-PCR and in situ hybridisation and with resources such as the European Pancreatic Reference Library database System (EPRLS) cDNA libraries will elucidate the eventual gene structure.

I have achieved a great deal of this as far as the anomalies on chromosomes 2 and 17 are concerned. With the alumorph IP-1 localised to 17p13.3, I have enjoyed more abundant resources available in comparison

to chromosome 2. Chromosome 17 mapping has been carried out at pace in recent years with research concentrating in the search for important cancer predisposing genes such as *NF-1, BRCA1* and Leukemia-associated breakpoints. STS mapping has placed the IP-1-specific cosmid (Cos 1A) contiguous with the *HIC-1* gene. Further, sequence analysis of cosmid fragments and isolated cDNA clones has demonstrated significant homology with the *PITPnα* gene as well as a possible processed pseudogene for *HER4*. The *PITPnα* gene has previously been localised to this chromosome location, however agreement on the exact physical location is unresolved.

With the chromosome 2 project, homology at the genomic level has resulted in the localisation of the differentially expressed transcript ART1 contiguous with the CYP1B1 gene already localised to 2p22. The expression of this 4.5 kb transcript has been analysed through Northern blotting and tissue *in situ* hybridisation experiments demonstrating a ductal adenocarcinoma-specific expression. The expression of this transcript is restricted to pancreatic adenocarcinoma, and possibly other tumour types such as breast and hepatocellular carcinoma, though expression was observed in normal duodenal and bladder epithelium. Further cDNA homology to a leucine-rich lectin-like protein located in the same 2cM interval has been observed. This gene (designated Lrp130) exhibits a similar sized transcript with the same expression profile seen with ART1. STS mapping has established that the 5 flanking sequences of the ART1/Lrp130 gene (including the original Alumorph AB1) and the upstream sequence of the *CYP1B1* gene are contiguous with Lrp130 probably transcribed on the opposing strand in the opposite direction.

Investigation of the prevalence of such mutational events may be studied by the design of a simple PCR-based diagnostic assay, which may be carried out on clinical samples including archival paraffin embedded tissue specimens.

The development of a suitable amplicon for an RT-PCR based diagnostic assay should be straightforward once the gene structure has been characterised. Studies of Lrp130 expression in greater numbers of cases need to be assessed to investigate any possible associations with disease progression or mortality figures. Should the mechanism for Lrp130 up-regulation in an apparently inappropriate state be due to hypomethylation of the upstream sequences, then it will be a challenge to design a simple PCR-based assay to analyse its prevalence. A PCR-based method to ascertain the methylation status of tissue has recently been developed and is being assessed in the laboratory with sequences from the AB1 region, however, they remain involved and prone to subtle technical difficulties affecting interpretation of the end result. Deletion of sequences on 17p13.3 would be a more suitable target for the development of a PCR-based kit, however the exact extent of such deletions is still to be fully ascertained . Deletion analysis in a small series of microdissected tumours has demonstrated three out of twelve cases exhibiting homozygous deletion of an area flanked by YNZ22/HIC-1 and the marker D17S379. This represents a distance of five hundred kilobases with three or more genes located within this interval. Preliminary single-stranded conformational polymorphism analysis (SSCP) has been carried out on exons two and three of the *PITPnα* gene with no apparent conformational changes being observed. There is believed to be a CpG island at the marker D17S379, however efforts to isolate further clones from this area have remained unsuccessful.

Likely scenarios for the absence of PITPnα protein being associated with tumour progression are abundant also. Inositol phosphates are an abundant, well characterised group of second messenger molecules produced by phosphoinositide hydrolysis. Their production triggers release of intracellular calcium ions (Ca^{2+}) and suggests a role in neurotransmitter release. Inositol 1,4,5-triphosphate (IP3) is such a well characterised member of this family and is readily metabolised in cells by several

phosphatases and kinases with various derivative inositol phosphates produced. Among these metabolites is 1,3,4,5-tetra-bisphosphate (IP4) which has been suggested as having a role in Ras signalling pathways. The IP4 binding proteins have much structural similarity to the GAP family of Ras GTPase-Activating Proteins. *PITPnα* is thought to be involved in transporting the phosphatidyl-inositol from intracellular compartments for the conversion to Phosphatidyl-inositol bisphosphate prior to hydrolysis by Phospholipase C (Thomas *et al.*, 1993).

References

Achille, A., Biasi, M.O., Zamboni, G., Bogina, G., Magalini, A. R., Pederzoli, P., Perucho, M. & Scarpa, A. (1996). Chromosome 7q allele losses in pancreatic carcinoma. *Cancer Res,.*,**56,** 3808-3813.

Almoguera, C., Shibata, D., Forrester, R., Martin, J. & Perucho, M. (1988). Most human carcinoma of exocrine pancreas contain mutant c-K-ras genes. *Cell.*, 53, 549-554.

Andersson, G., Svensson, A.C. & Setterblad, N. (1998). Retroelements in the human MHC class II region. *TIG* **14**, 109-114

Antequera, F. & Bird, A. (1993). CpG islands.*Proc. Natl. Acad. Sci. USA.*., **90**, 11995-11999.

Bansal, P. & Sonnenberg, A. (1995). Pancreatitis is a risk factor fo pancreatic cancer. *Gastroenterology.*, **109,** 247-251.

Bardi, G., Johansson, B., Pandis, N., Mandahl, N., Bak-Jensen, E., Andren-Sandberg, A., Mitelman, F. & Heim, S. (1993). Karyotypic abnormalities in tumours of the pancreas. *Br. J. Cancer.*, **67**, 1106-1112.

Bar-Sagi, D. (1994). The Sos Protein. *Trends Endocrinol. Metab..*, **5,** 165-169.

Barrett, M.T., Schutte, M., Kern, S.E. & Reid, B.J. (1996). Allelic loss and mutation analysis of the DPC4 gene in esophageal adenocarcinoma. *Cancer Res.*, **56,** 4351-4353.

Barton, C.M., Staddon, S.L., Hughes, C.M., Hall, P.A., O'Sullivan, C., Kloppel, G., Theis, B., Russell, R.C., Neoptolemos, J., Williamson, R.C.N. & Lemoine, N.R. (1991). Abnormalities of the p53 tumour suppressor gene in human

pancreatic cancer. *Br. J. Cancer.*, **6,** 1076-1082.

Bates, S., Phillips, A.C., Clark, P.A., Stott, F., Peters, G., Ludwig, R.L. & Vousden, K. (1998). P14/ARF links the tumour suppressors RB and p53. *Nature,* **395**, 124-125.

Batzer, M.A., Deininger, P.L., Hellmann-Blumberg, U., Jurka, J., Labuda, D., Rubin, C.M., Schmid, C.W., Zietkiewicz, E. & Zuckerlandl, E. (1996). Standerdised nomenclature for Alu repeats. *J. Mol. Evol.* **42**, 3-6

Baylin, S.B., Fearon, E.R., Vogelstein, B., de Bustros, A., Sharkis, S.J., Burke P.J., Staal, S.P. & Nelkin, B.D., (1987). Hypermethylation of the 5' region of the calcitonin gene is a property of human lymphoid and acute myeloid malignancies. *Blood.*, **70,** 412-417.

Bergman, W., Watson, P., de Jong, J., Lynch, H. T. & Fusaro, R. M. (1990). Systemic cancer and the FAMMM syndrome. *Br. J. Cancer.*, **61,** 932-936.

Berridge, M. (1993). Inositol triphosphate and calcium signalling. *Nature.*, **387,** 673-676.

Bestor, T. (1998). Methylation meets acetylation. *Nature.*, **393,** 311-312.

Birchmeier, W., Beherens, J., Weidner, K.M., Frixen, U.H. & Schipper, J.. (1991). Dominant and recessive genes involved in tumour cell invasion. *Curr. Opin. Cell. Biol.*, **3,** 832-840

Birkenmeier, E., Schneider, U. & Thurston, S. (1992). Fingerprinting genomes by use of PCR with primers that encode protein motifs or contain sequences that regulate gene expression. *Mammalian Genome.*, **3,** 537-545.

Boguski, M. and McCormick, F. (1993). Proteins regulating Ras and its

relatives. *Nature.*, **366**, 643-654.

Bollag, G. and McCormick, F. (1991). Differential regulation of rasGAPs and the neurofibromatosis gene product activities. *Nature.*, **351**, 576-580.

Bos, J.L. (1988). The ras family and human carcinogenesis. *Mutation Research*, **195**, 255-271.

Bos, J.L. (1989). Ras oncogenes in human cancer: a review. *Cancer Res.*, **49**, 4692-4699.

Bottger, T.C., Storkel, S., Wellek, S., Stockle, M. & Junginger, T. (1993). Factors influencing survival after resection of pancreatic cancer. *Cancer.*, **73**, 63-73.

Branch, P., Hampson, R. & Karran, P. (1995). DNA mismatch binding defects, DNA damage tolerance, and mutator phenotypes in human colorectal carcinoma cell lines. *Cancer Res.*, **55**, 2304-2369.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. & Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature.*, **371**, 435-438.

Bregula, U., Miller, O.J. & Harris, H. (1971). The analysis of malignancy by cell fusion II. *J. Cell. Sci.*, **8**, 673-680.

Brown, E. (1997). Adhesive interactions in the immune system. *Trends in Cell Biology.*, **7**, 289-295.

Caldas, C., Hahn, S.A., Da Costa, L.T.,Redston, M.S., Schutte, M., Seymour, A.B., Weinstein, C.L., Hruban, R.H., Yeo, C.J. & Kern, S.E. (1994a). Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat. Genet.*, **8**, 27-31.

Caldas, C., Hahn, S.A., Hruban, R.H., Redston, M.S., Yeo, C. J. & Kern, S. E. (1994b). Detection of K-ras mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. *Cancer Res.*, **54**, 3568-3573.

Carraway, K.I. & Cantley, L. (1994). A neu acquaintance for erbB3 and erbB4: a role for receptor heterodimerisation in growth signalling. *Cell.*, **78**, 5-8.

Cawthon, R.M., Weiss, R., Xu, G., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P. & White, R.A. (1990). A major segment of the Neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell.*, **62**, 193-201.

Chattopadhyay, P., Rathore, A., Mathur, M., Sarkar, C., Mahapatra, A.K. & Sinha, S. (1997). Loss of heterozygosity of a locus on 17p13.3, independant of p53, is associated with higher grades of astrocytic tumours. *Oncogene.*, **15**, 871-874.

Chen, J., Merechal, V. & Levine, A.J. (1993). Mapping of p53 and mdm-2 interaction domains. *Mol. Cell. Biol.*, **13**, 4107-4114.

Chen, X., Rubock, M.J. & Whitman, M. (1996). A transcriptional partner for MAD protein in TGF-beta signalling.*Nature.*, **383**, 691-696.

Cheng, J., Hiemstra, J., Schneider, S., Naumova, A. & Cheung, NK. (1993). Preferential amplification of the paternal allele of the N-Myc gene in human neuroblastomas. *Nat. Genet.*, **4**, 191-194.

Chowdary, D.R., Dermody, J.J., Jha, K.K. & Ozer, H.L. (1994). Accumulation of p53 in a mutant cell line deficient in ubiquitin pathway. *Mol. Cell. Biol.*, **14**, 1997-2003.

Cogen, P.H., Daneshvar, L. & Metzger, A.K. (1992). Involvement of multiple

chromosome 17p loci in medulloblastoma tumourigenesis. *Am. J. Hum. Genet.*, **50,** 584-589.

Coles, C., Thompson, A., Elder, P., Cohen, B.B., Mackenzie, IM., Cranston, G., Chetty, U., Mackay, J., Macdonald, M., Nakamura, Y., Hoyheim, B. & Steel, C.M. (1990). Evidence implicating at least two genes on chromosome 17p in breast carcinogenesis.*Lancet.*, **336,** 761-763.

Collins, F.S. (1991). Of needles and haystacks: finding human disease genes by positional cloning. *Clin. Res.*, **39,** 615-623

Cooper, D. & Youssoufian, H. (1988). The CpG dinucleotide and human genetic disease. *Human Genetics.*, **11,** 185-188.

Cornelis, R.S., van Vliet, M., Vos, C.B.J., Cleton-Jansen, A.M., van de Vijver, M.J., Peterse, J.L., Khan, P.M., Borrensen, A.L., Cornelisse, C.J. & Devilee, P. (1994). Evidence for a gene on 17p13.3, distal to TP53, as a target for allele loss in breast tumours without TP53 mutations. *Cancer Res.*, **54,** 4200-4206.

Craig, J. & Bickmore, W. (1994). The distribution of CpG islands in mammalian chromosomes. *Nat. Genet.*, **7,** 376-381.

Cross, S., Charlton, J., Xinsheng, N. & Bird, A. (1994). Purification of CpG islands using a methylated DNA binding column. *Nat. Genet..*,**6,** 236-244.

Csordas, M. (1990). The biological role of histone acetylation. *Biochem. J.*, **265,** 23-38.

Cubilla, A., Fortner, J. & Fitzgerald, P. (1978). Lymph node involvement in carcinoma of the head of the pancreas. *Cancer.*, **41,** 880-887.

de Camilli, P., Emr, S. & McPherson, P.N.P.(1996). Phosphoinositides as regulators in membrane traffic. *Science.*, **271,** 1533-1539.

Deininger, P., Batzer, M., Hutchison, C. & Edgel, M. (1992). Master genes in mammalian repetitive DNA amplification. *TIG*, **8**, 307-311.

Di Leonardo, A., Linke, S., Clarkin, K. & Wahl, G. (1994). DNA damage triggers a prolonged p53-dependent G1 arrest and long term induction of CIP1 in normal human fibroblasts. *Genes & Dev.*, **8**, 2540-2551.

Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgommery, C.A. Jr., Butel, J.S. & Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature.*, **356**, 215-221.

Drickamer, K. (1993). Ca2+-dependant carbohydrate-recognition domains in animal proteins. *Curr. Opinion. Struct. Biol.*, **3**, 393-400.

El-Deiry, W.S., Harper, J.W., O'Connor, P.M., Velculescu, V.E., Canman, C.E. Jackman, J. & Trent, J.M. (1994). WAF1/CIP1 is induced in p53 mediated G1 arrest and apoptosis. *Cancer Res.*, **54**, 1169-1174.

Evans, J.P., Burcke, W., Chen, R., Bennett, R.L., Dellinger, E.P., Kimmey, M., Crispin, D., Brentnall, T.A. & Byrd, D.R. (1995). Familial pancreatic adenocarcinoma: association with diabetes and early molecular diagnosis. *J. Med. Genet.*, **32**, 330-335.

Fearon, E.R. (1997). Human cancer syndromes: Clues to the origin and nature of cancer. *Science.*, **278**, 1043-1050.

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

Feinberg, A., Gehrke, C., Kuo, K. & Ehrlich, M. (1988). Reduced genomic 5-

methylcytosine content in human colonic neoplasias. *Cancer Res.*, **48**, 1159-1161.

Feinberg, A. & Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature.*, **310**, 89-91.

Feinberg, A.P. & Vogelstein, B. (1984). A Technique for the radiolabelling of DNA restriction endonuclease fragments to high specific activities. *Anal. Biochem.*, **137**, 266-267.

Fernandez-del castillo, C. & Warshaw, A.L. (1994) Pancreatic carcinoma. *Curr. Opin. Gastroenterol.*, **10**, 507-512.

Fill, W.L., Lameill, J.M. & Polk, N.O. (1979). The radiographic manifestations of Von Hippel-Lindau syndrome. *Radiology.*, **133**, 289-295.

Flanders, T.Y. & Foulkes, W.D. (1996). Pancreatic adenocarcinoma: epidemiology and genetics. *J. Med. Genet.*, **33**, 889-898.

Francis, F., Zehetner, G., Hoglund, M. & Lehrach, H. (1994). Construction and preliminary analysis of the ICRF human P1 library. *GATA.*, **11**, 148-157.

Friess, H., Yamanaka, T., Kobrin, M., Do, D.A., Buchler, M.W. & Korc, M. (1995). Enhanced erbB3 expression in human pancreatic cancers correlates with tumour progression. *Clin. Cancer Res.*, **1**, 1413-1420.

Fueyo, J., Gomez-Manzano, C., Bruner, J., Saito, Y., Zhang, B., Zhang, W., Levin, V.A., Yung, W.K.A. & Kyritis, AP. (1998). Hypermethylation of the CpG island of p16/CDKN2 correlates with gene inactivation in gliomas. *Oncogene.*, **13**, 1615-1619.

Fujii, H., Egami, H., Chaney, W., Pour, P. & Pelling, J. (1990). Pancreatic ductal

carcinomas induced in Syrian hamsters by N-nitrosobis (2-oxopropyl)amine contain a Ki-ras oncogene with point-mutated codon 12. *Carcinogenesis.*, **3**, 296-301.

Fukushige, S., Waldman, F.M., Kimura, M. & Abe, T. (1997). Frequent gain of copy number on the long arm of chromosome 20 in human pancreatic adenocarcinoma. *Genes, Chromosomes & Cancer.*, **19**, 161-169.

Funatomi, H., Itakura, J., Ishiwata, T., Pastan, I., Thompson, S.A., Johnson, G.R. & Korc, M. (1997). Amphiregulin anitsense oligonucleotides inhibit the growth of T3M4 human pancreatic cancer cells and sensitizes cells to EGF receptor-targeted therapy. *Int. J. Cancer.*, **72**, 512-517.

Goldstein, A.M., Fraser, M.C., Struewing, J.P., Hussassian, C.J., Ranade, K., Zametkin, D.P., Fontaine, L.S., Organic, S.M. & Dracopoli, N.C. (1995). Increased risk of pancreatic cancer in melanoma-prone kindreds with p16/INK4A mutations. *N. Eng. J. Med.*, **333**.

Gospodarowicz, D., Neufeld, G. & Schweigerer, L. (1986). Molecular and biological characterisation of the fibroblast growth factor. *Cell Differ.*, **19**, 1-17.

Greenblatt, M.S., Bennett, W., Hollstein, M. & Harris, C.C. (1994). Mutations in the p53 tumour suppressor gene : Clues to cancer etiology and molecular pathogenesis. *Cancer Res.*, **54**, 4855-4878.

Gress, T., Muller-Pillasch, F., Geng, M., Zehetner, G., Lehrach, H. & Adler, G. (1995). Molecular genetic resources for pancreatic cancer research. In *Pancreatic cancer: molecualr and clinical advances*, Neoptolemos, J. P. and Lemoine, N.R., (ed), Vol. 1. pp. 133-141. Blackwell: Oxford.

Gress, T.M., Muller-Pillasch, F., Geng, M., Zimmerhackl, F., Zehetner, G., Friess, H., Buchler, M., Adler, G. & Lehrach, H. (1996). A pancreatic cancer-

specific expression profile. *Oncogene.*, **13,** 1819-1830.

Griffin, C.A., Hruban, R.H., Morsberger, L.A., Ellingham, T., Long, P. P., Jaffee, E. M., Hauda, K. M., Bohlander, S. K. & Yeo, C. T. (1995). Consistent chromosome abnormalities in adenocarcinoma of the pancreas. *Cancer Res.*, **55,** 2394-2399.

Gumbiner, B. (1996). Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell.*, **84,** 345-357.

Hahn, S.A., Hoque, A.T., Moskaluk, C.A., da Costa, L.T., Schutte, M., Rozemblum, E., Seymour, A.B., Weinstein, C.L., Yeo, C.T., Hruban, R.H. & Kern, S.E. (1996a). Homozygous deletion map at 18q21.1 in pancreatic cancer. *Cancer Res.*, **56,** 490-494.

Hahn, S.A., Schutte, M., Hoque, A.T., Moskaluk, C.A., da Costa, L.T., Schutte, M., Rozemblum, E., Seymour, A.B., Weinstein, C.L., Yeo, C.T., Hruban, R.H. & Kern, S.E. (1996b). DPC4: A candidiate tumour suppressor gene at human chromosome 18q21.1. *Science.*, **271,** 350-353.

Hahn, S.A., Seymour, A.B., Shamsul-Hoque, A.T.M., Schutte, M., Da Costa, L. T., Redston, M. S., Caldas, C., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H. & Kern, S. E. (1995). Allelotype of pancreatic adenocarcinoma using xenograft enrichment. *Cancer Res.*, **55,** 4670-4675.

Halevy, O., Novitch, B.G., Spicer, D.B., Skapek, S.X., Rhee, J., Hannon, G.J., Beach, D. & Lassar, A.B. (1995). Correlation of terminal cell cycle arrest of skeletal muscle with induction of p21 by MyoD. *Science.*, **267,** 1018-1021.

Hall, P., Hughes, C., Staddon, S., Richman, P.I., Gullick, W.J. & Lemoine, N.R. (1990). The c-erbB2 proto-oncogene in human pancreatic cancer. *J. Pathol.*, **161,** 195-200.

Hanada, M., Delia, D., Aiello, A., Stadmauer, E. & Reed, J.C. (1993). Bcl-2 gene hypomethylation and high level expression in lymphocytic leukemia. *Blood.*, **82,** 1820-1828.

Harris, H., Miller, O.J., Klein, G., Worst, P. & Tachibana, T. (1969). Suppression of malignancy by cell fusion. *Nature.*, **223**.

Hibi, K., Nakamura, H., Hirai, A., Fujikake,Y. & Kasai, Y. (1996). Loss of H19 imprinting in esophageal tumours. *Cancer Res.*, **56,** 480-482.

Holton, T.A. & Graham, M.W. (1990). A simple and efficient method for the direct cloning of PCR products using ddT-tailed vectors. *Nucl. Acids. Res.*, **19,** 1156.

Horikoshi, N., Usheva, A., Chen, J., Levine, A.J., Weinmann, R. & Shenk, T. (1995). Two domains of p53 interact with the TATA-binding domain protein, and adenovirus 13S E1A protein disrupts the association. *Mol. Cell. Biol.*, **15,** 227-234.

Hou, J., Adams, P., Wang, F., Fast, D. & McKeehan, W. (1994). Molecular cloning and expression of the gene for a Leucine-rich protein from human hepatobastoma cells. *In Vitro Cell. Dev. Biol.*, **30A,** 111-114.

Hruban, R., van Mansfield, A., Offerhaus, G.,van Weering, D.H., Allison, D.C., Goodman, S.N., kensler, T.W., Bose, K.K., Cameron, J.L. & Bos, J.L. (1993). Ki-ras oncogene activation in adenocarcinoma of the human pancreas: a study of 82 carcinomas using a combination of mutant-enriched PCR analysis and allele-specific oligonucleotide hybridisation. *Am. J. Pathol.*, **143,** 545-554.

Hull, R. & Will, H. (1989). Molecular biology of viral and non-viral retroelements. *TIG.*, **5,** 357.

Humphries, M. & Newham, P. (1998). The structure of cell-adhesion molecules. *Trends in Cell Biology.*, **8,** 78-83.

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D. & Perucho, M. (1993). Ubiquitous somatic mutations in si,ple repeated sequences reveal a new mechanism for cholonic carcinogenesis. *Nature.*, **10,** 558-563.

Iounnou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M.A. & de Jong, P. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA inserts. *Nat. Genet.*, **6,** 84-89.

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

Iwamura, T., Katsuki, T. & Ide, K. (1987). Establishment and characterisation of a human pancreatic cancer cell line (SUIT-2) producing carcinoembryonic antigen and carbohydrate antigen 19-9. *Jpn J. Cancer Res.*, **78,** 54-62.

Jaenisch, R. (1997a). DNA methylation and imprinting: why bother? *TIG*, **13,** 323-329.

Jaenisch, R. (1997b). DNA methylation and imprinting:why bother. *TIG*, **13,** 323328.

Johansson, B., Bardi, G., Heim, S., Mandahl, N., Mertens, F., Bak-Jensen, E., Andren-Sandberg, A. & Mitelman, F. (1992). Non-random chromosomal rearrangements in pancreatic carcinoma. *Cancer.*, **69,** 1674-1681.

Jones, P. (1996). DNA methylation errors and cancer. *Cancer Res.*, **56,** 2463-2467.

Jones, P., Rideout, W., Shen, J.C., Spruck, C.H. & Tsai, Y.C. (1992). Methylation,

mutation and cancer. *BioEssays.*, **14,** 33-36.

Kamb, A. (1994). Sun protection factor p53. *Nature.*, **372,** 730-1.

Kamb, A., Gruis, N.A., Weaver-Feldhaus, J., Liu, Q., Harshman, K., Tavigian, S.V., Stockert, E., Day, R.S., Johnson, B.E. & Skolnick, M.H. (1994). A cell cycle regulator potentially involved in the genesis of many tumour types. *Science.*, **264,** 436-440.

Kass, S., Pruss, D. & Wolffe, A. (1997). How does methylation repress transcription. *TIG.*, **13,** 444-449.

Kautiainen, T. & Jones, P. (1986). DNA methylation levels in tumorigenic and nontumorigenic cells in culture. *J. Biol. Chem.*, **261,** 1594-1598.

Keim, V., Iovanna, J.-L., Orelle, B., Verdier, J-M., Busing, M., Hopt, U. & Dagorn, J-C. (1992). A novel exchange protein associated with pancreas transplantation in humans. *Gastroenterology.*, **103,** 248-254.

Kern, S., Pietenpol, J., Thiagalingam, S., Seymour, A.B., Kinzler, K. & Vogelstein, B. (1992). Oncogenic forms of p53 inhibit p53-regulated gene expression. *Science.*, **256,** 827-830.

Klein, V., Piontek, K., Brass, N., Subke, F., Zang, K. D. & Meese, E. (1993). Identification of chromosome-specific sequence-tagged sites by Alu-PCR. *GATA.*, **10,** 6-9.

Kojis, T.l., Heinzmann, C., Flodman, P., Ngo, J.T., Sparkes, R.S., Spence, M.A., Bateman, J.B. & Heckenlively, J.R. (1996). Map refinement of Locus RP13 to human chromosome 17p13.3 in a second family with autosomal dominant retinitis pigmentosa. *Am. J. Hum. Genet.*, **58,** 347-355

Kondo, M., Suzuki, H., Ueda, R., Osada, H., Takagi, K. & Takahashi, T. (1995).

Frequent loss of imprinting of the H19 gene is often associated with its overexpression in human lung cancers. *Oncogene.*, **10,** 1193-1198.

Korc, M., Meltzer, P. & Trent, J. (1986). Enhanced expression of epidermal growth factor receptor correlates with alterations of chromosome 7 in human pancreas cancer. *Proc. Natl. Acad. Sci. USA.*, **83,** 5141-5144.

Krajinovic, M., Richter, C., Labuda, D. & Sinnett, D. (1996). Detection of a mutator phenotype in cancer cells by inter-Alu PCR. *Cancer Res.*, **56,** 2733-2737.

Kraus, M., Issing, W., Miki, T.,Popescu, N.C. & Aaronson, S.A. (1989). Isolation and characterisation of ErbB3, a third member of the ERBB/epidermal growth factor family. *Proc. Natl. Acad. Sci. USA.*, **86,** 9193-9197.

Krontiris, T.G. & M.C.G. (1981). Transforming activity of human tumour DNAs. *Proc. Natl. Acad. Sci. USA.*, **78,** 1181-1184.

Ladanyi, M., Cha, C., Lewis, R., Jhanwar, S.C., Huvos, A.G. & Healey, J.H. (1993). MDM2 gene amplification in metastatic osteosarcoma. *Cancer Research*, **53,** 16-18.

Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massague, J. (1996). Partnership between DPC4 and Smad proteins in TGF-β signalling pathways. *Nature.*, **383,** 832-836.

Laird, P., Jackson-Grusby, L., Fazell, A., Dickenson, S., WE, J., Li, E., Weinberg, R. & Jaenisch, R. (1995). Suppression of interstitial neoplasia by DNA hypomethylation. *Cell.*, **81,** 197-205.

Laird, P. & Jaenisch, R. (1996). The role of DNA methylation in cancer genetics and epigenetics. *Annu. Rev. Genet.*, **30,** 441-464.

Lane, D.P. (1992). p53, Guardian of the genome. *Nature.*, **358,** 15-16.

Lane, D.P. and Hall, P.A. (1997). MDM2 - arbiter of p53's destruction. *Trends Biochem. Sci..*, **22**, 372-374.

Lasserre, C., Simon, M., Ishikawa, H.,Diriong, S., Nguyen, V.C., Christa, L., Vernier, P. & Brechot, C (1994). Structural organisation and chromosomal localisation of a human gene (HIP/PAP) encoding a C-type lectin overexpressed in primary liver cancer. *Eur. J. Biochem.*, **224**, 29-38.

Lemoine, N.R., Hughes, C., Barton, C.M., Poulsom, R., Jeffrey, R.E., Kloppel, G., Hall, P.A. & Gullick, W.J. (1992a). The epidermal growth factor receptor in human pancreatic cancer. *J. Pathol.*, **166**, 7-12.

Lemoine, N.R., Lobresco, M., Leung, H., Barton, C.M., Hughes, C., Prigent, S.A., Gullick, W.J. & Kloppel, G.(1992b). The erbB3 gene in human pancreatic cancer. *J. Pathol.*, **168**, 269-273.

Leung, H., Brown, R., Mckie, A.B., Sirivataneuksorn V. & Neal, D. (1999). Alu-PCR genotype analysis in prostate cancer. Manuscript in preparation.

Leung, H., Gullick, W. & Lemoine, N. (1994a). Expression and functional activity of the fibroblast growth factors and their receptors in human pancreatic cancers. *Int. J. Cancer.*, **59**, 6670675.

Leung, H., Hughes, C., Kloppel, G.,Williamson, R.C.N. & Lemoine, N.R. (1994b). Localisation of expression of the fibroblast growth factors and their receptors in pancreatic adenocarcinoma by in-situ hybridisation. *Int. J. Oncol.*, **4**, 1219-1223.

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

Liang, P. & Pardee, A.B. (1992). Differential display of eukaryotic messenger

RNA by means of the polymerase chain reaction. *Science.*, **257,** 967-971.

Lisitsyn, N., Lisitsyn, N. & Wigler, M. (1993). Cloning the differences between two complex genomes. *Science.*, **259,** 946-951.

Lo, K.W., Cheung, S.T., Leung, S.F., van Hasselt, A., Tsang, Y.S., Mak, K.f., Chung, Y.F., Woo, J.K., Lee, J.C. & Huang, D.P. (1996). Hypermethylation of the p16 gene in nasopharyngeal carcinoma. *Cancer Res.*, **56,** 2721-2725.

Lynch, H.T. & Fusaro, R.M. (1991). Pancreatic cancer and the familial atypical multiple mole (FAMMM) syndrome. *Pancreas.*, **6,** 127-131.

Lynch, H.T., Voorhees, G.J., Lanspa, S.J., McGreevy, P.S. & Lynch, J.F. (1985). pancreatic carcinoma and hereditary non-polyposis colorectal cancer: a family study. *Br. J. Cancer.*, **52,** 271-273.

Malkin, D., Li, F.P., Strong, L.C. Fraumeni, J.F.Jr., Nelson, C.E., Kim, D.H., Kassel, J., Gryka, M.A., Bischolf, F.Z., Tainsky, M.A. (1990). germline p53 mutations in a familial syndrome of breast cancer, sarcomas and other neoplasms. *Science.*, **250**. 1233-1238.

Mariyama, M., Kishi, K., Nakamura, K., Oata, H. & Nishimura, S. (1989). frequency and types of point mutation at the twelfth codon of the c-K-ras gene found in pancreatic cancers in Japanese patients. *Jpn. J. Cancer Res.*, **80,** 622-626.

McKie, A.B., Bhatia, A., Hurst, H.C. & Lemoine, N.R. (1995). Nucleic acid fingerprinting for the isolation of novel cancer-associated genes. In *Pancreatic cancer: Molecular and clinical advances*, Neoptolemos, J.P. & Lemoine, N.R. (eds). 142-151. Blackwell Science Ltd.: Oxford.

McKie, A.B., Filipe, M.I. & Lemoine, N.R. (1993). Abnormalities affecting the APC and MCC tumour suppressor gene loci in chromosome 5q occur frequently in gastric cancer but not in pancreatic cancer. *Int. J. Cancer.*, **55,** 598-603.

McKie, A.B., Iwamura, T., Leung, H., Hollingsworth, M.A. & Lemoine, N.R. (1997). Alu-Polymerase Chain Reaction genomic fingerprinting technique identifies multiple genetic loci associated with pancreatic tumourigenesis. *Genes, Chromosomes & Cancer..*,**18,** 30-41.

Meehan, R., Lewis, J. & Bird, A. (1992). Transcriptional repression by methylation of CpG. *Nucl. Acids. Res..*, **20,** 5085-5092.

Menegazzi, M., Scarpa, A. & Libonati, M. (1988). Analysis of the methylation pattern of H-ras oncogene in human prostatic cancer. *Ital. J. Biochem.*, **37,** 104-110.

Merlo, G.R., Venesio, T., Bernardi, A., Cropp, C.S., Diella, F., Cappa, A.P., Callahan, R. & Liscia, D.S. (1994). Evidence for a second tumour suppressor on 17p linked to high S-phase index in primary breast carcinoma. *Cancer Genet. Cytogenet.*, **76,** 106-111.

Mikulski, S., Viera, A., Darzynkiewicz, Z. & Shogen, K. (1992). Synergism between a novel amphibian oocyte ribonuclease and lovastatin in inducing cytostatic and cytotoxic effects in human lung and pancreatic carcinoma cell lines. *Br. J. Cancer.*, **66,** 304-310.

Nagai, M.A., Pacheco, M.M., Brentani, M.M., Marques, L. A., Brentani, R. R., Ponder, B. A. J. & Mulligan, L. M. (1994). Allelic loss on distal chromosome 17p is associated with poor prognosis in a group of Brazilian breast cancer patients. *Br. J. Cancer*, **69,** 754-758.

Nan, X., Tate, P., Li, E. & Bird, A. DNA methylation specifies chromosomal location of MeCP2. (1996). *Mol. Cell. Biol.*, **16,** 414-421.

Neiderhuber, J.E., Brennan, M.F. & Menck, H.R. (1995). Database report on pancreatic cancer. *Cancer.*, **76,** 1671-1677.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, Maureen F., Ramirez-Solis, R., Webster, T. D., Ledbetter, D.H. & Caskey, T.C. (1989). Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA.*, **86,** 6686-6690.

Neoptolemos, J. & Lemoine, N. (1995). Pancreatic Cancer: Molecular and clinical advances. *Pancreatic Cancer: Molecular and clinical advances,*. Neoptolemos, J.P. & Lemoine, N.R. (eds). 4-5. Blackwell Science Ltd.: Oxford.

Nobori, T., Miura, K., Lois, D.J., Takabayashi, K. & Carson, D.A. (1994). Deletions of the cyclin-dependant kinase-4 inhibitor gene in multiple human cancers. *Nature.*, **368,** 753-756.

Ogata, T., Ayusawa, D., Namba, M., takahashi, E., Oshimura, M. & Oishi, M. (1993). Chromosome 7 suppresses division of nontumourigenic immortilized human fibroblast cell lines KMST-6 and SUSM-1. *Mol. Cell. Biol.*, **13**, 6036-6043.

Oliner, J.D., Kinzler, K.W., Meltzer, P.S., George, D.L. & Vogelstein, B. (1992). Amplification of a gene encoding a p53-associated protein in human sarcomas. *Nature..*, **358,** 80-83.

Oliva, R., Bazett-Jones, D., Locklear, L. & Dixon, G. (1990). Histone acetylation can induce unfolding of the nucleosome core particle. *Nucl. Acids Res.*, **18,** 2739-2747.

Orita, M., Suzuki, Y., Sekiya, T. & Hayashi K. (1989). Rapid and sensitive detection of point mutations and DNA polynorphisms using the polymerase chain reaction. *Genomics*, **7**, 874-879.

Padgett, R.W., Das, P. and Krishna, S. (1998). TGF-b signaling, Smads, and tumour suppressors. *BioEssays*, **20**, 382-391

Parker, S., Eichele, G., Zhang, P., Rawls, A., Sands, A.T., Bradley, A., Olsen, E.N., Harper, J.W. & Elledge, S.J. (1995). P53-independant expression of p21 in muscle and other terminally differentiating cells. *Science.*, **267,** 1024-1027.

Perry, M. & Chalkley, R. (1981). The effects of histone hyperacetylation on nuclease sensitivity and the solubility of chromatin. *J. Biol. Chem.*, **256,** 3313-3318.

Phelan., M. C., Lancaster, J.M., Tonin, P., Gumbs, C., Cochran, C., Carter, R., Ghadirian, P., Perret, C., Moslehi, R., Dion, F., Faucher, M.C., Dole, K., Karimi, S., Foulkes, W., Lounis, H., Warner, E., Goss, P., Anderson, D., Larsson, C., Narod, S.A. & Futreal, P.A. (1996). Mutation analysis of the BRCA2 gene in 49 site-specific breast cancer families. *Nature Genet.*, **13,** 120-122.

Phillips, N.J., Ziegler, M.R., Radford, D.M., Kelly L., Steinbrueck, T., Xynos, F. P. & Donis-Keller, Helen. (1996). Allelic deletion on chromosome 17p13.3 in early ovarian cancer. *Cancer Res.*, **56,** 606-611.

Pierce, J.C., Sauer, B. & Sternberg, N. (1992). A positive selection vector for cloning high molecular weight DNA by the bactreiophage P1 system. *Proc. Natl. Acad. Sci USA.*, **89,** 2056-2060.

Pignatelli, M., Ansari, T., Gunter, P., Liu, D., Hirano, S., Takeichi, M., Kloppel, G. & Lemoine, N.R. (1994). Loss of membranous E-cadherin expression in pancreatic cancer-correlation with lymth node metastasis, high grade and advanced stage. *J. Pathology.*, **174,** 243-248.

Plowman, G., Culuscou, J., Whitney, G., Green, J.M., Carlton, G.W., Foy, L., Neubauer, M.G. & Shoyab, M. (1993). Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal growth factor receptor family. *Proc. Natl. Acad. Sci. USA.*, **90,** 1746-1750.

Pollock, P.M., Pearson, J.V. & Hayward, N.K. (1996). Compilation of somatic mutations of the CDKN2 gene in human cancers: Non-random distribution of base substitutions. *Genes, Chromosomes and Cancer.*, **15,** 77-88.

Poston, G.J., Gillespie, J. & Guillou, P.J. (1991). Biology of pancreatic cancer. *Gut.*, **32,** 800-812.

Pour, P., Konishi, Y., Kloppel, G. & Longnecker, D. (1994). Atlas of exocrine pancreatic tumours. In *Morphology, biology and diagnosis with an international guide for tumour classification*, Group, I.P.C.S. (ed) pp. 221-235. Springer-Verlag: Tokyo.

Queimado, L., Reis, A., Fonseca, I., Martins, C., Lovett, M., Soares, J. & Parreira, L. (1998). A refined Localisation of two deleted regions in chromosome 6q asociated with salivary gland carcinomas. *Oncogene.*, **16,** 83-88.

Quelle, D.E. (1995). Alternative reading frames for INK4A tumour suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell*, **83**, 993-1000.

Ray, J., Harbison, M., McClain, R. & Goodman, J. (1994). Alterations in the methylation status and expression of the raf oncogene in phenobarbitol-induced and spontaneous B6C3F1 mouse liver tumours. *Mol. Carcinog.enesis.*, **9,** 155-166.

Razin, A. & Cedar, H. (1994). DNA methylation and genomic imprinting. *Cell.*, **77**, 473-476.

Razin, A., Szyf, M., Kafri, T., Roll, M., Giloh, H., Scarpa, S., Carotti, D. & Cantoni, G.L. (1986). Replacement of 5-methylcytosine by cytosine : a possible mechanism for transient DNA demethylation. *Proc. Natl. Acad. Sci. USA.*, **83,** 2827-2831.

Redston, M.S., Caldas, C., Seymour, A.B., Hruban, R.H., Kinzler, K.W., Yeo, C.T. & Kern, S.E. (1994). P53 abnormalities in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. *Cancer Res.*, **54,** 3025-3033.

Riggs, M., Whittaker, R., Neumans, J. & Ingram, V. (1977). Nucleosome-associated proteins and phosphoproteins of differentiating friend erythroleukemia cells. *Nature.*, **268**, 462-464.

Rimm, D., Koslov, E., Kebriaei, P., Cianci, C.D. & Morrow, J.S. (1995). Catenin is an actin-binding and bundling protein mediating the attachment of F-actin to the cell adhesion complex. *Proc. Natl. Acad. Sci. USA.*, **92,** 8813-8817.

Rivenson, A., Hoffman, D., Prokopczyk B., Amin, S. & Hecht, S.S. (1988). Induction of lung and exocrine pancreas tumours in F344 rats by tobacco-specific and Areca-derived N-nitrosamines. *Cancer Res.*, **48,** 6912-6917.

Robles, S. & Adami, G. (1998). Agents that cause DNA double-strand breaks lead to p16/INK4a enrichment and the premature senescence of normal fibroblasts. *Oncogene.*, **16,** 1113-1123.

Rozenblum, E., Schutte, M., Goggins, M., Hahn, S.A., Hoque, A.T., Moskaluk, C.A., da Costa, L.T., Seymour, A.B., Weinstein, C.L., Yeo, C.T., Hruban, R.H. & Kern, S.E. (1997). Tumour-suppressive pathways in pancreatic carcinoma. *Cancer Res.*, **57** 1731-1734.

Ruggeri, B.A., Huang, L., Berger, D., Chang, H., Klein-Szanto, A J.P., Goodrow, T., Wood, M., Obara, T., Heath, C.W. & Lynch, H.T. (1997). Molecular pathology of primary and metastatic ductal pancreatic lesions. *American Cancer Society.*, **79,** 700-716.

Schutte, M., Da costa, L.T., Hahn, S.A., Moskaluk, C., Shamsul-Hoque, A.T.M.,

Rozenblum, E., Weinstein, C. L., Bittner, M., Meltzer, P.S.,Trent, J. M.,Yeo, C. J., Hruban, R. H. & Kern, S. E. (1995). Identification by representational difference analysis of a homozygous deletion in pancreatic carcinoma that lies within the BRCA2 region. *Proc. Natl. Acad. Sci. USA.*, **92,** 5950-5954.

Schutte, M., Hruban, R.R., Hedrick, L., Cho, K.R., Nadasdy, G.M., Weinstein, C.L., Bova, G.S., Isaacs, W.B., Cairns, P., Nawroz, H., Sidransky, D., Casero, R.A. Jr, Meltzer, P.S., Hahn, S.A. & Kern, S.E. (1996). DPC4 gene in various tumour types. *Cancer Res.*, **56**, 2527-2532.

Schutte, M. & Kern, S.E. (1995). The molecular genetics of pancreatic adenocarcinoma. In *Pancreatic cancer: Molecular and clinical advances*, Neoptolemos, J.P. & Lemoine, N.R. (eds). 115-131. Blackwell Science Ltd.: Oxford.

Schultz, D.C., vanderveer, L., Berman, D.B., Hamilton, T.C., Wong, A.J. & Godwin, A.K. (1996). Identification of two candidate tumour suppressor genes on chromosome 17p13.3 *Cancer Res.*, **56**, 1997-2002.

Sealey, P.G., Whittaker, P.A. & Southern, E.M. (1985). Removal of repeated sequences from hybridisation probes. *Nucl. Acids Res.*, **13,** 1905-1922.

Serrano, M., Lee, H.W., Chin, L., Cordon-Cardo, C., Beach, D. & DePinho, R.A. (1996). Role of INK4a locus in tumour suppression and cell mortality. *Cell.*, **85,** 27-37.

Seymour, A.B., Hruban, R.H., Redston, M.S. & Yeo, C. J. (1994). Allelotype of pancreatic adenocarcinoma. *Cancer Res.*, **54,** 2761-2764.

Sherr, C.J. (1996). Cancer cell cycles. *Science.*, **274**, 1672-1676.

Shih, C., Padhy, L.C., Murray, M., Weinberg, R.A. (1981). Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature*, **290**, 261-264,

Shizuya, H. (1992). Cloning and stable maintainance of 300 kb fragments of human DNA in E.Coli using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA.*, **89,** 8794-8797.

Singer M, B.P. (1991). In *Genes and Genomes. 1991.* University Science Books, Mill Way, CA.

Smit, A.F.A. (1996). Origins of interspersed repeats.*Curr. Opin. Genet. Dev.*, **6,** 743-748.

Speicher, M. (1995). Microsatellite instability in human cancer. *Oncol. Res.*, **7,** 267-275.

Srivastava, S., Zou, Z., Pirollo, K., Blattner, W. & Chang, E.H. (1990). Germline transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature.*, **348,** 747-749.

Stack, M., Jones, D., White, G., Liscia, D.,Venesio, T., Casey, G., Crichton, D., Varley, J., Mitchell, E., Heighway, J. & Santibanez-Koref, M. (1995). Detailed mapping and loss of heterozygosity analysis suggests a suppressor locus involved in sporadic breast cancer within a distal region of chromosome band 17p13.3. *Hum. Mol. Genet.*, **4,** 2047-2055.

Steenman, M., Rainier, S., Dobry, C., Grundy, P., Horon, I.L. & Feinberg, A.P. (1994). Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilm's tumour. *Nat. Genet.*, **7,** 433-439.

Strachan, T. & Read, A. (1996). Human multigene families and repetitive DNA. Chapter 8, *Hum. Mol.Genet.* Vol. 1. BIOS Scientific publishers.

Takeda, J., Yano, H., Eng, S., Zeng, Y. & Bell, G.I. (1993). A molecular inventory of human pancreatic islets: sequence analysis of 1000 cDNA clones. *Human Mol. Genetics.*, **2,** 1793-1798.

Tang, D. & Honn, K. (1994). Adhesion molecules and tumour metastasis: an update. *Invasion Metastasis.*, **95,** 109-122.

Taniguchi, S., Iwamura, T. & Katsuki, T. (1992). Correlation between spontaneous metastatic potential and type I collagenolytic activity in a human pancreas cancer cell line (SUIT-2) and sub-lines. *Clin. Exp. Metast.*, **10,** 259-266.

Temin, H. & Mizutani, S. (1970). RNA-dependant DNA polymerase in virions of Rous Sarcoma virus. *Nature.*, **226,** 1211.

Terazono, K., Yamamoto, H., Takasawa, S., Shiga, K., Yonemura, Y., Tochino, T. & Okamoto, H. (1988). A novel gene activated in regenerating islets. *J. Biol. Chem.*, **263,** 2111-2114.

Thiagalingam, S., Lengauer, C., 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,** 343-349.

Thibodeau, S., Bren, G. & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science.*, **363,** 558-561.

Thomas, G.M.H., Cunningham, E., Fensome. A., Ball, A. & Totty, N. F. (1993). An essential role for PITPn in phospholipase C-mediated inositol signalling. *Cell.*, **74,** 919-928.

Thorlacius, S., Olafsdottir, G., Tryggvadottir, L., Neuhausen, S., Jonasson, J.G., Tavtigian, S.V., Tulinius, H., Ogmundsdottir, H.M. & Eyfjord, J.E. (1996). A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes [see comments]. *Nature Genetics.*, **13,** 117-119.

Trent, J.M., Stanbridge, E.J., McBride, H.L., Meese, E.U., Casey, G., Araujo, D.E.,

Witkowski, C.M. & Nagle, R.B. (1990). Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6. *Science.*,**47**, 568-571.

Ullu, E. & Tschudi, C. (1984). Alu sequences are processed 7SL pseudogenes. *Nature.*, **312,** 171.

van Gurp, R., Oosterhuis, L., Kalscheuer, V., Marimann, E.C. & Looijega, L.H. (1994). Biallelic expression of the H19 and IGF2 genes in human testicular germ cell tumours. *J. Natl. Cancer Inst.*, **86,** 1071-1075.

Van Zee, K., Calvano, J. & Bisogna, M. (1998). Hypomethylation and increased gene expression of p16/INK4a in primary and metastatic breast carcinoma as compared to normal breast tissue. *Oncogene.*, **16,** 2723-2727.

Varesco, L., Thomas, H.J., Cottrell, S., Murday, V., Fennell, S.J. Williams, S., Searle, S., Sheer, D., Bodmer, W.F., Frischauf, A.M. & Solomon, E. (1989). CpG island clones from a deletion encompassing the gene for adenomatous polyposis coli. *Proc. Natl. Acad Sci. USA.*, **86**, 10018-10022

Vidaud, D., Vidaud, M., Bahnak, B.R.,Siguret, V., Gispert, S.S., Laurian, Y., Meyer, D., Goosens, M. & Lavergene, J.M. (1993). Hemophilia B due to a *de novo* insertion of a human-specific Alu subfamily memeber within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* **1**, 30-36.

Vorce, R. & Goodman, J. (1989). Hypomethylation of ras oncogenes in chemically induced and spontaneous B6C3F1 mouse liver tumors. *Mol. Toxicol.*, **2,** 99-116.

Vorechovsky, I., Luo, L., Lindblom, A., Negrini, M., Webster, D.B., Croce, C.M. & Hammarstrom, L. (1996). ATM mutations in cancer families. *Cancer Res.*, **56**, 4130-4133.

Wallace, M.R., Andersen, L.B. Saulino, A.M.Gregory, P.E., Glover, T.W. & Collins, F.S. (1991). A *de novo* Alu insertion results in neurofibromatosis type 1. *Nature* **353**, 864-866.

Wales, M.M., Biel, M.A., El Deiry, W., Nelkin, B. D., Issa, J-P., Cavenee, W. K., Kuerbitz, S. J. & Baylin, S. B. (1995). p53 activates expression of HIC-1, a new candidate tumour suppressor on 17p13.3. *Nat. Med.*, **1**, 570-577.

Weiss, A., Keshet, I., Razin, A. & Cedar, H. (1996). DNA demethylation in vitro: involvement of RNA. *Cell.*, **85**, 709-718.

Weizhen, J., Zhang, X.Y., Warshamana, G., Qu, G-Z.A. & Ehrlich, M. (1994). Effect of internal direct and inverted Alu repeat sequences on PCR. *PCR methods and applications*,., **4**, 109-116.

Welsh, J. & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids. Res.*, **18**, 7213-7218.

Wiley, L., Adamson, E. & Tsark, E. (1995). Epidermal growth factor receptor function in early mammalian development. *Bioessays.*, **17**, 839-846.

Willard, C., Nguyen, H.T., Scmid, C.W. (1987). Existence of at least three distinct Alu subfamilies. *J. Mol. Evol.*, **26**, 180-186.

Wilgenbus, K.K., Coy, J.F., Mincheva, A., Nicolai, H., Solomon, E., Lichter, P. & Poustka, A. (1996). Ordering of 66 STSs along the entire short arm of human chromosome 17 and chromosome assignment of a transcribed sequence (FMR1L2) homologous to FMR1. *Cytogenet. Cell. Genet.*, **73**, 240-243.

Wilkinson, D., Mager, D. & Leong, J.A.C. (1994). Endogenous human retroviruses. In *The Retroviridae*, Levy, J. (ed), Vol. 3, 465-521. Plenum Press: New York.

Williams, J., Brown, K., Mott, M. & Maitland, N. (1989). Maternal allele loss in Wilm's tumour. *Lancet.*, **1**, 283-284.

Williamson, M.P., Elder, P.A. and Knowles, M.A. (1994). The spectrum of TP53 mutations in bladder carcinoma *Genes Chrom. Cancer*. **9**, 108-118.

Wutz, A., Smrzka, O., Scweifer, N., Schellander, K., Wagner, E.F. & Barlow, D.P. (1997). *Nature.*, **389**, 745-749.

Yagle, M.K., Parruti, G., Xu. W., Black, D.M. & Solomon, E. (1990). Genetic and physical map of the von Reclinghausen neurofibromatosis (NF1) regions on chromosome 17.*Proc. Natl. Acad. Sci. USA.*, **87**, 7255-7259.

Yamanaka, Y., Friess, H., Buchler, M., Beger, H.G., Do, D.A., Kobrin, M.S. & Korc, M. (1993). Overexpression of acidic and basic fibroblast growth factors in human pancreatic cancer. *Cancer Res.*, **53**, 5289-5296.

Yanagisawa, A., Ohtake, K., Ohashi, K., Hori, M., Kitagawa, T., Sugano, H. & Kato, Y. (1993). Frequent K-ras oncogene activation in mucous cell hyperplasia of the pancreas suffering from chronic inflammation. *Cancer Res.*, **53**, 953-956.

Yasushi, T., Pencil, S.D., & Nicolson, G.L. (1994). A novel candidate metastasis-associated gene, *mta-1* differentially expressed in highly metastatic mammary adenocarcinoma cell lines. *J. Biol. Chem.*, **269**, 22958-22963

Yoder, J., Walsh, C. & Bestor, T. (1997). Cytosine methylation and the ecology of intragenomic parasites. *TIG.*, **13**, 335-341.

Zawel, L., Le Dai, J., Buckhaults, P., Zhou, S., Kinzler, K.W., Vogelstein, B. & Kern, S.E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. *Molecular Cell.*, **1**, 611-617.

Zehetner, G. & Lehrach, H. (1994). The reference library system-sharing biological material and experimental data. *Nature.*, **367,** 489-491.

Zenklusen, J.C., Bieche, I., Lindereau, R. & Conti, C.J. (1994a). The microsatellite marker D7S522 is the most commonly deleted region in human primary breast cancer. *Proc. Natl. Acad. Sci. USA.*, **91,** 12155-12158.

Zenklusen, J.C., Oshimura, M., Barret, J.C. & Conti, C.J. (1994b). Inhibition of tumourigenicity of a murine squamous cell carcinoma (SCC) cell line by putative tumour suppressor gene on human chromosome 7. *Oncogene.*, **9,** 2817-2825.

Zenklusen, J.C., Thompson, J.C., Klein-Szanto, A.J. & Conti, C.J. (1995). frequent LOH in human primary squamous cell and colon carcinomas at 7q31.1: evidence for a broad range tumour suppressor gene. *Cancer Res.*, **55,** 1347-1350.

Zenklusen, J.C., Thompson, J.C., Troncoso, P., Kagan, J. & Conti, C.J. (1994c). Loss of heterozygosity in human primary prostate carcinomas: a possible tumour suppressor gene at 7q31.1. *Cancer Research.*, **54,** 6370-6373.

Zhang, Y., Feng, X.H., Wu, R.Y. & Derynck, R. (1996). The tumour suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are co-activators for Smad3 in TGF-β induced transcriptional activation.*Nature.*, **383,** 168-172.

Zhou, Y., Magill, J., Newton, D. & Magill, C. (1997). Use of a single sequencing termination reaction to distinguish between cytosine and 5 methyl-cytosine in bisulphite-modified DNA.*Biotechniques.*, **22,** 850-854.

Appendix 1: Case details

Case name /(Centre)	Sex	Age (yrs)	Resection date	Macroscopic description	Microscopic description
N. Tk. (MDX)	M	56	02.11.91	3.5cm White cystic tumour.	P. diff. adenocarcinoma main panc. duct
R. Tnr. (MDX)	M	60	18.09.92	3.0cm large mass - head of pancreas	W. diff. adenocarcinoma main duct
J. Gt. (MDX)	F	60	27.02.91	2.2cm Firm mass.	M. diff. adenocarcinoma.
R. Nr. (MDX)	M	67	29.11.90	1.5cm Grey lesion.	W. diff. adenocarcinoma main panc. duct.
S.Atk. (MDX)	F	70	04.03.92	2.0cm fibrous mass	M. diff. adenocarcinoma-ductal type
A. Bn. (HAMM)	M	54	21.11.90	2.5cm Yellow, granular tumour.	M. diff. adenocarcinoma ductal type,
A. Gn. (HAMM)	M	58	29.11.90	1.5cm Fibrous tissue, carcinoma foci.	W. diff. adenocarcinoma.
D. Bd. (HAMM)	F	79	24.10.90	3.0cm hard mass.	P. diff. adenocarcinoma main panc. duct.
E. Mto. (HAMM)	M	54	13.03.90	0.1cm papillary framents	no evidence of malignancy
P.S. (BRHM-DR)	M	62	6.10.90	1.3cm ill-defined grey tumour mass	P. diff. adenocarcinoma main panc. duct. and Ampulla
M. Ch. (BRHM-DR)	M	51	31.10.90	3.0cm solid tumour mass	M. diff. primary adenocarcinoma
DS. (BRHM-DR)				NB.-No records traced	
				Microscopically dissected cases used in multiplex	
W. Cp. (MDX)	M	67	05.06.95	0.9cm mucin secreting tumour	Intraductal mucin-secreting tumour
J. Chu. (MDX)	F	59	05.02.96	1.4cm diffuse mass from Ampulla, extending throughout main panc. duct.	P. diff. periampullary carcinoma
C. Childs. (MDX)	F	55	06.10.95	2.0cm hard mass	Insulinoma
P. Stve. (MDX)	M	60	21.02.95	2.8cm fungating tumour-Ampulla	p. diff. periampullary carcinoma
D. Cor. (MDX)	F	47	17.12.92	0.7cm papillary mass	P. diff. adenocarcinoma main panc. duct
F. LaF. (MDX)	F	55	05.01.96	2.5cm cystic mass	Early mucinous adenocarcinoma
B. Fr (HAMM)	M	67	02.04.94	2.2cm solid mass-main duct	W. diff. adenocarcinoma
2922 (MDX)	M	56	12.06.95	1.5cm fibrous mass	M. diff. adenocarcinoma main duct
1083 (MDX)	M	60	10.12.95	0.6cm mucinous tumour/pancreatitis?	Intraductal mucin secreting tumour
10409 (MDX)	F	58	18.12.95	4.5cm grey diffuse mass-main duct	P. diff. adenocarcinoma ductal
763 (MDX)	M	58	23.10.92	1.2cm diffuse mass-Ampulla	P. diff. periampullary carcinoma
7099 (MDX)	M	64	16.05.95	2.5cm fibrous mass	M. diff. adenocarcinoma ductal
6173 (MDX)	F	59	06.09.95	1.7cm yellow/grey mass	M. diff. adenocarcinoma main duct
J. Bggs (HAMM)	M	60's	21.08.95		Adenocarcinoma - head of pancreas.
R. Pge. (MDX)	F	62	09.11.93	1.5cm diffuse mass - Ampulla	P. diff. periampullary carcinoma main duct

Appendix 2a: **AB1 sequence homolgy with Sp.0.3-23**

HREF="http://srs.ebi.ac.uk:5000/srs5bin/cgi-bin/wgetz?[{embl_SP_emblnew}-AccNumber:X07493]+-e">X07493</A> **Human DNA insert showing sperm-specific** (323 nt)
initn: 389 init1: 175 opt: 414 Z-score: 400.2 expect() 6.3e-15
81.944% identity in 144 nt overlap

```
                30        40        50        60        70        80
AB1-SP          CGCGCCTATTTGCTGAGTACGTCGTCGCGCCAGTGTACAGTGCCATTTGGTTAGAAATCA
                                                  ********** ******  **********
HSSPM8          CTCGGGATGACAGACGTGAGCCACTGCGTTCAGTGTACAGTACCATTT-CTTAGAAATCA
                210       220       230       240       250       260

                90        100       110       120       130       140
AB1-SP          CTCCTCACCGGGAACACACACTTACTGGTGAACCGTGTAGAGATTTTAGTTAGTTAGTTA
                ******* ** ***************    **** *   ****************************
HSSPM8          CTCCTCA-CGGGAACACACACTTATGGGTGA--CGTGTAGAGATTTTAGTTAGTTAGTTA
                270       280       290       300       310       320

                150       160       170       180       190
AB1-SP          GTTAGTT------ATTATCGATTGCCGCCCGGGGGGGGAGGTGAGGAGAGA
                ********      *****                      *******
HSSPM8          GTTAGTTAGTTATTATGTGC----------GCGGGGAGGTGGGGGGA--
                330       340       350       360       370
```

Appendix 2b: **17p13.3 cDNA clone7 homology with human PITP mRNA**

>>HSM73704 M73704 Human **phosphatidylinositol transfer protein** (2255 nt)
initn: 1266 init1: 1106 opt: 1269 Z-score: 1043.1 expect() 0
97.464% identity in 276 nt overlap

```
                                                  10        20        30
CHR17-                                            CCCGCCTGGCTCCTTGAGGGGAGGACCCCA
                                                  ***************************  ****
HSM73           TGTGGTGGGGATAAGAAGACCCTTCCAGATCCCGCCTGGCTCCTTGA-GGGAGGA--CCCA
                910       920       930       940       950       960

                40        50        60        70        80        90
CHR17-          CTGGGTGAAGCACCACCCACAAGATCTGAGCCAGGATGGGAGTGAATCCCAGGCGGAGGC
                ************************************************************
HSM73           CTGGGTGAAGCA-CAGCCAC-AGATCTGAGCCAGGATGGGAGTGAATCCCAGGCGGAGGC
                970       980       990       1000      1010      1020

                100       110       120       130       140       150
CHR17-          CAGAGCTGGCAGCCAGCTCTGCCTTTCCACTGCCGGGAGTGCTGGGGGCCCAGCCTGGCC
                ************************************************************
HSM73           CAGAGCTGGCAGCCAGCTCTGCCTTTCCACTGCCGGGAGTGCTGGGGGCCCAGCCTGGCC
                1030      1040      1030      1060      1070      1080

                160       170       180       190       200       210
CHR17-          CCCTGAAGAGATAGCCAAGTGTCGTCCACATACTCCTCCCAGAGTGAGCTCTAACCAGGC
                ************************************************************
HSM73           CCCTGAAGAGACAGCCAAGTGTCGTCCACATACTCCTCCCAGAGTGAGCTCTAACCAGGC
                1090      1100      1110      1120      1130      1140

                220       230       240       250       260       270
CHR17-          TCATTTGCTCTCTCCACTACTCATCTCTGGAATTAGCCGCTTAAATACAGGTTTTTGTTG
                ************************************************************
HSM73           TCATTTGCTCTCTCCACTACTCATCTCTGGAATTAGCCGCTTAAATACAGGTTTTTGTTG
                1150      1160      1170      1180      1190      1200      240
```

```
                280             290
CHR17-          CTGAAAAAAAAAAAAAAAAAAAAAAAAA
                * * * *
HSM73           CTGAGATGTGAGTGAAACCAGCTAGTGTGTCAACAGTGAAGACCTGGGGACAGTTCTGCG
                1210            1220            1230            1240            1250            1260
```

Appendix 2c: 17p13.3 COSMID IA (Eco RI fragment) homology with human HER4 mRNA

>>HSRETYKIN L07868 Homo sapiens receptor tyrosine kinase (5484 nt) **HER4**
initn: 1085 init1: 1085 opt: 1319 Z-score: 1097.5 expect() 0
96.271% identity in 295 nt overlap

```
                230         240         250         260         270         280
COSIA-          TTTTTTTTNGGGGGGGGNTNTTCCTTTNAACCCCCAAGGCCNAGTTAGTTTNGAAAAACTT
                                        * * * * * *   * * * * *  * * * * * *  * *  * * * * * *  *       * * * *
HSRETY          CCCTCTCTTTCTCTGGTGGTCTTCCTTCT-ACCCC-AAGGCC-AG-TAGTTTT-GCA---CTT
                4020        4030        4040        4050        4060        4070

                290         300         310         320         330         340
COSIA-          CCCAGGGGGGAAGGATACAGAGATGCAATGATAGTTATGTGCTTACCTAAACTTGAACATT
                * * * * *  * *   * * *   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
HSRETY          CCCAGTGG-AAGA--TACAGAGATGCAATGATAGTTATGTGCTTACCT-AACTTGAACATT
                4080        4090        4100        4110        4120        4130

                350         360         370         380         390         400
COSIA-          AGAGGGAAAGACTGAAAGAGAAAGATAGGAGGAACCACAATGTTTCTTCATTTCTCTGCA
                * * * * *  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
HSRETY          AGAGGGAAAGACTGAAAGAGAAAGATAGGAGGAACCACAATGTTTCTTCATTTCTCTGCA
                4130        4140        4150        4160        4170        4180

                410         420         430         440         450         460
COSIA-          TGGGTTGGTCAGGAGAATGAAACAGCTAGAGAAGGACCAGAAAATGTAAGGCAATGCTGC
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
HSRETY          TGGGTTGGTCAGGAGAATGAAACAGCTAGAGAAGGACCAGAAAATGTAAGGCAATGCTGC
                4190        4200        4210        4220        4230        4240

                470         480         490         500         510         520
COSIA-          CTACTATCAAACTAGCTGTCACTTTTTTTCTTTTTCTTTTTCTTTCTTTGTTTCTTTCTT
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
HSRETY          CTACTATCAAACTAGCTGTCACTTTTTTTCTTTTTCTTTTTCTTTCTTTGTTTCTTTCTT
                4250        4260        4270        4280        4290        4300

                530         540         550         560         570         580
COSIA-          CCTCTTCTTTTTTTTTTTTTTTTTATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCT
                * * * * * * * * * * * * * * * * * * * * * * * *
HSRETY          CCTCTTCTTTTTTTTTTTTTTTTTAAAGCAGATGGTTGAAACACCCATGCTATCTGTTCC
                4310        4320        4330        4340        4350        4360
```

Appendix 2d: AB4 Alumorph fragment homology with human homologue of Bzip transcription factor BACH1

mb:Af026200]+-e">Af026200</A> **Homo sapiens Bach1 protein homolog** m (207 nt)
initn: 420 init1: 235 opt: 820 Z-score: 786.1 expect() 0
91.163% identity in 215 nt overlap

```
                40          50          60          70          80          90
ART4M1          TCGGNTTGTCAGCGAAACAAGAAATGTGAGGTAAAACTGCCATTCAGTGCACAACGGGAT
                                                    ************************ ****
AF0262                                              GTAAAACTGCCATTCAATGCACAAC -GGAT
                                                    10          20          30

                100         110         120         130         140         150
ART4M1          AATTTCACTGTCTCGGAAATGATTTCAAGTCCCTGGTTGGAAATTGCAACAGGCTTACTC
                ************** ****************** ********* ** *** ************
AF0262          AATTTCACTGTCTC-GAAATGATTTTCAGTCC-TTGTTG-AAAAT-GC-ACAAGCTTACTC
                40          50          60      70      80          90          100

                160         170         180         190         200         210
ART4M1          CAGAACAAGCTGGATGGTTTCCATGATATTCGAAGGAAGAAGTAAAAACAGAATTGCTGC
                ******* ********  ** ***************** ****************************
AF0262          CAGAAC--AGCTGGATTTGTATCCATGATATTCGAA-GAAGAAGTAAAAACAGAATTGCTGC
                      110         120         130         140         150         160

                220         230         240         250         260         270
ART4M1A         CAGCGCTGTCGCAAGAGAAAAACTGGACTGTATACAGAATCTTGGATCAGAAATTGAGA
                ******************* ******** ******************* ****************** * *
AF0262          ACAGCGCTGTCGCAAGAG-AAAACTTG-ACTGTATACAGAATCTTGAATCAGAAATTGAGA
                          170         180         190         200         210

                280         290         300         310         320         330
ART4M1          AAGCTGGGTTGATGATAATTAGAAGCATGCTTTCCACTGAAACTTCCCGACAACATTTGTTA
                  *****
AF0262          -AGCTG
                220
```

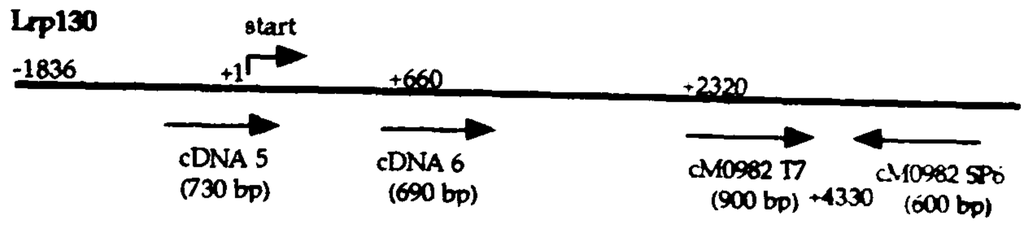
Appendix 2e: **2p22 shotgun clone P1sh19 (bases 303-433) homology with human cytochrome P450 subfamily IIIA.**
L35912 (1610 nt) 41% identity - 98 base match, 19 base mismatch.

```
                303         313         323         333         343
P1SH19          CTGTCCTTTATAATAGCGTGGCTGG - ACTGCAGGTCATTATG -
                ****** *** ***** * *** *** **** ************
P450-IIIA       CTGTCGTTTGTAATAACATGGATGGAACTGGAGGTCATTATGT
                142         152         162         172         182

                353         363         373         383         393
P1SH19          TTAGGTAAAATAAGCCAGGCACACAAAGACAGACATTG- CA
                **     ******   ** ************** ****** **  ***  **
P450-IIIA       TTT TTTAAAAT TAGCCAGGCACACANAGACAAACNTTGTCA
                  192         202         212         222         232

                            403         413         423         433
P1SH19          TGTTCTCACTTAT TTGTGGGATCTACAAATCAAAACAATT
                ************  * ********** *** * **** ****
P450-IIIA       TGTTCTCACTNACTTGTGGGACCTANACATCACAACATCACA
                            242         252         262         272
```

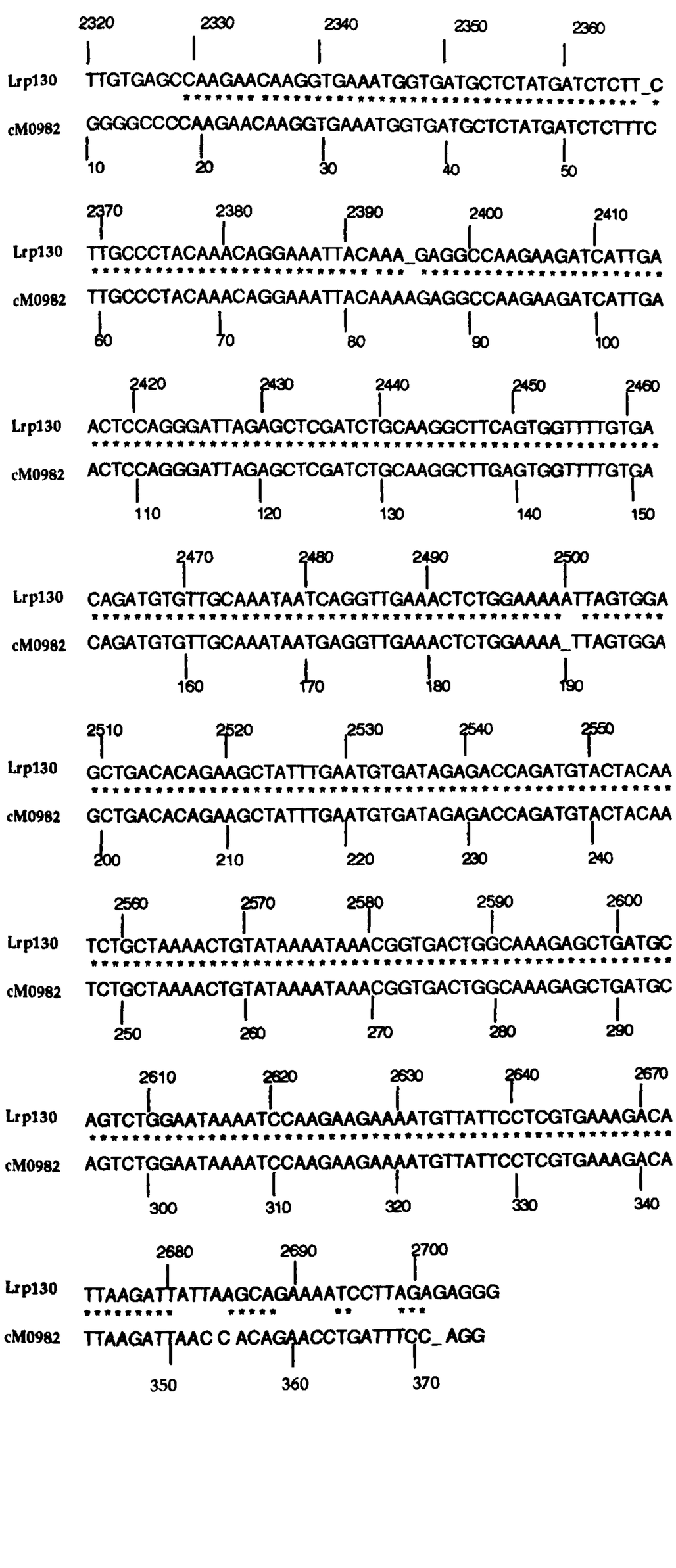
Appendix 2f: **2p22 shotgun clone P1sh19 (bases 177-292) homology with human cytochrome P450 subfamily IIIA - 5′ region.**
S74699 (1526 nt) 39% identity - 83 base match, 13 base mismatch.

```
                177         187         197         207         217
P1SH19          GGCTGGAACT - CAGGTCATTATG - TTAGGTAAAATAAGCCAGG
                * * ** * **** ************ ** ** ** ********
P450-IIIA5      GGATGGAACTGGAGGTCATTATGTTTTTTTTTAAATTAGCCAGG
                161         171         181         191         201

                227         237         247         257         267
P1SH19          GCACACAAAGACAGACATTG - CATGTTCTCACTTATTTGTGGG
                **   **** ****** ** *** ************** * ********
P450-IIIA5      GCNCACANAGACAAACNTTGTCATGTTCTCACTNACTTGTGGA
                  211         221         231         241         251

                        277         287         297
P1SH19          ATCTACAAATCAAATCAAAACAATT
                *  *** *  **** **** ** *****
P450-IIIA5      ACCTANACATCACATCACAACAATT
                        261         271         281
```

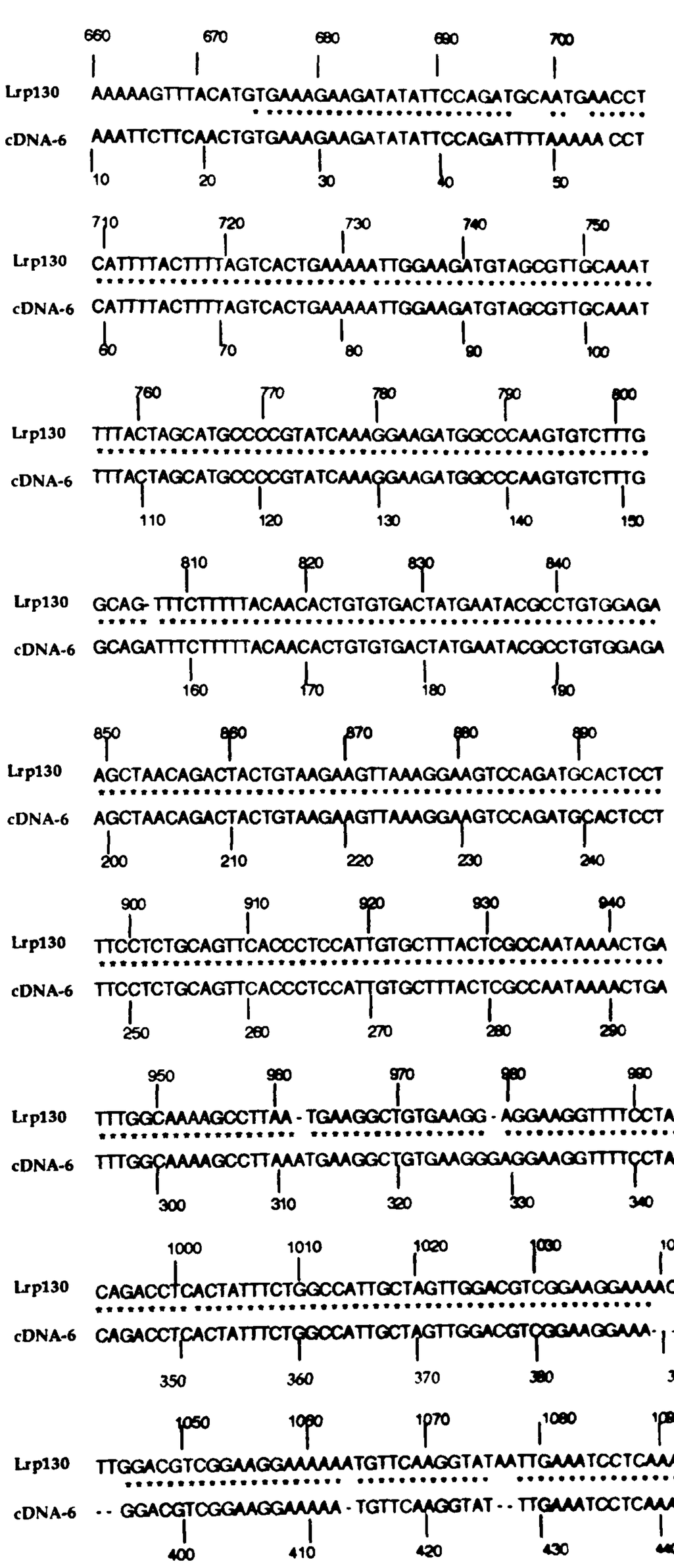
Appendix 2g: Overall alignment of cDNA fragments with the Lrp130 gene sequence



Appendix 2g-A: Alignment of cM0982-T7 cDNA with Lrp130 coding sequence



Appendix 2g-B: Alignment of cDNA clone6 (T7-end) with Lrp130 gene sequence



Appendix 3 : Lrp130 sequence including 5′ upstream sequences and homologies

(Linear) MAP of: LRP130.REP.UPSTR.SEQ check: 3403 from: 1 to: 6197

With 7 enzymes: ALUI HPHI AFLIII MSEI HPAII MSPI SACII July 7, 1998 13:40

..

```
                                                                TCTCTCTCTCTCTCTCACTCACTGTCTCTCCGTCTCTCTCTCTTTCTGTCTGTTTCTCTC
                                                                1
---------+---------+---------+---------+---------+---------+---------+
(-1836)
                                                                AGAGAGAGAGAGAGAGTGAGTGACAGAGAGGCAGAGAGAGAGAAAGACAGACAAAGAGAG

                                                                TGTCTCTGTCTTTCTGTGTGTCTGTCTGTCTCTCTCTCTATTTGTCTTTCTCCCTCCCTG
                                                                61
---------+---------+---------+---------+---------+---------+---------+
(-1776)
                                                                ACAGAGACAGAAAGACACACAGACAGACAGAGAGAGAGATAAACAGAAAGAGGGAGGGAC

                                                                TCTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTGTCTGTCTGTTTCTCTCTATCTCTCGC
                                                                121
---------+---------+---------+---------+---------+---------+---------+
(-1716)
                                                                AGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGACAGACAGACAAAGAGAGATAGAGAGCG

                                                                TGTCCATCTCTGTCTTTCTATGTCTGTCTCTTTCTCTGTCAGTCTGTCAGACACCCCGTG
                                                                181
---------+---------+---------+---------+---------+---------+---------+
(-1656)
                                                                ACAGGTAGAGACAGAAAGATACAGACAGAGAAAGAGACAGTCAGACAGTCTGTGGGGCAC

                                                                Hpall
                                                                 MspI
                                                                  I
                                                                 CCGGGTAGGCCCTGCCCCTTCCACGAAAGTGAGAAGCGCGTGCTTCGGTGCTTAGAGAGG
                                                                241
---------+---------+---------+---------+---------+---------+---------+
(-1596)
                                                                GGCCCATCCGGGACGGGGAAGGTGCTTTCACTCTTCGCGCACGAAGCCACGAATCTCTCC

                                                                CCGAGAGGAATCTAGACAGGCGGGCCTTGCTGGGCTTCCCCACTCGGTGTATGATTTCGG
                                                                301
---------+---------+---------+---------+---------+---------+---------+
(-1536)
                                                                GGCTCTCCTTAGATCTGTCCGCCCGGAACGACCCGAAGGGGTGAGCCACATACTAAAGCC
```

HpaII
MspI
I
GAGGTCGAGGCCGGGTCCCCGCTTGGATGCGAGGGGCATTTTCAGACTTTTCTCTCGGTC
361
---------+---------+---------+---------+---------+---------+
(-1476)
CTCCAGCTCCGGCCCAGGGGCGAACCTACGCTCCCCGTAAAAGTCTGAAAAGAGAGCCAG

AflIII
I
AluI
I
ACGTGTGGCGTCCGTACTTCTCCTATTTCCCCGATAAGCTCCTCGACTTCAACATAAACG
421
---------+---------+---------+---------+---------+---------+
(-1416)
TGCACACCGCAGGCATGAAGAGGATAAAGGGGCTATTCGAGGAGCTGAAGTTGTATTTGC

HphI
I
GCGTCCTAAGGGTCGATTTAGTGTCATGCCTCTTTCACCGCCACACCGAAGATGAAAGCA
481
---------+---------+---------+---------+---------+---------+
(-1376)
CGCAGGATTCCCAGCTAAATCACAGTACGGAGAAAGTGGCGGTGTGGCTTCTACTTTCGT

AAGATCGGCTAAATACCGCGTGTTCTCAATCTAGAAGTGGGAACTTACAGATGACAGTTC
541 LINE-1 HOMOLOGY
---------+---------+---------+---------+---------+---------+
(-1316)
TTCTAGCCGATTTATGGCGCACAAGAGTTAGATCTTCACCCTTGAATGTCTACTGTCAAG

HpaII
MspI
I
TTGCATGGGCAGAACGAGGGGGACCGGGGACGCGGAACGCTGCTTGAGGGAGGAGGGGTG
601 LINE-1 HOMOLOGY
---------+---------+---------+---------+---------+---------+
(-1256)
AACGTACCCGTCTTGCTCCCCCTGGCCCCTGCGCCTTGCGACGAACTCCCTCCTCCCCAC

AluI
I
GAAGGAGAGACAGCTTAGGAAGAAAACAAAACACGAATACTGTCGGACACAGCACTGACT
661 LINE-1 HOMOLOGY
---------+---------+---------+---------+---------+---------+
(-1196)
CTTCCTCTCTGTCGAATCCTTCTTTTGTTTTGTGCTTATGACAGCCTGTGTCGTGACTGA

HpaII
MspI
I
ACCCGGGTCATGAAATCATCTGCACACTGAACACCCCGTCACAAGATTTCCTATGTCACA
721 LINE-1 HOMOLOGY
---------+---------+---------+---------+---------+---------+---------+
(-1136)
TGGGCCCAGTACTTTAGTAGACGTGTGACTTGTGGGGCAGTGTTCTAAAGGATACAGTGT

GTCTTGCTCATGTATGCTTGAACGACAAATAAAAGTTCGGGGGGAGAAGAGAGGAGAGAG
781
---------+---------+---------+---------+---------+---------+---------+
(-1076)
CAGAACGAGTACATACGAACTTGCTGTTTATTTTCAAGCCCCCCTCTTCTCTCCTCTCTC

AGAGAGAGACGGGGAGAGAGGGGGGAGAGGGGGGGGAGAGAGAGAGAGAGAGAGAGAG
841
---------+---------+---------+---------+---------+---------+---------+
(-1016)
TCTCTCTCTGCCCCTCTCTCCCCCCTCTCCCCCCCCCTCTCTCTCTCTCTCTCTCTCTC

AGAGAGAGAGAGAGAGAGAGAAAGAGAAGTAAAACCAACACCACCTCCTTGACCTGAGTC
901
---------+---------+---------+---------+---------+---------+---------+
(-956)
TCTCTCTCTCTCTCTCTCTCTTTCTCTTCATTTTGGTTGTGGTGGAGGAACTGGACTCAG

AGGGGGTTTCTGGCCTTTTGGGAGAACGTTCAGCGACAATGCAGTATTTGGGCCCGTTCT
961
---------+---------+---------+---------+---------+---------+---------+
(-896)
TCCCCCAAAGACCGGAAAACCCTCTTGCAAGTCGCTGTTACGTCATAAACCCGGGCAAGA

HphI
I
TTTTTTCTTCTTCTTCTTTTCTTTCTTTTTTTTTGGACTGAGTCTCTCTCGCTCTGTCAC
1021 ALU Sg/Sx HOMOLOGY
---------+---------+---------+---------+---------+---------+---------+
(-836)
AAAAAAGAAGAAGAAGAAAAGAAAGAAAAAAAAACCTGACTCAGAGAGAGCGAGACAGTG

HpaII
MspI
I
CCAGGCTGCGGTGCGGTGGCGCTCTCTCGGCTCACTGAAATCTCTGTTTCCCGGGTTCCA
1081 ALU Sg/Sx HOMOLOGY
---------+---------+---------+---------+---------+---------+---------+
(-676)
GGTCCGACGCCACGCCACCGCGAGAGAGCCGAGTGACTTTAGAGACAAAGGGCCCAAGGT
HpaII

AluI MspI

GTGAGTTCTTCCTTCGGTAGCTGGGATTACAGGTGCGGACCATACGGCCGGCTCATGTTC

1141 ALU Sg/Sx HOMOLOGY

---------+---------+---------+---------+---------+---------+

(-616)

CACTCAAGAAGGAAGCCATCGACCCTAATGTCCACGCCTGGTATGCCGGCCGAGTACAAG

HpaII MspI

TATTTTAGTAGAGACGGGGTTTCTCCACGTTGGCCACTAATGAGACTGGATAATTCTGTC

1201

---------+---------+---------+---------+---------+---------+

(-556)

ATAAAATCATCTCTGCCCCAAAGAGGTGCAACCGGTGATTACTCTGACCTATTAAGACAG

CGGAGAACGGGCCGC-*ATTACCAAGGGGCTTCTGCAGAAAAGTCTTTGAGAGCACGTGCCG*

1261 ---------+---------+-----CpG island-+---------+---------+

(-496)

GCCTCTTGCCCGGCG-*TAATGGTTCCCCGAAGACGTCTTTTCAGAAACTCTCGTGCACGGC*

HphI AluI

CTCAGGTGAGCCGCAGGGAGTAACCAGGCCTTGCTTCTGCTTGCGGCAGCTGCGGCTCGC

1321 ---------+---------+-----CpG island-+---------+---------+

(-436)

GAGTCCACTCGGCGTCCCTCATTGGTCCGGAACGAAGACGAACGCCGTCGACGCCGAGCG

TTTCGCTGCGGCCGAACTAGAAGTCTTGCAGAGAGGACAGAATGTCCTGCTCATAAGATA

1381 ---------+---------+-----CpG island-+---------+---------+

(-376)

AAAGCGACGCCGGCTTGATCTTCAGAACGTCTCTCCTGTCTTACAGGACGAGTATTCTAT

TGGTTGACAAACCCTTCAGAAATGTCCCCGAGGCCGCTGACTTTCAATTAGCATCGCAGC

1441 ---------+---------+-----CpG island-+---------+---------+

(-316)

ACCAACTGTTTGGGAAGTCTTTACAGGGGCTCCGGCGACTGAAAGTTAATCGTAGCGTCG

AluI AluI

CCGAGGGAGCTGCTAGCTGCTACGTACCTCGTGCCAAACTCCGTCCCCC-TACCTCCCCTT

1501 ---------+------CpG island-+---------+---------+---------+

(-256)

GGCTCCCTCGACGATCGACGATGCATGGAGCACGGTTTGAGGCAGGGGG-ATGGAGGGGAA

HphI

CACATAATAACTAACTAACTAACTAACTAACTAAAATCTCTACACTGTCACCCATAAGTG

1561

(-196)

GTGTATTATTGATTGATTGATTGATTGATTGATTTTAGAGATGTGACAGTGGGTATTCAC

AB1 HOMOLOGY

TGTGTTCCCGTGAAGCAAACATTTCTAACCAAATCGAGTGACTGTACACAGAGATTGCAG

1621

(-136)

ACACAAGGGCACTTCGTTTGTAAAGATTGGTTTAGCTCACTGACATGTGTCTCTAACGTC

TTGCTCACGTCTGTCATCATCCTGTAATGTAGGAGATATTGAAGGTGCCCAGCAAGATGG

1681 ALU Sg/Sx HOMOLOGY

(-76)

AACGAGTGCAGACAGTAGTAGGACATTACATCCTCTATAACTTCCACGGGTCGTTCTACC

HphI
HpaII
MspI

TGAAACCCCGGATCTCTACTGAAAACTAAGGATCTCCCAGTTAGAGAGCGCAGTATTCAG

1741 ALU Sg/Sx HOMOLOGY

(-16) (+1)

ACTTTGGGGCCTAGAGATGACTTTTGATTCCTAGAGGGTCAATCTCTCGCGTCATAAGTC

AluI HphI

TGCACCTTGTAACCCCAGGGCATGCCAGAGCTGGTGATATGGAGAATGCAGAAAACATTC

1801

(+41)

ACGTGGAACATTGGGGTCCCGTACGGTCTCGACCACTATACCTCTTACGTCTTTTGTAAG

HpaII
MspI

TCACAGTGATGAGAGATGCCGGAATTGAGCCTGGTCCAGACACATACCTCGCATTATTGA

1861

(+101)

AGTGTCACTACTCTCTACGGCCTTAACTCGGACCAGGTCTGTGTATGGAGCGTAATAACT

MseI

ATGCATATGCTGAGAAGGGCGACATTGACCATGTTAAGCAGACTCTGGAGAAGGTGGAGA

1921

250

TACGTATACGACTCTTCCCGCTGTAACTGGTACAATTCGTCTGAGACCTCTTCCACCTCT

HphI AluI AluI AluI

AGTTCGAGCTTCACCTTATGGACCGTGATTTACTGCAAATTATTTTTAGCTTCAGTAAAG

1981

TCAAGCTCGAAGTGGAATACCTGGCACTAAATGACGTTTAATAAAAATCGAAGTCATTTC

AflIII

CTGGGTATCTCAGTATGTCTCAGAAATTTTGGAAAAAGTTTACATGTGAAAGAAGATATA

2041

GACCCATAGAGTCATACAGAGTCTTTAAAACCTTTTTCAAATGTACACTTTCTTCTATAT

TTCCAGATGCAATGAACCTCATTTTACTTTTAGTCACTGAAAAATTGGAAGATGTAGCGT

2101

AAGGTCTACGTTACTTGGAGTAAAATGAAAATCAGTGACTTTTTAACCTTCTACATCGCA

TGCAAATTTTACTAGCATGCCCCGTATCAAAGGAAGATGGCCCAAGTGTCTTTGGCAGTT

2161

ACGTTTAAAATGATCGTACGGGGCATAGTTTCCTTCTACCGGGTTCACAGAAACCGTCAA

AluI

TCTTTTTACAACACTGTGTGACTATGAATACGCCTGTGGAGAAGCTAACAGACTACTGTA

2221

AGAAAAAATGTTGTGACACACTGATACTTATGCGGACACCTCTTCGATTGTCTGATGACAT

MseI HphI

AGAAGTTAAAGGAAGTCCAGATGCACTCCTTTCCTCTGCAGTTCACCCTCCATTGTGCTT

2281

TCTTCAATTTCCTTCAGGTCTACGTGAGGAAAGGAGACGTCAAGTGGGAGGTAACACGAA

MseI
|
TACTCGCCAATAAAACTGATTTGGCAAAAGCCTTAATGAAGGCTGTGAAGGAGGAAGGTT
2341
---------+---------+---------+---------+---------+---------+

ATGAGCGGTTATTTTGACTAAACCGTTTTCGGAATTACTTCCGACACTTCCTCCTTCCAA

TTCCTATCAGACCTCACTATTTCTGGCCATTGCTAGTTGGACGTCGGAAGGAAAAAAATG
2401
---------+---------+---------+---------+---------+---------+

AAGGATAGTCTGGAGTGATAAAGACCGGTAACGATCAACCTGCAGCCTTCCTTTTTTTAC

TTCAAGGTATAATTGAAATCCTCAAAGGAATGCAAGAATTGGGAGTACATCCTGATCAGG
2461
---------+---------+---------+---------+---------+---------+

AAGTTCCATATTAACTTTAGGAGTTTCCTTACGTTCTTAACCCTCATGTAGGACTAGTCC

AAACATATACAGATTATGTGATTCCATGCTTTGATAGTGTAAACTCAGCACGAGCCATTT
2521
---------+---------+---------+---------+---------+---------+

TTTGTATATGTCTAATACACTAAGGTACGAAACTATCACATTTGAGTCGTGCTCGGTAAA

AluI
|
TGCAGGAAAATGGATGTCTGTCTGATAGTGATATGTTTTCTCAAGCTGGATTGAGAAGTG
2581
---------+---------+---------+---------+---------+---------+

ACGTCCTTTTACCTACAGACAGACTATCACTATACAAAAGAGTTCGACCTAACTCTTCAC

AAGCAGCAAATGGGAACTTAGACTTTGTATTATCATTTTTGAAATCAAATACATTGCCCA
2641
---------+---------+---------+---------+---------+---------+

TTCGTCGTTTACCCTTGAATCTGAAACATAATAGTAAAAACTTTAGTTTATGTAACGGGT

TCTCGCTGCAGTCTATAAGAAGTAGCCTACTGCTAGGCTTCAGGAGGTCTATGAATATAA
2701
---------+---------+---------+---------+---------+---------+

AGAGCGACGTCAGATATTCTTCATCGGATGACGATCCGAAGTCCTCCAGATACTTATATT

ATGTTTGGAGCGAGATAACAGAATTATTGTACAAGGATGGACGTTATTGCCAGGAGCCTC
2761
---------+---------+---------+---------+---------+---------+

TACAAACCTCGCTCTATTGTCTTAATAACATGTTCCTACCTGCAATAACGGTCCTCGGAG

AluI

GAGGACCGACGGAAGCTGTTGGCAATTTTCTTTATAACTTGATTGACAGCATGAGTGACT

2821

CTCCTGGCTGCCTTCGACAACCGTTAAAAGAAATATTGAACTAACTGTCGTACTCACTGA

AluI

CAGAGGTACAGGCCAAGGAGGAGCATTTGAGACAATACTTCCATCAGCTGGAGAAGATGA

2881

GTCTCCATGTCCGGTTCCTCCTCGTAAACTCTGTTATGAAGGTAGTCGACCTCTTCTACT

AluI

ATGTAAAAATTCCTGAAAATATCTACAGAGGCATTCGTAATCTCCTGGAAAGCTACCATG

2941

TACATTTTTAAGGACTTTTATAGATGTCTCCGTAAGCATTAGAGGACCTTTCGATGGTAC

MseI

TTCCTGAATTGATTAAGGATGCTCACTTGTTGGTTGAGCGTAAGAATTTAGACTTTCAAA

3001

AAGGACTTAACTAATTCCTACGAGTGAACAACCAACTCGCATTCTTAAATCTGAAAGTTT

AluI

AAACTGTGCAACTTACATCATCTGAATTGGAGTCAACACTTGAAACACTAAAAGCTGAAA

3061

TTTGACACGTTGAATGTAGTAGACTTAACCTCAGTTGTGAACTTTGTGATTTTCGACTTT

ATCAACCTATAAGAGATGTCCTAAAGCAACTCATATTAGTGCTTTGTTCAGAAGAGAATA

3121

TAGTTGGATATTCTCTACAGGATTTCGTTGAGTATAATCACGAAACAAGTCTTCTCTTAT

TGCAAAAAGCCCTTGAATTGAAAGCAAAATATGAATCCGACATGGTTACTGGTGGCTATG

3181

ACGTTTTTCGGGAACTTAACTTTCGTTTTATACTTAGGCTGTACCAATGACCACCGATAC

MseI
AluI I
I I
CAGCTTTAATAAATTTATGCTGTCGACATGATAAAGTAGAAGATGCCTTGAACTTGAAAG
3241

GTCGAAATTATTTAAATACGACAGCTGTACTATTTCATCTTCTACGGAACTTGAACTTTC

HpaII
MspI
I
AAGAATTTGACCGCTTAGATTCATCTGCTGTCCTTGACACCGGCAACTATCTAGGCCTTG
3301

TTCTTAAACTGGCGAATCTAAGTAGACGACAGGAACTGTGGCCGTTGATAGATCCGGAAC

AluI MseI
I I
TAAGAGTATTGGCAAAGCATGGCAAGCTCCAAGATGCTATTAAGATTCTGAAGGAGATGA
3361

ATTCTCATAACCGTTTCGTACCGTTCGAGGTTCTACGATAATTCTAAGACTTCCTCTACT

AAGAGAAGGATGTTCTTATCAAAGATACAACAGCCTTGTCCTTTTTCCACATGCTAAATG
3421

TTCTCTTCCTACAAGAATAGTTTCTATGTTGTCGGAACAGGAAAAAGGTGTACGATTTAC

MseI
AluI I HphI
I I I
GCGCAGCTTTAAGAGGTGAAATTGAAACAGTAAAACAGTTGCATGAAGCCATCGTGACTC
3481

CGCGTCGAAATTCTCCACTTTAACTTTGTCATTTTGTCAACGTACTTCGGTAGCACTGAG

TAGGGTTAGCAGAACCATCCACCAACATAAGTTTCCCATTGGTCACTGTACACTTGGAAA
3541

ATCCCAATCGTCTTGGTAGGTGGTTGTATTCAAAGGGTAACCAGTGACATGTGAACCTTT

AGGGCGACCTATCTACTGCTCTTGAGGTCGCCATTGACTGCTATGAAAAGTATAAAGTAT
3601
---------+---------+---------+---------+---------+---------+---------+

TCCCGCTGGATAGATGACGAGAACTCCAGCGGTAACTGACGATACTTTTCATATTTCATA

TACCAAGGATTCATGATGTCTTGTGTAAACTGGTAGAGAAAGGCGAGACTGATCTAATTC
3661
---------+---------+---------+---------+---------+---------+---------+

ATGGTTCCTAAGTACTACAGAACACATTTGACCATCTCTTTCCGCTCTGACTAGATTAAG

HphI HphI
| |
AGAAAGCAATGGACTTTGTGAGCCAAGAACAAGGTGAAATGGTGATGCTCTATGATCTCT
3721
---------+---------+---------+---------+---------+---------+---------+

TCTTTCGTTACCTGAAACACTCGGTTCTTGTTCCACTTTACCACTACGAGATACTAGAGA

TCTTTGCCTTCCTACAAACAGGAAATTACAAAGAGGCCAAGAAGATCATTGAGACTCCAG
3781
---------+---------+---------+---------+---------+---------+---------+

AGAAACGGAAGGATGTTTGTCCTTTAATGTTTCTCCGGTTCTTCTAGTAACTCTGAGGTC

AluI
|
GGATTAGAGCTCGATCTGCAAGGCTTCAGTGGTTTTGTGACAGATGTGTTGCAAATAATC
3841
---------+---------+---------+---------+---------+---------+---------+

CCTAATCTCGAGCTAGACGTTCCGAAGTCACCAAAACACTGTCTACACAACGTTTATTAG

AluI AluI
| |
AGGTTGAAACTCTGGAAAAATTAGTGGAGCTGACACAGAAGCTATTTGAATGTGATAGAG
3901
---------+---------+---------+---------+---------+---------+---------+

TCCAACTTTGAGACCTTTTTAATCACCTCGACTGTGTCTTCGATAAACTTACACTATCTC

AluI
HphI |
| |
ACCAGATGTACTACAATCTGCTAAAACTGTATAAAATAAACGGTGACTGGCAAAGAGCTG
3961
---------+---------+---------+---------+---------+---------+---------+

TGGTCTACATGATGTTAGACGATTTTGACATATTTTATTTGCCACTGACCGTTTCTCGAC

MseI

ATGCAGTCTGGAATAAAATCCAAGAAGAAAATGTTATTCCTCGTGAAAAGACATTAAGAT

4021

TACGTCAGACCTTATTTTAGGTTCTTCTTTTACAATAAGGAGCACTTTTCTGTAATTCTA

TATTAGCAGAAATCCTTAGAGAGGGTAACCAGGAAGTTCCGTTTGACGTACCTGAGTTGT

4081

ATAATCGTCTTTAGGAATCTCTCCCATTGGTCCTTCAAGGCAAACTGCATGGACTCAACA

GGTATGAAGATGAAAAACATTCCCTGAATTCTTCGTCAGCCTCAACCACAGAACCTGATT

4141

CCATACTTCTACTTTTTGTAAGGGACTTAAGAAGCAGTCGGAGTTGGTGTCTTGGACTAA

TCCAGAAAGATATATTGATTGCCTGCCGATTGAACCAAAAAAAAAGGGGCATATGATATTT

4201

AGGTCTTTCTATATAACTAACGGACGGCTAACTTGGTTTTTTTTCCCCGTATACTATAAA

MseI MseI

TCCTGAATGCAAAAGAGCAAAACATTGTGTTTAATGCTGAAACCTACAGCAATCTCATTA

4261

AGGACTTACGTTTTCTCGTTTTGTAACACAAATTACGACTTTGGATGTCGTTAGAGTAAT

AATTACTGATGTCAGAAGATTATTTTACACAAGCAATGGAAGTGAAAGCATTCGCGGAGA

4321

TTAATGACTACAGTCTTCTAATAAAATGTGTTCGTTACCTTCACTTTCGTAAGCGCCTCT

CCCACATCAAGGGCTTCACACTGAACGATGCTGCCAACAGCCGCCTCATCATAACGCAAG

4381

GGGTGTAGTTCCCGAAGTGTGACTTGCTACGACGGTTGTCGGCGGAGTAGTATTGCGTTC

TTAGGCGGGATTATTTGAAAGAGGCTGTGACAACACTGAAAACAGTATTGGATCAGCAGC

4441

AATCCGCCCTAATAAACTTTCTCCGACACTGTTGTGACTTTTGTCATAACCTAGTCGTCG

AGACCCCTTCTAGGTTAGCAGTGACCCGTGTCATCCAGGCATTGGCCATGAAGGGTGATG

4501

TCTGGGGAAGATCCAATCGTCACTGGGCACAGTAGGTCCGTAACCGGTACTTCCCACTAC

HphI MseI

TTGAAAACATAGAAGTAGTTCAGAAGATGTTAAATGGACTCGAAGACTCCATTGGACTTT

4561

AACTTTTGTATCTTCATCAAGTCTTCTACAATTTACCTGAGCTTCTGAGGTAACCTGAAA

CAAAAATGGTTTTCATCAATAACATTGCTTTGGCTCAAATAAAGAATAATAACATAGATG

4621

GTTTTTACCAAAAGTAGTTATTGTAACGAAACCGAGTTTATTTCTTATTATTGTATCTAC

CCGCAATAGAAAACATTGAAAATATGCTTACTTCAGAGAATAAAGTCATTGAACCCCAAT

4681

GGCGTTATCTTTTGTAACTTTTATACGAATGAAGTCTCTTATTTCAGTAACTTGGGGTTA

ACTTCGGCTTGGCATACTTATTCAGAAAAGTAATAGAGGAGCAGTTGGAACCAGCAGTTG

4741

TGAAGCCGAACCGTATGAATAAGTCTTTTCATTATCTCCTCGTCAACCTTGGTCGTCAAC

AAAAGATAAGCATCATGGCGGAGAGATTGGCCAATCAGTTTGCAATTTATAAACCTGTCA

4801

TTTTCTATTCGTAGTACCGCCTCTCTAACCGGTTAGTCAAACGTTAAATATTTGGACAGT

AluI

CTGATTTTTTTCCTTCAACTTGTGGATGCAGGCAAGGTGGATGATGCCAGAGCTCTCCTAC

4861

GACTAAAAAAGGAAGTTGAACACCTACGTCCGTTCCACCTACTACGGTCTCGAGAGGATG

AGAGATGTGGTGCAATTGCTGAACAAACCCCGATTTTGTTGTTGTTCCTCCTTAGGAATT
4921
---------+---------+---------+---------+---------+---------+---------+

TCTCTACACCACGTTAACGACTTGTTTGGGGCTAAAACAACAACAAGGAGGAATCCTTAA

MseI
|
CTAGGAAACAAGGAAAGGCATCAACTGTGAAATCTGTGTTAGAATTGATTCCTGAATTAA
4981
---------+---------+---------+---------+---------+---------+---------+

GATCCTTTGTTCCTTTCCGTAGTTGACACTTTAGACACAATCTTAACTAAGGACTTAATT

AluI
|
ATGAAAAGGAAGAAGCATACAATTCCCTCATGAAAAGCTATGTCTCAGAGAAAGATGTCA
5041
---------+---------+---------+---------+---------+---------+---------+

TACTTTTCCTTCTTCGTATGTTAAGGGAGTACTTTTCGATACAGAGTCTCTTTCTACAGT

CATCTGCTAAAGCACTGTATGAACATTTGACTGCAAAGAATACAAAATTGGATGATCTGT
5101
---------+---------+---------+---------+---------+---------+---------+

GTAGACGATTTCGTGACATACTTGTAAACTGACGTTTCTTATGTTTTAACCTACTAGACA

TTCTAAAGCGTTACGCATCTTTGCTGAAGTATGCTGGAGAGCCTGTCCCTTTCATTGAAC
5161
---------+---------+---------+---------+---------+---------+---------+

AAGATTTCGCAATGCGTAGAAACGACTTCATACGACCTCTCGGACAGGGAAAGTAACTTG

AluI AluI
| |
CCCCTGAAAGCTTTGAATTTTATGCACAGCAGCTAAGAAAATTGAGGGAAAACTCTTCTT
5221
---------+---------+---------+---------+---------+---------+---------+

GGGGACTTTCGAAACTTAAAATACGTGTCGTCGATTCTTTTAACTCCCTTTTGAGAAGAA

AflIII
|
GAAATAACCAGGCGATACTTTGTTTTGTATATATTTGTGATTCTGTGTCTACATGTTATT
5281
---------+---------+---------+---------+---------+---------+---------+

CTTTATTGGTCCGCTATGAAACAAAACATATATAAACACTAAGACACAGATGTACAATAA

TTGAAGTATATCTGAGGGAAAAATAAATGAAAATTTTCTTTATGTACTTATGTATGTGTG

5341

---------+---------+---------+---------+---------+---------+---------+

AACTTCATATAGACTCCCTTTTTATTTACTTTTAAAAGAAATACATGAATACATACACAC

ATGCATGTTCAAAGTCTTATTGACCATAACTCTGTGCACTTGGTTATTGGACATTTTTGG

5401

---------+---------+---------+---------+---------+---------+---------+

TACGTACAAGTTTCAGAATAACTGGTATTGAGACACGTGAACCAATAACCTGTAAAAACC

AGTTTTTTCTCTGGGAAAAATCGATAGTGTTTTCTTCAATGCTGCTGCTGTGTGAAGCCA

5461

---------+---------+---------+---------+---------+---------+---------+

TCAAAAAAGAGACCCTTTTTAGCTATCACAAAAGAAGTTACGACGACGACACACTTCGGT

TACTTTTCAGGATTCTTCCCTAATTGGCTCTTTGGTTTCCCTGCTCTGTTTCATTTATTT

5521

---------+---------+---------+---------+---------+---------+---------+

ATGAAAAGTCCTAAGAAGGGATTAACCGAGAAACCAAAGGGACGAGACAAAGTAAATAAA

MseI MseI

CATTAAAATGTTATTCCTTTATTTAAGATTCACTTATTAGTCTGCTGTTTCTCTGAAAAA

5581

---------+---------+---------+---------+---------+---------+---------+

GTAATTTTACAATAAGGAAATAAATTCTAAGTGAATAATCAGACGACAAAGAGACTTTTT

AluI

TTTTAGAGCTAGGTATAGTGACCGTGAACTTCTAACGCATAATATCTGTGATACAGCCAT

5641

---------+---------+---------+---------+---------+---------+---------+

AAAATCTCGATCCATATCACTGGCACTTGAAGATTGCGTATTATAGACACTATGTCGGTA

AflIII MseI

TCCGTACATGTGTGAGTCTGCATAACTTTCGAACTTTCGAACTTTGTTAAATGTTGGCAC

5701

---------+---------+---------+---------+---------+---------+---------+

AGGCATGTACACACTCAGACGTATTGAAAGCTTGAAAGCTTGAAACAATTTACAACCGTG

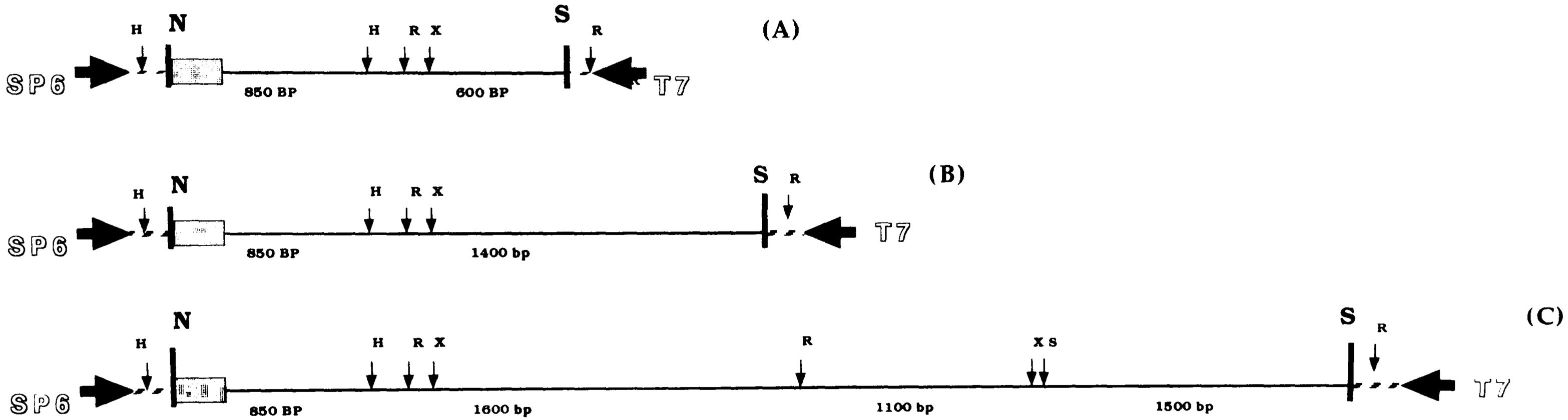
TAGGAGTCATCAGATCTAGGATTCATCATTTTCCAGTGAGAAGCAGAGACCCAAAGGCCT

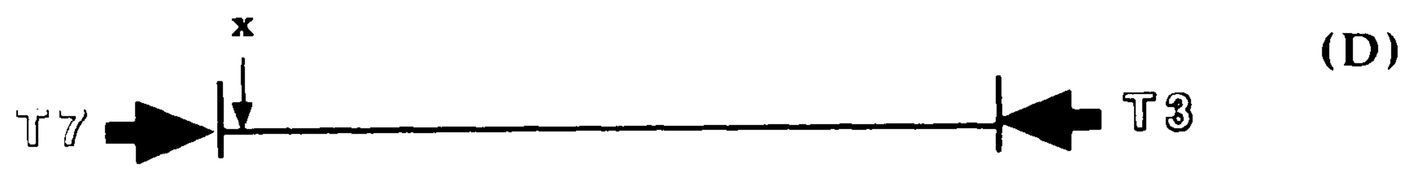
```
   5761
--------+---------+---------+---------+---------+---------+---------+-
     ATCCTCAGTAGTCTAGATCCTAAGTAGTAAAAGGTCACTCTTCGTCTCTGGGTTTCCGGA

     GTTACTTGTGCTTGGTCAGGGGACTGTCTGTCATGCCTGGAGGCTCTTCGGCACACTTCC
   5821
--------+---------+---------+---------+---------+---------+---------+
     CAATGAACACGAACCAGTCCCCTGACAGACAGTACGGACCTCCGAGAAGCCGTGTGAAGG

     CCATCTTTCCCTTCTGCCACTGTGGCTTCAAGCACCTCTGTTCATAGAGCGTCTCTGAAA
   5881
--------+---------+---------+---------+---------+---------+---------+
     GGTAGAAAGGGAAGACGGTGACACCGAAGTTCGTGGAGACAAGTATCTCGCAGAGACTTT

     TTGAGTCTCGGTCATGACTTATCCCGAAGTAGAGCAATGTGTTTCCTCTCATTGTAGTTT
   5941
--------+---------+---------+---------+---------+---------+---------+
     AACTCAGAGCCAGTACTGAATAGGGCTTCATCTCGTTACACAAAGGAGAGTAACATCAAA

                 AluI
                  |
     CAGGACTTTGTCAGTACAAGCTCTGCCCTAGGCTTCTTACTTTATACTCATATCCTGAAA
   6001
--------+---------+---------+---------+---------+---------+---------+
     GTCCTGAAACAGTCATGTTCGAGACGGGATCCGAAGAATGAAATATGAGTATAGGACTTT

                                  MseI
                                   |
     AGATGTGATTTCATCTATGAAGGGGTAAAATATTGGTTTGTATTTAATTGTTTGAAATAA
   6061
--------+---------+---------+---------+---------+---------+---------+
     TCTACACTAAAGTAGATACTTCCCCATTTTATAACCAAACATAAATTAACAAACTTTATT

     AAGTGATCCCTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
   6121
--------+---------+---------+---------+---------+---------+---------+
     TTCACTAGGGATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

     AAAAAAAAAAAAAAAA
   6181 --------+------- 6197
     TTTTTTTTTTTTTTTT
```

Enzymes that do cut:

AflIII AluI HpaII HphI MseI MspI

Appendix 4: Maps of cDNAs M0982 (A), GR-6 (B), GR-5 (C) and CoCa5 (D)





H = Hind III
N = Not I
R = Eco RI
S = Sal I
X = Xba I

Sp6 = Sp6 Promoter
T7 = T7 Promoter

A = M0982
B = Gr-6
C = Gr-5
D = CoCa 5

Poly-A tail

Vector polylinker

Appendix 5: STS and EST primer sequences used in thesis

Target sequence	Upstream primer 5' - 3'	Downstream primer 5' - 3'	Annealing temp. (oC)	Amplicon size (bp)
AB1-Alumorph	TTAGAAATCACTCCTCACCGG	AGGTACACGTGAACTGGCCTA	58	150
AB1- TETRA	TATTGATTGATTGATTGATTGATTTT	CCCATGCTGCTGTGTGTCC	58	300
M0982-(T7-SP6 PCR)	AAATCTGAACTCAGTGTCTGCTTCTT	TAAGAACAGTTATTTCCACAGCATAG	62	1300
CYP1B1	TTTTGAGTCAAAGACTTAAAGGGC	TATTTTCCTGAAAAGGTGAAAAGG	58	350
D2S177	AGCTCAGAGACACCTCTCCA	CTGTATTAGGATACTTGGCTATTGA	56	154
Leu-rich 130KD gene	GAATAAAGTCATTGAACCCC	CTCTGGCATCATCCACCTT	52	400
Na/Ca-UPSTR	CCGGAGGATTTTGAGGACAC	ACTAACAGATGAAATCCCATTG	58	220
Na/Ca-DWNSTR	GGAACTATCAGATATAGTAAA	TTGAACATGATGATGCCCC	50	190
rDNA-132 (R)	GCCACACAGTTAGTTGAACAA	GGTGCATTATATGTGAAAGACG	50	130
rDNA-281 (R)	GGGAGATGAGATCCCAG	GGTTTGATCAACTATATCT	52	280
G09101-sts	TTCTTCATCTGCGACGTAC	CGATCGCGCACGAGACAG	50	200
MTase	GATAGAGATCAAGCTGCC	ATGAGGGTGTTGAACTGC	50	1100
GAPDH	ATCAAGAAGGTGGTGAGCAGG	CACATACCAGGAAATGAGC	58	174
TP53 - EX5	TCCTTCCTCTTCCTACAGTACTCCCCT	GCCCCAGCTGCTCACCATCGCTA	58	212
TP53 - EX7	TCCTAGGTTGGCTCTGACTGT	AGTGGCTCCTGACCTGGAGTC	58	130
D17S379	GGAAATTTCACCTAGCCTGGGAG	CAGATTGCAAAGGTTCCCTTGAG	58	200
D17S1174	TGCACAGGTACCCCAGAACT	CCATTACCTCAGTCAGTAAGTGAACACAA	55	96
D17S5 (YNZ 22)	CGAAGAGTGAAGTGCACAGG	CAGAGTCTTTATTCTTCAGCG	58	168-938
D17S28 (YNH37.3)	TCCAGTGCCACCTGACACC	GGCTAAACAGGAGTTGGTGTG	59	98
UTAH 20	GCAGTGGGCCATCATCA	CCGCAGAAGGTGTTGT	50	250
TM45	CCTTCTCGCTGCTGGTGG	AGCGAGCAAGGGAGTGAG	56	170
HIC-1 3'UTR	CCTCATTTTGCACAAGTGGCAC	ACAAACGAAGAGGAGGGAGCTAAAG	59	300
HIC-1 5'	CCTTCTCGCTGCTGGTGG	AGCGAGCAAGGGAGTGAG	58	250
25BH6-RE	GAGGTTAATTGACTTGCCTAGAGCC	CAGGCAAGGTCGCCTTTTTTAA	57	70
25BH6-LE	CCAGACCAACAGAGAGAACTCTG	GGCCCAGAATTCAAGTTGGTTT	57	400
38B8E-RE	CCAGTCCAGGCAGTCCAC	TTACGGAATTCCTGGCCTC	57	117
38B8E-LE	CAGGAGAGAGTGGAGCAGGAA	CTTCGAACCCAAGGTGTTCTG	57	120
PI-TP EX2	GTAAAGGGTTGGACGCGACA	TGATATTCACAAGTACTCATGAA	54	176
PI-TP EX3	TGAACTGTTGTGAATAATGCT	GTGCCACCACTTGCAAAGCA	58	134

```
                                10        20        30        40        50        60
3-S2.Assemblage.1   AAAAAtCACTTTGGGAGGCCCAGGAAGGcGGA~TCATTAGGAGTCAACCTACGGGCAGAG

5-s2Bis          <-  AAAAA-CACTTTGGGAGGCCCAGGAAGGCGGA-TCATTAGGAGTCAACCTACGGGCAGAG
3-s2             <-  AAAAA-CACTTTGGGAGGCCCAGGAAGGCGGA-TCATTAGGAGTCAACCTACGGGCAGAG
3-s2Bis          <-  AAAAA-CACTTTGGGAGGCCCAGGAAGGCGGA-TCATTAGGAGTCAACCTACGGGCAGAG
1-t3bis          <-   AAAANCACTTTGGGAGGCCCAGGAAGGNGGA-TCATTAGGAGTCAACCTACGGGCAGAG
1-t3Bis          <-   AAAATCACTTTGGGAGGCCCAGGAAGGNGGA-TCATTAGGAGTCAACCTACGGGCAGAG
1-t3Bis (si)     <-   AAAATCACTTTGGGAGGCCCAGGAAGGNGGA-TCATTAGGAGTCAACCTACGGGCAGAG
A160960.SIMPLE   ->                                  ACTCA--AGGAGTCAACCTACGG-CAGAG

                                70        80        90        100       110       120
3-S2.Assemblage.1   AGGACTACGTCATTCTGAGCGTAAGGGCCACAGCGAAAGGTGGCAGGGCCCGCGCTTTaA

5-s2Bis          <-  AGGACTACGTCATTCTGAGCGTAAGGGCCACAGCGAAAGGTGGCAGGGCCCGCGCTTTAA
3-s2             <-  AGGACTACGTCATTCTGAGCGTAAGGGCCACAGCGAAAGGTGGCAGGGCCCGCGCTTTAA
3-s2Bis          <-  AGGACTACGTCATTCTGAGCGTAAGGGCCACAGCGAAAGGTGGCAGGGCCCGCGCTTTAA
1-t3bis          <-  AGGACTACGTCATTCTGAGCGTAAGGGCCACAGCGAAAGGTGGCAGGGCCCGCGCTTTNA
1-t3Bis          <-  AGGACTACGTCATTCTGAGCGTAAGGGCCACAGCGAAAGGTGGCAGGGCCCGCGCTTTNA
1-t3Bis (si)     <-  AGGACTACGTCATTCTGAGCGTAAGGGCCACAGCGAAAGGTGGCAGGGCCCGCGCTTTNA
A160960.SIMPLE   ->  AGGAC-ACGTCATTCTGAGCGTAAGGGCCACAGCGAAAGGTGGCAGGGCC-GCGCTTTTA

                                130       140       150       160       170       180
3-S2.Assemblage.1   AAGGGCTGAaATCcAAACGG~TCAGtACATACTCcGTTTCCAGCACTTTGGGAGGCCCAG

5-s2Bis          <-  AAGGGCTGAAATCCAAACGG-TCAG-ACATACT
3-s2             <-  AAGGGCTGAAATCCAAACGG-TCAG-ACATACT
3-s2Bis          <-  AAGGGCTGAAATCCAAACGG-TCAG-ACATACT
1-t3bis          <-  AAGGGCTGANATCNAAACGG-TCAGNACATACTCC
1-t3Bis          <-  AAGGGCTGANATCTAAACGG-TCAGTACATACTCC
1-t3Bis (si)     <-  AAGGGCTGANATCTAAACGG-TCAGTACATACTCC
A160960.SIMPLE   ->  AAGG-CTGAAATCCCGGCGGCTCAGG-CCTG-TC-GTTTCCAGCACTTTGGGAGGCCCAG

                                190       200       210       220       230       240
3-S2.Assemblage.1   GAAGGCGGATCATTTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACGTGCACAAACCCC

A160960.SIMPLE   ->  GAAGGCGGATCATTTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACGTGCACAAACCCC
```

Appendix 6: Sequence alignment of Bisulphite treated and non-treated DNA from the pancreatic cell lines SUIT2 (methylated) and T3M4 (hypomethylated). With the T3M4 DNA, there are instances of ambiguity at positions where there should be Cytosine nucleotides (base 29) and two instances of C-T base transitions, the latter occurring at a CpC dinucleotide rather than CpG however.

(Elelctrophoretograms were not available to print).

Appendix 7: Sequence of Alumorph fragments

Alumorph AB1 sequence - incomplete

TGTACAGTGC CATTTGGTTA GAAATCACTC CTCACCGGAA CACACACTTA
CTGGTACCGT GTAGAGATTT TAGTTAGTTA GTTAGTTAGT TATTATCGAT
TGCCGCCCNG GGGGAGGTGA GGAGACCGAG GTTTTGGCTC TTAGCTGCCC
AGGTACACGT GAACTGGCCT AGGGGACTCA AGGAGTAACT ACCTGGACAG
AGAGGACTGC ATTCCGAGT - **219 bp**

Alumorph IP-1 sequence - M13 forward primer

GGGGGCCCCT GGGGGGACTG ATGATCTCAA GGGACGTTCC GCATATGGAA
TTTTAAATTG GGCATCTGGG CACCCGGCTG TCAGAAGTTC CACCCGGATC
ACAAGTGGCA TAGCCCCAGG GCCAGGTGAT GGCAACACCA ACAGTGTTGC
CCAGCAGCAA TTGGGGTCAG TTACCCATGG GGATGGAACA TGTGGACTGT
CTTGAGGGTG ACCGCAGTCA AGCACATCGC ATGGTCCAGC AGTCATTCTG
CATCCGATGG GGAGCACATT CCGGGCGCCC TAG - **283 bp**

Alumorph Met-U complete sequence

CTGGGATTAC AGGCGTGACA GCAACACACC CAGCCTAAAG CAGGGTTTCT
CAACAGTGGC ACTACTGACC TGTGGGCCAG ATAGTGTTAT ATATATGTAG
AGCGTGTGAG CATCATAGAG ATTTAGCAGT ATCTGTTTTT ACACTAGCAC
GATGACACCC CACCCTCAAC TGTGACAATC CGGGTGTCTA CAGATATTAC
CAATGTCTCT TGGANNAAAA TTGTCCCCAG TTTATGGAAG TGACTTAGAT
ACTAGAACTG GGGGCGGGCG CACTGGTCAC GCCTGTAATC CCA - **293 bp**

Alumorph Met-L complete sequence

CTGGGATTAC AGGCGTGACG GCTGTGCCCG GCTGCACAAA CACTTTCAGA
AAGCAACAGA GGACGAAACA TTTTCTAACT CATTTATAAG CCAGCATAAC
CTTGATCCAA AGCCTAAGGA GATTAAATAA AATTGAGGAC CGAAACCTAG
TCCAATACGG ACCGAATTAT TTACTCAATC TTTTACAAAG CANGAGACAA
CGAAGACTTT CACAAACACG CTGGCCGTTC GCCAG - **234 bp**

Alumorph Met6 sequence - M13 reverse primer

AACCCTCACT AAAGGGAACA AAAGCTGGGT ACCGGGCCCC CCCTCGAGGT
CGACGGTATC GATAAGCTTG ATATCGAATT CAATCGGGGA TGTCTGAGGG
ACCTTCCGCG GCATCTATGC GGGCATGGTT ACTGCCTCTG GTGCCCCCCG
CAGCCGCGCG CAGGTACCGT GCGACATCGC GATGGCCCAG CTCCTCAGCC
AGGTCCACGG GCAGACGGGC CCAGGCATCG CGCTCGTCCA GCCGCGCCCC
GGCCCCGGTGC AGGACCACCA GCGTGTCCAG GAAGCCCTCC CGGGCAGAGT
CGTGCACGGG TCGGGTGAGA GTGGCGGGGT CGGCGCAAGT TGGGGCTCCG
CGCCGTGGGA GCAGCAGCAG CTCCGCCACT CGGGCGGTTG CCCATCATCA
ATGAACTGGG - **410 bp**

Alumorph AB3 complete sequence

AACTTTTCTT TTTTTTTTTT GAGACAGAGT CTTGCTCTGT CGCCCCAGGC
TGGAGTGCAG TGGCGCGATC TCGGCTCACT GCCACCTCAC TGCCACCTCT
GCCTCCCGGT TCACGCCATT CTCTCCTGCC TCAGCCTCCC CGAGTAGCTG
GCACTACAGG CGCCCCGCCA CCACGCCCGG CTAATTTTTT TGTATTTTTT
TAGTAGAGAT GGGGGTTTCA CCGTGTTCAC CAGGATGGTC TCGATCTCCT
GACCTCATGA TCCGCCCGCC TTGGCCTCCC AAAATGCTGG GATTACACGT
GTGAGGCAAG AACTTTTTAA AAGTGCATCT TGCGCACGGG TAGATCCATT
AAACC - **355 bp**

Alumorph AB4 - complete sequence

CCCCCCCGGG GNCNTTGGGC GCTACGCTCG ATATCGAATT CGGGTTGTCA
GCGAAACAAG AAATGTGAGG TAAAACTGCC ATTCAGTGCA CAACGGGATA
ATTTCACTGT CTCGGAAATG ATTTCAAGTC CCTGGTGGAA ATTGCAACAG
GCTTACTCCA GAACAAGCTG GATGGTTTCC ATGATATTCG AAGGAAGAAG
TAAAAACAGA ATTGCTGCAC AGCGCTGTCG CAAGAGAAAA ACTGGACTGT
ATACAGAATC TTGGATCAGA AATTGAGAAG CTGGTTGATG ATAATTAGAA
GCATGCTTTC CACTGAAACT TCCCGACAAC ATTTGTTATG CAGAATGTCT
CTGAGTGAGA ACTCGGTTTT GCCTATGAAT CTTCTGTGCA TAGCACCAAT
GTTTTACTCA GCCTTAATGA CCAGCGGAAG AAAGATGTGC TGTGCGATGT
CACCATCTTG GGGGNGGGAG AGAGGGTCCG CGCGCTAGGC CAGGCTGATG
GAGAGCTGAA CATTACTTTA CAGAGGGGTG AAAGGCTAAN GGGAATTGGT
TAAGGTNAAA TTGTGGGGTT GGAGTTGCAA TTTTTGGGGA GCGCTTTCAA
TCATTACTTC AGTTTGGGGT CCCTAGGCCC AACTTCCCAA AAAAAAAGCC
CTCTGGGGCC GGGGGCCTTC CTTGGGGCCC CCCCGGGGG - **689 bp**

Appendix 8: details of shotgun clones

Shotgun clone	Insert size (bp)	SK, T7 or M13 R	KS, T3 or M13F	Whole insert	Overlap	Homology
EcoR1-generated						
P1sh1	1,400	483				
P1sh2	2,800	453				
P1sh3	450	467				
P1sh4	1,400	368				
P1sh5	1,300	543				
P1sh6	1,400	501				
P1sh7	1,400					
P1sh8	0					
P1sh9	6,600					
P1sh10	1,500	498				
P1sh11	1,500	487				
P1sh12	2,700	551		P1sh16	P1sh4	
P1sh13	6,600					
P1sh14	1,600	349				
P1sh15	1,300	567				
P1sh16	1,300	620				
P1sh17	1,500	489				
P1sh18	1,500	486				
P1sh19	2,700	487	441	*AB1		rRNA spacer/CYP1B1
P1sh20	0					
Sau3A-generated						
S1	300	76	265	294		
S2	1600	480	479			
S3	1000	477	475	980	S34,S53,P1SH12	
S4	300	477	489	325		
S5	1100	223	227		S30	N1846-T7 END
S6	250	132	153	267		
S7	2500	475	487			
S8	900	132	291			
S9	1250	-	-			
S10	1200	480	475			
S11	1100	226	-			
S12	200	187	-		S22/S39	
S13	300	122	-		S53	
S14	500	0	260			
S15	400	122	178			
S16	1000	474	477			
S17	1700	487	488		S28	
S18	1000	488	476			
S19	1600	220	487		S118	Unknown protein
S20	200	-	-			
S21	0	168	-			
S22	400	0	-		S12/S39	
S23	300	170	140			
S24	0	0	-			
S25	400	201	132			
S26	800	227	-			
S27	1700	0	-			
S28	2100	480	890		S17	
S29	0	0	-			
S30	1000	478	486		S5	N1846-T7 END
S31	400	120	198			
S32	400	-	-		S103	Notch-2 (rat)
S33	300	-				
S34	1800	1032	474		S3,S53,P1SH12	Alu
S35	1500	478	479		S52	
S36	1000	477	483			
S37	250	489	481	223		
S38	1300	224	299			
S39	500	0	297	556		
S40	450	0	172			
S41	300	477	471	365		
S42	250	474	190	302		
S43	750	481	479	720		
S44	250	0	175			
S45	500	0	-	633		

S46	250	480	-			
S47	0	-	-			
S48	1400	484	465			
S49	200	0	-			
S50	250	484	481	288		
S51	900	484	482	792		
S52	700	783	782	715	S3,S34,S56,S114	H+ ATPase
S53	500	441	782	725	S3,S13,S34,P1SH12	Alu
S54	250	-	-			
S55	0	-	-			
S56	0	-	-		S52/S114	H+ ATPase
S57	300	-	-			
S58	750	0	-			
S59	600	787	806	902		
S60	250	0	-			
S61	0	-	-			
S62	1300	0	-			
S63	1300	0	-			
S64	1000	0	-			
S65	1100	0	215			
S66	1800	0	243			
S67	1000	0	261			
S68	2000	0	305			
S69	700	0	-			
S70	600	0	-			
S71	800	91	540			
S72	700	721	511			
S73	700	133	0			
S74	1300	146	0			
S75	1100	187	0			
S76	700	189	-			
S77	900	211	-			
S78	0	-	-			
S79	700	0	-			
S80	0	-	-			
S81	2700	748	433			
S82	0	-	-			
S83	1900	754	387			
S84	0	-	-			
S85	700	772	222			
S86	?	-	-			
S87	?	-	-			
S88	?	-	-			
S89	?	-	-			
S90	?		-			
S91	?		-			
S92	?		-			
S93	?		-			
S94	?		-			
S95	?		-			
S96	?		-			
S97	1000	797	682			
S98	500	798	434	827		
S99	350	-	-			
S100	800	-	-			
S101	800	-	606			
S102	450	-	-			
S103	500	779	411	532	S32	Notch-2 (rat)
S104	1200	-	748			
S105	500	234	412			
S106	300	231	208	297		
S107	0	-	-			
S108	500	272	245			
S109	400	-	-			
S110	150	-	-			
S111	300	-	-			
S112	400	-	-			
S113	400	-	-			
S114	800	760	762		S52/S56	
S115	250	-	-			
S116	1200	775	768			

S117	250	-	-		S19	Unknown protein
S118	700	798	740			
S119	0	-	-			
S120	400	-	-			

BIBL
LONDON

GENES, CHROMOSOMES & CANCER 18:30–41 (1997)

Alu-Polymerase Chain Reaction Genomic Fingerprinting Technique Identifies Multiple Genetic Loci Associated With Pancreatic Tumourigenesis

Arthur B. McKie,[1] Takeshi Iwamura,[2] Hing Y. Leung,[1] Michael A. Hollingsworth,[2] and Nicholas R. Lemoine[1]*

[1]Imperial Cancer Research Fund Oncology Unit, Royal Postgraduate Medical School, Hammersmith Hospital, London, United Kingdom
[2]Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska

DNA fingerprinting can be used to detect genetic rearrangements in cancer that may be associated with activation of oncogenes and inactivation of tumour suppressor genes. We have developed a fingerprinting strategy based on polymerase chain reaction (PCR) amplification of genomic DNA with primers specific for the Alu repeat sequences, which are highly abundant in the human genome. This has been applied to DNA from pancreatic cancer and paired normal samples to isolate and identify fragments of genomic DNA rearranged in the malignant cells. These fragments have been sequenced and used as probes to isolate hybridising clones from gridded bacteriophage P1, phage artificial chromosome, and cosmid libraries for fluorescent in situ hybridisation mapping and the identification of expressed sequences. Further characterisation has identified a putative novel gene (*ART1*) that is up-regulated specifically in pancreatic cancer as well as another sequence with similarity to genes involved in differentiation (POU domains). In conclusion, we suggest that Alu-PCR fingerprinting may be a useful technique for the identification of genes involved in tumourigenesis. *Genes Chromosom Cancer 18:30–41, 1997.* © 1997 Wiley-Liss, Inc.

INTRODUCTION

Pancreatic cancer is a major public health problem in the industrialised world, and it is now the fifth most common cause of cancer death in the USA and Europe (Niederhuber et al., 1995). It remains a disease that is difficult to diagnose and that is virtually incurable with current treatment modalities, including surgery, radiotherapy, and chemotherapy. In an effort to identify new genetic markers for early diagnosis and novel targets for rational therapy, several laboratories are attempting to define the molecular events involved in pancreatic tumourigenesis. A variety of candidate genes have been screened for mutations and/or changes in expression level, with the result that several consistent abnormalities have been identified. The first to be discovered was the remarkably high prevalence (at least 75% of cases) of mutations at codon 12 of the *KRAS* oncogene (Almoguera et al., 1988), and such mutations occur at an early stage of tumourigenesis (Lemoine et al., 1992), perhaps even in preneoplastic lesions (Yanagisawa et al., 1991; Caldas et al., 1994b; DiGiuseppe et al., 1994). The tumour suppressor genes *TP53* and *CDKN2* are inactivated very frequently by mutation and/or allelic deletion (Barton et al., 1991; Kalthoff et al., 1993; Scarpa et al., 1993; Caldas et al., 1994a; Redston et al., 1994; Bartsch et al., 1995; Liu et al., 1995), and abnormalities of *TP53* are found in early preinvasive cancers (Barton et al., 1991). In addition, a novel putative tumour suppressor gene, *DPC4*, has recently been reported to be involved in a significant proportion of cases of primary pancreatic cancer (Hahn et al., 1996). Although some controversy remains, most investigators find little evidence for the involvement of other tumour suppressor genes, such as *APC*, *RB1*, *DCC*, and genes encoding DNA repair enzymes, at least in Western populations (Lemoine et al., 1993; Schutte and Kern, 1995). A significant proportion of cases appear to lack some or all of the typical genetic abnormalities, suggesting that we still do not know the full complement of genes involved in transformation of human pancreatic epithelial cells.

Cytogenetic analysis is particularly difficult to achieve in pancreatic cancer specimens, but two groups have confirmed loss or rearrangement of chromosomal regions that are known to carry genes important in the disease as well as suggesting involvement of others, such as 6q (Bardi et al., 1993; Johansson et al., 1994; Griffin et al., 1995; Guronova et al., 1995). This has been taken a step further by molecular genetic approaches for the survey of loci throughout the genome that look for allelic imbalance, as revealed by DNA restriction length polymorphisms (RFLPs) and by more informative mi-

*Correspondence to: Nicholas Lemoine, ICRF Oncology Unit, RPMS, Hammersmith Hospital, London W12 0NN, United Kingdom.

Received 2 April 1996; Accepted 24 June 1996

© 1997 Wiley-Liss, Inc.

BIBL. LONDIN. UNIV.

crosatellite DNA repeats. Elimination of one allele of a tumour suppressor gene may be associated with concurrent loss of heterozygosity (LOH) at linked polymorphic genetic loci, and mapping of such losses in large series of cases may allow definition of a minimal area of overlap as a starting point for positional cloning or screening of candidate genes. Adenocarcinoma of the pancreas presents a serious technical problem in that tumours usually display a marked stromal reaction, producing a highly desmoplastic tumour. LOH studies, thus, are difficult to achieve convincingly without painstaking microdissection of the tumour cells away from fibrotic stromal elements or the use of xenograft enrichment protocols. The approach of looking for LOH as a starting point for the identification of candidate genes is laborious, and there are few reliable reports of allelic loss in carcinoma of the pancreas (Hahn et al., 1995).

Alternative strategies may be needed to identify pancreatic cancer-specific lesions. The last few years have seen the development of several novel polymerase chain reaction (PCR)-based techniques that scan the whole genome to identify genetic lesions in various diseases directly or to generate markers, which are associated specifically to such phenotypes. For instance, representational difference analysis (RDA), which is a technique described originally by Lisitsyn et al. (1993) that scans for RFLPs, has been applied successfully to pancreatic cancer for the identification of a deleted region that involves the *BRCA2* region at chromosome band 13q12 (Schutte et al., 1995).

Here, we describe a simple technique based on the PCR amplification of DNA fingerprints using primers specific to the Alu repeat element. Nelson and colleagues first used amplification of stretches of anonymous human DNA from somatic cell hybrids by employing human repeat sequence-specific primers (Nelson et al., 1989). In particular, they chose the Alu repeat element, which is distributed ubiquitously throughout the genome. Alu represents the major family of short interspersed repeat elements (SINEs) in the mammalian genome, with approximately 1,000,000 copies of a 300 bp sequence in humans (Schmid, 1982; Daniels et al., 1983). This is now a routine tool for the analysis of the human DNA content of somatic cell hybrids or yeast artificial chromosomes (YACs). For our application, however, ladders of random anonymous DNA fragments from human genomic DNA are generated, and the profile of bands are compared between tumour and normal DNA extracts from the same patient. This Alu-directed PCR strategy leads to amplification of random fragments of DNA that can show quantitative and/or qualitative differences between tumour and normal DNA samples. Including only one Alu-specific primer in a PCR amplification means that one may amplify sequences between two Alu repeats in an inverted orientation relative to each other. A "fingerprint" of random amplified fragments can be displayed by gel electrophoresis. The abundance of the Alu repeat element will result in such reactions yielding a complex pattern of Alu-generated intervening fragments that can be simplified by predigesting the genomic DNA with frequently cutting restriction endonucleases. These restriction enzymes will cleave the DNA template at a proportion of the sites between inverted Alu repeats, thus reducing the number of fragments generated to dozens and achieving a simpler fingerprint pattern. By performing the amplification with a particular primer on DNA samples cut with different restriction enzymes or, alternatively, by using one of eight different Alu-specific primers varying in their relative specificity to the Alu repeat element, one can produce fingerprint patterns that vary in complexity (and, hence, informativeness) and consistency between individuals.

MATERIALS AND METHODS

Clinical Specimens

Several surgical units provided tissues for investigations, including Prof. R.C.N. Williamson, Royal Postgraduate Medical School, London, and Mr. R.C.G. Russell, The Middlesex Hospital, London. Samples of the resected tissue were snap frozen in liquid nitrogen immediately after surgery, and DNA was prepared from tumour-rich tissue blocks as described previously (McKie et al., 1993).

SUIT 2 Cell Lines

The human pancreatic tumour cell line SUIT 2, which was derived from a liver metastasis, shows spontaneous metastasis to lung and regional lymph nodes after subcutaneous injection of the tumour cells into the lateral flank of nude mice. Twenty-eight sublines were cloned from SUIT 2 by culture in soft agar and were characterised for their morphological and metastatic properties (Iwamura et al., 1987; Taniguchi et al., 1992; Taniguchi, 1994). Four sublines were selected for further study: S2-007 (moderately differentiated tubular adenocarcinoma, highly metastatic), S2-020 (poorly differentiated tubular adenocarcinoma, moderately metastatic), S2-013 (well-differentiated tubular

adenocarcinoma, moderately metastatic), and S2-028 (papillotubular adenocarcinoma, rarely metastatic). Two additional cell lines were established by sequential selection either of lung colonies arising in nude mice following tail vein injection (SP-VP10, moderately differential tubular adenocarcinoma, high potential for lung colonisation) or of spontaneous metastases in the lung following subcutaneous injection (S2-CP9, moderately differentiated tubular adenocarcinoma, very highly metastatic to lungs). S2-VP10 and S2-CP9 are cell lines that were established after ten and nine cycles of selection, respectively. All of these cell lines have distinct metastatic and morphological properties. They were maintained in Dulbecco's modification of minimal essential medium (DMEM) supplemented with 5% fetal bovine serum (FBS).

Restriction Digestion

Restriction digests of tumour/normal DNA samples from each individual were performed overnight as recommended by the enzyme manufacturer (New England Biolabs). Ten units of restriction enzyme were used in a 20 µl reaction volume containing 1–5 µg of DNA (5 µl).

PCR Amplification

The conditions for PCR amplification were standardised as 50 mM KCl; 10 mM Tris-Cl, pH 8.3; 1.5 mM $MgCl_2$; 0.1% Triton X-100; 0.2 mM dNTP'S; and 1 U of *Taq* polymerase (Bioline Ltd). Approximately 500 ng of DNA were used as a template in 100 µl reaction volumes. The Alu primer concentration was 2 µM, a relatively high concentration that was designed to achieve preferential annealing of the primer to the repeat elements by displacing any rapidly forming duplexes in the template. PCR cycle parameters were 93°C for 3 minutes followed by 30 cycles, consisting of 57°C for 30 seconds, 72°C for 1 minute, and 94°C for 30 seconds. A final extension step of 72°C for 5 minutes was included to allow complete elongation of all products.

Radioactive labelling of PCR products was achieved by the addition of 1-2 µCi of α^{32}P-dATP and by using only 0.1 mM unlabelled dATP in the buffer. Alternatively, the PCR products were labelled by first tagging the primer using [^{32}P] γ-ATP and the enzyme T4 polynucleotide kinase (T4 PNK), see below. PCR amplification was carried out in the Hybaid Thermal cycler.

Electrophoresis and Autoradiography

For the purposes of displaying the Alu PCR-generated ladders, a nondenaturing 6% acrylamide gel matrix was used with an acrylamide:bisacrylamide ratio of 29:1 in Tris-borate-EDTA electrolyte (TBE). The gels were run slowly overnight at 400 V constant voltage on 50 cm Bio-Rad Sequigen systems, with which we achieved consistently good results. For sequencing, a denaturing 6% acrylamide gel matrix was used with an acrylamide: bisacrylamide ratio of 19:1 in TBE buffer equilibrated in 8 M urea. After electrophoresis, the apparatus was dismantled, and the gel was mounted wet on used x-ray film or Whatman chromatography paper, covered with Saran Wrap, then exposed to x-ray film within a cassette at −70°C for up to 4 days. After developing the x-ray film, any interesting bands were cut out carefully from the gel with a scalpel after precise orientation with the autoradiograph. The gel fragments were eluted into 100 µl of a buffer containing 1 × 10 mM Tris-Cl, pH 7.5, and 1 mM EDTA (TE buffer) supplemented with 0.1 M NaCl overnight at room temperature. The DNA fragments could then be amplified by using 10 µl of the eluate in a PCR with the same oligonucleotide primer used in the reaction that generated them.

Cloning of PCR Fragments

Amplified fragments were cloned into Bluescript by using the method of Holton and Graham (1990), based on the observation that *Taq* polymerase adds a single nontemplate-directed deoxyadenosine residue to the 3′ end of duplex PCR products.

Sequencing

Manual sequencing of selected clones was carried out by using the Sequenase 2.0 kit (United States Biochemicals) on double-stranded DNA template after denaturation with alkali. The reactions were then run out on 6% acrylamide gels (19:1 acrylamide: bisacrylamide ratio) saturated with 8 M urea, and autoradiographs were prepared from the dried gels.

Southern Hybridisation

Filters were prepared with paired tumour/normal DNA samples from individual patients. The digested DNA samples from each case were electrophoresed in 0.8% agarose gel for 16–20 hours. Alkaline blotting was carried out onto Hybond N+ nylon filters (Amersham Int. PLC), according to the manufacturer's instructions. The probes were labelled either by using radioactive PCR (for small fragments) or by the use of random hexamers (Feinberg and Vogelstein, 1984). After competition following an established regime (Sealey et al., 1985), filters were hybridised with the probes in

TABLE 1. Primers Used to Amplify from Alu Repeats in Genomic DNA

Cotter IV	5′-GCGACAGAGCCAGACTCCGTCTC-3′
BK 33	5′-CTGGGATTACAGGCGTGAGC-3′
Alu V	5′-AAGTCGCGGCGCTTGCAGTGAGCCGAGAT-3′
BK 34	5′-GATCGCGCCACTGCACTCC-3′
TC 65	5′-TTGCAGTGAGCCGAGAT-3′

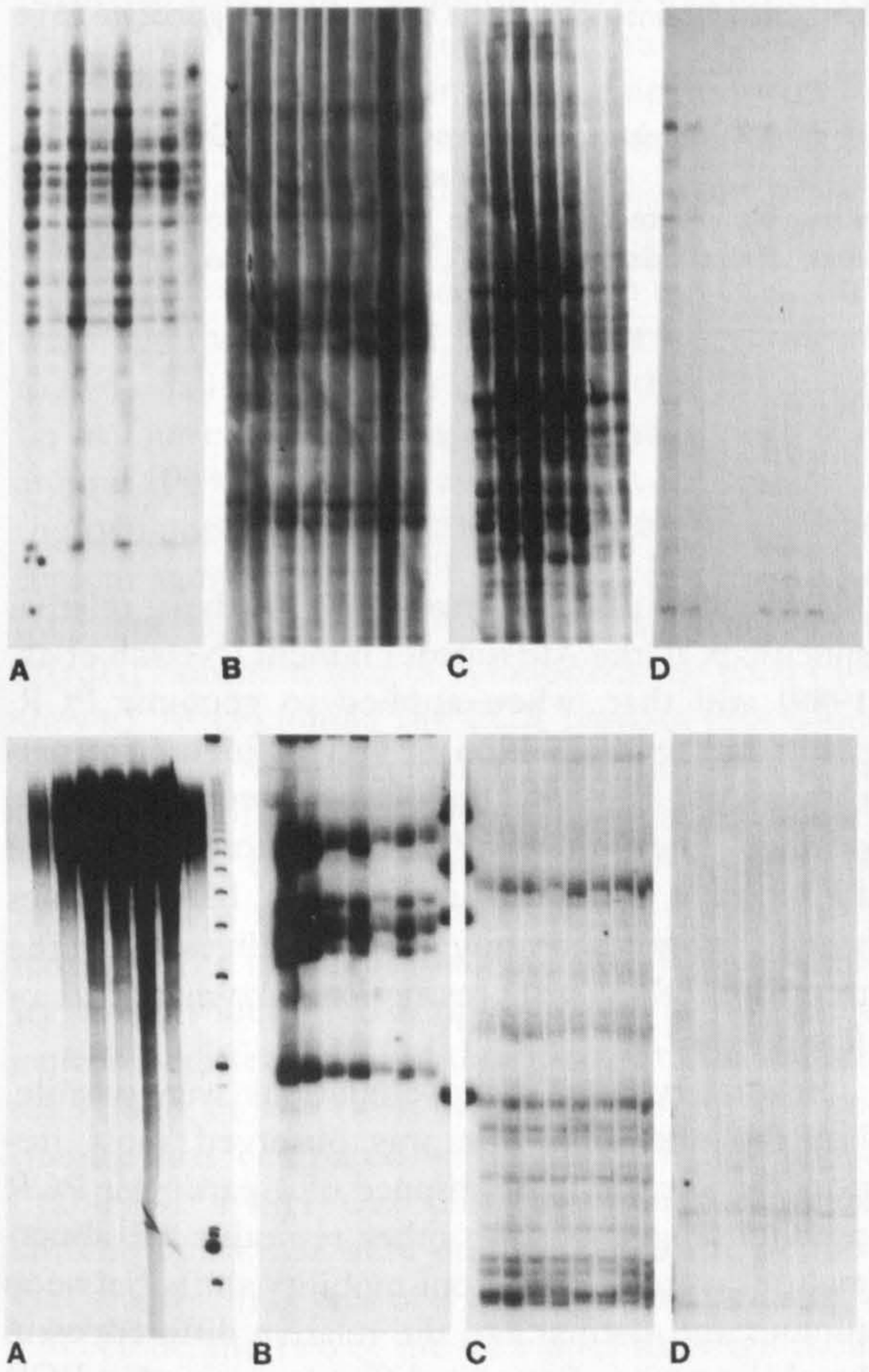


Figure 1. **a:** Autoradiographs illustrating fingerprint profiles achieved with the Alu-specific primers Cotter IV (**A**), Alu V (**B**), BK33 (**C**), and TC65 (**D**). Under each condition, the same four paired tumour/normal cases were digested with the restriction enzyme *Alu*I. **b:** Autoradiograph illustrating fingerprint profiles achieved with the Alu-specific primer BK33 following digestion of three paired tumour/normal cases with different enzymes; *Eco*RI (**A**), *Hin*fI (**B**), *Alu*I (**C**), and *Rsa*I (**D**).

5 × standard saline citrate (SSC), 5 × Denhardt's solution, 0.5% sodium dodecyl sulphate, and 1 mg/ml total yeast RNA at 65°C overnight. The filters were then washed in four successive washes of 2 × SSC/0.1% sodium dodecyl sulfate (SDS), 1 × SSC/0.1% SDS, 0.5 × SSC/0.1% SDS, and 0.2 × SSC/0.1% SDS at 65°C. Filters were exposed to Kodak X-OMAT film for up to 72 hours with intensifying screens.

TABLE 2. Results Using Individual Alu Primers to PCR Amplify Human Genomic DNA Template Predigested with Restriction Enzymes

	RsaI	AluI	HinfI	HaeIII
Cotter IV	10 bands 600–1400 bp	52 bands 250–1400 bp	n.d.	3 bands 250–600 bp
BK 33	7 bands 300–600 bp	43 bands 100–750 bp	9 bands 600–1500 bp	No bands
Alu V	14 bands 300–600 bp	31 bands 100–900 bp	n.d.	n.d.
BK 34	n.d.[a]	18 bands 300–600 bp	n.d.	n.d.
TC 65	n.d.	2 bands 400 & 450 bp	n.d.	n.d.

[a]n.d.: not determined.

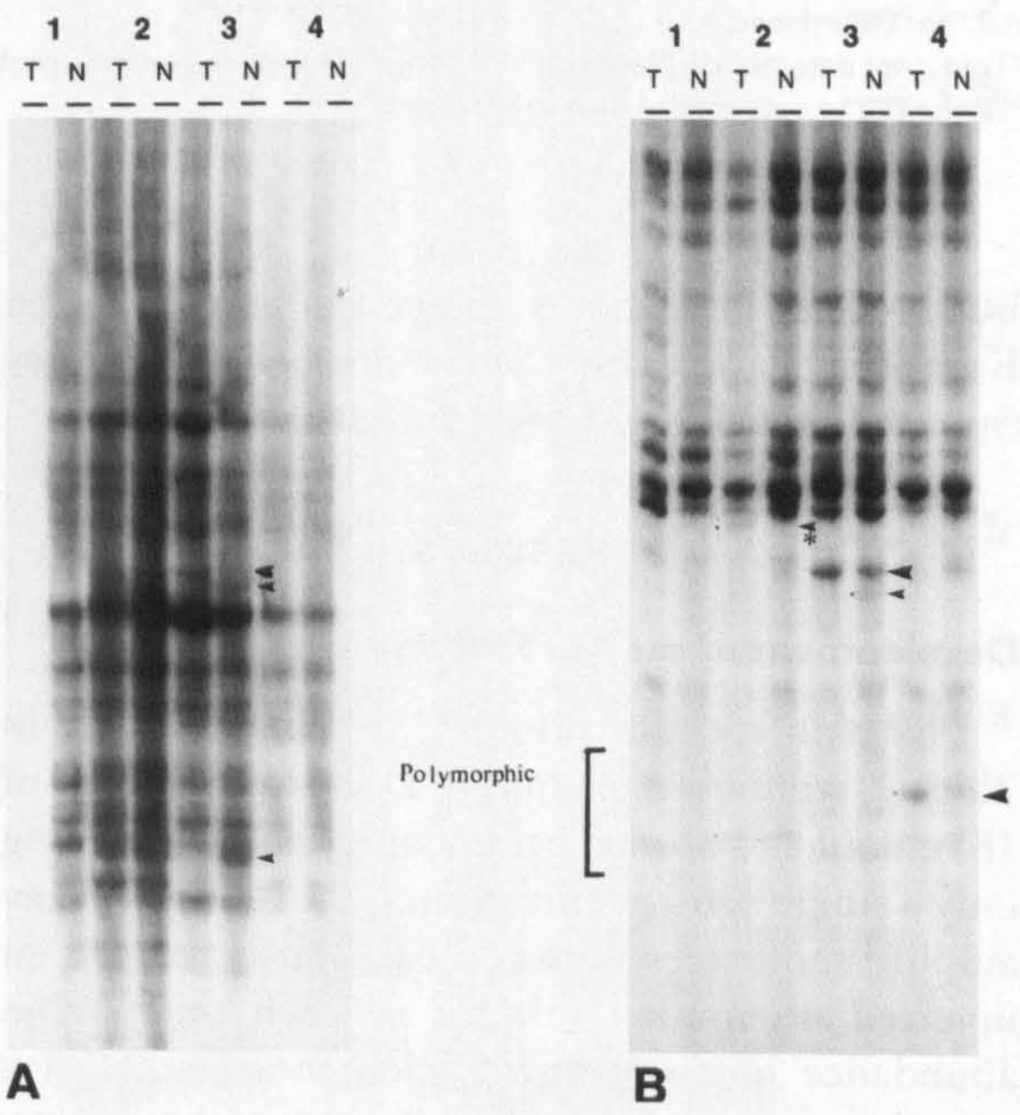


Figure 2. Two fingerprint profiles illustrating the different kinds of Alumorph observed. **A:** Profile generated with the BK33 primer on *Alu*I-digested DNA. Case 2 has an example of a band at 320 bp (single arrowhead), which is present in the normal lane but not in the tumour lane. Interestingly, the same case exhibits a pair of bands showing differences in relative intensity in tumour and normal DNA at 420 bp (double arrowheads). **B:** This profile was generated with the Cotter IV primer on *Alu*I-digested DNA. The figure illustrates polymorphic fragments on the lower half of the figure, and cases 3 and 4 illustrate two band species that show increased signal intensity in the tumour lane compared with the normal lane (large arrowheads). Case 3 has an extra band in the tumour DNA (small arrowhead), whereas case 2 has lost a band in the tumour DNA (small arrowhead with asterisk).

Genomic DNA and Pancreatic cDNA Libraries

The human genomic DNA library in the second generation P1 cloning vector pAd10-sacBII (Pierce et al., 1992) was obtained as gridded filters from the reference library database (RLDB) at the Max-Planck-Institut-fur Molekulare Genetik, Berlin (Lehrach et al., 1990). The pancreatic cancer cDNA

TABLE 3. Characterisation of Alumorph Clones Identified in Analysis of Pancreatic Cancer Samples

Alumorph (size)	Southern blot (genomic DNA)	Genomic clones	FISH localisation	Northern blot (pancreas RNA)	cDNA clones (pancreatic ca library)	Sequence similarities
AB1 (250 bp)	MspI polymorphism with LOH 2/18 cases	3 P1 1 YAC 2 cosmid 2 BAC	2p22	4.0 kb transcript	Four overlapping clones of ART1	Pancreatic ca EST
MET-L (235 bp)	No bands	3 P1	8q24	3.0 kb transcript	n.d.	No significant
MET-6 (500 bp)	HindIII polymorphism 6 kb, 1.8 or 1.6 kb	3 cosmid	11p12–13	10 kb transcript[a]	None isolated	Melanocyte EST
IP-1 (273 bp)	No bands	2 cosmid	17p13.3	2 kb transcript	Two overlapping clones	POU domain[b]
MET-U (295 bp)	No bands	n.d.	Not definitive	None detected	None isolated	No significant
AB4 (750 bp)	3 HindIII bands 10 kb, 6 kb, 3 kb	1 PAC	Not definitive	8 kb transcript[a]	One clone	No significant

n.d.: not determined.
[a]Transcript detected on Northern blot using Alumorph fragment as probe directly.
[b]Similarities to sequence of Alumorph fragment.

library was obtained as gridded filters from the European Pancreatic Cancer Reference Library System (Gress et al., 1994).

RESULTS

Development of the Alu Fingerprints

Five Alu-specific primers, which are listed in Table 1, were used to amplify DNA from a series of 18 tumour and normal paired samples. By including only a single Alu-specific primer in a PCR, one may amplify sequences between two Alu repeats in an inverted orientation relative to each other. The abundance of these repeat elements results in a complex pattern of amplified fragments that is seen as an continuous smear on gel electrophoresis. Simplification of the pattern can be achieved by predigesting the genomic DNA with frequent cutting restriction endonucleases, reducing the number of fragments generated to dozens that can be resolved as a fingerprint by gel electrophoresis. For our application, ladders of random anonymous DNA fragments from human genomic DNA were generated, and the profile of the bands achieved was compared between tumour and normal DNAs from the same cases. By performing the amplification with a particular primer on DNA samples cut with different restriction enzymes, different Alu fingerprint patterns can be generated (Fig. 1a). In addition, by using different restriction digests in combination with different Alu primers, the number of analyses one can carry out on DNA samples is greatly increased (Fig. 1b). We have used five Alu-specific primers that vary in their relative specificity to the Alu repeat element (Nelson et al., 1989) and that, when applied to genomic PCR, result in the production of varied fingerprint patterns depending, on the relative specificity of the primers. The Cotter IV, Alu V, and more degenerate BK33 gave the most informative band profiles when they were combined with predigestion of the template DNA with restriction enzymes *Alu*I or *Rsa*I (Table 2).

Several types of band alterations were visible, but the alteration that was observed most frequently was the total absence of a particular PCR product (Fig. 2A). Two other reproducible abnormalities were the apparent mobility shifts between tumour and normal and the relative differences in band intensity. Sixteen differences in Alu-PCR products, or Alumorphs, were identified after acrylamide gel electrophoresis and autoradiography of the gel. Most of these were bands that were present in the track of products from the normal DNA but that were absent in the paired tumour track. Two further Alumorphs were isolated that represented 1) a mobility shift between tumour and normal lanes and 2) bands that exhibited more intensity in the tumour lane compared with their constituent normal sample (Fig. 2B). All of these PCR products were labelled and were used as probes in hybridisations with Southern blots that carried tumour and normal pairs to demonstrate whether the fragment would hybridise to unique single copy sequences in the genome. The fragments were used to probe high-density gridded P1, PAC, and cosmid clones to obtain larger fragments, as summarised in

Table 3 for fluorescence in situ hybridisation (FISH) and other analysis.

Analysis of Alumorph AB1

Five of the 16 PCR products hybridised to discrete bands on Southern blots, two of which gave polymorphic band patterns (see Table 3). One particular fragment (Alumorph AB1) was observed originally as a 250 bp fragment that was absent from the fingerprint of tumour DNA compared with its paired normal DNA in one case. On elution from the gel, it was reamplified by using the same primer that was used to generate it and used as a probe in hybridisation with Southern blots. In ten cases that were digested with *Msp*I, the fragment hybridised to two small polymorphic bands of either 200 or 300 bp in four cases, and a constant band appeared at around 100 bp in all ten cases (Fig. 3). Interestingly, the two upper polymorphic bands were lost in the tumour samples of two of these paired cases. The fragment also hybridised to a 15 kb band on a *Hin*dIII Southern blot of tumour/normal pairs (not shown). After cloning the PCR product into Bluescript, it was sequenced and was found to contain five copies of a tetranucleotide repeat sequence, 5′-ATGG-3′, but variation in the copy number of this repeat is not responsible for the polymorphic nature of the fragment (data not shown). A similarity to ESTs identified as up-regulated in pancreatic cancer by differential hybridisation of cDNA libraries (Gress et al., 1994) was found when the sequence was compared with the EMBL database (Table 4). FISH was performed on metaphase chromosomes that were prepared from the peripheral blood leukocytes of a normal donor by using a YAC clone and two of the P1 clones to assign them to band p22 of chromosome 2 (not shown).

Once large genomic clones had been identified, hybridising fragments were then isolated, subcloned, and used as probes in Northern hybridisations. A 12 kb fragment from a P1 clone G09101 specific for the Alumorph AB1 identified a transcript at approximately 4.0 kb on filters carrying total RNA from pancreatic tumour cell lines (Fig. 4). The transcript appeared to be fairly restricted in expression pattern, because there was no evidence of a signal in three samples of RNA from normal pancreas nor in any of the seven other sources on a multiple human tissue blot. We then challenged gridded cDNA libraries made from foetal brain and a pancreatic tumour cell line with the 12 kb fragment isolating five overlapping clones ranging in size from 1.5 to 5 kb. In situ hybridisation was used to extend our studies of the expression of this

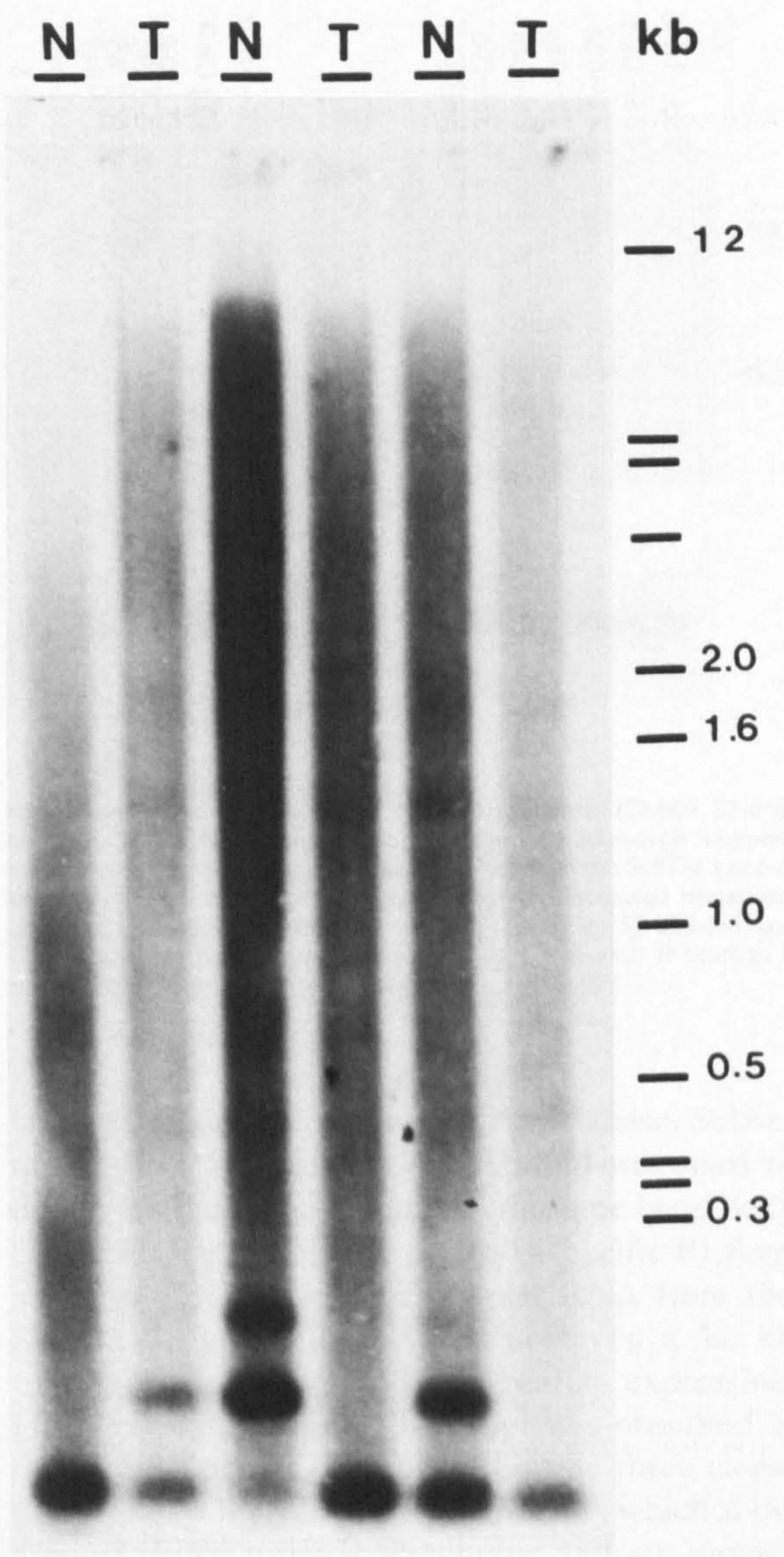


Figure 3. Southern hybridisation of *Msp*I-digested DNA from three individuals probed with the Alumorph fragment AB1. There are three hybridising bands, one constant and two polymorphic, which show loss in the tumour track of two of the cases. The middle hybridising band was present in all of the cases examined; however, the larger polymorphic band was present in only 1 of 10 cases.

sequence. The transcript (currently called *ART1* for Alu-PCR-recovered transcript 1) has been studied in tissue in situ hybridisation experiments (TISH) and in the pancreas appears to be expressed almost exclusively in ductal adenocarcinoma, with no expression in nonneoplastic epithelium and stroma of the pancreas (Fig. 5A). *ART1* expression in normal tissues was restricted to the islets of Langerhans, the surface epithelial cells of the small intestine, and the transitional epithelium of the urinary bladder (Fig. 5B,C). Detailed sequence analysis and studies to elucidate the nature of the original

TABLE 4. Sequence Similarities of Alumorph Fragments

		********** ***
ART1 vs EST over-expressed in pancreatic cancer (U46442, U46445)	17	GAATTCCCGGGTCGACCCACGCGTCGGAA 45
	1	GAATTCCCGGGTCGACCCACGCGTCCGAA 29
		********** * ********** **********
Met 6 vs EST expressed in human foreskin melanocytes (N25175)	21	TAACATGATTGGAGAGTTTCAGGGGAAAGGTTTAGA-TTTCTGAAACTGTATTTGGTATGTGACTCAATGTGGT ATTTCAGTCT 83
	2	TAACATGATTGGAGAGTTTCAGGGGAAAGGTTTACATTTTCTGAAACTGTATTTGGTATGTGACT CAATGTGGTATTTCAGTCT 65
		********** ********** ********** **
		TGTTAGTCACTTACATGACTGACGTTTGCAAGGATTTATTGCCAAGT--AATTTGACCAGAGTGCACTGAGAAT AGCTAC-TA-
		TGTTAGTCACTTACATGACTGACGTTTGCAAGGATTTATTGCCAAGTAAAATTTGACCAGAGTGCACTGAGAAT AGCTACATAA
		**** * ********** * **
		-GGGAAA-CTCTCAAAATTCCTTCTGTACTTT 213
		GGGGANATCTCTCAAAATTCCTTCTGTTCATT 201
		********** *******
IP-1 vs rat POU factor Brn-5 (L23204)	111	CCCAGGGCCAGGTGATGGCAACAC 134
	929	CCCAGGGCCAGGTGATCGCAACAC 952
		** ***** * ******** **** ** ***** * ****
IP-1 vs X. laevis X1POU3 (X64835)	88	GTCCACCCGGATCACGGCATGCACCCAGGGCCAGGTG-ATGG 128
	386	GTGCACCCTGCTCACGGCAACCACCATGGACCAGGGGCATGG 427

Database accession numbers are shown in brackets. *Represents sequence match; N represents ambiguous sequence.

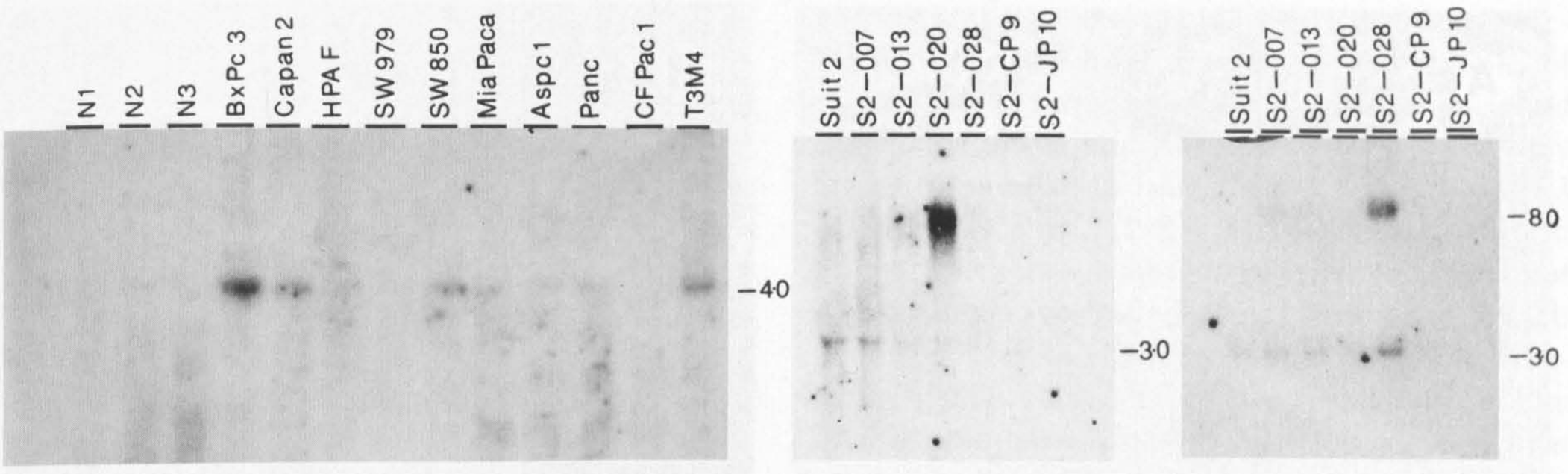


Figure 4. **A:** Northern analysis of *ART1* expression levels using total RNA from three primary cultures of normal pancreas tissue (N1–N3) and from ten pancreatic tumour cell lines, as indicated. The transcript appears to migrate at approximately 4.5 kbp and demonstrates varied expression, with little or no expression in the normal samples and up to five- or tenfold greater expression in cell lines BxPc3, Capan2, and T3M4. **B:** Northern analysis using a Met-L-specific P1 fragment exhibiting expression in the SUIT-2 parent tumour cell line and the clonal derivatives, which show increased metastatic potential (S2-007, S2-013, and S2-020). **C:** Northern analysis using the AB4 Alumorph fragment exhibiting faint expression of a 3.0 kbp transcript in the SUIT-2 tumour cell line and its clonal derivatives, which have increased metastatic potential, and more abundant expression in the cell line S2-028. In B and C, the lower autoradiographs illustrate the signal with β-actin as a control for amounts of RNA on the gel.

genomic loss and its relationship to the up-regulation of a transcript in a tumour-associated fashion are in progress and will be the subject of a separate report.

Analysis of Fragments from SUIT-2 Metastatic Cell Lines

A fragment termed Met-6 (the sixth Alumorph identified from a series of cell lines that differed in metastatic potential) demonstrated two polymorphic bands on Southern blots of *Hin*dIII-digested DNA samples. However, no clear losses were observed between DNAs from either the cell lines or the tumour/normal paired cases. We used Southern blot analysis to the ICRF somatic cell hybrid panel to localise this fragment to chromosome arm 11p, a result that was confirmed later with the subsequent isolation of a cosmid from a flow-sorted chromosome 11 library, which mapped to 11p12-p13 by FISH (data not shown). Hybridisation of the Met-6 fragment to Northern blots of pancreatic cancer RNA detected a transcript of approximately 10 kb.

Two further metastatic cell line-specific Alumorphs, Met-L and Met-U, were isolated as fragments that exhibited, respectively, altered mobility and loss in some of the SUIT-2 derived cell lines. Met-L was a fragment of 235 bp and, upon cloning and sequence analysis, revealed no significant similarity to sequences in the EMBL database. Subsequently, a P1 clone was isolated and was used to demonstrate localisation to chromosome band 8q24 by FISH. We hybridised a Met-L-specific P1 fragment to a Northern blot of total RNA from the SUIT-2-derived cell lines and observed a 3.0 kb transcript that exhibited differential expression. High expression of the transcript was observed in the SUIT-2 parent cell line and in the three clonal derivatives S2-007, S2-013, and S2-020, which show a poorly differentiated phenotype and are moderately to highly metastatic (S2-007 > S2-020). The Alumorph fragment Met-U was approximately 295 bp and had no similarity to sequences in the EMBL database.

Analyses of Alumorph Fragments IP-1 and AB4

Alumorph IP-1 was a 273 bp fragment that was observed as being lost in two pancreatic tumour lanes compared with their constituent normal lanes. The fragment was used as a probe in Southern hybridisations but did not detect any discrete bands. Sequence analysis revealed significant similarity to the human, mouse, and *Xenopus* sequences encoding the POU domain. The isolation of two IP-1-specific cosmid clones enabled us to localise the fragment to chromosome 17p13 by using FISH analysis (Fig. 6), and PCR analysis refined this to 17p13.3 in the region of a contig containing a

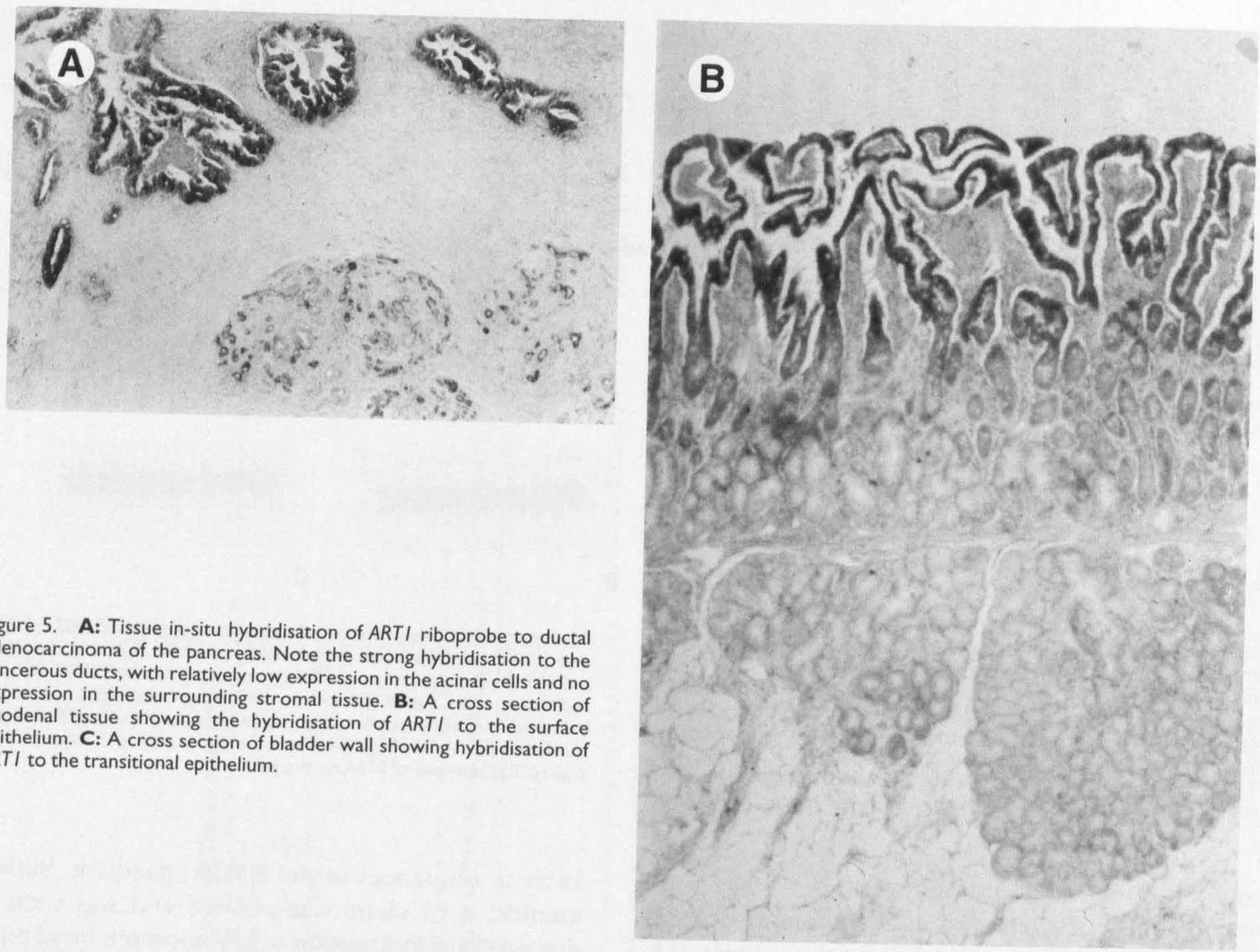


Figure 5. **A:** Tissue in-situ hybridisation of *ART1* riboprobe to ductal adenocarcinoma of the pancreas. Note the strong hybridisation to the cancerous ducts, with relatively low expression in the acinar cells and no expression in the surrounding stromal tissue. **B:** A cross section of duodenal tissue showing the hybridisation of *ART1* to the surface epithelium. **C:** A cross section of bladder wall showing hybridisation of *ART1* to the transitional epithelium.

Fig. 5.

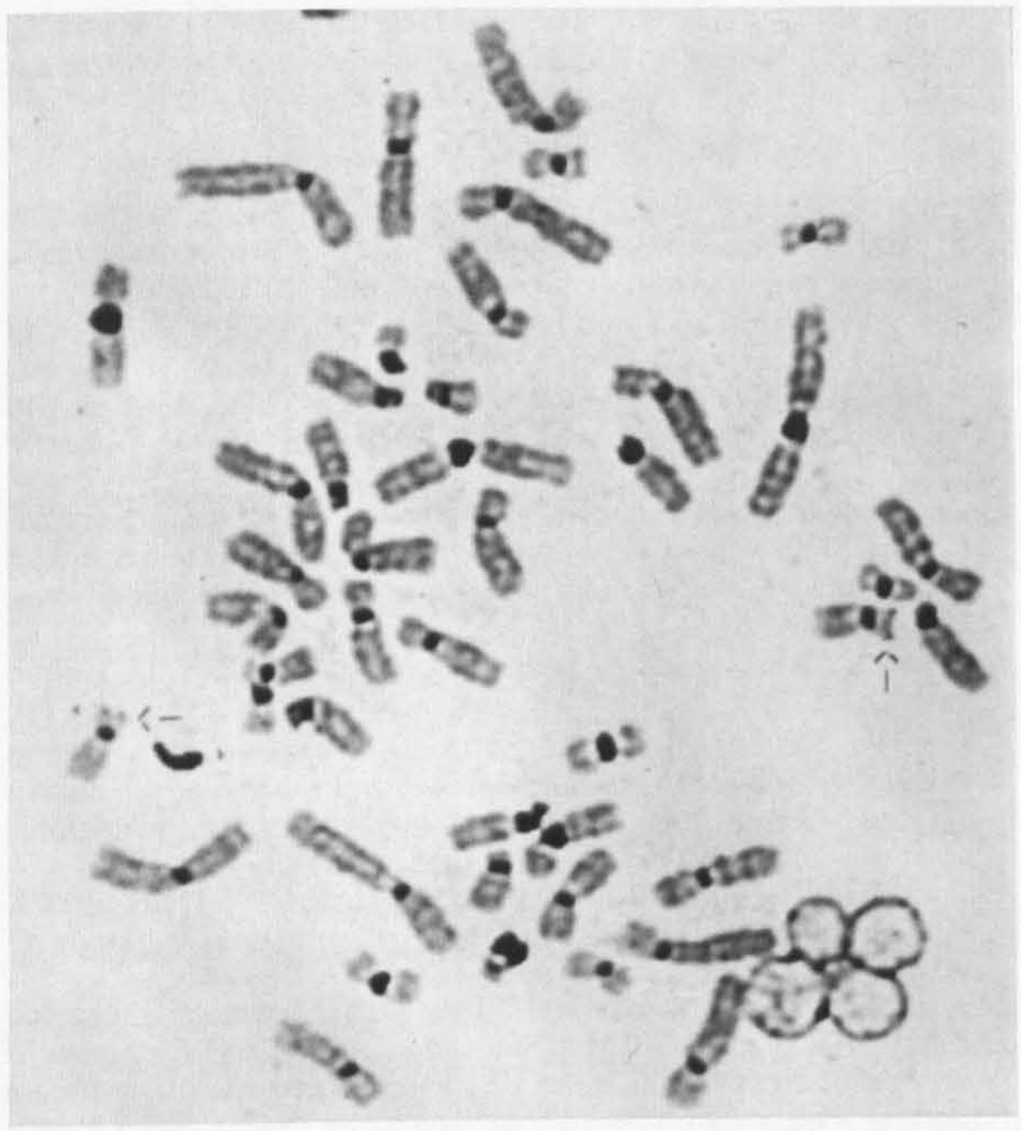


Figure 6. Fluorescence in situ hybridisation (FISH) on normal metaphase chromosomes demonstrating hybridisation of cosmid IP-cB specific for the Alumorph IP-1 to chromosome band 17p13.

potential tumour suppressor gene involved in breast cancer. This is an interesting fragment not only because of its sequence similarity but also because the reported frequency of allelic loss at 17p in pancreatic cancer approaches 100% (Hahn et al., 1995). A cosmid fragment was used to isolate from a pancreatic tumour library two cDNA clones (0.7 and 1.6 kb) that have been partially sequenced. These cDNAs and the original IP-1 Alumorph fragment identify a 2 kb transcript expressed at low levels in pancreatic tumour cell lines.

A sixth Alumorph, AB4, was also identified as a fragment lost in the tumour fingerprint profile compared with its paired normal sample. The fragment is the largest to be isolated thus far (approximately 750 bp) and has been used directly as a probe in Southern and Northern hybridisations. AB4 identifies three *Hin*dIII bands on Southern blots, but no losses between tumour and normal lanes are observed. A transcript band of 8.0 kb (Fig. 4) is seen on Northern filters of the SUIT-2-derived cell lines with expression more abundant in the papillotubular adenocarcinoma tumour cell line S2-028 (a subline of low metastatic potential).

DISCUSSION

We describe here a technique that allows scanning of the genome for genetic rearrangements and generation of PCR-generated tags for use as probes to isolate genes associated with these rearrangements. There are approximately 1,000,000 copies of the Alu repeat sequence in the human genome, and a combination of approaches has shown that they are distributed at an average spacing distance of 3 kb (Moyzis et al., 1989). However, it has been found that there is a 50-fold underrepresentation of Alu sequences in centric heterochromatin, which results in a relatively higher density in more interesting gene-rich regions (Moyzis et al., 1989). The highly ubiquitous distribution of the Alu repeat means that the coverage of the genome afforded by Alu sequence-primed amplification could potentially be of high resolution, and combinations of primers homologous to different regions of the Alu sequence and various restriction digests of the template DNAs can be used to vary the complexity of the DNA fingerprint generated. The fractionation of the template DNA by restriction digestion in our approach is analogous to the preparation of "representations" of the tester DNA necessary in RDA. In both approaches, it is designed to simplify the analysis of paired genomes by taking defined subsets for comparison. In Table 2, it can be seen that the initial screen was on a total of 189 bands that amplified consistently from various samples of normal human genomic DNA. Sixteen of those bands showed unequivocal differences when 18 paired samples of normal and tumour DNA were compared, and some of the bands were aberrant in more than one case (e.g., Alumorph IP-1). Hence, abnormalities were detectable in about 8.5% of the genomic fragments that were sampled in this analysis. Clearly, only a fraction of the genome is sampled in each Alu primer/restriction digest combination, and some are more useful than others (e.g., Cotter IV/*Alu*I combination). Because we use a single primer in the PCR reaction, only sequences lying between repeats in inverted orientation to one another will be amplified; because their orientation is essentially random, this could result in around 30% inter-Alu sequences being potential templates in each case. Like the alternative genome scanning technique of RDA, there is a bias towards the generation of smaller fragments, partly due to the predigestion of the template DNA and partly because of the kinetics of PCR amplification. However, these are tags that enable the subsequent isolation of larger rearranged fragments by hybridisation.

Our technique is related to arbitrarily primed PCR (Welsh and McClelland, 1990; Williams et al., 1990), which has also been used to detect and isolate tumour-specific allelic losses and gains (Pei-

nado et al., 1992) and to demonstrate the replication error phenotype (RER) associated with abnormalities of DNA mismatch repair systems (Ionov et al., 1993). Because Alu repeats contain runs of deoxyadenosine residues, they are often involved as targets for such ubiquitous somatic mutations, so that about one-third has deletion mutations in RER^+ cancers (Ionov et al., 1993). Interestingly, Alu repeats can even be involved in the genesis of inactivating mutations; for instance, one of the most common germline mutations in the DNA mismatch repair gene *MLH1* is a 3.5 kb genomic deletion that appears to result from Alu-mediated recombination (Nyström-Lahti et al., 1995). Alu elements are also involved in the production of oncogenes as a result of chromosomal translocation events joining the *BCR* and *ABL* loci and the *LCK* and *TCRB* loci (Tycko et al., 1991; Chissoe et al., 1995). Hence, our approach of using Alu sequences as the priming sites for the amplification of fingerprints from genomic DNA may have an advantage over arbitrarily primed PCR if such regions are particularly affected by the cancer-associated rearrangements we seek. However, it is likely that most rearrangements at repeat sequences are not significant to tumourigenesis. It may be possible to exploit other high-abundance repetitive elements in the human genome to similar ends, such as the L1 and MER families (Skowronski et al., 1988; Jurka, 1989).

Our results with the approach of using Alu-PCR fingerprinting suggest that it may be useful for the isolation of cancer-associated genetic events. For instance, the *ART1* gene, which was isolated following the identification in the DNA fingerprint of a deletion of a flanking sequence, shows up-regulated expression specifically in pancreatic cancer both in vitro and in vivo. The preliminary investigation of additional DNA fragments identified by this technique suggests that they are derived from chromosomal regions of significant allelic imbalance in pancreatic cancer (17p13 and 11p12-13). Moreover, two of these (Met-L and AB4) hybridise to transcripts that are expressed differentially by pancreatic tumour cells that show distinct morphological properties and varied metastatic potential. Further detailed characterisation of these clones is now underway.

REFERENCES

Almoguera C, Shibata D, Forrester R, Martin J, Perucho M (1988) Most human carcinomas of exocrine pancreas contain mutant c-K-*ras* genes. Cell 53:549–554.

Bardi G, Johansson B, Pandis N, Mandahl N, Bak JE, Andren SA, Mitelman F, Heim S (1993) Karyotypic abnormalities in tumours of the pancreas. Br J Cancer 67:1106–1112.

Barton CM, Staddon SL, Hughes CM, Hall PA, O'Sullivan C, Kloppel G, Theis B, Russell RC, Neoptolemos JP, Williamson RCN, Lane DP, Lemoine NR (1991) Abnormalities of the p53 tumour suppressor gene in human pancreatic cancer. Br J Cancer 64:1076–1082.

Bartsch D, Shevlin DW, Tung WS, Kisker O, Wells SA, Goodfellow PJ (1995) Frequent mutations of CDKN2 in primary pancreatic adenocarcinomas. Genes Chromosom Cancer 14:189–195.

Caldas C, Hahn SA, da Costa LT, Redston MS, Schutte M, Seymour AB, Weinstein CL, Hruban RH, Yeo CJ, Kern SE (1994a) Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. Nature Genet 8: 27–31.

Caldas C, Hahn SA, Hruban RH, Redston MS, Yeo CJ, Kern SE (1994b) Detection of K-*ras* mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. Cancer Res 54:3568–3573.

Chissoe SL, Bodenteich A, Wang YF, Wang YP, Burian D, Clifton SW, Crabtree J, Freeman A, Iyer K, Li JA, Ma YC, McLaury HJ, Pan HQ, Sarhan OH, Toth S, Wang ZL, Zhang GZ, Heisterkamp N, Groffen J, Roe BA (1995) Sequence and analysis of the human *abl* gene, the *bcr* gene, and regions involved in the Philadelphia chromosomal translocation. Genomics 27:67–82.

Daniels GR, Fox GM, Loewensteiner D, Schmid CW, Deininger PL (1983) Species-specific homogeneity of primate Alu family of repeated sequences. Nucleic Acids Res 11:7579–7593.

DiGiuseppe JA, Hruban RH, Offerhaus GJA, Clement MJ, van den Berg FM, Cameron JL, van Mansfeld ADM (1994) Detection of K-*ras* mutations in mucinous pancreatic duct hyperplasia from a patient with a family history of pancreatic cancer. Am J Pathol 144:889–895.

Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267.

Gress TM, Muller-Pillasch F, Adler G, Lehrach H (1994) European Pancreatic Cancer Reference Library System (EPCRLS). Eur J Cancer 30A:1391–1394.

Griffin CA, Hruban RH, Morsberger LA, Ellingham T, Long PP, Jaffee EM, Hauda KM, Bohlander SK, Yeo CJ (1995) Consistent chromosome abnormalities in adenocarcinoma of the pancreas. Cancer Res 55:2394–2399.

Guronova L, Johansson B, Dawiskiba S, Andren-Sandberg A, Jin YS, Mandahl N, Heim S, Mitelman F (1995) Massive cytogenetic heterogeneity in a pancreatic carcinoma—Fifty-four karyotypically unrelated clones. Genes Chromosom Cancer 14:259–266.

Hahn SA, Seymour AB, Hoque S, Schutte M, da Costa LT, Redston MS, Caldas C, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE (1995) Allelotype of pancreatic cancer using xenograft enrichment. Cancer Res 55:4670–4675.

Hahn SA, Schutte M, Hoque ATMS, Moskaluk CA, Dacosta LT, Rozenblum E, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE (1996) DPC4, a candidate tumor-suppressor gene at human-chromosome 18q21.1. Science 271:350–353.

Holton TA, Graham MW (1990) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res 19:1156.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colorectal carcinogenesis. Nature 363:558–561.

Iwamura T, Katsuki T, Ide K (1987) Establishment and characterization of a human pancreatic cancer cell line (SUIT-2) producing carcinoembryonic antigen and carbohydrate antigen 19–9. Jpn J Cancer Res 78:54–62.

Johansson B, Bardi G, Pandis N, Guronova L, Backman PL, Mandahl N, Dawiskiba S, Andren-Sandberg A, Heim S, Mitelman F (1994) Karyotypic pattern of pancreatic adenocarcinomas correlates with survival and tumor grade. Int J Cancer 58:8–13.

Jurka J (1989) Novel families of interspersed repetitive elements from the human genome. Nucleic Acids Res 18:137–141.

Kalthoff H, Schmiegel W, Roeder C, Kasche D, Schmidt A, Lauer G, Thiele HG, Honold G, Pantel K, Riethmuller G, Scherer E, Maurer J, Maacke H, Deppert W (1993) p53 and K-*ras* alterations in pancreatic epithelial cell lesions. Oncogene 8:289–298.

Lehrach H, Drmanac R, Hoheisel J, Larin Z, Lennon G, Monaco AP, Nizetic D, Zehetner G, Poustka A-M (1990) Hybridisation fingerprinting in genome mapping and sequencing. In Davies KE, Tilghman SM (eds): Genome Analysis, Vol 1: Genetic and Physical Mapping. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp 39–81.

nado et al., 1992) and to demonstrate the replication error phenotype (RER) associated with abnormalities of DNA mismatch repair systems (Ionov et al., 1993). Because Alu repeats contain runs of deoxyadenosine residues, they are often involved as targets for such ubiquitous somatic mutations, so that about one-third has deletion mutations in RER^+ cancers (Ionov et al., 1993). Interestingly, Alu repeats can even be involved in the genesis of inactivating mutations; for instance, one of the most common germline mutations in the DNA mismatch repair gene *MLH1* is a 3.5 kb genomic deletion that appears to result from Alu-mediated recombination (Nyström-Lahti et al., 1995). Alu elements are also involved in the production of oncogenes as a result of chromosomal translocation events joining the *BCR* and *ABL* loci and the *LCK* and *TCRB* loci (Tycko et al., 1991; Chissoe et al., 1995). Hence, our approach of using Alu sequences as the priming sites for the amplification of fingerprints from genomic DNA may have an advantage over arbitrarily primed PCR if such regions are particularly affected by the cancer-associated rearrangements we seek. However, it is likely that most rearrangements at repeat sequences are not significant to tumourigenesis. It may be possible to exploit other high-abundance repetitive elements in the human genome to similar ends, such as the L1 and MER families (Skowronski et al., 1988; Jurka, 1989).

Our results with the approach of using Alu-PCR fingerprinting suggest that it may be useful for the isolation of cancer-associated genetic events. For instance, the *ART1* gene, which was isolated following the identification in the DNA fingerprint of a deletion of a flanking sequence, shows up-regulated expression specifically in pancreatic cancer both in vitro and in vivo. The preliminary investigation of additional DNA fragments identified by this technique suggests that they are derived from chromosomal regions of significant allelic imbalance in pancreatic cancer (17p13 and 11p12-13). Moreover, two of these (Met-L and AB4) hybridise to transcripts that are expressed differentially by pancreatic tumour cells that show distinct morphological properties and varied metastatic potential. Further detailed characterisation of these clones is now underway.

REFERENCES

Almoguera C, Shibata D, Forrester R, Martin J, Perucho M (1988) Most human carcinomas of exocrine pancreas contain mutant c-K-*ras* genes. Cell 53:549–554.

Bardi G, Johansson B, Pandis N, Mandahl N, Bak JE, Andren SA, Mitelman F, Heim S (1993) Karyotypic abnormalities in tumours of the pancreas. Br J Cancer 67:1106–1112.

Barton CM, Staddon SL, Hughes CM, Hall PA, O'Sullivan C, Kloppel G, Theis B, Russell RC, Neoptolemos JP, Williamson RCN, Lane DP, Lemoine NR (1991) Abnormalities of the p53 tumour suppressor gene in human pancreatic cancer. Br J Cancer 64:1076–1082.

Bartsch D, Shevlin DW, Tung WS, Kisker O, Wells SA, Goodfellow PJ (1995) Frequent mutations of CDKN2 in primary pancreatic adenocarcinomas. Genes Chromosom Cancer 14:189–195.

Caldas C, Hahn SA, da Costa LT, Redston MS, Schutte M, Seymour AB, Weinstein CL, Hruban RH, Yeo CJ, Kern SE (1994a) Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. Nature Genet 8: 27–31.

Caldas C, Hahn SA, Hruban RH, Redston MS, Yeo CJ, Kern SE (1994b) Detection of K-*ras* mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. Cancer Res 54:3568–3573.

Chissoe SL, Bodenteich A, Wang YF, Wang YP, Burian D, Clifton SW, Crabtree J, Freeman A, Iyer K, Li JA, Ma YC, McLaury HJ, Pan HQ, Sarhan OH, Toth S, Wang ZL, Zhang GZ, Heisterkamp N, Groffen J, Roe BA (1995) Sequence and analysis of the human *abl* gene, the *bcr* gene, and regions involved in the Philadelphia chromosomal translocation. Genomics 27:67–82.

Daniels GR, Fox GM, Loewensteiner D, Schmid CW, Deininger PL (1983) Species-specific homogeneity of primate Alu family of repeated sequences. Nucleic Acids Res 11:7579–7593.

DiGiuseppe JA, Hruban RH, Offerhaus GJA, Clement MJ, van den Berg FM, Cameron JL, van Mansfeld ADM (1994) Detection of K-*ras* mutations in mucinous pancreatic duct hyperplasia from a patient with a family history of pancreatic cancer. Am J Pathol 144:889–895.

Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267.

Gress TM, Muller-Pillasch F, Adler G, Lehrach H (1994) European Pancreatic Cancer Reference Library System (EPCRLS). Eur J Cancer 30A:1391–1394.

Griffin CA, Hruban RH, Morsberger LA, Ellingham T, Long PP, Jaffee EM, Hauda KM, Bohlander SK, Yeo CJ (1995) Consistent chromosome abnormalities in adenocarcinoma of the pancreas. Cancer Res 55:2394–2399.

Guronova L, Johansson B, Dawiskiba S, Andren-Sandberg A, Jin YS, Mandahl N, Heim S, Mitelman F (1995) Massive cytogenetic heterogeneity in a pancreatic carcinoma—Fifty-four karyotypically unrelated clones. Genes Chromosom Cancer 14:259–266.

Hahn SA, Seymour AB, Hoque S, Schutte M, da Costa LT, Redston MS, Caldas C, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE (1995) Allelotype of pancreatic cancer using xenograft enrichment. Cancer Res 55:4670–4675.

Hahn SA, Schutte M, Hoque ATMS, Moskaluk CA, Dacosta LT, Rozenblum E, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE (1996) DPC4, a candidate tumor-suppressor gene at human-chromosome 18q21.1. Science 271:350–353.

Holton TA, Graham MW (1990) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res 19:1156.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colorectal carcinogenesis. Nature 363:558–561.

Iwamura T, Katsuki T, Ide K (1987) Establishment and characterization of a human pancreatic cancer cell line (SUIT-2) producing carcinoembryonic antigen and carbohydrate antigen 19–9. Jpn J Cancer Res 78:54–62.

Johansson B, Bardi G, Pandis N, Guronova L, Backman PL, Mandahl N, Dawiskiba S, Andren-Sandberg A, Heim S, Mitelman F (1994) Karyotypic pattern of pancreatic adenocarcinomas correlates with survival and tumor grade. Int J Cancer 58:8–13.

Jurka J (1989) Novel families of interspersed repetitive elements from the human genome. Nucleic Acids Res 18:137–141.

Kalthoff H, Schmiegel W, Roeder C, Kasche D, Schmidt A, Lauer G, Thiele HG, Honold G, Pantel K, Riethmuller G, Scherer E, Maurer J, Maacke H, Deppert W (1993) p53 and K-*ras* alterations in pancreatic epithelial cell lesions. Oncogene 8:289–298.

Lehrach H, Drmanac R, Hoheisel J, Larin Z, Lennon G, Monaco AP, Nizetic D, Zehetner G, Poustka A-M (1990) Hybridisation fingerprinting in genome mapping and sequencing. In Davies KE, Tilghman SM (eds): Genome Analysis, Vol 1: Genetic and Physical Mapping. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp 39–81.